Astérisque

JAMES SIMONS
DENNIS SULLIVAN

Structured bundles define differential *K*-theory

Astérisque, tome 321 (2008), p. 1-3
<http://www.numdam.org/item?id=AST_2008__321__1_0>

© Société mathématique de France, 2008, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (http://smf4.emath.fr/Publications/Asterisque/) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM
Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

Astérisque
321, 2008, p. 1–3

STRUCTURED BUNDLES DEFINE DIFFERENTIAL K-THEORY

by

James Simons & Dennis Sullivan

Abstract. — Complex bundles with connection up to isomorphism form a semigroup under Whitney sum which is far from being a group. We define a new equivalence relation (structured equivalence) so that stable isomorphism classes up to structured equivalence form a group which is describable in terms of the Chern character form plus some finite type invariants from algebraic topology. The elements in this group also satisfy two further somewhat contradictory properties: a locality or gluing property and an integrality property. There is interest in using these objects as pre-quantum fields in gauge theory and M-theory.

***Résumé* (Les fibrés structurés définissent la K-théorie différentielle).** — Les fibrés complexes à connexion forment, à isomorphisme près, un semi-groupe sous la somme de Whitney qui est loin d'être un groupe. Nous définissons une nouvelle relation d'équivalence (l'équivalence structurée) de manière à ce que les classes d'isomorphismes stable, à équivalence structurée près, forment un groupe qui puisse être décrit en termes de forme de caractère de Chern et de quelques invariants de type fini de la topologie algébrique. Les éléments de ce groupe satisfont également à deux propriétés en quelque sorte contradictoires : une propriété de localité ou de *gluing* et une propriété d'intégralité. Il semble intéressant d'utiliser ces objets en tant que champs pré-quantiques en théorie de gauge et en M-théorie.

Let M be the category whose objects are smooth manifolds and whose morphisms are smooth maps. We assume the manifolds are either compact manifolds possibly with boundary or diffeomorphic to those obtained from these by deleting some or all of the boundary components.

Let $K^\wedge$ denote the contravariant functor on M to abelian groups defined by equivalence classes of pairs (V, A) where V is a complex vector bundle and A is a connection on V. The equivalence relation is generated by stable isomorphisms of bundles with connection and by structured equivalence: namely any deformation of a connection A on a fixed bundle along any smooth path of connections so that the associated odd Chern Simons form is exact. Recall the exterior d of the Chern Simons form of a path

2000 Mathematics Subject Classification. — 57N65, 55N15, 19E08, 58J28, 14D21.
Key words and phrases. — Structured bundles, connections, Chern Simons form, differential K-theory.

© Astérisque 321, SMF 2008

of connections measures the change in the Chern Character form at the endpoints of the path.

We want to compute $K^\wedge$. Note an equivalence class in $K^\wedge$ has a precise total even form representing the Chern character plus some information related to a total odd cohomology class represented by Chern Simons forms which are closed but not exact.

Let $\mathrm{ch}(K)$ denote the contravariant functor on M to abelian groups defined by considering pairs $([V], C)$ where V represents an element of the K-theory of complex vector bundles and C is a total closed complex valued even dimensional form so that C represents the Chern character of V in rational cohomology.

We have the obvious map from $K^\wedge$ to $\mathrm{ch}(K)$ which assigns to the pair (V, A) of bundle with connection the pair $([V], C)$ where $[V]$ is the class of the stable vector bundle V in Atiyah's K-theory and C is the differential form defined by the Chern Weil curvature construction representing the total Chern character.

Let Torus be the functor on M to abelian complex Lie groups given by the odd cohomology with complex coefficients modulo the sublattice defined by considering all maps into G , the union over n of the n-dimensional complex linear groups, and by pulling back the desuspended Chern character class. This class is defined universally at level n defining G by desuspending the Chern character class of the bundle on suspension G defined using the identity map of G as a gluing function.

Theorem 1. — *The homomorphism from $K^\wedge$ to $\mathrm{ch}(K)$ is onto.*

The kernel of the homomorphism is the abelian complex Lie group Torus. We have the natural short exact sequence:

$$0 \to \text{Torus} \to K^\wedge \to \mathrm{ch}(K) \to 0. \tag{1}$$

Let k denote the kernel of the natural map from $K^\wedge$ to K, namely $(V, C) \to [V]$. Then from sequence (1), k maps with kernel Torus onto exact total even forms.

Let O = total odd forms modulo all closed forms in the cohomology classes of the sublattice above defining Torus. Then O maps via d with kernel Torus onto exact total even forms. The construction of $K^\wedge$ shows the kernel k is naturally isomorphic to O. In the detailed paper [1], O is denoted $\wedge^{odd}(X)/\wedge_G(X)$.

Theorem 2. — *There is the natural short exact sequence*

$$0 \to O \to K^\wedge \to K \to 0. \tag{2}$$

One may also show from the construction:

Theorem 3. — *$K^\wedge$ satisfies the Mayer-Vietoris property: if X is A union B with intersection C then given two elements a in $K^\wedge(A)$ and b in $K^\wedge(B)$ which restrict to the same element c in $K^\wedge(C)$, then there is an element x in $K^\wedge(X)$ which restricts to a and to b respectively.*

Consider E = all total even forms in the cohomology classes of the Chern characters of complex vector bundles. By Theorem 1 the map $K^\wedge \to E$ is surjective. Now consider k', the kernel of this map from $K^\wedge$ to E. Since $K^\wedge$ satisfies the Mayer-Vietoris property so does this kernel k'. One can show also that k' is a homotopy functor. Thus by Brown's representability theorem k' is represented by homotopy classes of maps into some space. Using this , the sequences above and side condition 1 in Remark 1 below leads to

Theorem 4. — *The kernel of the surjection of $K^\wedge$ onto E is naturally isomorphic to K-theory with coefficients in C/Z. Let us denote the latter by $K(C/Z)$. Then we have the natural short exact sequence:*

$$0 \to K(C/Z) \to K^\wedge \to E \to 0. \tag{3}$$

Now $K^\wedge$ is not a homotopy functor, but the change produced by an infinitesimal deformation v of a map can be computed. This change u is in $O =$ the kernel of $(K^\wedge \to K)$ because K is a homotopy functor. We know that du is the lie derivative of the Chern character form. So the following is natural and indeed true for $K^\wedge$:

Theorem 5. — *The change in $f^*(x)$ for x in $K^\wedge$ by an infinitesimal deformation v of a map f is obtained by contracting the Chern form of x byv and projecting it to O inside $K^\wedge$.*

Remark 1. — *We have omitted two natural side conditions in the statements of Theorems 2 and 4 which should be noted.*

1. *The composition $K(C/Z) \to K^\wedge \to K$ using* (2) *and* (3) *is the Bockstein map in the Bockstein exact sequence for K-theory.*
2. *The composition $O \to K^\wedge \to E$ using* (2) *and* (3) *is exterior d.*

Conjecture. — *There is at most one functor $K^\wedge$ up to natural equivalence satisfying Theorems 1, 2, 3, 4 and 5 and the side conditions 1 and 2 in Remark 1.*

The presence of the homotopy property expressed by Theorem 5 in the conjecture above was inspired by conversations with Moritz Wiethaup. This homotopy property was not needed in our axioms characterizing ordinary differential cohomology [**2**]. The details of the proofs of the results here will appear soon [**1**]. We close by expressing on this occasion our appreciation of and admiration for the geometer Jean Pierre Bourguignon.

References

[1] J. SIMONS & D. SULLIVAN – "Axiomatic characterization of ordinary differential cohomology", *J. Topol.* **1** (2008), p. 45–56.

[2] ———, "Structured bundles and differential K-theory", preprint arXiv:0810.4935 (submitted to *J. Topol.*).

J. SIMONS, Renaissance Technologies, New York, USA

D. SULLIVAN, Mathematics Department, Stony Brook University, Stony Brook, NY 11790, USA
E-mail : dennis@math.sunysb.edu

Astérisque

NIGEL HITCHIN

Einstein metrics and magnetic monopoles

Astérisque, tome 321 (2008), p. 5-29

<http://www.numdam.org/item?id=AST_2008__321__5_0>

© Société mathématique de France, 2008, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (http://smf4.emath.fr/Publications/Asterisque/) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

Astérisque
321, 2008, p. 5–29

EINSTEIN METRICS AND MAGNETIC MONOPOLES

by

Nigel Hitchin

For Jean Pierre Bourguignon on his 60th birthday

Abstract. — We investigate the geometry of the moduli space of centred magnetic monopoles on hyperbolic three-space, and derive using twistor methods some (incomplete) quaternionic Kähler metrics of positive scalar curvature. For the group SU(2) these have an orbifold compactification but we show that this is not the case for SU(3).

***Résumé* (Métriques d'Einstein et monopoles magnétiques).** — Nous étudions la géométrie des espaces de modules des monopoles maghétiques sur le 3-espace hyperbolique et nous en dérivons quelques métriques kähleriennes quaternioniques (incomplètes) de courbure scalaire positive, en utilisant des méthodes *twistor*. Celles-ci ont une compactification orbifolde pour le groupe SU(2) et nous montrons qu'il n'en est rien dans le cas du groupe SU(3).

1. Introduction

Over 20 years ago Jean Pierre Bourguignon and I were part of the team helping Arthur Besse to produce a state-of-the-art book on Einstein manifolds [**3**]. As might have been expected, the subject proved to be a moving target, and the contributors had to quickly assemble a number of appendices to cover material that came to light after all the initial planning. The last sentence of the final appendix refers to: "hyperkählerian metrics on finite dimensional moduli spaces", and so it seems appropriate to write here about some of the results which have followed on from that, and some questions that remain to be answered.

There is by now a range of gauge-theoretical moduli spaces which have natural hyperkähler metrics: the moduli space of instantons on $\mathbf{R}^4$ or the 4-torus or a K3

2000 Mathematics Subject Classification. — 53C28, 53C26, 51N30.
Key words and phrases. — Hyperkähler, quaternionic Kähler, moduli space, monopole, Bogomolny equations, spectral curve.

© Astérisque 321, SMF 2008

surface [**16**], magnetic monopoles on $\mathbf{R}^3$ [**2**] and Higgs bundles on a Riemann surface [**12**]. The latter structure features prominently in the recent work of Kapustin and Witten on the Geometric Langlands correspondence [**15**]. Some of these metrics, in low dimensions, can be explicitly calculated, but even when this is not possible, the fact that these spaces are moduli spaces enables us to observe some geometrical properties which reflect their physical origin. In this paper we shall concentrate on the case of magnetic monopoles.

For monopoles in Euclidean space $\mathbf{R}^3$, there exist in certain cases explicit formulae (for example [**5**]), but in general we cannot write the metric down. Instead we can seek a geometrical means to describe the metrics; such a technique is provided by the use of twistor spaces, spectral curves and the symplectic geometry of the space of rational maps. This is documented in [**2**]. We review this in Section 2, drawing on new approaches to the symplectic structure.

We then shift attention to the hyperbolic version. The serious study of monopoles in hyperbolic space $\mathbf{H}^3$ was initiated long ago by Atiyah [**1**], who showed that there were many similarities with the Euclidean case. Yet the differential-geometric structure of the moduli space is still elusive, despite recent efforts [**18**], [**19**]. One would expect some type of quaternionic geometry which in the limit where the curvature of the hyperbolic space becomes zero approaches hyperkähler geometry. In Section 3 we give one approach to this, and show, following [**17**], how to resolve one of the problems that arises in attempting this – assigning a centre to a hyperbolic monopole.

The other problem, concerning a real structure on the putative twistor space, can currently be avoided only in the case of charge 2 and in Section 4 we produce, for the groups SU(2) and SU(3), quaternionic Kähler metrics on the moduli spaces of centred hyperbolic monopoles, generalizing the Euclidean cases computed in [**2**] and [**8**]. These metrics are expressed initially in twistor formalism, using the holomorphic contact geometry of certain spaces of rational maps, but we obtain some very explicit formulae as well.

For SU(2), these concrete self-dual Einstein metrics, originally introduced in [**14**], have nowadays found a new life in the area of 3-Sasakian geometry, Kähler-Einstein orbifolds and manifolds of positive sectional curvature. We consider briefly these aspects in the final section, and suggest where new examples might be found.

2. Euclidean monopoles

All of the hyperkähler moduli spaces mentioned above arise through the hyperkähler quotient construction. Recall that a hyperkähler metric on a manifold M^{4n} is defined by three closed 2-forms $\omega_1, \omega_2, \omega_3$ whose joint stabilizer at each point is conjugate to $Sp(n) \subset GL(4n, \mathbf{R})$. If a Lie group G acts on M, preserving the forms, then there usually exists a hyperkähler moment map $\mu : M \to \mathfrak{g}^* \otimes \mathbf{R}^3$. The quotient construction is the statement that the induced metric on $\mu^{-1}(0)/G$ is also hyperkähler.

For the moduli space of monopoles we use an infinite-dimensional version of this. The objects consist of connections A on a principal G-bundle over $\mathbf{R}^3$ together with a Higgs field ϕ, a section of the adjoint bundle. There are boundary conditions at infinity [**2**], in particular that $\|\phi\| \sim 1 - k/2r$, which imply that the connection on the sphere of radius R approaches a standard homogeneous connection as $R \to \infty$. The manifold M to which we apply the quotient construction then consists of pairs (A, ϕ) which differ from this standard connection by terms which decay appropriately, and in particular are in $\mathcal{L}^2$. This is formally an affine flat hyperkähler manifold where the closed forms ω_i are given by

$$\omega_i((\dot{A}_1, \dot{\phi}_1), (\dot{A}_2, \dot{\phi}_2)) = \int_{\mathbf{R}^3} dx_i \wedge \mathrm{tr}(\dot{A}_1 \dot{A}_2) + \int_{\mathbf{R}^3} *dx_i \wedge [\mathrm{tr}(\dot{\phi}_1 \dot{A}_2) - \mathrm{tr}(\dot{\phi}_2 \dot{A}_1)].$$

For the symplectic action of a group we take the group of gauge transformations which approach the identity at infinity suitably fast.

The zero set of the moment map in this case consists of solutions to the Bogomolny equations $F_A = *d_A\phi$, and the hyperkähler quotient is a bundle over the true moduli space of solutions – it is a principal bundle with group the automorphisms of the homogeneous connection at infinity. This formal framework has to be supported by analytical results of Taubes to make it work rigorously.

When $G = \mathrm{SU}(2)$, the connection on a large sphere has structure group $U(1)$ and Chern class k, which is called the monopole charge. The hyperkähler quotient is a manifold of dimension $4k$ which is a circle bundle over the true moduli space. It has a complete metric which is invariant under the Euclidean group and the circle action (completeness comes from the Uhlenbeck weak compactness theorem, one use of gauge theoretical results to shed light on metric properties). The gauge circle action in fact preserves the hyperkähler forms $\omega_1, \omega_2, \omega_3$, and its moment map defines a centre in $\mathbf{R}^3$. The $(4k-4)$-dimensional hyperkähler quotient can then be thought of as the moduli space of centred monopoles.

For charge 2, by using a variety of techniques [**2**], one can determine the metric explicitly. It has an action of SO(3) and may be written as

$$g = (abc)^2 d\eta^2 + a^2\sigma_1^2 + b^2\sigma_2^2 + c^2\sigma_3^2 \tag{1}$$

where

$$ab = -2k(k')^2 K \frac{dK}{dk} \quad bc = ab - 2(k'K)^2 \quad ca = ab - 2(k'K)^2$$

$$\eta = -K'/\pi K \qquad K(k) = \int_0^{\pi/2} \frac{d\phi}{\sqrt{1 - k^2 \sin^2\phi}}$$

and $\sigma_1, \sigma_2, \sigma_3$ are the standard left-invariant forms on SO(3).

Differentiably, this manifold can be understood in terms of the unit sphere in the irreducible 5-dimensional representation space of SO(3). For each axis there is, up to a scalar multiple, a unique axially symmetric vector in this representation and these trace out two copies of $\mathbf{R}\mathrm{P}^2 \subset S^4$. The centred moduli space is the complement of one of these, the removed point being the axis joining two widely separated monopoles. The other $\mathbf{R}\mathrm{P}^2$ parametrizes axially symmetric monopoles, which are (for any value of charge) uniquely determined by their axis.

For the group $G = \mathrm{SU}(3)$ we consider a Higgs field which asymptotically has two equal eigenvalues. On a large sphere the eigenspace is a rank two bundle with first Chern class k, again called the charge. When $k = 2$, Dancer computed this metric [**7**]. For centred monopoles it is eight-dimensional with an $\mathrm{SO}(3)\times\mathrm{PSU}(2)$ action, the first factor from the geometric action of rotations and the second from the automorphisms of the connection at infinity. Explicitly it can be written as follows:

$$g = \frac{1}{4}\sum_i (x(1+px)m_im_i + y(1+py)n_in_i + 2pxym_in_i)$$

where

$$\begin{aligned}
m_1 &= -f_1df_1 + f_2df_2 \quad m_2 = (f_1^2 - f_2^2)\sigma_3 \\
m_3 &= \frac{1}{px}[(pyf_3^2 - (1+py)f_1^2)\sigma_2 + f_3f_1\Sigma_2] \\
m_4 &= -\frac{1}{1+px+py}[(pyf_3^2 - (1+py)f_2^2)\sigma_1 + f_2f_3\Sigma_1] \\
n_1 &= \frac{1}{py}(-pxf_2df_2 + (1+px)f_1df_1) \\
n_2 &= \frac{1}{py}[((1+px)f_1^2 - pxf_2^2)\sigma_3 - f_1f_2\Sigma_3] \\
n_3 &= (f_1^2 - f_3^2)\sigma_2 \\
n_4 &= \frac{1}{1+px+py}[(pxf_2^2 - (1+px)f_3^2)\sigma_1 + f_2f_3\Sigma_1]
\end{aligned}$$

with σ_i, Σ_i invariant one-forms on $\mathrm{SO}(3)\times\mathrm{SU}(2)$, and

$$f_1 = -\frac{D\,\mathrm{cn}(3D,k)}{\mathrm{sn}(3D,k)} \quad f_2 = -\frac{D\,\mathrm{dn}(3D,k)}{\mathrm{sn}(3D,k)} \quad f_3 = -\frac{D}{\mathrm{sn}(3D,k)},$$

$$x = \frac{1}{D^3}\int_0^{3D}\frac{\mathrm{sn}^2(u)}{\mathrm{dn}^2(u)}du \quad y = \frac{1}{D^3}\int_0^{3D}\mathrm{sn}^2(u)du.$$

and $p = f_1f_2f_3$ for $D < 2K/3$.

Clearly there are limits to extracting information from formulae like these. Nevertheless, the restriction to certain submanifolds can be useful.

2.1. Twistor spaces. — Penrose's twistor theory provides a method for transforming the equations of a hyperkähler metric into holomorphic geometry. The idea is that the three closed two-forms of a hyperkähler manifold M can be arranged as $\omega_1, \omega_2 + i\omega_3$ which define a complex structure I for which $\omega_2 + i\omega_3$ is a holomorphic symplectic two-form and ω_1 a Kähler form. The other choices give complex structures J, K; more generally for a point $\mathbf{x} \in S^2$, $(x_1 I + x_2 J + x_3 K)^2 = -1$ and defines a complex structure.

The twistor space is the product $Z = M \times S^2$. It has a complex structure $((x_1 I + x_2 J + x_3 K), \mathbf{I})$ where $\mathbf{I}$ is the complex structure on $S^2 = \mathbf{C}\mathrm{P}^1$. The projection $p : Z \to \mathbf{C}\mathrm{P}^1$ to the second factor is holomorphic, and the fibre is M with the structure of a holomorphic symplectic manifold. There is a real structure $(m, (x_1, x_2, x_3)) \to (m, -(x_1, x_2, x_3))$. To recover the space M one sees that for $m \in M$, (m, S^2) is a holomorphic section of the projection p and M is then a component of the space of real sections.

We shall describe here how to construct the twistor space for the moduli space of SU(2) monopoles on $\mathbf{R}^3$ (see [**2**]). This involves the link with rational maps. Consider a straight line $\mathbf{x} = \mathbf{a} + t\mathbf{u}$ and the ordinary differential equation along the line $(\nabla_{\mathbf{u}} - i\phi)s = 0$. Since asymptotically $\phi \sim \mathrm{diag}(i, -i)$, there is a solution s_0 which decays exponentially at $+\infty$. Choose another solution s_1 with $\langle s_0, s_1\rangle = 1$ using the SU(2)-invariant skew form. This is well-defined modulo the addition of a multiple of s_0. Now take s_0', a solution which decays at $-\infty$, then $s_0' = as_0 + bs_1$. There are normalizations at infinity which make the coefficient b well-defined.

Now take all lines in a fixed direction $(1, 0, 0)$. We split $\mathbf{R}^3 = \mathbf{C} \times \mathbf{R}$ with coordinates $(z, t) = (x_1 + ix_2, x_3)$, and then write $s_0'(z, t) = a(z)s_0(z, t) + b(z)s_1(z, t)$. The Bogomolny equations imply that the coefficients are holomorphic in z, and furthermore the boundary conditions tell us that for a charge k monopole $b(z)$ is a polynomial of degree k. Take $p(z)$ to be the unique polynomial of degree $k - 1$ such that $p(z) = a(z)$ modulo $b(z)$ and define

$$S(z) = \frac{p(z)}{b(z)} = \frac{a_0 + a_1 z + \cdots + a_{k-1} z^{k-1}}{b_0 + b_1 z + \ldots \cdots + b_{k-1} z^{k-1} + z^k}.$$

It is a theorem that this gives a diffeomorphism between the moduli space of monopoles and the space R_k of rational maps $S : \mathbf{C}\mathrm{P}^1 \to \mathbf{C}\mathrm{P}^1$ of degree k which take ∞ to 0. Note that the denominator vanishes when $s_0' = as_0$ – when a solution exists which decays at both ends of the line. Such lines are called spectral lines.

We can carry out the above isomorphism for lines in any direction in $\mathbf{R}^3$ which yields a 2-sphere of complex symplectic structures. The set of spectral lines then forms a k-fold cover of S^2 which is called the spectral curve. It is more than just an abstract Riemann surface, though: it sits naturally as an algebraic curve in the total space of $\mathcal{O}(2)$, which we can identify with the space of straight lines in $\mathbf{R}^3$.

In order to construct the twistor space we need to do two things: first identify the symplectic structure, and secondly see how to glue the different complex structures together to give a bundle over $\mathbf{CP}^1$.

The original approach of the authors of [**2**] was to think more in terms of the physics of the monopoles rather than the geometry of rational maps. For charge $k = 1$ we know that the moduli space is flat $S^1 \times \mathbf{R}^3$, simply a gauge phase S^1 and a centre, a point of $\mathbf{R}^3$. As a complex symplectic manifold this is $\mathbf{C} \times \mathbf{C}^*$ with holomorphic symplectic form $dz \wedge dw/w$. Now there are solutions to the Bogomolny equations (the original existence theorem of Jaffe and Taubes) which approximate k widely separated charge one monopoles, so it is reasonable to think that asymptotically the moduli space approximates the symmetric product $S^k(\mathbf{C} \times \mathbf{C}^*)$ with symplectic form

$$\sum_1^k dz_i \wedge \frac{dw_i}{w_i}. \tag{2}$$

This symmetric product is singular but R_k gives in fact a smooth resolution of it: if $S(z) = p(z)/q(z)$ and the zeros of q are $z_1, \dots, z_k$, then

$$S \mapsto ((z_1, p(z_1)), \dots, (z_k, p(z_k)))$$

is the map (note that p, q being coprime means that $p(z_i) \neq 0$).

It is shown in [**2**] that the symplectic form extends, but there is now a more attractive way of defining this form (see [**10**],[**22**],[**23**]). Note that fixing all z_i gives a Lagrangian submanifold, as does fixing all w_i. In other words fixing the numerator or denominator gives two transverse Lagrangian foliations. Given $x \in \mathbf{C}$ define $f_x(S) = p(x)$, $g_x(S) = q(x)$. Then from the previous remark the Poisson brackets $\{f_x, f_y\}, \{g_x, g_y\}$ vanish. We can determine the symplectic structure by Poisson brackets of the form $\{f_x, g_y\}$ and this is defined in [**10**] by

$$\{f_x, g_y\} = \frac{p(x)q(y) - q(x)p(y)}{x - y}$$

which is the classical invariant known as the Bezoutian. Taking k points in general give local coordinates to write down the form. Clearly as $y \to x$ we get the Wronskian $p'(x)q(x) - p(x)q'(x)$, so x and y don't have to be distinct.

The expression in (2) consists of taking the points to be very special – the zeros z_i of the denominator – for then $\{f_{z_i}, g_{z_j}\} = (p(z_i)q(z_j) - q(z_j)p(z_i))/(z_i - z_j) = 0$ if $i \neq j$ and $\{f_{z_i}, g_{z_i}\} = -p(z_i)q'(z_i)$ and hence the symplectic form is

$$\sum_i \frac{1}{p(z_i)q'(z_i)} df_{z_i} \wedge dg_{z_i} = \sum_i dz_i \wedge \frac{dp(z_i)}{p(z_i)}.$$

since $q(z_i) = 0$ implies $q'(z_i)dz_i + dg_{z_i} = 0$.

To define the monopole twistor space we stick together two copies of $R_k \times \mathbf{C}$ over $\mathbf{C}^*$ by the following patching:

$$\tilde{\zeta} = \frac{1}{\zeta} \qquad \tilde{q}\left(\frac{z}{\zeta^2}\right) = \frac{1}{\zeta^{2k}} q(z) \qquad p\left(\frac{z}{\zeta^2}\right) = e^{-2z/\zeta} p(z) \mod q(z).$$

To see that this preserves the symplectic form note that if $H : R_k \times \mathbf{C}^* \to \mathbf{C}$ is defined by

$$H(S, \zeta) = \frac{1}{\zeta} \sum z_i^2$$

we obtain the Hamiltonian vector field

$$\frac{dq(z)}{dt} = 0 \qquad \frac{dp(z)}{dt} = -\frac{2z}{\zeta} p(z) \mod q(z).$$

The transformation law for p is obtained by integrating this.

To find a holomorphic section of $p : Z \to \mathbf{CP}^1$, the transformation rule for $q(z)$ shows that we must have $q(z, \zeta) = z^k + a_1(\zeta)z^{k-1} + \cdots + a_k(\zeta)$ where $a_i(\zeta)$ is a polynomial of degree $2i$. The rule for $p(z)$ (and the fact that $p(z) \neq 0$) means that on the curve $q(z, \zeta) = 0$, the line bundle with transition function $e^{-2z/\zeta}$ must be trivial. Globally, noting the transformation $z \mapsto z/\zeta^2$, this makes sense in the total space of the line bundle $\mathcal{O}(2)$ over $\mathbf{CP}^1$. But the spectral curve of a monopole is defined by the equation $q(z, \zeta) = 0$ and satisfies precisely this constraint (see [**2**]). Since the spectral curve determines the monopole we can, then, in principle find the metric on the moduli space from just two pieces of information – the spectral curve and the symplectic geometry of the space of rational functions.

3. Hyperbolic monopoles

If we replace $\mathbf{R}^3$ by hyperbolic space $\mathbf{H}^3$, then some features of the Bogomolny equations remain the same, others are radically different. The main complication is that, with the analogous boundary conditions, the SU(2) connection A has a limiting $U(1)$ connection on the boundary two-sphere at infinity which is not homogeneous. In fact the solution is uniquely determined by its boundary value [**17**]. This means that there is no obvious $\mathcal{L}^2$ metric to define on the moduli space, and no analogue of the hyperkähler quotient to suggest what sort of geometric structure the moduli space might have. Another difference is the appearance of an extra parameter, the mass m, defined as the limit of $\|\phi\|$ as $R \to \infty$. In the Euclidean case one can rescale the metric to make $m = 1$ but in the hyperbolic case this will change the value of the curvature to $-1/m$. It is convenient to have the curvature of $\mathbf{H}^3$ fixed as -1 and vary the mass. Occasionally we shall consider a limit as $m \to \infty$, and interpret it as a limit through hyperbolic metrics with curvature tending to zero.

Some features are quite similar to the Euclidean case and discussed in the original paper [**1**]. In particular, the two end-points give a parametrization of the geodesics in hyperbolic space by $S^2 \times S^2 \setminus \{x = y\}$. We give this a complex structure by letting V be the standard 2-dimensional representation space of $\mathrm{SL}(2, \mathbf{C})$ (the isometries of $\mathbf{H}^3$) and take $P(V) \times P(\bar{V}) \setminus \{x = \bar{y}\}$. By considering the equation $(\nabla_{\mathbf{u}} - i\phi)s = 0$ along a geodesic we also obtain a spectral curve for an $\mathrm{SU}(2)$ monopole of charge k which is the divisor of a section of $\mathcal{O}(k, k)$ on $P(V) \times P(\bar{V})$, and is therefore given by $H \in S^kV^* \otimes S^k\bar{V}^*$ where S^kV is the kth symmetric power of V. By reality $\bar{H} = H^t$ but it is shown in [**17**] that H actually defines a positive definite Hermitian form on S^kV.

The spectral curve satisfies a constraint analogous to that of a Euclidean monopole – instead of the triviality of the line bundle with transition function $e^{-2z/\zeta}$ we have the triviality of $\mathcal{O}(k + 2m, -k - 2m)$. Note that by removing the graph of complex conjugation from $P(V) \times P(\bar{V})$, this line bundle makes sense for any real value of m. Nonetheless, there are special reasons for considering half-integral mass, in particular any formulas we derive will be algebraic in appropriate coordinates.

Given the lack of any direct introduction of a metric structure on the moduli space, we shall attempt to use the spectral curve to generate a metric by twistor means. But problems arise even here. In the Euclidean situation the one-monopole space was flat $S^1 \times \mathbf{R}^3$; in the hyperbolic case it is $S^1 \times \mathbf{H}^3$. This carries no $\mathrm{SL}(2, \mathbf{C})$-invariant Einstein metric. If one introduces singularities then, as pointed out by Kronheimer, the spectral curves for charge one Euclidean monopole moduli spaces generate non-trivial hyperkähler metrics of A_k ALF type, which is evidence for the type of geometry to be expected in general. In the hyperbolic case one obtains this way conformal structures related to LeBrun metrics [**19**] – non-trivial geometry but still not Einstein. These low-dimensional examples therefore provide no suggestions as to what geometry to expect. On the other hand, charge one is firmly rooted in the notion of a centre – each of these four-dimensional moduli spaces has a map to Euclidean or hyperbolic space which we can regard as assigning a centre to the monopole. The problem of centres for hyperbolic monopoles has a solution given in [**17**] which we describe (in slightly different terms) next.

3.1. Centres. — Let ϵ be a skew form on V preserved by $\mathrm{SL}(2, \mathbf{C})$. Hyperbolic space is the quotient $\mathrm{SL}(2, \mathbf{C})/\mathrm{SU}(2)$ which we interpret as the space of Hermitian forms ω on V such that $\omega^2 = -2\epsilon\bar{\epsilon}$. Thus the standard $\mathrm{SU}(2)$ preserves the two forms $\omega = dz_1 \wedge d\bar{z}_1 + dz_2 \wedge d\bar{z}_2$ and $\epsilon = dz_1 \wedge dz_2$.

From ω we can form

$$\omega^{\otimes k} \in S^kV^* \otimes S^k\bar{V}^*$$

and, using the isomorphism $V \cong V^*$ given by ϵ, define the real-valued function

$$h(\omega) = H(\omega^{\otimes k}).$$

Theorem 3.1. — *The function $h : \mathbf{H}^3 \to \mathbf{R}$ has a unique critical point which we call the centre of the monopole.*

Proof. — First consider the meaning of a critical point. The derivative of h in the direction $\dot{\omega}$ is $kH(\dot{\omega} \otimes \omega^{\otimes k-1})$. But the volume form of ω is fixed so using a Lagrange multiplier λ a critical point corresponds to

$$H(\omega^{\otimes k-1}) = \lambda\omega. \tag{3}$$

Here $H(\omega^{\otimes k-1}) \in V^* \otimes \bar{V}^*$ is the contraction of $H \in S^kV^* \otimes S^k\bar{V}^*$ with $\omega^{\otimes k-1} \in S^{k-1}V^* \otimes S^{k-1}\bar{V}^*$ using the skew form on V^*.

More explicitly, use ω at a critical point to identify $V \cong \bar{V}^*$, then $H \in S^kV \otimes S^kV$. The contraction using ω is now a contraction using the skew form ϵ which gives $H(\omega^{\otimes k-1}) \in V \otimes V$. The condition (3) says that this is a multiple of ϵ^{-1}. In other words, if $H \in S^kV \otimes S^kV$, the S^2V component in the Clebsch-Gordan decomposition of this tensor product vanishes. Now we essentially follow [**17**], showing that this condition is the vanishing of a moment map.

Choose a Hermitian metric ω on V (hence an origin in $\mathbf{H}^3$), then H can be considered as a self-adjoint endomorphism of S^kV. Since it is positive definite, we can write it as $H = Q^*Q$ for an invertible endomorphism Q. Now consider the right action of $\mathrm{SL}(2, \mathbf{C})$ on Q. This gives the transformation $H \mapsto A^*HA$ on H which is the natural isometric action of hyperbolic isometries.

Consider now $\operatorname{End} S^kV$ as a complex vector space with the right action of $\mathrm{SU}(2)$. Then the moment map is the projection of Q^*Q onto the Lie algebra of $\mathrm{SU}(2)$ in $\operatorname{End} S^kV$, and the vanishing of this is just the condition for a critical point above. Now, as shown in [**17**], an invertible Q is stable for the $\mathrm{SL}(2, \mathbf{C})$ action so by the theorem of Kempf and Ness there is a point on the $\mathrm{SL}(2, \mathbf{C})$ orbit of Q for which the moment map vanishes, and this point is unique modulo $\mathrm{SU}(2)$. Thus for any positive-definite H, an isometry, well-defined modulo the stabilizer of the origin, takes it to another H whose centre is the origin: in other words a monopole has a unique centre. $\square$

Remark 3.2. — Widely separated monopoles have a spectral curve which approximates the union of twistor lines for k distinct points, so we may try and apply the definition of centre above in this case. We therefore have k Hermitian forms $H_i \in V \otimes \bar{V}^*$ and the function h is then given by the product

$$h(\omega) = (H_1, \omega)(H_2, \omega) \dots (H_k, \omega).$$

If ω is a critical point then we rewrite the form as Hermitian matrices. Replacing h by $\log h$, the condition (3) is

$$\sum_1^k \frac{H_i}{\operatorname{tr} H_i} = \lambda I. \tag{4}$$

To reinterpret this, we use the projective (Beltrami-Klein) model of $\mathbf{H}^3$. Let P be the three-dimensional real projective space of the four-dimensional vector space of Hermitian 2×2 matrices with the quadric defined by $\det H = 0$. The interior of this quadric ($\det H > 0$) is hyperbolic space. The polar plane of the identity matrix is defined by $\operatorname{tr} H = 0$ (which of course lies outside the quadric). Removing this plane gives an affine space where the hermitian matrices H_i are represented as vectors $H_i / \operatorname{tr} H_i$. The centroid in the affine sense is

$$\frac{1}{k} \sum_1^k \frac{H_i}{\operatorname{tr} H_i}$$

and from (4) we see that with our definition this is the origin.

Given two points, there is a hyperbolic isometry interchanging the two points and preserving the geodesic joining them and whose fixed point on the line is the hyperbolic midpoint. In the projective model this is a projective transformation which preserves the polar plane of that midpoint and is hence affine, so it fixes the affine midpoint. For two points, it follows that our centre coincides with the hyperbolic midpoint of the geodesic joining the points.

3.2. Rational normal curves. — For a hyperbolic monopole, we can associate to the spectral curve a certain rational curve in projective space as follows.

Given $v \in \bar{V}$, define $H(-, v^{\otimes k}) \in S^k V^*$. Because H is invertible this defines a map of degree k from $P(\bar{V})$ to $P(S^k V^*)$, or using the $\mathrm{SL}(2, \mathbf{C})$-invariant isomorphism $V \cong V^*$, to $P(S^k V)$. The spectral curve S thus naturally defines a curve $C(S) \subset P(S^k V)$. This is a rational normal curve: the image of $v \mapsto v^{\otimes k}$ is a canonical rational normal curve Δ in $P(S^k \bar{V})$ (the diagonal when we identify $S^k P(\bar{V}) = P(S^k \bar{V})$) and $C(S)$ is its image under the projective transformation H.

We lose information in passing from S to $C(S)$ – indeed for $k = 1$ $C(S)$ is the whole space. However:

Proposition 3.3. — *The spectral curve of a centred monopole is uniquely determined by the rational normal curve $C(S)$.*

Proof. — Suppose H and H_0 define the same rational curve $C(S) = C(S_0)$ and have the same centre. Rescale the forms H, H_0 so that they have determinant one. Then H and H_0 define invertible maps from Δ to $C(S)$ and so differ by a projective transformation of Δ (this is the action of $A \in \mathrm{SL}(2, \mathbf{C})$ on the representation space

$S^k\bar{V}$). Thus, considering $H, H_0 : S^k\bar{V} \to S^kV^*$ we have $H_0 = HA$. But H, H_0 are Hermitian so

$$HA = H_0 = \bar{H}_0^t = \bar{A}^t H.$$

Equivalently, A is self-adjoint with respect to the Hermitian form H and in particular its eigenspaces are orthogonal with respect to H. But if A is not a scalar then its eigenspaces in S^kV are one-dimensional and A acts as μ^{k-2i} for $0 \leq i \leq k$. Since H and H_0 are positive definite, $\mu > 0$ and hence has a positive square root, so that $A = B^2$ for $B \in \mathrm{SL}(2, \mathbf{C})$ and B is self-adjoint with respect to H. But then

$$H_0(u, v) = H(u, B^2 v) = H(Bu, Bv)$$

so H_0 is obtained from H by the isometric action on $\mathbf{H}^3$ of $B \in \mathrm{SL}(2, \mathbf{C})$. However, from the centring argument this means $B \in \mathrm{SU}(2)$ and since it has positive eigenvalues $B = 1$ and $H = H_0$. □

The spectral curve $S \subset P(V) \times P(\bar{V})$ is constrained by the condition that the bundle $\mathcal{O}(k + 2m, -k - 2m)$ is trivial. This, as we shall see next, imposes a constraint on the curve $C(S)$. From Proposition 3.3 we may consider monopoles with fixed centre, which means that we have a chosen isomorphism $V \cong \bar{V}$ and so can consider H as a linear map from S^kV to S^kV, and the spectral curve S as lying in $P(V) \times P(V)$. Its equation is then $\langle H(v^{\otimes k}), w^{\otimes k}\rangle = 0$ where the brackets denote the $\mathrm{SL}(2, \mathbf{C})$-invariant bilinear form on S^kV built from the skew form ϵ on V.

Consider the map $p : P(V) \times P(S^{k-1}V) \to P(S^kV)$ defined by symmetrizing $v \otimes q$. This is a k-fold covering (in terms of polynomials in u this is the map $(z, q(u)) \mapsto (u - z)q(u)$ so the inverse image of $r(u)$ is defined by the k roots z_i). Now if $H(v^{\otimes k}) = \mathrm{Sym}(w \otimes q)$ then $\langle H(v^{\otimes k}), w^{\otimes k}\rangle = 0$ so, restricted to the rational normal curve $C(S)$, this map is the covering $\pi : S \to P(V)$ of $P(V)$ by the spectral curve $S \subset P(V) \times P(V)$ with respect to projection on the second factor.

For convenience set $n = k + 2m$. On the spectral curve S we have a non-vanishing section of $\mathcal{O}(n, -n)$. Let $\mathcal{O}(E) = \pi_*\mathcal{O}(n, -n)$ be the direct image sheaf on $P(V)$, so that E is a rank k vector bundle. Then tautologically there is a section s of E over $P(V)$, which defines a section of the projective bundle $P(E)$.

Now

$$\pi_*\mathcal{O}(n, 0) = \pi_*(\mathcal{O}(n, -n) \otimes \mathcal{O}(0, n)) = \pi_*(\mathcal{O}(n, -n) \otimes \pi^*\mathcal{O}(n)) = \pi_*\mathcal{O}(n, -n) \otimes \mathcal{O}(n)$$

so $P(E)$ can also be written as $P(\pi_*\mathcal{O}(n, 0))$. We can then extend this definition to define a bundle E_n over $P(S^kV)$ by taking the direct image of $\mathcal{O}(n, 0)$ on $P(V) \times P(S^{k-1}V)$ under the projection p to obtain a $2k - 1$-dimensional manifold $P(E_n)$. The constraint on the spectral curve S then defines a lift of the rational normal curve $C(S)$ to a rational curve $\tilde{C}$ in this projective bundle.

We calculate now the degree of the normal bundle of $\tilde{C}$. In the fibration $P(E_n) \to P(S^kV)$, the rational curve $\tilde{C}$ is a section over $C(S)$, so its normal bundle is an extension

$$0 \to T_F \to N \to N_P \to 0$$

where N_P is the normal bundle of $C(S)$ in $P(S^kV)$. But $C(S)$ has degree k and $c_1(T\mathbf{C}\mathrm{P}^k) = k+1$ so $\deg N_P = k(k+1) - 2$.

For T_F, by Grothendieck-Riemann-Roch for the map $S \to P(V)$, the degree of E is $-k^2 + k$. The tangent bundle along the fibres T_F fits into the Euler sequence

$$0 \to \mathcal{O} \to p^*E \otimes H \to T_F \to 0$$

where H is the fibrewise hyperplane bundle (H^{-1} is the tautological bundle). On $s(P(V)) \subset P(E)$, H^{-1} coincides with the trivial subbundle of E consisting of multiples of the non-vanishing section, and hence is trivial. From the Euler sequence it follows that, restricted to the section $s(P(V))$, $\deg(T_F) = \deg(E) = -k^2 + k$. Hence

$$\deg N = \deg N_P + \deg T_F = 2k - 2.$$

Generically, we expect the holomorphic structure of this rank $2k-2$ normal bundle to be $\mathbf{C}^{2k-2}(1)$ in which case the full space of deformations of the rational curve has complex dimension $4k-4$. Indeed, if there is a real structure on the complex manifold, then this is the situation where the twistor theory for a $4k-4$-dimensional quaternionic manifold becomes a theory of rational curves [**20**], a particular case being hyperkähler geometry. One might therefore expect that the complex manifold we have defined gives some type of quaternionic geometry for the moduli space of centred monopoles. There is a problem though, which involves the real structure.

By centring, we have a quaternionic structure on V and hence a real structure on $P(S^kV)$, and we have a rational normal curve $C(S) = H(\Delta)$ depending on a spectral curve S which has a real structure. However, $C(S)$ in general is not preserved by the real structure on $P(S^kV)$. In fact the reality condition on the spectral curve implies that $\bar{C}(S) = H^t(\Delta)$, so unless $H = H^t$ we do not have reality for $C(S)$.

However for charge $k = 2$, the Clebsch-Gordan decomposition is

$$S^2V \otimes S^2V = S^4V \oplus S^2V \oplus 1$$

where $S^4V \oplus 1$ are the symmetric forms and S^2V the skew-symmetric forms on S^2V. Centring sets the S^2V component to zero and so here we do in fact have $H = H^t$.

4. Charge 2 hyperbolic monopoles

4.1. SU(2) **monopoles.** — The programme in Section 3.2 for constructing a twistor space has been carried out in [**14**] to yield a quaternionic Kähler structure on the moduli space of centred charge 2 hyperbolic monopoles. Recall (see [**3**] Chapter

14) that a quaternionic Kähler manifold is of dimension $4n$ and has a rank three bundle $\mathcal{Q}$ of 2-forms $\omega_1, \omega_2, \omega_3$ whose stabilizer at each point is conjugate to $Sp(n) \cdot Sp(1) \subset GL(4n, \mathbf{R})$. The ideal generated by the ω_i should be closed under exterior differentiation. The bundle $\mathcal{Q}$ can also be thought of as the imaginary part of a bundle of quaternion algebras. The standard example is quaternionic projective space $\mathbf{HP}^n$.

The twistor space of a quaternionic Kähler manifold is a complex manifold Z^{2n-1} with a family of rational curves C with normal bundle $\mathbf{C}^{2n-2}(1)$ and a holomorphic contact form; there must also be a real structure compatible with these. It is the contact form which is the new feature here, replacing the symplectic geometry in the hyperkähler case.

Example 4.1. — A simple example is to take $Z = P(T^*\mathbf{CP}^{n+1})$, which has a canonical contact structure. Each rational curve is determined by a line $L \subset \mathbf{CP}^{n+1}$: using a Hermitian metric on $\mathbf{C}^{n+2}$, we take $L^{\perp}$ to be the orthogonal projective $(n-1)$-space and then for each $x \in L$, the join $x + L^{\perp}$ is a hyperplane in $\mathbf{CP}^{n+1}$ with a distinguished point x on it – hence a point in $P(T^*\mathbf{CP}^{n+1})$. As x moves along the line L this defines a rational curve in Z. The corresponding quaternionic Kähler manifold is the $4n$-dimensional Wolf space $U(n+2)/U(2) \times U(n)$ – the Grassmannian of lines L in $\mathbf{CP}^{n+1}$.

Here is how to derive the metric from the twistor data. A contact structure on Z is given invariantly by a holomorphic section φ of $T^* \otimes K^{-1/n}$ where $L = K^{-1/n}$ is a line bundle such that $L^n \cong K^{-1}$, the anticanonical bundle. In this formalism $\varphi \wedge (d\varphi)^{n-1}$ is a well-defined section of the trivial bundle and is therefore allowed to be everywhere non-vanishing, the contact condition. If the normal bundle N of C is isomorphic to $\mathbf{C}^{2n-2}(1)$ then the degree of K^{-1} on C is $2n$. Restricting φ to C gives a homomorphism from the tangent bundle T_C of the curve to $K^{-1/n}$. These are both of degree 2 and we consider rational curves for which this is non-zero and hence an isomorphism. It follows that $\varphi : TZ \to K^{-1/n} \cong T_C$ is a splitting of the sequence of bundles on C:

$$0 \to T_C \leftrightarrows TZ \to N \to 0.$$

A tangent vector to the space of rational curves at a curve C is a holomorphic section Y of N, which using the above splitting we can regard as a subbundle of $TZ|_C$. Again since $N \cong \mathbf{C}^{2n-2}(1)$, we have

$$H^0(C, N) \cong H^0(C, \mathcal{O}(1)) \otimes H^0(C, N(-1)) \tag{5}$$

where $\mathcal{O}(1) = K_C^{-1/2}$.

Now since φ is a contact form, $d\varphi$ restricted to the kernel of φ is a non-degenerate skew form with values in $K^{-1/n}$ which we have just identified with $T_C = \mathcal{O}(2)$. Hence it defines a skew form on $H^0(C, N(-1))$. There is a natural skew form (the

Wronskian) on $H^0(C, \mathcal{O}(1))$, and these two define a symmetric inner product on the tensor product, which from (5) is the tangent space to the space of rational curves.

Remark 4.2. — Given a point $x \in C$, there is, up to a constant multiple, a unique section v of $\mathcal{O}(1)$ which vanishes at x and so the sections of N which vanish at x are, in the decomposition (5), of the form $s \otimes v$. Since $\langle v, v\rangle = 0$ it follows from our description of the metric that such complex vectors are null.

The above is not the approach of [**14**], which is heavily focused on an alternative viewpoint: centred charge 2 monopoles form a 4-dimensional manifold with an isometric action of SO(3), which has generically codimension one orbits. Differentiably, and equivariantly, this is the same space as the Euclidean two-monopole space mentioned in Section 2: the 4-sphere with a copy of $\mathbf{R}\mathrm{P}^2$ removed.

The ODE which defines the metric is a particular form of the Painlevé VI equation. Together with some algebraic geometry [**13**],[**14**] it gives explicit formulae for the metric like the following (the charge 2, mass 2 case)

$$g = \frac{1+r+r^2}{r(r+2)^2(2r+1)^2}dr^2 + \frac{(1-r^2)^2}{(1+r+r^2)(r+2)(2r+1)}\sigma_1^2 + \\ + \frac{1+r+r^2}{(r+2)(2r+1)^2}\sigma_2^2 + \frac{r(1+r+r^2)}{(r+2)^2(2r+1)}\sigma_3^2$$

where $\sigma_1, \sigma_2, \sigma_3$ form a standard basis of left-invariant forms on SO(3). (The interested reader should beware of typos in some of the formulae in [**14**], and cross-check with [**24**], for example). The above defines a self-dual Einstein 4-manifold with positive scalar curvature, which is how one interprets the quaternionic Kähler condition in four dimensions.

The advantage of explicit formulas (in this case either algebraic or using elliptic functions) is that the behavior of the metric can be analyzed in more detail, and in [**14**] it is shown that, for all values of n, each of these (incomplete) metrics has an orbifold singularity of angle $2\pi/(n-2)$ around the removed $\mathbf{R}\mathrm{P}^2$. This is quite unlike the Euclidean moduli space, which is complete. Another feature is that, as $n \to \infty$ the metric approaches the Euclidean monopole metric, consistent with the idea that rescaling the mass to 1 is equivalent to changing the curvature to $-1/m$.

In fact, the orbifold behaviour can be detected without calculating the metric, as can be seen by looking at the twistor space afresh.

4.2. Orbifold twistor spaces. — We consider as in Section 3.2, for $k = 2$, the projective bundle $P(E_n) \to P(S^2V)$ obtained from the direct image of $\mathcal{O}(n, 0)$ under the projection $p : P(V) \times P(V) \to P(S^2V)$, which in this case is just the quotient by interchange of the two factors. These bundles were first introduced by Schwarzenberger

[21]. By Grothendieck-Riemann-Roch $c_1(E_n) = (n-1)x, c_2(E_n) = (n(n-1)/2)x^2$ where x is the positive generator of $H^2(P(S^2V), \mathbf{Z})$. In this case, the rational normal curve $C(S)$ is a conic in $P(S^2V)$ and the spectral curve is an elliptic curve, the double covering of $C(S)$ over the four points of intersection with the conic $\Delta \subset P(S^2V)$.

The twistor space itself is not the whole of $P(E_n)$, but an open subset. There is a divisor $D \subset P(E_n)$ defined by the section of $P(E_n)$ over $P(V) \times P(V)$ given by the kernel of the natural evaluation map

$$\text{ev} : p^*E_n \to \mathcal{O}(n, 0). \tag{6}$$

Because the lift $\tilde{C}$ of $C(S)$ is defined by a non-vanishing section of $\mathcal{O}(n, -n)$, $\tilde{C} \cap D = \varnothing$, and the twistor space Z is actually equal to $P(E_n) \setminus D$.

Given that the metric has an orbifold singularity around a copy of $\mathbf{R}\mathrm{P}^2$, there must be a singular compactification of this twistor space by adding in a 2-sphere. We shall see this next by using algebraic geometry instead of differential geometry, by showing that D can be blown down to a rational curve.

By the definition of D, the kernel of ev is naturally isomorphic on D to the tautological bundle H^{-1} and so from (6) $H^{-1} \cong \Lambda^2 p^*E_n(-n, 0)$ but $c_1(E_n) = (n-1)x$ thus $c_1(p^*E_n) = (n-1, n-1)$ and hence $H^{-1} \cong \mathcal{O}(-1, n-1)$. The cohomology class of D is of the form $ah + bx$ where $h = c_1(H)$, and as D intersects a generic fibre of $P(E_n)$ in two points, $a = 2$. Since $H \cong \mathcal{O}(1, -n+1)$ we have $h^2[D] = -2(n-1)$ and using $h^2 = c_1(E_n)h - c_2(E)$ we find $b = n$. Hence D is a divisor of the line bundle $H^2 \otimes p^*\mathcal{O}(n)$. Its normal bundle is therefore

$$H^2 \otimes p^*\mathcal{O}(n)|_D = \mathcal{O}(2, -2n+2) \otimes \mathcal{O}(n, n) = \mathcal{O}(n+2, -n+2).$$

For $n > 2$ the second degree is negative and we can therefore blow down the second $\mathbf{C}\mathrm{P}^1$ factor in $D \cong \mathbf{C}\mathrm{P}^1 \times \mathbf{C}\mathrm{P}^1$. For $n = 3$ the resulting manifold is smooth, but for $n > 3$ we have an orbifold singularity along a rational curve, locally modelled on a quotient of $\mathbf{C}^3$ by $\mathbf{Z}/(n-2)$ acting as $(z_1, z_2, z_3) \mapsto (z_1, \omega z_2, \omega z_3)$.

Remark 4.3. — In the case $n = 3$, the smooth blow-down is just $\mathbf{C}\mathrm{P}^3$. Take a rational normal curve $C \subset \mathbf{C}\mathrm{P}^3$ (a twisted cubic). It is well-known that through a generic point in $\mathbf{C}\mathrm{P}^3$ there passes a unique secant to C. The secant intersects C in two points and so this defines a rational map from $\mathbf{C}\mathrm{P}^3$ to the symmetric product $S^2C = \mathbf{C}\mathrm{P}^2$; the fibre is the secant itself. Blowing up C gives the projective bundle $P(E_3)$. Another way of looking at this is to view the family of secants as a map from S^2C to the Grassmannian of lines in $\mathbf{C}\mathrm{P}^3$ (its image is actually a Veronese surface in the Plücker embedding). The projective bundle is then the pull-back of the tautological bundle on the Grassmannian.

4.3. The contact form. — In the approach of [**14**], the contact form was defined using the Maurer-Cartan form of SO(3), but there is another way which makes contact with the symplectic geometry of rational maps discussed earlier, and gives an alternative viewpoint on the twistor space.

Consider an affine coordinate z on $\mathbf{C} \subset P(V)$ and a trivialization $dz^{-n/2}$ of $\mathcal{O}(n)$. Then for $(z_1, z_2) \in \mathbf{C} \times \mathbf{C} \subset P(V) \times P(V)$ with $z_1 \neq z_2$, the fibre of the bundle E_n defined by the direct image sheaf of $\mathcal{O}(n, 0)$ consists of a linear combination of $dz^{-n/2}$ at z_1 and z_2, and we can take local coordinates w_1, w_2 relative to this basis. On the complement of the divisor D we have w_1 and w_2 non-zero. Define a one-form by

$$\varphi = (z_1 - z_2)\left(\frac{dw_1}{w_1} - \frac{dw_2}{w_2}\right) + n(dz_1 + dz_2). \tag{7}$$

Note first that this is invariant under the exchange of z_1 and z_2 together with w_1 and w_2. Also, φ is homogeneous of degree 0 in the w_i and annihilates the Euler vector field

$$W = w_1 \frac{\partial}{\partial w_1} + w_2 \frac{\partial}{\partial w_2}$$

and so descends to the projective bundle $P(E_n)$.

As in Section 2, we can associate to the data z_1, z_2, w_2, w_2 a degree 2 rational map

$$S(z) = \frac{a_0 + a_1 z}{b_0 + b_1 z + z^2}$$

where the denominator is $(z - z_1)(z - z_2)$ and the numerator is the unique linear polynomial which takes the value w_1 at z_1 and w_2 at z_2. That part of the twistor space Z which lies over the open set of $P(S^2V)$ consisting of quadratic polynomials with finite roots can then be interpreted as the quotient of the space of rational maps R_2 by scalar multiplication. The form φ extends too, for it may be written as $\varphi = i_U\omega + ndx$ where $x = z_1 + z_2$, ω is the symplectic form described in Section 2.1 and U is the vector field

$$(z_1 - z_2)\left(\frac{\partial}{\partial z_1} - \frac{\partial}{\partial z_2}\right).$$

Using $x = z_1 + z_2$ and $y = z_1 z_2$ as local affine coordinates on $P(S^2V)$, we can write

$$U = (4y - x^2)\frac{\partial}{\partial y}$$

which is thus well-defined on the open set of $P(S^2V)$. It follows that φ is well-defined even where $z_1 = z_2$. Moreover,

$$\varphi \wedge d\varphi = -2ndz_1 \wedge dz_2 \wedge \left(\frac{dw_1}{w_1} - \frac{dw_2}{w_2}\right) = -2ni_W\omega \wedge \omega$$

and since ω is symplectic this defines a contact structure.

Now consider the action of a Möbius transformation $f(z) = (az+b)/(cz+d)$. It acts on $w \in \mathcal{O}(n,0)$ over z by $w \mapsto w(cz+d)^n$. A short calculation gives

$$f^*\varphi = \frac{1}{(cz_1+d)(cz_2+d)}\varphi.$$

It follows that we can use the same description replacing ∞ by $f(\infty)$. In particular, taking the three points $0, 1, \infty \in P(V)$, we cover $P(S^2V)$ by three corresponding affine open sets consisting of quadratic polynomials for which 0, or 1 or ∞ is not a root. It follows that φ extends as a line-bundle valued form over the whole of Z.

Our conclusion is that the twistor space has an alternative description: it is covered by open sets each of which is isomorphic to the quotient of the space R_2 of rational maps by scalar multiplication, where the identification preserves the contact structure (7).

4.4. SU(3) **monopoles.** — We now approach the question of what the moduli space for a hyperbolic SU(3) monopole with minimal symmetry breaking looks like. The Euclidean case was dealt with by Dancer [**7**]. There is again a spectral curve involved, which for charge 2 is an elliptic curve, but this time it is unconstrained: instead we have a choice of data, which is a pair of sections of $\mathcal{O}(\ell+1,-\ell)$ which are linearly independent at each point. This is a line bundle of degree 2 on an elliptic curve and so has a two-dimensional space of sections. When we centre the monopole we have five real degrees of freedom for the conic and an SU(2) gauge action which acts on the pair of sections, giving an 8-dimensional moduli space.

We shall try and defined a quaternionic Kähler metric from a twistor space. As we have seen before, the direct image sheaves of $\mathcal{O}(\ell+1,-\ell)$ and $\mathcal{O}(2\ell+1,0)$ have the same projective bundle, so a pair of sections of $\mathcal{O}(\ell+1,-\ell)$ over the conic $C(S)$ defines a section of $P(E_n \otimes \mathbf{C}^2)$, a 5-dimensional complex manifold, where $n = 2\ell+1$. These sections will be the twistor lines.

Remark 4.4. — The linear independence condition means that the twistor lines lie in an open set of $P(E_n \otimes \mathbf{C}^2)$, which gives an alternative description of Z as the principal $\mathrm{PSL}(2,\mathbf{C})$ frame bundle of $P(E_n)$.

We now have to introduce a contact form and here we take the lead from the Euclidean case treated by Dancer. In [**8**] he defines a symplectic form on the space of rational maps $z \mapsto [f_1(z), f_2(z), f_3(z)]$ of degree 2 from $\mathbf{C}\mathrm{P}^1$ to $\mathbf{C}\mathrm{P}^2$ which take ∞ to $[0,0,1]$, so that f_1, f_2 are linear and $f_3 = (z-z_1)(z-z_2)$. If $f_1(z_i) = p_i, f_2(z_i) = q_i$ and $z_1 \neq z_2$ then this form is, up to a constant,

$$\omega = -\frac{1}{z_1-z_2}dz_1 \wedge dz_2 + dz_1 \wedge \theta_1 + dz_2 \wedge \theta_2 + (z_1-z_2)d\theta_1$$

where

$$\theta_1 = \frac{q_1 dp_2 - p_1 dq_2}{p_1 q_2 - p_2 q_1} \qquad \theta_2 = \frac{p_2 dq_1 - q_2 dp_1}{p_1 q_2 - p_2 q_1}.$$

Following the rational map description of $P(E_n)\setminus D$, we can apply a similar argument here and define a one-form in the open set $z_1, z_2 \neq \infty$ by

$$\varphi = (z_1 - z_2)(\theta_1 - \theta_2) + nd(z_1 + z_2) \tag{8}$$

As in the previous case, the form φ may be written $\varphi = i_U\omega + ndx$ and the fact that ω is symplectic shows that this is a contact form on the quotient by the scalars acting on the rational maps. It has the same transformation properties as the contact form in the SU(2) case, and so extends.

The twistor lines again cover $C(S) \subset P(S^2V)$. We know that S has two sections s_1, s_2 of $\mathcal{O}(\ell+1, -\ell)$. In local coordinates $(z_1, z_2, p_1, p_2, q_1, q_2)$, p_1, p_2 are the values of s_1 at z_1, z_2 and q_1, q_2 the values of s_2, so the two sections give, in the rational map picture, the two numerators.

There are now two group actions – a geometrical action of SO(3), the hyperbolic isometries fixing the centre, and a gauge action by PSU(2) which changes the basis of sections of $\mathcal{O}(\ell+1, -\ell)$ (part of the principal bundle action according to Remark 4.4).

Given the twistor space we need to find the rational curves more explicitly, but there is a very concrete way of doing this in the case where $n = 2\ell+1$ is an odd integer. The spectral curve is an elliptic curve $S \subset P(V) \times P(V)$, a divisor of $\mathcal{O}(2,2)$, and projects to a conic $C(S) \subset P(S^2V)$. Choose a point $P_0 = (x_0, y_0) \in S \subset P(V) \times P(V)$ and take the line $\{x_0\} \times P(V)$ through this point. It intersects S in a point $Q_0 = (x_0, y_1)$ and so the divisor class $P_0 + Q_0 \sim \mathcal{O}(1,0)$. Now take the line $P(V) \times \{y_1\}$ which passes through Q_0 and intersects S again in $P_1 = (x_1, y_1)$, and continue. We have the divisor classes

$$P_0 + Q_0 \sim \mathcal{O}(1,0) \quad Q_0 + P_1 \sim \mathcal{O}(0,1) \quad P_1 + Q_1 \sim \mathcal{O}(1,0) \ldots$$

from which we get

$$P_0 + Q_0 + \cdots + P_\ell + Q_\ell \sim \mathcal{O}(\ell+1, 0)$$
$$Q_0 + P_1 + Q_1 + P_2 + \cdots + Q_{\ell-1} + P_\ell \sim \mathcal{O}(0, \ell)$$

and so

$$P_0 + Q_\ell \sim \mathcal{O}(\ell+1, -\ell).$$

Hence $P_\ell + Q_0$ is the zero set of a section of this bundle on S.

Down in $P(S^2V)$ we start at the image $X_0 \in C(S)$ of P_0, draw a tangent to the diagonal conic Δ to meet $C(S)$ at X_1, and continue. The Poncelet problem of the "in-and-circumscribed polygon" to two conics is the closure condition for this process and was the basis of the explicit formulas in [**13**],[**14**].

The question we ask ourselves now is whether this twistor data generates an orbifold quaternionic Kähler metric in eight dimensions. The twistor space has an open orbit under the complexified action of $\mathrm{SO}(3,\mathbf{C}) \times \mathrm{PSL}(2,\mathbf{C})$ but there seems no obvious way of equivariantly blowing down any lower dimensional orbits. In fact we shall see that there is no orbifold compactification in this case. We don't need to calculate the whole metric, just the induced metric on a certain totally geodesic submanifold.

4.5. Axially symmetric monopoles. — For each charge and mass there is, for the group $\mathrm{SU}(2)$, a unique monopole which is symmetric about a given axis. For $\mathrm{SU}(3)$ this is no longer the case and we shall compute the metric restricted to a surface of revolution which represents all such axially symmetric monopoles.

An axially symmetric spectral curve is of the form $(w - \mu z)(w - \mu^{-1}z) = 0$ and this defines the rational normal curve $z \mapsto w^2 - (\mu + \mu^{-1})wz + z^2$ in the space of quadratic polynomials in w. The parameter μ is real or complex depending on whether $\mu + \mu^{-1} - 2$ is positive or negative.

A section of $\mathcal{O}(\ell+1, -\ell)$ constructed as above and with $(x_0, y_0) = (\mu, 1)$ is given by
$$\frac{(w-\mu)(w-\mu^3)\dots(w-\mu^{(2\ell+1)})}{(z-\mu^2)(z-\mu^4)\dots(z-\mu^{2\ell})}$$
in the local trivialization $dw^{-(\ell+1)/2}dz^{-\ell/2}$. On the branch $w = \mu z$ it has the form $\mu^{\ell+1}(z-1)$ and on the branch $w = \mu^{-1}z$ is $\mu^{-(\ell+1)}(z - \mu^{2\ell+2})$. This is a section of a line bundle of degree 2 on a (degenerate) elliptic curve which therefore has two linearly independent sections. Changing the initial point (x_0, y_0) it is clear that this space is spanned by the two sections s_1 and s_2 where $s_1 = 1$ on each branch and $s_2 = \mu^{\ell+1}z$ on the first branch and $s_2 = \mu^{-(\ell+1)}z$ on the second.

The geometrical S^1-action $z \mapsto \lambda z$ acts as $(s_1, \lambda s_2)$ and the gauge action is $(s_1, s_2) \mapsto (\lambda^{-1/2}s_1, \lambda^{1/2}s_2)$, so coupling the two multiplies (s_1, s_2) by $\lambda^{1/2}$. This lifting of the geometric circle action means that we can consider the metric on the fixed point set, which is a totally geodesic surface of revolution.

To do the calculation we need to use coordinates for this data on a varying curve: we set $\mu = e^{2t}$ (where t is real in the first instance) and on the first branch $w = e^t u, z = e^{-t}u$ and on the second $z = e^t u, w = e^{-t}u$. Then u is a rational parametrization of the plane conic defined by the spectral curve: in fact u is an affine parameter on the diagonal conic Δ and we have transformed it by the hyperbolic isometry $\mathrm{diag}(e^t, e^{-t})$. The real structure is given by $u \mapsto -1/\bar{u}$. This provides a uniform parametrization of our family of conics.

The twistor line for an axisymmetric monopole is then given, with $n = 2\ell+1$, by:
$$z_1 = e^t u \quad z_2 = e^{-t}u \quad p_1 = 1 \quad p_2 = 1 \quad q_1 = e^{-i\phi - nt}u \quad q_2 = e^{-i\phi + nt}u.$$

(recall that p_i and q_i are the values of s_1 and s_2 at z_i). Differentiating with respect to u, the tangent to the line is spanned by

$$X = z_1\frac{\partial}{\partial z_1} + z_2\frac{\partial}{\partial z_2} + q_1\frac{\partial}{\partial q_1} + q_2\frac{\partial}{\partial q_2}.$$

Differentiating with respect to t and ϕ, an infinitesimal variation of the twistor line is given by:

$$Y = \dot{t}\left(z_1\frac{\partial}{\partial z_1} - z_2\frac{\partial}{\partial z_2}\right) - (i\dot{\phi} + n\dot{t})q_1\frac{\partial}{\partial q_1} - (i\dot{\phi} - n\dot{t})q_2\frac{\partial}{\partial q_2}.$$

At this stage we put into effect the description of the metric in Section 4.1. The tangent bundle of the rational curve C is defined by X and we must use the contact form to embed the normal bundle N in TZ. Thus Y is a section of TZ over C and

$$Y - \frac{\varphi(Y)}{\varphi(X)}X$$

is a section of the normal bundle.

We evaluate the contact form on the vectors X and Y to obtain

$$\varphi(X) = 2u(n\cosh t - \sinh t \coth nt) \qquad \varphi(Y) = 2i\dot{\phi}\sinh t \coth nt.$$

When $\dot{t} = 0, \dot{\phi} = 1$ the section of the normal bundle is then

$$Y_0 = -2i\frac{1}{\varphi(X)}\left[\sinh t \coth nt\left(z_1\frac{\partial}{\partial z_1} + z_2\frac{\partial}{\partial z_2}\right) + n\cosh t\left(q_1\frac{\partial}{\partial q_1} + q_2\frac{\partial}{\partial q_2}\right)\right]$$

and when $\dot{\phi} = 0, \dot{t} = 1$

$$Y_1 = z_1\frac{\partial}{\partial z_1} - z_2\frac{\partial}{\partial z_2} - n\left(q_1\frac{\partial}{\partial q_1} + q_2\frac{\partial}{\partial q_2}\right).$$

On our two-dimensional submanifold, whose tangent space is spanned by $\partial/\partial t$ and $\partial/\partial\phi$, the area form is $d\varphi(Y_1, Y_0)$:

$$n\frac{n\sinh t\cosh t \operatorname{cosech}^2 nt - \coth nt}{n\cosh t - \sinh t\coth nt}d\phi \wedge dt. \tag{9}$$

To obtain the metric, we also need the conformal structure which, from Remark 4.2 we can derive by considering complex variations of the twistor line which preserve a point. Our twistor lines are $\mathbf{C}^*$-invariant so we have to consider variations preserving a fixed point of the action. Unfortunately, this is where our local coordinates break down – we have to consider a description of the vector bundle E_n at a branch point of the covering $p : P(V)\times P(V) \to P(S^2V)$ restricted to the conic. The map p is the quotient by permuting the factors, so we can describe it as $p(w, z) = [1, w+z, wz] \in \mathrm{CP}^2$. The equation of the spectral curve is $w^2 - (\mu + \mu^{-1})wz + z^2 = 0$ and clearly any local function $f(w, z)$ can be written as $g(z) + wh(z)$ modulo the equation of the curve, or more conveniently as

$$f_0(w+z) + (w-z)f_1(w+z).$$

The direct image sheaf of $\mathcal{O}$ is then generated by 1 and $w - z$ over the functions on the conic.

We apply this now to our two sections s_1, s_2 of $\mathcal{O}(\ell+1, -\ell)$. The first is equal to 1 on both branches and gives $f_1 = 0, f_0 = 1$. The second is $\mu^{\ell+1}z$ if $w = \mu z$ and $\mu^{-(\ell+1)}z$ if $w = \mu^{-1}z$. This gives

$$f_0(u) = \frac{1}{2}\frac{\cosh(2\ell+1)t}{\cosh t}u \qquad f_1(u) = \frac{1}{2}\frac{\sinh(2\ell+1)t}{\sinh t}.$$

Thus a variation of the twistor line which keeps the point $u = 0$ fixed is obtained by fixing (again with $n = 2\ell+1$)

$$e^{-i\phi}\frac{\sinh nt}{\sinh t}.$$

It follows that the conformal structure is defined by

$$d\phi^2 + (n\coth nt - \coth t)^2 dt^2.$$

From (9), we conclude:

Proposition 4.5. — *The metric g restricted to the space of axially symmetric* SU(3) *monopoles is:*

$$n\frac{\coth nt - n\sinh t\cosh t\,\mathrm{cosech}^2\, nt}{(n\cosh t - \sinh t\coth nt)(n\coth nt - \coth t)}(d\phi^2 + (n\coth nt - \coth t)^2 dt^2).$$

When, in the equation of the spectral curve, $\mu+\mu^{-1} < 2$ then t becomes imaginary and we must replace the hyperbolic functions by the corresponding trigonometrical ones. Note that, near $\mu = 1$ (or $t = 0$), the metric is still regular and behaves like

$$\frac{2n}{n^2-1}\left(d\phi^2 + \left(\frac{n^2-1}{3}\right)^2 t^2 dt^2\right).$$

Fix n and consider the limit of the metric as $t \to \infty$. We find that the metric approximates

$$\frac{ne^{-t}}{(n-1)^2}(d\phi^2 + (n-1)^2 dt^2)$$

and putting $r^2 = e^{-t}$ this gives

$$\frac{n}{(n-1)^2}(r^2 d\phi^2 + 4(n-1)^2 dr^2)$$

which has an orbifold singularity: a quotient by $\mathbf{Z}/2(n-1)$. Here the spectral curve exactly corresponds to a pair of points, so the region is analogous to the orbifold singularity in the SU(2) case.

At the other extreme, consider the metric on the trigonometric branch:

$$n\frac{\cot nt - n\sin t\cos t\,\mathrm{cosec}^2\, nt}{(n\cos t - \sin t\cot nt)(n\cot nt - \cot t)}(d\phi^2 + (n\cot nt - \cot t)^2 dt^2)$$

as $t = \pi/n - u$ for u small. Then we obtain

$$n \cos \frac{\pi}{n}(d\phi^2 + u^{-2}du^2)$$

which is asymptotic to a cylinder.

Now suppose the moduli space had an equivariant orbifold compactification. Then the fixed point set of the circle action would extend to a compact orbifold and in particular would have finite area. But the cylinder has infinite area and so an orbifold compactification is impossible.

Remark 4.6. — Dancer explicitly wrote down the metric in the Euclidean case. If we fix t and ϕ and put $r = nt$, then the metric ng as $n \to \infty$ has a limit which is

$$\frac{r(\coth r - r \operatorname{cosech}^2 r)}{r \coth r - 1}\left(d\phi^2 + \left(\coth r - \frac{1}{r}\right)^2 dr^2\right).$$

This is precisely Dancer's metric (see [**7**] Theorem 5.1, or put $f_1 = -D \coth 3D, f_2 = f_3 = -D \operatorname{cosech} 3D$ in the formula in Section 2.) Thus in the infinite mass limit, or as the curvature of hyperbolic space tends to zero, our metric approaches the known Euclidean monopole metric. In the Euclidean case, the metric is asymptotically cylindrical where ours has an orbifold singularity, and asymptotically conical (with vertex angle $\pi/3$) where ours is cylindrical.

5. New metrics for old

The relationship between these metrics and their physical origins in the study of monopoles on hyperbolic space is not at all clear. We have proceeded by analogy and used spectral data rather than the fields themselves to provide a route to the metric. On the other hand they provide us also with a means for constructing other solutions to Einstein's equations. As the reader may find in [**3**], when a quaternionic Kähler manifold has positive scalar curvature, its twistor space has a natural Kähler-Einstein metric. Thus the singular spaces obtained in Section 4.2 are Fano varieties with explicit Kähler-Einstein metrics. But one can go further – the principal SO(3) bundle of the rank three bundle $\mathcal{Q}$ of imaginary quaternions on a quaternionic Kähler manifold also has a natural Einstein metric. This is a 3-Sasakian metric (which also means that by rescaling the SO(3) orbits one can find yet another Einstein metric). The $4n+3$-dimensional 3-Sasakian manifold is a principal S^1-bundle over the twistor space. One should read about these in the recently published book of Boyer and Galicki [**4**], in many respects a worthy successor to [**3**]. (The authors of that book note that "3-Sasakian manifolds are never mentioned in Besse" which is quite true, though had Arthur Besse known Bär's result that the metric cone on a 3-Sasakian manifold is hyperkähler he would almost certainly have taken them more seriously). For our orbifold examples, the 3-Sasakian manifold is actually smooth: the circle action is semi-free and has finite isotropy subgroup over the singular points of the twistor space. What is perhaps more interesting is that these 7-manifolds – as differentiable manifolds with a cohomogeneity one group action – have occurred in a completely different context, that of manifolds of positive curvature [**11**],[**24**]. A series of manifolds P_k and Q_k were found to be candidates for having metrics of positive sectional curvature. These manifolds are the (2-fold) universal covers of the 3-Sasakian manifolds associated to the moduli spaces of hyperbolic charge 2 monopoles of mass $(2k-1)/2$ and k respectively. Quite recently, Dearricott (unpublished) and, independently, Ziller [**24**] have shown that P_2 does indeed admit a positively curved metric.

From [**9**], the sectional curvature of the 3-Sasakian metric on the 7-manifold will be positive if the sectional curvature of the 4-manifold is positive. For the Einstein metrics described above this is true when the two monopoles are well separated but not when they are close to an axially symmetric one. Indeed, the scattering of Euclidean monopoles described in [**2**] involves some negative curvature behaviour which seems likely to persist in the hyperbolic case. The positively curved examples on P_2 are constructed by concretely deforming the 3-Sasakian metric.

There may however be other self-dual Einstein structures on the 4-dimensional spaces. Indeed, one of the spin-offs of Dancer's work on SU(3) monopoles was a hyperkähler deformation of the metric (1), obtained as a hyperkähler quotient of the Euclidean SU(3) moduli space. In the hyperbolic SU(3) case described in Section 4.4 we have an action of the rank two group $\mathrm{SO}(3)\times\mathrm{PSU}(2)$ and so we could attempt to take a quotient by a circle subgroup. Note that if the circle is in the gauge action PSU(2), then it has a commuting SO(3) action which descends to the quaternionic Kähler quotient, so already we know that this particular quotient is an SO(3)-invariant self-dual Einstein manifold. In fact this quotient is the SU(2) moduli space. To see this, recall [**4**] that from the twistor point of view quaternionic Kähler reduction proceeds by evaluating the contact form φ on the vector fields generated by the group action to get a section of $\mathfrak{g}^*\otimes K^{-1/n}$. The twistor space of the reduction is the quotient of the zero-set of this by the complexified group action. In our case the gauge circle action is $(p_1,p_2,q_1,q_2)\mapsto(p_1,p_2,e^{i\theta}q_1,e^{i\theta}q_2)$ which generates the vector field

$$X=q_1\frac{\partial}{\partial q_1}+q_2\frac{\partial}{\partial q_2}.$$

Evaluating the contact form (8) gives

$$\varphi(X)=(z_1-z_2)\left(\frac{p_2q_1+p_1q_2}{p_2q_1-p_1q_2}\right)$$

and the zero set of this is $p_2q_1+p_1q_2=0$. As a subset of $P(E_n\otimes\mathbf{C}^2)$ the equation $p_2q_1-p_1q_2=0$ is the quadric $P(E_n)\times\mathbf{CP}^1$ and the complement is the SU(3) twistor space. The quotient twistor space is defined by $p_2q_1 + p_1q_2 = 0$ modulo $(p_1, p_2, q_1, q_2) \mapsto (\lambda p_1, \lambda p_2, \mu q_1, \mu q_2)$. The projection to $[p_1, p_2] \in P(E_n)$ maps this isomorphically to the complement of $p_1 = 0$ and $p_2 = 0$ which is $P(E_n) \setminus D$, the twistor space for the SU(2) moduli space.

A more general circle subgroup of $\mathrm{SO}(3) \times \mathrm{PSU}(2)$ will yield a quotient with only a circle action, but whether it is an orbifold metric or not requires further investigation which we have no time to pursue here.

References

[1] M. Atiyah – "Magnetic monopoles in hyperbolic spaces", in *Vector bundles on algebraic varieties (Bombay, 1984)*, Tata Inst. Fund. Res. Stud. Math., vol. 11, Tata Inst. Fund. Res., 1987, p. 1–33.

[2] M. Atiyah & N. J. Hitchin – *The geometry and dynamics of magnetic monopoles*, M. B. Porter Lectures, Princeton University Press, 1988.

[3] A. L. Besse – *Einstein manifolds*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3), vol. 10, Springer, 1987.

[4] C. P. Boyer & K. Galicki – *Sasakian geometry*, Oxford Mathematical Monographs, Oxford University Press, 2008.

[5] S. A. Cherkis & N. J. Hitchin – "Gravitational instantons of type D_k", *Comm. Math. Phys.* **260** (2005), p. 299–317.

[6] S. A. Cherkis & A. Kapustin – "Singular monopoles and gravitational instantons", *Comm. Math. Phys.* **203** (1999), p. 713–728.

[7] A. S. Dancer – "Nahm's equations and hyper-Kähler geometry", *Comm. Math. Phys.* **158** (1993), p. 545–568.

[8] ———, "A family of hyper-Kähler manifolds", *Quart. J. Math. Oxford Ser.* **45** (1994), p. 463–478.

[9] O. Dearricott – "Positive sectional curvature on 3-Sasakian manifolds", *Ann. Global Anal. Geom.* **25** (2004), p. 59–72.

[10] L. Faybusovich & M. Gekhtman – "Poisson brackets on rational functions and multi-Hamiltonian structure for integrable lattices", *Phys. Lett. A* **272** (2000), p. 236–244.

[11] K. Grove, B. Wilking & W. Ziller – "Positively curved cohomogeneity one manifolds and 3-Sasakian geometry", *J. Differential Geom.* **78** (2008), p. 33–111.

[12] N. J. Hitchin – "The self-duality equations on a Riemann surface", *Proc. London Math. Soc.* **55** (1987), p. 59–126.

[13] ———, "Poncelet polygons and the Painlevé equations", in *Geometry and analysis (Bombay, 1992)*, Tata Inst. Fund. Res., 1995, p. 151–185.

[14] ———, "A new family of Einstein metrics", in *Manifolds and geometry (Pisa, 1993)*, Sympos. Math., XXXVI, Cambridge Univ. Press, 1996, p. 190–222.

[15] A. Kapustin & E. Witten – "Electric-magnetic duality and the geometric Langlands program", *Commun. Number Theory Phys.* **1** (2007), p. 1–236.

[16] A. Maciocia – "Metrics on the moduli spaces of instantons over Euclidean 4-space", *Comm. Math. Phys.* **135** (1991), p. 467–482.

[17] M. K. Murray, P. Norbury & M. A. Singer – "Hyperbolic monopoles and holomorphic spheres", *Ann. Global Anal. Geom.* **23** (2003), p. 101–128.

[18] O. Nash – "A new approach to monopole moduli spaces", *Nonlinearity* **20** (2007), p. 1645–1675.

[19] ______, "Singular hyperbolic monopoles", *Comm. Math. Phys.* **277** (2008), p. 161–187.

[20] S. M. SALAMON – "Differential geometry of quaternionic manifolds", *Ann. Sci. École Norm. Sup.* **19** (1986), p. 31–55.

[21] R. L. E. SCHWARZENBERGER – "Vector bundles on the projective plane", *Proc. London Math. Soc.* **11** (1961), p. 623–640.

[22] K. L. VANINSKY – "The Atiyah-Hitchin bracket and the open Toda lattice", *J. Geom. Phys.* **46** (2003), p. 283–307.

[23] ______, "The Atiyah-Hitchin bracket and the cubic nonlinear Schrödinger equation", *IMRP Int. Math. Res. Pap.* (2006), p. 17683, 1–60.

[24] W. ZILLER – "A new type of a positively curved manifold", preprint arXiv:0809.2304v2.

N. HITCHIN, Mathematical Institute, 24-29 St Giles, Oxford OX1 3LB, United Kingdom
E-mail : `hitchin@maths.ox.ac.uk` • *Url :* `http://www.maths.ox.ac.uk/~hitchin/`

Astérisque

KEFENG LIU
XIAOFENG SUN
SHING-TUNG YAU

Geometry of moduli spaces

Astérisque, tome 321 (2008), p. 31-50
<http://www.numdam.org/item?id=AST_2008__321__31_0>

© Société mathématique de France, 2008, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (http://smf4.emath.fr/Publications/Asterisque/) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM
Article numérisé dans le cadre du programme Numérisation de documents anciens mathématiques
http://www.numdam.org/

Astérisque
321, 2008, p. 31–50

GEOMETRY OF MODULI SPACES

by

Kefeng Liu, Xiaofeng Sun & Shing-Tung Yau

Dedicated to Jean Pierre Bourguignon

Abstract. — In this paper we describe some recent results on the geometry of the moduli space of Riemann surfaces. We surveyed new and classical metrics on the moduli spaces of hyperbolic Riemann surfaces and their geometric properties. We then discussed the Mumford goodness and generalized goodness of various metrics on the moduli spaces and their deformation invariance. By combining with the dual Nakano negativity of the Weil-Petersson metric we derive various consequences such that the infinitesimal rigidity, the Gauss-Bonnet theorem and the log Chern number computations.

***Résumé* (Géométrie des espaces de modules).** — Dans cet article nous décrivons certains résultats récents en géométrie de l'espace de modules des surfaces de Riemann. Nous parcourons un certain nombre de métriques classiques et nouvelles sur les les espaces de modules de surfaces de Riemann hyperboliques et leur propriétés géométriques. Ensuite nous discutons la bonté de Mumford et la bonté généralisée de différentes métriques sur l'espace de modules et leurs invariance de déformation. En combinant avec la négativité de Nakano duale de la métrique de Weil-Peterson nous en tirons différentes conséquences telles que la rigidité infinitésimale, le théorème de Gauss-Bonnet et les calculs de nombres logarithmiques de Chern.

1. Introduction

In this paper we describe our recent work on the geometry of the moduli space of Riemann surfaces $\mathcal{M}_g$. We will survey the properties of the canonical metrics especially the asymptotic behavior.

This paper is organized as follows. In the second section we will briefly recall the deformation theory of Riemann surfaces. In the third section we will recall the

2000 Mathematics Subject Classification. — 53C56, 14H15.
Key words and phrases. — Moduli spaces, good metrics, negativity.

The authors are supported by NSF grants.

© Astérisque 321, SMF 2008

Ricci and perturbed Ricci metrics as welll as the Kähler-Einstein metric which were discussed in [**5**] and [**6**].

In the fourth section we will discuss the notion of Mumford goodness and our generalizations to the p-goodness and intrinsic goodness. We then discuss the relation of the goodness and the complex Monge-Amperé equation as well as the Kähler-Ricci flow. In the last section we will discuss the applications of these fine properties of the canonical metrics.

2. Fundamentals of Teichmüller and Moduli Spaces

We briefly recall the fundamental theory of the geometry of Teichmüller and moduli spaces of hyperbolic Riemann surfaces in this section. Most of the results can be found in [**5**], [**6**], [**7**] and [**18**].

Let $\mathcal{M}_{g,k}$ be the moduli space of Riemann surfaces of genus g with k punctures such that $2g-2+k>0$. By the uniformization theorem we know there is a unique hyperbolic metric on such a Riemann surface. To simplify the computation, through out this paper, we will assume $k=0$ and $g\geq 2$ and work on $\mathcal{M}_g$. Most of the results can be trivially generalized to $\mathcal{M}_{g,k}$.

We first recall the local geometry of $\mathcal{M}_g$. For each point $s\in\mathcal{M}_g$, let X_s be the corresponding Riemann surface. By the Kodaira-Spencer deformation theory and Hodge theory, we know

$$T_s\mathcal{M}_g\cong H^1(X_s,T_{X_s})\cong H^{0,1}(X_s,T_{X_s}).$$

It follows direct from Serre duality that

$$T_s^*\mathcal{M}_g\cong H^0(X_s,K_{X_s}^2).$$

By the Riemann-Roch theorem, we know that the complex dimension of the moduli space is $n=\dim_{\mathbb{C}}\mathcal{M}_g=3g-3$. Given a Riemann surface X of genus $g\geq 2$, we denote by λ the unique hyperbolic (Kähler-Einstein) metric on X. Let z be local holomorphic coordinate on X. We normalize λ:

$$\partial_z\partial_{\overline{z}}\log\lambda=\lambda. \tag{2.1}$$

Let $\mathcal{T}_g$ be the Teichmüller space. It is well known that $\mathcal{T}_g$ is a domain of holomorphy and $\mathcal{M}_g$ is a quasi-projective orbifold. There are many canonical metrics on $\mathcal{T}_g$. These are the metrics where biholomorphisms are automatically isometries and thus these metrics descent down to $\mathcal{M}_g$.

There are three complex Finsler metrics on $\mathcal{T}_g$: The Teichmüller metric $\|\cdot\|_T$, the Kobayashi metric $\|\cdot\|_K$ and the Caratheódory metric $\|\cdot\|_C$. Each of these metrics defines a norm on the tangent space of $\mathcal{T}_g$. These metrics are non-Kähler. By the famous work of Royden we know that the Teichmüller metric coincides with the Kobayashi metric:

$$\|\cdot\|_T = \|\cdot\|_K.$$

We now describe the Kähler metrics. The first known Kähler metric is the Weil-Petersson metric ω_{WP}. Since $\mathcal{T}_g$ is a domain of holomorphy, there is a complete Kähler-Einstein metric on $\mathcal{T}_g$ due to the work of Cheng and Yau [**2**]. Since $\mathcal{M}_g$ is quasi-projective, there exist a Kähler metric on $\mathcal{M}_g$ with Poincaré growth. Furthermore, one has the Bergman metric associate to $\mathcal{T}_g$ and the Kähler metric defined by McMullen [**10**] by perturbing the Weil-Petersson metric.

In [**5**] and [**6**] we defined two new Kähler metrics: the Ricci and perturbed Ricci metrics which have very nice curvature and asymptotic properties. These metrics will be discussed in the following sections.

We now recall the construction of the Weil-Petersson metric. Let $(s_1, \cdots, s_n)$ be local holomorphic coordinates on $\mathcal{M}_g$ near a point p and let X_s be the corresponding Riemann surfaces. Let $\rho : T_s\mathcal{M}_g \to H^1(X_s, TX_s) \cong H^{0,1}(X_s, TX_s)$ be the Kodaira-Spencer map. Then the harmonic representative of $\rho\left(\frac{\partial}{\partial s_i}\right)$ is given by

$$\rho\left(\frac{\partial}{\partial s_i}\right) = \partial_{\bar z}\left(-\lambda^{-1}\partial_{s_i}\partial_{\bar z}\log\lambda\right)\frac{\partial}{\partial z}\otimes d\bar z = B_i. \tag{2.2}$$

If we let $a_i = -\lambda^{-1}\partial_{s_i}\partial_{\bar z}\log\lambda$ and let $A_i = \partial_{\bar z}a_i$, then the harmonic lift v_i of $\frac{\partial}{\partial s_i}$ is given by

$$v_i = \frac{\partial}{\partial s_i} + a_i\frac{\partial}{\partial z}. \tag{2.3}$$

The well-known Weil-Petersson metric $\omega_{WP} = \frac{\sqrt{-1}}{2}h_{i\bar j}ds_i \wedge d\bar s_j$ on $\mathcal{M}_g$ is the L^2 metric on $\mathcal{M}_g$:

$$h_{i\bar j}(s) = \int_{X_s} A_i\overline{A}_j\ dv \tag{2.4}$$

where $dv = \frac{\sqrt{-1}}{2}\lambda dz \wedge d\bar z$ is the volume form on X_s. It was proved by Ahlfors that the Ricci curvature of the Weil-Petersson metric is negative. The upper bound of the Ricci curvature of the Weil-Petersson metric was conjectured by Royden and was proved by Wolpert [**16**].

In our work [**5**] we defined the Ricci metric ω_τ:

$$\omega_\tau = -Ric\left(\omega_{WP}\right) \tag{2.5}$$

and the perturbed Ricci metric $\omega_{\tilde\tau}$:

$$\omega_{\tilde\tau} = \omega_\tau + C\omega_{WP} \tag{2.6}$$

where C is a positive constant. These new Kähler metrics have good curvature and asymptotic properties and play important roles in out study.

Now we describe the curvature formulas of the Weil-Petersson metric. Please see [**5**] and [**6**] for details. We denote by $f_{i\bar{j}} = A_i\overline{A_j}$ where each A_i is the harmonic Beltrami differential corresponding to the local holomorphic vector field $\frac{\partial}{\partial s_i}$. It is clear that $f_{i\bar{j}}$ is a function on X. We let $\square = -\partial_z\partial_{\bar{z}}$ be the Laplace operator, let $T = (\square + 1)^{-1}$ be the Green operator and let $e_{i\bar{j}} = T(f_{i\bar{j}})$. The functions $e_{i\bar{j}}$ and $f_{i\bar{j}}$ are building blocks of these curvature formula.

Theorem 2.1. — *The curvature formula of the Weil-Petersson metric was given by*

$$R_{i\bar{j}k\bar{l}} = -\int_{X_s} (e_{i\bar{j}}f_{k\bar{l}} + e_{i\bar{l}}f_{k\bar{j}})\, dv. \tag{2.7}$$

This formula was first established by Wolpert [**16**] and was generalized by Siu [**14**] and Schumacher [**13**] to higher dimensions. A short proof can be found in [**5**].

It is easy to derive information of the sign of the curvature of the Weil-Petersson metric from its curvature formula (2.7). However, the Weil-Petersson metric is incomplete and its curvature has no lower bound. Thus we need to look at its asymptotic behavior. We now recall geometric construction of the Deligne-Mumford (DM) moduli space and the degeneration of hyperbolic metrics. Please see [**5**] and [**16**] for details.

Let $\overline{\mathcal{M}}_g$ be the Deligne-Mumford compactification of $\mathcal{M}_g$ and let $D = \overline{\mathcal{M}}_g \setminus \mathcal{M}_g$. It was shown in [**3**] that D is a divisor with only normal crossings. A point $y \in D$ corresponds to a stable nodal surface X_y. A point $p \in X_y$ is a node if there is a neighborhood of p which is isometric to the germ $\{(u, v) \mid uv = 0,\ |u|, |v| < 1\} \subset \mathbb{C}^2$. Let $p_1, \cdots, p_m \in X_y$ be the nodes. X_y is stable if each connected component of $X_y \setminus \{p_1, \cdots, p_m\}$ has negative Euler characteristic.

Fix a point $y \in D$, we assume the corresponding Riemann surface X_y has m nodes. Now for any point $s \in \mathcal{M}_g$ lying in a neighborhood of y, the corresponding Riemann surface X_s can be decomposed into the thin part which is a disjoint union of m collars and the thick part where the injectivity radius with respect to the Kähler-Einstein metric is uniformly bounded from below.

There are two kinds of local holomorphic coordinate on a collar or near a node. We first recall the rs-coordinate defined by Wolpert in [**18**]. In the node case, given a nodal surface X with a node $p \in X$, we let a, b be two punctures which are glued together to form p.

Definition 2.1. — *A local coordinate chart (U, u) near a is called rs-coordinate if $u(a) = 0$ where u maps U to the punctured disc $0 < |u| < c$ with $c > 0$, and the restriction to U of the Kähler-Einstein metric on X can be written as $\frac{1}{2|u|^2(\log|u|)^2}|du|^2$. The rs-coordinate (V, v) near b is defined in a similar way.*

In the collar case, given a closed surface X, we assume there is a closed geodesic $\gamma \subset X$ such that its length $l = l(\gamma) < c_*$ where c_* is the collar constant.

Definition 2.2. — *A local coordinate chart (U,z) is called rs-coordinate at γ if $\gamma\subset U$ where z maps U to the annulus $c^{-1}|t|^{\frac12}<|z|<c|t|^{\frac12}$, and the Kähler-Einstein metric on X can be written as*

$$\frac{1}{2}\left(\frac{\pi}{\log|t|}\frac{1}{|z|}\csc\frac{\pi\log|z|}{\log|t|}\right)^2|dz|^2.$$

The existence of collar was due to Keen [**4**]. We formulate this theorem in the following:

Lemma 2.1. — *Let X be a closed surface and let γ be a closed geodesic on X such that the length l of γ satisfies $l<c_*$. Then there is a collar Ω on X with holomorphic coordinate z defined on Ω such that*

1. *z maps Ω to the annulus $\{\frac{1}{c}e^{-\frac{2\pi^2}{l}}<|z|<c\}$ for $c>0$;*
2. *the Kähler-Einstein metric on X restricted to Ω is given by*

$$\left(\frac{1}{2}u^2r^{-2}\csc^2\tau\right)|dz|^2 \tag{2.8}$$

   where $u=\frac{l}{2\pi}$, $r=|z|$ and $\tau=u\log r$;
3. *the geodesic γ is given by the equation $|z|=e^{-\frac{\pi^2}{l}}$;*
4. *the constant c has a lower bound such that the area of Ω is bounded from below by a universal constant.*

We call such a collar Ω a genuine collar.

Now we describe the pinching coordinate chart of $\overline{\mathcal{M}}_g$ near the divisor D [**18**]. Let X_0 be a nodal surface corresponding to a codimension m boundary point and let $p_1,\cdots,p_m$ be the nodes of X_0. Then $\widetilde{X}_0=X_0\setminus\{p_1,\cdots,p_m\}$ is a union of punctured Riemann surfaces. Fix rs-coordinate charts (U_i,η_i) and (V_i,ζ_i) at p_i for $i=1,\cdots,m$ such that all the U_i and V_i are mutually disjoint. Now pick an open set $U_0\subset\widetilde{X}_0$ such that the intersection of each connected component of $\widetilde{X}_0$ and U_0 is a nonempty relatively compact set and the intersection $U_0\cap(U_i\cup V_i)$ is empty for all i. Now pick Beltrami differentials $\nu_{m+1},\cdots,\nu_n$ which are supported in U_0 and span the tangent space at $\widetilde{X}_0$ of the deformation space of $\widetilde{X}_0$. Let $\Delta_\varepsilon^{n-m}\subset\mathbb{C}^{n-m}$ be the polydisc of radius ε. For $t''=(t_{m+1},\cdots,t_n)\in\Delta_\varepsilon^{n-m}$, let $\nu(t'')=\sum_{i=m+1}^n t_i\nu_i$. We assume $|t''|=(\sum_{i=m+1}^n|t_i|^2)^{\frac12}$ small enough such that $|\nu(t'')|<1$. The nodal surface $X_{0,t''}$ is obtained by solving the Beltrami equation $\overline{\partial}w=\nu(t'')\partial w$. Since $\nu(t'')$ is supported in U_0, (U_i,η_i) and (V_i,ζ_i) are still holomorphic coordinates on $X_{0,t''}$. By the theory of Ahlfors and Bers [**1**] and Wolpert [**18**] we can assume that there are constants $\delta,c>0$ such that when $|t''|<\delta$, η_i and ζ_i are holomorphic coordinates on $X_{0,t''}$ with $0<|\eta_i|<c$ and $0<|\zeta_i|<c$. Now we assume $t'=(t_1,\cdots,t_m)$ has small norm. We do the plumbing construction on $X_{0,t''}$ to obtain $X_t=X_{t',t''}$. For each $i=1,\cdots,m$, we remove the discs $\{0<|\eta_i|\le\frac{|t_i|}{c}\}$ and $\{0<|\zeta_i|\le\frac{|t_i|}{c}\}$ from $X_{0,t''}$

and identify $\{\frac{|t_i|}{c} < |\eta_i| < c\}$ with $\{\frac{|t_i|}{c} < |\zeta_i| < c\}$ by the rule $\eta_i\zeta_i = t_i$. This defines the surface X_t. The tuple $t = (t', t'') = (t_1, \cdots, t_m, t_{m+1}, \cdots, t_n)$ are the local pinching coordinates for the manifold cover of $\overline{\mathcal{M}}_g$. We call the coordinates η_i (or ζ_i) the plumbing coordinates on $X_{t,s}$ and the collar $\{\frac{|t_i|}{c} < |\eta_i| < c\}$ the plumbing collar.

Remark 2.1. — From the estimate of Wolpert [**17**], [**18**] on the length of short geodesic, we have $u_i = \frac{l_i}{2\pi} \sim -\frac{\pi}{\log|t_i|}$.

In [**5**] and [**6**] we derived the precise asymptotic of the Weil-Petersson metric and its curvature. This is one of the key components in the proof of its goodness. We have

Theorem 2.2. — *Let $(t, s) = (t_1, \cdots, t_m, s_{m+1}, \cdots, s_n)$ be the pinching coordinates near a codimension m boundary point in $\overline{\mathcal{M}}_g$. Let h be the Weil-Petersson metric. Then we have the asymptotic:*

1. $h^{i\bar{i}} = 2u_i^{-3}|t_i|^2(1 + O(u_0))$ *and* $h_{i\bar{i}} = \frac{1}{2}\frac{u_i^3}{|t_i|^2}(1 + O(u_0))$ *for* $1 \le i \le m$;
2. $h^{i\bar{j}} = O(|t_it_j|)$ *and* $h_{i\bar{j}} = O\left(\frac{u_i^3u_j^3}{|t_it_j|}\right)$, *if* $1 \le i, j \le m$ *and* $i \ne j$;
3. $h^{i\bar{j}} = O(1)$ *and* $h_{i\bar{j}} = O(1)$, *if* $m + 1 \le i, j \le n$;
4. $h^{i\bar{j}} = O(|t_i|)$ *and* $h_{i\bar{j}} = O\left(\frac{u_i^3}{|t_i|}\right)$ *if* $i \le m < j$;
5. $h^{i\bar{j}} = O(|t_j|)$ *and* $h_{i\bar{j}} = O\left(\frac{u_j^3}{|t_j|}\right)$ *if* $j \le m < i$

where $u_0 = \sum_{j=1}^m u_j + \sum_{j=m+1}^n |s_j|$.

The precise estimates of the asymptotic of the full curvature tensor of the Weil-Petersson metric, which will be used in the proof of its goodness, can be found in [**5**], [**6**] and [**7**].

3. Canonical Metrics on $\mathcal{M}_g$

Since the Weil-Petersson metric is incomplete and does not have bounded geometry, it is hard to use it to study the geometry of $\mathcal{M}_g$. In [**5**] we introduced the Ricci metric $\omega_\tau = -Ric(\omega_{WP})$ and the perturbed Ricci metric $\omega_{\widetilde{\tau}} = \omega_\tau + C\omega_{WP}$. It turns out that these new Kähler metrics have nice curvature and asymptotic properties. These new metrics are also closely related to the Kähler-Einstein metric. Especially the Ricci metric is cohomologous to the Kähler-Einstein metric as currents.

To describe the curvature formulae of the Ricci and perturbed Ricci metrics, we need to introduce several operators. We first define the operator $\xi_k : C^\infty(X_s) \to C^\infty(X_s)$ by

$$\xi_k(f) = \overline{\partial}^*(i(B_k)\partial f) = -\lambda^{-1}\partial_z(A_k\partial_z f) = -A_kK_1K_0(f) \tag{3.1}$$

where K_0, K_1 are the Maass operators [**16**], [**5**].

It was proved in[**5**] that ξ_k is the commutator of the Laplace operator and the Lie derivative in the direction v_k:

$$(\square+1)v_k - v_k(\square+1) = \square v_k - v_k\square = \xi_k. \tag{3.2}$$

We also need the commutator of the operator v_k and $\overline{v}_l$. In [**5**] we defined the operator $Q_{k\bar{l}} : C^\infty(X_s) \to C^\infty(X_s)$ by

$$Q_{k\bar{l}}(f) = [\overline{v}_l, \xi_k](f) = \overline{P}(e_{k\bar{l}})P(f) - 2f_{k\bar{l}}\square f + \lambda^{-1}\partial_z f_{k\bar{l}}\partial_{\bar{z}} f \tag{3.3}$$

where $P : C^\infty(X_s) \to \Gamma(\Lambda^{1,0}(T^{0,1}X_s))$ is the operator defined by $P(f) = \partial_z(\lambda^{-1}\partial_z f)$.

The terms appeared in the curvature formulae of the Ricci and perturbed Ricci metrics are formally symmetric with respect to indices. For convenience, we recall the symmetrization operator defined in [**5**].

Definition 3.1. — *Let U be any quantity which depends on indices i,k,α and $\bar{j},\bar{l},\bar{\beta}$. The symmetrization operator σ_1 is defined by taking the summation of all orders of the triple (i,k,α). Similarly, σ_2 is the symmetrization operator of $\bar{j}$ and $\bar{\beta}$ and $\widetilde{\sigma_1}$ is the symmetrization operator of $\bar{j}$, $\bar{l}$ and $\bar{\beta}$.*

In [**5**] we derived the curvature formulae of the new metrics. These formulae, although very complicated, are integral formulae along the fibers of the universal curve.

Theorem 3.1. — *Let $\widetilde{R}_{i\bar{j}k\bar{l}}$ and $P_{i\bar{j}k\bar{l}}$ be the curvature tensors of the Ricci and perturbed Ricci metrics respectively. In [**5**] we established the following curvature formulae of these metrics:*

$$\begin{aligned}
\widetilde{R}_{i\bar{j}k\bar{l}} =& -h^{\alpha\bar{\beta}}\left\{\sigma_1\sigma_2\int_{X_s}\left\{(T(\xi_k(e_{i\bar{j}}))\bar{\xi}_l(e_{\alpha\bar{\beta}}) + T(\xi_k(e_{i\bar{j}}))\bar{\xi}_\beta(e_{\alpha\bar{l}})\right\}\, dv\right\}\\
& - h^{\alpha\bar{\beta}}\left\{\sigma_1\int_{X_s} Q_{k\bar{l}}(e_{i\bar{j}})e_{\alpha\bar{\beta}}\, dv\right\}\\
& + \tau^{p\bar{q}}h^{\alpha\bar{\beta}}h^{\gamma\bar{\delta}}\left\{\sigma_1\int_{X_s}\xi_k(e_{i\bar{q}})e_{\alpha\bar{\beta}}\, dv\right\}\left\{\widetilde{\sigma}_1\int_{X_s}\bar{\xi}_l(e_{p\bar{j}})e_{\gamma\bar{\delta}})\, dv\right\}\\
& + \tau_{p\bar{j}}h^{p\bar{q}}R_{i\bar{q}k\bar{l}}
\end{aligned} \tag{3.4}$$

and

$$
\begin{aligned}
P_{i\bar{j}k\bar{l}} = & - h^{\alpha\bar{\beta}}\left\{\sigma_1\sigma_2\int_{X_s}\left\{T(\xi_k(e_{i\bar{j}}))\bar{\xi}_l(e_{\alpha\bar{\beta}}) + T(\xi_k(e_{i\bar{j}}))\bar{\xi}_\beta(e_{\alpha\bar{l}})\right\}\, dv\right\} \\
& - h^{\alpha\bar{\beta}}\left\{\sigma_1\int_{X_s} Q_{k\bar{l}}(e_{i\bar{j}})e_{\alpha\bar{\beta}}\, dv\right\} \\
& + \tilde{\tau}^{p\bar{q}}h^{\alpha\bar{\beta}}h^{\gamma\bar{\delta}}\left\{\sigma_1\int_{X_s}\xi_k(e_{i\bar{q}})e_{\alpha\bar{\beta}}\, dv\right\}\left\{\tilde{\sigma}_1\int_{X_s}\bar{\xi}_l(e_{p\bar{j}})e_{\gamma\bar{\delta}})\, dv\right\} \\
& + \tau_{p\bar{j}}h^{p\bar{q}}R_{i\bar{q}k\bar{l}} + CR_{i\bar{j}k\bar{l}}.
\end{aligned}
\tag{3.5}
$$

Unlike the case of the Weil-Petersson metric from which we can see the sign of the curvature directly, the above formulae are too complicated. On one hand we can see that these metrics are Kähler from these formulae. On the other hand, we need to look at the asymptotic of the curvature of these new metrics. In [**5**] and [**6**] we computed the asymptotic of these new metrics and their curvature:

Theorem 3.2. — *Let $u_0 = \sum_{j=1}^m u_j + \sum_{j=m+1}^n |s_j|$. The Ricci metric has the asymptotic:*

1. $\tau_{i\bar{i}} = \frac{3}{4\pi^2}\frac{u_i^2}{|t_i|^2}(1+O(u_0))$ *and* $\tau^{i\bar{i}} = \frac{4\pi^2}{3}\frac{|t_i|^2}{u_i^2}(1+O(u_0))$, *if* $i \leq m$;
2. $\tau_{i\bar{j}} = O\left(\frac{u_i^2u_j^2}{|t_it_j|}(u_i+u_j)\right)$ *and* $\tau^{i\bar{j}} = O(|t_it_j|)$, *if* $i,j \leq m$ *and* $i \neq j$;
3. $\tau_{i\bar{j}} = O\left(\frac{u_i^2}{|t_i|}\right)$ *and* $\tau^{i\bar{j}} = O(|t_i|)$, *if* $i \leq m$ *and* $j \geq m+1$;
4. $\tau_{i\bar{j}} = O(1)$, *if* $i,j \geq m+1$.

The holomorphic sectional curvature of the Ricci metric has the asymptotic:

1. $\widetilde{R}_{i\bar{i}i\bar{i}} = -\frac{3u_i^4}{8\pi^4|t_i|^4}(1+O(u_0))$ *if* $i \leq m$;
2. $\widetilde{R}_{i\bar{i}i\bar{i}} = O(1)$ *if* $i > m$.

We also have a weak curvature estimate of the Ricci metric. Let

$$
\Lambda_i = \begin{cases} \frac{u_i}{|t_i|} & \text{if } i \leq m \\ 1 & \text{if } i > m. \end{cases}
$$

Then

1. $\widetilde{R}_{i\bar{j}k\bar{l}} = O(1)$ *if* $i,j,k,l > m$;
2. $\widetilde{R}_{i\bar{j}k\bar{l}} = O(\Lambda_i\Lambda_j\Lambda_k\Lambda_l)O(u_0)$ *if at least one of these indices* i,j,k,l *is less than or equal to* m *and they are not all equal to each other.*

The asymptotic of the perturbed Ricci metric and its curvature can be found in [**5**] and [**6**]. Also, precise estimates of the full curvature tensor of the Ricci and perturbed Ricci metrics, which will also be used in the proof of their goodness, can be found in [**7**] and [**8**].

As a simple corollary of the curvature formulae and asymptotic analysis, in [**5**] we first proved the equivalence of canonical metrics on $\mathcal{M}_g$:

Theorem 3.3. — *All the canonical metrics on the moduli space $\mathcal{M}_g$: the Teichmüller-Kobayashi metric, the Carathéodory metric, the induced Bergman metric, the asymptotic Poincaré metric, the McMullen metric, the Ricci metric, the perturbed Ricci metric and the Kähler-Einstein metric are equivalent.*

The new metrics we defined have nice curvature properties which can be used to control the Kähler-Einstein metric. In [**5**] and [**6**] we proved

Theorem 3.4. — *Let $\mathcal{M}_g$ be the moduli space of genus $g \geq 2$ Riemann surfaces. Then*

- *The Ricci and perturbed Ricci metrics are complete Kähler metrics with Poincaré growth.*
- *The Ricci and perturbed Ricci metrics as well as the Kähler-Einstein metric have bounded geometry on the Teichmüller space $\mathcal{T}_g$.*
- *The Ricci and holomorphic sectional curvatures of the perturbed Ricci metric are bounded from above and below by negative constants.*
- *All the covariant derivatives of the curvature of the Kähler-Einstein metric are bounded.*

The finer asymptotic of these metrics, their local connection forms and curvature forms will lead to the Mumford goodness which is a set of growth conditions of these metrics and their derivatives modeled on the Poincaré metric on the punctured disk. These conditions will guarantee the behavior of the Chern forms of these complete metrics.

4. Notions of Goodness

In this section we will discuss various notions of goodness. The central idea is to control the Chern forms, as currents, of singular Hermitian metrics on holomorphic vector bundles over quasi-projective varieties.

Let M be a compact complex manifold and let (E, h) be a Hermitian vector bundle over M. We denote by $(z_1, \cdots, z_n)$ the local holomorphic coordinates on M and by $(e_1, \cdots, e_m)$ the local holomorphic frame of E. Let $h_{\alpha\overline{\beta}} = h(e_\alpha, e_\beta)$ and denote by θ and Θ the local connection and curvature forms of h. Then we have $\theta_\alpha^\gamma = \partial_i h_{\alpha\overline{\beta}} h^{\gamma\overline{\beta}} dz_i$ and $\Theta_\alpha^\gamma = R_{\alpha i\overline{j}}^\gamma dz_i \wedge d\overline{z}_j$ where $\partial_i = \frac{\partial}{\partial z_i}$ and

$$R_{\alpha i\overline{j}}^\gamma = -h^{\gamma\overline{\beta}} \left(\partial_i \partial_{\overline{j}} h_{\alpha\overline{\beta}} - h^{\tau\overline{\delta}} \partial_i h_{\alpha\overline{\delta}} \partial_{\overline{j}} h_{\tau\overline{\beta}} \right).$$

The k-th Chern form $c_k(h)$ of h is given by the coefficient of the term t^k in the polynomial $\det\left(I+\frac{\sqrt{-1}}{2\pi}\Theta\right)$. It is well known that

$$[c_k(h)] = c_k(E) \tag{4.1}$$

as cohomology classes. However, this is no long true in general when M is noncompact. One needs growth conditions on h and its derivatives. The class of noncompact manifolds we are interested in is the quasi-projective manifolds.

The first condition was given by Mumford in [**11**] which we will describe now. Let $\overline{X}^n$ be a projective manifold of complex dimension n and let $D\subset\overline{X}$ be a divisor of normal crossings. Let $X=\overline{X}\setminus D$.

We cover a neighborhood of $D\subset\overline{X}$ by finitely many polydiscs

$$\{U_\alpha=(\Delta^n,(z_1,\cdots,z_n))\}_{\alpha\in A}$$

such that $V_\alpha=U_\alpha\setminus D=(\Delta^*)^m\times\Delta^{k-m}$. Namely, $U_\alpha\cap D=\{z_1\cdots z_m=0\}$. We let $U=\bigcup_{\alpha\in A}U_\alpha$ and $V=\bigcup_{\alpha\in A}V_\alpha$. On each V_α we have the local Poincaré metric

$$\omega_{P,\alpha}=\frac{\sqrt{-1}}{2}\left(\sum_{i=1}^m\frac{1}{2|z_i|^2\left(\log|z_i|\right)^2}dz_i\wedge d\overline{z}_i+\sum_{i=m+1}^n dz_i\wedge d\overline{z}_i\right).$$

The Mumford goodness is a growth condition on differential forms. We recall the following definitions from [**11**]:

Definition 4.1. — *Let η be a smooth local p-form defined on V_α.*

- *We say η has Poincaré growth if there is a constant $C_\alpha>0$ depending on η such that*

$$|\eta(t_1,\cdots,t_p)|^2\le C_\alpha\prod_{i=1}^p\|t_i\|^2_{\omega_{P,\alpha}}$$

for any point $z\in V_\alpha$ and $t_1,\cdots,t_p\in T_zX$.
- *We say η is good if both η and $d\eta$ have Poincaré growth.*

Now let $\overline{E}$ be a holomorphic vector bundle of rank k over $\overline{X}$ and let E be the restriction of E to X. Let h be a Hermitian metric on E which may be singular along the divisor D.

Definition 4.2. — *An Hermitian metric h on E is good if for all $z\in V$, assuming $z\in V_\alpha$, and for all basis $(e_1,\cdots,e_k)$ of $\overline{E}$ over U_α we have*

- $\left|h_{\alpha\overline{\beta}}\right|,(\det h)^{-1}\le C\left(\sum_{i=1}^m\log|z_i|\right)^{2p}$ *for some $C>0$ and $p\ge1$.*
- *The local 1-forms $\left(\partial h\cdot h^{-1}\right)_\alpha^\gamma$ are good on V_α. Namely the local connection and curvature forms of h have Poincaré growth.*

Remark 4.1. — It is easy to see that the definition of Poincaré growth is independent of the choice of local data.

We collect the main properties of good metrics in the following theorem which is due to Mumford. Please see [**11**] for details.

Theorem 4.1. — *Let X and E be as above. Then*

- *A form $\eta \in A^p(X)$ with Poincaré growth defines a p-current $[\eta]$ on $\overline{X}$. In fact we have*
$$\int_X |\eta \wedge \xi| < \infty$$
for any $\xi \in A^{k-p}(\overline{X})$.
- *If both $\eta \in A^p(X)$ and $\xi \in A^q(X)$ have Poincaré growth, then $\eta \wedge \xi$ has Poincaré growth.*
- *For a good form $\eta \in A^p(X)$, we have $d[\eta] = [d\eta]$.*
- *Given an Hermitian metric h on E, there is at most one extension $\overline{E}$ of E to $\overline{X}$ such that h is good.*
- *If h is a good metric on E, the Chern forms $c_i(E,h)$ are good forms. Furthermore, as currents, they represent the corresponding Chern classes $c_i(\overline{E}) \in H^{2i}(\overline{X}, \mathbb{C})$.*

The most important feature of a good metric on E is that we can compute the Chern classes of $\overline{E}$ via the Chern forms of h as currents. Namely, with the growth assumptions on the metric and its derivatives, we can integrate by part, so Chern-Weil theory still holds. However, the Mumford goodness is very strong and hard to check. Also, there are only few examples. In [**7**] we showed that the canonical metrics on the moduli space of Riemann surfaces are Mumford good.

We now give weaker notions of goodness which still have the major properties of Mumford good metrics. The definition of Mumford on Poincaré growth and good forms is quite local. We first give a global formulation of these growth conditions. Please see [**7**] for details.

We call a Kähler metric ω_P on X a Poincaré type metric if ω_P is equivalent to $\omega_{P,\alpha}$ when restricted to V_α.

Remark 4.2. — It is easy to see that

- Any two Poincaré type metrics are equivalent.
- The quasi-projective Kähler manifold (X, ω_P) is complete and has finite volume.

Our first observation is

Lemma 4.1. — *A smooth form $\eta \in A^q(X)$ has Poincaré growth if and only if $\|\eta\|_{\omega_P} \leq C$ for some constant C and a Poincaré metric on X. Namely η has L^∞ bound with respect to Poincaré metrics.*

Parallel to the Poincaré growth and good forms, we know define the p-growth and p-good forms by replacing the L^∞ norm by L^p norm.

Definition 4.3. — *Let $p \geq 1$ be a real number. A differential form $\eta \in A^q(X)$ has p-growth if*

$$\|\eta\|_{\omega_P} \in L^p(X, \omega_P).$$

The form η is p-good if both η and $d\eta$ have p-growth.

We note here that the above definition is independent of the choice of ω_P. To study the currents of p-growth forms, we need a special cut-off functions. In [**9**] we construct a desirable cut-off function:

Proposition 4.1. — *There exists $\varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$, there is a function ρ_ε such that*

1. $0 \leq \rho_\varepsilon \leq 1$.
2. *For any open neighborhood N of D in $\overline{X}$, there is $\varepsilon > 0$ such that* $\operatorname{supp}(1-\rho_\varepsilon) \subset N$.
3. *For each $\varepsilon > 0$, there is a neighborhood N of D such that $\rho_\varepsilon|_N \equiv 0$.*
4. $\rho_{\varepsilon'} \geq \rho_\varepsilon$ *for* $\varepsilon' \leq \varepsilon$.
5. *There is a constant C, independent of ε such that*
$$-C\omega_P \leq \sqrt{-1}\partial\overline{\partial}\rho_\varepsilon \leq C\omega_P$$
and
$$|\nabla'\rho_\varepsilon| \leq C.$$
6. $$\lim_{\varepsilon\to 0}\rho_\varepsilon = 1.$$

The p-good forms have similar behavior to good forms.

Lemma 4.2. — *For $p \geq 1$, if $\eta \in A^q(X)$ has p-growth, then η defines a q-current. If η is p-good, then $d[\eta] = [d\eta]$. Furthermore, if η, η' have p and p' growth respectively, then $\eta \wedge \eta'$ has $\frac{pp'}{p+p'}$ growth.*

Now we can generalize the Mumford good metrics. Similar to Definition 4.2 we define

Definition 4.4. — *A Hermitian metric h on E is p-good if*

1. $\left|h_{\alpha\overline{\beta}}\right|, (\det h)^{-1} \leq C\left(\sum_{i=1}^m \log|z_i|\right)^{2s}$ *for some $C > 0$ and $s \geq 1$.*
2. *The local 1-forms $\left(\partial h \cdot h^{-1}\right)_\alpha^\gamma$ are p-good on V_α.*

We have

Theorem 4.2. — *For p large enough, if the Hermitian metric h on E is p-good, then the Chern forms of h represent the corresponding Chern classes of $\overline{E}$:*

$$[c_i(h)] = c_i(\overline{E}) \in H^{2i}(\overline{X}, \mathbb{C}).$$

The p-goodness is essentially integral conditions which is much easier to check than the Mumford goodness. Since the most important part of controlling the growth the singular metric h is to study its Chern forms, we can just take this as a definition.

Definition 4.5. — *A Hermitian metric h on E is intrinsically good if the Chern form $c_i(h)$ defines a $2i$-current and*

$$[c_i(h)] = c_i(\overline{E}).$$

It turns out that the intrinsic goodness is preserved by the continuity method and the Kähler-Ricci flow. We have the following relation:

good metrics $\Rightarrow$ p-good metrics for large p $\Rightarrow$ intrinsic good metrics

There are only few examples of Mumford good metrics. In [**11**] Mumford showed that the invariant metrics on Hermitian symmetric spaces are good. Later Wolpert [**18**] showed that the hyperbolic metric on the relative dualizing sheaf is good. In [**15**] Trapani proved that the metric on the logarithmic tangent bundle of $\overline{\mathcal{M}}_g$ is good. In the last cases, the holomorphic bundle involoved are line bundles. In [**7**] and [**8**] we prove:

Theorem 4.3. — *Let $\overline{E} = T_{\overline{\mathcal{M}}_g}(-\log D)$ be the logarithmic tangent bundle of the DM moduli space and let $E = \overline{E}\mid_{\mathcal{M})g}$. Then the metrics on E induced by the Weil-Petersson metric, the Ricci and perturbed Ricci metrics are good in the sense of Mumford.*

The moduli space $\mathcal{M}_g$ together with these metrics provide very interesting examples of good geometry. It is more interesting to study the goodness of the Kähler-Einstein metric since many consequences follows.

5. The Monge-Amperé Equation and the Goodness

As we described in last section, the Chern forms of various good singular Hermitian metrics represent corresponding Chern classes. Thus it is important to study the goodness of canonical metrics on the quasi-projective manifold X such as the the Kähler-Einstein metric.

Let X be a quasi-projective manifold obtained by removing a normal crossing divisor D from a projective manifold $\overline{X}$. Let $\overline{E} = T_{\overline{X}}(-\log D)$ be the logarithmic tangent bundle and let E be the restriction of $\overline{E}$ to X. In this section we will consider Hermitian metrics on E induced from a Kähler metric on X.

Let ω_g be a Kähler metric on X. Let $U,(z_1,\cdots,z_n)$ be a chart of $\overline{X}$ such that $U\cap D=\{z_1\cdots z_m=0\}$. It is clear that a local frame of E is given by

$$\underline{e}=(e_1,\cdots,e_n)=\left(z_1\frac{\partial}{\partial z_1},\cdots,z_m\frac{\partial}{\partial z_m},\frac{\partial}{\partial z_{m+1}},\cdots,\frac{\partial}{\partial z_n}\right).$$

Let h be the metric on E induced by ω_g. Then under this frame we have

$$g^*_{i\overline{j}}=\begin{cases} z_i\overline{z}_j g_{i\overline{j}} & i,j\le m\\ z_i g_{i\overline{j}} & i\le m<j\\ \overline{z}_j g_{i\overline{j}} & j\le m<i\\ g_{i\overline{j}} & i,j>m.\end{cases} \tag{5.1}$$

By using the above frame and the local formula of the metric h, we have

Lemma 5.1. — *The Chern forms of h and ω_g coincide. Namely,*

$$c_k(h)=c_k(g).$$

If we assume the background metric ω_g has Poincaré growth, then the induced metric h is good will imply that the metric g has bounded curvature. The converse is not true in general. But we can bound the Chern froms:

Lemma 5.2. — *If ω_g is a Kähler metric on X with bounded curvature and has Poincaré growth, then the Chern forms of the metric h on E induced by ω_g are good in the sense of Mumford.*

In the case when h is induced by the Kähler-Einstein metric on X, to ensure the Chern forms of h represent the correct Chern classes, we need control on the Kähler-Einstein metric.

The following result is a weaker version of our work. We state this version to illustrate the ideas.

Theorem 5.1. — *Let $\overline{X}$ be a projective manifold with $\dim_{\mathbb{C}}\overline{X}=n$. Let $D\subset\overline{X}$ be a divisor of normal crossings, let $X=\overline{X}\setminus D$, let $\overline{E}=T_{\overline{X}}(-\log D)$ and let $E=\overline{E}\,|_X$.*

Let ω_g be a Kähler metric on X with bounded curvature and Poincaré growth. Assume $Ric(\omega_g)+\omega_g=\partial\overline{\partial}f$ where f is a bounded smooth function. Then

- *There exist a unique Kähler-Einstein metric ω_{KE} on X with Poincaré growth.*
- *The curvature and covariant derivatives of curvature of the Kähler-Einstein metric are bounded.*
- *If ω_g is intrinsic good, then ω_{KE} is intrinsic good. Furthermore, all metrics along the paths of continuity and Kähler-Ricci flow are intrinsic good.*

Remark 5.1. — In [**8**] we will prove a stronger version of the above theorem by replacing the L^∞ bound of the Ricci potential f by L^p bound.

On the other hand, if we know the existence and properties of the Kähler-Einstein metric by other means, we can prove the above theorem by only assuming $f \in L^1(X, \omega_g)$.

Theorem 5.2. — *Let ω_g be a Kähler metric on X with Poincaré growth and bounded curvature. Assume $Ric(\omega_g) + \omega_g = \partial\overline{\partial} f$ where $f \in L^1(X, \omega_g)$ and there exist a Kähler-Einstein metric on X which is equivalent to ω_g. If ω_g is intrinsically good, then ω_{KE} is also intrinsically good.*

By combining Theorem 3.3, 4.3 and 5.2 we have

Theorem 5.3. — *Let ρ be the metric on the logarithmic tangent bundle over the moduli space $\mathcal{M}_g$ induced by the Kähler-Einstein metric on $\mathcal{M}_g$. Then ρ is intrinsically good.*

The intrinsic goodness of the Kähler-Einstein metric will imply stability of the log tangent bundle and a strong Chern number inequality. As a consequence we proved in [**6**] and [**7**]

Theorem 5.4. — *The logarithmic tangent bundle $\overline{E}$ of the DM moduli space $\overline{\mathcal{M}}_g$ is stable with respect to the canonical polarization. Furthermore, we have*

$$c_1(\overline{E})^2 \leq \frac{6g-4}{3g-3} c_2(\overline{E}).$$

We now briefly describe the proof of these two theorems. Please see [**7**] and [**8**] for details.

We first deform the background ω_g along the Kähler-Ricci flow for short time such that all the covariant derivatives of ω_g are bounded. In the case, the intrinsic goodness of ω_g is also preserved.

The existence of the Kähler-Einstein metric follows from the C^k estimates of the complex Monge-Amperé equation

$$\frac{(\omega_g + \partial\overline{\partial}\varphi)^n}{\omega_g^n} = e^{\varphi + f}$$

where we use Yau's generalized maximum principle. To prove that the intrinsic goodness of ω_g is preserved along the path of continuity, if we denote by g' the Kähler-Einstein metric, we need to show that

$$c_k(g) - c_k(g')$$

is the 0-current. Let R, R', Γ, Γ' be the curvatures and connections of g and g' respectively.

We first deal with renormalized Chern character forms. For a Hermitian metric h on a holomorphic vector bundle E with curvature Θ, the k-th Chern character form is defined by

$$ch_k(h) = Tr\left(\frac{\sqrt{-1}}{2\pi}\Theta\right)^k.$$

To simplify the notation, we drop the constant $\frac{\sqrt{-1}}{2\pi}$. As differential forms we have

$$ch_k(g) - ch_k(g') = d\left(\mathrm{Tr}\sum_{i=0}^{k-1} R^{k-1-i} \wedge (\Gamma - \Gamma') \wedge R'^i\right)$$

and

$$\Gamma'^p_{ik} - \Gamma^p_{ik} = g'^{p\bar{q}}\varphi_{;i\bar{q}k}.$$

By the C^2 and C^3 estimate we know

$$\mathrm{Tr}\left(\sum_{i=0}^{k-1} R^{k-1-i} \wedge (\Gamma - \Gamma') \wedge R'^i\right)$$

has Poincaré growth. Since both $ch_k(g)$ and $ch_k(g')$ has Poincaré growth it is easy to see $ch_k(g) - ch_k(g')$ is the 0 current.

This is proved by integration by part where we use the cut-off function as in Proposition 4.1. Finally, by the expression of $c_k(g)$ and $c_k(g')$ via $ch_k(g)$ and $ch_k(g')$ we see that $c_k(g) - c_k(g')$ is also the 0 current.

6. Rigidity and Gauss-Bonnet Theorem

In this final section we discuss the applications of the curvature and asymptotic properties of the canonical metrics on the curve moduli $\mathcal{M}_g$.

The Weil-Petersson metric has many negative curvature properties. Ahlfors showed that its Riemannian sectional curvature is negative. Later, it was proved by Wolpert that the bisectional curvature of the Weil-Petersson metric is negative. In [**12**] Schumacher showed that the curvature of the Weil-Petersson metric is strongly negative in the sense of Siu. In [**7**] we proved that the Weil-Petersson metric is dual-Nakano negative from which we will derive Nakano-type vanishing theorems.

We first recall the concept of dual Nakano negativity. Let (E^m, h) be a holomorphic vector bundle with a Hermitian metric over a complex manifold M^n. The curvature of E is given by

$$P_{i\bar{j}\alpha\bar{\beta}} = -\partial_\alpha\partial_{\bar{\beta}}h_{i\bar{j}} + h^{p\bar{q}}\partial_\alpha h_{i\bar{q}}\partial_{\bar{\beta}}h_{p\bar{j}}.$$

(E, h) is Nakano semi-positive if the curvature P defines a semi-positive form on the bundle $E \otimes T_M$. Namely,

$$P_{i\bar{j}\alpha\bar{\beta}}C^{i\alpha}\overline{C^{j\beta}} \geq 0 \tag{6.1}$$

for all $m \times n$ complex matrix C. The metric h is Nakano positive if (6.1) is a strict inequality whenever $C \neq 0$. E is dual Nakano (semi) negative if the dual bundle with the induced metric (E^*, h^*) is Nakano (semi) positive.

In [**7**] we showed

Theorem 6.1. — *Let $\mathcal{M}_g$ be the moduli space of Riemann surfaces of genus $g \geq 2$. Then $(T_{\mathcal{M}_g}, \omega_{WP})$ is dual Nakano negative.*

Let us briefly describe the idea. Please see [**7**] for details. By the definition of the dual-Nakano negativity, we only need to show that $(T^*\mathcal{M}_g, h^*)$ is Nakano positive. Let $R_{i\bar{j}k\bar{l}}$ be the curvature of $\mathcal{M}_g$ and $P_{i\bar{j}k\bar{l}}$ be the curvature of the cotangent bundle. We first have

$$P_{m\bar{n}k\bar{l}} = -h^{i\bar{n}} h^{m\bar{j}} R_{i\bar{j}k\bar{l}}.$$

Thus if we let $a_{kj} = \sum_m h^{m\bar{j}} C^{mk}$, we have

$$P_{m\bar{n}k\bar{l}} C^{mk} \overline{C^{nl}} = -\sum_{i,j,k,l} R_{i\bar{j}k\bar{l}} a_{kj} \overline{a_{li}} = -\sum_{i,j,k,l} R_{k\bar{j}i\bar{l}} a_{kj} \overline{a_{li}} = -\sum_{i,j,k,l} R_{i\bar{j}k\bar{l}} a_{ij} \overline{a_{lk}}.$$

Recall that at $X \in \mathcal{M}_g$ we have

$$R_{i\bar{j}k\bar{l}} = -\int_X \left(e_{i\bar{j}} f_{k\bar{l}} + e_{i\bar{l}} f_{k\bar{j}}\right) dv.$$

By combining the above two formulae, to prove that the WP metric is Nakano negative is equivalent to show that

$$\int_X \left(e_{i\bar{j}} f_{k\bar{l}} + e_{i\bar{l}} f_{k\bar{j}}\right) a_{ij} \overline{a_{lk}}\, dv \geq 0 \tag{6.2}$$

and the left hand side of the above formula is strictly positive if $A = [a_{ij}] \neq 0$.

We now describe the proof with the assumption that the matrix $[a_{ij}]$ is invertible. The general case can be found in [**7**] which follows from the same idea.

Recall that if we let $\square = -\lambda^{-1} \partial_z \partial_{\bar{z}}$ be the Laplace operator with respect to the KE metric λ on X and let $T = (\square + 1)^{-1}$, then $e_{i\bar{j}} = T\left(f_{i\bar{j}}\right)$ where $f_{i\bar{j}} = A_i \overline{A_j}$ and A_i is the harmonic representative of the Kodaira-Spencer class of $\frac{\partial}{\partial t_i}$ where $(t_1, \cdots, t_n)$ are local coordinates on $\mathcal{M}_g$ and z is the local coordinate on X_t.

Let $B_j = \sum_{i=1}^n a_{ij} A_i$. Then the inequality (6.2) is equivalent to

$$\sum_{j,k} R(B_j, \overline{B_k}, A_k, \overline{A_j}) = \sum_{j,k} \int_X \left(T\left(B_j \overline{A_j}\right) A_k \overline{B_k} + T\left(B_j \overline{B_k}\right) A_k \overline{A_j}\right) dv \geq 0. \tag{6.3}$$

Since $\{A_k\}$ is a basis of the space $H^{0,1}(X, T_X)$ and the matrix $\{a_{ij}\}$ is an arbitrary invertible matrix, we need to show that the inequality (6.3) holds for any two bases $\{A_i\}$ and $\{B_i\}$. Of course we can choose one basis, say $\{A_i\}$, and let the other basis vary freely.

Now we prove the inequality (6.3). Let $\mu = \sum_j B_j \overline{A_j}$. Then the first term in (6.3) is

$$\sum_{j,k} \int_X T\left(B_j \overline{A_j}\right) A_k \overline{B_k}\, dv = \int_X T(\mu)\overline{\mu}\, dv \geq 0.$$

To check the second term, we let $G(z,w)$ be the Green's function of the operator T. Namely, for any function $f \in C^\infty(X)$, we have $T(f) = \int_X G(z,w) f(w)\, dv(w)$. Now we let

$$H(z,w) = \sum_j \overline{A_j}(z) B_j(w).$$

We know the second term of (6.3) is

$$\begin{aligned}\sum_{j,k} \int_X T\left(B_j \overline{B_k}\right) A_k \overline{A_j}\, dv &= \sum_{j,k} \int_X \int_X G(z,w) B_j(w) \overline{B_k}(w) A_k(z) \overline{A_j}(z)\, dv(w) dv(z) \\ &= \int_X \int_X G(z,w) H(z,w) \overline{H}(z,w) dv(w) dv(z) \geq 0\end{aligned}$$

where the last inequality follows from the fact that the Green's function G is non-negative which was proved by Wolpert in [**16**].

The asymptotic of Weil-Petersson , Ricci and perturbed Ricci metrics give us good control of the L^2 cohomology with bundle twist. In [**7**] we showed

Theorem 6.2. — *Let $\mathcal{M}_g$ be the moduli space of genus g curves and let $\overline{\mathcal{M}}_g$ be its Deligne-Mumford compactification. Then*

$$H^*_{(2)}\left((\mathcal{M}_g, \omega_\tau), (T_{\mathcal{M}_g}, \omega_{WP})\right) \cong H^*(\overline{\mathcal{M}}_g, T_{\overline{\mathcal{M}}_g}(-\log D)).$$

Combining with the dual-Nakano negativity of the Weil-Petersson metric we have

Theorem 6.3. — *The Chern numbers of the log cotangent bundle $T^*_{\overline{\mathcal{M}}_g}(\log D)$ of the moduli spaces of Riemann surfaces are positive.*

More importantly, we proved that the complex structure of the moduli space is infinitesimally rigid:

Theorem 6.4. — *When $q \neq 3g-3$, the L^2 cohomology groups vanish*

$$H^{0,q}_{(2)}\left((\mathcal{M}_g, \omega_\tau), \left(T_{\overline{\mathcal{M}}_g}(-\log D), \omega_{WP}\right)\right) = 0.$$

One of the most important consequence of the curvature properties and goodness of the Ricci, perturbed Ricci and Kähler-Einstein metrics is the Gauss-Bonnet Theorem on $\mathcal{M}_g$. Together with L. Ji, we showed in [**7**]

Theorem 6.5. — *(Liu, Ji, Sun, Yau) The Gauss-Bonnet Theorem hold on the moduli space equipped with the Ricci, perturbed Ricci or Kähler-Einstein metrics:*

$$\int_{\mathcal{M}_g} c_n(\omega_\tau) = \int_{\mathcal{M}_g} c_n(\omega_{\tilde{\tau}}) = \int_{\mathcal{M}_g} c_n(\omega_{KE}) = \chi(\mathcal{M}_g) = \frac{B_{2g}}{4g(g-1)}.$$

Here $\chi(\mathcal{M}_g)$ is the orbifold Euler characteristic of $\mathcal{M}_g$ and $n = 3g-3$.

The computation of the Euler characteristic of the moduli space is due to Zagier.

In the proof of the Gauss-Bonnet Theorem we used the fact that the curvature of the Ricci, perturbed Ricci and Kähler-Einstein metrics are bounded. However, the curvature of the Weil-Petersson metric is not bounded. However, as an application of the Mumford goodness of the Weil-Petersson metric and the Ricci metric we have

Theorem 6.6. — *We have*

$$\chi(T_{\overline{\mathcal{M}}_g}(-\log D)) = \int_{\mathcal{M}_g} c_n(\omega_{\tau}) = \int_{\mathcal{M}_g} c_n(\omega_{WP}) = \frac{B_{2g}}{4g(g-1)}$$

where $n = 3g-3$.

This theorem gave us the first log Chern number of the DM moduli space $\overline{\mathcal{M}}_g$.

Corollary 6.1. — *We have*

$$\chi(\overline{\mathcal{M}}_g, T_{\overline{\mathcal{M}}_g}(-\log D)) = \chi(\mathcal{M}_g) = \frac{B_{2g}}{4g(g-1)}.$$

References

[1] L. AHLFORS & L. BERS – "Riemann's mapping theorem for variable metrics", *Ann. of Math.* **72** (1960), p. 385–404.

[2] S.-Y. CHENG & S.-T. YAU – "On the existence of a complete Kähler metric on non-compact complex manifolds and the regularity of Fefferman's equation", *Comm. Pure Appl. Math.* **33** (1980), p. 507–544.

[3] P. DELIGNE & D. MUMFORD – "The irreducibility of the space of curves of given genus", *Publ. Math. I.H.É.S.* **36** (1969), p. 75–109.

[4] L. KEEN – "Collars on Riemann surfaces", in *Discontinuous groups and Riemann surfaces (Proc. Conf., Univ. Maryland, College Park, Md., 1973)*, Princeton Univ. Press, 1974, p. 263–268. Ann. of Math. Studies, No. 79.

[5] K. LIU, X. SUN & S.-T. YAU – "Canonical metrics on the moduli space of Riemann surfaces. I", *J. Differential Geom.* **68** (2004), p. 571–637.

[6] ———, "Canonical metrics on the moduli space of Riemann surfaces. II", *J. Differential Geom.* **69** (2005), p. 163–216.

[7] ———, "Good metrics on the moduli space of Riemann surface I", preprint, 2008.

[8] ———, "Good metrics on the moduli space of Riemann surface II", preprint, 2008.

[9] Z. LU & X. SUN – "On the Weil-Petersson volume and the first chern class *o* of the Moduli Spaces of Calabi -Yau Manifolds", submitted to *Communications in Mathematical Physics.*

[10] C. T. MCMULLEN – "The moduli space of Riemann surfaces is Kähler hyperbolic", *Ann. of Math.* **151** (2000), p. 327–357.

[11] D. MUMFORD – "Hirzebruch's proportionality theorem in the noncompact case", *Invent. Math.* **42** (1977), p. 239–272.

[12] G. SCHUMACHER – "Harmonic maps of the moduli space of compact Riemann surfaces", *Math. Ann.* **275** (1986), p. 455–466.

[13] ———, "The curvature of the Petersson-Weil metric on the moduli space of Kähler-Einstein manifolds", in *Complex analysis and geometry*, Univ. Ser. Math., Plenum, 1993, p. 339–354.

[14] Y. T. SIU – "Curvature of the Weil-Petersson metric in the moduli space of compact Kähler-Einstein manifolds of negative first Chern class", in *Contributions to several complex variables*, Aspects Math., E9, Vieweg, 1986, p. 261–298.

[15] S. TRAPANI – "On the determinant of the bundle of meromorphic quadratic differentials on the Deligne-Mumford compactification of the moduli space of Riemann surfaces", *Math. Ann.* **293** (1992), p. 681–705.

[16] S. A. WOLPERT – "Chern forms and the Riemann tensor for the moduli space of curves", *Invent. Math.* **85** (1986), p. 119–145.

[17] ———, "Asymptotics of the spectrum and the Selberg zeta function on the space of Riemann surfaces", *Comm. Math. Phys.* **112** (1987), p. 283–315.

[18] ———, "The hyperbolic metric and the geometry of the universal curve", *J. Differential Geom.* **31** (1990), p. 417–472.

K. LIU, Center of Mathematical Sciences, Zhejiang University, Hangzhou, China; Department of Mathematics, University of California at Los Angeles, USA

X. SUN, Department of Mathematics, Lehigh University, USA

S.-T. YAU, Center of Mathematical Sciences, Zhejiang University, Hangzhou, China; Department of Mathematics, Harvard University, USA

Astérisque

ROBERT L. BRYANT

Gradient Kähler Ricci solitons

Astérisque, tome 321 (2008), p. 51-97
<http://www.numdam.org/item?id=AST_2008__321__51_0>

© Société mathématique de France, 2008, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (http://smf4.emath.fr/Publications/Asterisque/) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM
Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

Astérisque
321, 2008, p. 51–97

GRADIENT KÄHLER RICCI SOLITONS

by

Robert L. Bryant

To Jean Pierre Bourguignon, on the occasion of his 60th birthday.

Abstract. — Some observations about the local and global generality of gradient Kähler Ricci solitons are made, including the existence of a canonically associated holomorphic volume form and vector field, the local generality of solutions with a prescribed holomorphic volume form and vector field, and the existence of Poincaré coordinates in the case that the Ricci curvature is positive and the vector field has a fixed point.

***Résumé* (Solitons gradients de Kähler-Ricci).** — Nous proposons quelques observations sur les généralités locale et globale des solitons gradients de Kähler-Ricci, y compris l'existence d'une forme de volume holomorphe et d'un champ de vecteurs canoniquement associés, la généralité locale de solutions pour une forme de volume holomorphe et un champ de vecteurs donnés, et l'existence de coordonnées de Poincaré dans le cas où la courbure de Ricci est positive et le champ de vecteurs a un point fixe.

1. Introduction and Summary

This article concerns the local and global geometry of gradient Kähler Ricci solitons, i.e., Kähler metrics g on a complex n-manifold M that admit a *Ricci potential*, i.e., a function f such that $\mathrm{Ric}(g) = \nabla^2 f$ (where ∇ denotes the Levi-Civita connection of M.

These metrics arise as limiting metrics in the study of the Ricci flow $g_t = -2\,\mathrm{Ric}(g)$ applied to Kähler metrics. Under the Ricci flow, a gradient Kähler Ricci soliton g_0

2000 Mathematics Subject Classification. — 53C55, 58G11.
Key words and phrases. — Ricci flow, solitons, normal forms.

Thanks to Duke University for its support via a research grant, to the NSF for its support via grants DMS-8905207, DMS-0103884, and DMS-0604195 and to Columbia University for its support via an Eilenberg Visiting Professorship.

© Astérisque 321, SMF 2008

evolves by flowing under the vector field ∇f, i.e.,

$$g(t) = \exp_{(-t\nabla f)}{}^{*}(g_0). \tag{1.1}$$

In particular, if the flow of ∇f is complete, then the Ricci flow with initial value g_0 exists for all time.

The reader who wants more background on these metrics might consult the references and survey articles [**3, 5, 10**]. The references [**8, 9, 6, 14**] contain further important work in the area and will be cited further below.

1.1. Basic facts. — Unless the metric g admits flat factors, the equation $\mathrm{Ric}(g) = \nabla^2 f$ determines f up to an additive constant and it does no harm to fix a choice of f for the discussion. For simplicity, it does no harm to assume that g has no (local) flat factors and so this will frequently be done. Also, the Ricci-flat case (aka the Calabi-Yau case), in which $\mathrm{Ric}(g) = 0$, is a special case that is usually treated by different methods, so it will usually be assumed that $\mathrm{Ric}(g) \neq 0$. (Indeed, most of the latter part of this article will focus on the case in which $\mathrm{Ric}(g) > 0$).

1.1.1. *The associated holomorphic vector field Z.* — One of the earliest observations [**2**] made about gradient Kähler Ricci solitons is that the vector field ∇f is the real part of a holomorphic vector field and that, moreover, $J(\nabla f)$ is a Killing field for g. In this article, I will take $Z = \frac{1}{2}\big(\nabla f - \mathrm{i}J(\nabla f)\big)$ to be the holomorphic vector field *associated* to g.

1.1.2. *The holomorphic volume form Υ.* — In the Ricci-flat case, at least when M is simply connected, it is well-known that there is a g-parallel holomorphic volume form Υ, i.e., one which satisfies the condition that $\mathrm{i}^{n^2} 2^{-n}\, \Upsilon \wedge \overline{\Upsilon}$ is the real volume form determined by g and the J-orientation.

In § 2.2, I note that, for any gradient Kähler Ricci soliton g with Ricci potential f defined on a simply connected M, there is a holomorphic volume form Υ (unique up to a constant multiple of modulus 1) such that $\mathrm{i}^{n^2} 2^{-n}\, \mathrm{e}^{-f}\, \Upsilon \wedge \overline{\Upsilon}$ is the real volume form determined by g and the J-orientation. Of course, Υ is not g-parallel (unless g is Ricci-flat) but satisfies $\nabla \Upsilon = \frac{1}{2}\, \partial f \otimes \Upsilon$.

This leads to a notion of *special* coordinate charts for (g, f) i.e., coordinate charts (U, z) such that the associated coordinate volume form $\mathrm{d}z = \mathrm{d}z^1 \wedge \cdots \wedge \mathrm{d}z^n$ is the restriction of Υ to U. In such coordinate charts, several of the usual formulae simplify for gradient Kähler Ricci solitons.

1.1.3. *The Υ-divergence of Z.* — Given a vector field and and volume form, the divergence of the vector field with respect to the volume form is well defined. It turns out to be useful to consider this quantity for Z and Υ. The divergence in this case is the (necessarily holomorphic) function h that satisfies $\mathsf{L}_Z\, \Upsilon = h\, \Upsilon$.

By general principles, the scalar function h must be expressible in terms of the first and second derivatives of f. Explicit computation (Proposition 4) yields

$$2h = \operatorname{tr}_g(\nabla^2 f) + |\nabla f|^2 = R(g) + |\nabla f|^2, \tag{1.2}$$

where $R(g) = \operatorname{tr}_g(\operatorname{Ric}(g))$ is the scalar curvature of g. In particular, h is real-valued and therefore constant. Now, the constancy of $R(g) + |\nabla f|^2$ had already been noted and utilized by Hamilton and Cao [**6**]. However, its interpretation as a holomorphic divergence seems to be new.

1.2. Generality.

— An interesting question is: How many gradient Kähler Ricci solitons are there? Of course, this rather vague question can be sharpened in several ways.

The point of view adopted in this article is to start with a complex n-manifold M already endowed with a holomorphic volume form Υ and a holomorphic vector field Z and ask how many gradient Kähler solitons on M there might be (locally or globally) that have Z and Υ as their associated holomorphic data.

An obvious necessary condition is that the divergence h of Z with respect to Υ must be a real constant.

1.2.1. *Nonsingular extension.* — Away from the singularities (i.e., zeroes) of Z, this divergence condition turns out to be locally sufficient.

More precisely, I show (see Theorem 2) that if $H \subset M$ is an embedded complex hypersurface that is transverse at each of its points to Z, and g_0 and f_0 are, respectively, a real-analytic Kähler metric and function on H, then there is an open neighborhood U of H in M on which there exists a gradient Kähler Ricci soliton g with potential f whose associated holomorphic quantities are Z and Υ and such that g and f pull back to H to become g_0 and f_0. The pair (g, f) is essentially uniquely specified by these conditions. The real-analyticity of the 'initial data' g_0 and f_0 is necessary in order for an extention to exist since any gradient Kähler Ricci soliton is real-analytic anyway (see Remark 4).

Roughly speaking, this result shows that, away from singular points of Z, the local solitons g with associated holomorphic data (Z, Υ) depend on two arbitrary (real-analytic) functions of $2n-2$ variables.

1.2.2. *Singular existence.* — The existence of (local) gradient Kähler solitons in a neighborhood of a singularity p of Z is both more subtle and more interesting.

Even if the divergence of Z with respect to Υ is a real constant, it is not true in general that a gradient Kähler Ricci solition with Z and Υ as associated holomorphic data exists in a neighborhood of such a p.

I show (Proposition 6) that a necessary condition is that there exist p-centered holomorphic coordinates $z = (z^i)$ on a p-neighborhood $U \subset M$ and real numbers $h_1, \ldots, h_n$ such that, on U,

$$Z = h_1 \, z^1 \, \frac{\partial}{\partial z^1} + \cdots + h_n \, z^n \, \frac{\partial}{\partial z^n}. \tag{1.3}$$

In other words, Z must be holomorphically linearizable, with real eigenvalues. (1)

In such a case, if $\mathsf{L}_Z \Upsilon = h\Upsilon$ where h is a constant, then $h = h_1 + \cdots + h_n$. I show (Proposition 7) that, moreover, in this case, one can always choose Z-linearizing coordinates as above so that $\Upsilon = \mathrm{d}z^1 \wedge \cdots \wedge \mathrm{d}z^n$.

Thus, the possible local singular pairs (Z, Υ) that can be associated to a gradient Kähler Ricci soliton are, up to biholomorphism, parametrized by n real constants.

Using this normal form, one then observes that, by taking products of solitons of dimension 1, any set of real constants $(h_1, \ldots, h_n)$ can occur (see Remark 9). Since, for any gradient Kähler Ricci soliton g with associated holomorphic data (Z, Υ), the formula $\mathrm{Ric}(g) = \mathsf{L}_{\mathrm{Re}(Z)} \, g$ holds, it follows that if g is such a Kähler Ricci soliton defined on a neighborhood of a point p with $Z(p) = 0$, then $h_1, \ldots, h_n$ are the eigenvalues (each of even multiplicity) of $\mathrm{Ric}(g)$ with respect to g at p.

However, this does not fully answer the question of how 'general' the solitons are in a neighborhood of such a p. In fact, this very subtly depends on the numbers h_i. For example, if the $h_i \in \mathbb{R}$ are linearly independent over $\mathbb{Q}$, then any gradient Kähler Ricci soliton g with associated data (Z, Υ) defined on a neighborhood of p must be invariant under the compact n-torus action generated by the closure of the flow of the imaginary part of Z. This puts severe restrictions on the possibilities for such solitons.

At the conclusion of Section § 3, I discuss the local generality problem near a singular point of Z and explain how it can best be viewed as an elliptic boundary value problem of a certain type, but do not go into any further detail. A fuller discussion of this case may perhaps be undertaken at a later date.

1.3. The positive case. — In Section § 4, I turn to an interesting special case: The case where g is complete, the Ricci curvature is positive, and the scalar curvature $R(g)$ attains its maximum at some (necessarily unique) point $p \in M$.

This case has been studied before by Cao and Hamilton [**6**], who proved that this point p is a minimum of the Ricci potential f, that f is a proper plurisubharmonic exhaustion function on M (which is therefore Stein), and that, moreover, the Killing field $J(\nabla f)$ has a periodic orbit on 'many' of its level sets.

(1) Of course, it is by no means true that every holomorphic vector field is holomorphically linearizable at each of its singular points.

For simplicity, the Ricci potential f will be be normalized so that $f(p) = 0$, so that f is positive away from p.

I show (Theorem 3) that under these assumptions there exist global Z-linearizing coordinates $z = (z^i) : M \to \mathbb{C}^n$, so that M is biholomorphic to $\mathbb{C}^n$ (which generalizes an earlier result of Chau and Tam [**8**]). [2] Moreover, as a consequence, it follows that every positive level set of f has at least n periodic orbits of $J(\nabla f)$, a considerable sharpening of Cao and Hamilton's original results.

This global coordinate system has several other applications.

For example, I show that there is a Kähler potential ϕ for g that is invariant under the flow of $J(\nabla f)$ and that this potential is unique up to an additive constant. (Which can be normalized away by requiring that $\phi(p) = 0$.)

As another application, I show how to normalize the choice of Z-linearizing holomorphic coordinates up to an ambiguity that lies in a compact subgroup of $\mathrm{U}(n)$. This makes the function $|z|$ well-defined on M, so it is available for estimates.

As an illustration of such use, I show that there are positive constants r and a_1, a_2, b_1, b_2, c_1, and c_2 such that, whenever $|z| \geq r$,

$$\begin{aligned} a_1 \log|z| &\leq f(z) \leq a_2 \log|z|, \\ b_1 \log|z| &\leq d(z,0) \leq b_2 \log|z|, \\ c_1 \left(\log|z|\right)^2 &\leq \phi(z) \leq c_2 \left(\log|z|\right)^2. \end{aligned} \tag{1.4}$$

I also give some bounds for a_1 and a_2. Perhaps these will be useful in further work.

1.4. The toric case. — This section studies the geometry of the reduced equation in the case when a gradient Käher Ricci soliton g defined on a neighborhood of $0 \in \mathbb{C}^n$ has toric symmetry, i.e., is invariant under the action of $\mathbb{T}^n$, the diagonal subgroup of $\mathrm{U}(n)$. This may seem specialized, but, for example, if the associated holomorphic vector field is Z_h where $\mathsf{h} = (h_1, \ldots, h_n)$ and the real numbers $h_1, \ldots, h_n$ have the 'generic' property of being linearly independent over $\mathbb{Q}$, then g has toric symmetry. Thus, metrics with toric symmetry are the rule when Z has a 'generic' singularity.

I first derive the equation satisfied by the reduced potential, which turns out to be a singular Monge-Ampére equation. (The singularities are, of course, related to the singular orbits of the $\mathbb{T}^n$-action.) I then show that, nevertheless, this singular

(2) On 27 July 2004, about 12 hours before the first version of this article was posted on the arXiv, Chau and Tam posted the first version of their article arXiv:math.DG/0407449 in which they prove, under the same hypotheses as in Theorem 3, that M is biholomorphic to $\mathbb{C}^n$. I saw their posting just before I posted this article. Their method is different and does not produce Z-linearizing coordinates, but has the advantage that it applies in the case of expanding solitons. In the second (much shortened) version of their article, posted on 2 August, 2004, they deduce their biholomorphism result from already-known results about automorphisms of complex manifolds. See [**9**].

equation has good regularity and its singular initial value problem is well-posed in the sense of Gèrard and Tahara [**11**].

As a consequence (Corollary 5), it follows that, for any $\mathsf{h} \in \mathbb{R}^n$, any real-analytic $\mathbb{T}^{n-1}$-invariant Kähler metric on a neighborhood of $0 \in \mathbb{C}^{n-1}$ is the restriction to $\mathbb{C}^{n-1}$ of an essentially unique toric gradient Kähler Ricci soliton on an open subset of $\mathbb{C}^n$ with associated holomorphic vector field $Z = Z_{\mathsf{h}}$ and associated holomorphic volume form $\Upsilon = \mathrm{d}z$. In particular, it follows that, in a sense made precise in that section, the toric gradient Kähler Ricci solitons on $\mathbb{C}^n$ depend on one 'arbitrary' real-analytic function of $(n-1)$ (real) variables.

Next, I show that the reduced (singular Monge-Ampère) equation is of Euler-Lagrange type, at least, away from its singular locus, and discuss some of its conservation laws via an application of Noether's Theorem. (This is in contrast to the unreduced soliton equation, which is not variational).

1.5. Acknowledgement. — This work is mostly based on notes written after a conversation with Richard Hamilton during a visit he made to Duke University in 1991. Section 4 is more recent, having been written after further conversations with Hamilton during a semester I spent at Columbia University in the spring of 2004.

It is a pleasure thank Hamilton for his interest and to thank Columbia University for its hospitality.

2. Associated Holomorphic Quantities

In this section, constructions of some holomorphic quantities associated to a gradient Kähler Ricci soliton g on a complex n-manifold M^n with Ricci potential f will be described.

2.1. Preliminaries. — In order to avoid confusion because of various different conventions in the literature, I will collect the notations, conventions, and normalizations to be used in this article.

2.1.1. *Tensors and inner products.* — Factors of 2 are sometimes troubling and confusing in Kähler geometry.

For a and b in a vector space V, I will use the conventions $a \circ b = \frac{1}{2}(a \otimes b + b \otimes a)$ and $a \wedge b = a \otimes b - b \otimes a$. In particular, $a \otimes b = a \circ b + \frac{1}{2}\, a \wedge b$.

A real-valued inner product $\langle , \rangle$ on a real vector space V can be extended to $V^{\mathbb{C}} = \mathbb{C} \otimes V$ in several different ways. A natural way is to extend it as an Hermitian form, i.e., so that

$$\langle v_1 + \mathrm{i}v_2, w_1 + \mathrm{i}w_2 \rangle = \big(\langle v_1, w_1\rangle + \langle v_2, w_2\rangle\big) + \mathrm{i}\big(\langle v_2, w_1\rangle - \langle v_1, w_2\rangle\big) \tag{2.1}$$

and that is the convention to be adopted here.

If the real vector space V has a complex structure $J : V \to V$, then $V^{\mathbb{C}} = V^{1,0} \oplus V^{0,1}$ where $V^{1,0}$ is the $+\mathrm{i}$-eigenspace of J extended complex linearly to $V^{\mathbb{C}}$ while $V^{0,1}$ is the $(-\mathrm{i})$-eigenspace of J. It is common practice to identify $v \in V$ with $v^{1,0} = v - \mathrm{i}\, Jv \in V^{1,0}$, but some care must be taken with this.

For example, an inner product $\langle , \rangle$ on V is *compatible* with J if $\langle Jv, Jw\rangle = \langle v, w\rangle$ for all $v, w \in V$. Note the identity

$$\langle v^{1,0}, v^{1,0}\rangle = 2\langle v, v\rangle. \tag{2.2}$$

For any J-compatible inner product $\langle , \rangle$ on V (or equivalently, quadratic form) there is an *associated* 2-form η defined by

$$\eta(v, w) = \langle Jv, w\rangle\,. \tag{2.3}$$

2.1.2. *Coordinate expressions and the Ricci form.* — Let $z = (z^i) : U \to \mathbb{C}^n$ be a holomorphic coordinate chart on an open set $U \subset M$. The metric g restricted to U can be expressed in the form

$$g = g_{i\bar{\jmath}}\, \mathrm{d}z^i \circ \mathrm{d}\bar{z}^j \tag{2.4}$$

for some functions $g_{i\bar{\jmath}} = \overline{g_{j\bar{\imath}}}$ on U. The associated Kähler form Ω then has the coordinate expression

$$\Omega = \tfrac{\mathrm{i}}{2}\, g_{i\bar{\jmath}}\, \mathrm{d}z^i \wedge \mathrm{d}\bar{z}^j\,. \tag{2.5}$$

Note that $g_{i\bar{\jmath}}\, \mathrm{d}z^i \otimes \mathrm{d}\bar{z}^j = g - 2\mathrm{i}\,\Omega$.

The Ricci tensor $\mathrm{Ric}(g)$ is J-compatible since g is Kähler, and hence has a coordinate expression $\mathrm{Ric}(g) = R_{j\bar{k}}\, \mathrm{d}z^j \circ \mathrm{d}\bar{z}^k$ where $R_{j\bar{k}} = \overline{R_{k\bar{\jmath}}}$. Its associated 2-form ρ is computed by the formula

$$\rho = \tfrac{\mathrm{i}}{2}\, R_{i\bar{\jmath}}\, \mathrm{d}z^i \wedge \mathrm{d}\bar{z}^j = -\mathrm{i}\, \partial\bar{\partial}\, G \tag{2.6}$$

where

$$G = \log \det\bigl(g_{i\bar{\jmath}}\bigr). \tag{2.7}$$

While ρ is independent of the coordinate chart used to compute it, the function G does depend on the coordinate chart.

The scalar curvature $R(g) = \mathrm{tr}_g\bigl(\mathrm{Ric}(g)\bigr)$ has the coordinate expression

$$R(g) = 2g^{i\bar{\jmath}} R_{i\bar{\jmath}} \tag{2.8}$$

and satisfies $R(g)\, \Omega^n = 2n\, \rho \wedge \Omega^{n-1}$.

2.1.3. *The gradient Kähler Ricci soliton condition.* — The following equivalent formulation of the gradient Kähler Ricci soliton condition is well-known:

Proposition 1. — *A real-valued function f on M satisfies* $\mathrm{Ric}(g) = D^2 f$ *if and only if* $\rho = \mathrm{i}\,\partial\bar\partial f$ *and* $D^{0,2} f = 0$. *This latter condition is equivalent to the condition that the g-gradient of f be the real part of a holomorphic vector field on M.* □

2.2. The associated holomorphic volume form. — In this subsection, given a gradient Kähler Ricci soliton g with Ricci potential f on a simply-connected complex n-manifold M, a holomorphic volume form on M (unique up to a complex multiple of modulus 1) will be constructed.

2.2.1. *Existence of special coordinates.* — The following result shows that there are coordinate systems in which the Ricci potential is more closely tied to the local coordinate quantities.

Proposition 2. — *If g is a gradient Kähler Ricci soliton on M with Ricci potential f, then M has an atlas of holomorphic charts (U, z) satisfying* $\log\det\bigl(g_{i\bar\jmath}\bigr) = -f$.

Proof. — To begin, let (U, z) be any local holomorphic coordinate chart on M, with quantities $g_{i\bar\jmath}$ and G defined as above.

Since f is a Ricci potential for g, i.e., $\mathrm{Ric}(g) = D^2 f$, it follows from (2.6) and Proposition 1 that

$$-\,\mathrm{i}\,\partial\bar\partial\, G = \mathrm{i}\,\partial\bar\partial\, f\,. \tag{2.9}$$

Thus, $f + G$ is pluriharmonic. Assuming further that the domain U of the coordinate system z is simply connected, there exists a holomorphic function p on U so that

$$f = -G + p + \bar p. \tag{2.10}$$

Now let w be any other local coordinate system on the same simply connected domain U in M and write

$$\Omega = \tfrac{\mathrm{i}}{2}\, h_{i\bar\jmath}\, \mathrm{d}w^i \wedge \mathrm{d}\bar w^j. \tag{2.11}$$

Then $H = \log\det\bigl(h_{i\bar\jmath}\bigr)$ is of the form

$$H = G + J + \overline{J} \tag{2.12}$$

where J is the log-determinant of the Jacobian matrix of the change of variables from z to w, i.e.,

$$\mathrm{d}z^1 \wedge \mathrm{d}z^2 \wedge \cdots \wedge \mathrm{d}z^n = e^J\, \mathrm{d}w^1 \wedge \mathrm{d}w^2 \wedge \cdots \wedge \mathrm{d}w^n\,. \tag{2.13}$$

It follows that every point of U has an open neighborhood V on which there exists a coordinate chart w for which $-H = f$, the Ricci potential. □

***Definition 1* (Special coordinates).** — Let g be a gradient Kähler Ricci soliton on M with Ricci potential f. A coordinate chart (U, z) for which $\log\det(g_{i\bar{\jmath}}) = -f$ will be said to be *special* for (g, f).

***Remark 1* (The volume form in special coordinates).** — A coordinate chart (U, z) is special for (g, f) if and only if the volume form of g satisfies

$$\mathrm{dvol}_g = \frac{1}{n!}\,\Omega^n = \left(\frac{\mathrm{i}^n}{2}\right)^n \mathrm{e}^{-f}\,\mathrm{d}z \wedge \mathrm{d}\bar{z}. \tag{2.14}$$

***Theorem 1* (Existence of holomorphic volume forms).** — *Let M be a simply connected complex n-manifold endowed with a gradient Kähler Ricci soliton g with associated Kähler form Ω and a choice of Ricci potential f. Then there exists a holomorphic volume form Υ on M, unique up to muliplcation by a complex number of modulus* 1, *with the property that*

$$\mathrm{d}vol_g = \frac{1}{n!}\,\Omega^n = \left(\frac{\mathrm{i}^n}{2}\right)^n \mathrm{e}^{-f}\,\Upsilon \wedge \overline{\Upsilon}. \tag{2.15}$$

Proof. — For any two (g, f)-special coordinate charts z and w on the same domain U, the ratio of their corresponding holomorphic volume forms is a constant of modulus 1.

The volume forms of special coordinate systems are thus the sections of a flat connection ∇_0 on the canonical bundle of M, i.e., the bundle whose sections are the holomorphic volume forms on M. Since M is simply connected, the canonical bundle of M has a global ∇_0-flat section Υ that is unique up to a multiplicative constant.

By construction, Υ satisfies (2.15). Its uniqueness up to multiplication by a constant of modulus 1 is now evident. □

***Definition 2* (Associated holomorphic volume forms).** — Given a gradient Kähler Ricci soliton g with Ricci potential f, a holomorphic volume form Υ satisfying (2.15) will be said to be *associated* to the pair (g, f).

***Remark 2* (Scaling effects on Υ).** — Scaling a gradient Kähler Ricci soliton g by a constant produces another gradient Kähler Ricci soliton and adding a constant to f will produce another Ricci potential for g.

If Υ is associated to (g, f), then, for any real constants $\lambda > 0$ and c, the n-form $\lambda^n \mathrm{e}^c\,\Upsilon$ is associated to $(\lambda^2 g,\ f{+}2c)$.

2.3. The holomorphic flow. — Write the g-gradient of f as $Z + \bar{Z}$ where Z is of type $(1,0)$. Thus, $Z = \frac{1}{2}\big(\nabla f - \mathrm{i}\,J(\nabla f)\big)$.

2.3.1. *The infinitesimal symmetry.* — By the standard Kähler identities, Z is the unique vector field of type $(1,0)$ satisfying

$$\bar\partial f = -\mathrm{i}\, Z \lrcorner \Omega\,. \tag{2.16}$$

Writing $Z = X - i\,Y = X - i\,JX$, it follows that, in addition to X being the one-half the gradient of f, the vector field $Y = JX$ is Ω-Hamiltonian. Thus, the flow of Y preserves Ω.

Since Z is holomorphic by Proposition 1, the flow of Y also preserves the complex structure on M.

Hence, Y must be a Killing vector field for the metric g.

Thus, a gradient Kähler Ricci soliton that is not Ricci-flat always has a nontrivial infinitesimal symmetry.

Proposition 3. — *The singular locus of Z is a disjoint union of nonsingular complex submanifolds of M, each of which is totally geodesic in the metric g.*

Proof. — Since Z is holomorphic, its singular locus (i.e., the locus where it vanishes) is a complex subvariety of M. However, since this locus is also the zero locus of $Y = -\mathrm{Im}(Z)$, which is a Killing field for g, this locus is a submanifold that is totally geodesic with respect to g. In particular, it must be smooth and hence nonsingular as a complex subvariety. □

2.3.2. *Z in special coordinates.* — Assume (U, z) is a special local coordinate system. Since

$$\bar\partial G = g^{i\bar\jmath}\,\frac{\partial g_{i\bar\jmath}}{\partial \bar z^k}\,\mathrm{d}\bar z^k = -\bar\partial f\,, \tag{2.17}$$

the formula for Z in special coordinates is

$$Z = Z^\ell\,\frac{\partial}{\partial z^\ell} = -\left(2g^{\ell\bar k}g^{i\bar\jmath}\frac{\partial g_{i\bar\jmath}}{\partial \bar z^k}\right)\frac{\partial}{\partial z^\ell}\,. \tag{2.18}$$

Thus, the equations for a gradient Kähler Ricci soliton in special coordinates are that the functions Z^ℓ defined by (2.18) be holomorphic.

In fact, the expression in (2.18) can be simplified, since the closure of Ω is equivalent to the equations

$$\frac{\partial g_{i\bar\jmath}}{\partial \bar z^k} = \frac{\partial g_{i\bar k}}{\partial \bar z^j}. \tag{2.19}$$

Thus,

$$Z^\ell = -2\,g^{\ell\bar k}g^{i\bar\jmath}\,\frac{\partial g_{i\bar\jmath}}{\partial \bar z^k} = -2\,g^{i\bar\jmath}g^{\ell\bar k}\,\frac{\partial g_{i\bar k}}{\partial \bar z^j} = 2\,g^{i\bar\jmath}g_{i\bar k}\,\frac{\partial g^{\ell\bar k}}{\partial \bar z^j} = 2\,\frac{\partial g^{\ell\bar\jmath}}{\partial \bar z^j}, \tag{2.20}$$

where I have used the identity $g^{i\bar\jmath}g_{i\bar k} = \delta^{\bar\jmath}_{\bar k}$ and the identity $g_{i\bar k}g^{\ell\bar k} = \delta^\ell_i$ and its derivatives.

2.3.3. *The Υ-divergence of Z.* — Since Z is holomorphic, the Lie derivative of Υ with respect to Z must be of the form $h\,\Upsilon$ where h is a holomorphic function on M (usually called the divergence of Z with respect to Υ).

Replacing Υ by $\lambda\Upsilon$ for any $\lambda \in \mathbb{C}^*$ will not affect the definition of h, so the function h is intrinsic to the geometry of the soliton. On general principle, it must be computable in terms of the first and second covariant derivatives of f, which leads to the following interpretation of a result of Cao and Hamilton:

Proposition 4. — *The holomorphic function h is real-valued (and therefore constant). Moreover,*

$$2h = R(g) + 2|Z|^2 \tag{2.21}$$

where $R(g)$ is the scalar curvature of g and $|Z|^2$ is the squared g-norm of Z.

Proof. — In special coordinates, where $\Upsilon = \mathrm{d}z^1 \wedge \cdots \wedge \mathrm{d}z^n$, the function h has the expression

$$h = \frac{\partial Z^\ell}{\partial z^\ell}. \tag{2.22}$$

Thus, by (2.20),

$$h = 2\,\frac{\partial g^{\ell\bar{\jmath}}}{\partial z^\ell\,\partial \bar{z}^j}, \tag{2.23}$$

which shows that the holomorphic function h is real-valued and therefore constant. Moreover, since $\rho = \mathrm{i}\,\partial\bar{\partial}f$, it follows that

$$\begin{aligned}\left(\tfrac{\mathrm{i}}{2}\right) R_{j\bar{k}}\,\mathrm{d}z^j \wedge \mathrm{d}\bar{z}^k = \rho = \mathrm{i}\,\partial\bar{\partial}f &= \partial(Z \lrcorner\, \Omega)\\ &= \left(\tfrac{\mathrm{i}}{2}\right) \partial\left(g_{\ell\bar{k}} Z^\ell\,\mathrm{d}\bar{z}^k\right)\\ &= \left(\tfrac{\mathrm{i}}{2}\right)\left(g_{\ell\bar{k}}\frac{\partial Z^\ell}{\partial z^j} + Z^\ell\frac{\partial g_{\ell\bar{k}}}{\partial z^j}\right)\mathrm{d}z^j \wedge \mathrm{d}\bar{z}^k.\end{aligned} \tag{2.24}$$

In particular, in view of (2.19) and (2.18),

$$\begin{aligned}R(g) = 2g^{j\bar{k}}R_{j\bar{k}} = 2g^{j\bar{k}}\left(g_{\ell\bar{k}}\frac{\partial Z^\ell}{\partial z^j} + Z^\ell\frac{\partial g_{\ell\bar{k}}}{\partial z^j}\right) &= 2h + 2g^{j\bar{k}}Z^\ell\frac{\partial g_{\ell\bar{k}}}{\partial z^j}\\ = 2h + 2Z^\ell\,g^{j\bar{k}}\frac{\partial g_{j\bar{k}}}{\partial z^\ell} = 2h - g_{\ell\bar{k}}Z^\ell\bar{Z}^k = 2h - 2|Z|^2,&\end{aligned} \tag{2.25}$$

as claimed. □

***Remark 3* (Interpretations).** — It was Cao and Hamilton [**6**, Lemma 4.1] who first observed that the quantity $R(g) + |\nabla f|^2$ is constant for a (steady) gradient Kähler Ricci soliton. Since $Z = \frac{1}{2}\big(\nabla f - \mathrm{i}\,J(\nabla f)\big)$, one has $2|Z|^2 = |\nabla f|^2$, so their expression is the right hand side of (2.21).

The interpretation of $R(g) + |\nabla f|^2$ as the Υ-divergence of Z seems to be new. In fact, for any gradient Ricci soliton g (not necessarily Kähler) with Ricci potential f, one has the identity

$$\mathsf{L}_{\nabla f}\big(\mathrm{e}^f\, dvol_g\big) = 2h\ \mathrm{e}^f\, dvol_g\,. \tag{2.26}$$

where $R(g)+|\nabla f|^2 = 2h$ is a constant. This points out the importance of the modified volume form $\mathrm{e}^f\, dvol_g$ in the general case.

In a sense, this constancy can be regarded as a sort of conservation law for the Ricci flow. Note that, since $\Delta f = R(g)$, this relation is equivalent to the equation $\Delta_g(\mathrm{e}^f) = 2h\,\mathrm{e}^f$.

2.4. Examples. — The associated holomorphic objects constructed so far make it possible to simplify somewhat the usual treatment of the known explicit examples. The following examples will be useful in later discussions in this article.

***Example 1* (The one-dimensional case: Hamilton's cigar).** — Suppose that M is a Riemann surface. Then Υ is a nowhere vanishing 1-form on M and Z is a holomorphic vector field on M that satisfies $\mathrm{d}\big(\Upsilon(Z)\big) = h\,\Upsilon$, where h is a constant. There are essentially two cases to consider.

First, suppose that $h = 0$. Then $\Upsilon(Z)$ is a constant, say $\Upsilon(Z) = c$.

If $c = 0$, then Z is identically zero, and, from (2.20) it follows that, in special coordinates $z = (z^1)$ the real-valued function $g^{1\bar{1}}$ is constant. In particular, in special coordinates $g = g_{1\bar{1}}|\mathrm{d}z^1|^2$, so g is flat.

If $c \neq 0$, then Z is nowhere vanishing and, after adjusting Υ and the special coordinate system by a constant multiple, it can be assumed that $c = 2$, i.e., that $\Upsilon = \mathrm{d}z^1$ and $Z = 2\,\partial/\partial z^1$. Then (2.20) implies that $g^{1\bar{1}} = z^1 + \bar{z}^1 + C$ for some constant C. By adding a constant to z^1, it can be assumed that $C = 0$, so it follows that, in this coordinate system

$$g = \frac{|\mathrm{d}z^1|^2}{(z^1 + \bar{z}^1)}. \tag{2.27}$$

Since M is supposed to be simply connected, one can take z^1 to be globally defined. Thus M is immersed into the right half-plane in $\mathbb{C}$ in such a way that g is the pullback of the conformal metric defined by (2.27). Of course, this metric is not complete, even on the entire right half-plane.

Second, assume that h is not zero. Then $\Upsilon(Z)$ is a holomorphic function on M that has nowhere vanishing differential. Write $\Upsilon(Z) = hz^1$ for some (globally defined) holomorphic immersion $z^1 : M \to \mathbb{C}$. Then, by construction, $\Upsilon = \mathrm{d}z^1$ and $Z = hz^1\,\partial/\partial z^1$. By (2.20), it follows that

$$g^{1\bar{1}} = \tfrac{1}{2}(c + h\,|z^1|^2) \tag{2.28}$$

for some constant c, so $z^1(M) \subset \mathbb{C}$ must lie in the open set U in the w-plane on which $c + h|w|^2 > 0$. In fact, g must be the pullback under $z^1 : M \to U \subset \mathbb{C}$ of the metric

$$\frac{2\,|\mathrm{d}w|^2}{c + h\,|w|^2}. \tag{2.29}$$

This metric on the domain $U \subset \mathbb{C}$ is not complete unless both c and h are nonnegative and it is flat unless both c and h are positive. In this latter case, this metric is simply Hamilton's 'cigar' soliton [**12**].

Consequently, in dimension 1, the only complete gradient Kähler Ricci solitons are either flat or one of Hamilton's 'cigar' solitons (which are all homothetic to a single example).

Note that, under the Ricci flow $g_t = -2\,\mathrm{Ric}(g)$, the metric (2.29) evolves as

$$g(t) = \frac{2\,|\mathrm{d}w|^2}{\mathrm{e}^{2ht}c + h\,|w|^2} = \frac{2\,\left|\mathrm{d}(\mathrm{e}^{-ht}w)\right|^2}{c + h\,\left|\mathrm{e}^{-ht}w\right|^2} = \Phi(-t)^*(g_0) \tag{2.30}$$

where $\Phi(t)(w) = e^{ht}w$ is the flow of twice the real part of $Z = hw\,\partial/\partial w$.

***Example 2* (Products).** — By taking products of the 1-dimensional examples, one can construct a family of complete examples on $\mathbb{C}^n$: Let $h_1, \ldots, h_n$ and $c_1, \ldots, c_n$ be positive real numbers and consider the metric on $\mathbb{C}^n$ defined by

$$g = \sum_{k=1}^{n} \frac{2\,|\mathrm{d}w^k|^2}{(c_k + h_k\,|w^k|^2)}. \tag{2.31}$$

This is, of course, a gradient Kähler Ricci soliton, with associated holomorphic volume form and vector field

$$\Upsilon = \mathrm{d}w^1 \wedge \mathrm{d}w^2 \wedge \cdots \wedge \mathrm{d}w^n, \qquad Z = \sum_{k=1}^{n} h_k\, w^k\, \frac{\partial}{\partial w^k}. \tag{2.32}$$

The Ricci curvature is

$$\mathrm{Ric}(g) = \sum_{k=1}^{n} \frac{2c_k h_k\,|\mathrm{d}w^k|^2}{(c_k + h_k\,|w^k|^2)^2}. \tag{2.33}$$

Although these product examples are trivial generalizations of Hamilton's cigar soliton, they will be useful in observations to be made below.

Also, note that, even if the h_k are not positive, as long as the c_k are positive, the formula (2.31) defines an incomplete gradient Kähler Ricci soliton on the polycylinder defined by the inequalities $c_k + h_k|w^k|^2 > 0$.

***Example 3* (Cao's Soliton).** — One more case of an easily constructed example is the gradient Kähler Ricci soliton metric g on $\mathbb{C}^n$ that is invariant under $\mathrm{U}(n)$, discovered by H.-D. Cao [**2**]. The form of this metric can be derived as follows:

Suppose that such a metric g is given on $\mathbb{C}^n$. (One could do this analysis on any $\mathrm{U}(n)$-invariant domain in $\mathbb{C}^n$, and Cao does this, but I will not pursue this more general case further here.) The group $\mathrm{U}(n)$ must preserve the associated holomorphic volume form Υ up to a constant multiple and this implies that Υ must be a constant multiple of the standard volume form $\mathrm{d}z^1 \wedge \cdots \wedge \mathrm{d}z^n$. Since Υ is only determined up to a constant multiple anyway, there is no loss of generality in assuming that $\Upsilon = \mathrm{d}z^1 \wedge \cdots \wedge \mathrm{d}z^n$. Furthermore, the vector field Z must also be invariant under $\mathrm{U}(n)$, which implies that Z must be a multiple of the radial vector field. Since $\mathrm{d}(Z \lrcorner \Upsilon) = h\,\Upsilon$ where h is real, it follows that

$$Z = h \sum_{k=1}^{n} z^k \frac{\partial}{\partial z^k}. \tag{2.34}$$

Now, the condition that g be rotationally invariant with associated Kähler form closed implies that

$$g_{i\bar{\jmath}} = a(r)\delta_{ij} + a'(r)\,\bar{z}^i\,z^j \tag{2.35}$$

for some function a of $r = |z^1|^2 + \cdots + |z^n|^2$ that satisfies $ra'(r) + a(r) > 0$ and $a(r) > 0$ (when $n > 1$). Thus $G = \log\big(a(r)^{n-1}(ra'(r)+a(r))\big)$ in this coordinate system. Now, the identity $G = -f$, the equation (2.16), and the above formula for the coefficients of Ω combine to yield

$$\bar\partial G = \mathrm{i}\,Z \lrcorner \Omega = -\frac{h}{2}\big(ra'(r)+a(r)\big)\bar\partial r = -\frac{h}{2}\bar\partial\big(ra(r)\big). \tag{2.36}$$

Supposing that $n > 1$ (since the $n = 1$ case has already been treated), it follows that $G + \frac{h}{2}ra(r)$ must be constant, i.e., that

$$a(r)^{n-1}\big(ra(r)\big)' e^{(h/2)ra(r)} = a(0)^n. \tag{2.37}$$

Upon scaling Υ by a constant, it can be assumed that $a(0) = 1$, so assume this from now on. Also, one can assume that h is nonzero since, otherwise, the solution that is smooth at $r = 0$ is simply $a(r) \equiv a(0) = 1$, which gives the flat metric.

The ODE (2.37) for a is singular at $r = 0$, so the existence of a smooth solution near $r = 0$ is not immediately apparent.

Fortunately, (2.37) can be integrated by quadrature: Set $b(r) = (h/2)ra(r)$ and note that (2.37) can be written in terms of b as

$$b(r)^{n-1} e^{b(r)}\, b'(r) = (h/2)^n r^{n-1}. \tag{2.38}$$

Integrating both sides from 0 to $r > 0$ yields an equation of the form

$$(-1)^n (n-1)!\, e^{b(r)} \left(e^{-b(r)} - \sum_{k=0}^{n-1} \frac{\big(-b(r)\big)^k}{k!} \right) = \left(\frac{h}{2}\right)^n \frac{r^n}{n}. \tag{2.39}$$

Set

$$(2.40)\qquad F(b) = (-1)^n (n-1)!\, e^b \left(e^{-b} - \sum_{k=0}^{n-1} \frac{(-b)^k}{k!} \right) \simeq e^b \left(\frac{b^n}{n} - \frac{b^{n+1}}{n(n+1)} + \cdots \right).$$

Now, F has a power series of the form $F(b) = \frac{1}{n}\, b^n (1 + \frac{n}{n+1} b + \cdots)$, so F can be written in the form $F(b) = \frac{1}{n} f(b)^n$ for an analytic function of the form $f(b) = b(1 + \frac{1}{n+1} b + \cdots)$. The analytic function f is easily seen to satisfy $f'(b) > 0$ for all b and to satisfy the limits

$$(2.41)\qquad \lim_{b\to+\infty} f(b) = \infty \qquad \text{and} \qquad \lim_{b\to-\infty} f(b) = -\sqrt[n]{n!}\,.$$

In particular, f maps $\mathbb{R}$ diffeomorphically onto $\left(-\sqrt[n]{n!}, \infty\right)$ and is smoothly invertible. Of course, $f(0) = 0$.

Since (2.39) is equivalent to $f\big(b(r)\big)^n = \left(\frac{h}{2}\, r\right)^n$, when $h > 0$ it can be solved for $r \geq 0$ by setting $b(r) = f^{-1}\left(\frac{h}{2}\, r\right)$, yielding a unique real-analytic solution with a power series of the form

$$(2.42)\qquad b(r) = \frac{h}{2}\, r - \frac{h^2}{4(n+1)}\, r^2 + \cdots\ .$$

Consequently, when $h > 0$, the solution b is defined for all $r \geq 0$ and is positive and strictly increasing on the half-line $r \geq 0$. In particular, the function

$$(2.43)\qquad a(r) = \frac{2}{h}\, \frac{b(r)}{r} = 1 - \frac{h}{2(n+1)}\, r + \cdots\ .$$

is a positive real-analytic solution of (2.37) that is defined on the range $0 \leq r < \infty$ and satisfies $ra'(r) + a(r) = b'(r) > 0$ on this range, so that the expression (2.35) defines a gradient Kähler Ricci soliton on $\mathbb{C}^n$.

An ODE analysis of this solution (which Cao [**2**] does) shows that when $h > 0$ the resulting metric is complete on $\mathbb{C}^n$ and has positive sectional curvature.

When $h < 0$, the solution $b(r)$ only exists for $r < -\frac{2}{h}\sqrt[n]{n!}$. It is not difficult to see that the corresponding gradient Kähler Ricci soliton on a bounded ball in $\mathbb{C}^n$ is inextendible and incomplete.

Chau and Schnürer [**7**] have shown that Cao's example is stable in a certain sense and hence is 'isolated' in an appropriately defined neighborhood in the space of Kähler-Ricci solitons on $\mathbb{C}^n$.

3. Potentials and local generality

In this section, the question of 'how many' gradient Kähler Ricci soliton metrics could give rise to specified holomorphic data (Υ, Z) on a complex manifold M will be considered. While this question is not easy to answer globally, it is not so difficult to answer locally.

Thus, throughout this section, assume that a complex n-manifold M is specified, together with a nonvanishing holomorphic volume form Υ on M and a holomorphic vector field Z on M such that $\mathrm{d}\big(Z \lrcorner \Upsilon\big) = h\,\Upsilon$ for some real constant h.

3.1. Local potentials. — Suppose that $U \subset M$ is an open subset on which there exists a function ϕ such that $\Omega = \frac{\mathrm{i}}{2}\,\partial\bar\partial\phi$ is a positive definite $(1,1)$-form whose associated Kähler metric g is a gradient Ricci soliton with associated holomorphic data Υ and Z and Ricci potential f.

By (2.16),

$$
\begin{aligned}
2\bar\partial f &= -2\mathrm{i}Z \lrcorner \Omega = Z \lrcorner (\partial\bar\partial\phi) = -Z \lrcorner (\bar\partial\partial\phi)\\
&= -Z \lrcorner \big(\mathrm{d}(\partial\phi)\big) = -\,\mathsf{L}_Z(\partial\phi) + \mathrm{d}\big(\partial\phi(Z)\big)\\
&= \bar\partial\big(\partial\phi(Z)\big) - \big(\mathsf{L}_Z(\partial\phi) - \partial\big(\mathsf{L}_Z(\phi)\big)\big)
\end{aligned} \tag{3.1}
$$

By decomposition into type, it follows that

$$\bar\partial\big(2f - \partial\phi(Z)\big) = 0. \tag{3.2}$$

Consequently, $F = 2f - \partial\phi(Z) = 2f - \mathrm{d}\phi(Z)$ is a holomorphic function on U.

3.2. Nonsingular extension problems. — Suppose now that $p \in U$ is not a singular point of Z. Then, by shrinking U if necessary, F can be written in the form $F = dH(Z)$ for some holomorphic function H on the p-neighborhood U. Replacing ϕ by $\phi+H+\bar H$, gives a new potential for Ω that satisfies the stronger condition

$$\partial\phi(Z) = \mathrm{d}\phi(Z) = 2f. \tag{3.3}$$

This function ϕ is unique up to the addition of the real part of a holomorphic function that is constant on the orbits of Z.

Of course, (3.3) implies that $\mathrm{d}\phi(Y) = 0$, i.e., that ϕ is invariant under the flow of Y, the imaginary part of Z.

3.2.1. *Local reduction to equations.* — In local coordinates $z = (z^i)$ for which $\Upsilon = \mathrm{d}z^1 \wedge \cdots \wedge \mathrm{d}z^n$, one has $f = -G$ so ϕ satisfies the Monge-Ampère equation [3]

$$\det\left(\frac{\partial^2\phi}{\partial z^i \partial \bar z^j}\right) \mathrm{e}^{\frac{1}{2}\mathrm{d}\phi(X)} = 1 \tag{3.4}$$

as well as the equation

$$\mathrm{d}\phi(Y) = 0. \tag{3.5}$$

Conversely, if ϕ is a strictly pseudo-convex function defined on a p-neighborhood U that satisfies both (3.4) and (3.5), then the Kähler metric g whose associated Kähler

[3] It is interesting to note that this equation is not of Euler-Lagrange type, even locally, unless $Z \equiv 0$, i.e., the Ricci-flat case. Of course, in the Ricci-flat case, the variational nature of this equation is well-known.

form is $\Omega = \frac{i}{2}\,\partial\bar\partial\phi$ is a gradient Kähler Ricci soliton on U with associated holomorphic form Υ and holomorphic vector field Z.

***Remark 4* (Real-analyticity of solitons).** — Note that, because (3.4) is a real-analytic elliptic equation for the strictly pseudo-convex function ϕ, it follows by elliptic regularity that ϕ (and hence g) is real-analytic as well.

Now, (3.4) and (3.5) are two PDE for ϕ, the first of second order and the second of first order. While this is an overdetermined system, it is not difficult to show that it is involutive in Cartan's sense.

In fact, an analysis along the lines of exterior differential systems leads to the following result as a proper formulation of a 'Cauchy problem' for gradient Kähler Ricci solitons in the nonsingular case:

***Theorem 2* (Nonsingular extensions).** — *Let M^n be a complex n-manifold endowed with a holomorphic volume form Υ and a nonzero vector field Z satisfying $\mathrm{d}(Z \lrcorner \Upsilon) = h\,\Upsilon$ for some real constant h.*

Let $H^{n-1} \subset M$ be any embedded complex hypersurface that is transverse to Z, let Ω_0 be any real-analytic Kähler form on H, and let f_0 be any real-analytic function on H.

Then there is an open H-neighborhood $U \subset M$ on which there exists a gradient Kähler Ricci soliton g with associated Kähler form Ω, holomorphic volume form Υ, holomorphic vector field Z, and Ricci potential f that satisfy[(4)]

$$H^*\Omega = \Omega_0, \qquad \text{and} \qquad H^*f = f_0. \tag{3.6}$$

Moreover, g is locally unique in the sense that any other gradient Kähler Ricci soliton $\tilde g$ defined on an open H-neighborhood $\tilde U \subset M$ satisfying these initial conditions agrees with g on some open neighborhood of H in $U \cap \tilde U$.

Proof. — The first step in the proof will be to construct a special set of local 'flow-box' coordinate charts adapted to the hypersurface H, the holomorphic form Υ, and the holomorphic vector field Z.

To begin, note that, since, by hypothesis Z_p does not lie in $T_pH \subset T_pM$ for all $p \in H$, the $(n-1)$-form $Z \lrcorner \Upsilon$ is nonvanishing when pulled back to H.

Let $p \in H$ be fixed. Since $H^*\big(Z \lrcorner \Upsilon\big)$ does not vanish at p, there exist p-centered holomorphic coordinates $w^2, \dots, w^n$ on a p-neighborhood V in H such that $V^*\big(Z \lrcorner \Upsilon\big) = \mathrm{d}w^2 \wedge \cdots \wedge \mathrm{d}w^n$.

Since H is embedded in M, there exists an open neighborhood $U \subset M$ of $V \subset H$ with the property that $U \cap H = V$ and so that each complex integral curve $C \subset M$

(4) Notation: If $P \subset Q$ is a submanifold, and ψ is a differential form on Q, I use $P^*\phi$ to denote the pullback of ψ to P.

of Z that meets U does so in a connected open disk $U \cap C$ that intersects H in a unique point.

Consequently, there exist unique holomorphic functions $z^2, \ldots, z^n$ on U satisfying $\mathrm{d}z^2(Z) = \cdots = \mathrm{d}z^n(Z) = 0$ and $V^*(z^j) = w^j$ for $2 \leq j \leq n$. Moreover, there exists a unique function z^1 on U with the property that z^1 vanishes on $V = U \cap H$ and so that $U^*\Upsilon = \mathrm{d}z^1 \wedge \mathrm{d}z^2 \wedge \cdots \wedge \mathrm{d}z^n$. Since the functions $z^1, \ldots, z^n$ have independent differentials on U, it follows that by shrinking V (and hence U) if necessary, it can be assumed that (U, z) is a p-centered holomorphic coordinate chart whose image $z(U) \subset \mathbb{C}^n$ is a polycylinder of the form $|z^i| < \rho^i$ for some $\rho^1, \ldots, \rho^n > 0$. By shrinking ρ^1 if necessary, it can be arranged that $1{+}h\rho^1 > 0$.

By construction, $Z = F(z)\,\partial/\partial z^1$ for some holomorphic function F defined on $z(U) \subset \mathbb{C}^n$. Thus, $U^*(Z \lrcorner \Upsilon) = F(z)\,\mathrm{d}z^2 \wedge \cdots \wedge \mathrm{d}z^n$. Since $V^*(Z \lrcorner \Upsilon) = \mathrm{d}w^2 \wedge \cdots \wedge \mathrm{d}w^n$, it follows that $F(0, w^2, \ldots, w^n) = 1$ for $(0, w^2, \ldots, w^n) \in z(U)$. Moreover, since $\mathrm{d}\big(Z \lrcorner \Upsilon\big) = h\,\Upsilon$, it follows that $\partial F/\partial z^1 = h$. Consequently, in these coordinates $Z = (1{+}hz^1)\,\partial/\partial z^1$.

Now write $Z = X - \mathrm{i}Y$, where X and Y are commuting real vector fields. The integral curves of Y are transverse to the hypersurface H and there exists a real hypersurface $R \subset U$ that is the union of the integral curves of Y in U that pass through $V = U \cap H$. The vector field X is everywhere transverse to R in U.

Now let ψ_0 be a real-valued function on V such that $V^*(\Omega_0) = \frac{\mathrm{i}}{2}\,\partial\bar\partial\psi_0$. Such an Ω_0-potential ψ is unique up the the addition of the real part of a holomorphic function of $w^2, \ldots, w^n$. Extend ψ_0 to a function ψ_1 on R by making it constant along the integral curves of Y. Similarly, extend $V^*(f_0)$ to a function f_1 on R by making it constant along the integral curves of Y.

Finally, consider the initial value problem for a function ϕ on a neighborhood of R in U given by the real-analytic PDE

$$\det\left(\frac{\partial^2\phi}{\partial z^i \partial \bar z^j}\right) \mathrm{e}^{\frac{1}{2}\mathrm{d}\phi(X)} = 1 \tag{3.7}$$

subject to the real-analytic initial conditions

$$\begin{aligned} \phi(z) &= \psi_1(z) \\ \mathsf{L}_X(\phi)(z) &= 2f_1(z) \end{aligned} \qquad \text{for all } z \in R \subset U. \tag{3.8}$$

It is easy to check that (3.7) and (3.8) constitutes a noncharacteristic Cauchy problem. Hence, by the Cauchy-Kovalewski Theorem, there exists an open neighborhood $W \subset U$ containing R on which there exists a solution ϕ to this problem.

Now, the solution ϕ produced by the Cauchy-Kovalewski Theorem is real-analytic and strictly pseudo-convex. By uniqueness in the Cauchy-Kovalewski Theorem, ϕ is the unique real-analytic solution. Since, as has already been remarked, elliptic regularity implies that any strictly pseudo-convex solution of (3.7) must be real-analytic, it follows that ϕ is the unique solution of (3.7) that satisfies (3.8).

By its very construction, the $(1,1)$-form $\Omega = \frac{\mathrm{i}}{2}\partial\bar{\partial}\phi$ is then the Kähler form of a gradient Kähler Ricci soliton metric on $W \subset U$ that satisfies $V^*\Omega = V^*\Omega_0$, that has $W^*\Upsilon$ and W^*Z as the associated holomorphic volume form and vector field, respectively, and has $f = \frac{1}{2}\mathrm{d}\phi(X)$ as Ricci potential, which, of course, satisfies $V^*f = V^*f_0$.

Now, if one replaces ψ by $\psi + H + \bar{H}$ for some holomorphic function $H = H(w^2, \dots, w^n)$ on V, then one finds that the solution ϕ is replaced by by $\phi + H(z^2, \dots, z^n) + \overline{H(z^2, \dots, z^n)}$, so that Ω is unaffected.

The argument thus far has shown that every point $p \in H$ has an open neighborhood $U \subset M$ on which there exists a gradient Kähler-Ricci soliton g_U with the desired extension properties and associated holomorphic data. It has also shown that this extension is locally unique. Now a standard patching argument shows that there exists an open neighborhood $U \subset M$ of the entire complex hypersurface H on which such an extension exists and is unique in the sense described in the statement of the theorem. □

***Remark 5* (Local generality).** — Theorem 2 essentially says that the local gradient Kähler Ricci solitons depend on two real-analytic functions of $2n-2$ variables, namely the potential functions ψ_0 (which is assumed to be strictly pseudo-convex but otherwise arbitrary) and f_0 (which is arbitrary). There is, of course, some ambiguity in the choice of the holomorphic coordinates z^i, but this ambiguity turns out to depend on essentially $n-2$ holomorphic functions of $n-1$ holomorphic variables, which is negligible when compared with two arbitrary (real-analytic) functions of $2n-2$ real variables.

3.3. Near singular points of Z. — The situation near a singular point of Z is considerably more delicate and interesting.

3.3.1. *Linear parts and linearizability.* — Recall that, at a point $p \in M$ where Z vanishes, there is a well-defined linear map $Z'_p : T_pM \to T_pM$ (often called 'the linear part of Z at p') defined by setting $Z'_p(v) = w$ if $w = [V, Z](p)$ for some (and hence any) holomorphic vector field V defined near p and satisfying $V(p) = v \in T_pM$.

In local coordinates $z = (z^i)$ centered on p, if

$$Z = Z^j(z)\,\frac{\partial}{\partial z^j}, \tag{3.9}$$

where, by assumption $Z^j(0) = 0$ for $1 \le j \le n$, then

$$Z'_p\left(\frac{\partial}{\partial z^l}(p)\right) = \frac{\partial Z^j}{\partial z^l}(0)\,\frac{\partial}{\partial z^j}(p). \tag{3.10}$$

The linear map $Z'_p : T_pM \to T_pM$ has a Jordan normal form and this is an important invariant of the singularity. In particular, the set of eigenvalues of Z'_p is well-defined.

Proposition 5. — *Let Z be the holomorphic vector field associated to a gradient Kähler Ricci soliton g on M. At any singular point of Z, the linear part Z'_p is diagonalizable, with all eigenvalues real.*

Proof. — If the data (Υ, Z) is associated to a gradient Kähler Ricci soliton g in a neighborhood of a singular point p of Z, then (2.24) shows that, in special coordinates centered on p, one has

$$g^{i\bar{k}}(0)\, R_{j\bar{k}}(0) = \frac{\partial Z^i}{\partial z^j}(0). \tag{3.11}$$

Because the matrices $\big(g_{i\bar{j}}(0)\big)$ and $\big(R_{i\bar{j}}(0)\big)$ are Hermitian symmetric and $\big(g_{i\bar{j}}(0)\big)$ is positive definite, one can choose the special coordinates so that $\big(g_{i\bar{j}}(0)\big)$ is a multiple of the identity matrix and $\big(R_{i\bar{j}}(0)\big)$ is diagonal. □

Definition 3. — A holomorphic vector field Z on M is said to be *linearizable* near a singular point p if there exist p-centered coordinates $w = (w^i)$ on an open p-neighborhood W and constants a^i_j such that, on W, one has

$$Z = a^i_j w^j \,\frac{\partial}{\partial w^i}\,. \tag{3.12}$$

The coordinates $w = (w^i)$ are said to be *linearizing* or *Poincaré* coordinates for Z near p.

Not every holomorphic vector field is linearizable near its singular points, even if the linear part at such a point has all of its eigenvalues nonzero and distinct.

***Example 4* (A nonlinearizable singular point).** — The vector field

$$Z = z^1 \,\frac{\partial}{\partial z^1} + \big(2z^2 + (z^1)^2\big)\,\frac{\partial}{\partial z^2} \tag{3.13}$$

on $\mathbb{C}^2$ is not linearizable at the origin, even though its linear part there is diagonalizable with eigenvalues 1 and 2.

This nonlinearizability is perhaps most easily seen as follows: The flow $\Phi(t)$ of the vector field Z is

$$\Phi(t)(z^1, z^2) = \big(\mathrm{e}^t\, z^1,\, \mathrm{e}^{2t}(z^2 + (z^1)^2 t)\big). \tag{3.14}$$

In particular $\Phi(t + 2\pi \mathrm{i}) \neq \Phi(t)$, which would be true if Z were holomorphically conjugate to the linear vector field

$$Z'_{(0,0)} = z^1 \,\frac{\partial}{\partial z^1} + 2z^2 \,\frac{\partial}{\partial z^2}. \tag{3.15}$$

This phenomenon, however, does not happen for singular points of holomorphic vector fields associated to a gradient Kähler Ricci soliton:

Proposition 6. — *Let Z be a nonzero holomorphic vector field on the complex n-manifold M that is associated to a gradient Kahler Ricci soliton g. Then Z is linearizable at each of its singular points. Moreover, the linear part of Z at a singular point is diagonalizable and has all its eigenvalues real.*

Proof. — Let $p \in M$ be a singular point of Z. The diagonalizability of the linear part of Z at a singular point and the reality of the corresponding eigenvalues has already been demonstrated, so all that remains is to show that Z is linearizable near p.

To do this, write $Z = X - \mathrm{i}Y$ where X and Y are, as usual, real vector fields. As has already been remarked, the vector field Y is an infinitesimal isometry of g. In particular, the flow of Y is complete in the geodesic ball $B_r(p)$ for some $r > 0$ and is a 1-parameter group of isometries of the metric g restricted to $B_r(p)$ that fixes the center p. It follows that there is a compact, connected abelian subgroup $\mathbb{T} \subset \mathrm{U}(T_pM)$ whose Lie algebra is an abelian subalgebra $\mathfrak{t} \subset \mathfrak{u}(T_pM)$ that contains $Y_p' : T_pM \to T_pM$, the linearization of Y at p and is such that the 1-parameter subgroup $\exp(tY_p')$ is dense in $\mathbb{T}$.

Let $\Phi : \mathbb{T} \to \mathrm{Isom}\big(B_r(p), g\big)$ be the homomorphism induced by the exponential map, i.e., such that

$$\Phi(k)\big(\exp_p(v)\big) = \exp_p\big(k \cdot v\big) \tag{3.16}$$

for all $v \in B_r(0_p) \subset T_pM$. Then $\Phi(k)$ is a holomorphic isometry of g for all $k \in \mathbb{T}$.

Now let $\mathrm{d}\mu$ be Haar measure on $\mathbb{T}$ and choose any holomorphic mapping $\psi : B_r(p) \to T_pM \simeq \mathbb{C}^n$ with the property that $\psi(p) = 0$ and $\psi'(p) : T_pM \to T_{0_p}(T_pM)$ is the inverse of the exponential mapping $\exp_p' : T_{0_p}(T_pM) \to T_pM$. (It may be necessary to shrink r to do this.)

Define a holomorphic mapping $w : B_r(p) \to T_pM$ by the averaging formula

$$w(z) = \int_{\mathbb{T}} k^{-1} \cdot \psi\big(\Phi(k)z\big)\, \mathrm{d}\mu \tag{3.17}$$

for $z \in B_r(p)$. Then $w(p) = 0_p$ and, by construction, $w\big(\Phi(k)z\big) = k \cdot w(z)$ for all $z \in B_r(p)$ and all $k \in \mathbb{T}$. Moreover, also by construction, $w'(p) = \psi'(p)$. In particular, by shrinking r again, if necessary, it can be assumed that w defines a $\mathbb{T}$-equivariant holomorphic embedding of $B_r(p)$ into $T_pM \simeq \mathbb{C}^n$.

In particular, the holomorphic mapping $w : B_r(p) \to T_pM$ satisfies

$$w\big(\exp_{tY}(z)\big) = \exp(tY_p')\big(w(z)\big), \tag{3.18}$$

for all real t. Since w is holomorphic and Y is the imaginary part of the holomorphic vector field Z, it follows that, for $z \in B_r(p)$ and t complex and of sufficiently small modulus, the identity

$$w\big(\exp_{tZ}(z)\big) = \exp(tZ'_p)\big(w(z)\big) \tag{3.19}$$

holds. In particular, the coordinate system w linearizes Z at p. □

***Remark 6* (The exponential map).** — Of course, the exponential map $\exp_p : T_pM \to M$ of g also intertwines the flow of Y'_p on T_pM with the flow of Y on M, but the exponential map is not generally holomorphic and so cannot be used to linearize Z holomorphically.

***Remark 7* (Complex vs. real flows).** — The reader may want to remember that, for a holomorphic vector field $Z = X - \mathrm{i}Y$, the two real vector fields X and Y have commuting flows and that, moreover, the identity

$$\exp_{(a+\mathrm{i}b)Z} = \exp_{2aX} \circ \exp_{2bY} \tag{3.20}$$

holds. (The factors of 2 are neglected in some references.)

Corollary 1. — *Let g be a gradient Kähler Ricci soliton on M and let Z be its associated holomorphic vector field. Let $p \in M$ be a singular point of Z and let $\lambda \in \mathbb{R}^*$ be a nonzero eigenvalue of Z'_p of multiplicity $k \geq 1$. Then there exists a k-dimensional complex submanifold $N_\lambda \subset M$ that passes through p, to which Z is everywhere tangent, and on which Y is periodic of period $4\pi/|\lambda|$.*

***Remark 8* (Nonuniqueness of the N_λ).** — The reader should be careful not to confuse the submanifolds N_λ with the images under the exponential mapping of the eigenspaces of Z'_p acting on T_pM. Indeed, the N_λ need not be unique. For example, for the linear vector field

$$Z = z^1\,\frac{\partial}{\partial z^1} + 2z^2\,\frac{\partial}{\partial z^2}. \tag{3.21}$$

on $\mathbb{C}^2$, each of the parabolas $z^2 - c(z^1)^2 = 0$ for $c \in \mathbb{C}$ is tangent to Z and the imaginary part of Z has period 4π on all of $\mathbb{C}^2$, so each could be regarded as N_1.

On the other hand, the line $z^1 = 0$ is the only curve that could be regarded as N_2, since this is the union of the 2π-periodic points of Y.

***Remark 9* (Existence at singular points).** — Example 2 shows that diagonalizability with real eigenvalues is sufficient for a linear vector field to be the linear part of a vector field associated to a (locally defined) gradient Kähler Ricci soliton.

3.3.2. *Prescribed eigenvalues.* — Let $\mathsf{h} = (h_1, \ldots, h_n) \in \mathbb{R}^n$ be a nonzero real vector and define

$$\Lambda_{\mathsf{h}} = \{\mathsf{k} \in \mathbb{Z}^n \mid \mathsf{k} \cdot \mathsf{h} = 0\} = \mathbb{Z}^n \cap \mathsf{h}^{\perp} \subset \mathbb{R}^n. \tag{3.22}$$

Then Λ_{h} is a free abelian group of rank $n - k$ for some $1 \leq k \leq n$. The number k is the dimension over $\mathbb{Q}$ of the $\mathbb{Q}$-span of the numbers $h_1, \ldots, h_n$ in $\mathbb{R}$. Let $\Lambda_{\mathsf{h}}^{+} \subset \Lambda_{\mathsf{h}}$ consist of the $\mathsf{k} \in \Lambda_{\mathsf{h}}$ such that $\mathsf{k} = (k_1, \ldots, k_n)$ with each k_i nonnegative.

Consider the linear holomorphic vector field

$$Z_{\mathsf{h}} = \sum_{j=1}^{n} h_j z^j \frac{\partial}{\partial z^j} \tag{3.23}$$

on $\mathbb{C}^n$. Let $Z_{\mathsf{h}} = X_{\mathsf{h}} - \mathrm{i} Y_{\mathsf{h}}$ be the decomposition into real and imaginary parts.

The closure of the flow of Y_{h} is a connected compact abelian subgroup $\mathbb{T}_{\mathsf{h}} \subset \mathrm{U}(n)$ of dimension k. (In fact, in these coordinates, $\mathbb{T}_{\mathsf{h}}$ lies in the diagonal matrices in $\mathrm{U}(n)$.) Note that Z_{h} and (hence) X_{h} are invariant under the action of $\mathbb{T}_{\mathsf{h}}$.

3.3.3. *Normalizing volume forms.* — In addition to knowing that Z can be linearized near a singular point, it will be useful to know that this can be done in such a way that it simplifies the coordinate expression for Υ as well:

***Proposition** 7* **(Volume normalization at Z-singular points).** — *Set $h = h_1 + \cdots + h_n$ and let Υ be a nonvanishing holomorphic n-form defined on an open neighborhood U of the origin in $\mathbb{C}^n$ that satisfies $\mathrm{d}(Z_{\mathsf{h}} \lrcorner \Upsilon) = h\Upsilon$.*

Then there exist Z_{h}-linearizing coordinates $w = (w^i)$ near the origin in $\mathbb{C}^n$ such that, on the domain of these coordinates $\Upsilon = \mathrm{d}w^1 \wedge \cdots \wedge \mathrm{d}w^n$.

Proof. — There exists a nonvanishing holomorphic function F on U that satisfies

$$\Upsilon = F(z)\, \mathrm{d}z^1 \wedge \cdots \wedge \mathrm{d}z^n \tag{3.24}$$

and the function F must be invariant under the flow of Z_{h}. In particular, it follows that F has a power series expansion of the form

$$F(z) = c_0 + \sum_{\mathsf{k} \in \Lambda_{\mathsf{h}}^{+} \setminus \{0\}} c_{\mathsf{k}}\, z^{\mathsf{k}} \tag{3.25}$$

where z^{k} is the monomial $(z^1)^{k_1} \cdots (z^n)^{k_n}$ when $\mathsf{k} = (k_1, \ldots, k_n)$ and the c_{k} are constants, with $c_0 \neq 0$ (since, by hypothesis $F(0) \neq 0$).

Now, the series

$$G(z) = c_0 + \sum_{\mathsf{k} \in \Lambda_{\mathsf{h}}^{+} \setminus \{0\}} \frac{c_{\mathsf{k}}}{(k_1+1)}\, z^{\mathsf{k}} \tag{3.26}$$

converges on the same polycylinder that the series (3.25) does. The resulting holomorphic function G is evidently invariant under the flow of Z_{h} and satisfies

$$G + z^1\,\frac{\partial G}{\partial z^1} = F. \tag{3.27}$$

Because G satisfies (3.27), the function $w^1 = z^1 G(z)$ satisfies

$$\mathrm{d}w^1 \wedge \mathrm{d}z^2 \wedge \cdots \wedge \mathrm{d}z^n = F(z)\,\mathrm{d}z^1 \wedge \mathrm{d}z^2 \wedge \cdots \wedge \mathrm{d}z^n. \tag{3.28}$$

Moreover, since G is Z_{h}-invariant, the function w^1 satisfies $\mathsf{L}_{Z_{\mathsf{h}}}\,w^1 = h_1\,w^1$.

Thus, replacing z^1 by w^1 in the coordinate chart results in a new Z_{h}-linearizing coordinate chart in which $\Upsilon = \mathrm{d}z^1 \wedge \cdots \wedge \mathrm{d}z^n$. □

***Corollary 2* (Local normal form near singular points).** — *Let Z and Υ be a holomorphic vector field and volume form, respectively on a complex n-manifold M. Let $p \in M$ be a singular point of Z.*

If there exists a gradient Kähler Ricci soliton g with Ricci potential f on a neighborhood of p whose associated holomorphic vector field and volume form are Z and Υ, respectively, then there exists an $\mathsf{h} \in \mathbb{R}^n$ and a p-centered holomorphic chart $z = (z^i) : U \to \mathbb{C}^n$ such that, on U,

$$Z = h_i\,z^i\,\frac{\partial}{\partial z^i} \qquad \text{and} \qquad \Upsilon = \mathrm{d}z = \mathrm{d}z^1 \wedge \cdots \wedge \mathrm{d}z^n. \tag{3.29}$$

Proof. — Apply Propositions 6 and 7. □

3.3.4. *Local solitons near a singular point.* — In view of Corollary 2, questions about the local existence and generality of gradient Kähler Ricci solitons with prescribed Z and Υ near a singular point of Z can be reduced by a holomorphic change of variables to the study of solitons on an open neighborhood of $0 \in \mathbb{C}^n$ with $Z = Z_{\mathsf{h}}$ for some $\mathsf{h} \neq 0$ and $\Upsilon = \mathrm{d}z = \mathrm{d}z^1 \wedge \cdots \wedge \mathrm{d}z^n$.

***Proposition 8* (Solitons with a prescribed singularity).** — *Let ϕ be a strictly pseudoconvex function defined on a $\mathbb{T}_{\mathsf{h}}$-invariant, contractible neighborhood of $0 \in \mathbb{C}^n$ that satisfies*

$$\det\left(\frac{\partial^2\phi}{\partial z^i \partial \bar{z}^j}\right) \mathrm{e}^{\frac{1}{2}\mathrm{d}\phi(X_{\mathsf{h}})} = 1 \tag{3.30}$$

and

$$\mathrm{d}\phi(Y_{\mathsf{h}}) = 0. \tag{3.31}$$

Then $\Omega = \frac{i}{2}\,\partial\bar{\partial}\phi$ is the associated Kähler form of a gradient Kähler Ricci soliton with Ricci potential $f = \frac{1}{2}\mathrm{d}\phi(X_{\mathsf{h}})$ whose associated holomorphic vector field and volume form are Z_{h} and $\mathrm{d}z^1 \wedge \cdots \wedge \mathrm{d}z^n$, respectively.

Conversely, if g is a gradient Kähler Ricci soliton defined on a $\mathbb{T}_{\mathsf{h}}$-invariant, contractible neighborhood of $0 \in \mathbb{C}^n$ and f is a Ricci potential for g that satisfies $f(0) = 0$ such that the associated holomorphic vector field and volume form are Z_{h} and $\mathrm{d}z^1 \wedge \cdots \wedge \mathrm{d}z^n$, respectively, then g has a Kähler potential ϕ that satisfies (3.30) *and* (3.31).

Proof. — The first part of the proposition follows by computation, so nothing more needs to be said. It remains to establish the converse statement.

Thus, consider a gradient Kähler Ricci soliton g defined on a $\mathbb{T}_{\mathsf{h}}$-invariant, contractible neighborhood U of $0 \in \mathbb{C}^n$ with Ricci potential f satisfying $f(0) = 0$ whose associated holomorphic volume form and vector field are $\Upsilon = \mathrm{d}z$ and Z_{h}, respectively.

The metric g will necessarily be invariant under $\mathbb{T}_{\mathsf{h}}$, as will its associated Kähler form Ω. Since $\Omega^n = n!\, \mathrm{i}^{n^2}\, 2^{-n}\, \mathrm{e}^{-f}\, \Upsilon \wedge \overline{\Upsilon}$, it follows that f, too, must be invariant under $\mathbb{T}_{\mathsf{h}}$.

On U, there will exist some Kähler potential ϕ so that $\Omega = \frac{\mathrm{i}}{2}\partial\bar{\partial}\phi$. By averaging ϕ over $\mathbb{T}_{\mathsf{h}}$, it can be assumed that ϕ is $\mathbb{T}_{\mathsf{h}}$-invariant. By subtracting a constant, it can be assumed that $\phi(0) = 0$.

As has been already noted in § 3.1, the difference $F = 2f - \mathrm{d}\phi(Z_{\mathsf{h}})$ is a holomorphic function on U. By construction, F is also necessarily $\mathbb{T}_{\mathsf{h}}$-invariant and vanishes at 0. Since ϕ is $\mathbb{T}_{\mathsf{h}}$-invariant, it follows that $\mathrm{d}\phi(Y_{\mathsf{h}}) = 0$. Thus $F = 2f - \mathrm{d}\phi(Z_{\mathsf{h}}) = 2f - \mathrm{d}\phi(X_{\mathsf{h}})$ is real-valued and holomorphic and therefore constant. Thus, F vanishes identically, i.e., $f = \frac{1}{2}\mathrm{d}\phi(X_{\mathsf{h}})$.

Now, however, by construction, ϕ satisfies (3.30) and, since ϕ is invariant under the flow of Y_{h}, it also satisfies (3.31). □

***Remark 10* (Analyticity in the singular case).** — The equation (3.30) is a $\mathbb{T}_{\mathsf{h}}$-invariant real-analytic Monge-Ampère equation whose linearization at a strictly pseudo-convex solution ϕ is given by

$$\Delta u + 2\, \mathsf{L}_{X_{\mathsf{h}}}\, u = 0 \tag{3.32}$$

where Δ is the Laplacian with respect to the metric g associated to $\Omega = \frac{\mathrm{i}}{2}\, \partial\bar{\partial}\phi$. Of course, this is an elliptic equation.

It follows by elliptic regularity that any gradient Kähler Ricci soliton is real-analytic, even in the neighborhood of singular points of Z.

***Example 5* (Existence with prescribed h).** — By considering Example 2, one sees that, for any h, there is a sufficiently small ball around the origin on which there is at least one strictly pseudo-convex solution ϕ to (3.30).

3.3.5. *A boundary value formulation.* — Suppose now that ϕ is a strictly pseudo-convex solution of (3.30) defined on a $\mathbb{T}_{\mathsf{h}}$-invariant bounded neighborhood $D \subset \mathbb{C}^n$

of $0 \in \mathbb{C}^n$ with smooth boundary ∂D. Let g be the corresponding gradient Kähler Ricci soliton.

Any solution u of (3.32) in D that vanishes on the boundary will also satisfy

$$0 = \int_D |\nabla u|^2 + \tfrac{1}{2} R(g)\, u^2 \; dvol_g\,, \tag{3.33}$$

as follows by integration by parts using the identities $\rho = \mathsf{L}_{X_h}\,\Omega$ and $dvol_g = \frac{1}{n!}\Omega^n$.

In particular, by shrinking D if necessary, it can be assumed that any solution u to (3.32) in D that vanishes on ∂D must vanish on D.

It then follows, by the implicit function theorem, that any $\mathbb{T}_{\mathsf{h}}$-invariant function ψ on ∂D that is sufficiently close (in the appropriate norm) to ϕ on ∂D is the boundary value of a unique pseudo-convex solution $\tilde{\phi}$ of (3.30) that is near ϕ on D. The uniqueness then implies that $\tilde{\phi}$ must also be $\mathbb{T}_{\mathsf{h}}$-invariant and so must, in particular, satisfy (3.31).

Note that the metric g does not always uniquely determine ϕ by the construction given in Proposition 8 since one can add to ϕ the real part of any $\mathbb{T}_{\mathsf{h}}$-invariant holomorphic function that vanishes at $0 \in \mathbb{C}^n$. (Depending on h, there may or may not be any nonconstant $\mathbb{T}_{\mathsf{h}}$-invariant holomorphic functions on a neighborhood of $0 \in \mathbb{C}^n$.) However, this ambiguity is relatively small.

Thus, local gradient Kähler Ricci solitons near $0 \in \mathbb{C}^n$ with prescribed holomorphic data $(Z, \Upsilon) = (Z_{\mathsf{h}}, dz)$ do exist and have a 'degree of generality' that depends on the number k. The most constraints appear when k reaches its maximum value n and the least when k reaches its minimum value 1.

4. Poincaré coordinates in the positive case

Throughout this section, M will be a noncompact, simply connected complex manifold and g will be a complete gradient Kähler Ricci soliton with postive Ricci curvature. Moreover, it will be assumed that the scalar curvature $R(g)$ has at least one critical point.

4.1. First consequences. — Cao and Hamilton [**6**, Proposition 4.2] prove the following useful result:

Lemma 1. — *The scalar curvature $R(g)$ has only one critical point and it is both a local maximum and the unique critical point of f, which is a strictly convex proper function on M.*

Proof. — Since $R(g)+2|Z|^2 = 2h$ by Proposition 4, the function $R(g) \geq 0$ is bounded by the constant h and any critical point of $R(g)$ is a critical point of $|Z|^2 = \frac{1}{2}|\nabla f|^2$.

On the other hand, since $\nabla^2 f = \mathrm{Ric}(g)$, which is positive definite, the formula

$$\mathrm{d}\left(\tfrac{1}{2}|\nabla f|^2\right)(\nabla f) = \nabla^2 f(\nabla f, \nabla f) = \mathrm{Ric}(g)(\nabla f, \nabla f) \tag{4.1}$$

shows that $\frac{1}{2}|\nabla f|^2$ cannot have any critical point away from where $\nabla f = 0$. Moreover, any point p where ∇f vanishes satisfies $R(g)(p) = 2h$, which is the maximum possible value of $R(g)$.

Since $\nabla^2 f = \mathrm{Ric}(g)$ is positive definite, the function f is locally strictly convex. Since g is complete, f can have at most one critical point, i.e., point where $\nabla f = 0$, and it must be a nondegenerate minimum of f.

By hypothesis, there does exist a (unique) critical point of f; call it p. By adding a constant to f it can be assumed that $f(p) = 0$. It remains to show that f is proper, i.e., that $f^{-1}([a,b]) \subset M$ is compact for any closed interval $[a,b] \subset \mathbb{R}$.

Since $R(g) + 2|Z|^2 = 2h$ and since $R(g) > 0$, it follows that $|Z| \leq \sqrt{h}$, so that Z has bounded length. In particular, writing

$$Z = X - \mathrm{i}Y = \tfrac{1}{2}\big(\nabla f - \mathrm{i}J(\nabla f)\big) \tag{4.2}$$

one has $|X|^2 = |Y|^2 = \frac{1}{2}|Z|^2 < \frac{1}{2}h$, so X and Y have bounded lengths as well. Since g is complete, their flows are globally defined on M.

Let $\gamma : \mathbb{R} \to M$ be any nonconstant integral curve of ∇f, i.e., $\gamma'(t) = \nabla f\big(\gamma(t)\big) \neq 0$ for all $t \in \mathbb{R}$. Consider the function $\phi(t) = f\big(\gamma(t)\big)$. Straightforward computation yields $\phi'(t) = \big|\nabla f\big(\gamma(t)\big)\big|^2 > 0$ and

$$\phi''(t) = 2\,\mathrm{Ric}(g)\big(\nabla f\big(\gamma(t)\big), \nabla f\big(\gamma(t)\big)\big) > 0, \tag{4.3}$$

so $\phi : \mathbb{R} \to \mathbb{R}$ is strictly convex and increasing. It follows that ϕ increases without bound along γ.

Since $\nabla^2 f$ is positive definite, the critical point p is a source singularity of the vector field ∇f. Let $U \subset M$ be the open set that consists of p and all of the points q in M whose α-limit point under ∇f is equal to p. Since f strictly increases without bound on each integral curve of ∇f, it follows that f maps each integral curve of ∇f that lies in U diffeomorphically onto $(0,\infty)$. Moreover, for each $c > 0$, the set $f^{-1}(c) \cap U$ is compact and diffeomorphic to S^{2n-1}. Indeed, $f : U \to [0,\infty)$ is proper.

Now suppose that $U \neq M$. Then, by the connectedness of M, there exists a point $q \in M \setminus U$ that is not in the interior of $M \setminus U$, i.e., a point $q \notin U$ such that there exists a sequence $q_i \in U$ that converges to q. This implies, in particular, that $f(q_i) \geq 0$ converges to $f(q) = c$. Thus, $c \geq 0$ and, for i sufficiently large, q_i must lie in $f^{-1}\big([0,c{+}1]\big) \cap U$, which has been shown to be compact and must therefore contain its limit points. Thus q lies in $f^{-1}\big([0,c{+}1]\big) \cap U$, although, by construction, $q \notin U$. Thus, $U = M$ and f is proper, as claimed. □

***Remark 11* (M is Stein).** — As Cao and Hamilton remark, since $\rho = \mathrm{i}\partial\bar{\partial}f$ is the Ricci form of g, which is positive, the proof shows that f is a strictly plurisubharmonic proper exhaustion function on M. This implies that M is Stein and, as Cao points out in [**4**, Proposition 3.2], that M is diffeomorphic to $\mathbb{R}^{2n}$.

However, as will be seen in Theorem 3, one has the stronger result that M is biholomorphic to $\mathbb{C}^n$.

The following result, also known to Cao and Hamilton, [5] gives constraints on the rate of growth of the Ricci potential.

***Lemma 2* (Growth of f).** — *Let p be the critical point of $R(g)$ and let f be the Ricci potential, normalized so that $f(p) = 0$. There exist positive constants c_1 and c_2 such that, for all $x \in M$,*

$$\sqrt{1 + \big(c_1\, d(x,p)\big)^2} - 1 \leq f(x) \leq c_2\, d(x,p). \tag{4.4}$$

Proof. — Since g is complete, there exists a geodesic joining p to x whose length is $d(p,x)$. Let $\alpha : \mathbb{R} \to M$ be such a unit speed geodesic with $\alpha(0) = p$ and $\alpha(s) = x$ such that $d(p,x) = s$.

Consider the function $\phi(t) = f\big(\alpha(t)\big)$. By the Chain Rule, and the fact that α has unit speed,

$$\phi'(t) = \nabla f\big(\alpha(t)\big) \cdot \alpha'(t) \leq \big|\nabla f\big(\alpha(t)\big)\big| \leq \sqrt{2h}. \tag{4.5}$$

Since $\phi(0) = 0$, it follows that $f(x) = f\big(\alpha(s)\big) = \phi(s) \leq \sqrt{2h}\, s$. Thus, one can take $c_2 = \sqrt{2h}$.

For the other inequality, note that, again, by the Chain Rule,

$$\phi''(t) = \nabla^2 f\big(\alpha(t)\big)\big(\alpha'(t), \alpha'(t)\big) = \mathrm{Ric}(g)\big(\alpha(t)\big)\big(\alpha'(t), \alpha'(t)\big) \tag{4.6}$$

and the right hand side of this equation is positive since $\mathrm{Ric}(g)$ is positive. Moreover, if $\lambda_{\min}(g) > 0$ denotes the minimum eigenvalue of $\mathrm{Ric}(g)$, which is a positive continuous function on M, it follows that

$$\phi''(t) \geq \lambda_{\min}(g)\big(\alpha(t)\big) > 0. \tag{4.7}$$

In particular, ϕ is a convex function on $\mathbb{R}$.

Let $r_0 > 0$ be sufficiently small that it is below the injectivity radius of g at p and sufficiently small that $\lambda_{\min}(g)(y) \geq \frac{1}{2}\lambda_{\min}(g)(p)$ for all y lying within $B_{r_0}(p)$. Let $a = \frac{1}{2}\lambda_{\min}(g)(p) > 0$.

[5] H.-D. Cao, personal communication, 2 June 2004.

Then $\phi''(t) \geq a$ for $|t| \leq r_0$ while $\phi''(t) > 0$ for $|t| \geq r_0$. Because $\phi(0) = \phi'(0) = 0$, it follows that $\phi(t) \geq A(t)$ for all $t \in \mathbb{R}$ where

$$A(t) = \begin{cases} \frac{1}{2}at^2 & \text{for } |t| \leq r_0; \\ ar_0|t| - \frac{1}{2}a\,r_0{}^2 & \text{for } |t| \geq r_0. \end{cases} \tag{4.8}$$

Since there exists a positive constant c_1 such that $A(t) \geq \sqrt{1 + (c_1 t)^2} - 1$, the desired lower bound follows. □

***Remark 12* (An alternative growth formulation).** — Another formulation of Lemma 2 is that the function $c : M \setminus \{p\} \to \mathbb{R}$ defined by

$$c(x) = \frac{\sqrt{f(x)\big(f(x) + 2\big)}}{d(x,p)} > 0 \tag{4.9}$$

is bounded above and has a positive lower bound.

The bounds of Lemma 2 can be simplified somewhat if one stays sufficiently far from p:

Corollary 3. — *For every $r > 0$, there exist positive constants c_1 and c_2 such that, for all x outside the ball of radius r, one has*

$$c_1\, d(x,p) \leq f(x) \leq c_2\, d(x,p). \tag{4.10}$$

***Remark 13* (The growth rate of f).** — For any vector $v \in TM$, one has

$$\mathrm{Ric}(g)(v,v) \leq \lambda_{\max}(g)\,|v|^2 \tag{4.11}$$

where $\lambda_{\max}(g) : M \to \mathbb{R}$ is the maximum eigenvalue function for $\mathrm{Ric}(g)$. Since g is Kähler, the eigenvalues of $\mathrm{Ric}(g)$ occur in pairs and, since $\mathrm{Ric}(g) > 0$, it follows that $\lambda_{\max}(g) \leq \frac{1}{2}R(g)$. In particular, by Proposition 4, one has the more explicit inequality

$$\mathrm{Ric}(g)(v,v) \leq \tfrac{1}{2}R(g)\,|v|^2 \leq \tfrac{1}{2}\big(2h - |\nabla f|^2\big)\,|v|^2. \tag{4.12}$$

Now let $\gamma : (0,\infty) \to M$ be the arclength parametrization of a nonconstant integral curve of ∇f, such that p is the limit of $\gamma(s)$ as $s \to 0^+$. Thus, $\big|\nabla f\big(\gamma(s)\big)\big|\,\gamma'(s) = \nabla f\big(\gamma(s)\big)$ for all $s > 0$.

Let $\phi(s) = f\big(\gamma(s)\big)$. One then computes via the Chain Rule that

$$\phi'(s) = \big|\nabla f\big(\gamma(s)\big)\big| \leq \sqrt{2h}. \tag{4.13}$$

and hence that

$$\phi''(s) = \mathrm{Ric}(g)\left(\frac{\nabla f\big(\gamma(s)\big)}{\big|\nabla f\big(\gamma(s)\big)\big|}, \frac{\nabla f\big(\gamma(s)\big)}{\big|\nabla f\big(\gamma(s)\big)\big|}\right). \tag{4.14}$$

By the positivity of $\mathrm{Ric}(g)$ and (4.12), this implies

$$0 < \phi''(s) \leq \tfrac{1}{2}\left(2h - (\phi'(s))^2\right). \tag{4.15}$$

Moreover, it is clear that, as $s \to 0^+$, the quantity on the right hand side of (4.14) has $\lambda_{\min}(g)(0) > 0$ as a lower bound for its infimum limit. Thus, the infimum limit of $\phi''(s)$ as $s \to 0^+$ is positive.

From these relations, several conclusions can be drawn. The function ϕ is increasing and strictly convex up on $(0, \infty)$. On the other hand, since ϕ' is bounded above, it follows that ϕ grows at most linearly. Moreover, there must be a sequence of distances $s_k \to \infty$ such that $\phi''(s_k) \to 0$. Since, by (4.14)

$$\phi''(s_k) \geq \lambda_{\min}(g)\big(\gamma(s_k)\big), \tag{4.16}$$

it follows that $\lambda_{\min}(g)\big(\gamma(s_k)\big) \to 0$ as $k \to \infty$.

4.2. Poincaré coordinates. — Let Υ be the associated holomorphic volume form on M, normalized so that Υ has unit size at p. This determines Υ up to a complex multiple of modulus 1. Let Z be the associated holomorphic vector field.

Since Z vanishes at p, the eigenvalues of Z'_p are the eigenvalues of the Ricci tensor at p, which are real and positive, say $h_1, \ldots, h_n > 0$. Set $h = h_1 + \cdots + h_n > 0$, as usual.

***Theorem 3* (Poincaré coordinates).** — *There exists a global special coordinate system $z : M \to \mathbb{C}^n$ that linearizes Z. In particular, M is biholomorphic to $\mathbb{C}^n$.*

Proof. — By Proposition 6, there exists a small open ball U about p on which there exist p-centered holomorphic coordinates $w = (w^i) : U \to \mathbb{C}^n$ that linearize Z. By shrinking U if necessary, it can be assumed that $U = f^{-1}\big([0, \varepsilon)\big)$ for some small $\varepsilon > 0$. Note that, since the w^i linearize Z, the identity

$$w^i\big(\exp_{tZ}(q)\big) = \mathrm{e}^{h_i t}\, w^i(q) \tag{4.17}$$

holds for all $q \in U$ and all $t \in \mathbb{C}$ in the connected domain containing $0 \in \mathbb{C}$ for which $\exp_{tZ}(q)$ lies in U. In particular, this implies that

$$w^i\big(\exp_{2tX}(q)\big) = \mathrm{e}^{h_i t}\, w^i(q) \tag{4.18}$$

for all $q \in U$ and all $t \in \mathbb{R}$ in the interval containing $0 \in \mathbb{R}$ for which $\exp_{2tX}(q)$ lies in U.

Now, for $q \in M$ distinct from p, write $q = \exp_{2t'X}(q')$ for some $q' \in U$ and $t' \in \mathbb{R}$. Define

$$z^i(q) = \mathrm{e}^{h_i t'}\, w^i(q'). \tag{4.19}$$

If $\exp_{2t'X}(q') = \exp_{2t''X}(q'')$ for some $q'' \in U$ and $t'' \in \mathbb{R}$, then one sees from (4.18) that $\mathrm{e}^{h_i t''}\, w^i(q'') = \mathrm{e}^{h_i t'}\, w^i(q')$, so $z^i(q)$ is well-defined.

Since the flow of X is holomorphic and w^i is holomorphic on U, the function $z^i : M \to \mathbb{C}$ is also holomorphic. Moreover, by construction,

$$z^i\big(\exp_{2tX}(q)\big) = \mathrm{e}^{h_i t}\, z^i(q) \tag{4.20}$$

for all $q \in M$, which implies that

$$z^i\big(\exp_{tZ}(q)\big) = \mathrm{e}^{h_i t}\, z^i(q). \tag{4.21}$$

In particular, the Lie derivative of z^i by Z is $h_i z^i$.

The fact that the mapping $z = (z^i) : M \to \mathbb{C}^n$ is one-to-one and onto now follows immediately since, as was observed in the proof of Lemma 1, the gradient flow lines of $\nabla f = 2X$ all have p as α-limit point and the flow of ∇f exists for all time.

Finally, in these coordinates $\Upsilon = F(z)\,\mathrm{d}z^1 \wedge \cdots \wedge \mathrm{d}z^n$ for some nonvanishing entire holomorphic function F on $\mathbb{C}^n$. However, since $\mathrm{d}(Z \lrcorner \Upsilon) = h\,\Upsilon$, it follows immediately that $\mathrm{d}F(Z) = 0$. Since all of the eigenvalues of Z'_p are positive, this is only possible if F is a constant function. By scaling one of the z^i by a constant, it can be arranged that $F \equiv 1$.

Thus, the resulting global coordinate system (M, z) is special and linearizes Z, as desired. □

***Remark 14* (Previous results).** — Chau and Tam [**8**, Theorem 1.1] proved that M is biholomorphic to $\mathbb{C}^n$ under the additional hypothesis that all the eigenvalues h_i are equal. In a very recent posting to the arXiv [**8**], they prove a result that implies that M is biholomorphic to $\mathbb{C}^n$ under the hypotheses of Theorem 3. However, their result does not provide Z-linearizing coordinates, which is the main purpose of Theorem 3.

4.3. Coordinate ambiguities. — The reader may find it surprising that any local Z-linearizing coordinates z^i defined on a neighborhood of the Z-singular point p extend to global coordinates on $\mathbb{C}^n$ that are special for any gradient Käher-Ricci soliton defined on $\mathbb{C}^n$ with positive Ricci curvature whose associated holomorphic vector field is Z.

This is perhaps made less surprising by the following result:

Proposition 9. — *Let $\mathsf{h} = (h_1, \ldots, h_n) \in \mathbb{R}^n$ be a vector with $h_i > 0$ for $1 \le i \le n$. Consider the vector field*

$$Z_{\mathsf{h}} = h_i\, z^i\, \frac{\partial}{\partial z^i} \tag{4.22}$$

on $\mathbb{C}^n$. Then the set G_{h} of biholomorphisms $\psi : \mathbb{C}^n \to \mathbb{C}^n$ that preserve Z_{h} is a complex Lie group of dimension d_{h} where $d_{\mathsf{h}} \ge n$ is the number of vectors $\mathsf{k} = (k_1, \ldots, k_n) \in \mathbb{Z}^n$ that satisfy $k_i \ge 0$ and $\mathsf{k} \cdot \mathsf{h} \in \{h_1, \ldots, h_n\}$.

Moreover, if $U \subset \mathbb{C}^n$ is any connected open neighborhood of $0 \in \mathbb{C}^n$, then any locally defined biholomorphism $\psi : (U, 0) \to (\mathbb{C}^n, 0)$ that preserves Z_{h} is the restriction to U of an element of G_{h}.

Proof. — Let $U \subset \mathbb{C}^n$ be an open neighborhood of 0 and let $\psi = \bigl(w^i(z)\bigr) : U \to \mathbb{C}^n$ be a local biholomorphism that preserves Z. Since Z has only one singular point, namely $0 \in \mathbb{C}^n$, it follows that $\psi(0) = 0$. Moreover, by construction, the functions w^i must satisfy $\mathrm{d}w^i(Z) = h_i w^i$. It follows that each w^i has a power series expansion about $0 \in \mathbb{C}^n$ of the form

$$w^i = \sum_{\{\mathsf{k}\geq 0 \mid \mathsf{k}\cdot\mathsf{h} = h_i\}} c^i_{\mathsf{k}}\, z^{\mathsf{k}}. \tag{4.23}$$

Since the right hand side has only a finite number of terms, this power series is a polynomial and hence globally defined on $\mathbb{C}^n$. It remains to see that it is invertible.

Consider the n-form $\mathrm{d}w = \mathrm{d}w^1 \wedge \cdots \wedge \mathrm{d}w^n$. By the above analysis $\mathrm{d}w = F(z)\,\mathrm{d}z$ for some polynomial $F(z)$. By hypothesis, ψ is a local biholomorphism, so $F(0) \neq 0$. Since $\mathsf{L}_Z\,\mathrm{d}w = (h_1 + \cdots + h_n)\,\mathrm{d}w$ by construction, it follows that $\mathrm{d}F(Z) = 0$, i.e., that F is Z-invariant. This implies that F is constant and hence nowhere vanishing.

Now, by hypothesis, ψ is locally invertible, with, say, a local inverse $\psi^{-1} : (V, 0) \to (\mathbb{C}^n, 0)$. However, by construction, ψ^{-1} preserves Z, so, by the argument given above, ψ^{-1} is also a polynomial mapping and hence extends to a global polynomial mapping $\pi : (\mathbb{C}^n, 0) \to (\mathbb{C}^n, 0)$. Since $\psi \circ \pi : (\mathbb{C}^n, 0) \to (\mathbb{C}^n, 0)$ is a polynomial mapping that is the identity on some neighborhood of 0, it must be the identity everywhere on $\mathbb{C}^n$. In particular, π is the global inverse of ψ extended to $\mathbb{C}^n$, which is now revealed to be an element of G_{h}, which is what needed to be shown.

Finally, it is clear that, for any i and any choice of constants $c^i_{\mathsf{k}} \in \mathbb{C}$ for (i, k) such that $\mathsf{k} \in \mathbb{Z}^n$ satisfies $k_j \geq 0$ for $1 \leq j \leq n$ and $\mathsf{k} \cdot \mathsf{h} = h_i$, the formula (4.23) defines a polynomial w^i that satisfy $\mathsf{L}_Z\, w^i = h_i w^i$.

Moreover, for any choice of d_{h} constants $c = (c^i_{\mathsf{k}})$ where (i, k) satisfies $\mathsf{k} \in \mathbb{Z}^n$ with $k_j \geq 0$ for $1 \leq j \leq n$ and $\mathsf{k} \cdot \mathsf{h} = h_i$, the corresponding collection of functions w^i satisfies

$$\mathrm{d}w^1 \wedge \cdots \wedge \mathrm{d}w^n = F(c^i_{\mathsf{k}})\,\mathrm{d}z^1 \wedge \cdots \wedge \mathrm{d}z^n. \tag{4.24}$$

where F is a polynomial of degree n in the d parameters $c^i_{\mathsf{k}} \in \mathbb{C}$.

As long as $F(c^i_{\mathsf{k}}) \neq 0$, the polynomial mapping $\psi_c = (w^i)$ is a local (and therefore global) biholomorphism of $\mathbb{C}^n$ that preserves Z and hence lies in G_{h}. Thus, the c^i_{k} define global holomorphic coordinates on G_{h} that embed it into $\mathbb{C}^{d_{\mathsf{h}}}$ as an open set. $\square$

***Remark 15* (The structure of G_{h}).** — If $\mu_1, \ldots, \mu_k \geq 1$ are the multiplicities of the eigenvalues $(h_1, \ldots, h_n)$, then G_{h} is the semi-direct product of a reductive subgroup isomorphic to $\mathrm{GL}(\mu_1,\mathbb{C}) \times \cdots \mathrm{GL}(\mu_k,\mathbb{C})$ with a nilpotent subgroup biholomorphic to $\mathbb{C}^\mu$ where $\mu = d_{\mathsf{h}} - {\mu_1}^2 - \cdots - {\mu_k}^2$.

When $n = 1$, one has $G_{\mathsf{h}} \simeq \mathbb{C}^* = \mathrm{GL}(1,\mathbb{C})$. When $n = 2$, one has either

1. $d_{\mathsf{h}} = 2$ if $\mathsf{h} = (h_1, h_2)$ with neither h_1/h_2 nor h_2/h_1 an integer (in which case $G_{\mathsf{h}} = \mathbb{C}^* \times \mathbb{C}^*$);
2. $d_{\mathsf{h}} = 3$ if $\mathsf{h} = (h_1, h_2)$ with either h_1/h_2 or h_2/h_1 an integer greater than 1; or
3. $d_{\mathsf{h}} = 4$ if $\mathsf{h} = (h, h)$ (in which case $G_{\mathsf{h}} = \mathrm{GL}(2,\mathbb{C})$).

When $n > 2$, there is no upper bound for d_{h} that depends only on n. For example, when $n = 3$, one has $d_{(1,1,k)} = k + 6$ for any integer $k > 1$.

4.4. Global consequences. — Throughout this section, g will be a complete gradient Kähler Ricci soliton on $\mathbb{C}^n$ with positive Ricci curvature whose associated vector field Z is given by (4.22) where $\mathsf{h} = (h_1, \ldots, h_n)$ and

$$0 < h_1 \leq h_2 \leq \cdots \leq h_n\,. \tag{4.25}$$

The compact abelian group $\mathbb{T}_{\mathsf{h}} \subset \mathrm{U}(n)$ will denote the closure of the orbit of Y, the imaginary part of Z.

The existence of global linearizing coordinates for a gradient Kähler Ricci soliton gives elementary proofs and/or improvements of several known results.

4.4.1. *Periodic orbits.* — The first result sharpens Theorem 1.1 of the article [**6**] of Cao and Hamilton.

Proposition 10 **(Periodic orbits of $J(\nabla f)$).** — *For all $c > 0$, the flow of $J(\nabla f)$ preserves the (smooth) level set $f^{-1}(c) \subset M$ and has at least n periodic orbits on $f^{-1}(c)$.*

Proof. — Since $Z = \frac{1}{2}\big(\nabla f - \mathrm{i}\, J(\nabla f)\big)$, and since $h_i > 0$ for $1 \leq i \leq n$, it follows that $J(\nabla f)$ is periodic of period $2\pi/h_i$ on the z^i-axis. Moreover, since f increases without bound as $|z^i| \to \infty$, this axis meets each level set $f^{-1}(c)$ for $c > 0$ in a circle. Thus, there are at least n distinct periodic orbits of $J(\nabla f)$ within each such level set. $\square$

4.4.2. *An invariant potential.* — As has already been seen, the metric g is invariant under $\mathbb{T}_{\mathsf{h}}$. It turns out that one can canonically choose a Kähler potential for g:

Proposition 11 **(Canonical potentials).** — *There is a unique $\mathbb{T}_{\mathsf{h}}$-invariant Kähler potential $\phi : \mathbb{C}^n \to \mathbb{R}$ satisfying $\Omega = \frac{\mathrm{i}}{2}\,\partial\bar\partial\phi$ and $\phi(0) = 0$.*

Proof. — Since $M = \mathbb{C}^n$, there exists at least one Kähler potential ϕ for g, i.e., such that $\Omega = \frac{\mathrm{i}}{2}\,\partial\bar\partial\phi$. Since $\mathbb{T}_{\mathsf{h}}$ is compact, by averaging ϕ over $\mathbb{T}_{\mathsf{h}}$, one can assume that ϕ is $\mathbb{T}_{\mathsf{h}}$ invariant and by adding a constant, one can assume that $\phi(0) = 0$.

If $\tilde\phi$ were also $\mathbb{T}_{\mathsf{h}}$-invariant and satisfied $\Omega = \frac{\mathrm{i}}{2}\,\partial\bar\partial\tilde\phi$, then the difference $\tilde\phi - \phi$ would be the real part of a $\mathbb{T}_{\mathsf{h}}$-invariant holomorphic function H. In particular H would be invariant under the flow of Y and hence of Z. However, as has already been seen, the only holomorphic functions on $\mathbb{C}^n$ that are invariant under the flow of Z are the constants. Thus $\tilde{\phi} - \phi$ is constant. The normalization $\phi(0) = 0$ then guarantees the uniqueness of ϕ. □

4.4.3. *Normalized linearizing coordinates.* — The ambiguity in the linearizing coordinates for the vector field Z represented by the group G_{h} can be used to simplify the potential for g.

***Theorem 4* (Normalized coordinates).** — *Let ϕ be the unique $\mathbb{T}_{\mathsf{h}}$-invariant Kähler potential for g, normalized so that $\phi(0) = 0$. Then there exists an element $\Psi \in G_{\mathsf{h}}$, unique up to composition with an element of the compact group $\mathrm{U}(n) \cap G_{\mathsf{h}}$, such that*

$$\Psi^*(\phi) = |z^1|^2 + \cdots + |z^n|^2 + E_{\bar{\imath}\bar{\jmath}kl}(z)\bar{z}^i\bar{z}^j z^k z^l \tag{4.26}$$

for some real-analytic functions $E_{\bar{\imath}\bar{\jmath}kl} = E_{\bar{\jmath}\bar{\imath}kl} = E_{\bar{\imath}\bar{\jmath}lk} = \overline{E_{\bar{k}\bar{l}ij}}$ defined near $0 \in \mathbb{C}^n$.

Proof. — Let f be the Ricci potential for g, normalized so that $f(0) = 0$. Since f is $\mathbb{T}_{\mathsf{h}}$-invariant and since, by (3.2), the difference $2f - \mathrm{d}\phi(Z)$ is holomorphic and $\mathbb{T}_{\mathsf{h}}$-invariant, it follows by the same argument as above that $2f - \mathrm{d}\phi(Z)$ is constant and hence vanishes identically. Thus

$$\mathrm{d}\phi(Z) = \mathrm{d}\phi(X) = 2f. \tag{4.27}$$

Because ϕ and f are real-analytic they have convergent power series expansions near $0 \in \mathbb{C}^n$. Since $f(0) = 0$ and f has a critical point at 0, it has an expansion of the form

$$f = \tfrac{1}{2} f_{ij}\, z^i z^j + f_{i\bar{\jmath}}\, z^i \bar{z}^j + \tfrac{1}{2}\overline{f_{ij}}\, \bar{z}^i \bar{z}^j + O(|z|^3). \tag{4.28}$$

where $f_{ij} = f_{ji}$ and $f_{i\bar{\jmath}} = \overline{f_{j\bar{\imath}}}$. Because of the positivity of the h_i and the invariance of f under the flow of Y, it follows that $f_{ij} = 0$ and $(h_i - h_j)f_{i\bar{\jmath}} = 0$ for all i and j. Moreover, since f is strictly convex up at the origin, the Hermitian form $f_{i\bar{\jmath}}\, z^i \bar{z}^j$ is positive definite.

Thus, by making a linear change of variables that preserves Z (i.e., by applying a transformation in $\mathrm{GL}(n, \mathbb{C}) \cap G_{\mathsf{h}}$), it can be arranged that

$$f = \tfrac{1}{2} h_1\, |z^1|^2 + \cdots + \tfrac{1}{2} h_n\, |z^n|^2 + O(|z|^3). \tag{4.29}$$

Next, consider the part of f that is pure in z or $\bar{z}$, i.e., consider the expansion

$$f = \tfrac{1}{2} h_1\, |z^1|^2 + \cdots + \tfrac{1}{2} h_n\, |z^n|^2 + \sum_{\mathsf{k} \geq 0,\, |\mathsf{k}| \geq 3} \left(f_{\mathsf{k}}\, z^{\mathsf{k}} + \overline{f_{\mathsf{k}}}\, \bar{z}^{\mathsf{k}}\right) + f_{i\bar{\jmath}}(z) z^i \bar{z}^j. \tag{4.30}$$

where $f_{\mathsf{k}} \in \mathbb{C}$ and $f_{i\bar{\jmath}} = \overline{f_{j\bar{\imath}}}$ vanishes at $z = 0$. The invariance of f under the flow of Y implies that $f_{\mathsf{k}} = 0$ for all k, so these 'pure' terms do not appear after all.

Finally, consider the part of the remainder that is linear in the variables $\bar{z}^i$ or z^i and vanishes at $z = 0$ to order at least 3, i.e., write

$$f = \tfrac{1}{2}h_k\,|z^k|^2 + Q^i(z)\,\bar{z}^i + \overline{Q^i(z)}\,z^i + f_{\bar{\imath}\bar{\jmath}kl}(z)\bar{z}^i\bar{z}^j z^k z^l, \tag{4.31}$$

where $Q^i(z)$ is a holomorphic function of z that vanishes to order at least 2 at $z = 0$ and $f_{\bar{\imath}\bar{\jmath}kl} = f_{\bar{\jmath}\bar{\imath}kl} = f_{\bar{\imath}\bar{\jmath}lk} = \overline{f_{\bar{k}\bar{l}ij}}$.

Again, the fact that f is invariant under the flow of Y implies that Q^i must satisfy $\mathsf{L}_Z\,Q^i = h_i Q^i$, i.e., that Q^i has an expansion of the form

$$Q^i(z) = \sum_{\{\mathsf{k}\geq 0\ |\ \mathsf{k}\cdot\mathsf{h}=h_i\}} c^i_{\mathsf{k}} z^{\mathsf{k}} \tag{4.32}$$

with $c^i_{\mathsf{k}} = 0$ unless $|\mathsf{k}| = k_1 + \cdots + k_n > 1$. In particular, this implies that Q^i is a polynomial in z since the right hand side of (4.32) can contain only finitely many terms. Now consider the change of variables defined by

$$w^i = z^i + \frac{2}{h_i}\,Q^i(z) \tag{4.33}$$

This transformation belongs to G_{h} by definition and satisfies

$$f = \tfrac{1}{2}h_k\,|w^k|^2 + f^*_{\bar{\imath}\bar{\jmath}kl}(w)\,\bar{w}^i\bar{w}^j w^k w^l, \tag{4.34}$$

for some functions $f^*_{\bar{\imath}\bar{\jmath}kl}$ with the same symmetry and reality properties as the corresponding $f_{\bar{\imath}\bar{\jmath}kl}$.

Since $\mathsf{L}_X\,\phi = 2f$ and $\phi(0) = 0$, it follows that ϕ has a power series expansion

$$\phi = |w^k|^2 + E_{\bar{\imath}\bar{\jmath}kl}(w)\,\bar{w}^i\bar{w}^j w^k w^l, \tag{4.35}$$

as desired. The uniqueness of the transformation $\Psi = (w^i)$ up to composition with an element of $\mathrm{U}(n) \cap G_{\mathsf{h}}$ is now evident. □

4.4.4. *Totally geodesic submanifolds.* — Since the fixed locus of an isometry of g must be totally geodesic, one has the following result:

***Proposition 12* (Geodesic subspaces).** — *If h_i has multiplicity $\mu_i > 0$ and has the property that, for all k, $h_k \neq m h_i$ for any integer $m > 1$, then the μ_i-plane in $\mathbb{C}^n$ defined by $z^j = 0$ when $h_j \neq h_i$ is totally geodesic.*

More generally, if Y has a periodic point q with period $T > 0$, then the union of the T-periodic points is a nontrivial totally geodesic linear subspace of $\mathbb{C}^n$ generated by the z^i-axis lines for which h_i is an integer multiple of $4\pi/T$. □

***Remark 16* (Geodesic axes).** — The reader might wonder whether or not the hypothesis of h_i having no 'supermultiples' is necessary in order for the h_i-eigenspace of Z_{h} in $\mathbb{C}^n$ to be totally geodesic.

The answer is clearly 'yes' in general Z-linearizing coordinates: For example, if $n = 2$ and $\mathsf{h} = (1, k)$ for some integer k, then, any of the curves $z^2 = \lambda(z^1)^k$ could be taken to be the z^1-axis in Z_{h}-linearizing coordinates. They all have the same tangent space at the origin, so at most one of them could be geodesic for a given gradient Kähler Ricci soliton g defined near $0 \in \mathbb{C}^2$ with associated holomorphic vector field Z_{h}.

However, if one uses g-normalized coordinates as provided by Theorem 4, there is a canonical $\mathbb{C}^{\mu_i} \subset \mathbb{C}^n$ associated to the eigenvalue h_i of multiplicity μ_i by the equations $z^j = 0$ when $h_j \neq h_i$. It is still not clear to me whether this canonical subspace is totally geodesic unless h_i satisfies the 'no supermultiples' condition.

4.4.5. *Growth of f in linearizing coordinates.* — Now that global linearizing coordinates are available, it makes sense to ask about the growth of the metric g and its related quantities in those coordinates.

One particularly useful quantity to estimate will be the size of $|\nabla f|^2(z)$ as $|z| \to \infty$. Note that, because of (4.3), the function $|\nabla f|^2$ is strictly increasing on the nonconstant flow lines of ∇f. On the other hand, $|\nabla f|^2 = 2h - R(g)$ is bounded by $2h$. Define

$$\lambda_- = \liminf_{|z|\to\infty} |\nabla f|^2(z) > 0 \quad \text{and} \quad \lambda_+ = \sup_z |\nabla f|^2(z) \leq 2h. \tag{4.36}$$

Proposition 13. — *For any $r > 0$, there exist constants $a_1 > 0$, $a_2 > 0$, b_1, and b_2 such that, for all $z \in \mathbb{C}^n$ with $|z| \geq r$,*

$$a_1 \log|z| + b_1 \leq f(z) \leq a_2 \log|z| + b_2\,. \tag{4.37}$$

Explicitly, one can take

$$a_1 = \frac{1}{h_n} \inf_{|z|=r} |\nabla f(z)|^2(z) > 0 \quad \textit{and} \quad a_2 = \frac{\lambda_+}{h_1} \leq \frac{2h}{h_1}. \tag{4.38}$$

Proof. — Fix $r > 0$ and note that there exist constants $m_r > 0$ and $M_r > 0$ such that

$$m_r \leq f(z) \leq M_r \quad \text{when } |z| = r. \tag{4.39}$$

Moreover, taking a_1 and a_2 as defined in (4.38) and using the fact that $|\nabla f|^2(z)$ and $|z|$ both increase along the flow lines of ∇f, one sees that

$$h_n\, a_1 \leq |\nabla f(z)|^2 \leq h_1\, a_2 \quad \text{when } |z| \geq r. \tag{4.40}$$

Now, the flow of $\nabla f = 2\mathrm{Re}(Z)$ in Z-linearizing coordinates is

$$\exp_{t\nabla f}(z^1, \ldots, z^n) = (\mathrm{e}^{h_1 t} z^1, \ldots, \mathrm{e}^{h_n t} z^n), \tag{4.41}$$

so, since $0 < h_1 \leq \cdots \leq h_n$, it follows that

$$\mathrm{e}^{h_1 t}|z| \leq \left|\exp_{t\nabla f}(z^1, \ldots, z^n)\right| \leq \mathrm{e}^{h_n t}|z|. \tag{4.42}$$

In particular, it follows that, for $t \geq 0$.

$$t \leq \frac{1}{h_1}\left(\log\left(\left|\exp_{t\nabla f}(z^1, \ldots, z^n)\right|\right) - \log|z|\right). \tag{4.43}$$

and

$$\frac{1}{h_n}\left(\log\left(\left|\exp_{t\nabla f}(z^1,\dots,z^n)\right|\right)-\log|z|\right)\leq t. \tag{4.44}$$

On the other hand, since $\mathsf{L}_{\nabla f}\, f = |\nabla f|^2$, it follows that

$$f(z)+h_n\, a_1\, t \leq f\big(\exp_{t\nabla f}(z^1,\dots,z^n)\big)\leq f(z)+h_1\, a_2\, t \tag{4.45}$$

for all $t \geq 0$ and z satisfying $|z| = r$. Combining this with the above inequality gives, for all $t \geq 0$ and z satisfying $|z| = r$,

$$f\big(\exp_{t\nabla f}(z^1,\dots,z^n)\big)-a_2\log\left|\exp_{t\nabla f}(z^1,\dots,z^n)\right|\leq f(z)-a_2\,\log|z|. \tag{4.46}$$

Since every $w \in \mathbb{C}^n$ with $|w| \geq r$ is of the form $w = \exp_{t\nabla f}(z)$ for some $t \geq 0$ and z with $|z| = r$, it follows that

$$f(w)\leq a_2\,\log|w|+(M_r-a_2\,\log r) \tag{4.47}$$

for all $w \in \mathbb{C}^n$ with $|w| \geq r$. Thus, taking $b_2 = M_r - a_2\,\log r$ verifies the claimed upper bound on f.

The lower bound follows by combining the lower bound on t with the lower bound on f:

$$m_r + a_1\left(\log\left(\left|\exp_{t\nabla f}(z^1,\dots,z^n)\right|\right)-\log|z|\right)\leq f\big(\exp_{t\nabla f}(z^1,\dots,z^n)\big), \tag{4.48}$$

which gives

$$(m_r - a_1\log r)+a_1\log|w|\leq f(w), \tag{4.49}$$

for all $w \in \mathbb{C}^n$ with $|w| \geq r$. □

Note that, as a function of r, the expression a_1 defined in (4.38) is increasing and its limit as $r \to \infty$ is λ_-/h_n.

Corollary 4. — *For any $\varepsilon > 0$, there exists $r > 0$ such that, for $z \in \mathbb{C}^n$ with $|z| \geq r$,*

$$\left(\frac{\lambda_-}{h_n}-\varepsilon\right)\log|z|\leq f(z)\leq\left(\frac{\lambda_+}{h_1}+\varepsilon\right)\log|z|. \tag{4.50}$$

In particular, there exist constants $b_1 > 0$ and $b_2 > 0$ such that, for all $z \in \mathbb{C}^n$ with $|z| \geq r$,

$$b_1\log|z|\leq d(z,p)\leq b_2\log|z|. \tag{4.51}$$

Proof. — The first statement follows by elementary reasoning from Proposition 13 while the second follows by combining the first with Corollary 3. □

Note that Corollary 4 implies that the ratio $f(z)/\log|z|$ is bounded above and has a positive lower bound as $|z| \to \infty$. Set

$$\mu_- = \liminf_{|z|\to\infty}\frac{f(z)}{\log|z|} \qquad\text{and}\qquad \mu_+ = \limsup_{|z|\to\infty}\frac{f(z)}{\log|z|}. \tag{4.52}$$

Then Corollary 4 implies

$$\frac{\lambda_-}{h_n} \leq \mu_- \leq \mu_+ \leq \frac{\lambda_+}{h_1}. \tag{4.53}$$

Proposition 14. — *One has the bounds $\mu_- \leq 2n \leq \mu_+$, in other words*

$$\liminf_{|z|\to\infty} \frac{f(z)}{\log|z|} \leq 2n \leq \limsup_{|z|\to\infty} \frac{f(z)}{\log|z|}. \tag{4.54}$$

Proof. — Suppose these bounds do not hold and let $R > 0$ be fixed large enough so that there exist positive constants a_1 and a_2 where either $a_2 < 2n$ or else $a_1 > 2n$ and positive constants b_1 and b_2 so that

$$a_1 \log|z| \leq f(z) \leq a_2 \log|z| \tag{4.55}$$

and

$$b_1 \log|z| \leq d(z,0) \leq b_2 \log|z| \tag{4.56}$$

hold whenever $|z| \geq R$. (Remember that, in these linearizing coordinates $p = 0$.)

Let $M > 0$ be sufficiently large that $d(z,0) \leq M$ when $|z| \leq R$, and consider any real number ρ that is larger than both $\log R$ and M/b_2.

Consider the g-metric ball $B_{b_1\rho}(0)$. Since $d(z,0) \leq b_1\rho$ for $z \in B_{b_1\rho}(0)$, it follows that either $|z| \leq R$ or $b_1 \log|z| \leq b_1\rho$, i.e., $|z| \leq \mathrm{e}^\rho$. Since $\mathrm{e}^\rho > R$, in either case it follows that $|z| \leq \mathrm{e}^\rho$. Thus $B_{b_1\rho}(0)$ is contained in the flat metric ball $B^0_{\mathrm{e}^\rho}(0)$.

On the other hand, if $|z| \leq \mathrm{e}^\rho$, then either $|z| \leq R$ or else $d(z,0) \leq b_2\rho$. In the former case, $d(z,0) \leq M \leq b_2\rho$, by construction. In either case, z lies in the g-metric ball $B_{b_2\rho}(0)$.

Thus, one has inclusions

$$B_{b_1\rho}(0) \subseteq B^0_{\mathrm{e}^\rho}(0) \subseteq B_{b_2\rho}(0). \tag{4.57}$$

Now, the volume form for g on $\mathbb{C}^n$ is

$$\mathrm{vol}_g = \mathrm{e}^{-f}\,\mathrm{vol}_0 \tag{4.58}$$

where $\mathrm{vol}_0 = \mathrm{i}^{n^2} 2^{-n} \mathrm{d}z \wedge \mathrm{d}\bar{z}$ is the volume form of the flat metric on $\mathbb{C}^n$.

Consequently, the g-volume of the g-metric ball $B_{b_2\rho}(0)$ is at least as large as the g-volume of the flat metric ball $B^0_{\mathrm{e}^\rho}(0)$ which is given by the integral

$$\begin{aligned}
\int_{|z|\leq \mathrm{e}^\rho} \mathrm{e}^{-f}\,\mathrm{vol}_0 &= \int_{|z|\leq R} \mathrm{e}^{-f}\,\mathrm{vol}_0 + \int_{|z|=R}^{|z|=\mathrm{e}^\rho} \mathrm{e}^{-f}\,\mathrm{vol}_0 \\
&\geq \int_{|z|\leq R} \mathrm{e}^{-f}\,\mathrm{vol}_0 + \int_{|z|=R}^{|z|=\mathrm{e}^\rho} |z|^{-a_2}\,\mathrm{vol}_0 \\
&= \int_{|z|\leq R} \mathrm{e}^{-f}\,\mathrm{vol}_0 + \mathrm{vol}(S^{2n-1}) \int_{s=R}^{s=\mathrm{e}^\rho} s^{2n-1-a_2}\,\mathrm{d}s
\end{aligned} \tag{4.59}$$

Now, if $a_2 < 2n$, then the above would imply

$$\operatorname{vol}\bigl(B_{b_2\rho}(0), g\bigr) \geq \int_{|z|\leq R} \mathrm{e}^{-f} \operatorname{vol}_0 + \frac{\operatorname{vol}(S^{2n-1})}{2n-a_2}\bigl(\mathrm{e}^{(2n-a_2)\rho} - R^{2n-a_2}\bigr). \tag{4.60}$$

However, because g has positive Ricci curvature, by the Bishop Comparison Theorem [**13**, Theorem 1.3] the g-volume of $B_{b_2\rho}(0)$ is bounded by a constant times ρ^{2n}. Obviously, such a bound is not compatible with (4.60) for all ρ sufficiently large. Thus, $a_2 \geq 2n$.

In the other direction, the g-volume of the g-metric ball $B_{b_1\rho}(0)$ is at most as large as the g-volume of the flat metric ball $B^0_{\mathrm{e}^\rho}(0)$, which obeys the upper bound

$$\begin{aligned}
\int_{|z|\leq \mathrm{e}^\rho} \mathrm{e}^{-f} \operatorname{vol}_0 &= \int_{|z|\leq R} \mathrm{e}^{-f} \operatorname{vol}_0 + \int_{|z|=R}^{|z|=\mathrm{e}^\rho} \mathrm{e}^{-f} \operatorname{vol}_0 \\
&\leq \int_{|z|\leq R} \mathrm{e}^{-f} \operatorname{vol}_0 + \int_{|z|=R}^{|z|=\mathrm{e}^\rho} |z|^{-a_1} \operatorname{vol}_0 \\
&= \int_{|z|\leq R} \mathrm{e}^{-f} \operatorname{vol}_0 + \operatorname{vol}(S^{2n-1}) \int_{s=R}^{s=\mathrm{e}^\rho} s^{2n-1-a_1}\, \mathrm{d}s
\end{aligned} \tag{4.61}$$

If $a_1 > 2n$, then this would imply

$$\operatorname{vol}\bigl(B_{b_1\rho}(0), g\bigr) \leq \int_{|z|\leq R} \mathrm{e}^{-f} \operatorname{vol}_0 + \frac{\operatorname{vol}(S^{2n-1})}{a_1-2n}\bigl(R^{2n-a_1} - \mathrm{e}^{(2n-a_1)\rho}\bigr). \tag{4.62}$$

and the right hand side is bounded as a function of ρ. Thus, $\operatorname{vol}\bigl(B_{b_1\rho}(0), g\bigr)$ would be bounded, independent of ρ, which, because g is complete and of positive Ricci curvature on the noncompact manifold $\mathbb{C}^n$, violates Theorem 4.1 of [**13**], which asserts that g must have at least linear volume growth. Thus $a_1 \leq 2n$. □

Remark 17 **(Growth of f in examples).** — In the case of Hamilton's soliton (Example 1) and, more generally Cao's soliton (Example 3), one has $h_1 = h_n$ and $\lambda_- = \lambda_+ = 2nh_1$, so equality holds in the bounds of Proposition 14.

On the other hand for the product examples (Example 2),

$$f(z) = \sum_{k=1}^{n} \log\bigl(1 + (h_k/c_k)|z^k|^2\bigr) \tag{4.63}$$

which satisfies

$$\liminf_{|z|\to\infty} \frac{f(z)}{\log|z|} = 2 \qquad \text{while} \qquad \limsup_{|z|\to\infty} \frac{f(z)}{\log|z|} = 2n. \tag{4.64}$$

In particular, note that this implies $\lambda_- \leq 2h_n < 2h$.

Remark 18 **(Growth of the potential ϕ).** — Let ϕ be the $\mathbb{T}_{\mathsf{h}}$-invariant potential for g, i.e., $\Omega = \frac{\mathrm{i}}{2}\,\partial\bar{\partial}\phi$, and assume that ϕ is normalized so that $\phi(0) = 0$.

Since $\mathsf{L}_{\nabla f}\,\phi = f$, it follows that ϕ is determined in terms of f and that Corollary 4 implies growth bounds for ϕ as well. For example, one sees that there exist positive constants r, c_1, and c_2 so that, whenever $|z| \geq r$, one has

$$c_1\left(\log|z|\right)^2 \leq \phi(z) \leq c_2\left(\log|z|\right)^2. \tag{4.65}$$

It should be possible to derive C^2-bounds on ϕ (and hence on g) using the fact that ϕ satisfies an elliptic Monge-Ampére equation, but I do not see, at present, a good way to do this so as to get any useful information.

5. The toric case

In this last section, some remarks will be made about the reduction of the gradient Kähler Ricci soliton equation in the 'toric' case, which will now be defined.

Throughout this section, $\mathbb{T}^n$ will denote the maximal abelian subgroup of $\mathrm{U}(n)$ that consists of diagonal matrices. Although there is no symplectic form specified on $\mathbb{C}^n$, the mapping $\mu_n : \mathbb{C}^n \to \mathbb{R}^n$ defined by

$$\mu_n(z^1,\dots,z^n) = \left(|z^1|^2,\dots,|z^n|^2\right) \tag{5.1}$$

will sometimes be referred to as the 'momentum mapping' of $\mathbb{T}^n$.

***Definition 4* (Toric metrics).** — A $\mathbb{T}^n$-invariant Kähler metric g that is defined on a connected $\mathbb{T}^n$-invariant open neigborhood of $0 \in \mathbb{C}^n$ will be said to be *toric*.

***Remark 19* (Toric ubiquity).** — While, at first glance, the toric condition seems to be rather special, note that any gradient Kähler Ricci soliton g on a neighborhood of $0 \in \mathbb{C}^n$ that has $(Z, \Upsilon) = (Z_{\mathsf{h}}, \mathrm{d}z)$ as its associated holomorphic data is invariant under the torus $\mathbb{T}_{\mathsf{h}}$. If h is 'generic' in the sense that the real numbers $h_1,\dots,h_n$ are linearly independent over $\mathbb{Q}$, then $\mathbb{T}_{\mathsf{h}} = \mathbb{T}^n$ and hence g is toric.

Thus, in some sense, the toric case is 'generic' among complete gradient Kähler Ricci solitons with positive Ricci curvature.

5.1. Symmetry reduction in the toric case.

— Assuming an n-torus symmetry allows one to reduce the number of independent variables in the gradient Kähler Ricci soliton equation (3.4).

Proposition 15. — *Let g be a toric gradient Kähler Ricci solition defined on a connected open neighborhood of $0 \in \mathbb{C}^n$ with a nonzero associated holomorphic vector field Z and holomorphic volume form Υ (defined with respect to a Ricci potential f satisfying $f(0) = 0$). Then*

1. *The vector field Z is linearized in the coordinates $z = (z^i)$, so that $Z = Z_{\mathsf{h}}$ for some nonzero $\mathsf{h} = (h_1,\dots,h_n) \in \mathbb{R}^n$;*

2. *The n-form* Υ *is* $c\,\mathrm{d}z^1 \wedge \cdots \mathrm{d}z^n$ *for some nonzero constant* c*; and*
3. g *has a unique Kähler potential satisfying* $\phi(0) = 0$ *of the form*

$$\phi(z) = u\big(|z^1|^2, |z^2|^2, \ldots, |z^n|^2\big) \tag{5.2}$$

for some real-analytic function u *defined on an open neighborhood of* $0 \in \mathbb{R}^n$*. Moreover,* u *satisfies the singular real Monge-Ampère equation*

$$\det_{1\leq i,j\leq n}\left(r^i \frac{\partial}{\partial r^i}\left(r^j \frac{\partial u}{\partial r^j}\right)\right) \exp\left(\frac{1}{2}\sum_{j=1}^{n} h_j\, r^j \frac{\partial u}{\partial r^j}\right) = |c|^2\, r^1 r^2 \cdots r^n\,. \tag{5.3}$$

where

$$\prod_{j=1}^{n} \frac{\partial u}{\partial r^j}(0) = |c|^2 \qquad \textit{and} \qquad \frac{\partial u}{\partial r^j}(0) > 0, \qquad 1 \leq j \leq n. \tag{5.4}$$

Conversely, for any nonzero $\mathsf{h} \in \mathbb{R}^n$ *and any nonzero complex constant* c*, if* u *is a real-analytic function defined on an open neighborhood of* $0 \in \mathbb{R}^n$ *that satisfies* (5.3) *and* (5.4)*, then the function* ϕ *defined on a* $\mathbb{T}^n$*-invariant neighborhood of* $0 \in \mathbb{C}^n$ *by* (5.2) *is the Kähler potential of a toric gradient Kähler Ricci soliton on the open neighborhood of* $0 \in \mathbb{C}^n$ *where it is strictly pseudo-convex.*

Proof. — To begin with, let me point out a fact that will be used several times in the following argument: Any $\mathbb{T}^n$-invariant holomorphic function defined on a connected open neighborhood of $0 \in \mathbb{C}^n$ is constant there. This follows, for example, by examining the effect of $\mathbb{T}^n$ on the individual terms in the power series of such a function.

Now, since g is toric, its associated holomorphic vector field Z is invariant under the action of $\mathbb{T}^n$ and hence must vanish at $0 \in \mathbb{C}^n$ and commute with each of the scaling vector fields $Z_i = z^i \frac{\partial}{\partial z^i}$. It follows easily that $Z = Z_{\mathsf{h}}$ for some $\mathsf{h} \in \mathbb{R}^n$. (For the definition of Z_{h}, see (3.23).)

Let f be the unique $\mathbb{T}^n$-invariant Ricci potential for g that satisfies $f(0) = 0$ and let Υ be a holomorphic volume form associated to g and f. Since Υ is uniquely determined up to a complex number of modulus 1, it follows that, under the action of $\mathbb{T}^n$, Υ must transform multiplicitively by a character of $\mathbb{T}^n$. It then follows easily that $\Upsilon = c\,\mathrm{d}z$ for some nonzero constant c.

Let ϕ be the unique $\mathbb{T}^n$-invariant Kähler potential for g that satisfies $\phi(0) = 0$. As has already been remarked, ϕ is real-analytic and so can be expanded as a convergent power series in the variables z^i and $\bar{z}^i$. However, $\mathbb{T}^n$-invariance evidently implies that this power series can be collected in terms of the quantities $r^i = |z^i|^2$. Thus, the existence of a function u satisfying (5.2) follows.

As argued in § 3.2, the quantity $2f - \partial\phi(Z_{\mathsf{h}})$ is a holomorphic function on a neighborhood of $0 \in \mathbb{C}^n$. By construction, it, too, is $\mathbb{T}^n$-invariant and vanishes at $0 \in \mathbb{C}^n$, which implies that it vanishes identically. Thus, $\partial\phi(Z_{\mathsf{h}}) = \mathrm{d}\phi(X_{\mathsf{h}}) = 2f$.

The rest of the argument follows by substituting the formula (5.2) into (3.4), multiplying by $r^1 \cdots r^n$, and rearranging terms, which gives (5.3).

Note that the stated positivity conditions on the first derivatives of u are needed in order that the corresponding ϕ be strictly pseudo-convex in a neighborhood of $0 \in \mathbb{C}^n$ and the relation with $|c|^2$ follows by computing the coefficient of $r^1 \cdots r^n$ in the power series expansion of the left hand side of (5.3).

The converse statement follows by computation. □

***Remark 20* (Normalizations).** — Given a solution u to (5.3) that satisfies $u(0) = 0$, one can obviously scale in the individual coordinates so as to arrange that

$$\phi = r^1 + \cdots + r^n + O\big(|r|^2\big), \tag{5.5}$$

thereby reducing to the case $|c| = 1$, so it suffices to consider this case. Note also that the resulting Kähler soliton g is already in the normalized form guaranteed by Theorem 4.

***Remark 21* (pseudo-convexity of toric potentials).** — A $\mathbb{T}^n$-invariant function ϕ of the form (5.2), i.e., $\phi = u \circ \mu_n$ for some u defined on a domain $V \subset \mathbb{R}^n$, is strictly pseudo-convex on the domain $(\mu_n)^{-1}(V) \subset \mathbb{C}^n$ if and only if the symmetric matrix

$$\left(\delta_{ij}\,\frac{\partial u}{\partial r^j} + \sqrt{r^i r^j}\,\frac{\partial^2 u}{\partial r^i \partial r^j}\right) \tag{5.6}$$

is positive definite on the part of V that lies in the orthant defined by the inequalities $r^i \geq 0$.

5.1.1. *A singular initial value problem.* — Although (5.3) is singular along the hypersurfaces $r^i = 0$ in $\mathbb{R}^n$, it turns out that the methods of Gérard and Tahara [**11**] can be used to prove an extension theorem.

Theorem 5. — *Let v be a real-analytic function on an open subset $V \subset \mathbb{R}^{n-1}$ with the property that $\psi = v \circ \mu_{n-1}$ is strictly pseudo-convex on $(\mu_{n-1})^{-1}(V) \subset \mathbb{C}^{n-1}$.*

Then there exists an open neighborhood $U \subset \mathbb{R}^n$ of $V \times \{0\}$ and a real-analytic function u on U with the properties

1. *$u(r^1, \ldots, r^{n-1}, 0) = v(r^1, \ldots, r^{n-1})$ for $(r^1, \ldots, r^{n-1}) \in V$;*
2. *u satisfies (5.3) with $|c| = 1$; and*
3. *$\phi = u \circ \mu_n$ is strictly pseudo-convex on ${\mu_n}^{-1}(U) \subset \mathbb{C}^n$.*

Moreover, u is locally unique in the sense that any for any other pair $(\tilde{U}, \tilde{u})$ with these properties, there is an open neigborhood W of $V \times \{0\}$ contained in $U \cap \tilde{U}$ such that u and $\tilde{u}$ agree on W.

Proof. — For the sake of clarity, write $t = r^n$ and let the lower case latin indices run from 1 to $n-1$. Then after dividing both sides of (5.3) (with $|c| = 1$) by $r^1 \cdots r^{n-1}$ and the exponential factor, this equation takes the form

$$\det \begin{pmatrix} \delta_{ij}\dfrac{\partial u}{\partial r^i} + r^j \dfrac{\partial^2 u}{\partial r^i \partial r^j} & \dfrac{\partial(tu_t)}{\partial r^i} \\ r^j \dfrac{\partial(tu_t)}{\partial r^j} & (t\,\partial_t)^2 u \end{pmatrix} = t\,\mathrm{e}^{\left(-\frac{h_n}{2}(tu_t) - \frac{1}{2}\sum_{j=1}^{n-1} h_j\, r^j \frac{\partial u}{\partial r^j}\right)}. \tag{5.7}$$

Note the first crucial aspect of this equation, which is that the t-derivatives of u occur as either tu_t or $t(tu_t)_t = (t\partial_t)^2 u$, i.e., as the 'regular singular' versions of the t-derivatives at $t = 0$.

Expanding the left hand side of (5.7) along the last column shows that this equation can be written in the form

$$\begin{aligned} \det\left(\delta_{ij}\frac{\partial u}{\partial r^i} + r^j \frac{\partial^2 u}{\partial r^i \partial r^j}\right)\left((t\,\partial_t)^2 u\right) &= t\,\mathrm{e}^{\left(-\frac{h_n}{2}(tu_t) - \frac{1}{2}\sum_{j=1}^{n-1} h_j\, r^j \frac{\partial u}{\partial r^j}\right)} \\ &\quad + Q_{ij}\left(r, \frac{\partial u}{\partial r}, \frac{\partial^2 u}{\partial r^2}\right)\frac{\partial(tu_t)}{\partial r^i}\,\frac{\partial(tu_t)}{\partial r^j} \end{aligned} \tag{5.8}$$

where $Q_{ij} = Q_{ji}$ are certain polynomials in the variables r^i and the first and second derivatives of u with respect to the variables r^i.

In particular, note that the right hand side of (5.8) is an entire analytic function of the variables r^i and t, the first and second derivatives of u with respect to the variables r^i, the expression tu_t and its first derivatives with respect to the r^i.

In what follows, it will be particularly important that this right hand side is also in the ideal generated by t and the quadratic expressions $\frac{\partial(tu_t)}{\partial r^i}\,\frac{\partial(tu_t)}{\partial r^j}$.

Now, set

$$u(r^1, \ldots, r^{n-1}, t) = v(r^1, \ldots, r^{n-1}) + z(r^1, \ldots, r^{n-1}, t) \tag{5.9}$$

and define

$$F_{ij}(r^1, \ldots, r^{n-1}, t) = \delta_{ij}\frac{\partial v}{\partial r^i} + r^j \frac{\partial^2 v}{\partial r^i \partial r^j} \tag{5.10}$$

Note that, by hypothesis, $\det\big(F_{ij}(r, 0)\big) \neq 0$ for $r \in V \subset \mathbb{R}^{n-1}$. In particular, the expression

$$\det\left(F_{ij}(r, t) + \delta_{ij}\frac{\partial z}{\partial r^i} + r^j \frac{\partial^2 z}{\partial r^i \partial r^j}\right), \tag{5.11}$$

which is what the coefficient of $(t\partial_t)^2 u$ on the left hand side of (5.8) becomes when one substitutes $u = v + z$ into that equation, is an analytic expression in $r \in V$, t, and the partials of z that is non-vanishing on V when one sets $t = z = 0$.

Thus, substituting $u = v + z$ into (5.8) and dividing by the determinant factor yields an equation for z of the form

$$(5.12)\qquad (t\,\partial_t)^2 z = E\left(r, t, z, \frac{\partial z}{\partial r^i}, tz_t, \frac{\partial^2 z}{\partial r^i \partial r^j}, \frac{\partial(tz_t)}{\partial r^i}\right)$$

where the function E is

1. real-analytic on an open neighborhood of $V \times \{0\}$ in $V \times \mathbb{R} \times \mathbb{R}^{1+n+\frac{1}{2}n(n+1)}$ and
2. in the ideal generated by t and the products of *pairs* of the last $(n-1)$ variables (i.e., the 'slots' containing the entries $\frac{\partial(tz_t)}{\partial r^i}$).

Now, turning to Chapter 8 of Gèrard and Tahara [**11**], one sees that (5.12) is of the form to which their Theorem 8.0.3 applies. [(6)] Consequently, (5.12) has a unique real-analytic solution $z(r,t)$ (defined on some neighborhood of $V \times \{0\} \subset \mathbb{R}^n$) that satisfies the initial condition

$$(5.13)\qquad z(r^1, \ldots, r^{n-1}, 0) = 0 \qquad \text{for } (r^1, \ldots, r^{n-1}) \in V.$$

Using this solution z to define u via (5.9), one sees that (5.7) has a correspondingly unique real-analytic solution satisfying the initial condition

$$(5.14)\qquad u(r^1, \ldots, r^{n-1}, 0) = v(r^1, \ldots, r^{n-1}) \qquad \text{for } (r^1, \ldots, r^{n-1}) \in V,$$

as claimed. The existence of an open neighborhood U of $V \times \{0\}$ such that $\phi = u \circ \mu_n$ is strictly pseudo-convex on $(\mu_n)^{-1}(U) \subset \mathbb{C}^n$ is routine. □

***Corollary 5* (Singular initial value problem for toric solitons).** — *Let g' be a real-analytic toric Kähler metric on a $\mathbb{T}^{n-1}$-invariant, connected open neighborhood $V \subset \mathbb{C}^{n-1}$ of 0.*

Then, for any $\mathsf{h} \in \mathbb{R}^n$ there exists a $\mathbb{T}^n$-invariant open neighborhood $U_\mathsf{h} \subset \mathbb{C}^n$ of $V \times \{0\}$ and a toric gradient Kähler Ricci soliton g_h on U_h whose pullback to V is g', whose associated vector field is Z_h, and whose associated holomorphic volume form with respect to its $\mathbb{T}^n$-invariant Ricci potential f_h vanishing at $0 \in \mathbb{C}^n$ is $\Upsilon = \mathrm{d}z^1 \wedge \cdots \wedge \mathrm{d}z^n$.

Moreover, g_h is locally unique in that any extension of g' with these properties agrees with g_h on some open neighborhood of $V \times \{0\}$.

***Remark 22* (Contrast in initial value problems).** — Note that Corollary 5 has a very different character from Theorem 2. Not only is the nature of the initial data different, but, in the case of Corollary 5, one is imposing initial conditions along a submanifold that is everywhere *tangent* to the holomorphic vector field $Z = Z_\mathsf{h}$, rather than

(6) While I do not want to state their full theorem here, I will give the gist: The two properties listed for the function E are easily seen to imply that there exists a unique formal power series solution of the form $z(r,t) = z_1(r)\,t + z_2(r)\,t^2 + \cdots$ to (5.12). The main import of the quoted Theorem 8.0.3 is that this series actually converges to an analytic solution on some open neighborhood of $V \times \{0\}$. (The need for a theorem is caused by the singularity at $t = 0$, which renders the standard method of majorants ineffective in proving the convergence of the formal series.)

everywhere *transverse.* The difference, of course, is that Corollary 5 addresses a singular initial value PDE problem that is, in many ways the analogue of the sort of ODE problem one encounters in the theory of regular singular points of ODE.

Because the generalization of the ODE concept of 'regular singular point' to the case of PDE is very delicate (cf. the book of Gèrard and Tahara), it is somewhat remarkable that this theory actually applies in this case.

5.1.2. *A Lagrangian formulation.* — While the reduced equation (5.3) is singular along the hypersurfaces $r^i = 0$, it is regular on the open simplicial cone defined by $r^i > 0$. Indeed, setting $r^i = \mathrm{e}^{\rho^i}$, the equation (5.3) with $|c|^2 = 1$ can be written in the form

$$\det_{1\le i,j\le n}\left(\frac{\partial^2 u}{\partial\rho^i\partial\rho^j}\right)\mathrm{e}^{\left(\frac{h_1}{2}\frac{\partial u}{\partial\rho^1}+\cdots+\frac{h_n}{2}\frac{\partial u}{\partial\rho^n}\right)} = \mathrm{e}^{\rho^1+\cdots+\rho^n}. \tag{5.15}$$

Setting $u_i = \frac{\partial u}{\partial\rho^i}$, this can be further rewritten into the form

$$\mathrm{e}^{\left(\frac{h_1}{2}u_1+\cdots+\frac{h_n}{2}u_n\right)}\,\mathrm{d}u_1\wedge\cdots\wedge\mathrm{d}u_n = \mathrm{e}^{\rho^1+\cdots+\rho^n}\,\mathrm{d}\rho^1\wedge\cdots\wedge\mathrm{d}\rho^n. \tag{5.16}$$

Thus, on $\mathbb{R}^{2n+1}$ with coordinates u, ρ^i, u_i, if one defines the contact form

$$\theta = \mathrm{d}u - u_i\,\mathrm{d}\rho^i \tag{5.17}$$

and the closed θ-primitive [(7)] n-form

$$\Psi = \mathrm{e}^{\left(\frac{h_1}{2}u_1+\cdots+\frac{h_n}{2}u_n\right)}\,\mathrm{d}u_1\wedge\cdots\wedge\mathrm{d}u_n - \mathrm{e}^{\rho^1+\cdots+\rho^n}\,\mathrm{d}\rho^1\wedge\cdots\wedge\mathrm{d}\rho^n, \tag{5.18}$$

Then the solutions of the original equation (5.3) correspond to the integral manifolds of the Monge-Ampère ideal

$$\mathcal{I} = \langle\theta, \mathrm{d}\theta, \Psi\rangle. \tag{5.19}$$

Since Ψ is closed and $\mathrm{d}\theta\wedge\Psi = 0$, the $(n{+}1)$-form $\Pi = \theta\wedge\Psi$ is closed and hence is the Poincaré-Cartan form (see [**1**]) of a contact Lagrangian for the function u. In particular, it follows by Noether's Theorem that the symmetries of the Poincaré-Cartan form give conservation laws for the reduced equation.

This is interesting because this system turns out to have a number of symmetries that are not apparent from the symmetries of the original equation.

***Remark 23* (Affine symmetries and equivalences).** — For example, consider the affine transformations on $\mathbb{R}^{2n+1}$ of the form

$$\begin{aligned}\bar u &= s\,u + a_iB^i_k\,\rho^k + c\,,\\ \bar u_i &= A^j_i\,u_j + a_i\,,\\ \bar\rho^i &= B^i_j\,\rho^j + b^i\end{aligned} \tag{5.20}$$

(7) If (M^{2n+1},θ) is a contact manifold of dimension $2n{+}1$, then an n-form Ψ on M is said to be *θ-primitive* if $\mathrm{d}\theta\wedge\Psi\equiv 0 \mod \theta$.

where A^i_j, B^i_j, $s \neq 0$, a_i, b^i, and c are real constants satisfying the n^2+2n+1 equations

$$
\begin{aligned}
A^j_i B^i_k &= s\,\delta^j_i\,,\\
\sum_i h_i A^j_i &= h_j \quad \text{for } 1 \le j \le n\,,\\
\sum_i B^i_j &= 1 \quad \text{for } 1 \le j \le n\,,\\
\mathrm{e}^{\left(\frac{h_1}{2}a_1+\cdots+\frac{h_n}{2}a_n\right)} \det(A) &= \mathrm{e}^{b^1+\cdots+b^n} \det(B)\,.
\end{aligned} \tag{5.21}
$$

Such transformations, which constitute a Lie group of dimension n^2+1, preserve the forms θ and Υ up to constant multiples and hence preserve the system $\mathcal{I}$.

Obviously, the system depends on the vector $\mathsf{h} = (h_1, \ldots, h_n)$. However, by leaving off the second of the above four conditions, one finds transformations that define equivalences between any two systems with $h = h_1 + \cdots + h_n \neq 0$ and any two systems with $h = h_1 + \cdots + h_n = 0$ but $\mathsf{h} \neq 0$. (The system corresponding to $\mathsf{h} = 0$ is, of course, the system that gives Ricci-flat toric Kähler metrics.)

***Remark 24* (Algebraic coordinates).** — The function u is, in some sense, not that important, since only the derivatives of u appear in the formula for the metric. Thus, one can actually formulate the essential part of the exterior differential system as a system on $\mathbb{R}^{2n}$.

Assuming that none of the h_i are zero, one can coordinatize the system algebraically as follows: Set $v_i = \mathrm{e}^{\frac{1}{2}h_i u_i}$. Then the form Υ can, after multiplying by a constant, be written in the form

$$\Upsilon = \mathrm{d}v_1 \wedge \cdots \wedge \mathrm{d}v_n - \frac{h_1 \cdots h_n}{2^n}\,\mathrm{d}r^1 \wedge \cdots \wedge \mathrm{d}r^n, \tag{5.22}$$

and the contact condition that $\mathrm{d}u - u_i\,\mathrm{d}\rho^i = 0$ can be replaced by the condition

$$\sum_{i=1}^{n} \frac{2}{h_i}\frac{\mathrm{d}v_i}{v_i} \wedge \frac{\mathrm{d}r^i}{r^i} = 0. \tag{5.23}$$

References

[1] R. Bryant, P. Griffiths & D. Grossman – *Exterior differential systems and Euler-Lagrange partial differential equations*, Chicago Lectures in Mathematics, University of Chicago Press, 2003.

[2] H.-D. Cao – "Existence of gradient Kähler-Ricci solitons", in *Elliptic and parabolic methods in geometry (Minneapolis, MN, 1994)*, A K Peters, 1996, p. 1–16.

[3] ———, "Limits of solutions to the Kähler-Ricci flow", *J. Differential Geom.* **45** (1997), p. 257–272.

[4] ———, "On dimension reduction in the Kähler-Ricci flow", *Comm. Anal. Geom.* **12** (2004), p. 305–320.

[5] H.-D. Cao & B. Chow – "Recent developments on the Ricci flow", *Bull. Amer. Math. Soc. (N.S.)* **36** (1999), p. 59–74.

[6] H.-D. CAO & R. S. HAMILTON – "Gradient Kähler-Ricci solitons and periodic orbits", *Comm. Anal. Geom.* **8** (2000), p. 517–529.

[7] A. CHAU & O. C. SCHNÜRER – "Stability of gradient Kähler-Ricci solitons", *Comm. Anal. Geom.* **13** (2005), p. 769–800.

[8] A. CHAU & L.-F. TAM – "A note on the uniformization of gradient Kähler Ricci solitons", *Math. Res. Lett.* **12** (2005), p. 19–21.

[9] ______, "Gradient Kähler-Ricci solitons and complex dynamical systems", in *Recent progress on some problems in several complex variables and partial differential equations*, Contemp. Math., vol. 400, Amer. Math. Soc., 2006, p. 43–52.

[10] X. CHEN – "Recent progress in Kähler geometry", in *Proceedings of the International Congress of Mathematicians, Vol. II (Beijing, 2002)*, Higher Ed. Press, 2002, p. 273–282.

[11] R. GÉRARD & H. TAHARA – *Singular nonlinear partial differential equations*, Aspects of Mathematics, Friedr. Vieweg & Sohn, 1996.

[12] R. S. HAMILTON – "The Ricci flow on surfaces", in *Mathematics and general relativity (Santa Cruz, CA, 1986)*, Contemp. Math., vol. 71, Amer. Math. Soc., 1988, p. 237–262.

[13] R. SCHOEN & S.-T. YAU – *Lectures on differential geometry*, Conference Proceedings and Lecture Notes in Geometry and Topology, I, International Press, 1994.

[14] G. TIAN & X. ZHU – "Uniqueness of Kähler-Ricci solitons", *Acta Math.* **184** (2000), p. 271–305.

R. BRYANT, Duke University Mathematics Department, P.O. Box 90320, Durham, NC 27708-0320, U.S.A. • *E-mail :* bryant@math.duke.edu • *Url :* http://www.math.duke.edu/~bryant

Astérisque

DENIS AUROUX

Special Lagrangian Fibrations, Mirror Symmetry and Calabi-Yau Double Covers

Astérisque, tome 321 (2008), p. 99-128
<http://www.numdam.org/item?id=AST_2008__321__99_0>

© Société mathématique de France, 2008, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (http://smf4.emath.fr/Publications/Asterisque/) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM
Article numérisé dans le cadre du programme Numérisation de documents anciens mathématiques
http://www.numdam.org/

Astérisque
321, 2008, p. 99–128

SPECIAL LAGRANGIAN FIBRATIONS, MIRROR SYMMETRY AND CALABI-YAU DOUBLE COVERS

by

Denis Auroux

To Jean Pierre Bourguignon on his 60^th birthday, with my most sincere gratitude for the time he spent guiding me through the process of becoming a mathematician.

Abstract. — The first part of this paper is a review of the Strominger-Yau-Zaslow conjecture in various settings. In particular, we summarize how, given a pair (X, D) consisting of a Kähler manifold and an anticanonical divisor, families of special Lagrangian tori in $X \setminus D$ and weighted counts of holomorphic discs in X can be used to build a Landau-Ginzburg model mirror to X. In the second part we turn to more speculative considerations about Calabi-Yau manifolds with holomorphic involutions and their quotients. Namely, given a hypersurface H representing twice the anticanonical class in a Kähler manifold X, we attempt to relate special Lagrangian fibrations on $X \setminus H$ and on the (Calabi-Yau) double cover of X branched along H; unfortunately, the implications for mirror symmetry are far from clear.

***Résumé* (Fibrations lagrangiennes spéciales, symétrie miroir et revêtements doubles de Calabi-Yau)**

La première partie de cet article concerne la conjecture de Strominger-Yau-Zaslow dans diverses situations. En particulier nous décrivons comment, étant donnés une variété kählerienne X et un diviseur anticanonique D, un miroir de X dans la catégorie des modèles de Landau-Ginzburg peut être construit en considérant une famille de tores lagrangiens spéciaux dans $X \setminus D$ et en comptant des disques holomorphes dans X. La seconde partie est consacrée à des considérations plus spéculatives concernant les variétés de Calabi-Yau équipées d'une involution holomorphe et leurs quotients. Autrement dit, étant donnée une hypersurface H représentant le double de la classe anticanonique dans une variété kählerienne X, nous tentons d'établir un lien entre les fibrations lagrangiennes spéciales sur $X \setminus H$ et sur le revêtement double de X ramifié le long de H, qui est une variété de Calabi-Yau ; malheureusement, les conséquences pour la symétrie miroir sont loin d'être évidentes.

2000 Mathematics Subject Classification. — 14J32, 53D12.
Key words and phrases. — Mirror symmetry, special Lagrangian fibrations, Strominger-Yau-Zaslow conjecture, Calabi-Yau double covers.

Partially supported by NSF grants DMS-0600148 and DMS-0652630.

© Astérisque 321, SMF 2008

1. Introduction

The phenomenon of mirror symmetry was first evidenced for Calabi-Yau manifolds, i.e. Kähler manifolds with holomorphically trivial canonical bundle. Subsequently it became apparent that mirror symmetry also holds in a more general setting, if one enlarges the class of objects under consideration (see e.g. [**14**]); namely, one should allow the mirror to be a *Landau-Ginzburg model*, i.e. a pair consisting of a non-compact Kähler manifold and a holomorphic function on it (called *superpotential*).

Our motivation here is to understand how to *construct* the mirror manifold, starting from examples where the answer is known and extrapolating to less familiar situations; generally speaking, the verification of the mirror symmetry conjectures for the manifolds obtained by these constructions falls outside the scope of this paper.

The geometric understanding of mirror symmetry in the Calabi-Yau case relies on the Strominger-Yau-Zaslow (SYZ) conjecture [**28**], which roughly speaking postulates that mirror pairs of Calabi-Yau manifolds carry dual fibrations by special Lagrangian tori, and its subsequent refinements (see e.g. [**10, 21**]). This program can be extended to the non Calabi-Yau case, as suggested by Hori [**12**] and further investigated in [**3**]. In that case, the input consists of a pair (X, D) where X is a compact Kähler manifold and D is a complex hypersurface representing the anticanonical class. Observing that the complement of D carries a holomorphic n-form with poles along D, we can think of $X \setminus D$ as an open Calabi-Yau manifold, to which one can apply the SYZ program. Hence, one can attempt to construct the mirror of X as a (complexified) moduli space of special Lagrangian tori in $X \setminus D$, equipped with a Landau-Ginzburg superpotential defined by a weighted count of holomorphic discs in X. However, exceptional discs and wall-crossing phenomena require the incorporation of "instanton corrections" into the geometry of the mirror (see [**3**]).

One notable feature of the construction is that it provides a bridge between mirror symmetry for the Kähler manifold X and for the Calabi-Yau hypersurface $D \subset X$. Namely, the fiber of the Landau-Ginzburg superpotential is expected to be the SYZ mirror to D, and the two pictures of homological mirror symmetry (for X and for D) should be related via restriction functors (see Section 7 of [**3**] for a sketch).

In this paper, we would like to consider a slightly different situation, which should provide another relation with mirror symmetry for Calabi-Yau manifolds. The union of two copies of X glued together along D can be thought of as a singular Calabi-Yau manifold, which can be smoothed to a double cover of X branched along a hypersurface H representing twice the anticanonical class and contained in a neighborhood of D. This suggests that one might be able to think of mirror symmetry for X as a $\mathbb{Z}/2$-invariant version of mirror symmetry for the Calabi-Yau manifold Y. Unfortunately, this proposal comes with several caveats which make it difficult to implement.

Let (X, ω, J) be a compact Kähler manifold, and let H be a complex hypersurface in X representing twice the anticanonical class. Then the complement of H carries a nonvanishing section Θ of $K_X^{\otimes 2}$ with poles along H. We can think of Θ as the square of a holomorphic volume form defined up to sign. In this context, we can look for special Lagrangian submanifolds of $X \setminus H$, i.e. Lagrangian submanifolds on which the restriction of Θ is real. The philosophy of the SYZ conjecture suggests that, in favorable cases, one might be able to construct a foliation of $X \setminus H$ in which the generic leaves are special Lagrangian tori. Indeed, denote by Y the double cover of X branched along H: then Y is a Calabi-Yau manifold with a holomorphic involution. If Y carries a special Lagrangian fibration that is invariant under the involution, then by quotienting we could hope to obtain the desired foliation on $X \setminus H$; unfortunately the situation is complicated by technicalities involving the symplectic form.

Conjecture 1.1. — *For a suitable choice of H, $X \setminus H$ carries a special Lagrangian foliation whose lift to the Calabi-Yau double cover Y can be perturbed to a $\mathbb{Z}/2$-invariant special Lagrangian torus fibration.*

If $-K_X$ is effective, we can consider a situation where H degenerates to a hypersurface D representing the anticanonical class in X, with multiplicity 2. As explained above, this corresponds to the situation where Y degenerates to the union of two copies of X glued together along D. One could hope that under such a degeneration the foliation on $X \setminus H$ converges to a special Lagrangian torus fibration on $X \setminus D$. Using the mirror construction described in [**3**], one can then try to relate a Landau-Ginzburg mirror $(X^\vee, W)$ of X to a Calabi-Yau mirror $Y^\vee$ of Y. The simplest case should be when $K_{X|D}$ is holomorphically trivial (which in particular requires $c_1(X)^2 = 0$). Then $W : X^\vee \to \mathbb{C}$ is expected to have trivial monodromy around infinity (see Remark 2.11), so that $\partial X^\vee \approx S^1 \times D^\vee$ where $D^\vee$ is mirror to D. It is then tempting to conjecture that, considering only the complex structure of the mirror (and ignoring its symplectic geometry), $Y^\vee$ can be obtained by gluing together two copies of the mirror $X^\vee$ to X along their boundary $S^1 \times D^\vee$. Unfortunately, as we will see in § 3.5 this is not compatible with instanton corrections.

The rest of this paper is organized as follows. In Section 2 we review the geometry of mirror symmetry from the perspective of the SYZ conjecture, both in the Calabi-Yau case and in the more general case (relatively to an anticanonical divisor). We then turn to more speculative considerations in Section 3, where we discuss the geometry of Calabi-Yau double covers, clarify the statement of Conjecture 1.1, and consider various examples.

Acknowledgements. — I would like to thank Mohammed Abouzaid, Paul Seidel, Ludmil Katzarkov, and Dima Orlov for many fruitful discussions. I am also grateful to Ron Donagi, whose interest in this topic prompted the writing of this paper, and to the referee for insightful comments. This research was partially supported by NSF grants DMS-0600148 and DMS-0652630.

2. The SYZ conjecture and mirror symmetry

2.1. Motivation. — One of the most spectacular mathematical predictions of string theory is the phenomenon of mirror symmetry, i.e. the existence of a broad dictionary under which the symplectic geometry of a given manifold X can be understood in terms of the complex geometry of a mirror manifold $X^\vee$, and vice-versa. This dictionary works at several levels, among which perhaps the most exciting is Kontsevich's homological mirror conjecture, which states that the derived Fukaya category of X should be equivalent to the derived category of coherent sheaves of its mirror $X^\vee$ [**19**]; in the non Calabi-Yau case the categories under consideration need to be modified appropriately [**20**] (see also [**1, 13, 18, 26, 27**]).

The main goal of the Strominger-Yau-Zaslow conjecture [**28**] is to provide a geometric interpretation of mirror symmetry. Roughly speaking it says that mirror manifolds carry dual fibrations by special Lagrangian tori. In the Calabi-Yau case, one way to motivate the conjecture is to observe that, given any point p of the mirror $X^\vee$, mirror symmetry should put the skyscraper sheaf $\mathcal{O}_p$ in correspondence with some object $\mathcal{L}_p$ of the Fukaya category of X. As a graded vector space $\operatorname{Ext}^*(\mathcal{O}_p, \mathcal{O}_p)$ is isomorphic to the cohomology of T^n; therefore the most likely candidate for $\mathcal{L}_p$ is a (special) Lagrangian torus in X, equipped with a rank 1 unitary local system (a flat $U(1)$ bundle). This suggests that one should try to construct $X^\vee$ as a moduli space of pairs (L, ∇) where L is a special Lagrangian torus in X and ∇ is a flat unitary connection on the trivial line bundle over L. Since for each torus L the moduli space of flat connections can be thought of as a dual torus, we arrive at the familiar picture.

When X is not Calabi-Yau but the anticanonical class $-K_X$ is effective, we can still equip the complement of a hypersurface $D \in |-K_X|$ with a holomorphic volume form, and thus consider special Lagrangian tori in $X \setminus D$. However, in this case, holomorphic discs in X with boundary in L cause Floer homology to be *obstructed* in the sense of Fukaya-Oh-Ohta-Ono [**6**]: to each object $\mathcal{L} = (L, \nabla)$ we can associate an obstruction $\mathfrak{m}_0(\mathcal{L})$, given by a weighted count of holomorphic discs in (X, L), and the Floer differential on $CF^*(\mathcal{L}, \mathcal{L}')$ squares to $\mathfrak{m}_0(\mathcal{L}') - \mathfrak{m}_0(\mathcal{L})$. Moreover, even when the Floer homology groups $HF^*(\mathcal{L}, \mathcal{L})$ can still be defined, they are often zero, so that $\mathcal{L}$ is a trivial object of the Fukaya category. On the mirror side, these features of the theory can be replicated by the introduction of a *Landau-Ginzburg superpotential*, i.e. a holomorphic function $W : X^\vee \to \mathbb{C}$. Without getting into details, W can be thought of as an obstruction term for the B-model on $X^\vee$, playing the same role as

$\mathfrak{m}_0$ for the A-model on X. In particular, a point of $X^\vee$ defines a nontrivial object of the category of B-branes $D^b_{sing}(X^\vee, W)$ only if it is a critical point of W [**18, 24**].

2.2. Special Lagrangian fibrations and T-duality. — Let (X, ω, J) be a smooth compact Kähler manifold of complex dimension n. If X is Calabi-Yau, i.e. the canonical bundle K_X is holomorphically trivial, then X carries a globally defined holomorphic volume form $\Omega \in \Omega^{n,0}(X)$: this is the classical setting for mirror symmetry. Otherwise, assume that K_X^{-1} admits a nontrivial holomorphic section σ, vanishing along a hypersurface D. Typically we will assume that D is smooth, or with normal crossing singularities. Then $\Omega = \sigma^{-1}$ is a nonvanishing holomorphic $(n, 0)$-form over $X \setminus D$, with poles along D.

The restriction of Ω to a Lagrangian submanifold $L \subset X \setminus D$ does not vanish, and can be expressed in the form $\Omega_{|L} = \psi \operatorname{vol}_g$, where $\psi \in C^\infty(L, \mathbb{C}^*)$ and vol_g is the volume form induced on L by the Kähler metric $g = \omega(\cdot, J\cdot)$.

Definition 2.1. — *A Lagrangian submanifold $L \subset X \setminus D$ is* special Lagrangian *if the argument of ψ is constant.*

The value of the constant depends only on the homology class $[L] \in H_n(X \setminus D, \mathbb{Z})$, and we will usually arrange for it to be a multiple of $\pi/2$. For simplicity, in the rest of this paragraph we will assume that $\Omega_{|L}$ is a real multiple of vol_g.

The following classical result is due to McLean [**23**] (at least when $|\psi| \equiv 1$, which is the case in the Calabi-Yau setting; see § 9 of [**17**] or Proposition 2.5 of [**3**] for the case where $|\psi| \neq 1$):

***Proposition 2.2* (McLean).** — *Infinitesimal special Lagrangian deformations of L are in one to one correspondence with cohomology classes in $H^1(L, \mathbb{R})$. Moreover, the deformations are unobstructed.*

More precisely, a section of the normal bundle $v \in C^\infty(NL)$ determines a 1-form $\alpha = -\iota_v\omega \in \Omega^1(L, \mathbb{R})$ and an $(n-1)$-form $\beta = \iota_v \operatorname{Im}\Omega \in \Omega^{n-1}(L, \mathbb{R})$. These satisfy $\beta = \psi *_g \alpha$, and the deformation is special Lagrangian if and only if α and β are both closed. Thus special Lagrangian deformations correspond to "ψ-harmonic" 1-forms $-\iota_v\omega \in \mathcal{H}^1_\psi(L) = \{\alpha \in \Omega^1(L, \mathbb{R}) |\, d\alpha = 0,\ d^*(\psi\alpha) = 0\}$ (recall $\psi \in C^\infty(L, \mathbb{R}_+)$ is the ratio between the volume elements determined by Ω and g).

In particular, special Lagrangian tori occur in n-dimensional families, giving a local fibration structure provided that nontrivial ψ-harmonic 1-forms have no zeroes.

The base B of a special Lagrangian torus fibration carries two natural affine structures, which we call "symplectic" and "complex". The first one, which encodes the symplectic geometry of X, is given by locally identifying B with a domain in $H^1(L, \mathbb{R})$

(where $L \approx T^n$). At the level of tangent spaces, the cohomology class of $-\iota_v\omega$ provides an identification of TB with $H^1(L,\mathbb{R})$; integrating, the local affine coordinates on B are the symplectic areas swept by loops forming a basis of $H_1(L)$. The other affine structure encodes the complex geometry of X, and locally identifies B with a domain in $H^{n-1}(L,\mathbb{R})$. Namely, one uses the cohomology class of $\iota_v \mathrm{Im}\,\Omega$ to identify TB with $H^{n-1}(L,\mathbb{R})$, and the affine coordinates are obtained by integrating $\mathrm{Im}\,\Omega$ over the n-chains swept by cycles forming a basis of $H_{n-1}(L)$.

In practice, B can usually be compactified to a larger space $\bar{B}$ (with non-empty boundary in the non Calabi-Yau case), by also considering singular special Lagrangian submanifolds that arise as limits of degenerating families of special Lagrangian tori; however the affine structures are only defined on the open subset $B \subset \bar{B}$.

Ignoring singular fibers and instanton corrections, the first candidate for the mirror of X is therefore a moduli space M of pairs (L,∇), where L is a special Lagrangian torus in X (or $X \setminus D$) and ∇ is a flat $U(1)$ connection on the trivial line bundle over L (up to gauge). The local geometry of M is well-understood [**11, 22, 8, 3**], and in particular we have the following result (see e.g. § 2 of [**3**]):

Proposition 2.3. — *M carries a natural integrable complex structure $J^\vee$ arising from the identification*

$$T_{(L,\nabla)}M = \{(v,\alpha) \in C^\infty(NL) \oplus \Omega^1(L,\mathbb{R}) \mid -\iota_v\omega + i\alpha \in \mathcal{H}^1_\psi(L) \otimes \mathbb{C}\},$$

a holomorphic n-form

$$\Omega^\vee((v_1,\alpha_1),\dots,(v_n,\alpha_n)) = \int_L (-\iota_{v_1}\omega + i\alpha_1) \wedge \dots \wedge (-\iota_{v_n}\omega + i\alpha_n),$$

and a compatible Kähler form

$$\omega^\vee((v_1,\alpha_1),(v_2,\alpha_2)) = \int_L \alpha_2 \wedge \iota_{v_1}\mathrm{Im}\,\Omega - \alpha_1 \wedge \iota_{v_2}\mathrm{Im}\,\Omega$$

(this formula for $\omega^\vee$ assumes that $\int_L \mathrm{Re}\,\Omega$ has been suitably normalized).

The moduli space of pairs M can be viewed as a complexification of the moduli space of special Lagrangian submanifolds; forgetting the connection gives a projection map $f^\vee$ from M to the real moduli space B. The fibers of this projection are easily checked to be special Lagrangian tori in $(M,\omega^\vee,\Omega^\vee)$.

The special Lagrangian fibrations $f : X \to \bar{B}$ (or rather, its restriction to the open subset $f^{-1}(B)$) and $f^\vee : M \to B$ can be viewed as fiberwise dual to each other. In particular, it is easily checked that the affine structure induced on B by the symplectic geometry of $f^\vee$ coincides with that induced by the complex geometry of f, and vice-versa. Giving priority to the symplectic affine structure, we will often implicitly equip B with the affine structure induced by the symplectic geometry of X, and denote by $B^\vee$ the same manifold equipped with the other affine structure (induced by the complex geometry of X, or the symplectic geometry of M).

Thus, the philosophy of the Strominger-Yau-Zaslow conjecture is that, in first approximation (ignoring instanton corrections), mirror symmetry amounts simply to exchanging the two affine structures on B. However, in general it is not at all obvious how to extend the picture to the compactification $\bar{B}$. The reader is referred to [**28**], [**8**], [**22**] for more details in the Calabi-Yau case, and to [**12**] and [**3**] for the non Calabi-Yau case.

2.3. Mirror symmetry for Calabi-Yau manifolds. — Constructing a special Lagrangian fibration on a Calabi-Yau manifold is in general a challenging task, but there are a few situations where it can be done explicitly, for instance in the case of flat tori, or for hyperkähler manifolds. We give two well-known examples.

***Example 2.4* (Elliptic curves).** — Consider an elliptic curve $E = \mathbb{C}/(\mathbb{Z} \oplus \tau\mathbb{Z})$, where $\tau = i\gamma \in i\mathbb{R}_+$, equipped with the holomorphic volume form $\Omega = dz$ and a Kähler form ω such that $\int_E \omega = \lambda \in \mathbb{R}_+$. (The reason why we assume τ to be pure imaginary is that for simplicity we are suppressing any discussion of B-fields). Then the family of circles parallel to the real axis $\{\mathrm{Im}\,(z) = c\}$ defines a special Lagrangian fibration on E, with base $B \simeq S^1$. One easily checks that the length of B with respect to the affine metric is equal to λ for the symplectic affine structure, and γ for the complex affine structure. The mirror elliptic curve $E^\vee$ is obtained by exchanging the two affine structures on B; accordingly, it has modular parameter $\tau^\vee = i\lambda$ and symplectic area $\int_{E^\vee} \omega^\vee = \gamma$. (The reader is referred to [**25**] for a verification of homological mirror symmetry for the mirror pair $E, E^\vee$.)

***Example 2.5* (K3 surfaces).** — In the case of K3 surfaces, special Lagrangian fibrations can be built using hyperkähler geometry. Let (X, J) be an elliptically fibered K3 surface, for example obtained as the double cover of $\mathbb{CP}^1 \times \mathbb{CP}^1$ branched along a suitably chosen algebraic curve of bidegree $(4,4)$: composing the covering map with projection to the first $\mathbb{CP}^1$ factor, we obtain an elliptic fibration $f : X \to \mathbb{CP}^1$ with 24 nodal singular fibers. Equip X with a Calabi-Yau metric g, and denote the corresponding Kähler form by ω_J. Denote by Ω_J a holomorphic $(2,0)$-form on X, suitably normalized, and let $\omega_K = \mathrm{Re}(\Omega_J)$ and $\omega_I = \mathrm{Im}(\Omega_J)$: then $(\omega_I, \omega_J, \omega_K)$ is a hyperkähler triple for the metric g. Now switch to the complex structure $I = g^{-1}\omega_I$ determined by the Kähler form ω_I, and with respect to which $\Omega_I = \omega_J + i\omega_K$ is a holomorphic volume form. Since the fibers of $f : X \to \mathbb{CP}^1$ are calibrated by ω_J, the map f is a special Lagrangian fibration on (X, ω_I, Ω_I).

The affine structures on the base of f are only defined away from the singularities of the fibration. Thus the geometry of (X, ω_I, Ω_I) is characterized by a pair of affine structures on the open subset $B \simeq S^2 \setminus \{24 \text{ points}\}$ of $\bar{B} \simeq S^2$. The monodromies of the two affine structures around each singular point are the transpose of each other, and each individual monodromy is conjugate to the standard matrix $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

The mirror of (X, ω_I, Ω_I) is another K3 surface, carrying a special Lagrangian fibration whose base differs from B by an exchange of the two affine structures. In fact, under certain assumptions (e.g., existence of a section) and for a specific choice of $[\omega_J]$, the mirror may be obtained simply by performing another hyperkähler rotation to get $(X, -\omega_K, \Omega_{-K} = \omega_J + i\omega_I)$; see e.g. § 7 of [**15**]. The reader is also referred to § 7 of [**8**] for more details on the SYZ picture for K3 surfaces.

In the above examples, one can avoid confronting heads-on the delicate issues that arise when trying to reconstruct the mirror from the affine geometry of B. In general, however, the compactification of the mirror fibration over the singularities of the affine structure and the incorporation of instanton corrections are two extremely challenging aspects of this approach. The reader is referred to [**21**] and [**10**] for two attempts at tackling this problem.

Another even more important issue is constructing a special Lagrangian torus fibration on X in the first place. When there is no direct geometric construction as in the above examples, the most promising approach seems to be Gross and Siebert's program to understand mirror symmetry via toric degenerations [**9, 10**]. The main idea is to degenerate X to a union X_0 of toric varieties glued together along toric strata; toric geometry then provides a special Lagrangian fibration on X_0, whose base is a polyhedral complex formed by the union of the moment polytopes for the components of X_0. Gross and Siebert then analyze carefully the behavior of this special Lagrangian fibration upon deforming X_0 back to a smooth manifold, showing how to insert singularities into the affine structure to compensate for the nontriviality of the normal bundles to the singular strata along which the smoothing takes place. Moreover, they also show that, in the toric degeneration limit, exchanging the affine structures on the base of the special Lagrangian fibration can be understood as a combinatorial process called *discrete Legendre transform* [**9**].

Remark 2.6. — The affine geometry of B is a remarkably powerful tool to understand the symplectic and complex geometry of X (and, by exchanging the affine structures, of its mirror). Namely, away from the singularities, the two affine structures on $B = B^\vee$ each determine an integral lattice in the tangent bundle TB; denoting these lattices by Λ for the symplectic affine structure and $\Lambda^\vee$ for the complex affine structure, locally X can be identified with either one of the torus bundles T^*B/Λ^* (with its standard symplectic form) and $TB^\vee/\Lambda^\vee$ (with its standard complex structure). Thus, locally, an integral affine submanifold of B (i.e., a submanifold described by linear equations with integer coefficients in local affine coordinates with respect to the symplectic affine structure) determines a Lagrangian submanifold of X by the conormal construction. Similarly, an integral affine submanifold with respect to the complex affine structure $B^\vee$ locally determines a complex submanifold of X (by considering its tangent bundle). More generally, *tropical subvarieties* of B or $B^\vee$ determine piecewise smooth Lagrangian or complex subvarieties in X; whether these can be smoothed is a difficult problem whose answer is known only in simple cases.

To give a concrete example, let us return to K3 surfaces (Example 2.5) and the corresponding affine structures on $B \simeq S^2 \setminus \{24 \text{ points}\}$. Each singular fiber of the special Lagrangian torus fibration $f : X \to \mathbb{CP}^1$ has a nodal singularity obtained by collapsing a circle in the smooth fiber. The homology class of this vanishing cycle determines a pair of rays in B (straight half-lines emanating from the singular point), with the property that the conormal bundles to these rays compactify nicely to Lagrangian discs in X (possibly after a suitable translation within the fibers). Similarly, the nodal singularity determines a pair of rays in $B^\vee$ (different from the previous ones), whose tangent bundles (again after a suitable translation) compactify to holomorphic discs in X. When two singularities of the affine structure lie in a position such that the corresponding rays in B (resp. in $B^\vee$) align with each other (and assuming the translations in the fibers also match), the line segment joining them in B (resp. $B^\vee$) determines a Lagrangian sphere (resp. a rational curve with normal bundle $\mathcal{O}(-2)$) in X. In the mirror $X^\vee$ the same alignment produces a rational -2-curve (resp. a Lagrangian sphere). In fact, using the hyperkähler structure on X and remembering that the elliptic fibration f is J-holomorphic, these spheres correspond to (special) Lagrangian spheres in (X, ω_J) which arise from the matching path construction and additionally are calibrated by ω_K (resp. ω_I).

2.4. Mirror symmetry in the complement of an anticanonical divisor. — We now consider special Lagrangian torus fibrations in the complement $X \setminus D$ of an anticanonical divisor D in a Kähler manifold X. We start with a very easy example to make the following discussion more concrete:

Example 2.7. — Let $X = \mathbb{CP}^1$, equipped with any Kähler form invariant under the standard S^1-action. Equip the complement of the anticanonical divisor $D = \{0, \infty\}$, namely $\mathbb{CP}^1 \setminus \{0, \infty\} = \mathbb{C}^*$, with the standard holomorphic volume form $\Omega = dz/z$. It is easy to check that the circles $|z| = r$ are special Lagrangian (with phase $\pi/2$). Thus we have a special Lagrangian fibration $f : \mathbb{CP}^1 \setminus D \to B$, whose base B is homeomorphic to an interval. As seen above, B carries two affine structures. With respect to the symplectic affine structure, the special Lagrangian fibration is simply the moment map for the S^1-action on $\mathbb{CP}^1$ (up to a factor of 2π). Thus B is an open interval of length equal to the symplectic area of $\mathbb{CP}^1$, and can be compactified by adding the end points of the interval, which correspond to the S^1 fixed points, i.e. the points of D. On the other hand, with respect to the complex affine structure, B is an infinite line: from this point of view, the special Lagrangian fibration is given by the map $z \mapsto \log|z|$.

We can start building a mirror to X by considering the dual special Lagrangian torus fibration M as in § 2.2. M is a non-compact Kähler manifold and, after taking instanton corrections into account, it is in fact the mirror to the open Calabi-Yau manifold $X \setminus D$. Thus, some information is missing from this description. As explained at the end of §2.1, adding in the divisor D very much affects the special Lagrangian tori $X \setminus D$ from a Floer-theoretic point of view, and the natural way to account for the resulting obstructions is to make the mirror a Landau-Ginzburg model by introducing a superpotential $W : M \to \mathbb{C}$.

Recall that a point of M is a pair (L, ∇), where $L \subset X \setminus D$ is a special Lagrangian torus, and ∇ is a flat connection on the trivial line bundle over L. Given a homotopy class $\beta \in \pi_2(X, L)$, we can consider the moduli space of holomorphic discs in X with boundary on L, representing the class β. The virtual dimension (over $\mathbb{R}$) of this moduli space is $n - 3 + \mu(\beta)$, where $\mu(\beta) \in \mathbb{Z}$ is the Maslov index; in our case, the Maslov index is twice the algebraic intersection number $\beta \cdot [D]$ (see e.g. Lemma 3.1 of [**3**]). When $\mu(\beta) = 2$, in favorable cases we can define a (virtual) count $n_\beta(L)$ of holomorphic discs in the class β whose boundary passes through a generic point $p \in L$, and define

$$(2.1)\quad W(L, \nabla) = \sum_{\substack{\beta \in \pi_2(X,L)\\ \mu(\beta)=2}} n_\beta(L)\, z_\beta(L, \nabla), \quad \text{where} \quad z_\beta(L, \nabla) = \exp(-\textstyle\int_\beta \omega)\, \mathrm{hol}_\nabla(\partial\beta).$$

Thus, W is a weighted count of holomorphic discs of Maslov index 2 with boundary in L, with weights determined by the symplectic area of the disc and the holonomy of the connection ∇ along its boundary.

For example, in the case of $\mathbb{CP}^1$ (Example 2.7), each special Lagrangian fiber separates $\mathbb{CP}^1$ into two discs, one containing 0 and the other one containing ∞. The classes β_1 and β_2 represented by these discs satisfy $\beta_1 + \beta_2 = [\mathbb{CP}^1]$, and hence the corresponding weights satisfy $z_{\beta_1} z_{\beta_2} = \exp(-\int_{\mathbb{CP}^1} \omega)$. One can check that $n_{\beta_1} = n_{\beta_2} = 1$, so that using $z = z_{\beta_1}$ as coordinate on M we obtain the well-known formula for the superpotential, $W = z + e^{-\Lambda} z^{-1}$, where Λ is the symplectic area of $\mathbb{CP}^1$.

While the example of $\mathbb{CP}^1$ is straightforward, several warnings are in order. First, unless X is Fano the sum (2.1) is not known to converge. More importantly, if L bounds non-constant holomorphic discs of Maslov index 0 (i.e., discs contained in $X \setminus D$), then the counts $n_\beta(L)$ depend on auxiliary data, such as the point $p \in L$ through which the discs are required to pass, or an auxiliary Morse function on L.

An easy calculation shows that the weights z_β are local holomorphic functions on M (with respect to the complex structure defined in Proposition 2.3), and once all ambiguities are lifted the disc counts $n_\beta(L)$ are locally constant, so that W is locally a holomorphic function on M. However, Maslov index 0 discs determine "walls" in M, across which the counts $n_\beta(L)$ jump and hence the quantity (2.1) presents discontinuities. In terms of the affine geometry of the base of the special Lagrangian fibration, an important mechanism for the generation of walls comes from the rays in $B^\vee$ (the base with its complex affine structure) that emanate from the vanishing cycles at the singular fibers of the special Lagrangian fibration: indeed, by definition any special Lagrangian fiber that lies on such a ray bounds a holomorphic disc in $X \setminus D$ (see Remark 2.6). Intersections between these "primary" walls then generate further walls (which can be visualized as rigid tropical configurations in $B^\vee$).

Fukaya-Oh-Ohta-Ono's results [**6**] imply that the formulas for W in adjacent chambers of M differ by a holomorphic substitution of variables (see also Proposition 3.9 in [**3**]). The guiding principle that governs *instanton corrections* is that the various chambers of M should be glued to each other not in the naive manner suggested by the geometry of B, but rather via the holomorphic gluing maps that arise in the wall-crossing formulas. Thus, the instanton-corrected mirror is precisely the analytic space on which the weighted count (2.1) of holomorphic discs in (X, L), and more generally the "open Gromov-Witten invariants" of (X, L) (yet to be defined in the most general setting), become single-valued quantities. The reader is referred to [**21**] and [**10**] for more details on instanton corrections.

One final issue is that, according to Hori and Vafa [**14**], the mirror obtained by T-duality needs to be enlarged. The holomorphic volume form Ω has poles along D, which causes B equipped with the complex affine structure to have infinite diameter (after adding in the singular fibers, $B^\vee$ is complete). On the other hand, the fact that ω extends smoothly across D means that, with respect to the symplectic affine structure, B has finite diameter, and compactifies to a singular affine manifold with boundary. The consequence is that, after exchanging the affine structures, the Kähler metric on the mirror is complete but its complex structure is "incomplete": for instance, in Example 2.7 the mirror of $\mathbb{CP}^1$ is naturally a bounded annulus (of modulus equal to the symplectic area of $\mathbb{CP}^1$), rather than the expected $\mathbb{C}^*$. Hori and Vafa's suggestion (assuming that X is Fano) is to symplectically "enlarge" $X \setminus D$ by considering a family of Kähler forms $(\omega_k)_{k\to\infty}$ obtained by symplectic inflation along D, with the property that $[\omega_k] = [\omega] + k\, c_1(X)$, and simultaneously rescaling the superpotential by a factor of e^k; see also § 4.2 of [**3**]. (In some cases, this process can also be viewed as a flow that should converge to a complete Ricci-flat metric on $X \setminus D$.) However, this "renormalization" procedure does not seem desirable in the geometric setting considered in Section 3, so we do not consider it further.

We end here our discussion of the various delicate points that come up in the construction of the mirror and its superpotential, and simply refer the reader to [**3**] for more details. Instead, we return to examples.

***Example 2.8* (Toric varieties).** — Let (X, ω, J) be a toric variety of complex dimension n, and consider the toric anticanonical divisor D (i.e., the divisor of points where the T^n-action is not free). Recall that $X \setminus D$ is biholomorphic to $(\mathbb{C}^*)^n$, and equip it with the holomorphic $(n,0)$-form $\Omega = d\log z_1 \wedge \cdots \wedge d\log z_n$, which has poles along D. Then the orbits of the standard T^n-action define a special Lagrangian fibration on $X \setminus D \simeq (\mathbb{C}^*)^n$. With respect to the symplectic affine structure, the base B of this fibration is the moment polytope for (X, ω), or rather its interior, and the special Lagrangian fibration is simply given by the moment map. On the other hand, the complex affine structure on B naturally identifies it with $\mathbb{R}^n$; from this point of view the special Lagrangian fibration is the Log map $(z_1, \ldots, z_n) \mapsto (\log|z_1|, \ldots, \log|z_n|)$.

Exchanging the two affine structures, the mirror of X is naturally a bounded domain in $(\mathbb{C}^*)^n$ (the subset of points whose image under the Log map lies in the moment polytope of X), equipped with a complete Kähler metric and a superpotential W defined by a Laurent polynomial consisting of one term for each component of D. Details can be found in [**5**] and [**7**] (see also § 4 of [**3**] for a brief overview, and [**1**] for a partial verification of homological mirror symmetry).

***Example 2.9* ($\mathbb{CP}^2$).** — Consider $\mathbb{CP}^2$ equipped with the Fubini-Study Kähler form ω_0. Let $D \subset \mathbb{CP}^2$ be a smooth elliptic curve defined by a homogeneous polynomial of degree 3, and let Ω be a holomorphic volume form on $\mathbb{CP}^2$ with poles along D.

Conjecture 2.9. — *$\mathbb{CP}^2 \setminus D$ carries a special Lagrangian torus fibration over the disc with (generically) three nodal singular fibers.*

Tentatively, the construction of this special Lagrangian fibration proceeds as follows. Start with the toric setting, i.e. equip $\mathbb{CP}^2$ with a holomorphic volume form with poles along the toric anticanonical divisor D_0 consisting of the three coordinate lines ($\Omega_0 = dx \wedge dy/xy$ in an affine chart). As mentioned above, the orbits of the standard T^2-action define a special Lagrangian fibration on $(\mathbb{C}^*)^2 = \mathbb{CP}^2 \setminus D_0$; with respect to the symplectic affine structure, the base B_0 of this fibration is the moment polytope for $\mathbb{CP}^2$, i.e. a triangle. Deforming this situation to the case of a holomorphic volume form Ω' with poles along a smooth cubic curve D' obtained by smoothing out the three nodal points of D_0 modifies the structure of the special Lagrangian fibration near the three toric fixed points. A local model for what happens near each of these points is described in § 5 of [**3**]. Namely, if we replace Ω_0 by $\Omega_\varepsilon = dx \wedge dy/(xy - \varepsilon)$, then the complement of the anticanonical divisor D_ε formed by the conic $xy = \varepsilon$ and the line at infinity carries a special Lagrangian torus fibration with one nodal singular fiber: the fibers are formed by intersecting the level sets of the moment map for the S^1-action $e^{i\theta} \cdot (x, y) = (e^{i\theta}x, e^{-i\theta}y)$ with the level sets of the function $|xy - \varepsilon|^2$, and the singularity is at the origin [**3**]. If ε is small then this family is close to the toric family away from the origin. Therefore, general considerations about deformations of families of special Lagrangians suggest that, if the smooth elliptic curve D' lies in a sufficiently small neighborhood of D_0, then $(\mathbb{CP}^2 \setminus D', \omega_0, \Omega')$ carries a special Lagrangian fibration with three nodal singular fibers. From the point of view of the affine geometry of the base B' of this fibration, the smoothing of each node of D_0 amounts to replacing a corner of the triangle B_0 by a singular point in the interior of B' (so that B' is a singular affine manifold with boundary but without corners). This construction can be thought of as a special Lagrangian version of a trick studied by Margaret Symington [**29**].

The special Lagrangian fibers over points close to the boundary of B' lie in a tubular neighborhood of D', and collapse to closed loops in D' as one approaches the boundary. Thus their first homology group is generated by a meridian m (the boundary of a small disc that intersects D' transversely once) and by a longitude ℓ (a curve that runs parallel to a closed loop on D'). The monodromy of the special Lagrangian fibration along $\partial B'$ fixes m, but because the normal bundle to D' has degree 9 it maps ℓ to $\ell + 9m$. Thus, in a suitable basis the monodromy along the boundary of B' can be expressed by the matrix $\begin{pmatrix} 1 & 9 \\ 0 & 1 \end{pmatrix}$ (see equation (7.2) in [**3**]).

The general case, where the cubic curve D is not necessarily close to the singular toric configuration $D_0 \subset \mathbb{CP}^2$, should follow from a suitable result on deformations of two-dimensional special Lagrangian torus fibrations with nodal singularities. (To our knowledge such a result hasn't been proved yet; however it should follow from an explicit analysis of the deformations of the nodal singularities and the implicit function theorem applied to the smooth part of the fibration. In our case one also needs to control the behavior of the fibration near the boundary of B.)

When constructing the mirror, the singular fibers create walls, which require instanton corrections. In the case of a cubic D' obtained by a small deformation of the toric configuration D_0, the local model for a single smoothing suggests that the walls run parallel to the boundary of the base B'. In fact, the special Lagrangian fibers which lie sufficiently far from D' are Floer-theoretically equivalent to standard product tori. Thus, in the "main" chamber the superpotential is given by the same formula as in the toric case, $W = x + y + e^{-\Lambda}/xy$ in suitable coordinates (where $\Lambda = \int_{\mathbb{CP}^1} \omega$); in the other chambers it is given by some analytic continuation of this expression (see § 5 of [**3**] for an explicit formula in the case of smoothing a single node of D_0). In fact, ignoring completeness issues (e.g., looking only at $|W| \ll 1$), the overall effect of deforming D_0 to a smooth cubic curve on the complex geometry of the Landau-Ginzburg mirror is expected to be a fiberwise compactification. Simultaneously, the symplectic area of the fiber of the Landau-Ginzburg model, which is infinite in the toric case, is expected to become finite and equal to the imaginary part of the modular parameter of the elliptic curve D' (see also [**4**]).

***Example 2.10* (Rational elliptic surface).** — Let X be a rational elliptic surface obtained by blowing up $\mathbb{CP}^2$ at the nine base points of a pencil of cubics, equipped with a Kähler form $\hat{\omega}$. Let $\hat{D} \subset X$ be a smooth elliptic fiber (the proper transform of a cubic of the pencil), and let $\hat{\Omega}$ be a holomorphic $(2,0)$-form on X with poles along $\hat{D}$. We expect:

Conjecture 2.10. — *$X \setminus \hat{D}$ carries a special Lagrangian torus fibration over the disc with (generically) 12 nodal singular fibers. The monodromy of the affine structure around each singularity is conjugate to $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, and the monodromy along $\partial \hat{B}$ is trivial.*

The construction starts with $(\mathbb{CP}^2, D, \omega_0, \Omega)$, where $D \subset \mathbb{CP}^2$ is an elliptic curve and Ω is a holomorphic (2,0)-form with poles along D, as in Example 2.9 above. By Conjecture 2.9, we expect $\mathbb{CP}^2 \setminus D$ to carry a special Lagrangian torus fibration with three nodal singular fibers. Now we blow up $\mathbb{CP}^2$ at nine points on the cubic D, to obtain the rational elliptic surface X. Pulling back Ω under the blowup map yields a holomorphic (2,0)-form $\hat{\Omega}$ on X, with poles along an elliptic curve $\hat{D} \subset X$ (the proper transform of D). On the other hand, the Kähler form $\hat{\omega}$ on X is not canonical, and depends in particular on the choice of the symplectic areas of the exceptional divisors. We claim that, provided these areas are sufficiently small, the blowup should carry a special Lagrangian torus fibration with 12 nodal singular fibers.

The local model for each blowup operation is as follows [**2**]. Consider a neighborhood of the origin in $\mathbb{C}^2$ equipped with the standard symplectic form, the holomorphic volume form $dx \wedge dy/y$ with poles along $\mathbb{C} \times \{0\}$, and the family of special Lagrangian cylinders $\{\mathrm{Re}(x) = t_1,\ \frac{1}{2}|y|^2 = t_2\} \subset \mathbb{C} \times \mathbb{C}^*$. Equip the blowup $\hat{\mathbb{C}}^2$ with a toric Kähler form $\hat{\omega}_0$ (invariant under the standard T^2-action) for which the area of the exceptional divisor is $\epsilon > 0$, and the holomorphic volume form $\hat{\Omega}_0$ obtained by pulling back $dx \wedge dy/y$ under the blowup map $\pi : \hat{\mathbb{C}}^2 \to \mathbb{C}^2$. The lift to $\hat{\mathbb{C}}^2$ of the S^1-action $e^{i\theta} \cdot (x, y) = (x, e^{i\theta} y)$ preserves $\hat{\omega}_0$ and $\hat{\Omega}_0$; its fixed point set consists of on one hand the proper transform $\hat{D}_0$ of $\mathbb{C} \times \{0\}$, and on the other hand the point where the proper transform of $\{0\} \times \mathbb{C}$ hits the exceptional divisor. Denote by $\mu : \hat{\mathbb{C}}^2 \to \mathbb{R}$ the moment map for this S^1-action, normalized to equal 0 on $\hat{D}_0$ and ϵ at the isolated fixed point. Then it is easy to check that the submanifolds $\{\mathrm{Re}(\pi^* x) = t_1,\ \mu = t_2\} \subset \hat{\mathbb{C}}^2 \setminus \hat{D}_0$ are special Lagrangian with respect to $\hat{\omega}_0$ and $\hat{\Omega}_0$ [**2**]. This family of special Lagrangians presents one nodal singular fiber – the fiber which corresponds to $(t_1, t_2) = (0, \epsilon)$ and passes through the isolated S^1-fixed point. Moreover, if ϵ is small then away from a neighborhood of the exceptional divisor this family is close to the initial family of special Lagrangians in $\mathbb{C} \times \mathbb{C}^*$.

Even though the local model is only an asymptotic description of the geometry of the special Lagrangian fibration on $\mathbb{CP}^2 \setminus D$ near a point of D, it should be possible to glue this local construction into the fibration of Conjecture 2.9, and thereby construct a special Lagrangian fibration on the rational elliptic surface X obtained by blowing up $\mathbb{CP}^2$ at 9 points on the elliptic curve D. Each blow-up operation inserts a nodal singular fiber into the fibration; thus the base $\hat{B}$ of the special Lagrangian fibration on X presents 12 singular points. (The whole process can again be viewed as a special Lagrangian version of Symington's construction [**29**].) From the point of view of the symplectic affine structure, an easy calculation on the local model shows that each new singular point lies at a distance from the boundary of $\hat{B}$ equal to the symplectic area of the exceptional curve of the corresponding blowup; in fact the exceptional curve can be seen as a complex ray that runs from the singular point to the boundary of $\hat{B}$. Moreover, the monodromy of the fibration along the boundary of $\hat{B}$ is trivial, reflecting the fact that the anticanonical divisor $\hat{D} \subset X$ has trivial normal bundle.

The general case, where the exceptional divisors of the blowups are not assumed to have small symplectic areas, should again follow from a careful analysis of deformations of two-dimensional special Lagrangian torus fibrations with nodal singularities (with the same caveats as in the case of $\mathbb{CP}^2$).

Remark 2.11. — Assume D is smooth. Then the holomorphic $(n,0)$-form Ω on $X \setminus D$ induces a holomorphic volume form $\Omega_D = \mathrm{Res}_D(\Omega)$ on D: the *residue* of Ω along D. It is reasonable to expect that, as is the case in the various examples considered above, in a neighborhood of D the special Lagrangian fibration on $(X \setminus D, \omega, \Omega)$ consists of tori which are S^1-bundles over special Lagrangian submanifolds of $(D, \omega_{|D}, \Omega_D)$. As a toy example, consider $X = D \times \mathbb{C}$, $\omega = \omega_D + \frac{i}{2} dz \wedge d\bar{z}$, and $\Omega = \Omega_D \wedge dz/z$: then the product any special Lagrangian submanifold of D with a circle centered at the origin in $\mathbb{C}$ is easily seen to be special Lagrangian. We conjecture that the qualitative behavior is the same in the general case; see § 7 of [**3**] for more details.

Assuming that this picture holds, the special Lagrangian fibration $f : X \setminus D \to B$ can be extended over the boundary of B by a special Lagrangian fibration on D. In particular, the boundary of B, with the induced affine structures, is the base B_D of an SYZ fibration on D. More precisely: with respect to the symplectic affine structure, the compactified base $\bar{B}$ is a singular affine manifold with boundary (and corners if D has normal crossings), and its boundary is B_D. With respect to the complex affine structure, $B^\vee$ (after adding in the interior singular fibers) is a complete singular affine manifold, isomorphic to $\mathbb{R}_+ \times B_D^\vee$ outside of a compact subset.

As already seen in Example 2.9, near ∂B the monodromy of the affine structures on B is determined explicitly by the affine structures on B_D and by the first Chern class of the normal bundle to D. Indeed, given a fiber of f near the boundary of B, i.e. an S^1-fibered special Lagrangian $L \subset X \setminus D$, the action of the monodromy on $H_1(L)$ can be determined by working in a basis consisting of a meridian loop linking D and $n-1$ longitudes running parallel to D; from this one deduces the corresponding actions on $H^1(L)$ (monodromy of B) and $H^{n-1}(L)$ (monodromy of $B^\vee$).

Next, we look at the mirror, and observe that near its boundary M consists of pairs (L, ∇) where L is an S^1-fibered special Lagrangian contained in a neighborhood of D. Denote by $\delta \in \pi_2(X, L)$ the homotopy class of a small meridian disc intersecting D transversely once (with boundary the meridian loop), and let $z_\delta(L, \nabla)$ be the corresponding weight as in equation (2.1). Then z_δ is a holomorphic function on M near its boundary. (In fact, z_δ is the dominant term in the expression of the superpotential W near ∂M, as the meridian discs have the smallest symplectic area among all Maslov index 2 holomorphic discs.) By construction, the boundary of M corresponds to the case where the area of the meridian disc reaches zero, i.e. $\partial M = \{|z_\delta| = 1\}$.

Consider the complex hypersurface $M_D = \{z_\delta = 1\}$ ($\subset \partial M$). Geometrically, M_D corresponds to limits of sequences of pairs (L, ∇) where L collapses onto a special Lagrangian torus $\Lambda \subset D$ and the connection ∇ has trivial holonomy along the collapsed S^1-factor in L, i.e. is pulled back from a flat connection on the trivial bundle over Λ. Thus M_D is none other than the SYZ mirror to D. Moreover, the restriction of z_δ to ∂M induces a locally trivial fibration $z_\delta : \partial M \to S^1$ with fiber M_D. The monodromy of this fibration can be realized geometrically as follows. Start with a pair (L, ∇) where L is almost collapsed onto $\Lambda \subset D$ and ∇ has trivial holonomy along the meridian loop (so $z_\delta \in \mathbb{R}_+$): then we can change the holonomy of ∇ along the meridian loop by adding to it a multiple of $\sigma^{-1}\nabla\sigma$, where σ is the defining section of D and ∇ is a suitable connection on K_X^{-1}. From there it follows easily that the monodromy of the fibration $z_\delta : \partial M \to S^1$ is a symplectomorphism of M_D which geometrically realizes (as a fiberwise translation in the special Lagrangian fibration $M_D \to B_D$ dual to the SYZ fibration on D) the mirror to the autoequivalence $- \otimes K_{X|D}^{-1}$ of $D^b\mathrm{Coh}(D)$.

This rich geometric picture naturally leads to a formulation of mirror symmetry for the pairs (X, D) and (M, M_D); see § 7 of [**3**] for details.

3. Special Lagrangian fibrations and double covers

3.1. Special Lagrangians and Calabi-Yau double covers. — Let (X, ω, J) be a smooth compact Kähler manifold of complex dimension n, and let s be a nontrivial holomorphic section of K_X^{-2}. Unless otherwise specified we assume that the hypersurface $H = s^{-1}(0)$ is smooth. $\Theta = s^{-1}$ is a nonvanishing section of $K_X^{\otimes 2}$ over $X \setminus H$, with poles along H, and locally $\Omega = \Theta^{1/2}$ is a nonvanishing holomorphic n-form, defined up to sign. The restriction of Θ to a Lagrangian submanifold $L \subset X \setminus H$ does not vanish, and can be expressed in the form $\eta \operatorname{vol}_g^2$, where $\eta \in C^\infty(L, \mathbb{C}^*)$. By analogy with the situation considered previously, we make the following definition:

Definition 3.1. — *A Lagrangian submanifold $L \subset X \setminus H$ is* special Lagrangian *if the argument of η is constant. (In fact Θ will usually be normalized so that η is real).*

It is easy to see that, if $L \subset X \setminus H$ is an orientable special Lagrangian submanifold, then over L the holomorphic quadratic differential Θ admits a globally defined square root Ω. Therefore Proposition 2.2 still applies in this setting; since $\Omega_{|L} = \eta^{1/2}\operatorname{vol}_g$, special Lagrangian deformations are now given by $\eta^{1/2}$-harmonic 1-forms on L.

As before, the base B of a special Lagrangian torus fibration carries two natural affine structures, one arising from the symplectic geometry of X and the other one arising from its complex geometry.

We now turn to the Calabi-Yau double cover of X branched along H, namely the unique double cover $\pi : Y \to X$ with the property that $\tilde{\Theta} = \pi^*\Theta$ admits a globally defined square root $\tilde{\Omega} \in \Omega^{n,0}(Y)$. More explicitly, the obstruction for Θ to admit a globally defined square root is given by an element of $H^1(X \setminus H, \mathbb{Z}/2) \simeq \operatorname{Hom}(\pi_1(X \setminus H), \mathbb{Z}/2)$, and we consider the branched cover with this monodromy.

The complex geometry of Y is fairly straightforward, as the complex structure $\tilde{J}$ and the holomorphic volume form $\tilde{\Omega}$ are simply lifted from those of X via π. In particular, it is easy to check that $\tilde{\Omega}$ is well-behaved along the ramification divisor. (To give the simplest example, consider the map $z \mapsto z^2$ from $\mathbb{C}$ to itself: the pullback of $\Theta = z^{-1} dz^{\otimes 2}$ is $\tilde{\Theta} = 4\, dz^{\otimes 2}$, which has a well-defined square root $\tilde{\Omega} = 2\, dz$.)

On the other hand, constructing a Kähler form on Y requires some choices, because the pullback form $\pi^*\omega$ is degenerate along the ramification locus $\tilde{H} = \pi^{-1}(H)$. One approach is to view Y as a complex hypersurface in the total space of the line bundle K_X^{-1} over X, equipped with a suitable Kähler metric. More directly, one can equip Y with a Kähler form $\tilde{\omega} = \pi^*\omega + \epsilon\,\lambda$, where $\epsilon > 0$ is a sufficiently small constant and λ is an exact real $(1,1)$-form whose restriction to the complex line $\operatorname{Ker}(d\pi)$ is positive at every point of the ramification locus. Any two forms obtained in this manner are symplectically isotopic; for example one can take $\lambda = -i\partial\bar{\partial}\phi$ where $\phi : Y \to [0,1]$ is supported in a neighborhood of $\tilde{H}$, equal to 1 on $\tilde{H}$, and strictly concave in the normal directions at every point of $\tilde{H}$.

Thus, given a compact special Lagrangian submanifold $L \subset X \setminus H$, the two lifts of L are in general not special Lagrangian submanifolds of Y, even though the restriction of $\tilde{\Omega}$ has constant phase, because they are not necessarily Lagrangian for $\tilde{\omega}$. In very specific cases (for instance in dimension 1 or in product situations) this is not an issue, but in general one needs to deform the lift of L to a nearby special Lagrangian submanifold $\tilde{L} \subset Y$, whose existence is guaranteed by the unobstructedness of deformations (Proposition 2.2) as long as $\tilde{\omega}$ is sufficiently close to $\pi^*\omega$.

When considering not just one submanifold but a whole special Lagrangian fibration on $X \setminus H$, it is natural to ask whether the lifts can be similarly deformed to a special Lagrangian fibration on Y. Away from H and from the singular fibers, we can rely on an implicit function theorem for special Lagrangian fibrations which again follows from unobstructedness. In spite of the wealth of results that have been obtained on singularities of special Lagrangians and their deformations (see e.g. [**16**]), to our knowledge there is no general result that would yield a special Lagrangian fibration on Y from one on $X \setminus H$. Nonetheless, it seems reasonable to expect that such a result might hold at least in low dimensions if the Kähler form on Y is chosen suitably and the family of special Lagrangians only presents generic singularities.

Thus, Conjecture 1.1 can be stated more precisely as follows:

Conjecture 3.2

1. *X carries a special Lagrangian fibration (or rather, foliation) $f : X \to \bar{B}$, where $\bar{B}$ is a singular affine manifold with boundary (with two affine structures), such that the generic fibers of f are special Lagrangian tori in $X \setminus H$, and the fibers of f above $\partial\bar{B}$ are special Lagrangians with boundary in H.*

2. *Y carries a special Lagrangian torus fibration $\tilde{f} : Y \to \tilde{B}$, where $\tilde{B}$ is a singular affine manifold without boundary (with two affine structures), obtained by gluing together two copies of $\bar{B}$ along their boundary.*

Note that the boundaries of the two copies of $\bar{B}$ are identified using the identity map, whereas the normal direction is reflected; thus this is an orientation-reversing gluing, and the resulting singular affine manifold $\tilde{B}$ admits an orientation-reversing involution whose fixed point locus is the "seam" of the gluing.

3.2. Example: $\mathbb{CP}^1$ and elliptic curves. — As our first example, we consider $X = \mathbb{CP}^1$ equipped with any Kähler form and a holomorphic quadratic differential Θ with poles at a subset $H \subset \mathbb{CP}^1$.

We first consider the special case $\Theta = dz^2/(z^2 - a^2)$, with simple poles at $\pm a$ and a double pole at infinity. Setting $a = 0$, we recover the classical situation discussed in Example 2.7, in which the circles centered at the origin are special Lagrangian. For arbitrary a, it follows from classical geometry that every ellipse with foci $\pm a$

is special Lagrangian with phase $\pi/2$ for $\Omega = \Theta^{1/2} = dz/\sqrt{z^2 - a^2}$. Thus we get a special Lagrangian foliation of $\mathbb{C} \setminus \{\pm a\}$ by this family of ellipses, the sole noncompact leaf being the real interval $(-a, a)$. The general case is less explicit but essentially amounts to modifying the special Lagrangian family in the same manner not only near zero but also near infinity.

More precisely, equip $\mathbb{CP}^1$ with a generic holomorphic quadratic differential $\Theta = z^2\,dz^2/(z-a)(z-b)(z-c)(z-d)$ with poles at $H = \{a, b, c, d\}$. Then, for a suitable choice of phase, $\mathbb{CP}^1 \setminus H$ admits a special Lagrangian foliation in which all the leaves are closed loops with the exception of two noncompact leaves, each connecting two of the points of H (say a and b on one hand, and c and d on the other hand). For instance, if $a < b < c < d$ are real, then we have such a foliation (with phase $\pi/2$) in which the two noncompact leaves are the real line segments (a, b) and (c, d). Indeed, after removing the two intervals $[a, b]$ and $[c, d]$, the quadratic differential Θ admits a well-defined square root Ω, which is a closed 1-form and hence has the same period (easily checked to be pure imaginary) on any homotopically nontrivial embedded curve. The general case follows from the same argument.

From a symplectic point of view, the base B of this foliation is again an interval of length equal to the symplectic area of $\mathbb{CP}^1$. However, unlike the situation of Example 2.7, the affine structure induced on B by the holomorphic volume form identifies it with a *finite* interval: if we normalize Ω so that the integral of $\operatorname{Im}\Omega$ over each special Lagrangian fiber is 1, then the length of this interval is equal to $\int_b^c \operatorname{Re}\Omega$.

The double cover of $\mathbb{CP}^1$ branched at H is an elliptic curve Y, and the family of special Lagrangians in $\mathbb{CP}^1 \setminus H$ lifts to a smooth special Lagrangian fibration on Y. The base $\tilde{B} \simeq S^1$ of this fibration, and its two affine structures, are obtained by doubling B along its boundary. For instance, the symplectic area of Y (which is the length of $\tilde{B}$ with respect to the symplectic affine structure, cf. Example 2.4) is twice that of $\mathbb{CP}^1$, whereas the integral of $\operatorname{Re}\tilde{\Omega}$ over a section of the special Lagrangian fibration (which is the length of $\tilde{B}$ with respect to the complex affine structure) is twice $\int_b^c \operatorname{Re}\Omega$.

Remark 3.3. — With respect to the complex affine structure, the base B of the special Lagrangian foliation on $(\mathbb{CP}^1 \setminus H, \Omega)$ is a finite interval, whereas the base B_0 of the special Lagrangian fibration on $(\mathbb{CP}^1 \setminus \{0, \infty\}, \Omega_0 = dz/z)$ has infinite size. The reason is that, as $a, b \to 0$ and $c, d \to \infty$, the elliptic curve Y degenerates to a curve with two nodal singularities, and the base $\tilde{B}$ of its special Lagrangian fibration degenerates to a union of two infinite intervals. On the other hand, the symplectic structure on Y, which determines the length of the base with respect to the other affine structure, is unaffected by the degeneration.

3.3. Example: Elliptic surfaces. — We revisit Example 2.10, and again denote by X a rational elliptic surface obtained by blowing up $\mathbb{CP}^2$ at the 9 base points of a pencil of cubics, equipped with a Kähler form $\hat{\omega}$. We previously considered a holomorphic volume form $\hat{\Omega}$ on X with poles along an elliptic fiber $\hat{D}$. Now we equip X with a section Θ of $K_X^{\otimes 2}$, with poles along the union $H = D_+ \cup D_-$ of two smooth fibers of the elliptic fibration; for simplicity we assume that $D_\pm$ lie close to a same smooth fiber $\hat{D}$, so that away from a neighborhood of $\hat{D}$ the quadratic volume element Θ is close to the square $\hat{\Omega}^{\otimes 2}$ of the volume form considered in Example 2.10.

Conjecture 3.4. — *The special Lagrangian fibration on $X\setminus\hat{D}$ constructed in Conjecture 2.10 deforms to a special Lagrangian family on $X \setminus H$. The base B of this family is homeomorphic to a closed disc, and over its interior the fibers are special Lagrangian tori, with the exception of 12 nodal singular fibers. The fibers above ∂B are special Lagrangian annuli with one boundary component on D_+ and the other on D_-.*

We now explain the geometric intuition behind this conjecture by considering a simplified local model in which everything is explicit. The actual geometry of X near $\hat{D}$ differs from this local model by higher order terms; however the local model is expected to accurately describe all the qualitative features of the special Lagrangian families of Conjectures 2.10 and 3.4 in a small neighborhood of $\hat{D}$.

In a small neighborhood of the fiber $\hat{D}$, the elliptic fibration $X \to \mathbb{CP}^1$ is topologically trivial, and even though it is not holomorphically trivial, in first approximation we can consider a local model of the form $E\times U$, where E is an elliptic curve ($E \simeq \hat{D}$) and U is a neighborhood of the origin in $\mathbb{C}$ (with coordinate z). In this simplified local model, the holomorphic volume form $\hat{\Omega}$ can be written in the form $dw \wedge dz/z$, where dw is a holomorphic 1-form on E (in fact, the *residue* of $\hat{\Omega}$ along $\hat{D}$), the symplectic form $\hat{\omega}$ is a product form, and the special Lagrangian family of Conjecture 2.10 consists of product tori, where the first factor is a special Lagrangian circle in (E, dw) and the second factor is a circle centered at the origin.

We now equip $E\times U$ with the quadratic volume element $\Theta = (dw\wedge dz)^{\otimes 2}/(z^2-\epsilon^2)$, with poles along $H = E \times \{\pm\epsilon\}$. Then the previous family of special Lagrangians deforms to one where each submanifold is again a product: the first factor is still a special Lagrangian circle in (E, dw), and the second factor is now an ellipse with foci at $\pm\epsilon$ (in the degenerate case, the line segment $[-\epsilon, \epsilon]$).

The bases of these two special Lagrangian fibrations on $E\times U$, equipped with their symplectic affine structures, are naturally isomorphic, as each ellipse with foci at $\pm\epsilon$ can be used interchangeably with the circle that encloses the same symplectic area (in fact, the corresponding product Lagrangian tori in $E \times U$ are Hamiltonian isotopic to each other). In this sense, passing from $X \setminus \hat{D}$ to $X \setminus H$ (i.e., from $\hat{B}$ to B) is expected to be a trivial operation from the symplectic point of view. However, the complex affine structures on $\hat{B}$ and B are very different: from that perspective $\hat{B}$ is "complete" (its boundary lies "at infinity", since the affine structure blows up near $\partial\hat{B}$ due to the singular behavior of $\hat{\Omega}$ along $\hat{D}$), whereas B has finite diameter. This is most easily seen in terms of the local model near $\hat{D}$, which allows us to reduce to the one-dimensional case (see Remark 3.3).

Finally, we consider the double cover Y of the rational elliptic surface X branched along H. It is easy to see that Y is an elliptically fibered K3 surface, carrying a holomorphic involution under which the holomorphic volume form $\tilde{\Omega} = (\pi^*\Theta)^{1/2}$ is anti-invariant. By Conjecture 1.1 we expect that Y, equipped with a suitably chosen Kähler form in the class $[\pi^*\hat{\omega}]$, carries a special Lagrangian fibration with 24 nodal singular fibers, whose base $\tilde{B} \simeq S^2$ is obtained by doubling B along its boundary.

In fact, it is well-known that such a fibration can be readily obtained using hyperkähler geometry as in Example 2.5. Indeed, consider an elliptically fibered K3 surface with a real structure for which the real part consists of two tori. For example, let Y' be the double cover of $\mathbb{CP}^1 \times \mathbb{CP}^1$ branched along the zero set of a generic real homogeneous polynomial of bidegree $(4,4)$ without any real roots. Composing the covering map with projection to the first $\mathbb{CP}^1$ factor, we obtain an elliptic fibration $f: Y' \to \mathbb{CP}^1$ with 24 singular fibers. Complex conjugation lifts to an involution ι on Y' which is antiholomorphic with respect to the given complex structure J, and whose fixed point locus is the trivial (disconnected) double cover of $\mathbb{RP}^1 \times \mathbb{RP}^1$ (i.e., two tori). The involution ι maps each fiber of f to the fiber above the complex conjugate point of $\mathbb{CP}^1$, and in particular it interchanges pairs of complex conjugate singular fibers.

Equip Y' with a Calabi-Yau metric, such that the Kähler form ω_J is anti-invariant under ι (this is guaranteed by uniqueness of the Calabi-Yau metric if one imposes $[\omega_J]$ to be the pullback of a Kähler class on $\mathbb{CP}^1 \times \mathbb{CP}^1$ and hence anti-invariant). Denote by Ω_J a holomorphic $(2,0)$-form on Y': then $\iota^*\Omega_J$ is a scalar multiple of $\bar{\Omega}_J$, because $\dim H^{0,2}_J(Y) = 1$. So after normalization we can assume that $\iota^*\Omega_J = -\bar{\Omega}_J$, i.e. $\omega_K := \mathrm{Re}(\Omega_J)$ is anti-invariant and $\omega_I := \mathrm{Im}(\Omega_J)$ is invariant.

Now switch to the complex structure I determined by the Kähler form ω_I. Then ι becomes a holomorphic involution, and the holomorphic volume form $\Omega_I = \omega_J + i\omega_K$ is anti-invariant. Since the fibers of $f: Y' \to \mathbb{CP}^1$ are calibrated by ω_J, the map f is a special Lagrangian fibration on (Y', ω_I, Ω_I), compatible with the involution ι.

It seems likely that this construction can be used as an alternative approach to Conjecture 3.4, by considering the quotient of this special Lagrangian fibration by the involution ι.

Remark 3.5. — The elliptic surface X contains nine exceptional spheres, arising from the nine blow-ups performed on $\mathbb{CP}^2$; these spheres intersect H in two points, so their preimages in the double cover Y are rational curves with normal bundle $\mathcal{O}(-2)$. These curves can be seen by looking at the complex affine structures on the bases B and $\tilde{B}$ of the special Lagrangian fibrations on X and Y, as discussed in Remark 2.6. Namely, the exceptional curves in X correspond to complex rays that run from singularities of the affine structure of B to its boundary (as in Example 2.10). Doubling B along its boundary to form $\tilde{B}$ creates alignments between pairs of singular points lying symmetrically across from each other. For at least 9 of the 12 pairs of points (those which correspond to the blowups) the corresponding complex rays match up to yield -2-curves in Y.

3.4. Example: $\mathbb{CP}^2$ and K3. — We now revisit Example 2.9, and now equip $\mathbb{CP}^2$ with a section Θ of $K^{\otimes 2}$ with poles along a smooth curve H of degree 6. We assume that H lies in a small neighborhood of a cubic D, i.e. it is defined by a homogeneous polynomial of the form $p = \sigma^2 - \epsilon q$, where $\sigma \in H^0(\mathcal{O}(3))$ is the defining section of D and ϵ is a small constant. Thus, away from a neighborhood of D the quadratic volume element Θ is close to the square $\Omega^{\otimes 2}$ of the volume form considered in Example 2.9.

Conjecture 3.6. — *The special Lagrangian fibration on $\mathbb{CP}^2 \setminus D$ constructed in Conjecture 2.9 deforms to a special Lagrangian family on $\mathbb{CP}^2 \setminus H$. The base B of this family is homeomorphic to a closed disc, and over its interior the fibers are special Lagrangian tori, with the exception of three nodal singular fibers. The fibers above ∂B are special Lagrangian annuli with boundary on H, with the exception of 18 pinched annuli (with one arc connecting the two boundaries collapsed to a point).*

While we do not have a complete picture to propose, the rough idea is as follows. Looking at the defining section $p = \sigma^2 - \epsilon q$ of H, away from the zeroes of q we can think of H as two parallel copies of D, and special Lagrangians are expected to behave as in the previous example. Namely, near D a special Lagrangian in $\mathbb{CP}^2 \setminus D$ looks like the product of a special Lagrangian $\Lambda\,(\simeq S^1)$ in D with a small circle in the normal direction, and the corresponding special Lagrangian in $\mathbb{CP}^2 \setminus H$ should be obtained by replacing the circle factor by a family of ellipses whose foci lie on H. In the degenerate limit case, the ellipses become line segments joining the two foci, forming an annulus; when Λ passes through a zero of q, the corresponding line segment is collapsed to a point, giving a pinched annulus.

In fact, we are unable to provide an explicit local model for this behavior on $X \setminus H$. However, Conjecture 3.6 can be corroborated by calculations on a local model for the double cover Y of X branched along H.

Near a point of D, we can consider local coordinates (u, v) on a domain in $\mathbb{C}^2$ such that D is defined by the equation $u = 0$, and H is defined by the equation $u^2 - \epsilon q(v) = 0$ for some holomorphic function q. The corresponding section of $K_X^{\otimes 2}$ is given by $\Theta = (u^2 - \epsilon q(v))^{-1}\,(du \wedge dv)^{\otimes 2}$. As $\epsilon \to 0$, this converges to the square of the holomorphic volume form $u^{-1}\,du \wedge dv$, for which the cylinders $\{\mathrm{Re}\, v = a,\ |u|^2 = r\}$ are special Lagrangians (the circle factor corresponds to the direction normal to D, while the other factor corresponds to a local model for a special Lagrangian in D).

In this local model the double cover of $\mathbb{C}^2$ branched along H is the hypersurface $Y \subset \mathbb{C}^3$ defined by the equation $z^2 = u^2 - \epsilon q(v)$. The pullback of Θ under the projection map $(z,u,v) \mapsto (u,v)$ admits the square root

$$\tilde{\Omega} = z^{-1}\,du \wedge dv = u^{-1}\,dz \wedge dv.$$

It is worth noting that $\tilde{\Omega}$ is the natural holomorphic volume form induced on Y by the standard volume form of $\mathbb{C}^3$: denoting by $f = z^2 - u^2 + \epsilon q(v)$ the defining function of Y, we have $df \wedge \tilde{\Omega} = dz \wedge du \wedge dv$. We equip Y with the restriction of the standard Kähler form $\omega_0 = \frac{i}{2}\,dz \wedge d\bar{z} + \frac{i}{2}\,du \wedge d\bar{u} + \frac{i}{2}\,dv \wedge d\bar{v}$, which differs from the pullback of the standard Kähler form of $\mathbb{C}^2$ by the extra term $\frac{i}{2}\,dz \wedge d\bar{z} = \frac{i}{2}\partial\bar{\partial}|u^2 - \epsilon q(v)|^2$. We claim that the (possibly singular) submanifolds

$$\tilde{L}_{a,b} = \{(z,u,v) \in Y \mid \mathrm{Re}\,(v) = a,\ \mathrm{Re}\,(u\bar{z}) = b\} \qquad (a,b) \in \mathbb{R}^2$$

are special Lagrangian with respect to $\tilde{\Omega}$ and ω_0. Indeed, the vector field $\xi(z,u,v) = (iu, iz, 0)$ is tangent to the submanifolds $\tilde{L}_{a,b}$, and the 1-forms $\iota_\xi \mathrm{Im}\,\tilde{\Omega} = \mathrm{Re}\,dv$ and $\iota_\xi \omega_0 = -d\,\mathrm{Re}(u\bar{z}) + \frac{i}{2}dv \wedge d\bar{v}$ both vanish on $\tilde{L}_{a,b}$. Moreover, $\tilde{L}_{a,b}$ is singular if and only if it passes through a point $(0,0,v_0)$ with v_0 a root of q.

The involution $(z,u,v) \mapsto (-z,u,v)$ maps $\tilde{L}_{a,b}$ to $\tilde{L}_{a,-b}$. Thus, the special Lagrangian fibration $(z,u,v) \mapsto (\mathrm{Re}\, v, \mathrm{Re}\,(u\bar{z}))$ descends to a family of submanifolds in $\mathbb{C}^2$, parameterized by the quotient of $\mathbb{R}^2$ by the reflection $(a,b) \mapsto (a,-b)$, i.e. the closed upper half-plane. The image of $\tilde{L}_{a,b}$ under this projection is

$$L_{a,b} = \{(u,v) \in \mathbb{C}^2 \,|\, \mathrm{Re}\,(v) = a,\ \mathrm{Re}\,(\bar{u}\sqrt{u^2 - \epsilon q(v)}) = \pm b\},$$

and behaves exactly as described above: fixing a value of v (i.e., a point of D), the intersection of $L_{a,b}$ with $\mathbb{C} \times \{v\}$ is an ellipse with foci the two square roots of $\epsilon q(v)$ (i.e. the two points where H intersects $\mathbb{C} \times \{v\}$). For $b = 0$ the ellipse degenerates to a line segment; when v is a root of q the ellipses become circles and the line segment collapses to a point. However, a quick calculation shows that $L_{a,b}$ is not Lagrangian with respect to the standard Kähler form on $\mathbb{C}^2$.

Thus, it may well be easier to construct a special Lagrangian fibration on the double cover of $\mathbb{CP}^2$ branched at H (namely, a K3 surface) than on $\mathbb{CP}^2 \setminus H$. In fact, as in the previous example, the easiest way to construct such a fibration is probably through hyperkähler geometry, starting from an elliptically fibered K3 surface with a real structure for which the real part is a smooth connected surface of genus 10. Let P be a real homogeneous polynomial of bidegree $(4,4)$ whose zero set in $\mathbb{RP}^1 \times \mathbb{RP}^1$ consists of nine homotopically trivial circles $C_1, \dots, C_9$ bounding mutually disjoint discs D_i, and let Y' be the double cover of $\mathbb{CP}^1 \times \mathbb{CP}^1$ branched along the zero set of P (over $\mathbb{C}$). Then complex conjugation lifts to a J-antiholomorphic involution ι of Y', whose fixed point locus is a connected surface of genus 10, namely the preimage of $\mathbb{RP}^1 \times \mathbb{RP}^1 \setminus (D_1 \cup \cdots \cup D_9)$ (whereas the fixed point set of the composition of ι with the nontrivial deck transformation consists of 9 spheres, the preimages of $D_1, \dots, D_9$). After performing a hyperkähler rotation as in § 3.3, we obtain a new complex structure I on Y' with respect to which ι is holomorphic and the elliptic fibration induced by projection to a $\mathbb{CP}^1$ factor is special Lagrangian.

Remark 3.7. — The curve $H \subset \mathbb{CP}^2$ bounds a number of Lagrangian discs, arising as relative vanishing cycles for degenerations of H to a nodal curve. For instance, considering a degeneration of H to two intersecting cubics singles out 9 such discs. The preimages of these discs are Lagrangian spheres in the double cover Y, and can be seen by looking at the symplectic affine structure on the bases B and $\tilde{B}$ of the special Lagrangian fibrations on $\mathbb{CP}^2$ and Y. Namely, $\tilde{B}$ is obtained by doubling B along its boundary, and 18 of its singular points are aligned along the "seam" of this gluing. The rays emanating from these singular points run along the seam, and match with each other to give rise to Lagrangian spheres.

Remark 3.8. — Consider a singular K3 surface Y_0 with 9 ordinary double point singularities, obtained as the double cover of $\mathbb{CP}^2$ branched along the union H_0 of two intersecting cubics. The singularities of Y_0 can be either smoothed, which amounts to smoothing H_0 to a smooth sextic curve, or blown up, which is equivalent to blowing up $\mathbb{CP}^2$ at the 9 intersection points between the two components of H_0. These two procedures yield respectively the K3 surface considered in the above discussion, and the K3 surface considered in § 3.3. Y_0 admits a special Lagrangian fibration whose base $\tilde{B}_0$ presents 9 singularities with monodromy conjugate to $\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$; viewing $\tilde{B}_0$ as two copies of a disc glued along the boundary, these 9 singularities all lie along the seam of the gluing. Smoothing Y_0 replaces each ordinary double point by a Lagrangian sphere, and resolves the corresponding singularity of $\tilde{B}_0$ into a pair of singular points aligned along the seam. Blowing up Y_0 replaces each ordinary double point by an exceptional curve, and resolves the corresponding singularity of $\tilde{B}_0$ into a pair of singular points lying symmetrically across from each other on either side of the seam.

3.5. Towards mirror symmetry for double covers. — Conjecture 3.2 suggests that a mirror $Y^\vee$ of the Calabi-Yau double cover Y of X branched along H can be obtained by gluing two copies of the mirror of $X \setminus H$ along their boundary. From the point of view of affine geometry, we start with a special Lagrangian fibration $f^\vee : M \to B$ (T-dual to the special Lagrangian fibration on $X \setminus H$), and glue together two copies of M using an orientation-reversing diffeomorphism of ∂M which induces a reflection in each fiber of $f^\vee$ above ∂B.

Arguably the "usual" mirror of X arises from considering the complement of an anticanonical divisor D, rather than the hypersurface H. Consider a degeneration of H under which it collapses onto D (with multiplicity 2). At the level of double covers, this amounts to degenerating Y to the union of two copies of X glued together along D. By Moser's theorem, this deformation affects the complex geometry of Y but not its symplectic geometry. Hence, the special Lagrangian fibrations on $X \setminus H$ and $X \setminus D$ can reasonably be expected to have the same base B, as long as we only consider the symplectic affine structure. (The complex affine structures are very different: in the case of $X \setminus D$ the complex affine structure blows up near the boundary of B, while in the case of $X \setminus H$ it doesn't. See e.g. Remark 3.3.) So, as long as we only consider the complex geometry of the mirror and not its symplectic structure, it should be possible to construct the mirror of Y simply by gluing two copies of the mirror of $X \setminus D$ (which is also the mirror of X without its superpotential).

Remark 3.9. — What we are considering is a special case of a more general construction, in which one degenerates a Calabi-Yau manifold to a reducible configuration of manifolds of negative Kodaira dimension. For instance, as pointed out by the referee, one could extend the example of § 3.3 to that of a K3 surface degenerating to a union of two (different) rational elliptic surfaces glued together along a smooth elliptic fiber.

As pointed out by the referee, given a degeneration of Y to two copies of X glued together along D (or another reducible configuration), one can try to build a Calabi-Yau metric on Y by truncating and gluing together complete Ricci-flat metrics on the open pieces $X \setminus D$ (in the example of § 3.3 those exist by the work of Tian and Yau). This differs somewhat from our perspective, where the Kähler metric on Y is not required to be Ricci-flat (i.e., Y is only "almost Calabi-Yau"), and hence it can be obtained more directly from a non-singular Kähler metric on X.

A complication arises when the normal bundle to D is not holomorphically trivial. In that case, the family of special Lagrangians in $X \setminus H$ presents additional singularities at the boundary of B; these singularities are not directly visible in the special Lagrangian fibration on $X \setminus D$. An example of this phenomenon is presented in § 3.4 (compare Conjecture 3.6 with Conjecture 2.9). Thus, when building $\tilde{B}$ out of two copies of the base B of the special Lagrangian fibration on $X \setminus D$, we need to introduce extra singularities into the affine structure along the seam of the gluing. This is essentially the same phenomenon as in Gross and Siebert's program (where singularities of the affine structure also arise from the nontriviality of the normal bundles to the codimension 1 toric strata along which the smoothing takes place).

For simplicity, let us just consider the case where D has trivial normal bundle. In that case, the discussion in Remark 2.11 implies that the boundary of the (uncorrected) mirror M of $X \setminus D$ is the product of S^1 with a complex hypersurface

$M_D \subset \partial M$ (the uncorrected SYZ mirror to D). In fact, we have a trivial fibration $z_\delta : \partial M \approx M_D \times S^1 \to S^1$, where z_δ is the weight associated to the homotopy class of a meridian disc (collapsing to a point as the special Lagrangian torus L collapses onto a special Lagrangian submanifold of D, whence $|z_\delta| = 1$ on ∂M). The orientation-reversing diffeomorphism $\varphi : \partial M \to \partial M$ used to glue the two copies of M together corresponds to a reversal of the coordinate dual to the class of the meridian loop. More precisely, view a point of ∂M as a pair (Λ, ∇) where Λ is a special Lagrangian torus in D and ∇ is a flat unitary connection on the trivial bundle over $\Lambda \times S^1$ (here we use the triviality of the normal bundle to D to view nearby special Lagrangians in $X \setminus D$ as products $\Lambda \times S^1$ rather than S^1-bundles over Λ). Then the gluing diffeomorphism φ is given by $(\Lambda, \nabla) \mapsto (\Lambda, \bar{\nabla})$, where $\bar{\nabla}$ is the pullback of ∇ by the diffeomorphism $(p, e^{i\theta}) \mapsto (p, e^{-i\theta})$ of $\Lambda \times S^1$. Thus, under the identification of ∂M with $M_D \times S^1$, the diffeomorphism φ is the product of the identity map in M_D and the complex conjugation map $z_\delta \mapsto \bar{z}_\delta = z_\delta^{-1}$ from S^1 to itself.

At this point it would be tempting to conclude that, if $K_{X|D}$ is holomorphically trivial, then a mirror of Y can be obtained (at least as a complex manifold) by gluing together two copies of the mirror of X along their boundary $S^1 \times M_D$, to obtain a Calabi-Yau variety with a holomorphic involution given near the "seam" of the gluing by $z_\delta \mapsto z_\delta^{-1}$. Unfortunately, in the presence of instanton corrections this seems to always fail; in particular, the fibers of $z_\delta : \partial X^\vee \to S^1$ above two complex conjugate points are not necessarily biholomorphic. The following example in complex dimension 2 (inspired by calculations in [**2**]) illustrates a fairly general phenomenon.

Example 3.10. — We consider again the local model for blow-ups mentioned in Example 2.10, modified so the special Lagrangian fibers are tori rather than cylinders [**2**]. Start with $\mathbb{C}^* \times \mathbb{C}$ equipped the holomorphic volume form $d\log x \wedge d\log y$ with poles along $\mathbb{C}^* \times \{0\}$, and blow up the point $(1, 0)$ to obtain a complex manifold X equipped the holomorphic volume form $\Omega = \pi^*(d\log x \wedge d\log y)$, with poles along the proper transform D of $\mathbb{C}^* \times \{0\}$. Observe that the S^1-action $e^{i\theta} \cdot (x, y) = (x, e^{i\theta} y)$ lifts to X, and consider an S^1-invariant Kähler form ω for which the area of the exceptional divisor is ϵ. Denote by $\mu : X \to \mathbb{R}$ the moment map for the S^1-action, normalized to equal 0 on D and ϵ at the isolated fixed point. The S^1-invariant tori $L_{t_1,t_2} = \{\log |\pi^* x| = t_1,\ \mu = t_2\}$ define a special Lagrangian fibration on $X \setminus D$, with one nodal singularity at the isolated fixed point (for $(t_1, t_2) = (0, \epsilon)$) [**2**].

The base B of this special Lagrangian fibration is a half-plane, with a singular point at distance ϵ from the boundary (and nontrivial monodromy around the singularity), as pictured in Figure 1; we place the cut above the singular point in order to better visualize wall-crossing phenomena near the boundary of B. The complex rays emanating from the singular point (one of which corresponds precisely to the exceptional divisor of the blowup) are responsible for wall-crossing jumps in holomorphic disc counts, and split the mirror M into two chambers, which are essentially the preimages of the left and right halves of the figure.

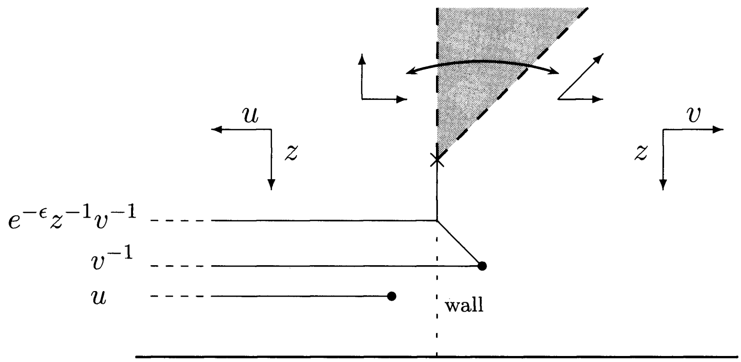

FIGURE 1. A special Lagrangian fibration on the blowup of $\mathbb{C}^* \times \mathbb{C}$

Denote by z $(= z_\delta)$ the holomorphic coordinate on M which corresponds to the holomorphic disc $\{\pi^*x = e^{t_1},\ \mu < t_2\}$ in (X, L_{t_1,t_2}); it can be thought as a complexified and exponentiated version of the downward-pointing affine coordinate pictured on Figure 1. In one of the two chambers of M, denote by u the holomorphic coordinate that similarly corresponds to the leftward-pointing affine coordinate represented in the figure. For instance, if we partially compactify X to allow π^*x to become zero (i.e., if we had blown up $\mathbb{C}^2$ at $(1,0)$ rather than $\mathbb{C}^* \times \mathbb{C}$), then u becomes (up to a scaling factor) the weight associated to a disc that runs parallel to the x-axis. Similarly, denote by v the holomorphic coordinate in the other chamber of M corresponding to a rightward-pointing affine coordinate, normalized so that, if we ignore instanton corrections, the gluing across the wall is given by $u = v^{-1}$.

Imagine that L_{t_1,t_2} in the "left" chamber $(t_1 < 0)$ bounds a holomorphic disc with associated weight u (such a disc doesn't exist in X, but it exists in a suitable partial compactification), and increase the value of t_1 past zero, keeping t_2 less than ϵ: then this holomorphic disc deforms appropriately (and its weight is now called v^{-1}), but it also generates a new disc with weight $e^{-\epsilon}z^{-1}v^{-1}$, obtained by attaching an exceptional disc (the part of the exceptional divisor where $\mu > t_2$) as one crosses the wall. This phenomenon is pictured on Figure 1 (where the various discs are abusively represented as tropical curves, which actually should be drawn in the complex affine structure). Thus the instanton-corrected gluing is given by $u = v^{-1} + e^{-\epsilon}z^{-1}v^{-1}$, i.e.,

$$uv = 1 + e^{-\epsilon}z^{-1}. \tag{3.1}$$

Actually the portion of the wall where $t_2 > \epsilon$ also gives rise to the same instanton-corrected gluing, so that the corrected mirror is globally given by (3.1); see [**2**].

Now replace D by the union $H = D_+ \cup D_-$ of two disjoint complex curves, e.g. the proper transforms of two complex lines intersecting transversely at the blown up point $(1, 0)$, and consider the double cover Y of X branched along H. (We leave the details unspecified, as the construction should arguably be carried out in a global setting such as that of Conjecture 3.4 rather than in the local setting.)

Conjecture 1.1 suggests that Y should carry a special Lagrangian fibration whose base (considering only the symplectic affine structure) is obtained by doubling B along its boundary. Pictorially, this corresponds to flipping Figure 1 about the horizontal axis and gluing the two pictures together. On the mirror, before instanton corrections this amounts to reflecting the z variable via $z \mapsto z^{-1}$, and gluing M and its reflected copy along their common boundary $|z| = 1$. However, the gluing via $z \mapsto z^{-1}$ is not compatible with the instanton corrections discussed above; this is because when we cross the wall there are now two different exceptional discs to consider. Namely, Y contains a -2-curve C (the preimage of the exceptional curve in X), corresponding to the alignment between the walls that come out of the two singular fibers on either side of the seam. Special Lagrangian fibers which lie on the wall intersect C in a circle and split it into two Maslov index 0 discs, which both contribute to instanton corrections. A careful calculation shows that the instanton-corrected gluing is now

$$uv = (1 + e^{-\epsilon} z^{-1})(1 + e^{-\epsilon} z). \tag{3.2}$$

Thus the instanton-corrected mirror to Y does carry a holomorphic involution defined by $z \mapsto z^{-1}$, but restricting to the subset $|z| < 1$ does not yield the instanton-corrected mirror to X.

References

[1] M. Abouzaid – "Morse homology, tropical geometry, and homological mirror symmetry for toric varieties", preprint arXiv:math.SG/0610004.

[2] M. Abouzaid, D. Auroux & L. Katzarkov – in preparation.

[3] D. Auroux – "Mirror symmetry and T-duality in the complement of an anticanonical divisor", *J. Gökova Geom. Topol. GGT* **1** (2007), p. 51–91.

[4] D. Auroux, L. Katzarkov & D. Orlov – "Mirror symmetry for del Pezzo surfaces: vanishing cycles and coherent sheaves", *Invent. Math.* **166** (2006), p. 537–582.

[5] C.-H. Cho & Y.-G. Oh – "Floer cohomology and disc instantons of Lagrangian torus fibers in Fano toric manifolds", *Asian J. Math.* **10** (2006), p. 773–814.

[6] K. Fukaya, Y.-G. Oh, H. Ohta & K. Ono – "Lagrangian intersection Floer theory: Anomaly and obstruction", preprint, second expanded version, 2006.

[7] ———, "Lagrangian Floer theory on compact toric manifolds I", preprint arXiv:0802.1703.

[8] M. Gross – "Special Lagrangian fibrations. II. Geometry. A survey of techniques in the study of special Lagrangian fibrations", in *Winter School on Mirror Symmetry, Vector Bundles and Lagrangian Submanifolds (Cambridge, MA, 1999)*, AMS/IP Stud. Adv. Math., vol. 23, Amer. Math. Soc., 2001, p. 95–150.

[9] M. Gross & B. Siebert – "Affine manifolds, log structures, and mirror symmetry", *Turkish J. Math.* **27** (2003), p. 33–60.

[10] ———, "From real affine geometry to complex geometry", preprint arXiv:math.AG/0703822.

[11] N. J. Hitchin – "The moduli space of special Lagrangian submanifolds", *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* **25** (1997), p. 503–515 (1998).

[12] K. Hori – "Mirror symmetry and quantum geometry", in *Proceedings of the International Congress of Mathematicians, Vol. III (Beijing, 2002)*, Higher Ed. Press, 2002, p. 431–443.

[13] K. Hori, A. Iqbal & C. Vafa – "D-branes and mirror symmetry", preprint arXiv:hep-th/0005247.

[14] K. Hori & C. Vafa – "Mirror symmetry", preprint arXiv:hep-th/0002222.

[15] D. Huybrechts – "Moduli spaces of hyperkähler manifolds and mirror symmetry", in *Intersection theory and moduli*, ICTP Lect. Notes, XIX, Abdus Salam Int. Cent. Theoret. Phys., Trieste, 2004, p. 185–247.

[16] D. Joyce – "Special Lagrangian submanifolds with isolated conical singularities. V. Survey and applications", *J. Differential Geom.* **63** (2003), p. 279–347.

[17] ———, "Lectures on Calabi-Yau and special Lagrangian geometry", preprint arXiv:math.DG/0108088.

[18] A. Kapustin & Y. Li – "D-branes in Landau-Ginzburg models and algebraic geometry", *J. High Energy Phys.* **12** (2003).

[19] M. Kontsevich – "Homological algebra of mirror symmetry", in *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994)*, Birkhäuser, 1995, p. 120–139.

[20] ———, "Lectures at ENS, Paris, Spring 1998", notes taken by J. Bellaiche, J.-F. Dat, I. Marin, G. Racinet and H. Randriambololona, unpublished.

[21] M. Kontsevich & Y. Soibelman – "Affine structures and non-Archimedean analytic spaces", in *The unity of mathematics*, Progr. Math., vol. 244, Birkhäuser, 2006, p. 321–385.

[22] N. C. Leung – "Mirror symmetry without corrections", preprint arXiv:math.DG/0009235.

[23] R. C. McLean – "Deformations of calibrated submanifolds", *Comm. Anal. Geom.* **6** (1998), p. 705–747.

[24] D. Orlov – "Triangulated categories of singularities and D-branes in Landau-Ginzburg models", preprint arXiv:math.AG/0302304.

[25] A. Polishchuk & E. Zaslow – "Categorical mirror symmetry: the elliptic curve", *Adv. Theor. Math. Phys.* **2** (1998), p. 443–470.

[26] P. Seidel – "Vanishing cycles and mutation", in *European Congress of Mathematics, Vol. II (Barcelona, 2000)*, Progr. Math., vol. 202, Birkhäuser, 2001, p. 65–85.

[27] ———, *Fukaya categories and Picard-Lefschetz theory*, Zurich Lectures in Advanced Mathematics, European Mathematical Society (EMS), Zürich, 2008.

[28] A. Strominger, S.-T. Yau & E. Zaslow – "Mirror symmetry is T-duality", *Nuclear Phys. B* **479** (1996), p. 243–259.

[29] M. Symington – "Four dimensions from two in symplectic topology", in *Topology and geometry of manifolds (Athens, GA, 2001)*, Proc. Sympos. Pure Math., vol. 71, Amer. Math. Soc., 2003, p. 153–208.

D. Auroux, Department of Mathematics, M.I.T., Cambridge MA 02139, USA
E-mail : auroux@math.mit.edu

Astérisque

JEFF CHEEGER

BRUCE KLEINER

Characterization of the Radon-Nikodym property in terms of inverse limits

Astérisque, tome 321 (2008), p. 129-138

<http://www.numdam.org/item?id=AST_2008__321__129_0>

© Société mathématique de France, 2008, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (http://smf4.emath.fr/Publications/Asterisque/) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

Astérisque
321, 2008, p. 129–138

CHARACTERIZATION OF THE RADON-NIKODYM PROPERTY IN TERMS OF INVERSE LIMITS

by

Jeff Cheeger & Bruce Kleiner

Abstract. — In this paper we clarify the relation between inverse systems, the Radon-Nikodym property, the Asymptotic Norming Property of James-Ho [**10**], and the GFDA spaces introduced in [**5**].

***Résumé* (Caractérisation de la propriété de Radon-Nikodym en termes de limites inverses)**
Dans cet article nous clarifions la relation entre les systèmes inverses, la propriété de Radon-Nikodym, la propriété normative asymptotique de James-Ho [**10**] et les espaces GFDA, introduits dans [**5**].

1. Introduction

A Banach space V is said to have the *Radon-Nikodym Property* (RNP) if every Lipschitz map $f : \mathbf{R} \to V$ is differentiable almost everywhere. By now, there are a number of characterizations of Banach spaces with the RNP, the study of which goes back to Gelfand [**7**]; for additional references and discussion, see [**1**, Chapter 5], [**8**]. Of particular interest here is the characterization of the RNP in terms of the Asymptotic Norming Property; [**10, 8**].

In this paper we will show that a variant of the GFDA property introduced in [**5**] is actually equivalent to the Asymptotic Norming property of James-Ho, and hence by [**10, 8**], is equivalent to the RNP. In addition, we observe that the GFDA spaces of [**5**] are just spaces which are isomorphic to a separable dual space.

Definition 1.1. — An inverse system

$$W_1 \xleftarrow{\theta_1} W_2 \xleftarrow{\theta_2} \ldots \xleftarrow{\theta_{i-1}} W_i \xleftarrow{\theta_i} \ldots, \tag{1.2}$$

2000 Mathematics Subject Classification. — 46XX, 51XX.
Key words and phrases. — Radon-Nikodym Property, Asymptotic Norming Property, inverse limit, separable dual space.

The first author was partially supported by NSF Grant DMS 0105128 and the second by NSF Grant DMS 0701515.

© Astérisque 321, SMF 2008

is *standard* if the W_i's are finite dimensional Banach spaces and the θ_i's are linear maps of norm ≤ 1. We let $\pi_j : \varprojlim W_i \to W_j$ denote the projection map.

Definition 1.3. — Let $\{(W_i, \theta_i)\}$ be a standard inverse system and $V \subset \varprojlim W_i$ be a subspace. The pair $(\varprojlim W_i, V)$ has the *Determining Property* if a sequence $\{v_k\} \subset V$ converges strongly provided the projected sequences $\{\pi_j(v_k)\} \subset W_j$ converge for every j, the sequence $\{\|v_k\|\}$ is bounded, and the convergence $\|\pi_j(v_k)\| \to \|v_k\|$ is uniform in k. A Banach space U has the *Determining Property* if there is a pair $(\varprojlim W_i, V)$ with Determining Property, such that V is isomorphic to U.

We have:

Theorem 1.4. — *A separable Banach space has the RNP if and only it has the Determining Property.*

Since a Banach space has the RNP if and only if every separable subspace has the RNP, Theorem 1.4 yields a characterization of the RNP for nonseparable Banach spaces as well.

To prove the theorem, we first observe in Proposition 2.8 that the inverse limit $\varprojlim W_i$ is the dual space of a separable Banach space. Then, by a completely elementary argument, we show that a Banach space has the Determining Property if and only if it has the Asymptotic Norming Property (ANP) of James-Ho [**10**]. Since a separable Banach space U has the RNP if and only if it has the ANP [**10, 8**], the theorem follows. We remark that there is a simple direct proof that if V has the ANP (or the Determining Property), then every Lipschitz map $f : \mathbf{R} \to V$ is differentiable almost everywhere.

Characterizations of the RNP using inverse limits are useful for applications; see [**5**], the discussion below concerning metric measure spaces, and [**6**].

Relation with previous work. — In slightly different language, our earlier paper [**5**] also considered pairs $(\varprojlim W_i, V)$, where $\varprojlim W_i$ is the inverse limit of a standard inverse system, and $V \subset \varprojlim W_i$ is a closed subspace. A Good Finite Dimensional Approximation (GFDA) of a Banach space V, a notion introduced in [**5**], is a pair $(\varprojlim W_i, V)$ with the Determining Property such that $\pi_i\big|_V : V \to W_i$ is a quotient map for every i.

It follows immediately from Lemma 3.8 of [**5**] that if $(\varprojlim W_i, V)$ is a GFDA of V, then $V = \varprojlim W_i$. Since such inverse limits are dual spaces by Proposition 2.8, V is a separable dual space in this case. Conversely, using the Kadec-Klee renorming Lemma [**11, 12**], it was shown in [**5**] that every separable dual space is isomorphic to a Banach space which admits a GFDA. Thus, a Banach space admits a GFDA if and only if it is isomorphic to a separable dual space.

Applications to metric measure spaces. — We will call a metric measure space (X, μ) a *PI space* if the measure is doubling, and a Poincaré inequality holds in the sense of upper gradients [**9, 4**]. In [**5**], differentiation and bi-Lipschitz non-embedding theorems were proved for maps $f : X \to V$ from PI spaces into GFDA targets V, generalizing results of [**4**] for finite dimensional targets. As explained above, it turns out that these targets are just separable dual spaces, up to isomorphism.

As an application of the inverse limit framework, we will show in [**6**] that the differentiation theorem [**5**, Theorem 4.1] and bi-Lipschitz non-embedding theorem [**5**, Theorem 5.1] hold whenever the target has the RNP.

Acknowledgement. — We are very grateful to Bill Johnson for sharing an observation which helped give rise to this paper. We are much indebted to Nigel Kalton for immediately catching a serious error in an earlier version.

2. Inverse systems

In this section, we recall some basic facts concerning direct and inverse systems, and the duality between them. Then we show that inverse limits of standard inverse systems are precisely duals of separable spaces.

The following conventions will be in force throughout the remainder of the paper.

Definition 2.1. — An *standard direct system* is a sequence of finite dimensional Banach spaces $\{E_i\}$ and 1-Lipschitz linear maps $\iota_i : E_i \to E_{i+1}$.

Definition 2.2. — An *standard inverse system* is a sequence of finite dimensional Banach spaces $\{W_i\}$ and 1-Lipschitz linear maps $\theta_i : W_{i+1} \to W_i$.

Definition 2.3. — A standard direct system is *isometrically injective* if the maps $\iota_i : E_i \to E_{i+1}$ are isometric injections.

Definition 2.4. — A standard inverse system is *quotient* if the maps $\theta_i : W_{i+1} \to W_i$ are quotient maps.

By a *quotient map* of normed spaces, we mean a surjective map $\pi : U \to V$ for which the norm on the target is the quotient norm, i.e. for every $v \in V$,

$$\|v\| = \inf\{ \|u\| \mid u \in \pi^{-1}(v)\}.$$

We will refer to the maps ι_i and θ_i as *bonding maps.*

There is a duality between the objects in Definitions 2.1 and 2.2, respectively, 2.3 and 2.4: if $\{(E_i, \iota_i)\}$ is a standard direct system, then $\{(E_i^*, \iota_i^*)\}$ is a standard inverse system and conversely; similarly, isometrically injective direct systems are dual to quotient systems. To see this, one uses the facts that the adjoint of a 1-Lipschitz map of Banach spaces is 1-Lipschitz and the the adjoint of an isometric embedding is a quotient map. (This follows from the Hahn-Banach theorem.) In particular, since the spaces in our systems are assumed to be finite dimensional (hence reflexive) every inverse system arises as the dual of its dual direct system and conversely. The same holds for quotient inverse systems.

We now recall the definitions of direct and inverse limits.

Given a standard direct system $\{(E_i, \iota_i)\}$ we form the direct limit Banach space $\varinjlim E_i$ as follows. We begin with the disjoint union $\sqcup_i E_i$, and declare two elements $e \in E_i$, $e' \in E_{i'}$ to be equivalent if their images in E_j coincide for some $j \geq \max\{i, i'\}$. Since the bonding maps are 1-Lipschitz, the set of equivalence classes inherits an obvious vector space structure with a pseudo-norm. The direct limit $\varinjlim E_i$ is defined to be the completion of the quotient of this space by the closed subspace of elements whose pseudo-norm is zero. Clearly, there are 1-Lipschitz maps

$$\tau_i : E_i \to \varinjlim E_i \,,$$

which in the case of isometrically injective direct systems, are isometric injections. The union $\bigcup_i \tau_i(E_i)$ is dense in $\varinjlim E_i$.

The inverse limit $\varprojlim W_i$ of a standard inverse system $\{(W_i, \theta_i)\}$ is defined as follows. The underlying set consists of the collection of elements $(w_i) \in \prod_i W_i$ which are compatible with the bonding maps, i.e. $\theta_i(w_i) = w_{i-1}$ for all i, and which satisfy $\sup_i \|w_i\| < \infty$. This is equipped with the obvious vector space structure and the norm

$$\|\{w_i\}\| := \lim_{j\to\infty} \|w_j\| \,. \tag{2.5}$$

The map

$$\pi_j : \varprojlim W_i \to W_j \tag{2.6}$$

given by

$$\pi_j(\{w_i\}) = w_j$$

is 1-Lipschitz, and

$$\lim_{j\to\infty} \|\pi_j(\{w_i\})\| = \|\{w_i\}\| \,.$$

An inverse limit $\varprojlim W_i$ has a natural *inverse limit topology*, namely the weakest topology such that every projection map $\pi_j : \varprojlim W_i \to W_j$ is continuous. Thus a sequence $\{v_k\} \subset \varprojlim W_i$ converges in the inverse limit topology to $v \in \varprojlim W_i$ if and only if for every i, we have $\pi_i(v_k) \to \pi_i(v)$ as $k \to \infty$.

If $\{v_k\} \subset \varprojlim W_i$ and $\{v_k\} \overset{\text{invlim}}{\longrightarrow} v \in \varprojlim W_i$, then

$$\|v\| \;\leq\; \liminf_k \|v_k\| \,. \tag{2.7}$$

Also, every norm bounded sequence $\{v_k\} \subset \varprojlim W_i$ has a subsequence which converges with respect to the inverse limit topology; this follows from a diagonal argument, because $\{\pi_i(v_k)\}$ is contained in a compact subset of W_i, for all i.

Proposition 2.8. — *Given a standard inverse system $\{(W_i, \theta_i)\}$, there is an isometric isomorphism*

$$C : \varprojlim W_i \equiv (\varinjlim W_i^*)^* . \tag{2.9}$$

In particular, $\varprojlim W_i$ is the dual of the separable Banach space $\varinjlim W_i^*$.

Proof. — Pick a compatible sequence $\{x_i\} \in \varprojlim W_i$. We get a map

$$\sqcup W_j^* \to \mathbf{R}$$

by sending $\phi \in W_j^*$ to $\phi(x_j)$. Since $\{x_i\}$ is compatible with bonding maps and

$$|\phi(x_j)| \leq \|\phi\| \, \|x_j\| \leq \|\phi\| \, \|\{x_j\}\|,$$

this defines a linear functional of norm $\leq \|\{x_j\}\|$ on $\varinjlim W_i^*$. Therefore we get a 1-Lipschitz map

$$C : \varprojlim W_i \longrightarrow \left(\varinjlim W_i^*\right)^* .$$

We now verify that C is an isometry.

Pick $(x_i) \in \varprojlim W_i$, and choose $n \in \mathbb{N}$ such that $\|x_n\| \geq \|(x_i)\| - \epsilon$. If $\phi \in W_n^*$ has norm 1 and $\phi(x_n) = \|x_n\|$, then

$$\|C((x_i))\| \, \|\tau_n(\phi)\| \geq C((x_i))(\tau_n(\phi)) = \phi(x_n) = \|x_n\| \geq \|(x_i)\| - \epsilon,$$

where $\tau_n : W_n^* \to \varinjlim W_i^*$ is the canonical 1-Lipschitz map described above. This shows that C is an isometric embedding.

If $\Phi \in (\varinjlim W_i^*)^*$, then we define $\Phi_i \in W_i^{**} = W_i$ to be the composition

$$W_i^* \longrightarrow \varinjlim W_i^* \xrightarrow{\Phi} \mathbf{R}.$$

This defines a compatible sequence $(\Phi_i) \in \varprojlim W_i$, such that $\|(\Phi_i)\| = \|\Phi\|$ and $C((\Phi_i)) = \Phi$. Hence C is onto. $\square$

Corollary 2.10. — 1) *A separable Banach space Y is isomorphic to the direct limit of an isometrically injective direct system (E_i, ι_i).*

2) *The dual space Y^* of the separable Banach space Y (as in 1)) is isometric to the inverse limit $\varprojlim E_i^*$ of the a quotient inverse system $\{(E_i^*, \iota_i^*)\}$.*

Proof. — To see that 1) holds, start with a countable increasing sequence $E_1 \subset E_2 \subset \cdots \subset Y$ of finite dimensional subspaces whose union is dense in Y, and take the bonding maps $\iota_i : E_i \to E_{i+1}$ to be the inclusions. Clearly the inclusion maps $E_i \to Y$ induce an isometry $\varinjlim E_i \to Y$.

Assertion 2) follows from 1) and Proposition 2.8. $\square$

Let C be the isometry in Proposition 2.8.

Lemma 2.11. — 1) *Suppose $\{v_k\} \subset \varprojlim W_i$ is a sequence such that $\{C(v_k)\} \subset (\varinjlim W_i^*)^*$ weak* converges to some $y \in (\varinjlim W_i^*)^*$. Then $\{v_k\}$ is convergent with respect to the inverse limit topology, and its limit $v_\infty \in \varprojlim W_i$ satisfies $C(v_\infty) = y$; in particular, $y \in C(\varprojlim W_i)$.*

2) *If $\{v_k\} \subset \varprojlim W_i$ converges in the inverse limit topology, and has uniformly bounded norm, then $\{C(v_k)\}$ is weak* convergent.*

Proof. — Assertions 1) and 2) follow readily from the assumption that the W_i are finite dimensional together with the density of compatible sequences in $\varprojlim W_i$. □

3. The proof of Theorem 1.4

The proof of Theorem 1.4 is based on the Asymptotic Norming Property, which we now recall.

Let Y denote a separable Banach space and $V \subset Y^*$ a separable subspace of its dual. (Here Y^* need not be separable.)

Definition 3.1. — The pair (Y^*, V) has the *Asymptotic Norming Property* (ANP) if a sequence $\{v_k\} \subset V$ converges strongly provided it is weak* convergent and the sequence of norms $\{\|v_k\|\}$ converges to the norm of the weak* limit.

A Banach space U is said to have the *Asymptotic Norming Property* if there is a pair (Y^*, V) with the ANP such that U is isomorphic to V.

***Theorem 3.2* ([10, 8]).** — *For separable Banach spaces, the RNP is equivalent to the ANP.*

Hence to prove Theorem 1.4, it suffices to show that for separable Banach spaces, the ANP is equivalent to the Determining Property. By Corollary 2.10, every separable Banach space Y is isometric to the direct limit of a standard direct system, and Y^* is isometric to the inverse limit of the dual inverse system. Hence the proof of Theorem 1.4 reduces to:

Proposition 3.3. — *Let $\{(W_i, \theta_i)\}$ be a standard inverse system, and V be a closed separable subspace of $\varprojlim W_i$. Then the pair $(\varprojlim W_i, V)$ has the ANP if and only if it has the Determining Property. Here we are identifying $\varprojlim W_i$ with the dual of $\varinjlim W_i^*$, see Proposition 2.8.*

Proof. — Let $\{v_k\} \subset V$ be a sequence with bounded norm. By Lemma 2.11, the sequence $\{v_k\}$ is weak* convergent if and only if it converges in the inverse limit topology. Therefore, to prove the equivalence of the ANP and the Determining Property for the pair $(\varprojlim W_i, V)$, it suffices to show that when

$$v_k \xrightarrow{w^*} w \in \varprojlim W_i \,, \tag{3.4}$$

the sequence of norms $\{\|v_k\|\}$ converges to the $\|w\|$ if and only if the convergence $\|\pi_j(v_k)\| \to \|v_k\|$ is uniform in k. Although this is completely elementary, we will write out the details.

We have

$$(3.5)\quad \|v_k\| - \|w\| = (\|v_k\| - \|\pi_i(v_k)\|) + (\|\pi_i(v_k)\| - \|\pi_i(w)\|) + (\|\pi_i(w)\| - \|w\|)\,.$$

Assume first that $\lim_{k\to\infty} \|v_k\| = \|w\|$. Given $\epsilon > 0$, there exists I_1 such that $\|w\| - \|\pi_i(w)\| < \epsilon/3$, for $i \geq I_1$. By (3.4) there exists K_1 such that $\|\pi_{I_1}(v_k) - \pi_{I_1}(w)\| < \epsilon/3$, for $k \geq K_1$. Also, there exists K_2 such that $\big|\|v_k\| - \|w\|\big| < \epsilon/3$, if $k \geq K_2$. Set $K = \max(K_1, K_2)$.

From (3.5), with $i = I_1$, we get $\|v_k\| - \|\pi_{I_1}(v_k)\| < \epsilon$, for all $k \geq K$. Since, $\|v_k\| - \|\pi_i(v_k)\|$ is a nonnegative decreasing function of i, this implies, $\|v_k\| - \|\pi_i(v_k)\| < \epsilon$, for all $i \geq I_1$, $k \geq K$.

Finally, there exists I_2 such that $\|v_k\| - \|\pi_i(v_k)\| < \epsilon$ for all $i \geq I_2$, $k = 1, \ldots, K-1$, Thus, if $i \geq \max(I_1, I_2)$ then $\|v_k\| - \|\pi_i(v_k)\| < \epsilon$, for all k.

Conversely, suppose the convergence $\|\pi_i(v_k)\| \to \|v_k\|$ is uniform in k. Given $\epsilon > 0$, there exists I such that $\|v_k\| - \|\pi_i(v_k)\| < \epsilon/3$, for $i \geq I$ and all k. Also, there exists I_1 such that $\|w\| - \|\pi_i(w)\| < \epsilon/3$, for $i \geq I_1$. Set $I' = \max(I, I_1)$. By (3.4), there exists K such that $\|\pi_{I'}(v_k) - \pi_{I'}(w)\| < \epsilon/3$.

From (3.5), with $i = I'$, we get $\big|\|v_k| - \|w\|\big| < \epsilon$, for all $k \geq K$. □

4. A variant of the Determining Property

In this section we discuss a variant of the Determining Property, which was introduced in [**5**] (with a different name). A compactness argument implies that it is equivalent to Definition 1.3, see Proposition 4.6.

For the remainder of this section, we fix a standard inverse system $\{(W_i, \theta_i)\}$ and a closed subspace $V \subset \varprojlim W_i$.

Definition 4.1. — A positive nonincreasing finite sequence $1 \geq \rho_1 \geq \cdots \geq \rho_N$ is *ϵ-determining* if for any pair $v, v' \in V$, the conditions

$$(4.2)\quad \|v\| - \|\pi_i(v)\| < \rho_i \cdot \|v\|, \qquad \|v'\| - \|\pi_i(v')\| < \rho_i \cdot \|v'\|, \qquad 1 \leq i \leq N\,,$$

and

$$(4.3)\quad \|\pi_N(v) - \pi_N(v')\| < N^{-1} \cdot \max(\|v\|, \|v'\|)\,,$$

imply

$$(4.4)\quad \|v - v'\| < \epsilon \cdot \max(\|v\|, \|v'\|)\,.$$

Observe that by dividing by $\max(\|v\|, \|v'\|)$, it suffices to consider pairs v, v' for which $\max(\|v\|, \|v'\|) = 1$.

This leads to the alternate definition of the Determining Property:

Definition 4.5. — The pair $(\varprojlim W_i, V)$ has the *Determining Property* if for every $\epsilon > 0$ and every infinite nonincreasing sequence

$$1 \geq \rho_1 \geq \cdots \geq \rho_i \geq \ldots$$

with $\rho_i \to 0$, some finite initial segment $\rho_1 \geq \cdots \geq \rho_N$ is ϵ-determining.

Proposition 4.6. — *The pair $(\varprojlim W_i, V)$ satisfies Definition 1.3 if and only if it satisfies Definition 4.5.*

Proof. — First we show that the property in Definition 4.5 implies the property in Definition 1.3. So assume that the sequence $\{\|v_k\|\}$ is bounded and the convergence, $\|\pi_i(v_k)\| \to \|v_k\|$ is uniform in k.

Suppose that there exists a sequence, a positive sequence, $\rho_i \searrow 0$, such that $\|v_k\| - \|\pi_i(v_k)\| \leq \rho_i$. By applying the condition in Definition 4.5 to this sequence and using convergence in the inverse limit topology together with (4.3) it is clear from (4.4) that we obtain strong convergence.

Without loss of essential loss of generality, we can assume $\|v_k\| \leq 1$ for all k. Since the convergence, $\|\pi_i(v_k)\| \to \|v_k\|$ is uniform in k, it follows that there exists a strictly increasing sequence, $N_1 < N_2 < \ldots$, such that for all k, we have

$$\|v_k\| - \|\pi_{N_\ell}(v_k)\| < \frac{1}{\ell}\,.$$

Then $\|v_k\| - \|\pi_i(v_k)\| \leq \rho_i$, for the sequence, ρ_i given by

$$\rho_i = \frac{1}{\ell} \qquad (N_\ell \leq i < N_{\ell+1})\,.$$

Conversely, suppose that the property in Definition 1.3 holds, but not the property in Definition 4.5. Then for some decreasing sequence $\{\rho_i\} \subset (0,\infty)$ with $\rho_i \to 0$, and some $\epsilon > 0$, there are sequences $\{v_k\}, \{v'_k\} \subset V$, such that for all $k < \infty$,

$$\|v_k\|,\ \|v'_k\| \leq 1\,, \tag{4.7}$$

$$\max\left(\|v_k\| - \|\pi_i(v_k)\|,\ \|v'_k\| - \|\pi_i(v'_k)\|\right) < \rho_i \quad \text{for } 1 \leq i \leq k\,, \tag{4.8}$$

$$\|\pi_i(v_k) - \pi_i(v'_k)\| < \frac{1}{k}\,, \tag{4.9}$$

$$\|v_k - v'_k\| \geq \epsilon\,. \tag{4.10}$$

By the Banach-Alaoglu theorem, we can pass to weak* convergent subsequences, with respective limits v_∞ and v'_∞. From (4.9), it follows that $v_\infty = v'_\infty$.

It follows from (4.7), (4.8), that the sequences, $\|v_k\|$, $\|v'_k\|$, are bounded and the convergence $\|\pi_i(v_k)\| \to \|v_k\|$, $\|\pi_i(v'_k)\| \to \|v'_k\|$ is uniform in k. Since we assume the property in Definition 1.3, it follows $v_k \to v_\infty$, $v'_k \to v'_\infty$, is actually strong. Since, $v_\infty = v'_\infty$, this contradicts (4.10). □

We remark that proof of the implication Definition 1.3 $\Longrightarrow$ Definition 4.5 is similar to the proof of Proposition 3.11 in [**5**].

5. GFDA versus ANP

We conclude with some remarks about the relation between the ANP and GFDA's; see the introduction and [**5**].

Suppose Y is a separable Banach space and (Y^*, V) has the ANP. By Lemma 2.10, we may realize Y^* – up to isometry – as the inverse limit of a quotient system $\{(W_i, \theta_i)\}$.

Viewing V as a subspace of $\varprojlim W_i$, one might be tempted to modify the inverse system to produce a GFDA of V. For instance, one could restrict the projection maps $\pi_j : \varprojlim W_i \to W_j$ to V, and replace W_j with $\pi_j(V) \subset W_j$. However, the resulting maps $\pi_j|_V : V \to \pi_j(V)$ will usually not be quotient maps. One could also try renorming the spaces $\pi_j(V) \subset W_j$ so that the restrictions $\pi_j|_V : V \to \pi_j(V)$ become quotient maps. This will typically destroy the Determining Property, however. In any case, V will not admit any GFDA unless it is a separable dual space, whereas many Banach spaces with the RNP are not separable dual spaces.

In fact, there are seperable spaces with the RNP which are not isomorphic subspaces of dual spaces with the RNP; see [**13**], [**2**]; compare also [**3**]. We are indebted to the referee for providing these references.

References

[1] Y. Benyamini & J. Lindenstrauss – *Geometric nonlinear functional analysis. Vol. 1*, American Mathematical Society Colloquium Publications, vol. 48, Amer. Math. Soc., 2000.

[2] J. Bourgain – *New classes of $\mathcal{L}^p$-spaces*, Lecture Notes in Math., vol. 889, Springer, 1981.

[3] R. D. Bourgin – *Geometric aspects of convex sets with the Radon-Nikodým property*, Lecture Notes in Math., vol. 993, Springer, 1983.

[4] J. Cheeger – "Differentiability of Lipschitz functions on metric measure spaces", *Geom. Funct. Anal.* **9** (1999), p. 428–517.

[5] J. Cheeger & B. Kleiner – "On the differentiability of Lipschitz maps from metric measure spaces to Banach spaces", in *Inspired by S. S. Chern*, Nankai Tracts Math., vol. 11, World Sci. Publ., Hackensack, NJ, 2006, p. 129–152.

[6] ______, "Differentiability of Lipschitz maps from metric measure spaces to Banach spaces with the Radon-Nikodym Property", preprint, 2008.

[7] I. M. Gelfand – "Abstrakte Funktionen und lineare Operatoren", *Mat. Sbornik* **46** (1938), p. 235–284.

[8] N. Ghoussoub & B. Maurey – "The asymptotic-norming and the Radon-Nikodým properties are equivalent in separable Banach spaces", *Proc. Amer. Math. Soc.* **94** (1985), p. 665–671.

[9] J. Heinonen & P. Koskela – "From local to global in quasiconformal structures", *Proc. Nat. Acad. Sci. U.S.A.* **93** (1996), p. 554–556.

[10] R. C. James & A. Ho – "The asymptotic-norming and Radon-Nikodým properties for Banach spaces", *Ark. Mat.* **19** (1981), p. 53–70.

[11] M. Ĭ. Kadec′ – "On the connection between weak and strong convergence", *Dopovidi Akad. Nauk Ukraïn. RSR* **1959** (1959), p. 949–952.

[12] V. Klee – “Mappings into normed linear spaces”, *Fund. Math.* **49** (1960/1961), p. 25–34.

[13] P. W. McCartney & R. C. O'Brien – “A separable Banach space with the Radon-Nikodým property that is not isomorphic to a subspace of a separable dual”, *Proc. Amer. Math. Soc.* **78** (1980), p. 40–42.

J. Cheeger, Courant Institute of Mathematical Sciences, 251 Mercer Street, New York, NY 10012, USA

B. Kleiner, Mathematics Department, Yale University, New Haven, CT 06520, USA

Astérisque

XIUXIONG CHEN
YUDONG TANG

Test configuration and geodesic rays

Astérisque, tome 321 (2008), p. 139-167
<http://www.numdam.org/item?id=AST_2008__321__139_0>

© Société mathématique de France, 2008, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (http://smf4.emath.fr/Publications/Asterisque/) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM
Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

Astérisque
321, 2008, p. 139–167

TEST CONFIGURATION AND GEODESIC RAYS

by

Xiuxiong Chen & Yudong Tang

Dedicated to Professor J. P. Bourguignon,
with affection, gratitude and admiration

Abstract. — This paper presents recent research findings on the connection between test configuration and geodesic ray in Kähler metric space. The purpose was to gain insight on the degeneration of Kähler metrics along geodesic rays. A result associating every smooth test configuration a $C^{1,1}$ geodesic ray is proved and exemplified with toric degenerations. Furthermore, we show that the ¥ invariant agrees with Futaki invariant, thus acts as a good substitute in general $C^{1,1}$ geodesic rays without a background test configuration. Based on the assumption of simple test configuration, we extend Donaldson's correspondence between solutions of Monge-Ampère equation and holomorphic discs. Results indicate that Chen and Tian's analysis on Monge-Ampère equation via holomoprhic discs could apply in simple test configuration.

***Résumé* (Configuration de test et rayons géodésiques).** — Cet article présente les dernières découvertes sur la connexion entre la configuration de test et les rayons géodésiques dans les espaces métriques kähleriens. Un résultat qui associe à chaque configuration de test lisse un $C^{1,1}$-rayon géodésique est démontré, et nous fournissons des exemples avec des dégénérations toriques. D'autre part, nous montrons que l'invariant ¥ s'accorde avec celui de Futaki, et forme ainsi un bon substitut dans le cas de $C^{1,1}$-rayons géodésiques généraux sans configuration de test. En nous basant sur l'hypothèse d'une configuration de test simple, nous étendons la correspondance de Donaldson entre les solution de l'équation de Monge-Ampère et les disques holomorphes. Les résultats indiquent que l'analyse de Chen et Tian sur l'équation de Monge-Ampère par le biais des disques holomorphes pourrait s'applique dans les configurations de test simples.

1. Introduction

The purpose of this paper is to explore the connection between geodesic rays in the space of Kähler metrics and test configurations in algebraic manifold [**15**]. This

2000 Mathematics Subject Classification. — 53C55.
Key words and phrases. — Monge-Ampère equation, test configuration, geodesic ray, Futaki invariant, ¥ invariant, toric degeneration.

Partially supported by a NSF grant.

© Astérisque 321, SMF 2008

is a continuation of [**9**] in some aspects. In [**7**], the first named author and E. Calabi proved that the space of Kähler potentials is a non-positive curved space in the sense of Alexanderov. As a consequence, they proved that for any given geodesic ray and any given Kähler potential outside of the given ray, there always exists a geodesic ray in the sense of metric distance (L^2 in the Kähler potentials) which initiates from the given Kähler potential and parallel to the initial geodesic ray. The initial geodesic ray, plays the role of prescribing an asymptotic direction for the new geodesic ray out of any other Kähler potential. When the initial geodesic ray is smooth and is tamed by a bounded ambient geometry, the first named author [**9**] proved the existence of relative $C^{1,1}$ geodesic ray from any initial Kähler potential. (These definitions can be found in Section 2.) Similarly, as remarked in [**9**], a test configuration should play a similar role. One would like to know if it induces a relative $C^{1,1}$ geodesic ray from any other Kähler potential in the direction of test configuration. In [**3**], Arezzo and Tian proved a surprising result that for a smooth test configuration with analytic (smooth) central fiber, there always exists a general fiber sufficiently closed to the central fiber, such that there exists a smooth geodesic ray initiated from that fiber metric, and be asymptotically closed to the test configuration (or approximating to some analytic metric in the central fiber). A natural question, motivated by Arezzo-Tian's work, is if there exists a relative geodesic ray from arbitrary initial Kähler metric which also reflects the same geometry (i.e., degenerations) of the underlying test configuration. In section 3, we prove

Theorem 1.1. — *Every smooth test configuration induces a relative $C^{1,1}$ geodesic ray from any Kähler potential in the given class.* (1)

Test configurations can be viewed as algebraic rays, which are geodesics in a finite dimensional subspace(with new metric) of space of Kähler metrics. The geodesic rays induced by a test configuration are the rays parallel to the algebraic ray. They automatically have bounded ambient geometry introduced by the first named author [**9**].

Theorem 1.2. — *For simple test configuration* (2)*, if the induced geodesic ray is smooth regular* (3)*, then the generalized Futaki invariant agrees with the ¥ invariant* (4)*.*

In 1982, E. Calabi asked if there always exists an extremal Kähler metric in every Kähler class [**5**]. This is a very ambitious conjecture which includes his famous conjecture on Kähler Einstein metric (when the first Chern class has a definite sign) as

(1) Following ideas of [**9**], the smooth assumption can be reduced to a lower bound of the Riemannian curvature of the total space.
(2) Definition 2.3.
(3) Definition 2.1, it is also equivalent to Definition 6.2 in this case.
(4) The ¥ invariant is defined by the first named author [**9**].

a special case. It was soon pointed out by Levine [**19**] that Calabi's conjecture can not hold for general Kähler class. However, it is understood among the experts that, with some modification, Calabi's conjecture might hold for general Kähler manifolds. Unfortunately, it is truely subtle and elusive to search/fromulate a correct statement regarding the existence of constant scalar curvature Kähler (cscK) metrics.

The generalized Futaki invariant or algebraic Futaki invariant is an algebraic notion which relates to the stability of projective manifolds. In the late 1990s, S. T. Yau conjectured that the existence of Kähler Einstein metrics in Fano manifolds is equivalent to some form of Stability of the underlying polarized Kähler class. Even though what stability notion to use is also part of puzzle, this is indeed a fundamental conjecture with respect to Kähler Einstein metrics. According to G. Tian [**34**] and Donaldson [**12**], this equivalence relation should be extended to include the case of the constant scalar curvature (cscK) metric in a general Kähler class. In [**34**] , G. Tian introduced the notion of K-Stability and in the same paper, he proved that the existence of KE metric implies weak K stability. In [**13**], Donaldson proved that, in algebraic manifold with discrete automorphism group, the existence of cscK metrics implies that the underlying Kähler class is Chow-Stable. In this paper, Donaldson actually formulated a new version (but equivalent) of K-Stability in terms of weights of Hilbert points. In Kähler toric varieties, the existence of cscK metrics implies that the underlying Kähler class is Semi-K stable [**15**]. Now it is a well-known conjecture that the existence of constant scalar curvature metrics, is equivalent to the K stability of the underlying complex polarization (the so called "Yau-Tian-Donaldson conjecture").

In [**9**], the first named author used the ¥ invariant to define geodesic stability. Theorem 1.2 states that geodesic stability in the algebraic manifold, is a proper generalization of K stability, at least conceptually. The first named author believes that the existence of KE metrics is equivalent to the geodesic stability introduced in [**9**]. Note that the geodesic stability introduced in [**9**] is a mild modification of a similar concept of S. K. Donaldson [**12**].

The Yau-Tian-Donaldson conjecture is a central problem in Kähler geometry now. Through the hard work of many mathematicians, we now know more about one direction (from existence to stability), cf. Tian [**34**], Donaldson [**16**] , Mabuchi [**22**], Paul-Tian [**23**], Phong-Sturm [**24**], Chen-Tian [**10**]... But on the direction from algebraic stability to existence, few progress has been made though. However, in toric manifolds, there has been special results of Donaldson [**15**] and Zhou-Zhu [**37**].

There is a recent intriguing work by V.Apostolov, D.Calderbank, P.Gauduchon and C.W.Tonnesen-Friedman [**2**]. They constructed an example which is suspected to be algebraically K stable [5], but admits no extremal Kähler metric. Perhaps one might speculate that, the geodesic stability aforementioned is one of the possible alternatives since it appears to be stronger than K stability and it is a non algebraic notion in nature.

The converse to Theorem 1.1 is widely open. In other words, it is hard to compactify a geodesic ray. The rays induced by any test configuration is very special in many aspects. For instance, generally speaking, the foliation of a smooth geodesic ray is a family of open strips which cover the base punctured disc. However, for the smooth geodesic rays induced from a test configurations, the strips always close up as punctured disc, or we may say that, the orbits are periodic. Unfortunately, having a periodic orbit does not appear to be enough to construct a test configuration. It would be a very intriguing problem to find a sufficient condition so that we can "construct" a test configuration out of a "good" geodesic ray.

Question A. — *Is there a canonical method to construct some test configuration/algebraic ray such that it reflects the same degeneration of a geodesic ray? What is natural geometric conditions on the "good" geodesic ray?*

Our second main result is to establish the correspondence between smooth regular solutions of Homogeneous complex Monge-Ampère equation (HCMA) on simple test configurations and some family of holomorphic discs in an ambient space $\mathcal{W}$ which will be explicitly constructed. We prove, in section 5:

Theorem 1.3. — *There is a one to one correspondence between smooth regular solutions of HCMA on simple test configuration $\mathcal{M}$ and families of holomorphic discs in $\mathcal{W}$ with proper boundary condition.* [6]

Note that in the case of disc, S. K. Donaldson [**14**] and S. Semmes [**30**] established first such a correspondence between the regularity of the solution of the HCMA equation and the smoothness of the moduli space of holomorphic discs whose boundary lies in some totally real sub-manifold. The theorem above is a generalization of Donaldson's result. Following this point of view, the regularity of the solution is essentially the same as the smoothness of the moduli space of these holomorphic discs under perturbation. As in [**14**], we proved the openness of smooth regular solutions in Section 6.

Theorem 1.4. — *Let $\rho(t)$ be a smooth regular geodesic ray induced by a simple test configuration. Then there exists a parallel smooth regular geodesic ray for any initial point sufficiently close to $\rho(0)$ in C^∞ sense.*

(5) Generalized K stable for extremal Kähler metrics, cf. [**32**].

(6) In a followup work, we expect to extend this to all smooth test configurations.

An immediate corollary is that the smooth geodesic ray constructed by Arezzo-Tian is open for small deformation of the initial Kähler potential. One may wonder what about the closeness of these solutions? Note that the first named author and Tian [**10**] studied the compactness of these holomorphic discs in the disc setting and we believe that the technique of [**10**] can be extended over here.

In Section 7, as a special case, we explore the geodesic rays induced by toric degenerations [**15**]. In particular, we found plenty of geodesic rays whose regularity is at most $C^{1,1}$ globally. We prove:

Theorem 1.5. — *The geodesic ray induced by a toric degeneration has the initial direction equal to the extremal function in the polytope representation.*

More interestingly, we can write down the geodesic ray explicitly in polytope representation. Thus, the various invariants and energies can be calculated explicitly. This should have general interest since there are very few non-trivial examples of geodesic segments or rays in the literature.

Acknowledgments. — Both authors are grateful to G. Tian for many insightful discussions. The first named author is grateful to S. K. Donaldson for many discussions in this subject.

The first named author has been lecturing on these theorems since spring of 2007. In particular, he lectured in a week long conference on geometric analysis (June 17–22, 2007) held at Luminy, France.

When we are ready to post our paper, the authors noticed Phong-Sturm's work [**27**] which overlaps with our theorem 1.1.

2. Preliminary

2.1. Geodesic rays in Kähler potential space. — Let (M, ω, J) be a compact Kähler manifold of complex dimension n. This means J is an integrable complex structure and the symplectic form ω is compatible with J. In another word, $\omega(J\cdot, J\cdot) = \omega(\cdot,\cdot)$, and $g = \omega(\cdot, J\cdot)$ is a metric.

In local complex coordinates $z_\alpha = x_\alpha + i y_\alpha$, denote the metric $g = \omega(\cdot, J\cdot)$ by $g_{\alpha\bar\beta} dz^\alpha \otimes dz^{\bar\beta}$. Then $g_{\alpha\bar\beta}$ is the complexification of the real metric g_{ij}.

By definition, $\omega = \frac{\sqrt{-1}}{2} g_{\alpha\bar\beta} dz^\alpha \wedge dz^{\bar\beta}$. Let

$$\mathcal{H} = \{\phi \in C^\infty(M) : g_{\alpha\bar\beta} + \frac{\partial^2\phi}{\partial z_\alpha \partial z_{\bar\beta}} > 0\}. \tag{1}$$

It follows from the $\partial\bar\partial$ lemma that $\mathcal{H}$ is the moduli space of all Kähler metrics in the class $[\omega]$.

$\mathcal{H}$ is an infinite dimensional manifold with formal tangent space $T\mathcal{H}_\phi = C^\infty(M)$. T. Mabuchi [**21**] defined a metric as the following: Let $\phi_1, \phi_2 \in T\mathcal{H}_\phi$.

$$(2) \qquad < \phi_1, \phi_2 >_{\omega_\phi} = \int_M \phi_1 \phi_2 d\mu = \int_M \phi_1 \phi_2 \frac{\omega_\phi^n}{n!} = \int_M \phi_1 \phi_2 \frac{(\omega + i\partial\bar{\partial}\phi)^n}{n!}.$$

This metric was also defined in S. Semmes [**29**] and S.K. Donaldson [**12**]. Under this metric, the geodesic equation for curve $\phi(t) \in \mathcal{H}$ is the following:

$$(3) \qquad \ddot{\phi} - g_\phi^{\alpha\bar{\beta}} \dot{\phi}_\alpha \dot{\phi}_{\bar{\beta}} = 0.$$

It is just the Euler-Lagrange equation of the energy $E(\phi(t)) = \int_0^1 \int \dot{\phi}^2 \frac{\omega_\phi^n}{n!} dt$. According to Semmes [**29**], the geodesic equation can be transferred into a Complex Monge-Ampère equation: Let $\Sigma = [0,1] \times S^1$, a Riemann surface. Now ϕ is originally defined for $t \in [0,1]$. Extend ϕ to be S^1 invariant function on Σ. Let $z = t + is$ be complex coordinate of Σ, $\{w_\alpha, 1 \leq \alpha \leq n\}$ be a local coordinates on M. Then the geodesic equation is transformed into

$$(4) \qquad \det \begin{pmatrix} g_{\alpha\bar{\beta}} + \phi_{\alpha\bar{\beta}} & \phi_{\alpha\bar{z}} \\ \phi_{z\bar{\beta}} & \phi_{z\bar{z}} \end{pmatrix} = 0.$$

In another word, it is $(\Omega + i\partial\bar{\partial}\phi)^{n+1} = 0$ on $M \times \Sigma$, where $\Omega = \pi^*\omega$ is the pull back of ω by the projection $\pi : M \times \Sigma \to M$.

A geodesic segment connecting two points ϕ_0 and ϕ_1 is the solution of the following Drichelet boundary value problem.

$$(5) \qquad \det \begin{pmatrix} g_{\alpha\bar{\beta}} + \phi_{\alpha\bar{\beta}} & \phi_{\alpha\bar{z}} \\ \phi_{z\bar{\beta}} & \phi_{z\bar{z}} \end{pmatrix} = 0 \text{ on } M \times \Sigma,$$

$$(6) \qquad \phi = \phi_0 \text{ on } M \times 0 \times S^1,$$

$$(7) \qquad \phi = \phi_1 \text{ on } M \times 1 \times S^1.$$

Definition 2.1. — Smooth regular solution: We call ϕ a smooth regular solution (sometimes smooth solution for simplicity) of the Monge-Ampère equation, if ϕ is smooth and if $g_{\alpha\bar{\beta}} + \phi_{\alpha\bar{\beta}} > 0$ hold on all fibers.

In [**8**], The first named author proved the existence of a $C^{1,1}$ solution to above equation. He used the continuity method to solve $\det = \epsilon f$ equation, and proved the following: For every $\epsilon > 0$, there is a unique smooth solution ϕ_ϵ with $|\partial\bar{\partial}\phi_\epsilon| < C$. The C only depends on the background metric and the manifold. In fact, his proof works for Monge-Ampère equation on general compact complex manifold with boundary. He also proved the uniqueness of the limit when $\epsilon \to 0$. Notice that the uniqueness is expected since $\mathcal{H}$ is negatively curved space. T. Mabuchi [**21**], S. Semmes [**29**] and Donaldson [**12**] showed that $\mathcal{H}$ is negatively curved in formal sense and later, the first named author and Calabi [7] proved it is negatively curved in the sense of Alexanderov.

The regularity beyond $C^{1,1}$ is missing. Our example in section 7 shows a solution with no global C^3 bound. A similar setup [14] to the geodesic equation is concerned Monge-Ampère equation on $M \times D$ instead of $M \times (I \times S^1)$. In that setup, Donaldson showed that there exists boundary value such that there is no smooth regular solution. In this direction, a deep analytic result is [10]. The first named author and Tian characterized the singularity in detail by analyzing the holomorphic discs associated to a solution.

In the geodesic ray case, the equation holds on $M \times [0,\infty) \times S^1$ instead of $M \times I \times S^1$. By changing variable: $z = e^{-(t+is)}$, the strip $[0,\infty) \times S^1$ goes to a punctured disc. The equation becomes $(\Omega + i\partial\bar{\partial}\phi)^{n+1} = 0$ on $M \times (D-0)$. The well posed question for geodesic ray is a "starting potential", as well as prescribing an "asymptotic direction." This "asymptotic direction" is usually given by either a known geodesic ray with bounded geometry or a smooth test configuration. In [9], we study the existence of geodesic ray with given geodesic ray as "asymptotic direction." Part of the goal of this paper is to established the existence result with respect to test configuration and to explore the relation of geodesic rays with test configurations.

2.2. Test configuration and equivariant embedding. — Test configuration is defined first by Donaldson [15]. He used test configurations to study the relation between stability of projective manifolds and the existence of extremal Kähler metrics. Test configuration is parallel to the notion "special degeneration" introduced by Tian [34] earlier. Both notions describe a certain degeneration of Kähler manifolds. As discussed already in [12], the geodesic ray represents also degeneration of Kähler metrics. Therefore, it is natural to relate these notions together.

Following Donaldson's definition,

Definition 2.2. — Let $L \to M$ be an ample line bundle over a compact complex manifold. A test configuration $\mathcal{M}$ consists of:

1. a scheme $\mathcal{M}$ with a C^*−action.
2. a C^*−equivariant line bundle $\mathcal{L} \to \mathcal{M}$.
3. a flat C^*−equivariant map $\pi : \mathcal{M} \to C$, where C^* acts on C by multiplication. Any fiber $M_t = \pi^{-1}(t)$ for $t \neq 0$ is isomorphic to M. The pair (L^r, M) is isomorphic to $(\mathcal{L}|_{M_t}, M_t)$ for some $r > 0$, in particular, $(L^r, M) = (L_1, M_1)$.

Test configuration is more explicit in the view of equivariant embedding [28]. Without loss of generality, assume $r = 1$. For large k, $\mathcal{L}^k \to \mathcal{M} \to C$ can be embedded into $\mathcal{O}(1) \to P^N \times C \to C$ equi-variantly. It means there is a C^* action on $\mathcal{O}(1) \to P^N \times C \to C$, which restricts to the C^* action of the embedded

$\mathcal{L}^k \to \mathcal{M} \to C$. In fact, the embedding of each fiber M_t is just the Kodaira embedding by the linear system $H^0(M_t, \mathcal{L}^k|_{M_t})$. Moreover, one can make the S^1 action on $\mathcal{O}(1) \to P^N \times C \to C$ unitary.

In the rest of the paper, we always treat test configurations as equi-variantly embedded with $r = 1, k = 1$. Therefore, we work on a subspace of $P^N \times C$. Also, in geodesic ray problem, there is no loss of generality to only look at truncated test configuration $\mathcal{M} \to D$.

At last, we define a special kind of test configuration.

Definition 2.3. — Simple test configuration: A test configuration $\mathcal{M} \subset P^N \times D$ is called simple if the total space is smooth ($\mathcal{M}$ is a smooth sub-manifold of $P^N \times D$) and the projection $\pi : \mathcal{M} \to D$ is submersion everywhere.

By definition, the central fiber of a simple test configuration is automatically smooth.

3. Relative $C^{1,1}$ geodesic ray from smooth test configuration

3.1. Existence. — As mentioned before, test configuration represents some degeneration of a Kähler manifold along a C^* action. Geodesic ray represents a degeneration of Kähler metrics along a punctured disc. So it is natural to relate the truncated test configuration to a geodesic ray. We have the following theorem:

Theorem 3.1. — *A smooth truncated test configuration $\mathcal{M} \to D$ induces a relative $C^{1,1}$ geodesic ray from any given initial point $p \in \mathcal{H}$.*

The existence is a direct application of the first named author's result [**8**]. The key ingredient of this theorem is the boundary estimate in [**8**]. For Homogenous complex Monge-Ampère equation, there is an extensive literature in the subject (cf. [**4**], [**18**], [**35**]...).

At present, we assume that the total space of the test configuration is smooth. We expect that these results can be extended to singular test configurations accordingly. For instance, in [**9**], the first named author took another approach to construct the geodesic ray. Using techniques in [**9**], the smoothness condition here can be reduced to a uniform lower bound of the Riemannian curvature of the total space.

Proof. — Consider a smooth test configuration over a disc: $(\mathcal{L} \to \mathcal{M} \to D) \hookrightarrow (\mathcal{O}(1) \to P^N \times D \to D)$. Assume the total space is smooth. i.e, $\mathcal{M} \subset P^N \times D$ is smooth. Let Ω be the Fubini-study metric on $P^N \times D$. Actually, it means the pull back of Fubini-study metric on P^N by projection: $P^N \times D \to P^N$.

Now solve the equation

$$(8)\qquad (\Omega + \sqrt{-1}\partial\bar{\partial}\psi)^{n+1} = 0 \text{ on } \mathcal{M},$$

$$(9)\qquad \psi = 0 \text{ on } \partial\mathcal{M}.$$

According to [**8**], this equation has a $C^{1,1}$ solution (it is not exactly the same situation as in [**8**], but the techniques are the same). The following shows that: This solution corresponds to a geodesic ray in the Kähler class $c_1(L)$.

The C^* action on $\mathcal{M}$ induces a biholomorphic map $i : (L_1, M_1) \times (D - 0) \to (\mathcal{L}, \mathcal{M}) - M_0$. Now i maps $(e, x, z) \in (L_1, M_1) \times (D - 0)$ to $z \circ (e, x, 1) \subset (\mathcal{L}, \mathcal{M})$. $z\circ$ is the C^* action of test configuration, and $(e, x, 1) \in (L_1, M_1)$. The map i pulls the equation to

$$(10)\qquad (i^*\Omega + \sqrt{-1}\partial\bar{\partial} i^*\psi)^{n+1} = 0.$$

on $M_1 \times (D - 0)$, with boundary condition $i^*\psi = 0$ on $M_1 \times S^1$.

Let $\omega = \Omega|_{M_1}$, and $\pi : M_1 \times (D - 0) \to M_1$ be the projection, then

Proposition 3.2. — *$i^*\Omega = \pi^*\omega + \sqrt{-1}\partial\bar{\partial}\eta$ for some smooth function η.*

Proof. — Let h be the Fubini-Study hermitian metric on $\mathcal{O}(1) \to P^N$. So $\Omega = -\sqrt{-1}\partial\bar{\partial}\log h$ and $i^*\Omega = -\sqrt{-1}\partial\bar{\partial}\log i^*h$. Note $\pi^*\omega = -\sqrt{-1}\partial\bar{\partial}\log h_1$. h_1 is the pull back of the hermitian metric on line bundle $L_1 \to M_1$ by trivial projection $\pi : (L_1, M_1) \times (D - 0) \to (L_1, M_1)$. So $i^*\Omega = \pi^*\omega + \sqrt{-1}\partial\bar{\partial}\log\frac{h_1}{i^*h}$ and $\eta = \log\frac{h_1}{i^*h}$. □

Proposition 3.3. — *$\varphi = \eta + i^*\psi$ is a geodesic ray.*

Proof. — We have shown $(\pi^*\omega + \sqrt{-1}\partial\bar{\partial}\varphi)^{n+1} = 0$ on $M \times (D - 0)$. It remains to show the S^1 invariance of φ. First, we check the S^1 invariance of η. By assumption, S^1 action on $\mathcal{O}(1) \to P^N \times C$ is unitary. So the h is preserved by S^1 action. This immediately implies that $\eta = \log\frac{h_1}{i^*h}$ is S^1 invariant. Now we check ψ. ψ is S^1 invariant because the boundary condition $\psi = 0$ is S^1 invariant, and the uniqueness of Monge-Ampère solution. In another word, for the unique solution, the S^1 symmetric on the boundary will force the S^1 symmetry in the interior. Now both η and ψ are S^1 invariant, so is φ. □

Back to the proof of the theorem 3.1: At this moment, we have associated a relative $C^{1,1}$ geodesic ray to the test configuration. The ray starts from a fixed point p, because we solved the equation with boundary condition $\psi = 0$. However, for another arbitrary point q, one can go back to the equation 8, solve $\psi = \psi_0$ on $\partial\mathcal{M}$ and obtain the relative $C^{1,1}$ ray from q. ψ_0 is the S^1 extension of the potential difference between q and p. □

In [3], Arezzo and Tian constructed an analytic geodesic ray from a test configuration when the central fiber is analytic. Such test configurations in [3] are simple test configuration (cf. Defi. 2.3). Using the openness Theorem 6.5, we know that there are smooth geodesic rays near the ray they constructed.

When the test configuration is simple (cf. Defi. 2.3), one may expect some better regularity of the induced geodesic ray. Using the correspondence in section 5, the techniques developed by the first named author and Tian in [10] would apply. We expect a similar regularity result here: For any boundary condition $\phi \in C^{k,\alpha}$, there exists nearby perturbation $\phi_\epsilon, |\phi_\epsilon - \phi|_{C^{k,\alpha}} < \epsilon$, such that the HCMA with boundary value ϕ_ϵ has a almost smooth solution [(7)]. When the test configuration is not simple, bad regularity may appear, maybe due to lack of the correspondence in section 5. For example, in the case of toric degenerations: The total space is smooth when the total polytope is delzant, but the central fiber is never smooth. The geodesic ray is piece wise smooth and has no global C^3 bound. The singularity set on polytope representation has real codimension 1.

Back to the question raised in the introduction: given a geodesic ray, how to construct a test configuration which represents the same degeneration? Donaldson's construction of toric degenerations [15] is very inspiring: He chose piecewise linear functions to approximate an arbitrary direction. A piecewise linear function can lead to a well defined test configuration. In principle, one might view the degenerations represented by a test configuration are dense in all possible geometrical degenerations. Donaldson's construction suggests a way to choose a good approximation, which reflects the same character of degeneration.

3.2. Special cases: geodesic line and Toric variety. — One example of geodesic ray is the geodesic line generated by a holomorphic vector field. Let M be a Kähler manifold with Kähler form ω. Let X be a holomorphic vector field such that: $X = f^{,\alpha}\frac{\partial}{\partial w^\alpha}$ for some real potential f. It is well known that $Im(X)$ is killing vector field. Let $\sigma(t)$ be the flow generated by $Re(X) = \nabla_\omega f$. Then, the 1-parameter family $\omega_{\rho(t)} = \sigma(t)^*\omega$ is a geodesic line, $t \in (-\infty, \infty)$.

Nontrivial example of geodesic rays can be explicitly constructed in toric varieties. For a toric variety, there is an associated polytope. More specifically, there is a biholomorphic map $f : M^\circ = C^n/2\pi i Z^n \to P^\circ \times T^n$. Here M° is an open dense subset of $\mathcal{M}$ where the toric action is free. P is a polytope in R^n satisfying Delzant conditions. Represent a toric-invariant Kähler metric as $\omega|_{M^\circ} = i\partial\bar{\partial}f$, then there is a map f from

$$C^n/2\pi i Z^n \longrightarrow P^\circ \times T^n, \qquad \forall\,(u,v) \to (x = \frac{\partial f}{\partial u}, y = v).$$

(7) For definition of almost smooth solution, see the first named author and Tian [10].

Under this map , the Kähler form ω is translated into $dx \wedge dy$. The complex structure is translated into

$$J = \begin{pmatrix} 0 & G \\ G^{-1} & 0 \end{pmatrix}. \tag{11}$$

where

$$(G_{ij}) = \left(\frac{\partial^2 g}{\partial x_i \partial x_j}\right), \qquad \text{and } g(x) + f(u) = \sum x_i\, u_i, \text{ at } x = \frac{\partial f}{\partial u}.$$

In another word, in the symplectic chart, the complex structure has a potential g.

This transformation is really helpful for the geodesic equation. The geodesic equation, in the polytope representation, is linear for complex structure potential $g(t)$. In other words,

$$\ddot{g}(t) = 0. \tag{12}$$

This immediately implies the existence of smooth geodesics segment connecting any two toric metrics. It is just the linear interpolation of the two end potentials.

4. Connection between algebraic notions and geometric notions

4.1. Algebraic ray and geodesic ray. — Test configurations can be viewed as algebraic rays. The induced geodesic rays are parallel to the algebraic ray.

Definition 4.1. — Two rays $\rho_1(t)$ and $\rho_2(t)$ in the space of Kähler metrics are called parallel if $\rho_1(t) - \rho_2(t)$ is uniformly bounded.

The equality $\varphi = \eta + i^*\psi$ can be interpreted geometrically. η represents the degeneration of the metric from the algebraic C^* action. ψ is the difference between the algebraic ray and the differential geometric ray. Notice that ψ is $C^{1,1}$ bounded. We will elaborate above statement in the following:

Recall that $(L, M) = (L_1, M_1) \hookrightarrow (\mathcal{O}(1), P^N)$ is embedding. The group $GL(N+1, C)$ acts on $(\mathcal{O}(1), P^N)$. If one looks at the dual bundle of $\mathcal{O}(1)$ (i.e. the universal bundle $\{(e, x) \in C^{N+1} \times P^N : e = \lambda x\}$), the action is simply $A(e, x) = (Ae, Ax), \forall\, A \in GL(N+1, C)$. The natural dual map between $\mathcal{O}(1)$ and universal bundle passes the action from one to the other.

Consequently, the action acts on the Hermitian metric of $\mathcal{O}(1)$, thus on its curvature. The following lemma shows that this action preserves the positivity of the Hermitian curvature.

Lemma 4.2. — *Let $A \in GL(N+1, C)$ and h be the Fubini-Study hermitian metric on $\mathcal{O}(1)$. Then, $-i\partial\bar{\partial} \log A^* h > 0$.*

Proof. — It suffices to prove that the action preserves the negativity of curvature on the universal bundle. Under the action A, the metric of $e = (X_0, X_1, ..., X_N) \in \mathcal{O}(-1)$ changes into $||Ae||^2$ from standard Fubini-Study metric $||e||^2$. Notice that the action $A^{-1}UA$ for $U \in U(N+1)$ is transitive on P^N and this action preserves A^*h. Thus, one only need to show the negativity at one point. Let's consider the point $p = A^{-1}(1, 0, ..., 0)^t$, and $e = (X_0, ..., X_{i-1}, 1, X_{i+1} ..., X_N)$. At this point p, we have

$$-\sqrt{-1}\partial\bar{\partial}\log ||Ae||^2 = -\sqrt{-1}\sum_{j=1}^{n}\sum_{k,l\neq i} A_{jk}\bar{A}_{jl}dX_k \wedge d\bar{X}_l. \tag{13}$$

To show the positivity, it suffices to show that the null space of the matrix $A_{jk}, j \neq 1, k \neq i$ must be empty. If $v = (\alpha_0, ..., \alpha_{i-1}, \alpha_{i+1}, ...\alpha_N)$ is a null vector, then the vector Av^t must be of form $(c \neq 0, 0, 0, ..., 0)$, because of non-singularity of A. By scaling $c = 1$, A will map two vectors to $(1, 0, ..., 0)$, which is a contradiction. □

As a consequence, the $GL(N+1, C)$ action induces a finite dimensional subspace $\mathcal{H}_N \subset \mathcal{H}$. Note that $\mathcal{H}_N$ consists of those metrics obtained by the $GL(N+1, C)$ action.

The space $\mathcal{H}_N$ is a symmetric space. Its dual is the unitary group $U(N+1)$. Under the natural metric of symmetric spaces, the C^* action (as a 1-parameter family of metrics) is a geodesic ray in $\mathcal{H}_N$. It is interesting to consider the limit of these algebraic rays when one raises the dimension of ambient space P^N(we can raise the power k of $\mathcal{L}^k$ and do Kodaira embedding, then pull the ray back to the class $c_1(L)$ by dividing out the scalar k). First, it is easy to derive that all the embedding induces the same geometric geodesic ray.

Lemma 4.3. — *Different embedding of a test configuration into projective spaces induce the same geodesic ray provided the rays start at the same point.*

Proof. — By different embedding, one essentially raises the power k of $\mathcal{L}^k \to \mathcal{M} \to D$ first. Then, we use sections of $H^0(\mathcal{M}, \mathcal{L}^k)$ to embed $\mathcal{L}^k \to \mathcal{M}$ into $\mathcal{O}(1) \to P^N \times D$. The Fubini-Study metric naturally induces a metric on $\mathcal{L}^k$, which has curvature in class $kc_1(\mathcal{L})$. To get a geodesic ray in the Kähler class $c_1(L)$, one takes the k-th root of the Fubini metric on $\mathcal{L}^k$ to get a Hermitian metric h_k on $\mathcal{L}$. Notice that $\log \frac{h_k}{h_n}$ is the potential difference of the background metric Ω_k and Ω_n. When we solve the Monge-Ampère equation, by uniqueness of the solution, the potential difference $\log \frac{h_k}{h_n}$ goes into the difference between the $C^{1,1}$ solutions ϕ_k and ϕ_n, such that the ray potential $\eta_k + i^*\phi_k = \eta_n + i^*\phi_n$. □

As $k \to \infty$, it is expected that these algebraic rays converge to some geometric geodesic rays. This is a natural extension of the classical problem: Use Bergman metrics to approximate a given Kähler metric. There is extensive literature on this topic, c.f. Tian [**33**], Zelditch [**36**], Lu [**20**], Phong-Sturm [**25**] and Song [**31**].

4.2. Bounded ambient geometry and test configuration. — In [**9**], the first named author introduced the notion of "bounded ambient geometry" to study geodesic rays. Briefly speaking, a geodesic ray is called to have bounded ambient geometry if the following holds: There exists a metric $\tilde{g}$ on $M \times S^1 \times [0, \infty)$ such that the ray has a $C^{1,1}$ relative potential under $\tilde{g}$, and $\tilde{g}$ has uniformly bounded curvature.

The geodesic ray induced by a smooth test configuration always has bounded ambient geometry. To see this, one restricts the metric $\Omega + idz \wedge d\bar{z}$ to the punctured part $\mathcal{M} - M_0$. Since $\Omega + idz \wedge d\bar{z}$ has bounded geometry on $\mathcal{M}$, the restriction clearly has bounded geometry. The punctured part is holomorphically identified with $M \times S^1 \times [0, \infty)$. Thus the ray has bounded ambient geometry. Actually, it is a stronger version of bounded ambient geometry since the metric $\tilde{g}$ on $M \times S^1 \times [0, \infty)$ can be compactified into a fibration over a disc. In general, this is not necessarily true.

In [**9**], it is proved that: Let $\rho(t)$ be a geodesic ray with bounded ambient geometry, then for any other potential ϕ_0, there is a unique relative $C^{1,1}$ geodesic ray starting from ϕ_0 and parallel to $\rho(t)$. Alternatively, we can use this to derive the existence of geodesic rays, based on the algebraic ray.

4.3. Futaki invariant, ¥ invariant and geodesic stability. — The classical definition of Futaki invariant is the following: Let M be a Kähler manifold with Kähler metric ω. Let X be a holomorphic vector field on M. Let h be the solution of $\Delta h = R - \underline{R}$. Futaki invariant is a linear functional: $\mathcal{F}(X) = \int_M X(h)\omega^n$. The definition is independent with the metric ω chosen in a fixed class. In particular, when $X = f^{,\alpha} \frac{\partial}{\partial w^\alpha}$, $\mathcal{F}(X) = \int_M f^{,\alpha} h_{,\alpha} \omega^n = \int_M f(\underline{R} - R)\omega^n$.

Ding and Tian [**11**] generalized the Futaki invariant to a class of singular varieties. Briefly speaking, they embed the variety into a projective space P^N, and consider the restriction of ambient holomorphic vector fields tangent to the variety on regular points. Also they consider the restriction of ambient Fubini-study metric ω and define Futaki invariant in similar fashion.

In test configuration, Donaldson's algebraic definition of Futaki invariant is: Let $\mathcal{L} \to \mathcal{M} \to D$ be a test configuration. Consider the C^* action on the central fiber $L_0 \to M_0$, and its powers $L_0^k \to M_0$. Let $d_k = \dim H_k = \dim H^0(M_0; L_0^k)$ and w_k be the weight of the C^* action on highest exterior power of H_k. Then $F(k) = w_k/kd_k$ has an expansion

$$F(k) = F_0 + F_1 k^{-1} + F_2 k^{-2} + \ldots\ldots \tag{14}$$

The coefficient F_1 is called the Futaki invariant of the C^* action on (L_0, M_0). He proved that if the central fiber is smooth, then the algebraic Futaki invariant agrees with the classical Futaki invariant.

Using Futaki invariant, Donaldson defined stability. A pair (L, M) is K-stable if: For each test configuration for (L, M) (i.e, $(L_1, M_1) = (L, M)$), the Futaki invariant of the C^* action on (L_0, M_0) is less than or equal to zero, and the equality only occurs when the configuration is a product configuration.

This algebraic definition agrees with an early geometric definition of K-stability by Ding and Tian. In [**11**], they used a C^* action of P^N to obtain the limit of the varieties M_t, then studied the Futaki invariant of the limiting variety M_0. The spirit is similar to Donaldson's setup of test configuration.

Notice that in test configuration, the stability is to check the Futaki invariant of the central fiber. However, one would like to have some criterion that doesn't need a specific central fiber. Just as the bounded ambient geometry only concerns behavior before reaching the limit, the ¥ invariant is a nice notion parallel to Futaki invariant and doesn't need a specific central fiber.

Definition 4.4. — [**9**] For a smooth geodesic ray $\rho(t)$, ¥ invariant is defined to be

$$¥ = \lim_{t\to\infty} \frac{dE}{dt} = \lim_{t\to\infty} \int \frac{\partial \rho}{\partial t}(\underline{R} - R)\omega_\rho^n. \tag{15}$$

The K-engery is convex along geodesics. So $\frac{dE}{dt}$ is monotone and the limit exists (either it is positive ∞ or a finite number).

The first named author defined the notion of geodesic stability by ¥ invariant: M is weakly geodesically stable if every geodesic ray has nonnegative ¥ invariant. M is geodesically stable if every geodesic ray has positive ¥ invariant. Conceptually, this is parallel to K-stability for test configurations. However, geodesic rays represent all possible geometrical degenerations. Therefore, it would not be a total surprise if geodesic rays detect some instabilities that test configuration method can't detect.

To clarify this analogy further, we prove the following.

Theorem 4.5. — *For simple test configuration, if the induced geodesic ray is smooth regular, then ¥ invariant agrees with Futaki invariant*[8].

Proof. — By definition of simple test configuration, the central fiber is smooth. Following [**15**], the algebraic Futaki invariant is exactly the classical Futaki-invariant applying to holomorphic vector field (induced by the C^* action) in the central fiber.

Denote the associated HCMA on $\mathcal{M}$ by $(\Omega + i\partial\bar{\partial}\phi)^{n+1} = 0$, ϕ is the smooth regular solution. Let ω_c be the restriction of $\Omega + i\partial\bar{\partial}\phi$ on M_0. The S^1 action of the C^* action

(8) It is the same up to a sign.

is a Hamiltonian action on M_0. Let f be the hamiltonian potential. In another word, $df = i_v\omega_c$, where v is the S^1 action vector field. The Futaki-invariant of the C^* action is

$$\nu = \int f(\underline{R} - R)\omega_c^n. \tag{16}$$

Now we look at $¥ = \lim_{t\to\infty} \int \frac{\partial\rho}{\partial t}(\underline{R} - R)\omega_\rho^n$.

The C^* action induces a diffeomorphism $i : \mathcal{M} - M_0 \to M \times [0,\infty) \times S^1$. Identify $M \times [0,\infty) \times S^1$ with $\mathcal{M} - M_0$ this way, then

$$\lim_{t\to\infty} i^*\omega_\rho = \omega_c, \qquad \lim_{t\to\infty} i^*R_\rho = R_{\omega_c}. \tag{17}$$

So it suffices to show,

$$\lim_{t\to\infty} i^*\frac{\partial\rho}{\partial t} = -f + \text{const}. \tag{18}$$

The assumption ϕ is smooth regular means $(\Omega + i\partial\bar\partial\phi)|_{M_\tau} > 0$ for all fiber $M_\tau \subset \mathcal{M}$. So it induces a smooth foliation F by holomorphic discs on $\mathcal{M}$.[9] Translate into $M \times [0,\infty) \times S^1$, i_*F is a foliation by holomorphic punctured discs. i_*F in turns induces an S^1 action on $M \times [0,\infty) \times S^1$, which is moving along the leaf of i_*F in S^1 direction. By identifying the fiber M_t with $M_{t\theta}$ ($|\theta| = 1$) trivially in $M \times [0,\infty) \times S^1$, the S^1 action is Hamiltonian action with hamiltonian $\frac{\partial\rho}{\partial t}$, under the symplectic form ω_ρ. In $M \times [0,\infty) \times S^1$, notice that the identification between M_t and $M_{t\theta}$ preserves the symplectic form since $\omega_\rho|_{M_t} = \omega_\rho|_{M_{t\theta}}$ for $|\theta| = 1$.

Translate this into the context of $\mathcal{M}$, we have: If we identify the fiber M_t with $M_{t\theta}$ in $\mathcal{M}$, via the S^1 action of the C^* action, then the S^1 action induced by foliation F(on $\mathcal{M}$) is hamiltonian action with hamiltonian $i^*\frac{\partial\rho}{\partial t}$, under symplectic form $i^*\omega_\rho$. Now we take limit towards the central fiber. Under this limit, the central fiber M_0 should be identified with itself via the S^1 rotation of the C^* action. Also, originally, the S^1 action induced by F is trivial on M_0. But, under the identification (which is distorted in M_0), the limit S^1 action should be the reverse of S^1 action of the C^* action on central fiber.

At last, we can take the limit of $i^*\frac{\partial\rho}{\partial t}$. Because the leaf vector on $M \times [0,\infty) \times S^1$ is $\frac{\partial}{\partial t} - g_\rho^{\alpha\bar\beta}(\frac{\partial\rho}{\partial t})_{\bar\beta}\frac{\partial}{\partial z^\alpha}$ and

$$\frac{\partial^2\rho}{\partial t^2} - g_\rho^{\alpha\bar\beta}\left(\frac{\partial\rho}{\partial t}\right)_{\bar\beta}\left(\frac{\partial\rho}{\partial t}\right)_\alpha = 0 \tag{19}$$

So the $\frac{\partial\rho}{\partial t}$ is constant along leaves. Therefore, when passing into $\mathcal{M}$, the $i^*\frac{\partial\rho}{\partial t}$ is constant along leaves of F. But F is foliation of discs and well defined on the central fiber, so $i^*\frac{\partial\rho}{\partial t}$ converges smoothly as moving towards the central fiber in $\mathcal{M}$. The limit

(9) See 5.2 for foliation induced by smooth regular solution of HCMA.

of the hamiltonian $i^* \frac{\partial \rho}{\partial t}$ is the hamiltonian of the limiting action. So $\lim_{t\to\infty} i^* \frac{\partial \rho}{\partial t} = -f + \text{const.}$ and the theorem is proved. □

5. Monge-Ampère equation on Simple test configurations

Following Donaldson's idea [**14**], we want to extend the correspondence in [**14**] to the case of Monge-Ampère equation on simple test configurations.

But to explain the background and the motive, we start with a review on Donaldson's result. M is a Kähler manifold with a given Kähler form ω. We solve the equation $(\pi^*\omega + i\partial\bar{\partial}\phi)^{n+1} = 0$ on $M \times D$ with boundary condition $\phi = \phi_0$ on $M \times \partial D$. π is the natural projection to M.

Donaldson and Semmes independently constructed the following manifold $W \to M$. W is glued by local holomorphic cotangent bundle over M. There exists a lifting of M into W for every Kähler metric $\omega + i\partial\bar{\partial}\phi$. If one take the lifting of $M \times D$ into $W \times D$ by the solution $\omega + i\partial\bar{\partial}\phi$, then one will obtain a family of holomorphic discs. These discs are the lifting of the foliation induced by the degenerated form $\pi^*\omega + i\partial\bar{\partial}\phi$. Conversely, if one has the family, then it can induce a solution to Monge-Ampère equation. This correspondence is powerful. It relates the regularity of a solution of HCMA equation to the regularity of moduli space of holomorphic discs in the sense of Fredholm theory.

The construction of Donaldson and Semmes works for a product manifold like $M \times D$. However, a test configuration of real interest is not a product space. So the previous construction would not work here directly. We solve this problem by taking a new point of view on the old construction: View $W \times D$ as a global construction over $M \times D$. Then we can derive an analogy in non-product case. This viewpoint might potentially be generalized to other cases.

5.1. Construction of $\mathcal{W} \to \mathcal{M}$. — Recall a test configuration is simple (Defi. 2.3) if: The total space $\mathcal{M}$ is smooth ($\mathcal{M}$ is a smooth sub-manifold of $P^N \times D$) and the projection $\pi : \mathcal{M} \to D$ is submersion everywhere.

From the definition, any simple test configuration is a fibration over the disc. Each fiber is smooth because $\pi : \mathcal{M} \to D$ is submersion everywhere.

Let $\mathcal{M}$ be a simple test configuration. We solve $(\Omega + i\partial\bar{\partial}\phi)^{n+1} = 0$ on $\mathcal{M}$. Since $\pi : \mathcal{M} \to D$ is submersion everywhere, so $\mathcal{M}$ is locally product space. To see this explicitly in the complex coordinates: First, choose a complex coordinate $\{x_0, ..., x_n\}$ for $U \subset \mathcal{M}$. The projection $z = z(x_0, ..., x_n)$ is holomorphic and $\frac{\partial z}{\partial x_i} \neq 0$ by assumption of submersion. Now one can easily cook up a tuple $\{z, x_{i_1}, ..., x_{i_n}\}$ such that the transition between $\{z, x_{i_1}, ..., x_{i_n}\}$ and $\{x_0, ..., x_n\}$ is non-degenerate. $\{z, x_{i_1}, ..., x_{i_n}\}$ is the product holomorphic coordinate we are looking for.

In the future, such product coordinate is denoted by (z, w) with $z \in D$ and $w \in M_z$. Cover $\mathcal{M}$ with local product charts U_i. On U_i, suppose that the $\Omega = i\partial\bar{\partial}\rho_i$. Write $T^*\mathcal{M}/T^*C$ over U_α by local coordinates (z, w, q). We glue these charts together, and define the transition between (z, w, q) over U_α and (v, x, p) over U_β:

$$\begin{aligned} z &= v, \\ w &= w(v, x) \text{ as defined in} \mathcal{M}, \\ q_j &= p_i \frac{\partial x_i}{\partial w_j} + \frac{\partial(\rho_\beta - \rho_\alpha)}{\partial w_j}. \end{aligned} \tag{20}$$

One can verify these local charts (z, w, q) glue up to a complex manifold $\mathcal{W} \to \mathcal{M}$. Define a form Θ on each fiber of $\mathcal{W} \to D$,

$$\Theta|_{W_t} = dq_i \wedge dw_i. \tag{21}$$

Here Θ is well defined only on the fiber, so $\Theta|_{W_t}$ is a family of forms.

The real part of Θ is a symplectic form on W_t. So W_t is a symplectic manifold and we can talk about Lagrangian sub-manifolds of W_t.

Definition 5.1. — For a Lagrangian sub-manifold L_t, L_t is called LS-submanifold if $\Theta|_{L_t}$ is non-degenerate. L_t is called LS-graph if it is LS-submanifold and also be a graph over M_t.

By straightforward calculation, one can see: Locally, LS-graphs are of forms $\partial\phi$ for some real potential ϕ on M_t, and $\Theta|_{L_t} = \partial\bar{\partial}\phi$. Our main result in this Section is:

Theorem 5.2. — *Let $\mathcal{M}$ be a simple test configuration. There is an associated manifold $\mathcal{W} \to \mathcal{M}$ such that:*

1. *A smooth solution ϕ of $(\Omega + i\partial\bar{\partial}\phi)^{n+1} = 0, \phi = \phi_0$ on $\partial\mathcal{M}$ induces a family of holomorphic discs $G : M \times D \to \mathcal{M} \to \mathcal{W}$ factoring through the foliation on $\mathcal{M}$, such that the image of $G(\cdot, z)$ is a LS-graph in $W_z \to M_z$ for all z and $\bigcup_{z \in \partial D} G(\cdot, z)$ is a totally real sub-manifold of $\mathcal{W}$.*
2. *If a family of holomorphic discs $G : M \times D \to \mathcal{W}$ respects the projection $\mathcal{W} \to D$, i.e, $\pi \circ G : M \times D \to D$ is a projection to D. Also assume it satisfies the boundary condition $G(\cdot, z) = \Lambda_{z,\phi_0}$ for $z \in \partial D$, where Λ_{z,ϕ_0} is the lifting of M_z by metric $\Omega + i\partial\bar{\partial}\phi_0$, then the image of $G(\cdot, z)$ is a LS-submanifold in W_z for all z. Moreover, if assuming these images are LS-graphs, then the family projects to a foliation of $\mathcal{M}$ and induces a smooth solution ϕ to $(\Omega + i\partial\bar{\partial}\phi)^{n+1} = 0$ with $\phi = \phi_0$ on $\partial\mathcal{M}$.*

Following Donaldson [**14**], we prove this theorem by discussion from both side of this correspondence in next two subsections.

5.2. One side of the Correspondence. — Now suppose there is a smooth solution ϕ for $(\Omega + i\partial\bar{\partial}\phi)^{n+1} = 0$ on $\mathcal{M}, \phi = \phi_0$ on $\partial\mathcal{M}$, with $\Omega + i\partial\bar{\partial}\phi$ positive on M_t.

In local product coordinates (z, w) of $\mathcal{M}$, write $\Omega + i\partial\bar{\partial}\phi = i\partial\bar{\partial}f$. Since $\Omega + i\partial\bar{\partial}\phi$ has rank n, it has a 1-complex dimension kernel. Let $X = \frac{\partial}{\partial z} + \eta^\alpha \frac{\partial}{\partial w^\alpha}$ be in kernel of $i\partial\bar{\partial}f$, then

$$\partial\bar{\partial}f\left(\frac{\partial}{\partial z} + \eta^\alpha \frac{\partial}{\partial w^\alpha}\right) = (\eta^\alpha f_{\alpha\bar{\beta}} + f_{z\bar{\beta}})\, dw^{\bar{\beta}} + (\eta^\alpha f_{\alpha\bar{z}} + f_{z\bar{z}})\, d\bar{z} \tag{22}$$
$$= 0.$$

Thus,

$$\eta^\alpha = -f_{z\bar{\beta}} f^{\alpha\bar{\beta}}, \tag{23}$$

$$f_{z\bar{z}} = -\eta^\alpha f_{\alpha\bar{z}}. \tag{24}$$

A direct calculation shows

$$[X, \bar{X}] = \left(\frac{\partial \eta^{\bar{\beta}}}{\partial z} + \eta^\alpha \frac{\partial \eta^{\bar{\beta}}}{\partial w^\alpha}\right) \frac{\partial}{\partial w^{\bar{\beta}}} - \left(\frac{\partial \eta^\alpha}{\partial \bar{z}} + \eta^{\bar{\beta}} \frac{\partial \eta^\alpha}{\partial w^{\bar{\beta}}}\right) \frac{\partial}{\partial w^\alpha} \tag{25}$$
$$= 0.$$

This means that the kernel distribution is holomorphically parametrized by $z \in D$. Therefore a smooth regular solution implies a foliation of $\mathcal{M}$ by holomorphic discs.

The $\mathcal{M}$ can be lifted to a graph in $\mathcal{W}$, using the form $\Omega + i\partial\bar{\partial}\phi$. On local product charts U_i, $\Omega = i\partial\bar{\partial}\rho_i$, we can lift $\mathcal{M}$ to graph $\partial(\rho_i + \phi)$ in each fiber. The lift is well defined globally due to the way we glue $\mathcal{W}$.

In [**14**], Donaldson showed in the lifting of $\mathcal{M}$, the foliation is lifted up to a family of holomorphic discs in $\mathcal{W}$. More importantly, these holomorphic discs take boundary value in a totally real sub-manifold Λ_{ϕ_0}. The same technique can be extended to our case.

Theorem 5.3. — *For a simple test configuration, the smooth solution of the HCMA equation induces a foliation of holomorphic discs on $\mathcal{M}$ which can be lifted up to a family of holomorphic discs with in $\mathcal{W}$. These discs have boundary in a totally real sub-manifold.*

Proof. — As above. □

5.3. The other side of the correspondence. — It is reasonable to consider the reverse correspondence locally. We have the following theorem:

Theorem 5.4. — *Suppose $G : D \times U \to \mathcal{W}$ is a smooth map which respects the projection and holomorphic in D. Assume for all $\tau \in \partial D$, U is mapped to be LS-graph* *and this LS-graph family has a global potential ϕ_0. Then for each $\tau \in D$, G maps U to an immersed LS-submanifold in $\mathcal{W}$. Moreover, if assuming these LS-submanifolds are LS-graphs*[10]*, then this family induces a smooth solution to the Monge-Ampère equation with boundary condition $\phi = \phi_0$.*

In above theorem, U is an open set of real dimension $2n$. $G : D \times U \to \mathcal{W}$ is smooth and respects the projection. In another word, for $\pi : \mathcal{W} \to D$, $\pi \circ G$ is identity on D. G is holomorphic in D variable. For each $\tau \in \partial D$, U is mapped to be a LS graph over M_τ and this LS-graph family have a global potential ϕ_0. This just means these LS-graphs are lifting of $\mathcal{M}$ using $\Omega + i\partial\bar{\partial}\phi_0$ on the boundary.

Proof. — Consider $G^*\Theta$ on $D \times U$. Θ is well defined on fibers W_t, so $G^*\Theta$ is well defined on fibers U_t in $D \times U$. We should view $G^*\Theta$ as a family of forms on U_t. Denote real coordinates on U by q_j, write $G^*\Theta = (r_{jk} + i s_{jk}) dq_j \wedge dq_k$. It is straightforward to show $r_{jk} + i s_{jk}$ is holomorphic function over D: Let (z, q) be coordinates on $D \times U$. Let (v, x, p) be a local coordinates in $\mathcal{W}$. The map G is $v = z, x = x(z, q), p = p(z, q)$. G is holomorphic, so $\frac{\partial x}{\partial \bar{z}} = 0$, $\frac{\partial p}{\partial \bar{z}} = 0$. Now $\Theta|_{W_t} = dp_i \wedge dx_i, G^*\Theta|_{U_t} = \frac{\partial p_i}{\partial q_j} \frac{\partial x_i}{\partial q_k} dq_j \wedge dq_k$, therefore $\frac{\partial}{\partial \bar{z}}(r_{jk} + i s_{jk}) = \frac{\partial}{\partial \bar{z}} \frac{\partial p_i}{\partial q_j} \frac{\partial x_i}{\partial q_k} = 0$.

On the boundary $\tau \in \partial D$, G maps U to LS-graphs. But Θ is purely imaginary on LS-graphs. Thus, $G^*\Theta$ is also purely imaginary. A holomorphic function on the disc with pure imaginary value on ∂D must be constant, so $r_{ij} + i s_{ij}$ must be constant on every disc in $D \times U$. This also implies the Jacobi of the map $G(\tau, \cdot) : U \to W_\tau$ is non-degenerate, since the pull back image $G^*\Theta$ is non-degenerate. It follows that the image $G(\tau, U)$ is an immersed LS-submanifold.

Now assume $G(\tau, U)$ is actually a LS-graph, i.e, the projection $\pi \circ G(\tau, \cdot)$ is diffeomorphism. Following [**10**], we find a global potential for this family of LS-graphs (modulo the local potential of the background metric).

First, consider the case when U is a very small open ball. Let D_α be a small open set in D. Without loss of generality, G maps $D_\alpha \times U$ into a single chart in $\mathcal{W}$. Since they are LS-graphs, one can solve a real potential φ_α for this family in the local product chart by $\frac{\partial \varphi_\alpha}{\partial x_i} = p_i$. φ_α is unique up to a smooth function in $z \in D$.

Choose a finite covering $D_\alpha \subset D$, and make U so small such that $D_\alpha \times U$ all fit in single charts in $\mathcal{W}$. This can be done if one fixes a finite chart covering of $\mathcal{W} \to D$ in first place and then replace U by small subset if necessary. Solve the potential φ_α respectively in each $D_\alpha \times U$, and the geometry of $\mathcal{W}$ implies $\partial(\varphi_\alpha - \rho_\alpha) = \partial(\varphi_\beta - \rho_\beta)$ on every fiber M_t of $\mathcal{M}$. So on each fiber, the difference $(\varphi_\alpha - \rho_\alpha) - (\varphi_\beta - \rho_\beta)$ must be constant. It follows that $\varphi_\alpha - \rho_\alpha$ differ with $\varphi_\beta - \rho_\beta$ by a smooth real function of

(10) Thanks to Song Sun, we noticed that the interior LS graphs are exact because the boundary LS graphs are exact. For definition of exact LS graphs, cf, Donaldson [**14**].

z on intersection. The fact $H^1(D,\mathcal{S})=0$, ($\mathcal{S}$ is the sheaf of C^∞ functions) implies one can adjust φ_α by function of z such that $\varphi_\alpha-\rho_\alpha=\varphi_\beta-\rho_\beta$. Therefore they give the global potential $\phi=\varphi_\alpha-\rho_\alpha$. ϕ is unique up to a function of z on D.

The next step is to make ϕ satisfy the boundary condition $\phi=\phi_0$. Let $X=\frac{\partial}{\partial z}+\eta^\alpha\frac{\partial}{\partial w^\alpha}$ be the tangential vector of the foliation $\pi\circ G: D\times U\to\mathcal{M}$. There exists a 1-1 form Ω' on $\mathcal{M}$ such that $i_X\Omega'=0$ and its restriction to M_t is $i\partial\bar\partial\varphi_\alpha|_{M_t}=\Theta|_{L_t}$. Locally, $\Omega'=i(\frac{\partial^2\varphi}{\partial w_\alpha w_{\bar\beta}}dw_\alpha dw_{\bar\beta}+\zeta^\alpha dw_\alpha d\bar z+\zeta^{\bar\beta}dw_{\bar\beta}dz+hdzd\bar z)$, where $\zeta^\alpha=-\eta^{\bar\beta}\varphi_{\alpha\bar\beta}$ and $h=\eta^\alpha\eta^{\bar\beta}\varphi_{\alpha\bar\beta}$.

Let (v,q) be coordinates on $D\times U$, q as real coordinates. (z,w) are local coordinates on $\mathcal{M}$. We have $\eta^{\bar\beta}=\frac{\partial w^{\bar\beta}}{\partial\bar v}$. Let ρ be local potential for background metric Ω, and $\varphi=\rho+\phi$. The disc family in $\mathcal{W}$ is holomorphic implies $\frac{\partial}{\partial\bar v}\frac{\partial\varphi}{\partial w_\alpha}=0$, therefore

$$0=\frac{\partial}{\partial\bar v}\frac{\partial\varphi}{\partial w_\alpha}=\frac{\partial^2\varphi}{\partial w_\alpha\partial\bar z}+\frac{\partial^2\varphi}{\partial w_\alpha\partial w_{\bar\beta}}\eta^{\bar\beta}\tag{26}$$

So $\zeta^\alpha=\frac{\partial^2\varphi}{\partial w_\alpha\partial\bar z}$, $\Omega'=i(\partial\bar\partial\varphi+(h-\varphi_{z\bar z})dzd\bar z)=i(\partial\bar\partial(\rho+\phi)+(h-\rho_{z\bar z}-\phi_{z\bar z})dzd\bar z)=\Omega+i\partial\bar\partial\phi+i(h-\rho_{z\bar z}-\phi_{z\bar z})dzd\bar z$.

On the other hand, Ω' is a closed form. To see this: Let $i: M_t\to\mathcal{M}$ be the embedding of fibers, then $i^*d\Omega'=d(i^*\Omega')=0$. It suffices to show $i_Xd\Omega'=0$ since the restriction of $d\Omega'$ to the fiber is zero already. Now we show $i_Xd\Omega'=L_X\Omega'-di_X\Omega'=L_X\Omega'=0$. Notice that Ω' is determined by $\Theta|_{L_t}$ and the condition $i_X\Omega'=0$. If we can show $\Theta|_{L_t}$ and X are preserved by X-flow, then immediately we obtain $L_X\Omega'=0$ by uniqueness. The fact $\Theta|_{L_t}$ is preserved follows $G^*\Theta$ is constant along leaves and the fact X is preserved follows $[X,\bar X]=0$. So Ω' is closed form on $\mathcal{M}$, and $i(h-\rho_{z\bar z}-\phi_{z\bar z})dzd\bar z=\Omega'-\Omega-i\partial\bar\partial\phi$ is closed. This implies $(h-\rho_{z\bar z}-\phi_{z\bar z})$ is just a function of z. Also, since Ω' and Ω and ϕ are globally defined, so $(h-\rho_{z\bar z}-\phi_{z\bar z})dzd\bar z$ is defined globally and doesn't depend on the local representation. Therefore, the function $h-\rho_{z\bar z}-\phi_{z\bar z}$ is globally defined, since $dzd\bar z$ is defined on the whole disc. (Notice that the z stands for a coordinate in a local product chart, so in different product charts, $\phi_{z\bar z}$ is not the same though the function ϕ is the same.)

Now let $H=h-\rho_{z\bar z}-\phi_{z\bar z}$. H is defined globally on $\pi\circ G(D\times U)$, but solely depends on $z\in D$. One can solve the following equation on disc:

$$\partial_{z\bar z}\phi'=H\tag{27}$$

with $\phi'=\phi_0-\phi$ on the ∂D. Now replace ϕ by $\phi+\phi'$, then one get $\Omega'=\Omega+i\partial\bar\partial\phi$ and $\phi=\phi_0$ on ∂D. (Note that in different local charts, (z,w) and (v,x) in $\mathcal{M}$, where z,v project down to the same disc variable. $\partial_{z\bar z}\phi'=\partial_{v\bar v}\phi'$ since ϕ' is constant fiber-wise.) This finishes the proof of finding potential ϕ if U is sufficiently small.

Now for arbitrary U, one can always decompose it into small open balls U_i which admit potential ϕ_i. Let ρ be a local potential for the Ω on $\mathcal{M}$. Then on the leaf, we have

Lemma 5.5. — *We have* $\Delta(\rho+\phi_i) = X\bar{X}(\rho+\phi_i) = 0$.

Proof. — Let $f = \rho + \phi_i$,

$$X\bar{X}f = X(\eta^{\bar{\beta}})f_{\bar{\beta}} + \partial\bar{\partial}f(X,\bar{X}) \tag{28}$$
$$= 0. \qquad \square$$

This implies $\Delta(\phi_i - \phi_j) = 0$ on the leaf. Now with the extra condition $\phi_i = \phi_j = \phi_0$ on the ∂D, it implies $\phi_i = \phi_j$ on the intersection. The global potential is immediately obtained from this. $\square$

Remark 5.6. — The above correspondence is constructed only on simple test configurations. In these configurations, central fiber are smooth. However, we believe the techniques should work for some mild singularities in the central fiber.

Another point is that the correspondence has nothing to do with the C^* action.

6. Openness of super regular solution

In simple test configurations, we can study regularity of the solution ϕ by the associated holomorphic disc family in $\mathcal{W} \to \mathcal{M}$.(11) Donaldson's definition [**14**] of super regular discs and the linearized model could be extended to our case as well. In detail,

Definition 6.1. — In the moduli map $G : D \times U \to \mathcal{W}$, a disc $G(D,x)$ is called super regular at $z \in D$ if $d(\pi \circ G_z)_x : TU \to TM$ is isomorphism. A disc $G(D,x)$ is called super regular if it is super regular at every $z \in D$.

Definition 6.2. — A geodesic ray induced from a simple test configuration is called super regular if the disc family in $\mathcal{W}$ is super regular.(12).

For a disc $G_x = G(\cdot,x)$ in the moduli map $G : D \times U \to \mathcal{W}$, one can consider the holomorphic perturbation of G_x that satisfies the totally real boundary condition (the boundary is in the Λ_ϕ, i.e., the lifting of $M_t, t \in \partial D$ by $\Omega + i\partial\bar{\partial}\phi$). Also, we normalize the perturbation such that it preserves the projection property. In another word, $\pi \circ G : D \times U \to D$ is identity on D variable. Following Donaldson [**14**], the linearized problem is

(11) However, the existence so far only requires smoothness of total space.

(12) I.e.: the solution is smooth regular to the Monge-Ampère equation on the test configuration $\mathcal{M}$.

Theorem 6.3. — *In the moduli map $G : D \times U \to \mathcal{W}$ corresponding to a smooth solution ϕ, the linearized perturbation equation for a disc $G(\cdot, x)$ is*

$$v = Su + A\bar{u} \quad \textit{on } \partial D, \tag{29}$$

$$\bar{\partial} u = 0, \tag{30}$$

$$\bar{\partial} v = 0, \tag{31}$$

where S and A are maps from ∂D to complex symmetric matrices and positive hermitian matrices respectively; while u, v are C^n valued functions on D.

Proof. — The idea is the same to Donaldson [**14**]: Trivialize the exact sequence $0 \to (\pi \circ G_x)^*(T^*\mathcal{M}) \to G_x^*(T\mathcal{W}) \to (\pi \circ G_x)^*(T\mathcal{M}) \to 0$. □

In [**14**], it is showed that the problem is Fredholm and the index is $2n$. Consequently, if the disc is regular in Fredholm sense, then $G : D \times U \to \mathcal{W}$ is indeed an open set in the universal moduli space.

Regarding on the criterion of regularity for a disc, a mild modification of Donaldson's argument leads to the following:

Theorem 6.4. — *If a disc is super regular at any point $p \in \partial D$, then the disc is regular.*

Proof. — We look at the linearized model since the general case can be reduced to this simple model.

First, define $\Omega(s_1, s_2) = u_1^t v_2 - u_2^t v_1$. This is a symplectic form for $s = (u, v) \in C^{2n}$. In particular, for $s_1, s_2 \in \ker \bar{\partial}_{S,A}$, $i\Omega(s_1(\tau), s_2(\tau))$ is real and independent of τ. To see this, just notice that $i\Omega(s_1, s_2)$ is holomorphic function and on ∂D, $i\Omega(s_1, s_2) = i[u_1^t(Su_2 + A\bar{u}_2) - u_2^t(Su_1 + A\bar{u}_1)] = i(u_1^t A\bar{u}_2 - u_2^t A\bar{u}_1)$ is real.

The super regularity at $p \in \partial D$ means there are $2n$ elements $s_j = (u_j, v_j) \in \ker \bar{\partial}_{S,A}$ such that $u_j(p)$ form a R-basis for C^n. By continuity, it implies $u_j(\tau)$ form a R-basis for C^n in a neighborhood $\tau \in U_p$.

We claim $s_i(\tau)$ are generically C-linearly independent. It is equivalent to claim $\det[s_j]_{1 \le j \le 2n}$ has discrete zeros. Notice det is holomorphic, so the zeros are either discrete or the whole disc. Suppose it is the whole disc for contradiction. In the neighborhood U_p, assume the maximal rank of $[s_j]_{1 \le j \le 2n}$ for $\tau \in U_p$ is achieved at p without loss of generality, and the rank is $k < 2n$. Assume $s_1, s_2, \ldots, s_k$ form a basis for $span\{s_i\}$ at p, then near p, $s_{k+1} = \sum \lambda_i s_i, 1 \le i \le k$. λ_i is holomorphic, since it satisfies $\sum \lambda_i s_i^t s_j = s_{k+1}^t s_j, 1 \le i, j \le k$. In another word, it is obtained by solving the holomorphic matrix equation $\lambda[s_i^t s_j] = s_{k+1}^t s_j$. Now one finds holomorphic functions $\lambda_1, \ldots, \lambda_k, \lambda_{k+1} = -1, \lambda_{k+2} = 0, \ldots, \lambda_{2n} = 0$ near p, such that $\sum \lambda_i s_i = 0$. On the boundary ∂D near p,

$$0 = \sum \lambda_j v_j = S(\sum \lambda_j u_j) + A(\sum \lambda_j \bar{u}_j) = A(\sum \lambda_j \bar{u}_j). \tag{32}$$

So $\sum \lambda_j \bar{u}_j = 0$ and we also have $\sum \lambda_j u_j = 0$, so

$$\sum \Im(\lambda_j) u_j = 0 = \sum \Re(\lambda_j) u_j. \tag{33}$$

Since u_j form R-basis near p, one has $\lambda_j = 0$ on ∂D near p, which contradicts the choice of λ_j. Therefore, the $\det[s_j]_{1 \le j \le 2n}$ has discrete zero.

Now suppose the $\ker \bar{\partial}_{S,A}$ has dimension strictly greater than $2n$. Then one can choose s_0 not in $span\{s_i\}, 1 \le i \le 2n$. Now in the $2n+1$ dimensional vector space $span\{s_i\}$, $i\Omega$ as a skew form, must be singular. So there is a vector $s \in span\{s_0, ..., s_{2n}\}$ such that $i\Omega(s, span\{s_1, ..., s_{2n}\}) = 0$. Notice we proved $s_1, ..., s_{2n}$ form a C-basis generically, this implies $s = 0$ generically on D. Thus it implies $s = 0$, contradiction. □

In particular, since the holomorphic discs associated to smooth solution ϕ are automatically super regular, above theorem proves that they are all regular and the moduli space M in the map $G : D \times M \to \mathcal{W}$ is a compact connected component of the universal moduli space. It readily implies the following theorem.

Theorem 6.5. — *Openness: If the equation $(\Omega + i\partial\bar{\partial}\phi)^{n+1} = 0, \phi = \phi_0$ on $\partial\mathcal{M}$ admits a smooth solution ϕ with $\Omega + i\partial\bar{\partial}\phi > 0$ on fibers, then for any small perturbation $\delta\phi_0 \in C^\infty(\partial\mathcal{M})$, the new boundary value problem still has smooth solution ϕ' which is close to ϕ in $C^\infty(\mathcal{M})$ and $(\Omega + i\partial\bar{\partial}\phi') > 0$ on fibers.*

Proof. — We refer the proof to [**14**], which essentially asserts that compact families of regular normalized discs are stable under small perturbations. □

7. Geodesic ray from Toric degenerations

7.1. Basics of Toric degeneration. — For toric varieties, there has been extensive literature in extremal metrics. Abreu [**1**] initiated to study complex geometry on toric variety by symplectic coordinates. Afterwards, there has been much work in extremal metrics on toric variety, c.f. Donaldson [**15**], Zhou-Zhu [**37**], Gabor [**32**]. For completeness, we describe Donaldson's construction of Toric degenerations [**15**] in the following:

Let $P \subset R^n$ be a polytope associated to a toric variety M. For simplicity, let us assume P is Delzant. Let f be a rational piecewise linear function on M. One can associate it with a polytope $\hat{P} = \{(x, y) : x \in P, 0 \le y \le K - f\} \subset R^{n+1}$, $K = \max f$. By re-scaling $\hat{P}$, we can assume $\hat{P}$ is integral. In other words, all vertices of $\hat{P}$ are integral points.

It is a classical fact that $\hat{P}$ as above induces a toric variety $\mathcal{M}$ with a positive line bundle $\mathcal{L}$. Each integral point p in $\hat{P}$ corresponds to a section s_p of $\mathcal{L} \to \mathcal{M}$. The correspondence is compatible with addition of integral points in $\hat{M}$ and multiplication of sections in $\mathcal{L}$. In other words, if $p_1 + p_2 = p_3 + p_4$, then $s_{p_1}s_{p_2} = s_{p_3}s_{p_4}$.

One can view $\mathcal{M}$ as a sub-variety in P^N by Kodaira embedding: $x \in \mathcal{M}, x \to [s_1(x) : s_2(x) : ... : s_i(x)...]$. i runs through the integral points of $\hat{P}$. So $\mathcal{M} \subset P^N$ is defined by homogeneous equations $F(X_i) = 0$. These equations are induced by the relations of s_i, or equivalently, by the relations of the integral points in $\hat{P}$.

There is a map $\pi : \mathcal{M} \to P^1$, defined by $\pi : x \to [s_p(x) : s_q(x)]$. $p = (t_1, ..., t_n, t_{n+1}), q = (t_1, ..., t_n, t_{n+1} + 1) \in \hat{P}$. Also, there is a natural C^* action on $\mathcal{M}$ from the torus $T^{n+1} = T^n \times C^*$. It transforms section s_p to $t^k s_p$. k is the height of p. i.e, $p = (t_1, ..., t_n, k)$. So the C^* action can be lifted to $\pi : \mathcal{M} \to P^1$ by defining $t \circ [x : y] = [x : ty]$ on P^1.

The toric degeneration is just $\mathcal{M} - \pi^{-1}([1 : 0])$. The following example shows the construction in detail.

Example. — Let $P = [0, 2] \in R$ be the base polytope. $f = \max\{0, x - 1\}$ is the piece wise linear function on P. $\hat{P} = ([0, 1] \times [0, 1]) \bigcup \{1 \leq x \leq 2, x + y \leq 2\}$. Denote the integral points $X = (0, 0), Y = (1, 0), Z = (2, 0), U = (0, 1), V = (1, 1)$. Then the toric degenerations is the sub-variety in P^4 defined by

$$XZ = Y^2, XV = UY \tag{34}$$

The C^* action on $\mathcal{M}$ is $t : [X : Y : Z : U : V] \to [X : Y : Z : tU : tV]$. Notice that in order to get nontrivial test configuration, we only consider the part $\mathcal{M} - \pi^{-1}([1 : 0])$. In another word, we consider the asymptotic direction when $t \to \infty$ on C^*.

The central fiber is defined by $[Y : V] = [0 : 1]$. It is the toric variety associated to the segment $y = 1, x \in [0, 1]$ and $x \in [1, 2], x + y = 2$. Geometrically, the central fiber is the union of two P^1 which intersect at one point. Notice that the ambient space $\mathcal{M}$ is smooth here, so the induced geodesic ray has ambient bounded geometry automatically.

7.2. Explicit calculation of the $C^{1,1}$ geodesic ray. — We calculate the induced geodesic ray of previous example. The idea is to first calculate the geodesic segment connecting the fiber at $[1 : 1]$ to the fiber at $[1 : e^t], t \in R \times S^1$. Then, taking the limit of these segments when $t \to \infty$, we obtain a geodesic ray.

Equipped with the natural background metric of P^4, the fiber at $w = [1 : e^t] \in P^1$ has metric potential $\frac{1}{2}\log(|X|^2 + |Y|^2 + |Z|^2 + |U|^2 + |V|^2)$. Pulling this metric to the fixed fiber M at $w = [1 : 1] \in P^1$, the potential becomes

$$\frac{1}{2}\log(|X|^2 + |Y|^2 + |Z|^2 + e^{2t}|U|^2 + e^{2t}|V|^2). \tag{35}$$

Since the fiber M is at $[1:1]$, so $Y = V, X = U$. After proper normalization, the potential is

$$\frac{1}{2}\log(|X|^2 + |Y|^2 + (e^{2t}+1)^{-1}|Z|^2). \tag{36}$$

Now we calculate the geodesic segment connecting these two metrics.

Choose $[A, B]$ as standard P^1 coordinate on M. Thus, $X = B^2, Y = AB, Z = A^2$. Using $C^* = R \times S^1$ coordinate of P^1, $A = e^y, B = 1, y \in R \times S^1$. The Kähler potential is

$$h_{0,t} = \frac{1}{2}\log(1 + e^{2y} + e^{4y}(e^{2t}+1)^{-1}). \tag{37}$$

One can verify that the Legendre transform of $h_{0,t}$ maps R to $(0, 2)$ for each fixed t.

Notice that in polytope representation, the geodesic is just a straight line of convex functions. Now by straightforward calculation, one just computes the two end points associated to the two metrics in polytope representation and then take the linear interpolation. Passing to limit, one gets the $C^{1,1}$ ray in polytope representation

$$u_t = u_0 + t\max(0, x-1) \ \ , \ \ t \in [0, \infty). \tag{38}$$

In the standard picture of $M \times [0, \infty)$, we transform the u_t by Legendre transform and get the potential

$$h_t(y) = \begin{cases} h_0(y), & \text{when } y < \frac{\log 2}{4}, \\ h_0(\frac{\log 2}{4}) + y - \frac{\log 2}{4}, & \text{when } \frac{\log 2}{4} < y < \frac{\log 2}{4} + t, \\ h_0(y-t) + t, & \text{when } \frac{\log 2}{4} + t < y. \end{cases} \tag{39}$$

One can verify that $h_t - h_{0,t}$ is uniformly bounded. This confirms that the geometric ray is parallel to the algebraic ray.

It is natural to extend this observation to general toric degenerations.

Theorem 7.1. — *Let $\mathcal{M}$ be a toric degeneration with extremal piece wise linear function f. Suppose the ambient polytope $\hat{P}$ is integral and the base P is delzant. Then the induced geodesic ray is $u = u_0 + tf$ in polytope representation.*

Proof. — Similar to the previous set up, we calculate the geodesic segment connecting the fiber at $[1, 1]$ to the fiber at $[1, e^t]$. Then we pass the directions of these geodesic segments to the limit as $t \to \infty$.

Under the $(C^*)^n$ coordinates of $M = M_{[1:1]}$, the projective coordinates can be represented by $[\ldots : \exp(\sum_1^n X_i y_i) : \ldots]$. Let $(X_1, \ldots, X_n, X_{n+1}) = p$ be coordinates of those integral points in $\hat{P}$. Therefore, after proper normalization, the metric potential of the algebraic ray is:

$$h_{0,t} = \frac{1}{2}\log\left(\sum_{p\in\hat{P}} \exp 2(-Kt + X_{n+1}t + \sum_1^n X_i y_i)\right). \tag{40}$$

$K = \max f$, $p = (X_1, ..., X_{n+1}) \in \hat{P}$ are integral points.

Let $x = (x_1, ..., x_n) \in P$. Assume the extremal function $f = c_k x_k + d$ near x. i.e, we consider x in interior of a single definition domain of f. Under the Legendre transform of $h_{0,t}$, the pre-image $\tilde{y}$ of x satisfies

$$\frac{\sum_{p\in\hat{P}} X_j \exp 2(-Kt + X_{n+1}t + \sum_1^n X_i\tilde{y}_i)}{\sum_{p\in\hat{P}} \exp 2(-Kt + X_{n+1}t + \sum_1^n X_i\tilde{y}_i)} = x_j. \tag{41}$$

In particular, we denote the pre-image of x at time $t = 0$ by y.

By the Legendre transform, the potential u_t in polytope representation is:

$$u_t(x) = x\tilde{y} - h_{0,t}. \tag{42}$$

So, the limit direction is:

$$\lim \frac{u_t - u_0}{t} = \lim_{t\to\infty} \frac{\sum x_k(\tilde{y}_k - y_k) - \frac{1}{2}\log \frac{\sum_{p\in\hat{P}} \exp 2(-Kt+X_{n+1}t+\sum_1^n X_i\tilde{y}_i)}{\sum_{p\in\hat{P}} \exp 2(\sum_1^n X_i y_i)}}{t}. \tag{43}$$

If we can prove $\lim \frac{\tilde{y}_k - y_k}{t} = c_k$ and $\lim \frac{-\frac{1}{2}\log \frac{\sum_{p\in\hat{P}} \exp 2(-Kt+X_{n+1}t+\sum_1^n X_i\tilde{y}_i)}{\sum_{p\in\hat{P}} \exp 2(\sum_1^n X_i y_i)}}{t} = d$, then the theorem is proved. Now, we prove that the second is an implication of the first. Assuming $\lim \frac{\tilde{y}_k - y_k}{t} = c_k$, i.e, $\tilde{y}_k - y_k = c_k t + \epsilon_k t$ where $\epsilon_k \to 0$ as $t \to \infty$. We have the following

$$-Kt + X_{n+1}t + \sum_1^n X_i\tilde{y}_i = \left(-K + X_{n+1} + \sum_1^n c_i X_i + d\right) t \tag{44}$$

$$+ \left(-d + \sum_1^n \epsilon_i X_i\right) t + \sum_1^n X_i y_i. \tag{45}$$

For integral points p in the area where $f = c_k x_k + d$, the $L(p) = -K + X_{n+1} + \sum_1^n c_i X_i + d = X_{n+1} - h(X) \le 0$, where $h(X) = h(X_1, ..., X_n)$ is the height of $\hat{P}$ over the base point $(X_1, ..., X_n)$. i.e, $h(X) = K - f(X)$. For integral points p not in the area where $f = c_k x_k + d$, by definition of $f = \max(f_1, f_2, .., f_l)$ (f_l are linear functions), it is clear that $L(p) < -\delta$ for a fixed $\delta > 0$ Therefore,

$$\sum_{p\in\hat{P}} \exp 2\left(-Kt + X_{n+1}t + \sum_1^n X_i\tilde{y}_i\right) \tag{46}$$

$$\neq \exp(-2dt)\left(\sum_{p\in A} \exp 2(-\delta_p t)\exp\left(2\sum_1^n X_i y_i\right) + \sum_{p\in B} \exp\left(2\sum_1^n X_i y_i\right)\right) \tag{47}$$

B contains integral points p in $\hat{P}$ such that their projection $(X_1, ..., X_n)$ are in the area where $f = c_k x_k + d$ and $X_{n+1} = h(X)$. A contains the rest integral points in $\hat{P}$ but not in B. The condition $\hat{P}$ is integral guarantees that B is not empty.

Now we can calculate

$$\lim_{t\to\infty} \frac{-\frac{1}{2}\log \frac{\sum_{p\in\hat{P}} \exp 2(-Kt+X_{n+1}t+\sum_1^n X_i\tilde{y}_i)}{\sum_{p\in\hat{P}} \exp 2(\sum_1^n X_i y_i)}}{t} \tag{48}$$

$$= d + \lim_{t\to\infty} \frac{-\frac{1}{2}\log \frac{\sum_A \exp 2(-\delta_p t)\exp(2\sum_1^n X_i y_i)+\sum_B \exp(2\sum_1^n X_i y_i)}{\sum_{p\in\hat{P}} \exp 2(\sum_1^n X_i y_i)}}{t} \tag{49}$$

$$= d. \tag{50}$$

So it remains to prove $\lim_{t\to\infty} \frac{\tilde{y}_k - y_k}{t} = c_k$.

Let $y'_k = \tilde{y}_k - c_k t$, our purpose next is to prove y' is uniformly bounded for t sufficiently large. The equation 41 can be rewritten as:

$$x_k = \frac{\sum_B X_k \exp(2\sum X_i y'_i) + \sum_A X_k \exp(2\sum X_i y'_i)\exp(-\delta_p t)}{\sum_B \exp(2\sum X_i y'_i) + \sum_A \exp(2\sum X_i y'_i)\exp(-\delta_p t)}. \tag{51}$$

Define a map $\phi : y' \to x$ by $x_k = \frac{\sum_B X_k \exp(2\sum X_i y'_i)}{\sum_B \exp(2\sum X_i y'_i)}$. Let $P' \subset P$ be the polytope where $f = c_k x_k + d$. We need the following lemma:

***Lemma** 7.2*. — *$\phi : R^n \to P'$ is a diffeomorphism from R^n to the interior of P'*

Proof. — The lemma is a special case of a more general fact: Let $S = \{p_1, ..., p_m\}$ be a set of arbitrary points in R^n. $p_i = (X_1^i, ..., X_n^i)$. If the convex hull P spanned by S has dimension n, then the map:

$$\phi : (y_1, ..., y_n) \to (x_1, ..., x_n), x_k = \frac{\sum_S X_k^i \exp(2\sum X_j^i y_j)}{\sum_S \exp(2\sum X_j^i y_j)} \tag{52}$$

is a diffeomorphism from R^n to the interior of P.

Notice that B projects to be a grid G on P'. G contains all the vertices of P' due to the integral condition of $\hat{P}$. P' is convex since f is convex. So P' is the convex hull spanned by G. Therefore, the above fact applies exactly. □

Now, using this lemma, we can prove $\lim_{t\to\infty} \frac{\tilde{y}_k - y_k}{t} = c_k$: Choose a small closed ball $B_p \subset P'$ near $p = (x_1, ..., x_n)$. The pre-image $\phi^{-1}(B_p)$ is bounded closed set in R^n. Now consider the family of maps $\phi_t : y' \to x$ defined by equation 51. Notice that each ϕ_t is a diffeomorphism since equation 51 is just another form of equation 41, which defines the standard identification between R^n and P.

Since $\phi^{-1}(B_p)$ is bounded, it is straightforward from the equation 51 that: For any $\epsilon > 0$, there exists T such that $|\phi_t(y) - \phi(y)| < \epsilon$ for $y \in \phi^{-1}(B_p)$ and $t > T$. Thus the image $\phi_t(\phi^{-1}(B_p))$ is a ball close to B_p and contains p for t sufficiently large.

So above argument proves: For any B_p contains p and lies in interior of P', there exists $T > 0$ such that $y' = \phi_t^{-1}(p) \in \phi^{-1}(B_p)$ for $t > T$. Since $\phi^{-1}(B_p)$ is bounded,

$$\lim_{t\to\infty} \frac{\tilde{y}_k - y_k}{t} = c_k + \lim_{t\to\infty} \frac{y'_k - y_k}{t} = c_k. \tag{53}$$

□

These geodesic rays show some bad regularity. In general, they behave like the following: First, they break the manifold M into several pieces. As time evolves, they will tear these pieces apart, but keep metrics on each part. The space between the teared parts has degenerated metrics and zero volume. In particular, one can verify that the 2nd derivative of these rays are piece wise smooth function on fibers. At the broken points, these 2nd derivatives have jumps, so there is no global C^3 bound for the relative geodesic ray potential.

References

[1] M. ABREU – "Kähler geometry of toric varieties and extremal metrics", *Internat. J. Math.* **9** (1998), p. 641–651.

[2] V. APOSTOLOV, D. M. J. CALDERBANK, P. GAUDUCHON & C. W. TØNNESEN-FRIEDMAN – "Hamiltonian 2-forms in Kähler geometry. III. Extremal metrics and stability", *Invent. Math.* **173** (2008), p. 547–601.

[3] C. AREZZO & G. TIAN – "Infinite geodesic rays in the space of Kähler potentials", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* **2** (2003), p. 617–630.

[4] E. BEDFORD & B. A. TAYLOR – "The Dirichlet problem for a complex Monge-Ampère equation", *Invent. Math.* **37** (1976), p. 1–44.

[5] E. CALABI – "Extremal Kähler metrics", in *Seminar on Differential Geometry*, Ann. of Math. Stud., vol. 102, Princeton Univ. Press, 1982, p. 259–290.

[6] ______, "Extremal Kähler metrics. II", in *Differential geometry and complex analysis*, Springer, 1985, p. 95–114.

[7] E. CALABI & X. CHEN – "The space of Kähler metrics. II", *J. Differential Geom.* **61** (2002), p. 173–193.

[8] X. CHEN – "The space of Kähler metrics", *J. Differential Geom.* **56** (2000), p. 189–234.

[9] ______, "Space of Kähler metrics. III—On the lower bound of the Calabi energy and geodesic distance", *Invent. math.* (2008), DOI: 10.1007/s00222-008-0153-7.

[10] X. CHEN & G. TIAN – "Geometry of Kähler metrics and foliations by holomorphic discs", *Publ. Math. Inst. Hautes Études Sci.* **107** (2008), p. 1–107.

[11] W. Y. DING & G. TIAN – "Kähler-Einstein metrics and the generalized Futaki invariant", *Invent. Math.* **110** (1992), p. 315–335.

[12] S. K. DONALDSON – "Symmetric spaces, Kähler geometry and Hamiltonian dynamics", in *Northern California Symplectic Geometry Seminar*, Amer. Math. Soc. Transl. Ser. 2, vol. 196, Amer. Math. Soc., 1999, p. 13–33.

[13] ______, "Scalar curvature and projective embeddings. I", *J. Differential Geom.* **59** (2001), p. 479–522.

[14] ______, "Holomorphic discs and the complex Monge-Ampère equation", *J. Symplectic Geom.* **1** (2002), p. 171–196.

[15] ______, "Scalar curvature and stability of toric varieties", *J. Differential Geom.* **62** (2002), p. 289–349.

[16] ______, "Lower bounds on the Calabi functional", *J. Differential Geom.* **70** (2005), p. 453–472.

[17] A. FUTAKI – "An obstruction to the existence of Einstein Kähler metrics", *Invent. Math.* **73** (1983), p. 437–443.

[18] B. GUAN – "The Dirichlet problem for complex Monge-Ampère equations and regularity of the pluri-complex Green function", *Comm. Anal. Geom.* **6** (1998), p. 687–703.

[19] M. Levine – "A remark on extremal Kähler metrics", *J. Differential Geom.* **21** (1985), p. 73–77.

[20] Z. Lu & G. Tian – "The log term of the Szegő kernel", *Duke Math. J.* **125** (2004), p. 351–387.

[21] T. Mabuchi – "Some symplectic geometry on compact Kähler manifolds. I", *Osaka J. Math.* **24** (1987), p. 227–252.

[22] ———, "Stability of extremal Kähler manifolds", *Osaka J. Math.* **41** (2004), p. 563–582.

[23] S. T. Paul & G. Tian – "Analysis of geometric stability", *Int. Math. Res. Not.* **48** (2004), p. 2555–2591.

[24] D. H. Phong & J. Sturm – "Stability, energy functionals, and Kähler-Einstein metrics", *Comm. Anal. Geom.* **11** (2003), p. 565–597.

[25] ———, "The Monge-Ampère operator and geodesics in the space of Kähler potentials", *Invent. Math.* **166** (2006), p. 125–149.

[26] ———, "Test configurations for K-stability and geodesic rays", *J. Symplectic Geom.* **5** (2007), p. 221–247.

[27] ———, "On the regularity of geodesic rays associated to test configurations", preprint arXiv:0707.3956.

[28] J. Ross & R. Thomas – "A study of the Hilbert-Mumford criterion for the stability of projective varieties", *J. Algebraic Geom.* **16** (2007), p. 201–255.

[29] S. Semmes – "Complex Monge-Ampère and symplectic manifolds", *Amer. J. Math.* **114** (1992), p. 495–550.

[30] ———, "The homogeneous complex Monge-Ampère equation and the infinite-dimensional versions of classic symmetric spaces", in *The Gelfand Mathematical Seminars, 1993–1995*, Gelfand Math. Sem., Birkhäuser, 1996, p. 225–242.

[31] J. Song & S. Zelditch – "Bergman metrics and geodesics in the space of Kähler metrics on toric varieties", preprint arXiv:0707.3082.

[32] G. Székelyhidi – "Extremal metrics and K-stability", *Bull. Lond. Math. Soc.* **39** (2007), p. 76–84.

[33] G. Tian – "On a set of polarized Kähler metrics on algebraic manifolds", *J. Differential Geom.* **32** (1990), p. 99–130.

[34] ———, "Kähler-Einstein metrics with positive scalar curvature", *Invent. Math.* **130** (1997), p. 1–37.

[35] S. T. Yau – "On the Ricci curvature of a compact Kähler manifold and the complex Monge-Ampère equation. I", *Comm. Pure Appl. Math.* **31** (1978), p. 339–411.

[36] S. Zelditch – "Szegő kernels and a theorem of Tian", *Internat. Math. Res. Notices* **6** (1998), p. 317–331.

[37] B. Zhou & X. H. Zhu – "A note on the K-stability on toric manifolds", preprint arXiv:0706.0505.

X. Chen, Department of Mathematics, University of Wisconsin-Madison, Van Vleck Hall, 480 Lincoln Dr Madison WI 53706-1388, USA • *E-mail :* `xiu@math.wisc.edu`

Y. Tang, Department of Mathematics, University of Wisconsin-Madison, Van Vleck Hall, 480 Lincoln Dr Madison WI 53706-1388, USA • *E-mail :* `tang@math.wisc.edu`

Astérisque

Rafe Mazzeo

Flexibility of singular Einstein metrics

Astérisque, tome 321 (2008), p. 169-193
<http://www.numdam.org/item?id=AST_2008__321__169_0>

© Société mathématique de France, 2008, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (http://smf4.emath.fr/Publications/Asterisque/) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM
Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

Astérisque
321, 2008, p. 169–193

FLEXIBILITY OF SINGULAR EINSTEIN METRICS

by

Rafe Mazzeo

Dedicated to Jean Pierre Bourguignon on his 60^{th} birthday.

Abstract. — This is a survey of a collection of related results about the deformation properties of Einstein metrics on a certain class of spaces with stratified singular structure. The results in low dimensions are particularly clean, and are motivated by applications in hyperbolic and convex geometry. The three-dimensional setting is related to an old conjecture by Stoker about flexibility of convex hyperbolic polyhedra, and we report on a partial answer. We also review some of the analytic methods used to prove these results.

***Résumé* (Flexibilité des métriques d'Einstein singulières).** — Cet article constitue un compte-rendu d'une collection de résultats autour des propriétés de déformation des métriques d'Einstein sur une certaine classe d'espaces à structure singulière stratifiée. Les résultats en basse dimension sont particulièrement intéressants, et ils sont motivés par des applications en géométrie hyperbolique et convexe. La configuration 3-dimensionnelle est reliée à une vieille conjecture de Stoker sur la flexibilité des polyèdres convexes hyperboliques et nous proposons une réponse partielle. Nous examinons également certaines méthodes analytiques utilisées pour démontrer ces résultats.

1. Introduction

The construction and study of canonical metrics on smooth Riemannian manifolds is a longstanding central theme in geometric analysis. The term 'canonical' can be interpreted in many ways; we shall take it here to mean Einstein, so we study metrics satisfying $\mathrm{Ric}^g = \lambda g$ for some constant λ. Beyond the basic existence questions, one of the main problems in this subject is to understand whether a given Einstein metric is rigid or flexible, i.e. admits nontrivial deformations amongst Einstein metrics.

2000 Mathematics Subject Classification. — 58D27, 53C25, 53C21, 35J70.
Key words and phrases. — Einstein metrics, conic singularities, moduli problems.

Supported by the NSF under Grant DMS-0505709.

© Astérisque 321, SMF 2008

As a rule of thumb, negative curvature usually implies rigidity of while positive (or even just nonnegative) curvature often allows 'flexibility'. Our goal here is to discuss how Einstein metrics on a certain class of stratified singular spaces are sometimes flexible precisely because of the geometry of the singular set. There are several well-known instances of this, for example the classical problem of determining the flexibility of convex polyhedra in space forms, to which some of the theory discussed below is directly applicable. This provides one motivation for the more general study of Einstein metrics on stratified spaces proposed here.

This paper is intended as a brief survey of some small part of a broader subject, focusing on one interesting class of stratified spaces – the iterated cone-edge spaces – and presenting some recent results about the local deformation theory of Einstein metrics on these, particularly in low dimensions where it is closely related to many investigations in geometric topology concerning the class of 'cone-manifolds' introduced by Thurston. The new results reported here are parts of various ongoing collaborations by the author with Gregoire Montcouquiol, Frank Pacard and Hartmut Weiss, and are also very closely related to his work with Olivier Biquard. The intention is to indicate the beginnings of a coherent 'story', and one which seems worthy of further development, albeit from a very personal point of view. Due to limitations of space and the author's expertise, we do not touch on many interesting situations where singular Einstein metrics have already been studied by others, e.g. for metrics with special holonomy, particularly in complex geometry. Finally, we also do not discuss any global aspects of this moduli problem, in particular the compactification theory, though this is likely to be both important and very interesting.

Let us first mention a few facts about Einstein metrics on smooth manifolds. Recall that a deformation of an Einstein metric g is a (smooth) one-parameter family of metrics g_t with $g_0 = g$; it is called a trivial deformation if there exists a one-parameter family of diffeomorphisms ϕ_t of the underlying manifold such that $g_t = \phi_t^* g_0$. In other words, the moduli space $\mathcal{E}(M)$ of Einstein metrics on a given manifold M is the space of all metrics satisfying the Einstein condition modulo diffeomorphisms. (Just as for surfaces, one may mod out by all diffeomorphisms or by those isotopic to the identity, but since our focus is on local aspects of the deformation problem, we do not emphasize this here.) As usual, it is more convenient to study an auxiliary equation whose solution space yields all (nearby) Einstein metrics without the diffeomorphism redundancy; this is done by introducing an auxiliary gauge condition to make the problem elliptic; we describe this later. There is a well-known result due to Koiso [**22**] which states that if M is compact, then $\mathcal{E}(M)$ is a finite dimensional analytic set. (This means that it can be covered by neighbourhoods, each of which is identified by a real analytic diffeomorphism with the zero set of an analytic function in a finite dimensional Euclidean space.) The subtlety in proving this, and the reason that its conclusion is not more specific (e.g. with respect to the dimension or smoothness of the moduli space) is that when the manifold is compact, the deformation theory is either trivial or obstructed. Indeed, one standard approach to such a result is to apply an implicit function theorem, for which one needs surjectivity of the linearization of the relevant operator, and if this holds, then the space of solutions of the nonlinear geometric problem is locally parametrized by elements of the nullspace of this linearization. (We describe this in greater detail in §6 below.) This linearization is a self-adjoint elliptic operator, so when M is compact, its surjectivity is equivalent to its injectivity. Thus if the linearization is surjective, it is injective too and the Einstein metric is rigid; on the other hand, if the linearization has nontrivial nullspace, then it has cokernel too, so the implicit function theorem does not directly apply. There is a standard trick to handle situations of this sort, known as Ljapunov-Schmidt reduction, but one can then only deduce much less precise conclusions.

Despite the fact that the 'formal dimension' of the moduli space of Einstein metrics on a compact manifold is zero, there are many manifolds M for which $\mathcal{E}(M)$ is positive dimensional and sometimes even smooth. The best known examples in higher dimensions are the families of flat tori, and less trivially, the family of Calabi-Yau K3 surfaces, for which very detailed results may be obtained using algebraic geometric techniques. Also worthy of note are the recent results of [**11**] about the existence of smooth high dimensional families of Einstein metrics on the sphere, all far from the standard metric, obtained using an integrable systems approach. On the other hand, as suggested above, if M is compact and the sectional curvatures of g are everywhere nonpositive, and negative somewhere, then g is rigid; this can be proved using the Bochner technique. The special case where (M, g) is locally symmetric was proved in various settings of increasing generality by Weil, Calabi, and Matsushita-Murakami. For quite different reasons it is known that the standard metric on the sphere is also rigid. We refer to the outstanding expository monographs [**6**], [**21**], and the collection [**23**], for more about these facts and their proofs.

When the manifold is noncompact or incomplete, this rigidity or deformation theory has a very different flavour. At one end is the study of the asymptotic boundary problems associated to Einstein metrics with certain types of asymptotically symmetric geometries, in particular the much-studied case of asymptotically hyperbolic Einstein metrics (also called Poincaré-Einstein metrics), see [**17**], [**24**], [**2**], as well as the work by Biquard on complex and quaternionic analogues, [**8**], [**7**], and more recently, some 'higher rank' analogues studied by the author in collaboration wtih Biquard, [**9**], [**10**]. As the name 'asymptotic boundary problem' suggests, complete Einstein metrics with these various types of asymptotic conditions come in infinite dimensional families, and the emphasis changes to parametrizing them by some appropriate type of boundary data, which in these cases are the associated parabolic geometries on the ideal boundary at infinity. The parabolic geometry associated to an asymptotically real hyperbolic Einstein metric is a conformal structure on the boundary of the geodesic compactification. The classical analogue of this, when $\dim M = 3$ and M is a quasiFuchsian convex cocompact hyperbolic manifold, was developed by Ahlfors and Bers; here, hyperbolic structures satisfying these hypotheses are in bijective correspondence with the space of conformal structures on the boundary at infinity, which is then a compact surface with two components. In the asymptotically complex or quaternionic settings, the parabolic geometries are the CR and quaternionic CR structures; for the higher rank cases, the relevant asymptotic boundary structures are somewhat less familiar but quite explicit, see [**9**]. There is also some recent progress by Anderson on the boundary problem in the usual sense for incomplete Einstein metrics on manifolds with boundary [**3**].

Of a different nature is the study of Einstein metrics which are asymptotically locally Euclidean (ALE), or which satisfy other more intricate but related asymptotic conditions, but which in any case are complete and have polynomial volume growth. Almost all known examples of these are metrics with restricted holonomy group, e.g. Kähler-Einstein or even hyperKähler, and that extra structure provides a substantial key to unlocking their properties. These have only finite dimensional deformation spaces, which are in some cases very well understood; we refer again to [**21**] for more on this.

On the other hand, there does not seem to have been any systematic study of Einstein metrics on various classes of spaces with 'geometrically structured' singularities, e.g. manifolds with conic points, edges and iterated edges, or more general stratified spaces, despite their ubiquity 'in nature'. As indicated above, we focus on the local rigidity/flexibity question, and in particular how geometric data at the singular locus can provide at least some of the moduli parameters. There is nothing approaching a comprehensive understanding of this phenomenon yet; rather, we simply present several recent results in this area in order to explain what is possible with current techniques and to emphasize this as an interesting area of study.

To be more specific, we first recall a particular class of Riemannian stratified spaces obtained by an iterated coning procedure and a class of Riemannian metrics on their principal smooth strata which induce metrics on each of the substrata. The general problem we pose is to study Einstein metrics in this class of singular spaces. In successive sections we consider this problem in the two, three, and higher-dimensional settings. Not surprisingly, the results are of decreasing specificity. The case of conic surfaces is certainly well-motivated through its association with marked Teichmüller theory, and serves as an excellent test-case for refining techniques for the more general settings. The results on this discussed here are joint work with H. Weiss. The three-dimensional case has also appeared in many other guises before, in particular through the study of 'cone-manifolds' (or conifolds as we shall call them) by Thurston and many others, cf. the exposition in [**14**]. There is another application, however, concerning deformations of three-dimensional convex hyperbolic polyhedra; one result discussed here, obtained recently with G. Montcouquiol, answers the infinitesimal version of an old question due to Stoker in polyhedral geometry, and is of independent interest to that community. The local version of this same question has been treated more recently, using methods from geometric topology, independently by Montcouquiol and Weiss. There is quite a large literature about various aspects of these three-dimensional problems, however, and we shall mention only a few other related results. Finally, the situation in higher dimensions is much less complete; we discuss one result concerning isolated conic singularities, with Pacard and Weiss, and another (very special) set of examples of Einstein metrics bending along codimension two edges, but can mostly point to what are likely to be the tractable interesting directions. The final section contains some discussion of the analytic underpinnings of the proofs of these results: first, a reminder of one convenient gauge choice, and second, an overview of the analysis of elliptic operators in the conic and iterated edge settings. I am grateful to all these collaborators for allowing me to report on these ongoing projects here. I have also learned much from conversations with Steve Kerckhoff and Igor Rivin, and through my long-standing collaboration with Frank Pacard. Finally, the referee provided some very helpful comments about the exposition and relevant literature.

2. Iterated cone-edge spaces

Let (N, h) be a compact stratified Riemannian space with top-dimensional stratum an open dense subset. We refer to [**35**] for generalities on stratified spaces; by Riemannian we mean that each stratum S carries a Riemannian metric h_S, which extends smoothly to the closure of this stratum, and that this collection of metrics satisfies the obvious compatibility relationships: if S_1 and S_2 are any two strata with $\iota : S_1 \hookrightarrow \overline{S_2}$, then $\iota^* h_{S_2} = h_{S_1}$. We are interested in the subclass consisting of iterated cone-edge spaces; these are spaces obtained locally by an iterated coning process, starting from smooth compact manifolds. First, recall that the (complete) cone over N, $C(N)$, is the space $([0,\infty)_r \times N) / \sim$, where $\sim$ is the equivalence relation collapsing $\{0\} \times N$ to a point, endowed with the metric $dr^2 + r^2 h$. The truncated cone where $r \leq 1$ is denoted $C_1(N)$. Any singular stratum $S \subset N$ induces a singular stratum $C(S)$ in $C(N)$, with $\dim C(S) = \dim S + 1$. Now we can make the

Definition 2.1. — *We define, for each $k \geq 0$, the class $\mathcal{I}_k$ of compact iterated cone-edge spaces of depth k. This is done by induction on k. An iterated cone-edge space of depth 0 is a compact smooth manifold. A stratified space M lies in $\mathcal{I}_k$ if for any $p \in M$, if S is the open singular stratum containing p and $\dim S = \ell$, then there exists a neighbourhood $\mathcal{U}$ of p in M diffeomorphic to the product $\mathcal{V} \times C_1(N)$ where $\mathcal{V} \subset \mathbb{R}^\ell$ is an open Euclidean ball diffeomorphic to a neighbourhood in S and $N \in \mathcal{I}_j$ for some $j < k$. We assume furthermore that the integer $n = \ell + \dim C(N)$ is independent of the point $p \in M$; this number is called the dimension of M.*

If $\dim S > 0$, then we say that the stratum S is an edge in M with link N; some neighbourhood of S in M is diffeomorphic to a bundle of cones over S with fibre $C(N)$. If $\dim S = 0$, then we call it a conic point, but note that if N is itself singular, then there are edges of lower depth which terminate at this point.

An iterated cone-edge metric g on M is by definition one which respects this diffeomorphism, i.e. is locally quasi-isometric to one of the form $g \sim dr^2 + r^2 h + \kappa$, where h is an iterated cone-edge metric on N and κ is a metric on S.

To simplify the name a bit, we shall often call these iterated edge spaces. They are much simpler than general stratified spaces, both geometrically and analytically. To our knowledge, they were first singled out for the tractability of analysis on them in Cheeger's famous paper [**12**].

We shall be discussing Einstein metrics on iterated edge spaces, but one should note that the precise definition of an Einstein metric on such a singular space is not necessarily clear. Obviously any such metric g should be Einstein on the principal open stratum, but it is not clear whether one should also require special conditions on the restrictions of these metrics to the lower dimensional strata. This might be clarified, for example, by examining what it means for a metric on an iterated edge space to be critical for the Einstein-Hilbert action. In the low dimensional cases we shall be focusing on mostly, this issue does not arise, while the higher dimensional examples discussed in §5 are so special that they are not necessarily indicative of the general case. In any case, this seems like an important issue to clarify.

3. Surfaces with conic singularities

The simplest setting for our general problem is the existence and deformation theory of compact constant curvature surfaces with isolated conic singularities. This can be approached by various different methods, but we follow one modelled on the presentation developed by Tromba [**41**] to study Teichmüller theory on compact surfaces without singular points since it generalizes to higher dimensions more readily.

We first recall some facts about 'marked Teichmüller theory'. Let M be a compact, oriented two-dimensional surface with genus γ. Any conformal class $[g]$ on M contains a constant curvature metric g_0 which is unique after some choice of normalization (when $\gamma > 1$, it is unique if we fix the curvature to be -1; on the other hand, requiring that the area equals 1, say, yields a unique solution for $\gamma \geq 1$; for $\gamma = 0$ there is the usual nonuniqueness due to the Möbius group). For $\gamma \geq 1$, the genus γ Riemann moduli space $\mathcal{R}_\gamma$ is thus identified with the space of all constant curvature metrics (completed in some Banach topology) with area 1 modulo the space of all diffeomorphisms (of appropriate regularity); the genus γ Teichmüller space $\mathcal{T}_\gamma$ is the quotient of the same space of metrics by the identity component of this group of diffeomorphisms, i.e. the subgroup of diffeomorphisms which are isotopic to the identity. Finally, the marked Teichmüller space $\mathcal{T}_{\gamma,k}$ is the quotient of the same space of metrics by the still smaller subgroup of diffeomorphisms which are isotopic to the identity and which fix a specified collection of points $\{p_1, \ldots, p_k\} \subset M$. When $\chi(M) - k < 0$, it again follows from the classical uniformization theorem that in each marked conformal class there is a unique complete, hyperbolic, finite area metric. When $\chi(M) - k = 0$, this uniformizing metric is flat. These metrics are the ones most commonly associated to marked conformal structures.

Another choice of canonical metric in this setting is obtained as follows: given a conformal class $\mathfrak{c}$ on M, a collection of distinct points $\{p_1, \ldots, p_k\} \subset M$, and a collection of positive numbers $\{\alpha_1, \ldots, \alpha_k\}$, find a metric with constant curvature K on $M \setminus \{p_1, \ldots, p_k\}$ which has an isolated conic singularity at each p_j, with specified cone angle $2\pi\alpha_j$ there.

In two dimensions, the local geometry of a constant curvature metric around a conic point is quite simple. Define the function $\mathrm{sn}_{\mathrm{K}}(r)$ to be the unique solutions to the initial value problem $f'' + \mathrm{K}f = 0$ satisfying $\mathrm{sn}_{\mathrm{K}}(0) = 0$, $\mathrm{sn}'_{\mathrm{K}}(0) = 1$. Then the metric

$$g = dr^2 + \mathrm{sn}^2_{\mathrm{K}}(r)\, dy^2, \qquad 0 < r < r_0, \quad y \in \mathbb{R}/2\pi\alpha$$

is a two-dimensional conic metric with curvature K and cone angle $2\pi\alpha$; when $\mathrm{K} \leq 0$ we can take $r_0 = \infty$, while $r_0 < \pi/\sqrt{\mathrm{K}}$ when $\mathrm{K} > 0$. There is another useful representation,

$$g = e^{2\phi}|z|^{2\beta}|dz|^2, \qquad \alpha = 1 + \beta,$$

in local holomorphic coordinates near 0 in the disk in $\mathbb{C}$; here ϕ is some explicit function (which equals 0 when $\mathrm{K} = 0$). More generally, if ϕ is any reasonably smooth function, we say that a metric of this form has isolated conic singularity at 0 with cone angle $2\pi\alpha = 2\pi(1 + \beta)$.

This existence problem translates into finding a solution of the following semilinear elliptic PDE. Let $\overline{g}$ be any fixed metric in the conformal class $\mathfrak{c}$, and fix a function

$Z(p_1,\dots,p_k,\beta_1,\dots,\beta_k)$ which depends smoothly on the p_j and β_j, is everywhere positive and smooth away from the p_j, and such that near each of these points, in a local holomorphic coordinate z, equals $|z|^{2\beta_j}|dz|^2$. Now write

$$\widehat{g} = Z(p_1,\dots,p_k,\beta_1,\dots,\beta_k)\overline{g};$$

the metric we seek can be written

$$g = e^{2\phi}\widehat{g}$$

and it has curvature $K_g \equiv \mathrm{K}$ if and only if

$$\Delta_{\widehat{g}}\phi - K_{\widehat{g}} + \mathrm{K}e^{2\phi} = 0. \tag{3.1}$$

The solvability of (3.1) in general, i.e. for arbitrary values of the cone angle parameters $\beta_j > -1$, is not known. One immediate constraint is obtained by applying the Gauss-Bonnet formula to the surfaces with boundary $M \setminus \cup_j B_\epsilon(p_j)$ and letting $\epsilon \searrow 0$; this shows that if a solution exists, then

$$\mathrm{K} \times \mathrm{Area}\,(M,g)) = 2\pi(\chi(M) + \sum_{j=1}^{k}\beta_j). \tag{3.2}$$

Since the term on the right, which we call the conic Euler characteristic, changes sign as the cone angles vary, it is more convenient to fix the area and let K be determined by (3.2). Solutions are obtained easily when $\mathrm{K} \leq 0$ using barrier techniques, but we pass out of this regime as soon as some of the β_j become sufficiently large. There is a complete existence theory when $\mathrm{K} > 0$ only if we restrict each β_j to lie in the interval $(-1,0)$, corresponding to each cone angle lying in the interval $(0,2\pi)$.

Theorem 3.1. — *Suppose that each $\beta_j \in (-1,0)$. Then there is a solution of (3.1) if and only if for each $i = 1,\dots,k$,*

$$\chi(M) + \sum_{j\neq i}\beta_j < 2 + \beta_i, \tag{3.3}$$

and moreover, if we require its area to equal 1, *then this solution is unique. The Gauss curvature* K *of this solution is equal to the conic Euler characteristic $\chi(M) + \sum\beta_j$.*

Notice that by adding β_i to each side, (3.3) is equivalent to

$$\chi(M) + \sum_{j=1}^{k}\beta_j < 2(1+\beta_i);$$

since $\beta_i \in (-1,0)$, the right hand side is always positive, so this condition presents a genuine obstruction only when conic Euler characteristic is positive. Existence and uniqueness when $\mathrm{K} \leq 0$ is due to Troyanov [**42**] and also McOwen [**31**]; Troyanov used variational methods and was also able to obtain existence in the spherical case ($\mathrm{K} > 0$), assuming (3.3). Later, Luo and Tian [**25**] proved that (3.3) is necessary and that the solution obtained by Troyanov is unique. We shall say that the k-tuple $(\beta_1, \dots, \beta_k) \in (-1,0)^k$ lies in the Troyanov region if it satisfies (3.3).

There are a few results concerning existence and uniqueness when some of the cone angles are larger than 2π, cf. [**16**], [**44**], but the situation is still far from being well understood. A very interesting recent survey paper by Troyanov [**43**] provides a lot of information about the flat case.

Although it is implicit in these existence proofs that the solutions in Theorem 3.1 depend smoothly on the underlying parameters, i.e. the marked conformal structure and cone angles, it is still of interest to understand the way in which all these metrics fit together. There are some analytic subtleties, and overcoming them in this context is good preparation for understanding the higher dimensional situation. Furthermore, it is hoped that these methods will eventually produce a much better picture of the existence theory when the cone angles are larger than 2π. This was carried out several years ago in joint work with Hartmut Weiss [**30**] (but only now finally being written). The basic result is the

Theorem 3.2. — *Let M be a compact orientable surface, as above. Let $\mathcal{T}^{\mathrm{conic}}_{g,k}$ denote the space of all constant curvature metrics on M with area equal to* 1 *and with conic singularities at k distinct points on M with cone angles $2\pi(1+\beta_j)$, and $\mathcal{T}^{\mathrm{conic}}_{g,k,o}$ the subset where the k-tuple $(\beta_1, \dots, \beta_k)$ satisfy the Troyanov constraint* (3.3) *and $\beta_j \in (-1,0)$ for all j. Then $\mathcal{T}^{\mathrm{conic}}_{g,k,o}$ is a smooth open manifold of dimension $6g-6+3k$; it contains as an open submanifold the subset of metrics with negative curvature, and as a hypersurface the subset of flat conic metrics.*

The complete result contains other statements about the limiting behaviour of these metrics as $(\beta_1, \dots, \beta_k)$ approaches the boundary of the Troyanov region; we refer to [**30**] for more details.

The proof involves constructing coordinate charts for this space, which we do by regarding its elements as satisfying the Einstein equations (just the constant Gauss curvature equation in this dimension) along with an auxiliary gauge condition. The new feature, however, is that these equations are singular at the conic points, so one must substitute other techniques to handle them. The gauge condition and some discussion of elliptic theory adapted to conic spaces will be given at the end of this paper.

There are several intriguing open questions. First, although there is no constant curvature metric when the cone angles are still less than 2π but the β_j lie outside the Troyanov region, is there still some sort of canonical metric with these specified cone angles? Many years ago, Tian suggested that in these cases the canonical metric should be a Ricci soliton with prescribed cone angles; there has been no good progress on this yet. Second, when extending this result into the region where some of the cone angles are greater than 2π, it is likely that one will need to face the issue of bifurcations; for example a conic point with cone angle 3π may split into either two or three points. It will be interesting to put this on solid analytic footing.

4. Conifolds in dimension 3

Iterated cone-edge spaces with constant curvature (or more generally, with a (G, X) structure) were introduced by Thurston as a generalization of orbifolds. He called these 'cone-manifolds', but this is not a very satisfactory name, so we opt for the alternate moniker 'conifold'. Thus, for us, a conifold is an iterated cone-edge space (M, g) such that the induced metric g_S on any stratum has constant sectional curvature K, and each stratum is totally geodesic in an appropriate sense in all higher dimensional ones for which it lies in the frontier. As in the surface case, we call a conifold hyperbolic, flat or spherical depending on whether K is negative, zero or positive. We restrict attention to the 3-dimensional case, and mainly the hyperbolic case. This has been intensively studied due to many applications to the theory of smooth hyperbolic 3-manifolds, stemming from Thurston's proof of the orbifold theorem and various hopes to use similar methods to prove the full hyperbolization theorem. The monograph [**14**] provides a good introduction.

The singular locus of M, denoted Σ, is a union of 1- and 0-dimensional strata which constitute the edges and vertices of a graph (with the slightly nonstandard convention that it may have components which are closed loops). Near each edge of the singular locus, M is a bundle of cones with cone angle constant along that edge; this bundle is trivial unless the edge is a closed loop. Near each vertex, M is identified with the cone over a space Y, which is a copy of S^2 with k conic points, where k is the valence of that vertex. (Thus the edges of Σ are depth 1 singularities, while its vertices are depth 2 singular points.) We shall denote the vertex set of Σ by $\mathcal{V}$ and its edge set by $\mathcal{E}$. and we write the valence of a vertex v as the integer $n(v) \geq 3$.

In a neighbourhood of the interior of any edge of Σ, the metric g has a standard form; this involves the function $\mathrm{sn}_{\mathrm{K}}(\rho)$ used in the surface case, and its companion, $\mathrm{cs}_{\mathrm{K}}(\rho)$, which is the unique solution to $f'' + \mathrm{K}f = 0$, $\mathrm{cs}_{\mathrm{K}}(0) = 1$, $\mathrm{cs}'_{\mathrm{K}}(0) = 0$. Now, with ρ equal to the distance from that edge (in a sufficiently small neighbourhood), we have

$$g = d\rho^2 + \mathrm{sn}^2_{\mathrm{K}}(\rho)\, dy^2 + \mathrm{cs}^2_{\mathrm{K}}(\rho)\, dt^2, \qquad y \in \mathbb{R}/2\pi\alpha = S^1_{2\pi\alpha}, \quad t \in (-a, a). \tag{4.4}$$

We call this the constant curvature cylinder with cone angle $2\pi\alpha$. On the other hand, near a vertex $p \in \mathcal{V}$, M is a constant curvature cone over a spherical cone surface (N, h), so g has the form

$$dr^2 + \mathrm{sn}^2_{\mathrm{K}} r\, h, \tag{4.5}$$

where r is the distance to the vertex. Of course, h in turn has the form described in the previous section with $\mathrm{K} = +1$ near each one of its singular points. Note that the cone points of each link N correspond to edges of M.

Now let us identify natural geometric parameters. These are of two types: along each edge e there is the cone angle $2\pi\alpha(e)$ (or equivalently, the parameter $\beta(e) = \alpha(e) - 1$), the length $\ell(e)$ of the edge and also a certain twist parameter $\tau(e)$, which will be described below; at each vertex v the parameter is in fact the spherical cone metric on S^2 with $n(v)$ conic points, i.e. an element of $\mathcal{T}^{\mathrm{conic}}_{0,n(v)}$. Hence the total set of free parameters lies in some subset of the space

$$(0,\infty)^{|\mathcal{E}|}_{\alpha} \times (0,\infty)^{|\mathcal{E}|}_{\ell} \times (0,\infty)^{|\mathcal{E}|}_{\tau} \times \prod_{v\in\mathcal{V}} \mathcal{T}^{\mathrm{conic}}_{0,n(v)}.$$

There are some obvious constraints: the cone angle parameter $\alpha(e)$ associated to each edge e determines the angles at the cone points of the spherical links at the terminal vertices of that edge. The length and twist parameters do not satisfy any such 'local' constraints, nor apparently does the marked conformal structure on each spherical link. We denote the set of parameters satisfying these 'obvious' constraints by $\mathcal{P}$. For the same reasons as in the last section, namely the poor understanding of spherical cone surfaces with cone angles larger than 2π or outside the Troyanov region, we restrict attention to the subset $\mathcal{P}_o \subset \mathcal{P}$ where the cone angles satisfy (3.3) and are all less than 2π.

To each element $\zeta = (\alpha(e), \ell(e), \tau(e), h(v)) \in \mathcal{P}_o$, where $\alpha(e), \ell(e)$ and $\tau(e)$ are the cone angle, length and twist parameters associated to each edge $e \in \mathcal{E}$ and $h(v)$ is the spherical cone metric with $n(v)$ conic points in S^2 associated to each vertex v, we can associate a local conifold 'thickening' of the graph Σ as follows. First define the cones with constant curvature K over each spherical cone metric $h(v)$, $v \in \mathcal{V}$ by the formula (4.5), for $0 < r < r_0$. Next, over each edge $e \in \mathcal{E}$ construct the cylindrical metric (4.4), again only up to some small radius. Take the core geodesic of this cylinder to have length $\ell(e)$. The twist parameter $\tau(e)$ provides a way of measuring how these cylinders are attached at either end. It is only a relative parameter unless the edge e is a closed loop, but in that case it is equal to the holonomy around that loop. Let us call this thickened graph the singular germ associated to the data ζ (and with curvature K), and denote it by $\mathcal{S}(\zeta)$.

Here are the two main questions:

i) Which singular germs $\mathcal{S}(\zeta)$ arise as the restriction to a neighbourhood of the singular set Σ of a compact conifold; alternately, to which elements of $\mathcal{P}_o$ does $S(\zeta)$ extend to a compact conifold?
ii) Given a compact conifold (M, g) of curvature K, let $S(\zeta_0)$ denote the associated singular germ. Describe the local, or even just the infinitesimal, structure of

the space of nearby conifolds with the same curvature, or equivalently, of their geometric parameter sets.

Problem i) is much more subtle, and we do not have anything to say about it here. Problem ii), on the other hand, can be treated by analytic methods, much like for constant curvature conic surfaces.

Before proceeding, we describe one special case which is of interest in polyhedral geometry. Let A be a polyhedron in either the sphere S^3, Euclidean space $\mathbb{R}^3$ or hyperbolic space $\mathbb{H}^3$. Then A has a set of edges and vertices and its faces are totally geodesic. At each edge e we may associate a dihedral angle, $\delta(e)$, which is the inner angle between the two faces meeting at that point; similarly, at each vertex v we may associate a 'solid angle', which is a spherical polygon $B_v \subset S^2$ consisting of the set of interior normal directions at v (i.e. it is just the spherical link). The polyhedron A has other geometric parameters as well, namely the lengths of each edge, but in this context there is no twist parameter since there is a unique way of choosing a wedge with opening angle $\delta(e)$ along a geodesic of length $\ell(e)$, and the local structure at each vertex is obtained uniquely by intersecting these wedges associated to all incoming edges.

To pass from such a polyhedron to a conifold, double A simultaneously across all of its faces. Since these faces are totally geodesic, the resulting space M is singular only along a 1-skeleton. Its angle along each edge is given by angle $2\pi\alpha = 2\delta(e)$, while its link at each vertex v is the double of the spherical polygon B_v, hence a spherical cone surface. This conifold has a natural involution, for which A is a fundamental domain. Convexity is a natural condition; if A is a convex polyhedron, then each $\delta(e) < \pi$, hence the cone angles along each edge in the conifold M are all less than 2π. Furthermore, these cone angles lie in the Troyanov region simply because the link at each vertex is a spherical cone surface with all angles less than 2π, and by Luo-Tian [**25**], this can only exist when its cone angles satisfy (3.3). We have therefore proved that if the conifold M is the double of a convex polyhedron, then its geometric parameters lie in $\mathcal{P}_o$.

In an influential 1967 paper [**40**], J.J. Stoker studied the flexibility of convex polyhedra in $\mathbb{R}^3$, and made the conjecture that the dihedral angles of a convex polyhedron determine the angles in each face. (The polyhedron itself is not determined even up to homothety since translating any face parallel to itself leaves all dihedral angles unchanged.) This has become known as the Stoker conjecture. The analogous conjecture in hyperbolic space, that convex polyhedra in $\mathbb{H}^3$ are determined by their dihedral angles, was made explicit by Igor Rivin in his thesis. (Note the stronger statement than in the Euclidean setting; one does not have the same ambiguity from parallel translation of the faces.) Andreev [**4**] settled this when all dihedral angles are less than $\pi/2$, and it was proved by Rivin for ideal polyhedra [**37**] and later by Bao and

Bonahon for hyperideal polyhedra [**5**], with further extensions by Schlenker [**39**]. In the restricted setting of ideal and hyperideal polyhedra, the parameter space is convex, but one of the main difficulties in the general case is that this is no longer true, see [**15**]. There are counterexamples for spherical polyhedra due to Schlenker [**38**], and it is known that any corresponding assertion for conifolds will be more complicated, see [**20**].

One good place to start is to study the infinitesimal or local version of this conjecture, either for polyhedra or conifolds, and for this, analytic methods turn out to be very well suited. One can state the infinitesimal conjecture in the hyperbolic setting as follows:

> If (M, g_t) is a smooth one-parameter family of *hyperbolic* conifold structures with geometric parameters lying in $\mathcal{P}_o$ which preserves the cone angles at each edge to first order, then there is a one-parameter family of diffeomorphisms ϕ_t of the stratified space M such that $g_t - \phi_t^* g_0$ vanishes to second order.

Said more plainly, any nontrivial infinitesimal variation of conifold structures includes a nontrivial infinitesimal variation of some of the dihedral angles; likewise, in any nontrivial variation of convex polyhedra in $\mathbb{H}^3$, the set of dihedral angles must vary. The conjecture in the Euclidean setting is slightly more intricate since it must allow for the phenomenon of families of nonisometric polyhedra with the same dihedral angles which are obtained by parallel translations of the faces.

Several papers in the last decade have addressed special cases. The first, by Hodgson and Kerckhoff [**19**], concerns the case of hyperbolic conifolds with singular set a finite union of loops (hence, no vertices), and they settled the infinitesimal and local conjectures for cone angles less than 2π. More recently, Weiss [**45**] in his thesis generalized their methods to prove the same result for conifolds for which the singular set is allowed to have trivalent vertices and all cone angles are less than π. Some other nice results in this direction have been obtained by Porti and Weiss [**36**] and Huesener, Porti and Suarez [**18**].

The point of view of all of these is to study this from the point of view of deforming representations of the fundamental group into the Möbius group. However, it is also possible to approach these problems using methods from global analysis similar to those used in other dimensions, and this has led to the following result by the author in collaboration with G. Montcouquiol:

***Theorem 4.1* (Infinitesimal conifold Stoker conjecture [27]).** — *Let (M, g) be a hyperbolic conifold with parameters lying in $\mathcal{P}_o$. Then any nontrivial variation of g amongst hyperbolic conifolds changes at least one cone angle to first order.*

More recently still, Montcouquiol and Weiss, independently, have established a local (rather than infinitesimal) result. One formulation is that there is a local parametrization of the set of hyperbolic conifold structures in some neighbourhood of (M, g) by an analytic set (i.e. the zero set of an analytic function) in an ϵ-ball B_ϵ in the space of cone angle parameters $(\beta_1, \ldots, \beta_k)$ around those of g. In other words, all nearby conifolds are parametrized by letting β vary in some analytic subset of the space of cone angles. Both of these authors use techniques similar to the ones employed earlier by Hodgson-Kerckhoff, Weiss, et al.; it is quite likely that the approach of [**27**] can be extended to handle this as well, but this is still work in progress.

There are analogous results in the Euclidean case and also in both the infinitesimal and local setting for convex hyperbolic polyhedra, but we shall not state any of these explicitly here. One subtle point is that if A is a convex hyperbolic polyhedron and (M, g) its conifold double, then there may be conifold variations of (M, g) which are not doubles of hyperbolic polyhedra. This would be very interesting to understand better. The full Stoker conjecture in the polyhedral or conifold setting (either Euclidean or hyperbolic) remains open. A substantial new difficulty which must be faced in the global problem for conifolds is that as the cone angles vary, the topology of the singular set might be forced to change. For example, under a family of deformations, edges might shrink and disappear, or conversely, be generated and grow, or disjoint 'skew' edges might move toward each other and touch. As in the surface case, it is also important to try to push these techniques and results to when the cone angles are larger than 2π.

5. Higher dimensions and codimensions

It is possible to obtain reasonably explicit results about local deformation theory for singular Einstein metrics in low dimensions simply because these metrics have constant sectional curvature. This allows the analytic problem to be reduced to a finite dimensional one. In higher dimensions the situation is quite different. Even though the gauged Einstein equation seems formally well-posed, it becomes highly overdetermined on an iterated cone-edge space, at least near edges of positive dimension, and with codimension at least two. Because of this, very few singular Einstein spaces with interesting singular sets are known in higher dimensions. In this section we first describe the local structure theory of the space of Einstein metrics with isolated conic singularities in general dimensions, and then go on to discuss a few examples of Einstein metrics with higher dimensional singular sets. These examples have a lot of symmetry, and although it is reasonable to think that there might be many other Einstein metrics with similar singular structure, this is quite unknown and seems to be a very difficult problem. We do not discuss any examples where the singular set is itself stratified.

A standard computation, see [**6**], shows that the exact conic metric $g = dr^2 + r^2 h$ on $\mathbb{R}^+ \times N$ is Einstein if and only if the link (N, h) is itself Einstein, with $\mathrm{Ric}^h = (n-2)h$ (where $\dim N = n-1$). This generalizes the standard picture of $\mathbb{R}^n$ with its flat metric as a cone over the sphere with unit radius; cones over spheres of other radii are no longer even Ricci flat (except when $\dim N = 1$; in this case the condition on the link is satisfied by a circle of any radius). Based on this, we see immediately that Einstein deformations of the cone $C(N)$ can be obtained by deforming the link (N, h) in its own Einstein moduli space. If $\dim N = 3$, the link is either the sphere or a spherical space form, neither of which admits Einstein deformations; on the other hand, when $\dim N \geq 4$, it is sometimes possible to obtain a finite dimensional family of Einstein cones this way. More generally, if (M, g) is an Einstein space with isolated conic singularities $p_1, \ldots, p_k$, and if $n = \dim M \geq 5$, then denote by (N_j, h_j) the link at p_j and $\mathcal{E}(N_j)$ the Einstein deformation space of this link. We shall need to impose an extra integrability condition: for each such link, suppose that κ is an infinitesimal Einstein deformation on the entire cone $C(N_j)$ which is homogeneous of degree 0 with respect to radial dilations. Then we assume that κ is the derivative of a one-parameter family of conic Einstein metrics.

Theorem 5.1. — *Let (M, g) be as above, and suppose that the integrability condition is satisfied at each p_j. If the sectional curvature of (M, g) is nonpositive and negative somewhere, then the local Einstein deformation space can be identified with an analytic subset in the product $\prod_j \mathcal{E}(N_j)$. If this curvature condition is not satisfied, the local Einstein deformation space is contained in an analytic subset in the larger space $\prod_j \mathcal{E}(N_j) \times \mathbb{R}^\ell$, where ℓ is the dimension of the space of infinitesimal Einstein deformations which decay at each p_j.*

The integrability condition is a bit of a surprise. It holds in all known situations, but seems to be necessary, at least using our approach.

This theorem, joint with Frank Pacard and Hartmut Weiss [**28**], is a direct analog to the two-dimensional case, but with the important proviso that we know very little about the spaces $\mathcal{E}(N_j)$ beyond the fact that they too are finite dimensional analytic spaces. This last fact is a classical result due to Koiso, cf. [**6**]. The second result in this theorem, about the deformation space in the 'degenerate' situation where there are decaying infinitesimal Einstein deformations, follows by a standard adaptation of the proof of the first part, using Ljapunov-Schmidt reduction.

One motivation for studying this type of singular Einstein space is the fact that Einstein metrics with conic singularities arise naturally as limits in the compactification theory of the Einstein moduli space in four dimensions. This is due originally to

Anderson [1] and Nakajima [33], but see the more recent work by Cheeger and Tian [13]. The precise mechanism by which a conic singularity arises is that a Ricci flat ALE space 'pinches off' in the limit.

We finally turn to the case where (M, g) is an Einstein metric with higher dimensional singular set. Of particular interest is the case when the singular set has a stratum of codimension two, partly because this is quite natural in complex geometry, but also because this should correspond to the greatest flexibility. We mostly discuss this case. There are various examples of this phenomenon known; the simplest arise as quotients of smooth Einstein spaces. In particular, it is not hard to construct examples of hyperbolic manifolds singular along a codimension two edge. One may also construct cohomogeneity one metrics, for which the Einstein condition reduces to an ODE, and which have a singular edge. The paper [29] shows how to adapt an ansatz by Page and Pope to produce families of singular Einstein metrics with simple edge singularities along a smooth codimension two stratum. The examples emphasized there are actually noncompact (their other end is asymptotically hyperbolic), but this is immaterial for the present discussion. To write these down, fix a holomorphic line bundle L with Hermitian metric and connection 1-form θ over a compact Kähler-Einstein manifold $(X, \widehat{g})$ with $c_1 > 0$. The metrics are defined on the complement of a ball around the zero section in L by the formula

$$g = (r^2 - 1)^n P(r)^{-1} dr^2 + c^2 P(r)(r^2 - 1)^{-n}\theta^2 + c(r^2 - 1)\widehat{g};$$

here $P(r)$ satisfies the ODE

$$\frac{d}{dr}(r^{-1}P(r)) = r^{-2}\left(|\Lambda|(r^2 - 1)^{n+1} + c^{-1}\lambda(r^2 - 1)^n\right),$$

and Λ, c and λ are parameters. The issue is to show that there are choices for these parameters, including the initial condition for P, which yield metrics with the stated properties. We refer to [29] for more details.

These metrics have a number of interesting features, but their definition relies on many strong hypotheses and it is unclear whether these features are in any way necessary. Following the approach of this paper, one should be able to discern some of this from the local deformation theory. One attack on this is in the thesis by Montcouquiol [32], who proved that for higher dimensional hyperbolic conifolds with smooth codimension two singular set, all nontrivial infinitesimal deformations must vary the cone angle. However, this result does not in any obvious way imply a local rigidity statement: in the language explained in the final section of this paper, the defect space is infinite dimensional, and does not seem to integrate to families of Einstein metrics, even those just defined near the singular set.

6. Methods

After the geometric descriptions in the earlier parts of this paper, the reader is owed some indication of the methods used to prove these results. We begin with the fairly standard formalism of turning the Einstein deformation problem into an elliptic partial differential equation, and then discuss the extensions of ordinary elliptic theory to manifolds with conic and iterated cone-edge singularities needed for this problem.

6.1. The Einstein equation and Bianchi gauge. — Let M be a smooth compact manifold with $\dim M = n$ and define $\mathcal{M}^{k,\alpha}$ as the space of all $\mathcal{C}^{k,\alpha}$ metrics on M. The mapping

$$g \longmapsto \mathrm{Ric}^g$$

is a second order quasilinear differential operator which is polynomial in the components of g, g^{-1}, ∇g and $\nabla^2 g$, hence is a real analytic mapping $\mathcal{M}^{k+2,\alpha}(M) \to \mathcal{C}^{k,\alpha}(M, S^2T^*M)$. Fixing $\lambda \in \mathbb{R}$, the metrics which are Einstein with this given constant λ are the solutions of

$$\mathcal{E}_\lambda(g) \;:=\; \mathrm{Ric}^g - \lambda g = 0. \tag{6.6}$$

Taking traces of both sides yields $\lambda = R^g/n$, where R^g is the scalar curvature. From now on we fix λ and drop it from the notation.

This equation is not elliptic because of its invariance under diffeomorphisms, i.e. if $\mathcal{E}(g) = 0$, then for any diffeomorphism ϕ of M, $\mathcal{E}(\phi^* g) = 0$. Equivalently, the gauge group $\mathcal{G}^{k+1,\alpha}(M)$ of $\mathcal{C}^{k+1,\alpha}$ diffeomorphisms acts on $\mathcal{M}^{k,\alpha}$ by pullback, and the zero set of $\mathcal{E}$ consists of the orbits of this action. This action is not $\mathcal{C}^1$, so the orbits are not in general smooth, which complicates the global analysis slightly.

Fix g with $\mathcal{E}(g) = 0$. To study the Einstein deformations of g, consider the mapping

$$h \mapsto E^g(h) := \mathrm{Ric}^{g+h} - \lambda(g+h). \tag{6.7}$$

From [**6**, p.63],

$$DE^g\,|_{h=0} = \frac{1}{2}(\nabla^*\nabla - 2\overset{o}{R}{}^g) - (\delta^g)^*(\delta + \frac{1}{2}d\mathrm{tr}^{\,g}); \tag{6.8}$$

here $\overset{o}{R}{}^g$ is the curvature operator acting as a symmetric endomorphism on symmetric two-tensors,

$$(\overset{o}{R}{}^g h)_{ij} = R_{ipjq}h^{pq}$$

and $((\delta^g)^*\omega)_{ij} = \frac{1}{2}(\omega_{i;j} + \omega_{j;i})$. For simplicity we set

$$L^g = \frac{1}{2}(\nabla^*\nabla - 2\overset{o}{R}{}^g), \qquad B^g = \delta^g + \frac{1}{2}d\mathrm{tr}^{\,g}.$$

so that (6.8) takes the simpler form

$$DE^g\,|_{h=0} = \frac{1}{2}L^g - (\delta^g)^*B^g. \tag{6.9}$$

The operator B^g is called the Bianchi operator, and appears in two important identities:

$$B^g(g) = 0, \qquad \text{and} \qquad B^g(\mathrm{Ric}^g) = 0. \tag{6.10}$$

The first is trivial, and the second follows from the contracted second Bianchi identity. Note that this yields

$$h \longmapsto B^{g+h}E^g(h) \equiv 0$$

for any g, h. Now suppose that g is Einstein; linearizing this identity at $h = 0$ gives

$$0 = B^g \; DE^g|_0 = B^g L^g - B^g(\delta^g)^* B^g. \tag{6.11}$$

This means in particular that

$$\mathrm{ran}\,(DE^g\,|_0) \subset \ker(B^g);$$

in other words, on any compact manifold, the Einstein equation is always obstructed since its linearization has range lying in a proper subspace (in fact, the nullspace of the underdetermined differential operator B^g).

The orbit of the diffeomorphism group has tangent space at g given by the range of the mapping $(\delta^g)^*$; the restriction of DE^g to the orthogonal complement of this subspace, i.e. to the nullspace of δ^g, is elliptic. We shall use a slight variant of this procedure, restricting DE^g instead to the nullspace of B^g. This 'Bianchi gauge', introduced in [**8**], is very convenient for calculations.

The system $h \mapsto (DE^g(h), B^g(h))$ is elliptic in the sense of Agmon-Douglis-Nirenberg, and so one can look for gauge group representatives for all Einstein metrics near to g as solutions of $E^g(h) = 0$, $B^g(h) = 0$. We consider instead the operator

$$h \longmapsto N^g(h) \;:=\; E^g(h) + (\delta^{g+h})^* B^g(h). \tag{6.12}$$

Its linearization when g is Einstein is

$$DN^g|_{h=0} \;:=\; L^g = \frac{1}{2}(\nabla^g)^*\nabla^g - \overset{o}{R}{}^g. \tag{6.13}$$

Clearly $(E^g(h), B^g(h)) = (0,0)$ implies $N^g(h) = 0$, and the converse is almost true as well:

Proposition 6.1. — *If $N^g(h) = 0$ and $\mathrm{Ric}^{g+h} < 0$, then $g+h$ is Einstein and h satisfies the gauge condition $B^g(h) = 0$.*

Proof. — Let $\gamma = B^g(h)$. Applying δ^{g+h} to $N^g(h) = 0$ gives $(\delta^{g+h}(\delta^{g+h})^* - \frac{1}{2}d\delta^{g+h})\gamma = 0$. Now recall the Weitzenböck formula on 1-forms

$$B^k(\delta^k)^* = \delta^k(\delta^k)^* - \frac{1}{2}d\delta^k = \frac{1}{2}\left((\nabla^k)^*\nabla^k - \mathrm{Ric}^k\right) \tag{6.14}$$

for any metric k (where the first equality uses $\operatorname{tr}^k(\delta^k)^* = -\delta^k$), and so the equation above becomes $((\nabla^{g+h})^*\nabla^{g+h} - \mathrm{Ric}^{g+h})\gamma = 0$. Because $\mathrm{Ric}^{g+h} < 0$, this operator is an isomorphism, and so $\gamma = 0$ as desired. □

As a final comment, if h is an arbitrary (small) solution of $N^g(h) = 0$, then the metric $g + h$ is a Ricci soliton: it satisfies the equation $E(g+h) = (\delta^{g+h})^*\omega$ where $\omega = -B^g(h)$. This suggests that a problem which may be somewhat less obstructed than the deformation problem for Einstein metrics is the deformation problem for Ricci solitons.

6.2. Conic and edge operators. — Implementing this analytic formalism for the Einstein deformation problem on singular spaces requires an understanding of the mapping properties for linear elliptic operators on such spaces. There is a good theory to draw upon for spaces with isolated conic or simple edge singularities, which we describe briefly below, and this can be extended to the depth 2 singularities which appear in the three-dimensional theory. However, its full extension to the general iterated cone-edge setting does not yet exist. Rather than presenting this linear theory in any sort of generality, we present the main results quickly in two and three-dimensions and then indicate the general picture. As a reference for this material we list [**26**]: it does contain all the results quoted below (at least when the singular set is either a point or a smooth submanifold), albeit in a very general form. There are other more accessible and direct approaches for some of this, which work particularly well in low dimensions, see [**34**] and [**45**], for example.

Let M be a surface with isolated conic points, and suppose that L is a 'generalized Laplacian' associated to the metric. In other words $L = \nabla^*\nabla + R$ where R is a naturally defined symmetric endomorphism depending only on the curvature tensor and its covariant derivatives; rather than being precise about this we turn always to the special operators which were described in the last subsection, e.g. the scalar Laplacian, the linearized gauged Einstein operator, etc. Near a conic point p we can choose coordinates (r, y) in terms of which

$$L = \frac{\partial^2}{\partial r^2} + \frac{A(r,y)}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}H \tag{6.15}$$

where $A(r,y)$ is smooth on $r \geq 0$, $A(0,y) \equiv 1$, and $H = H(r) = \partial_y^2 + R_0(r,y)$ is a family of operators acting on (sections over the) circle, also depending smoothly on r. Associated to L is the set Λ of indicial weights. We say that γ is an indicial root of L if there exists a function $\phi(y)$ such that $L(r^\gamma\phi(y)) = \mathcal{O}(r^{\gamma-1})$. This is one order better than the expected rate of blowup or decay, $r^{\gamma-2}$, which indicates that there is some 'leading order' cancellation. Indeed, we see directly that

$$L(r^\gamma\phi(y)) = r^{\gamma-2}\left(\gamma^2 + \partial_y^2 + R_0(0,y)\right)\phi(y) + \mathcal{O}(r^{\gamma-1}),$$

so $-\gamma^2$ is equal to an eigenvalue of $\partial_y^2 + R_0(0,y)$ and ϕ its corresponding eigenfunction. A priori, these indicial roots may be real or complex, and we define the set Λ to consist of the real parts of these indicial roots. It is easy to see that Λ is infinite and discrete. For example, if L is the scalar Laplacian, or the linearized gauged Einstein operator acting on trace-free symmetric two-tensors, for a metric g which has a single conic point with cone angle $2\pi\alpha$, then $\Lambda = \{j/\alpha : j \in \mathbb{Z}\}$ and $\{\pm(2 \pm j/\alpha) : j \in \mathbb{Z}\}$, respectively. For generalized Laplacians in two dimensions, Λ is symmetric about 0. In all the examples of interest here, all indicial roots are real, so we think of Λ as being precisely equal to the set of indicial roots.

Consider the action of L on weighted Hölder spaces $r^\nu \mathcal{C}_g^{k,\alpha}(M)$ consisting of functions of the form $u = r^\nu v$ where v is in the 'geometric Hölder space' with respect to the metric g, i.e. computed using derivatives and distance functions for g. The basic result is

Proposition 6.2. — *The mapping*

$$L : r^\nu \mathcal{C}_g^{2,\alpha}(M) \longrightarrow r^{\nu-2}\mathcal{C}_g^{0,\alpha}(M) \tag{6.16}$$

is Fredholm if and only if $\nu \notin \Lambda$. This mapping is surjective for $\nu \notin \Lambda$, $\nu \ll 0$, and injective when $\nu \gg 0$. Finally, (6.16) is injective for some value $\nu \notin \Lambda$ if and only if the corresponding mapping with weight $-\nu$ is surjective.

If γ is an indicial root, then a sequence of appropriate cutoffs of the approximate solution $r^\gamma \phi(y)$ can be used to show that (6.16) does not have closed range when $\nu = \gamma$. The more subtle parts of this result are to show Fredholmness when $\nu \notin \Lambda$, and to prove the final statement. We comment on this last part especially a bit further. Both assertions are proved by constructing, for each $\nu \notin \Lambda$, a generalized inverse G for L. This is done using L^2 based methods, but the key is to show that the Schwartz kernel of this operator has a precise structure as a smooth (or rather, polyhomogeneous) function, which allows one to pass easily between weighted L^2 and weighted Hölder estimates. This is what makes it possible to prove the 'duality statement' in a Hölder setting.

For all the relevant operators in the Einstein deformation problem, one can prove that (6.16) is injective for $\nu > \nu_0 \geq 0$, hence surjective for $-\nu < -\nu_0 \leq 0$. In order to apply the inverse function theorem or any related contraction mapping arguments, the operator should be surjective, hence by this result we should be working on a Hölder space with negative weight. However, it is impossible to let the nonlinear PDE act on functions unbounded at the conic points. The resolution of this dilemma rests on the

Proposition 6.3. — *Let L be as above, and suppose that $\Lambda \ni \nu > 0$ is such that (6.16) is injective. List the indicial roots of L with real parts in the interval $(-\nu, \nu)$ by*

$\{\gamma_j : -N \leq j \leq N\}$ with the convention that $\gamma_{-j} = -\gamma_j$. Then for any $f \in r^{\nu-2}\mathcal{C}^{0,\alpha}_g$, there exists a solution $u \in r^{-\nu}\mathcal{C}^{2,\alpha}_g$ to $Lu = f$, and this u has a decomposition

$$u = \sum_{j=-N}^{N} u_j(y) r^{\gamma_j} + v, \qquad v \in r^{\nu}\mathcal{C}^{2,\alpha}_g, \tag{6.17}$$

where each $u_j(y)$ is an eigenfunction associated with that indicial root.

The finite dimensional span of terms $u_j(y)r^{\gamma_j}$ which appear in this partiale xpansion is called the defect space. This result is general, but the crucial observation is that for our particular problem, every element of this defect space can be identified with an infinitesimal variation of a one-parameter family of solutions of the nonlinear gauged Einstein operator. All the geometric moduli for the problem appear in this way: the underlying Teichmüller parameter on the compact surface, the location of the conic points and the cone angles. We can then 'solve' the problem $Lu = f$ with $f \in r^{\nu-2}\mathcal{C}^{0,\alpha}_g$, $\nu > \nu_0$, by first altering these geometric parameters and applying the operator L corresponding to the new metric to the remainder term v. In this way we can set up an iteration scheme to solve the nonlinear perturbation problem.

Suitable generalizations of this idea are behind all of the other deformation results discussed in the earlier parts of this paper. (Indeed, this type of idea has been applied in very many other circumstances.) The result about higher dimensional Einstein spaces with isolated conic singularities uses essentially the same linear theory, and there are direct analogues of the Propositions 6.2 and 6.3. The calculational aspects are substantially different, however, and unfortunately much more complicated. On the other hand, the main indicial term to understand is the one corresponding to the indicial root $\gamma = 0$. The integrability hypothesis we imposed is that this does indeed correspond to a one-parameter family of Einstein deformations of the cone $C(N)$. This appears to be generically true, and can be checked explicitly in several cases of interest, but it is unclear if it holds in general (chances are that it does not).

When the singular set is a submanifold Y of dimension $d > 0$, the linear theory is more complicated. It suffices to work in neighbourhoods diffeomorphic to $\mathcal{U} \times C_1(N)$, where $\mathcal{U} \subset Y$ is a coordinate neighbourhood. Each of the operators of interest have the form

$$L = \frac{\partial^2}{\partial r^2} + \frac{A(r,y)}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}H + K;$$

where H is much as before, an elliptic operator acting the link N, while K restricts to an elliptic operator on Y. We can define the indicial roots of L exactly as in the previous case; the operator K does not appear at the lowest level in terms of a formal count of powers of r. It does play a role in a new model operator that we need to consider in this setting, called the normal operator of L. This is defined as

$$N(L) = \frac{\partial^2}{\partial r^2} + \frac{A(0,y)}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}H(0) + \Delta_{\mathbb{R}^d},$$

acting on function (or sections) on $C(N) \times \mathbb{R}^d$. We roll the requisite results into the one

Proposition 6.4. — *The mapping*

$$L : r^{\nu}\mathcal{C}^{2,\alpha}_g(M) \longrightarrow r^{\nu-2}\mathcal{C}^{0,\alpha}_g(M)$$

is Fredholm if and only if ν does not lie in the indicial weight set Λ, and addition,

$$N(L) : r^{\nu}\mathcal{C}^{2,\alpha}_g(C(N) \times \mathbb{R}^d) \longrightarrow r^{\nu-2}\mathcal{C}^{0,\alpha}_g(C(N) \times \mathbb{R}^d)$$

is an isomorphism. If the nullspace of $N(L)$ is nontrivial, then it is automatically infinite dimensional and the same is true for the nullspace of L for the space with the same weight; the analogous statement is true for the cokernel. As in the conic setting, this mapping is surjective from $r^{\nu}\mathcal{C}^{2,\alpha}_g$ if and only if it is injective from $r^{-\nu}\mathcal{C}^{2,\alpha}_g$.

Suppose $\Lambda \ni \nu > 0$ is such that (6.16) is injective, and list the indicial roots of L in the interval $(-\nu, \nu)$ as $\{\gamma_j : |j| \leq N\}$ with $\gamma_{-j} = -\gamma_j$. Then for any $f \in r^{\nu-2}\mathcal{C}^{0,\alpha}_g$, there exists a solution $u \in r^{-\nu}\mathcal{C}^{2,\alpha}_g$ such that

$$u = \sum_{j=-N}^{N} u_j(y) r^{\gamma_j} + v. \tag{6.18}$$

with each $u_j(y)$ equal to a multiple of the eigenfunction associated with that indicial root.

It is no longer true that v or the coefficients $u_j(y)$ in this decomposition are as smooth as formal considerations would dictate, and this leads to some considerable, and perhaps insurmountable, analytic difficulties when attempting to apply this linear theory to our nonlinear problem. More plainly, when the codimension of Y in M is equal to 2, then the defect space corresponding to crossing the indicial root 0 is infinite dimensional, and its elements do not correspond in any reasonable way to geometric motions.

When M is 3-dimensional, it is possible to overcome this using the fact that there is a simple correspondence between the overall geometric parameters for the problem in a neighbourhood of the singular set and their 'traces' in the asymptotic expansions of infinitesimal Einstein deformations along the singular set.

The final issue to discuss is the general three-dimensional case, when M is a conifold whose singular set contains not only edges but also vertices. One now needs an extension of the theory of conic and edge operators discussed above. Fortunately, the generalization needed is the simplest one possible, where the depth 2 points are isolated vertices of cones where the links are themselves spaces with isolated conic singularities. The idea is to try to adapt the conic theory at these vertices, even though the links are not smooth; in general one would expect to have difficulties with lack of smoothness in the asymptotic expansions along the edges which terminate at this vertex, and the main new steps are to control these expansions uniformly on approach to these depth 2 vertices.

References

[1] M. T. ANDERSON – "Ricci curvature bounds and Einstein metrics on compact manifolds", *J. Amer. Math. Soc.* **2** (1989), p. 455–490.

[2] ______, "Topics in conformally compact Einstein metrics", in *Perspectives in Riemannian geometry*, CRM Proc. Lecture Notes, vol. 40, Amer. Math. Soc., 2006, p. 1–26.

[3] ______, "On boundary value problems for Einstein metrics", *Geom. Topol.* **12** (2008), p. 2009–2045.

[4] E. M. ANDREEV – "Convex polyhedra in Lobačevskiĭ spaces", *Mat. Sb. (N.S.)* **81 (123)** (1970), p. 445–478.

[5] X. BAO & F. BONAHON – "Hyperideal polyhedra in hyperbolic 3-space", *Bull. Soc. Math. France* **130** (2002), p. 457–491.

[6] A. L. BESSE – *Einstein manifolds*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3), vol. 10, Springer, 1987.

[7] O. BIQUARD – "Métriques autoduales sur la boule", *Invent. Math.* **148** (2002), p. 545–607.

[8] ______, "Asymptotically symmetric Einstein metrics", SMF/AMS Texts and Monographs, vol. 13, 2006.

[9] O. BIQUARD & R. MAZZEO – "Parabolic geometries as conformal infinities of Einstein metrics", *Arch. Math. (Brno)* **42** (2006), p. 85–104.

[10] ______, "A nonlinear Poisson transform for Einstein metrics on product spaces", preprint arXiv:math/0701868.

[11] C. P. BOYER, K. GALICKI & J. KOLLÁR – "Einstein metrics on spheres", *Ann. of Math.* **162** (2005), p. 557–580.

[12] J. CHEEGER – "Spectral geometry of singular Riemannian spaces", *J. Differential Geom.* **18** (1983), p. 575–657.

[13] J. CHEEGER & G. TIAN – "Curvature and injectivity radius estimates for Einstein 4-manifolds", *J. Amer. Math. Soc.* **19** (2006), p. 487–525.

[14] D. COOPER, C. D. HODGSON & S. P. KERCKHOFF – *Three-dimensional orbifolds and cone-manifolds*, MSJ Memoirs, vol. 5, Mathematical Society of Japan, 2000.

[15] R. DÍAZ – "A generalization of Andreev's theorem", *J. Math. Soc. Japan* **58** (2006), p. 333–349.

[16] A. EREMENKO – "Metrics of positive curvature with conic singularities on the sphere", *Proc. Amer. Math. Soc.* **132** (2004), p. 3349–3355.

[17] C. R. GRAHAM & J. M. LEE – "Einstein metrics with prescribed conformal infinity on the ball", *Adv. Math.* **87** (1991), p. 186–225.

[18] M. HEUSENER, J. PORTI & E. SUÁREZ – "Regenerating singular hyperbolic structures from Sol", *J. Differential Geom.* **59** (2001), p. 439–478.

[19] C. D. HODGSON & S. P. KERCKHOFF – "Rigidity of hyperbolic cone-manifolds and hyperbolic Dehn surgery", *J. Differential Geom.* **48** (1998), p. 1–59.

[20] ———, "Universal bounds for hyperbolic Dehn surgery", *Ann. of Math.* **162** (2005), p. 367–421.

[21] D. D. JOYCE – *Compact manifolds with special holonomy*, Oxford Mathematical Monographs, Oxford University Press, 2000.

[22] N. KOISO – "Rigidity and infinitesimal deformability of Einstein metrics", *Osaka J. Math.* **19** (1982), p. 643–668.

[23] C. LEBRUN & M. WANG (eds.) – *Surveys in differential geometry, vol. VI: Essays on Einstein manifolds*, International Press, 2001.

[24] J. M. LEE – "Fredholm operators and Einstein metrics on conformally compact manifolds", *Mem. Amer. Math. Soc.* **183** (2006), p. 83.

[25] F. LUO & G. TIAN – "Liouville equation and spherical convex polytopes", *Proc. Amer. Math. Soc.* **116** (1992), p. 1119–1129.

[26] R. MAZZEO – "Elliptic theory of differential edge operators. I", *Comm. Partial Differential Equations* **16** (1991), p. 1615–1664.

[27] R. MAZZEO & G. MONTCOUQUIOL – "Infinitesimal rigidity of hyperbolic cone manifolds and the infinitesimal Stoker problem for hyperbolic polyhedra", in preparation.

[28] R. MAZZEO, F. PACARD & H. WEISS – "Einstein metrics and Ricci solitons with conic singularities", in preparation.

[29] R. MAZZEO & M. SINGER – "Some remarks on conic degeneration and bending of Poincaré-Einstein metrics", preprint arXiv:0709.1498.

[30] R. MAZZEO & H. WEISS – "Teichmüller theory of conic surfaces", in preparation.

[31] R. C. MCOWEN – "Point singularities and conformal metrics on Riemann surfaces", *Proc. Amer. Math. Soc.* **103** (1988), p. 222–224.

[32] G. MONTCOUQUIOL – "Déformations de métriques Einstein sur des variétés à singularité coniques", Ph.D. Thesis, Thesis, Université Paul Sabatier – Toulouse III, 2005.

[33] H. NAKAJIMA – "Hausdorff convergence of Einstein 4-manifolds", *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **35** (1988), p. 411–424.

[34] F. PACARD – "Lectures on connected sum constructions in analysis and geometry", preprint `http://perso-math.univ-mlv.fr/users/pacard.frank/Lecture-Part-I.pdf`.

[35] M. J. PFLAUM – *Analytic and geometric study of stratified spaces*, Lecture Notes in Math., vol. 1768, Springer, 2001.

[36] J. PORTI & H. WEISS – "Deforming Euclidean cone 3-manifolds", *Geom. Topol.* **11** (2007), p. 1507–1538.

[37] I. RIVIN – "A characterization of ideal polyhedra in hyperbolic 3-space", *Ann. of Math.* **143** (1996), p. 51–70.

[38] J.-M. SCHLENKER – "Dihedral angles of convex polyhedra", *Discrete Comput. Geom.* **23** (2000), p. 409–417.

[39] ———, "Hyperideal circle patterns", *Math. Res. Lett.* **12** (2005), p. 85–102.

[40] J. J. STOKER – "Geometrical problems concerning polyhedra in the large", *Comm. Pure Appl. Math.* **21** (1968), p. 119–168.

[41] A. J. TROMBA – *Teichmüller theory in Riemannian geometry*, Lectures in Mathematics ETH Zürich, Birkhäuser, 1992.

[42] M. TROYANOV – "Prescribing curvature on compact surfaces with conical singularities", *Trans. Amer. Math. Soc.* **324** (1991), p. 793–821.

[43] ———, "On the moduli space of singular Euclidean surfaces", in *Handbook of Teichmüller theory. Vol. I*, IRMA Lect. Math. Theor. Phys., vol. 11, Eur. Math. Soc., Zürich, 2007, p. 507–540.

[44] M. UMEHARA & K. YAMADA – "Metrics of constant curvature 1 with three conical singularities on the 2-sphere", *Illinois J. Math.* **44** (2000), p. 72–94.

[45] H. WEISS – "Local rigidity of 3-dimensional cone-manifolds", *J. Differential Geom.* **71** (2005), p. 437–506.

R. MAZZEO • *E-mail :* `mazzeo@math.stanford.edu`, Department of Mathematics, Stanford University, Stanford, CA 94305, USA

Astérisque

PIOTR T. CHRUSCIEL
JOÃO LOPES COSTA

On uniqueness of stationary vacuum black holes

Astérisque, tome 321 (2008), p. 195-265
<http://www.numdam.org/item?id=AST_2008__321__195_0>

© Société mathématique de France, 2008, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (http://smf4.emath.fr/Publications/Asterisque/) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM
Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

Astérisque
321, 2008, p. 195–265

ON UNIQUENESS OF STATIONARY VACUUM BLACK HOLES

by

Piotr T. Chruściel & João Lopes Costa

It is a pleasure to dedicate this work to J.-P. Bourguignon on the occasion of his 60^{th} birthday.

Abstract. — We prove uniqueness of the Kerr black holes within the connected, non-degenerate, analytic class of regular vacuum black holes.

Résumé (*Sur l'unicité de trous noirs stationnaires dans le vide*). — On démontre l'unicité de trous noirs de Kerr dans la classe de trous noirs connexes, analytiques, réguliers, non-dégénérés, solutions des équations d'Einstein du vide.

1. Introduction

It is widely expected that the Kerr metrics provide the only stationary, asymptotically flat, sufficiently well-behaved, vacuum, four-dimensional black holes. Arguments to this effect have been given in the literature [**12, 84**] (see also [**51, 77, 91**]), with the hypotheses needed not always spelled out, and with some notable technical gaps. The aim of this work is to prove a precise version of one such uniqueness result for analytic space-times, with detailed filling of the gaps alluded to above.

The results presented here can be used to obtain a similar result for electro-vacuum black holes (compare [**13, 71**]), or for five-dimensional black holes with three commuting Killing vectors (see also [**56, 57**]); this will be discussed elsewhere [**31**].

We start with some terminology. The reader is referred to Section 2.1 for a precise definition of asymptotic flatness, to Section 2.2 for that of a domain of outer communications $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$, and to Section 3 for the definition of mean-non-degenerate horizons. A Killing vector K is said to be complete if its orbits are complete, i.e., for every $p \in \mathscr{M}$ the orbit $\phi_t[K](p)$ of K is defined for all $t \in \mathbb{R}$; in an asymptotically flat context, K is called *stationary* if it is timelike at large distances.

2000 Mathematics Subject Classification. — 83C57.
Key words and phrases. — Stationary black holes, no-hair theorems.

© Astérisque 321, SMF 2008

A key definition for our work is the following:

Definition 1.1. — *Let $(\mathscr{M}, \mathfrak{g})$ be a space-time containing an asymptotically flat end $\mathscr{S}_{\rm ext}$, and let K be stationary Killing vector field on $\mathscr{M}$. We will say that $(\mathscr{M}, \mathfrak{g}, K)$ is I^+–*regular *if K is complete, if the domain of outer communications $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$ is globally hyperbolic, and if $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$ contains a spacelike, connected, acausal hypersurface $\mathscr{S} \supset \mathscr{S}_{\rm ext}$, the closure $\overline{\mathscr{S}}$ of which is a topological manifold with boundary, consisting of the union of a compact set and of a finite number of asymptotic ends, such that the boundary $\partial\overline{\mathscr{S}} := \overline{\mathscr{S}} \setminus \mathscr{S}$ is a topological manifold satisfying*

$$\partial\overline{\mathscr{S}} \subset \mathscr{E}^+ := \partial\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle \cap I^+(\mathscr{M}_{\rm ext}), \tag{1.1}$$

with $\partial\overline{\mathscr{S}}$ meeting every generator of $\mathscr{E}^+$ precisely once. (See Figure 1.1.)

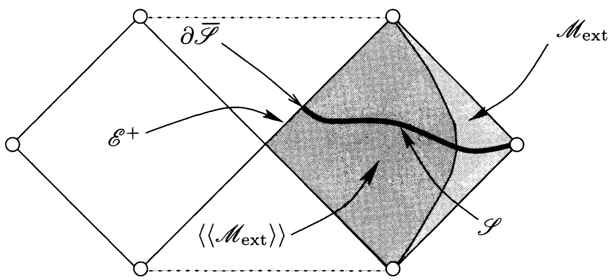

FIGURE 1.1. The hypersurface $\mathscr{S}$ from the definition of I^+–regularity.

In Definition 1.1, the hypothesis of asymptotic flatness is made for definiteness, and is not needed for several of the results presented below. Thus, this definition appears to be convenient in a wider context, e.g. if asymptotic flatness is replaced by Kaluza-Klein asymptotics, as in [**20, 23**].

Some comments about the definition are in order. First we require completeness of the orbits of the stationary Killing vector because we need an action of $\mathbb{R}$ on $\mathscr{M}$ by isometries. Next, we require global hyperbolicity of the domain of outer communications to guarantee its simple connectedness, to make sure that the area theorem holds, and to avoid causality violations as well as certain kinds of naked singularities in $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$. Further, the existence of a well-behaved spacelike hypersurface gives us reasonable control of the geometry of $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$, and is a prerequisite to any elliptic PDEs analysis, as is extensively needed for the problem at hand. The existence of compact cross-sections of the future event horizon prevents singularities on the future part of the boundary of the domain of outer communications, and eventually guarantees the smoothness of that boundary. (Obviously I^+ could have been replaced by I^-

throughout the definition, whence $\mathscr{E}^+$ would have become $\mathscr{E}^-$.) We find the requirement (1.1) somewhat unnatural, as there are perfectly well-behaved hypersurfaces in, e.g., the Schwarzschild space-time which do not satisfy this condition, but we have not been able to develop a coherent theory without assuming some version of (1.1). Its main point is to avoid certain zeros of the stationary Killing vector K at the boundary of $\mathscr{S}$, which otherwise create various difficulties; e.g., it is not clear how to guarantee then smoothness of $\mathscr{E}^+$, or the static-or-axisymmetric alternative. [(1)] Needless to say, all those conditions are satisfied by the Schwarzschild, Kerr, or Majumdar-Papapetrou solutions.

We have the following, long-standing conjecture, it being understood that both the Minkowski and the Schwarzschild space-times are members of the Kerr family:

Conjecture 1.2. — *Let $(\mathscr{M}, \mathfrak{g})$ be a vacuum, four-dimensional space-time containing a spacelike, connected, acausal hypersurface $\mathscr{S}$, such that $\overline{\mathscr{S}}$ is a topological manifold with boundary, consisting of the union of a compact set and of a finite number of asymptotically flat ends. Suppose that there exists on $\mathscr{M}$ a complete stationary Killing vector K, that $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$ is globally hyperbolic, and that $\partial\overline{\mathscr{S}} \subset \mathscr{M} \setminus \langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$. Then $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$ is isometric to the domain of outer communications of a Kerr space-time.*

In this work we establish the following special case thereof:

Theorem 1.3. — *Let $(\mathscr{M}, \mathfrak{g})$ be a stationary, asymptotically flat, I^+-regular, vacuum, four-dimensional analytic space-time. If each component of the event horizon is mean non-degenerate, then $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$ is isometric to the domain of outer communications of one of the Weinstein solutions of Section 6.7. In particular, if $\mathscr{E}^+$ is connected and mean non-degenerate, then $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$ is isometric to the domain of outer communications of a Kerr space-time.*

In addition to the references already cited, some key steps of the proof are due to Hawking [**48**], and to Sudarsky and Wald [**89**], with the construction of the candidate solutions with several non-degenerate horizons due to Weinstein [**93, 94**]. It should be emphasized that the hypotheses of analyticity and non-degeneracy are highly unsatisfactory, and one believes that they are not needed for the conclusion.

One also believes that no candidate solutions with more than one component of $\mathscr{E}^+$ are singularity-free, but no proof is available except for some special cases [**69, 92**].

A few words comparing our work with the existing literature are in order. First, the event horizon in a smooth or analytic black hole space-time is a priori only a Lipschitz surface, which is way insufficient to prove the usual static-or-axisymmetric alternative.

(1) In fact, this condition is not needed for *static* metric if, e.g., one assumes at the outset that all horizons are non-degenerate, as we do in Theorem 1.3 below, see the discussion in the Corrigendum to [**18**].

Here we use the results of [**22**] to show that event horizons in regular stationary black hole space-times are as differentiable as the differentiability of the metric allows. Next, no paper that we are aware of adequately shows that the "area function" is non-negative within the domain of outer communications; this is due both to a potential lack of regularity of the intersection of the rotation axis with the zero-level-set of the area function, and to the fact that the gradient of the area function could vanish on its zero level set *regardless of whether or not the event horizon itself is degenerate.* The second new result of this paper is Theorem 5.4, which proves this result. The difficulty here is to exclude *non-embedded Killing prehorizons* (for terminology, see below), and we have not been able to do it without assuming analyticity or axisymmetry, *even for static solutions.* Finally, no previous work known to us establishes the behavior, as needed for the proof of uniqueness, of the relevant harmonic map at points where the horizon meets the rotation axis. The third new result of this paper is Theorem 6.1, settling this question for non-degenerate black-holes. (This last result requires, in turn, the Structure Theorem 4.5 and the Ergoset Theorem 5.24, and relies heavily on the analysis in [**19**].) Last but not least, we provide a coherent set of conditions under which all pieces of the proof can be combined to obtain the uniqueness result.

We note that various intermediate results are established under conditions weaker than previously cited, or are generalized to higher dimensions; this is of potential interest for further work on the subject.

1.1. Static case. — Assuming *staticity*, i.e., stationarity and hypersurface-orthogonality of the stationary Killing vector, a more satisfactory result is available in space dimensions less than or equal to seven, and in higher dimensions on manifolds on which the Riemannian rigid positive energy theorem holds: non-connected configurations are excluded, without any *a priori* restrictions on the gradient $\nabla(\mathfrak{g}(K,K))$ at event horizons.

More precisely, we shall say that a manifold $\widehat{\mathscr{S}}$ is of *positive energy type* if there are no asymptotically flat complete Riemannian metrics on $\widehat{\mathscr{S}}$ with positive scalar curvature and vanishing mass except perhaps for a flat one. This property has been proved so far for all n–dimensional manifolds $\widehat{\mathscr{S}}$ obtained by removing a finite number of points from a compact manifold of dimension $3 \le n \le 7$ [**86**], or under the hypothesis that $\widehat{\mathscr{S}}$ is a spin manifold of any dimension $n \ge 3$, and is expected to be true in general [**14, 70**].

We have the following result, which finds its roots in the work of Israel [**61**], with further simplifications by Robinson [**85**], and with a significant strengthening by Bunting and Masood-ul-Alam [**10**]:

Theorem 1.4. — *Under the hypotheses of Conjecture 1.2, suppose moreover that* $(\langle\langle \mathscr{M}_{\text{ext}}\rangle\rangle, \mathfrak{g})$ *is analytic and* K *is hypersurface-orthogonal. Let* $\widehat{\mathscr{S}}$ *denote the manifold obtained by doubling* $\mathscr{S}$ *across the non-degenerate components of its boundary and compactifying, in the doubled manifold, all asymptotically flat regions but one to a point. If* $\widehat{\mathscr{S}}$ *is of positive energy type, then* $\langle\langle \mathscr{M}_{\text{ext}}\rangle\rangle$ *is isometric to the domain of outer communications of a Schwarzschild space-time.*

Remark 1.5. — As a corollary of Theorem 1.4 one obtains non-existence of black holes as above with some components of the horizon degenerate. In space-time dimension four an elementary proof of this fact has been given in [**26**], but the simple argument there does not seem to generalize to higher dimensions in any obvious way.

Remark 1.6. — Analyticity is only needed to exclude non-embedded degenerate prehorizons within $\langle\langle \mathscr{M}_{\text{ext}}\rangle\rangle$. In space-time dimension four it can be replaced by the condition of axisymmetry and I^+–regularity, compare Theorem 5.2.

Proof. — We want to invoke [**18**], where $n = 3$ has been assumed; the argument given there generalizes immediately to those higher dimensional manifolds on which the positive energy theorem holds. However, the proof in [**18**] contains one mistake, and one gap, both of which need to be addressed.

First, in the case of degenerate horizons $\mathscr{H}$, the analysis of [**18**] assumes that the static Killing vector has no zeros on $\mathscr{H}$; this is used in the key Proposition 3.2 there, which could be wrong without this assumption. The non-vanishing of the static Killing vector is justified in [**18**] by an incorrectly quoted version of Boyer's theorem [**8**], see [**18**, Theorem 3.1]. Under a supplementary assumption of I^+–regularity, the zeros of a Killing vector which could arise in the closure of a degenerate Killing horizon can be excluded using Corollary 3.3. In general, the problem is dealt with in the addendum to the arXiv versions vN, $N \geq 2$, of [**18**] in space-dimension three, and in [**20**] in higher dimensions.

Next, neither the original proof, nor that given in [**18**], of the Vishveshwara-Carter Lemma, takes properly into account the possibility that the hypersurface $\mathscr{N}$ of [**18**, Lemma 4.1] could fail to be embedded.(2) This problem is taken care of by Theorem 5.4 below with $s = 1$, which shows that $\langle\langle \mathscr{M}_{\text{ext}}\rangle\rangle$ cannot intersect the set where $W := -\mathfrak{g}(K, K)$ vanishes. This implies that K is timelike on $\langle\langle \mathscr{M}_{\text{ext}}\rangle\rangle \supset \mathscr{S}$, and null on $\partial\overline{\mathscr{S}}$. The remaining details are as in [**18**]. □

(2) This problem affects points 4c,d,e and f of [**18**, Theorem 1.3], which require the supplementary hypothesis of existence of an embedded closed hypersurface within $\mathscr{N}$; the remaining claims of [**18**, Theorem 1.3] are justified by the arguments described here.

2. Preliminaries

2.1. Asymptotically flat stationary metrics. — A space-time $(\mathscr{M}, \mathfrak{g})$ will be said to possess an *asymptotically flat end* if $\mathscr{M}$ contains a spacelike hypersurface $\mathscr{S}_{\mathrm{ext}}$ diffeomorphic to $\mathbb{R}^n \setminus B(R)$, where $B(R)$ is an open coordinate ball of radius R, with the following properties: there exists a constant $\alpha > 0$ such that, in local coordinates on $\mathscr{S}_{\mathrm{ext}}$ obtained from $\mathbb{R}^n \setminus B(R)$, the metric γ induced by $\mathfrak{g}$ on $\mathscr{S}_{\mathrm{ext}}$, and the extrinsic curvature tensor K_{ij} of $\mathscr{S}_{\mathrm{ext}}$, satisfy the fall-off conditions

$$\gamma_{ij} - \delta_{ij} = O_k(r^{-\alpha}), \qquad K_{ij} = O_{k-1}(r^{-1-\alpha}), \tag{2.1}$$

for some $k > 1$, where we write $f = O_k(r^\alpha)$ if f satisfies

$$\partial_{k_1} \dots \partial_{k_\ell} f = O(r^{\alpha-\ell}), \quad 0 \le \ell \le k. \tag{2.2}$$

For simplicity we assume that the space-time is vacuum, though similar results hold in general under appropriate conditions on matter fields, see [**4, 25**] and references therein. Along any spacelike hypersurface $\mathscr{S}$, a Killing vector field X of $(\mathscr{M}, \mathfrak{g})$ can be decomposed as

$$X = Nn + Y,$$

where Y is tangent to $\mathscr{S}$, and n is the unit future-directed normal to $\mathscr{S}_{\mathrm{ext}}$. The vacuum field equations, together with the Killing equations imply the following set of equations on $\mathscr{S}$, where $R_{ij}(\gamma)$ is the Ricci tensor of γ:

$$D_i Y_j + D_j Y_i = 2NK_{ij}, \tag{2.3}$$

$$R_{ij}(\gamma) + K^k{}_k K_{ij} - 2K_{ik}K^k{}_j - N^{-1}(\mathscr{L}_Y K_{ij} + D_i D_j N) = 0. \tag{2.4}$$

Under the boundary conditions (2.1) with $k \ge 2$, an analysis of (2.3)-(2.4) provides detailed information about the asymptotic behavior of (N, Y). In particular, one can prove that if the asymptotic region $\mathscr{S}_{\mathrm{ext}}$ is contained in a hypersurface $\mathscr{S}$ satisfying the requirements of the positive energy theorem, and if X is timelike along $\mathscr{S}_{\mathrm{ext}}$, then $(N, Y^i) \to_{r\to\infty} (A^0, A^i)$, where the A^μ's are constants satisfying $(A^0)^2 > \sum_i (A^i)^2$. One can then choose adapted coordinates so that the metric can, locally, be written as

$$\mathfrak{g} = -V^2(dt + \underbrace{\theta_i dx^i}_{=\theta})^2 + \underbrace{\gamma_{ij} dx^i dx^j}_{=\gamma}, \tag{2.5}$$

with

$$\partial_t V = \partial_t \theta = \partial_t \gamma = 0 \tag{2.6}$$

$$\gamma_{ij} - \delta_{ij} = O_k(r^{-\alpha}), \quad \theta_i = O_k(r^{-\alpha}), \quad V - 1 = O_k(r^{-\alpha}), \tag{2.7}$$

for any $k \in \mathbb{N}$. As discussed in more detail in [**7**], in γ-harmonic coordinates, and in e.g. a maximal time-slicing, the vacuum equations for $\mathfrak{g}$ form a quasi-linear elliptic system with diagonal principal part, with principal symbol identical to that of the scalar Laplace operator. Methods known in principle show that, in this "gauge", all metric functions have a full asymptotic expansion (3) in terms of powers of $\ln r$ and inverse powers of r. In the new coordinates we can in fact take

$$\alpha = n - 2. \tag{2.8}$$

By inspection of the equations one can further infer that the leading order corrections in the metric can be written in a Schwarzschild form, which in "isotropic" coordinates reads

$$\mathfrak{g}_m = -\left(\frac{1 - \frac{m}{2|x|^{n-2}}}{1 + \frac{m}{2|x|^{n-2}}}\right)^2 dt^2 + \left(1 + \frac{m}{2|x|^{n-2}}\right)^{\frac{4}{n-2}} \left(\sum_{i=1}^{n} dx_i^2\right),$$

where $m \in \mathbb{R}$.

2.2. Domains of outer communications, event horizons. — A key notion in the theory of black holes is that of the *domain of outer communications*: A space-time $(\mathscr{M}, \mathfrak{g})$ will be called stationary if there exists on $\mathscr{M}$ a complete Killing vector field K which is *timelike* in the asymptotically flat region $\mathscr{S}_{\mathrm{ext}}$. (4) For $t \in \mathbb{R}$ let $\phi_t[K] : \mathscr{M} \to \mathscr{M}$ denote the one-parameter group of diffeomorphisms generated by K; we will write ϕ_t for $\phi_t[K]$ whenever ambiguities are unlikely to occur. The exterior region $\mathscr{M}_{\mathrm{ext}}$ and the *domain of outer communications* $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$ are then defined as (5) (compare Figure 2.1)

$$\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle = I^+(\underbrace{\cup_t \phi_t(\mathscr{S}_{\mathrm{ext}})}_{=:\mathscr{M}_{\mathrm{ext}}}) \cap I^-(\cup_t \phi_t(\mathscr{S}_{\mathrm{ext}})). \tag{2.9}$$

The *black hole region* $\mathscr{B}$ and the *black hole event horizon* $\mathscr{H}^+$ are defined as

$$\mathscr{B} = \mathscr{M} \setminus I^-(\mathscr{M}_{\mathrm{ext}}), \quad \mathscr{H}^+ = \partial \mathscr{B}.$$

The *white hole region* $\mathscr{W}$ and the *white hole event horizon* $\mathscr{H}^-$ are defined as above after changing time orientation:

$$\mathscr{W} = \mathscr{M} \setminus I^+(\mathscr{M}_{\mathrm{ext}}), \quad \mathscr{H}^- = \partial \mathscr{W}, \quad \mathscr{H} = \mathscr{H}^+ \cup \mathscr{H}^-.$$

(3) One can use the results in, e.g., [**15**] together with a simple iterative argument to obtain the expansion. This analysis holds in any dimension.

(4) In fact, in the literature it is always implicitly assumed that K is *uniformly timelike* in the asymptotic region $\mathscr{S}_{\mathrm{ext}}$, by this we mean that $\mathfrak{g}(K, K) < -\epsilon < 0$ for some ϵ and for all r large enough. This uniformity condition excludes the possibility of a timelike vector which asymptotes to a null one. This involves no loss of generality in well-behaved space-times: indeed, uniformity always holds for Killing vectors which are timelike for all large distances if the conditions of the positive energy theorem are met [**5, 25**].

(5) Recall that $I^-(\Omega)$, respectively $J^-(\Omega)$, is the set covered by past-directed timelike, respectively causal, curves originating from Ω, while $\dot{I}^-$ denotes the boundary of I^-, etc. The sets I^+, etc., are defined as I^-, etc., after changing time-orientation.

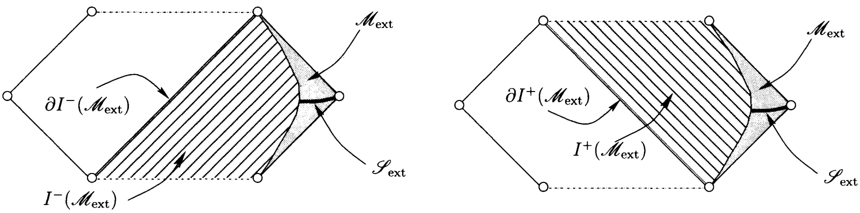

FIGURE 2.1. $\mathscr{S}_{\text{ext}}$, $\mathscr{M}_{\text{ext}}$, together with the future and the past of $\mathscr{M}_{\text{ext}}$. One has $\mathscr{M}_{\text{ext}} \subset I^{\pm}(\mathscr{M}_{\text{ext}})$, even though this is not immediately apparent from the figure. The domain of outer communications is the intersection $I^+(\mathscr{M}_{\text{ext}}) \cap I^-(\mathscr{M}_{\text{ext}})$, compare Figure 1.1.

It follows that the boundaries of $\langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle$ are included in the event horizons. We set

$$\mathscr{E}^{\pm} = \partial \langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle \cap I^{\pm}(\mathscr{M}_{\text{ext}}), \qquad \mathscr{E} = \mathscr{E}^+ \cup \mathscr{E}^- . \tag{2.10}$$

There is considerable freedom in choosing the asymptotic region $\mathscr{S}_{\text{ext}}$. However, it is not too difficult to show, using Lemma 3.6 below, that $I^{\pm}(\mathscr{M}_{\text{ext}})$, and hence $\langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle$, $\mathscr{H}^{\pm}$ and $\mathscr{E}^{\pm}$, are independent of the choice of $\mathscr{S}_{\text{ext}}$ whenever the associated $\mathscr{M}_{\text{ext}}$'s overlap.

Several results below hold without assuming asymptotic flatness: for example, one could assume that we have a region $\mathscr{S}_{\text{ext}}$ on which K is timelike, and carry on with the definitions above. An example of interest is provided by Kaluza-Klein metrics with an asymptotic region of the form $(\mathbb{R}^n \setminus B(R)) \times \mathbb{T}^p$, with the space metric asymptotic to a flat metric there. However, for definiteness, and to avoid unnecessary discussions, we have chosen to assume asymptotic flatness in the definition of I^+-regularity.

2.3. Killing horizons, bifurcate horizons. — A null hypersurface, invariant under the flow of a Killing vector K, which coincides with a connected component of the set

$$\mathscr{N}(K) := \{\mathfrak{g}(K,K) = 0,\ K \neq 0\},$$

is called a *Killing horizon* associated to K.

A set will be called a *bifurcate Killing horizon* if it is the union of four Killing horizons, the intersection of the closure of which forms a smooth submanifold S of co-dimension two, called the *bifurcation surface.* The four Killing horizons consist then of the four null hypersurfaces obtained by shooting null geodesics in the four distinct null directions normal to S. For example, the Killing vector $x\partial_t + t\partial_x$ in Minkowski space-time has a bifurcate Killing horizon, with the bifurcation surface $\{t = x = 0\}$.

The *surface gravity* κ of a Killing horizon $\mathscr{N}$ is defined by the formula

$$d\,(\mathfrak{g}(K,K))\,|_{\mathscr{N}} = -2\kappa K^{\flat}\ , \tag{2.11}$$

where $K^\flat = \mathfrak{g}_{\mu\nu}K^\nu dx^\mu$. A fundamental property is that the surface gravity κ is constant over each horizon in vacuum, or in electro-vacuum, see e.g. [**51**, Theorem 7.1]. The proof given in [**90**] generalizes to all space-time dimensions $n+1 \geq 4$; the result also follows in all dimensions from the analysis in [**55**] when the horizon has compact spacelike sections. (The constancy of κ can be established without assuming any field equations in some cases, see [**62, 82**].) A Killing horizon is called *degenerate* if κ vanishes, and *non-degenerate* otherwise.

2.3.1. *Near-horizon geometry.* — Following [**74**], near a smooth event horizon one can introduce *Gaussian null coordinates*, in which the metric takes the form

$$\mathfrak{g} = r\varphi dv^2 + 2dvdr + 2rh_a dx^a dv + h_{ab}dx^a dx^b. \tag{2.12}$$

(These coordinates can be introduced for any null hypersurface, not necessarily an event horizon, in any number of dimensions). The horizon is given by the equation $\{r=0\}$; replacing r by $-r$ if necessary we can without loss of generality assume that $r>0$ in the domain of outer communications. Assuming that the horizon admits a smooth compact cross-section S, the *average surface gravity* $\langle\kappa\rangle_S$ is defined as

$$\langle\kappa\rangle_S = -\frac{1}{|S|}\int_S \varphi d\mu_h, \tag{2.13}$$

where $d\mu_h$ is the measure induced by the metric h on S, and $|S|$ is the volume of S. We emphasize that this is defined regardless of whether or not some Killing vector K is tangent to the horizon generators; but if K is, and if the surface gravity κ of K is constant on S, then $\langle\kappa\rangle_S$ equals κ.

On a degenerate Killing horizon the surface gravity vanishes by definition, so that the function φ in (2.12) can itself be written as rA, for some smooth function A. The vacuum Einstein equations imply (see [**74**, eq. (2.9)] in dimension four and [**67**, eq. (5.9)] in higher dimensions)

$$\mathring{R}_{ab} = \frac{1}{2}\mathring{h}_a\mathring{h}_b - \mathring{D}_{(a}\mathring{h}_{b)}, \tag{2.14}$$

where $\mathring{R}_{ab}$ is the Ricci tensor of $\mathring{h}_{ab} := h_{ab}|_{r=0}$, and $\mathring{D}$ is the covariant derivative thereof, while $\mathring{h}_a := h_a|_{r=0}$. The Einstein equations also determine $\mathring{A} := A|_{r=0}$ uniquely in terms of $\mathring{h}_a$ and $\mathring{h}_{ab}$:

$$\mathring{A} = \frac{1}{2}\mathring{h}^{ab}\left(\mathring{h}_a\mathring{h}_b - \mathring{D}_a\mathring{h}_b\right) \tag{2.15}$$

(this equation follows again e.g. from [**74**, eq. (2.9)] in dimension four, and can be checked by a calculation in all higher dimensions). We have the following: [6]

[6] Some partial results with a non-zero cosmological constant have also been proved in [**26**].

Theorem 2.1 **([26]).** — *Let the space-time dimension be $n+1$, $n \geq 3$, suppose that a degenerate Killing horizon $\mathscr{N}$ has a compact cross-section, and that $\mathring{h}_a = \partial_a \lambda$ for some function λ (which is necessarily the case in vacuum static space-times). Then* (2.14) *implies $\mathring{h}_a \equiv 0$, so that $\mathring{h}_{ab}$ is Ricci-flat.*

Theorem 2.2 **([47, 67]).** — *In space-time dimension four and in vacuum, suppose that a degenerate Killing horizon $\mathscr{N}$ has a spherical cross-section, and that $(\mathscr{M}, \mathfrak{g})$ admits a second Killing vector field with periodic orbits. For every connected component $\mathscr{N}_0$ of $\mathscr{N}$ there exists an embedding of $\mathscr{N}_0$ into a Kerr space-time which preserves $\mathring{h}_a$, $\mathring{h}_{ab}$ and $\mathring{A}$.*

It would be of interest to understand fully (2.14), in all dimensions, without restrictive conditions.

In the four-dimensional static case, Theorem 2.1 enforces toroidal topology of cross-sections of $\mathscr{N}$, with a flat $\mathring{h}_{ab}$. On the other hand, in the four-dimensional axisymmetric case, Theorem 2.2 guarantees that the geometry tends to a Kerr one, up to errors made clear in the statement of the theorem, when the horizon is approached. (Somewhat more detailed information can be found in [**47**].) So, in the degenerate case, the vacuum equations impose strong restrictions on the near-horizon geometry.

It seems that this is not the case any more for non-degenerate horizons, at least in the analytic setting. Indeed, we claim that for any triple $(N, \mathring{h}_a, \mathring{h}_{ab})$, where N is a two-dimensional analytic manifold (compact or not), $\mathring{h}_a$ is an analytic one-form on N, and $\mathring{h}_{ab}$ is an analytic Riemannian metric on N, there exists a vacuum space-time $(\mathscr{M}, \mathfrak{g})$ with a bifurcate (and thus non-degenerate) Killing horizon, so that the metric $\mathfrak{g}$ takes the form (2.12) near each Killing horizon branching out of the bifurcation surface $S \approx N$, with $\mathring{h}_{ab} = h_{ab}|_{r=0}$ and $\mathring{h}_a = h_a|_{r=0}$; in fact $\mathring{h}_{ab}$ is the metric induced by $\mathfrak{g}$ on S. When N is the two-dimensional torus $\mathbb{T}^2$ this can be inferred from [**73**] as follows: using [**73**, Theorem (2)] with $(\phi, \beta_a, g_{ab})|_{t=0} = (0, 2\mathring{h}_a, \mathring{h}_{ab})$ one obtains a vacuum space-time $(\mathscr{M}' = S^1 \times \mathbb{T}^2 \times (-\epsilon, \epsilon), \mathfrak{g}')$ with a compact Cauchy horizon $S^1 \times \mathbb{T}^2$ and Killing vector K tangent to the S^1 factor of $\mathscr{M}'$. One can then pass to a covering space where S^1 is replaced by $\mathbb{R}$, and use a construction of Rácz and Wald [**82**, Theorem 4.2] to obtain the desired $\mathscr{M}$ containing the bifurcate horizon. This argument generalizes to any analytic $(N, \mathring{h}_a, \mathring{h}_{ab})$ without difficulties.

2.4. Globally hyperbolic asymptotically flat domains of outer communications are simply connected. — Simple connectedness of the domain of outer communication is an essential ingredient in several steps of the uniqueness argument below. It was first noted in [**28**] that this stringent topological restriction is a consequence of the "topological censorship theorem" of Friedman, Schleich and Witt [**37**] for asymptotically flat, stationary and globally hyperbolic domains of outer communications satisfying the null energy condition:

$$R_{\mu\nu}Y^\mu Y^\nu \geq 0 \ \text{ for null } \ Y^\mu. \tag{2.16}$$

In fact, stationarity is not needed. To make things precise, consider a space-time $(\mathscr{M}, \mathfrak{g})$ with several asymptotically flat regions $\mathscr{M}^i_{\mathrm{ext}}$, $i = 1, \ldots, N$, each generating its own domain of outer communications. It turns out [**41**] (compare [**42**]) that the null energy condition prohibits causal interactions between distinct such ends:

Theorem 2.3. — *If $(\mathscr{M}, \mathfrak{g})$ is a globally hyperbolic and asymptotically flat space-time satisfying the null energy condition* (2.16), *then*

$$\langle\langle \mathscr{M}^i_{\mathrm{ext}} \rangle\rangle \cap J^{\pm}(\langle\langle \mathscr{M}^j_{\mathrm{ext}} \rangle\rangle) = \varnothing \ \textit{for}\ i \neq j. \tag{2.17}$$

A clever covering/connectedness argument [(7)] [**41**] shows then: [(8)]

Corollary 2.4. — *A globally hyperbolic and asymptotically flat domain of outer communications satisfying the null energy condition is simply connected.*

In space-time dimension four this, together with standard topological results [**76**], leads to a spherical topology of horizons (see [**28**] together with Proposition 4.4 below):

Corollary 2.5. — *In I^+-regular, stationary, asymptotically flat space-times satisfying the null energy condition, cross-sections of $\mathscr{E}^+$ have spherical topology.*

3. Zeros of Killing vectors

Let $\mathscr{S}$ be a spacelike hypersurface in $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$; in the proof of Theorem 1.3 it will be essential to have no zeros of the stationary Killing vector K on $\overline{\mathscr{S}}$. Furthermore, in the axisymmetric scenario, we need to exclude zeros of Killing vectors of the form $K_{(0)} + \alpha K_{(1)}$ on $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$, where $K_{(0)} = K$ and $K_{(1)}$ is a generator of the axial symmetry. The aim of this section is to present conditions which guarantee that; for future reference, this is done in arbitrary space-time dimension.

We start with the following:

Lemma 3.1. — *Let $\mathscr{S}_{\mathrm{ext}} \subset \mathscr{S} \subset \langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$, and suppose that $\mathscr{S}$ is achronal in $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$. Then for any $p \in \mathscr{M}_{\mathrm{ext}}$ there exists $t_0 \in \mathbb{R}$ such that*

$$\overline{\mathscr{S}} \cap I^+(\phi_{t_0}(p)) = \varnothing.$$

---

(7) Under more general asymptotic conditions it was proved in [**44**] that inclusion induces a surjective homeomorphism between the fundamental groups of the exterior region and the domain of outer communications. In particular, $\pi_1(\mathscr{M}_{\mathrm{ext}}) = 0 \Rightarrow \pi_1(\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle) = 0$.

(8) Strictly speaking, our applications below of [**41**] require checking that the conditions of asymptotic flatness in [**41**] coincide with ours; this, however, can be avoided by invoking directly [**28**].

Proof. — Let $p \in \mathscr{M}_{\text{ext}}$. There exists t_0 such that $r := \phi_{t_0}(p) \in \mathscr{S}_{\text{ext}}$. Suppose that $\overline{\mathscr{S}} \cap I^+(\phi_{t_0}(p)) \neq \varnothing$. Then there exists a timelike future directed curve γ from r to $q \in \overline{\mathscr{S}}$. Let $q_i \in \mathscr{S}$ converge to q; then $q_i \in I^+(r)$ for i large enough, which contradicts achronality of $\mathscr{S}$ within $\langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle$. □

Lemma 3.2. — *Let $S \subset I^+(\mathscr{M}_{\text{ext}})$ be compact.*

1. *There exists $p \in \mathscr{M}_{\text{ext}}$ such that S is contained in $I^+(p)$.*
2. *If $S \subset \partial\langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle \cap I^+(\mathscr{M}_{\text{ext}})$ and if $(\overline{\langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle}, \mathfrak{g})$ is strongly causal at S,*(9) *then for any $p \in \mathscr{M}_{\text{ext}}$ there exists $t_0 \in \mathbb{R}$ such that $S \cap I^+(\phi_{t_0}(p)) = \varnothing$.*

Proof. — 1: Let $q \in S$; there exists $p_q \in \mathscr{M}_{\text{ext}}$ such that $q \in I^+(p_q)$, and since $I^+(p_q)$ is open there exists an open neighborhood $\mathscr{O}_q \subset S$ of q such that $\mathscr{O}_q \subset I^+(p_q)$. By compactness there exists a finite collection $\mathscr{O}_{q_i}$, $i = 1, \ldots, I$, covering S, thus $S \subset \cup_i I^+(p_{q_i})$. Letting $p \in \mathscr{M}_{\text{ext}}$ be any point such that $p_{q_i} \in I^+(p)$ for $i = 1, \ldots, I$, the result follows.

2: Suppose not. Then $\phi_i(p) \in I^-(S)$ for all $i \in \mathbb{N}$, hence there exists $q_i \in S$ such that $q_i \in I^+(\phi_i(p))$. By compactness there exits $q \in S$ such that $q_i \to q$. Let $\mathscr{O}$ be an arbitrary neighborhood of q; since $q \in \mathscr{E}^+$; there exists $r \in \mathscr{O} \cap \langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle$, $p_+ \in \mathscr{M}_{\text{ext}}$, and a future directed causal curve γ from r to p_+. For all i large, this can be continued by a future directed causal curve from p_+ to $\phi_i(p)$, which can then be continued by a future directed causal curve to q_i. But $q_i \in \mathscr{O}$ for i large enough. This implies that every small neighborhood of q meets a future directed causal curve entirely contained within $\langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle$ which leaves the neighborhood and returns, contradicting strong causality of $\overline{\langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle}$. □

It follows from Lemma 3.1, together with point 1 of Lemma 3.2 with $S = \{r\}$, that

Corollary 3.3. — *If $r \in \overline{\mathscr{S}} \cap I^+(\mathscr{M}_{\text{ext}})$, then the stationary Killing vector K does not vanish at r. In particular if $(\mathscr{M}, \mathfrak{g})$ is I^+-regular, then K has no zeros on $\overline{\mathscr{S}}$.* □

To continue, we assume the existence of a commutative group of isometries $\mathbb{R} \times \mathbb{T}^{s-1}$, $s \geq 1$. We denote by $K_{(0)}$ the Killing vector tangent to the orbits $\mathbb{R}$ factor, and we assume that $K_{(0)}$ is timelike in $\mathscr{M}_{\text{ext}}$. We denote by $K_{(i)}$, $i = 1, \ldots, s-1$ the Killing vector tangent to the orbits of the i'th S^1 factor of $\mathbb{T}^{s-1}$. We assume that each $K_{(i)}$ is spacelike in $\langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle$ wherever non-vanishing, which will necessarily be the case if $\langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle$ is chronological. Note that asymptotic flatness imposes $s - 1 \leq n/2$, though most of the results of this section remain true without this hypothesis, when properly formulated.

(9) In a sense made clear in the last sentence of the proof below.

We say that a Killing orbit $\gamma : \mathbb{R} \to \mathscr{M}$ is future-oriented if there exist numbers $\tau_1 > \tau_0$ such that $\gamma(\tau_1) \in I^+(\gamma(\tau_0))$. Clearly all orbits of a Killing vector K are future-oriented in the region where K is timelike. A less-trivial example is given by orbits of the Killing vector $\partial_t + \Omega\partial_\varphi$ in Minkowski space-time. Similarly, in stationary axisymmetric space-times, those orbits of this last Killing vector on which ∂_t is timelike are future-oriented (let $\tau_0 = 0$ and $\tau_1 = 2\pi/\Omega$).

We have:

Lemma 3.4. — *Orbits through $\mathscr{M}_{\mathrm{ext}}$ of Killing vector fields K of the form $K_{(0)} + \sum \alpha_{(i)} K_{(i)}$ are future-oriented.*

Proof. — Recall that for any Killing vector field Z we denote by $\phi_t[Z]$ the flow of Z. Let

$$Y := \sum \alpha_{(i)} K_{(i)}.$$

Suppose, first, that there exists $\tau > 0$ such that $\phi_\tau[Y]$ is the identity. Since $K_{(0)}$ and Y commute we have

$$\phi_\tau[K] = \phi_\tau[K_{(0)} + Y] = \phi_\tau[K_{(0)}] \circ \phi_\tau[Y] = \phi_\tau[K_{(0)}].$$

Setting $\tau_0 = 0$ and $\tau_1 = \tau$, the result follows.

Otherwise, there exists a sequence $t_i \to \infty$ such that $\phi_{t_i}[Y](p)$ converges to p. Since $I^+(p)$ is open there exists a neighborhood $\mathscr{U}^+ \subset I^+(p)$ of $\phi_1[K_{(0)}](p)$. Let $\mathscr{V}^+ = \phi_{-1}[K_{(0)}](\mathscr{U}^+)$, then every point in $\mathscr{U}^+$ lies on a future directed timelike path starting in $\mathscr{V}^+$, namely an integral curve of $K_{(0)}$. There exists $i_0 \geq 1$ so that $t_i \geq 1$ and $\phi_{t_i}[Y](p) \in \mathscr{V}^+$ for $i \geq i_0$. We then have

$$\phi_{t_i}[K](p) = \phi_{t_i}[K_{(0)} + Y](p) = \phi_{t_i - 1}[K_{(0)}]\big(\underbrace{\phi_1(\underbrace{\phi_{t_i}[Y](p)}_{\in \mathscr{V}^+})}_{\in \mathscr{U}^+ \subset I^+(p)}\big) \in I^+(p).$$

The numbers $\tau_0 = 0$ and $\tau_1 = t_{i_0}$ satisfy then the requirements of the definition. □

For future reference we note the following:

Lemma 3.5. — *The orbits through $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$ of any Killing vector K of the form $K_{(0)} + \sum \alpha_{(i)} K_{(i)}$ are future-oriented.*

Proof. — Let $p \in \langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$, thus there exist points $p_\pm \in \mathscr{M}_{\mathrm{ext}}$ such that $p_\pm \in I^\pm(p)$, with associated future directed timelike curves $\gamma_\pm$. It follows from Lemma 3.4 together with asymptotic flatness that there exists τ such that $\phi_\tau[K](p_-) \in I^+(p_+)$ for some τ, as well as an associated future directed curve γ from p_+ to $\phi_\tau[K](p_-)$. Then the curve $\gamma_+ \cdot \gamma \cdot \phi_\tau[K](\gamma_-)$, where $\cdot$ denotes concatenation of curves, is a timelike curve from p to $\phi_\tau(p)$. □

The following result, essentially due to [**27**], turns out to be very useful:

Lemma 3.6. — *Let $\alpha_i \in \mathbb{R}$. For any set C invariant under the flow of $K = K_{(0)} + \sum_i \alpha_i K_i$, the set $I^{\pm}(C) \cap \mathscr{M}_{\rm ext}$ coincides with $\mathscr{M}_{\rm ext}$, if non-empty.*

Proof. — The null achronal boundaries $\dot I^{\mp}(C) \cap \mathscr{M}_{\rm ext}$ are invariant under the flow of K. This is compatible with Lemma 3.4 if and only if $\dot I^{\mp}(C) \cap \mathscr{M}_{\rm ext} = \varnothing$. If C intersects $I^+(\mathscr{M}_{\rm ext})$ then $I^-(C) \cap \mathscr{M}_{\rm ext}$ is non-empty, hence $I^-(C) \supset \mathscr{M}_{\rm ext}$ since $\mathscr{M}_{\rm ext}$ is connected. A similar argument applies if C intersects $I^-(\mathscr{M}_{\rm ext})$. □

We have the following strengthening of Lemma 3.2:

Lemma 3.7. — *Let $\alpha_i \in \mathbb{R}$. If $(\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle, \mathfrak{g})$ is chronological, then there exists no nonempty set N which is invariant under the flow of $K_{(0)} + \sum_i \alpha_i K_i$ and which is included in a compact set $C \subset \langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$.*

Proof. — Assume that $N \subset \langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$ is not empty. From Lemma 3.6 we obtain $\mathscr{M}_{\rm ext} \subset I^+(N)$, hence $I^+(\mathscr{M}_{\rm ext}) \subset I^+(N)$. Arguing similarly with I^- we infer that

$$\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle \subset I^+(N) \cap I^-(N).$$

Hence every point q in $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$ is in $I^+(p)$ for some $p \in N$. We conclude that $\{I^+(p) \cap C\}_{p\in N}$ is an open cover of C. Assuming compactness, we may then choose a finite subcover $\{I^+(p_i) \cap C\}_{i=1}^I$. This implies that each p_i must be in the future of at least one p_j, and since there is a finite number of them one eventually gets a closed timelike curve, which is not possible in chronological space-times. □

Since each zero of a Killing vector provides a compact invariant set, from Lemma 3.7 we conclude

Corollary 3.8. — *Let $\alpha_i \in \mathbb{R}$. If $(\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle, \mathfrak{g})$ is chronological, then Killing vectors of the form $K_{(0)} + \sum_i \alpha_i K_i$ have no zeros in $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$*

4. Horizons and domains of outer communications in regular space-times

In this section we analyze the structure of a class of horizons, and of domains of outer communications.

4.1. Sections of horizons. — The aim of this section is to establish the existence of cross-sections of the event horizon with good properties.

By standard causality theory the future event horizon $\mathscr{H}^+ = \dot I^-(\mathscr{M}_{\rm ext})$ (recall that $\dot I^{\pm}$ denotes the boundary of $I^{\pm}$) is the union of Lipschitz topological hypersurfaces. Furthermore, through every point $p \in \mathscr{H}^+$ there is a future inextendible null geodesic entirely contained in $\mathscr{H}^+$ (though it may leave $\mathscr{H}^+$ when followed to the past of p). Such geodesics are called *generators*. A topological submanifold S of $\mathscr{H}^+$ will be called a *local section*, or simply *section*, if S meets the generators of $\mathscr{H}^+$ transversally; it will be called a *cross-section* if it meets all the generators precisely once. Similar definitions apply to any null achronal hypersurfaces, such as $\mathscr{H}^-$ or $\mathscr{E}^\pm$.

We start with the proof of existence of sections of the event horizon which are moved to their future by the isometry group. The existence of such sections has been claimed in Lemma 5.2 of [**16**]; here we give the proof of a somewhat more general result:

Proposition 4.1. — *Let $\mathscr{H}_0 \subset \mathscr{H} := \mathscr{H}^+ \cup \mathscr{H}^- \equiv \dot{I}^-(\mathscr{M}_{\rm ext}) \cup \dot{I}^+(\mathscr{M}_{\rm ext})$ be a connected component of the event horizon $\mathscr{H}$ in a space-time $(\mathscr{M}, \mathfrak{g})$ with stationary Killing vector $K_{(0)}$, and suppose that there exists a compact cross-section S of $\mathscr{H}_0$ satisfying*

$$S \subset \mathscr{E}_0 := \mathscr{H}_0 \cap I^+(\mathscr{M}_{\rm ext}).$$

Assume that

1. *either*
$$\overline{\langle\langle \mathscr{M}_{\rm ext} \rangle\rangle} \cap I^+(\mathscr{M}_{\rm ext}) \text{ is strongly causal,}$$
2. *or there exists in $\langle\langle \mathscr{M}_{\rm ext} \rangle\rangle$ a spacelike hypersurface $\mathscr{S} \supset \mathscr{S}_{\rm ext}$, achronal in $\langle\langle \mathscr{M}_{\rm ext} \rangle\rangle$, so that S above coincides with the boundary of $\overline{\mathscr{S}}$:*
$$S = \partial \overline{\mathscr{S}} \subset \mathscr{E}^+.$$

Then there exists a compact Lipschitz hypersurface S_0 of $\mathscr{E}_0$ which is transverse to both the stationary Killing vector field $K_{(0)}$ and to the generators of $\mathscr{E}_0$, and which meets every generator of $\mathscr{E}_0$ precisely once; in particular

$$\mathscr{E}_0 = \cup_t \phi_t(S_0).$$

Proof. — Changing time orientation if necessary, and replacing $\mathscr{M}$ by $I^+(\mathscr{M}_{\rm ext}) \setminus (\mathscr{H} \setminus \mathscr{H}_0)$, we can without loss of generality assume that $\mathscr{E} = \mathscr{E}_0 = \mathscr{H}_0 = \mathscr{H} = \mathscr{H}^+$. Choose a point $p \in \mathscr{M}_{\rm ext}$, where the Killing vector $K_{(0)}$ is timelike, and let

$$\gamma_p = \cup_{t\in\mathbb{R}} \phi_t(p)$$

be the orbit of $K_{(0)}$ through p. Then $I^-(S)$ must intersect γ_p (since $\mathscr{E}_0$ is contained in the future of $\mathscr{M}_{\rm ext}$). Further, $I^-(S)$ cannot contain all of γ_p, by Lemma 3.1 or by part 2 of Lemma 3.2. Let $q \in \gamma_p$ lie on the boundary of $I^-(S)$, then $I^+(q)$ cannot contain any point of S, so it does not contain any complete null generator of $\mathscr{E}_0$. On the other hand, if $I^+(q)$ failed to intersect some generator of $\mathscr{E}_0$, then (by invariance under the flow of $K_{(0)}$) each point of γ_p would also fail to intersect some generator. By considering a sequence, $\{q_n = \phi_{t_n}(q)\}$, along γ_p with $t_n \to -\infty$, one would obtain a corresponding sequence of horizon generators lying entirely outside the future of $\{q_n\}$. Using compactness, one would get an "accumulation generator" that lies outside the future of all $\{q_n\}$ and thus lies outside of $I^+(\gamma_p) = I^+(\mathscr{M}_{\rm ext})$, contradicting the fact that S lies to the future of $\mathscr{M}_{\rm ext}$.

Set
$$S_0 := \dot{I}^+(q) \cap \mathscr{E}_0,$$
and we have just proved that every generator of $\mathscr{E}_0$ intersects S_0 at least once.

The fact that the only null geodesics tangent to $\mathscr{E}_0$ are the generators of $\mathscr{E}_0$ shows that the generators of $\dot{I}^+(q)$ intersect $\mathscr{E}_0$ transversally. (Otherwise a generator of $\dot{I}^+(q)$ would become a generator, say Γ, of $\mathscr{E}_0$. Thus Γ would leave $\mathscr{E}_0$ when followed to the past at the intersection point of $\dot{I}^+(q)$ and $\mathscr{E}_0$, reaching q, which contradicts the fact that $\mathscr{E}_0$ lies at the boundary of $I^-(\mathscr{M}_{\rm ext})$.) As in [**22**], Clarke's Lipschitz implicit function theorem [**29**] shows now that S_0 is a Lipschitz submanifold intersecting each horizon generator; while the argument just given shows that it intersects each generator at most one point. Thus, S_0 is a cross-section with respect to the null generators. However, S_0 also is a cross-section with respect to the flow of $K_{(0)}$, because for all t we have
$$\phi_t(S_0) = \dot{I}^+(\phi_t(q)) \cap \mathscr{E},$$
and for $t > 0$ the boundary of $I^+(\phi_t(q))$ is contained within $I^+(q)$. In other words, $\phi_t(S_0)$ cannot intersect S_0, which is equivalent to saying that each orbit of the flow of $K_{(0)}$ on the horizon cannot intersect S_0 at more than one point. On the other hand, each orbit must intersect S_0 at least once by the type of argument already given — one will run into a contradiction if complete Killing orbits on the horizon are either contained within $I^+(q)$ or lie entirely outside of $I^+(q)$. □

Now, both S and S_0 are compact cross-sections of $\mathscr{E}_0$. Flowing along the generators of the horizon, one obtains:

Proposition 4.2. — *S is homeomorphic to S_0.*

We note that so far we only have a $C^{0,1}$ cross-section of the horizon, and in fact this is the best one can get at this stage, since this is the natural differentiability of $\mathscr{E}_0$. However, if $\mathscr{E}_0$ is smooth, we claim:

Proposition 4.3. — *Under the hypotheses of Proposition 4.1, assume moreover that $\mathscr{E}_0$ is smooth, and that $\langle\langle\mathscr{M}_{\rm ext}\rangle\rangle$ is globally hyperbolic. Then S_0 can be chosen to be smooth.*

Proof. — The result is obtained with the following regularization argument: Choose a point $p \in \mathscr{M}_{\rm ext}$, such that the section S of Proposition 4.1 does *not* intersect the future of p. Let the function u be the retarded time associated with the orbit γ_p through p parameterized by the Killing time from p; this is defined as follows: For any $q \in \mathscr{M}$ we consider the intersection $J^-(q) \cap \gamma_p$. If that intersection is empty we set $u(q) = \infty$. If $J^-(q)$ contains γ_p we set $u(q) = -\infty$. Otherwise, as $\dot{J}^-(q)$ is achronal, the set $\dot{J}^-(q) \cap \gamma_p$ contains precisely one point $\phi_\tau(p)$ for some τ. We then set $u(q) = \tau$. Note that, with appropriate conventions, this is the same as setting

$$u(q) = \inf\{t : \phi_t(p) \in J^-(q)\}. \tag{4.1}$$

It follows from the definition of u that we have, for all r,

$$u(\phi_t(r)) = u(r) + t. \tag{4.2}$$

In particular, u is differentiable in the direction tangent to the orbits of $K_{(0)}$, with

$$K_{(0)}(u) = \mathfrak{g}(K_{(0)}, \nabla u) = 1, \tag{4.3}$$

everywhere.

The proof of Proposition 4.1 shows that u is finite in a neighborhood of $\mathscr{E}_0$; let

$$S_0 = u^{-1}(0) \cap \mathscr{E}_0,$$

and let $\mathscr{O}$ denote a conditionally compact neighborhood of S_0 on which u is finite; note that S_0 here is a $\phi_t[K_{(0)}]$–translate of the section S_0 of Proposition 4.1.

Let n be the field of future directed tangents to the generators of $\mathscr{E}_0$, normalized to unit length with some auxiliary smooth Riemannian metric on $\mathscr{M}$. For $q \in S_0$ let $\mathscr{N}_q \subset T_q\mathscr{M}$ denote the collection of all similarly normalized null vectors that are tangent to an achronal past directed null geodesic γ from q to $\phi_{u(q)}(p)$, with γ contained in $\langle\langle \mathscr{M}_{\rm ext} \rangle\rangle$ except for its initial point. (If u is differentiable at q then $\mathscr{N}_q$ contains one single element, proportional to ∇u, but $\mathscr{N}_q$ can contain more than one null vector in general.) We claim that there exists $c > 0$ such that

$$\inf_{q\in S_0, l_q \in \mathscr{N}_q} \mathfrak{g}(l_q, n_q) \geq c > 0. \tag{4.4}$$

Indeed, suppose that this is not the case; then there exists a sequence $q_i \in S_0$ and a sequence of past directed null achronal geodesic segments γ_i from q_i to p, with tangents l_i at q_i, such that $\mathfrak{g}(l_i, n) \to 0$. Compactness of S_0 implies that there exists $q \in S_0$ such that $q_i \to q$.

Let γ be an accumulation curve of the γ_i's passing through q. By hypothesis, $\mathscr{E}_0$ is a smooth null hypersurface contained in the boundary of $\langle\langle \mathscr{M}_{\rm ext} \rangle\rangle$, with $q \in \mathscr{E}_0$. This implies that either γ immediately enters $\langle\langle \mathscr{M}_{\rm ext} \rangle\rangle$, or γ is a subsegment of a generator of $\mathscr{E}_0$ through q. In the latter case γ intersects S when followed from q towards the past, and therefore the γ_i's intersect $\dot{J}^-(S) \cap \langle\langle \mathscr{M}_{\rm ext} \rangle\rangle$ for all i large enough. But this is not possible since $S \cap J^+(p) = \varnothing$. We conclude that there exists $s_0 > 0$ such that $\gamma(s_0) \in \langle\langle \mathscr{M}_{\rm ext} \rangle\rangle$. Thus a subsequence, still denoted by $\gamma_i(s_0)$, converges to $\gamma(s_0)$, and global hyperbolicity of $\langle\langle \mathscr{M}_{\rm ext} \rangle\rangle$ implies that the γ_i's converge to an achronal null geodesic segment γ through p, with tangent l at S_0 satisfying $\mathfrak{g}(l, n) = 0$. Since both

l and n are null we conclude that l is proportional to n, which is not possible as the intersection must be transverse, providing a contradiction, and establishing (4.4).

Let $\mathscr{O}_i$, $i = 1, \dots, N$, be a family of coordinate balls of radii $3r_i$ such that the balls of radius r_i cover $\overline{\mathscr{O}}$, and let φ_i be an associated partition of unity; by this we mean that the φ_i's are supported in $\mathscr{O}_i$, and they sum to one on $\mathscr{O}$. For $\epsilon \leq r := \min r_i$ let $\varphi_\epsilon(x) = \epsilon^{-n-1}\varphi(x/\epsilon)$ (recall that the dimension of $\mathscr{M}$ is $n+1$), where φ is a positive smooth function supported in the ball of radius one, with integral one. Set

$$u_\epsilon := \sum_{i=1}^{N} \varphi_i \, \varphi_\epsilon * u, \tag{4.5}$$

where $*$ denotes a convolution in local coordinates. Strictly speaking, φ_ϵ should be denoted by $\varphi_{\epsilon,i}$, as it depends explicitly on the local coordinates on $\mathscr{O}_i$, but we will not overburden the notation with yet another index. (10) Then u_ϵ tends uniformly to u. Further, using the Stokes theorem for Lipschitz functions [**75**],

$$\begin{aligned} du_\epsilon &= \sum_{i=1}^{N} \Big\{ \varphi_\epsilon * u \, d\varphi_i + \varphi_i \, \varphi_\epsilon * du \Big\} \\ &= \sum_{i=1}^{N} \Big\{ \underbrace{(\varphi_\epsilon * u - u)}_{I} \, d\varphi_i + \varphi_i \, \underbrace{\varphi_\epsilon * du}_{II} \Big\}, \end{aligned} \tag{4.6}$$

where we have also used $\sum_i d\varphi_i = d\sum_i \varphi_i = d1 = 0$. It immediately follows that the term I uniformly tends to zero as ϵ goes to zero. Now, the term II, when contracted with $K_{(0)}$, gives a contribution

$$\begin{aligned} i_{K_{(0)}}(\varphi_\epsilon * du)(x) &= \int_{|y-x|\leq\epsilon} K^i_{(0)}(x)\, \partial_i u(y) \varphi_\epsilon(x-y) d^{n+1}y \\ &= \int_{|y-x|\leq\epsilon} \Big[\underbrace{(K^i_{(0)}(x) - K^i_{(0)}(y))}_{=O(\epsilon)} \partial_i u(y) \\ &\qquad + \underbrace{K^i_{(0)}(y)\partial_i u(y)}_{=1 \text{ by } (4.3)} \Big] \varphi_\epsilon(x-y) d^{n+1}y \\ &= 1 + O(\epsilon). \end{aligned} \tag{4.7}$$

It follows that, for all ϵ small enough, the differential du_ϵ is nowhere vanishing, and that $K_{(0)}$ is transverse to the level sets of u_ϵ.

To conclude, let n denote any future directed causal smooth vector field on $\mathscr{O}$ which coincides with the field of tangents to the null generators of $\mathscr{E}_0$ as defined

(10) This is admittedly somewhat confusing since, e.g., $\sum_{i=1}^{N} \varphi_i \, \varphi_\epsilon * u \neq (\sum_{i=1}^{N} \varphi_i) \, \varphi_\epsilon * u$.

above. By (4.4) the terms II in the formula for du_ϵ, when contracted with n, will give a contribution

(4.8)
$$\begin{aligned} i_n(\varphi_\epsilon * du)(x) &= \int_{|y-x|\le\epsilon} [\underbrace{(n^i(x)-n^i(y))}_{=O(\epsilon)}\partial_i u(y) + \underbrace{n^i(y)\partial_i u(y)}_{\ge c}]\varphi_\epsilon(x-y)d^{n+1}y \\ &\ge c+O(\epsilon), \end{aligned}$$

and transversality of the generators of $\mathscr{E}_0$ to the level sets of u_ϵ, for ϵ small enough, follows. □

4.2. The structure of the domain of outer communications. — The aim of this section is to establish the product structure of I^+–regular domains of outer communication, Theorem 4.5 below. The analysis here is closely related to that of [**27**].

As in Section 3, we assume the existence of a commutative group of isometries $\mathbb{R}\times\mathbb{T}^{s-1}$ with $s\ge 1$. We use the notation there, with $K_{(0)}$ timelike in $\mathscr{M}_{\text{ext}}$, and each $K_{(i)}$ spacelike in $\langle\langle\mathscr{M}_{\text{ext}}\rangle\rangle$.

Let $r=\sqrt{\sum_i (x^i)^2}$ be the radius function in $\mathscr{M}_{\text{ext}}$. By the asymptotic analysis of [**25**] there exists R so that for $r\ge R$ the orbits of the $K_{(i)}$'s are entirely contained in $\mathscr{M}_{\text{ext}}$, so that the function

$$\hat r(p)=\int_{g\in\mathbb{T}^{s-1}} r(g(p))d\mu_g,$$

is well defined, and invariant under $\mathbb{T}^{s-1}$. Here $d\mu_g$ is the translation invariant measure on $\mathbb{T}^{s-1}$ normalized to total volume one, and $g(p)$ denotes the action on $\mathscr{M}$ of the isometry group generated by the $K_{(i)}$'s. Similarly, let t be any time function on $\langle\langle\mathscr{M}_{\text{ext}}\rangle\rangle$, the level sets of which are asymptotically flat Cauchy surfaces. Averaging over $\mathbb{T}^{s-1}$ as above, we obtain a new time function $\hat t$, with asymptotically flat level sets, which is invariant under $\mathbb{T}^{s-1}$. (The interesting question, whether or not the level sets of $\hat t$ are Cauchy, is irrelevant for our further considerations here.) It is then easily seen that, for σ large enough, the level sets

$$\hat S_{\tau,\sigma} := \{\hat t=\tau, \hat r=\sigma\}$$

are smooth embedded spheres included in $\mathscr{M}_{\text{ext}}$.

Throughout this section we assume that $(\mathscr{M},\mathfrak{g})$ is I^+–regular. Let $\mathscr{S}$ be as in the definition of regularity, thus $\mathscr{S}$ is an asymptotically flat spacelike acausal hypersurface in $\langle\langle\mathscr{M}_{\text{ext}}\rangle\rangle$ with compact boundary, the latter coinciding with a compact cross-section of $\mathscr{E}^+$. Deforming $\mathscr{S}$ if necessary, without loss of generality we may assume that $\mathscr{S}\cap\mathscr{M}_{\text{ext}}$ is a level set of $\hat t$. We choose R large enough so that $\hat S_{0,R}$ is a smooth sphere, and so that the slopes of light cones on the $\hat{S}_{\tau,\sigma}$'s, for $\sigma \geq R$, are bounded from above by two, and from below by one half, and redefine $\mathscr{S}_{\text{ext}}$ so that $\partial \mathscr{S}_{\text{ext}} = \hat{S}_{0,R}$.

Consider

$$\mathscr{C}^+ := (\dot{J}^+(\hat{S}_{0,R}) \setminus \mathscr{M}_{\text{ext}}) \cap \langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle .$$

Then $\mathscr{C}^+$ is a null, achronal, Lipschitz hypersurface generated by null geodesics initially orthogonal to $\hat{S}_{0,R}$. Let us write ϕ_t for $\phi_t[K_{(0)}]$, and set

$$\mathscr{C}^+_t := \phi_t(\mathscr{C}^+) \;;$$

we then have

$$\mathscr{C}^+_t := (\dot{J}^+(\hat{S}_{t,R}) \setminus \mathscr{M}_{\text{ext}}) \cap \langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle ,$$

(recall that the flow of $K_{(0)}$ consists of translations in t in $\mathscr{M}_{\text{ext}}$) which implies that every orbit of $K_{(0)}$ intersects $\mathscr{C}^+$ at most once.

Since $\mathscr{S}$ is achronal it partitions $\langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle$ as

$$\langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle = \mathscr{S} \cup I^+(\mathscr{S}; \langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle) \cup I^-(\mathscr{S}; \langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle) \text{ (disjoint union)}. \tag{4.9}$$

Indeed, as $\langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle$ is globally hyperbolic, the boundaries $(\dot{I}^\pm(\mathscr{S}) \setminus \mathscr{S}) \cap \langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle$ are generated by null geodesics with end points on $\text{edge}(\mathscr{S}) \cap \langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle = \varnothing$.

We claim that every orbit of $K_{(0)}$ intersects $\mathscr{S}$. For this, recall that for any q in $\langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle$ there exist points $p_\pm \in \mathscr{M}_{\text{ext}}$ such that $q \in I^\mp(p_\pm)$. Since the flow of $K_{(0)}$ in $\mathscr{M}_{\text{ext}}$ is by time translations there exist $t_\pm \in \mathbb{R}$ so that $\phi_{t_\pm}(p_\pm) \in \mathscr{S}_{\text{ext}}$. Hence $\phi_{t_\pm}(q) \in I^\mp(\mathscr{S}_{\text{ext}})$, which shows that every orbit of $K_{(0)}$ meets both the future and the past of $\mathscr{S}$. By continuity and (4.9) every orbit meets $\mathscr{S}$ (perhaps more than once). Hence

$$\langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle = \cup_t \phi_t(\mathscr{S}), \quad \overline{\langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle} \cap I^+(\mathscr{M}_{\text{ext}}) = \cup_t \phi_t(\overline{\mathscr{S}}) \tag{4.10}$$

(for the second equality Proposition 4.1 has been used). Setting $\mathscr{M}_{\text{int}} = \langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle \setminus \mathscr{M}_{\text{ext}}$, one similarly obtains

$$\mathscr{M}_{\text{int}} = \mathscr{C}^+ \cup I^+(\mathscr{C}^+; \mathscr{M}_{\text{int}}) \cup I^-(\mathscr{C}^+; \mathscr{M}_{\text{int}}) \text{ (disjoint union)}, \tag{4.11}$$

$$\mathscr{M}_{\text{int}} = \cup_t \phi_t(\mathscr{C}^+). \tag{4.12}$$

By hypothesis $\overline{\mathscr{S} \setminus \mathscr{S}_{\text{ext}}}$ is compact and so, by the first part of Lemma 3.2, there exists $p_- \in \mathscr{M}_{\text{ext}}$ such that

$$\overline{\mathscr{S} \setminus \mathscr{S}_{\text{ext}}} \subset I^+(p_-). \tag{4.13}$$

Choose $t_- < 0$ so that $p_- \in I^+(\hat{S}_{t_-,R})$; we obtain that $\overline{\mathscr{S} \setminus \mathscr{S}_{\text{ext}}} \subset I^+(\hat{S}_{t_-,R})$, hence

$$\overline{\mathscr{S} \setminus \mathscr{S}_{\text{ext}}} \subset I^+(\mathscr{C}^+_{t_-}).$$

Since $\hat{S}_{0,R} \subset \mathscr{S}$ we have $\mathscr{C}^+ \subset I^+(\mathscr{S})$. By acausality of $\mathscr{S}$ and (4.9) we infer that $\overline{\mathscr{S} \setminus \mathscr{S}_{\text{ext}}} \subset I^-(\mathscr{C}^+)$, and hence $\phi_{t_-}(\overline{\mathscr{S} \setminus \mathscr{S}_{\text{ext}}}) \subset I^-(\mathscr{C}^+_{t_-})$.

So, for $p \in \overline{\mathscr{S} \setminus \mathscr{S}_{\mathrm{ext}}}$ the orbit segment

$$[t_-, 0] \ni t \mapsto \phi_t(p)$$

starts in the past of $\mathscr{C}^+_{t_-}$ and finishes to its future. From (4.10) we conclude that

$$\overline{\mathscr{C}^+_{t_-}} \subset \cup_{t \in [t_-, 0]} \phi_t(\overline{\mathscr{S} \setminus \mathscr{S}_{\mathrm{ext}}}) \; ; \tag{4.14}$$

equivalently,

$$\overline{\mathscr{C}^+} \subset \cup_{t \in [0, -t_-]} \phi_t(\overline{\mathscr{S} \setminus \mathscr{S}_{\mathrm{ext}}}).$$

As the set at the right-hand-side is compact, we have established:

Proposition 4.4. — *Suppose that $(\mathscr{M}, \mathfrak{g})$ is I^+-regular, then $\overline{\mathscr{C}^+}$ is compact.*

We are ready to prove now the following version of point 2 of Lemma 5.1 of [**16**]:

***Theorem 4.5* (Structure theorem)**. — *Suppose that $(\mathscr{M}, \mathfrak{g})$ is an I^+-regular stationary space-time invariant under a commutative group of isometries $\mathbb{R} \times \mathbb{T}^{s-1}$, $s \geq 1$, with the stationary Killing vector $K_{(0)}$ tangent to the orbits of the $\mathbb{R}$ factor. There exists on $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$ a smooth time function t, invariant under $\mathbb{T}^{s-1}$, which together with the flow of $K_{(0)}$ induces the diffeomorphisms*

$$\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle \approx \mathbb{R} \times \mathring{\mathscr{S}}, \qquad \overline{\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle} \cap I^+(\mathscr{M}_{\mathrm{ext}}) \approx \mathbb{R} \times \overline{\mathring{\mathscr{S}}}, \tag{4.15}$$

where $\mathring{\mathscr{S}} := t^{-1}(0)$ is asymptotically flat, (invariant under $\mathbb{T}^{s-1}$), with the boundary $\partial\overline{\mathring{\mathscr{S}}}$ being a compact cross-section of $\mathscr{E}^+$. The smooth hypersurface with boundary $\overline{\mathring{\mathscr{S}}}$ is acausal, spacelike up-to-boundary, and the flow of $K_{(0)}$ is a translation along the $\mathbb{R}$ factor in (4.15).

Proof. — From what has been said, every orbit of $K_{(0)}$ through $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle \setminus \mathscr{M}_{\mathrm{ext}}$ intersects $\mathscr{C}^+$ precisely once. For $p \in \langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle \setminus \mathscr{M}_{\mathrm{ext}}$ we let $u(p)$ be the unique real number such that $\phi_{u(p)}(p) \in \mathscr{C}^+$, while for $p \in \mathscr{M}_{\mathrm{ext}}$ we let $u(p)$ be the unique real number such that $\phi_{u(p)}(p) \in \mathscr{S}_{\mathrm{ext}}$. The function $u : \langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle \to \mathbb{R}$ is Lipschitz, smooth in $\mathscr{M}_{\mathrm{ext}}$, with achronal level sets transverse to the flow of $K_{(0)}$, and provides a homeomorphism

$$\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle \setminus \mathscr{M}_{\mathrm{ext}} \approx \mathbb{R} \times \mathscr{C}^+, \qquad \langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle \approx \mathbb{R} \times (\mathscr{C}^+ \cup \mathscr{S}_{\mathrm{ext}}).$$

The desired hypersurface $\mathring{\mathscr{S}}$ will be a small spacelike smoothing of $u^{-1}(0)$, obtained by first deforming the metric $\mathfrak{g}$ to a metric $\mathfrak{g}_\epsilon$, the null vectors of which are spacelike for $\mathfrak{g}$. The associated corresponding function u_ϵ will have Lipschitz level sets which are uniformly spacelike for $\mathfrak{g}$. A smoothing of u_ϵ will provide the desired function t. The details are as follows:

We start by finding a smooth hypersurface, not necessarily spacelike, transverse to the flow of K. We shall use the following general result, pointed out to us by R. Wald (private communication):

Proposition 4.6. — *Let S_0 be a two-sided, smooth, hypersurface in a manifold M with an open neighborhood $\mathscr{O}$ such that $M \setminus \mathscr{O}$ consists of two disconnected components M_- and M_+. Let X be a complete vector field on M and suppose that there exists $T > 0$ such that for every orbit $\phi_t(p)$ of X, $t \in \mathbb{R}$, $p \in M$, there is an interval $[t_0, t_1]$ with $(t_1 - t_0) < T$ such that $\phi_t(p)$ lies in M_- for all $t < t_0$, and $\phi_t(p)$ lies in M_+ for all $t > t_1$. If M has a boundary, assume moreover that $\partial S_0 \subset \partial M$, and that X is tangent to ∂M. Then there exists a smooth hypersurface $S_1 \subset M$ such that every orbit of X intersects S_1 once and only once.*

Proof. — Let f be a smooth function with the property that $f = 0$ in M_-, $0 \le f \le 1$ in $\mathscr{O}$, and $f = 1$ in M_+; such a function is easily constructed by introducing Gauss coordinates, with respect to some auxiliary Riemannian metric, near S_0. For $t \in \mathbb{R}$ and $p \in M$ let $\phi_t(p)$ denote the flow generated by X. Define $F : M \to \mathbb{R}$ by

$$F(p) = \int_{-\infty}^{0} f \circ \phi_s(p) ds.$$

Then F is a smooth function on M increasing monotonically from zero to infinity along every orbit of X. Furthermore F is strictly increasing along the orbits at points at which $F \ge T$ (since such points must lie in M_+, where $f = 1$). In particular, the gradient of F is non-vanishing at all points where $F \ge T$. Setting $S_1 = \{F = T\}$, the result follows. □

Returning to the proof of Theorem 4.5, we use Proposition 4.6 with $X = K_{(0)}$,

$$M = \overline{\overline{\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle} \cap I^+(\mathscr{M}_{\mathrm{ext}}) \setminus \mathscr{M}_{\mathrm{ext}}},$$

and $S_0 = \mathscr{S} \cap M$. Letting t_- be as in (4.14) we set

$$\mathscr{O} := \cup_{t \in (t_-, -t_-)} \phi_t(\mathscr{S}) \ ;$$

by what has been said, $\mathscr{O}$ is an open neighborhood of $\mathscr{S}$. Finally

$$M_- := \cup_{t \in (-\infty, t_-]} \phi_t(\mathscr{S}), \quad M_+ := \cup_{t \in [-t_-, \infty)} \phi_t(\mathscr{S}).$$

It follows now from Proposition 4.6 that there exists a hypersurface $S_1 \subset M$ which is transverse to the flow of $K_{(0)}$.

Let $\hat{T}$ be any smooth, timelike vector field defined along S_1, and define the smooth timelike vector field T on M as the unique solution of the Cauchy problem

$$\mathscr{L}_{K_{(0)}} T = 0, \quad T = \hat{T} \text{ on } S_1. \tag{4.16}$$

Since the flow of $K_{(0)}$ acts by time translations on $\mathscr{M}_{\mathrm{ext}}$, it is straightforward to extend T to a smooth vector field defined on $\mathscr{M}$, timelike wherever non vanishing, still denoted by T, which is invariant under the flow of $K_{(0)}$, the support of which on $\mathscr{S}$ is compact. Replacing T by its average over $\mathbb{T}^{s-1}$, we can assume that T is invariant under the action of $\mathbb{T}^{s-1}$.

For all $\epsilon \geq 0$ sufficiently small, the formula

$$\mathfrak{g}_\epsilon(Z_1, Z_2) = \mathfrak{g}(Z_1, Z_2) - \epsilon \mathfrak{g}(T, Z_1)\mathfrak{g}(T, Z_2). \tag{4.17}$$

defines a Lorentzian, $\mathbb{R} \times \mathbb{T}^{s-1}$ invariant metric on the manifold with ($\mathfrak{g}_\epsilon$–timelike) boundary $\overline{\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle} \cap I^+(\mathscr{M}_{\rm ext})$. By definition of $\mathfrak{g}_\epsilon$, vectors which are causal for $\mathfrak{g}$ are timelike for $\mathfrak{g}_\epsilon$. Wherever $T \neq 0$ the light cones of $\mathfrak{g}_\epsilon$ are spacelike for $\mathfrak{g}$, provided $\epsilon \neq 0$.

Since $\mathfrak{g}$-causal curves are also $\mathfrak{g}_\epsilon$-causal, $(\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle, \mathfrak{g}_\epsilon)$ is also a domain of outer communications with respect to $\mathfrak{g}_\epsilon$.

Set

$$\mathscr{C}^+_\epsilon = (\dot J^+_\epsilon(\hat S_{0,R}) \setminus \mathscr{M}_{\rm ext}) \cap \langle\langle \mathscr{M}_{\rm ext}\rangle\rangle,$$

where we denote by $J^+_\epsilon(\Omega)$ the future of a set Ω with respect to the metric $\mathfrak{g}_\epsilon$. Then the $\mathscr{C}^+_\epsilon$'s are Lipschitz, $\mathfrak{g}$-spacelike wherever differentiable, $\mathbb{T}^{s-1}$ invariant, hypersurfaces. Continuous dependence of geodesics upon the metric together with Proposition 4.4 shows that the $\mathscr{C}^+_\epsilon$'s accumulate at $\mathscr{C}^+$ as ϵ tends to zero.

Let $u_\epsilon : M \to \mathbb{R}$ be defined as in (4.1) using the metric $\mathfrak{g}_\epsilon$ instead of $\mathfrak{g}$. As before we have

$$u_\epsilon(\phi_t(p)) = u_\epsilon(p) + t, \text{ so that } K_{(0)}(u_\epsilon) = 1. \tag{4.18}$$

We perform a smoothing procedure as in the proof of Proposition 4.3, with $\mathscr{O}$ there replaced by a conditionally compact neighborhood of $\mathscr{C}^+$. The vector field $\hat T$ in (4.16) is chosen to be timelike on $\overline{\mathscr{O}}$; the same will then be true of T. Analogously to (4.5) we set

$$u_{\epsilon,\eta} := \sum_{i=1}^N \varphi_i \, \varphi_\eta * u_\epsilon, \tag{4.19}$$

so that the $u_{\epsilon,\eta}$'s converge uniformly on $\mathscr{O}$ to u_ϵ as η tends to zero. The calculation in (4.7) shows that

$$K_{(0)}(u_{\epsilon,\eta}) \geq \frac{1}{2}$$

for η small enough, so that the level sets of $u_{\epsilon,\eta}$ near $\mathscr{C}^+$ are transverse to the flow of $K_{(0)}$.

It remains to show that the level sets of $u_{\epsilon,\eta}$ are spacelike. For this we start with some lemmata:

Lemma 4.7. — *Let $\mathfrak{g}$ be a Lipschitz-continuous metric on a coordinate ball $B(p, 3r_i) \equiv \mathscr{O}_i$ of coordinate radius $3r_i$. There exists a constant C such that for any $q \in B(p, r_i)$ and for any timelike, respectively causal, vector $N_q = N^\mu_q \partial_\mu \in T_q\mathscr{M}$ satisfying*

$$\sum_\mu (N^\mu_q)^2 = 1 \tag{4.20}$$

there exists a timelike, respectively causal, vector field $N = N^\mu \partial_\mu$ *on* $B(p, 2r_i)$ *such that for all points* $y, z \in B(p, 2r_i)$ *we have*

$$|N^\mu_y - N^\mu_z| \le C|y-z|, \quad C^{-1} \le \sum_\mu (N^\mu_y)^2 \le C. \tag{4.21}$$

Proof. — We will write both N^μ_q and $N^\mu(q)$ for the coordinate components of a vector field at q. For $\nu = 0, \ldots, n$, let $e_{(\nu)} = e^\mu_{(\nu)}\partial_\mu$ be any Lipschitz-continuous ON basis for $\mathfrak{g}$ on $\mathscr{O}_i$. there exists a constant c such that on $B(p, 2r_i)$ we have

$$|e^\mu_{(\nu)}(y) - e^\mu_{(\nu)}(z)| \le c|y-z|.$$

Decompose N_q as $N_q = N^{(\nu)}_q e_{(\nu)}(q)$, and for $y \in \mathscr{O}_i$ set $N_y = N^{(\nu)}_q e_{(\nu)}(y)$; (4.21) easily follows. □

Lemma 4.8. — *Under the hypotheses of Lemma 4.7, let* f *be differentiable on* $\mathscr{O}_i$. *Then* ∇f *is timelike past directed on* $B(p, 2r_i)$ *if and only if* $N^\mu \partial_\mu f < 0$ *on* $\mathscr{O}_i$ *for all causal past directed vector fields satisfying* (4.20) *and* (4.21).

Proof. — The condition is clearly necessary. For sufficiency, suppose that there exists $q \in B(p, 2r_i)$ such that ∇f is null, let $N_q = \lambda \nabla f(q)$, where λ is chosen so that (4.20) holds, and let N be as in Lemma 4.7; then $N^\mu \partial_\mu f$ vanishes at q. If ∇f is spacelike at q the argument is similar, with N_q chosen to be any timelike vector orthogonal to $\nabla f(q)$ satisfying (4.20). □

Let N be any $\mathfrak{g}$–timelike past directed vector field satisfying (4.20) and (4.21). Returning to (4.6) we find,

$$i_N du_{\epsilon,\eta} = \sum_{i=1}^N \Big\{ \underbrace{(\varphi_\eta * u_\epsilon - u_\epsilon)}_{I} \; i_N d\varphi_i + \varphi_i \; \underbrace{i_N(\varphi_\eta * du_\epsilon)}_{II} \Big\}. \tag{4.22}$$

For any fixed ϵ, and for any $\delta > 0$ we can choose η_δ so that the term I is smaller than δ for all $0 < \eta < \eta_\delta$.

To obtain control of II, we need uniform spacelikeness of du_ϵ:

Lemma 4.9. — *There exists a constant* c *such that, for* N *as in Lemma 4.7,*

$$N^\mu \partial_\mu u_\epsilon < -c\epsilon \tag{4.23}$$

almost everywhere, for all $\epsilon > 0$ *sufficiently small.*

Proof. — Let $\{e_{(\nu)}\}$ be an $\mathfrak{g}$–ON frame in which the vector field T of (4.17) equals $T^{(0)}e_{(0)}$. Let $\alpha_{(\nu)}$ denote the components of du_ϵ in a frame dual to $\{e_{(\nu)}\}$. In this frame we have

$$\mathfrak{g} = \mathrm{diag}(-1, 1, \ldots, 1), \qquad \mathfrak{g}_\epsilon = \mathrm{diag}(-(1 + (T^{(0)})^2\epsilon), 1, \ldots, 1).$$

Since du_ϵ is $\mathfrak{g}_\epsilon$–null and past pointing we have

$$\alpha_{(0)} = \sqrt{1+(T^{(0)})^2\epsilon}\sqrt{\sum \alpha_{(i)}^2}.$$

The last part of (4.18) reads

$$K_{(0)}^{(0)}\alpha_{(0)} + K_{(0)}^{(i)}\alpha_{(i)} = 1.$$

It is straightforward to show from these two equations that there exists a constant c_1 such that, for all ϵ sufficiently small,

$$\alpha_{(0)} > c_1^{-1}, \quad \sqrt{\sum \alpha_{(i)}^2} > c_1^{-1}, \quad \sum |\alpha_{(\mu)}| \le c_1.$$

Since N is $\mathfrak{g}_\epsilon$ causal past directed, (4.20) and (4.21) together with the construction of N show that there exists a constant c_2 such that

$$N^{(0)} < -c_2.$$

We then have

$$\begin{aligned}
N^\mu\partial_\mu u_\epsilon &= N^{(0)}\alpha_{(0)} + N^{(i)}\alpha_{(i)}\\
&= N^{(0)}\sqrt{1+(T^{(0)})^2\epsilon}\sqrt{\sum \alpha_{(i)}^2} + N^{(i)}\alpha_{(i)}\\
&= N^{(0)}(\sqrt{1+(T^{(0)})^2\epsilon}-1)\sqrt{\sum \alpha_{(i)}^2} + \underbrace{N^{(0)}\sqrt{\sum \alpha_{(i)}^2} + N^{(i)}\alpha_{(i)}}_{<0 \text{ by Cauchy-Schwarz, as } N \text{ is } \mathfrak{g}\text{–timelike}}\\
&< -\frac{c_2}{4c_1}\inf_{\overline{\mathscr{O}}}(T^{(0)})^2\,\epsilon =: -c\epsilon,
\end{aligned}$$

for ϵ small enough. □

Now, calculating as in (4.8), using (4.23),

$$\begin{aligned}
i_N(\varphi_\eta * du_\epsilon)(x) &= \int_{|y-x|\le\eta} [\underbrace{(N^\mu(x)-N^\mu(y))}_{\le C\eta}\partial_\mu u_\epsilon(y) + \underbrace{N^\mu(y)\partial_\mu u_\epsilon(y)}_{\le -c\epsilon}]\varphi_\eta(x-y)d^{n+1}y\\
&\le -c\epsilon + O(\eta),
\end{aligned}$$

so that for η small enough each such term will give a contribution to (4.22) smaller than $-c\epsilon/2$. Timelikeness of $\nabla u_{\epsilon,\eta}$ on $\overline{\mathscr{O}}$ follows now from Lemma 4.8.

Summarizing, we have shown that we can choose ϵ and η small enough so that the function $u_{\epsilon,\eta} : M \to \mathbb{R}$ is a time function near its zero level set. It is rather straightforward to extend $u_{\epsilon,\eta}$ to a function on $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle \to \mathbb{R}$, with smooth spacelike zero-level-set, which coincides with $\mathscr{S}$ at large distances. Letting $\mathring{\mathscr{S}}$ be this zero level set, the function $t(p)$ is defined now as the unique value of parameter t so that $\phi_t(p) \in \mathring{\mathscr{S}}$; since the level sets of t are smooth spacelike hypersurface, t is a smooth time function. This completes the proof of Theorem 4.5. □

4.3. Smoothness of event horizons. — The starting point to any study of event horizons in stationary space-times is a corollary to the area theorem, essentially due to [**22**], which shows that event horizons in well-behaved stationary space-times are as smooth as the metric allows. In order to proceed, some terminology from that last reference is needed; we restrict ourselves to asymptotically flat space-times; the reader is referred to [**22**, Section 4] for the general case. Let $(\widetilde{\mathscr{M}}, \widetilde{\mathfrak{g}})$ be a C^3 completion of $(\mathscr{M}, \mathfrak{g})$ obtained by adding a null conformal boundary at infinity, denoted by $\mathscr{I}^+$, to $\mathscr{M}$, such that $\mathfrak{g} = \Omega^{-2}\widetilde{\mathfrak{g}}$ for a non-negative function Ω defined on $\widetilde{\mathscr{M}}$, vanishing precisely on $\mathscr{I}^+$, and $d\Omega$ without zeros on $\mathscr{I}^+$. Let $\mathscr{E}^+$ be the future event horizon in $\mathscr{M}$. We say that $(\widetilde{\mathscr{M}}, \widetilde{\mathfrak{g}})$ is $\mathscr{E}^+$–regular if there exists a neighborhood $\mathscr{O}$ of $\mathscr{E}^+$ such that for every compact set $C \subset \mathscr{O}$ for which $I^+(C; \widetilde{\mathscr{M}}) \neq \varnothing$ there exists a generator of $\mathscr{I}^+$ intersecting $\overline{I^+(C; \widetilde{\mathscr{M}})}$ which leaves this last set when followed to the past. (Compare Remark 4.4 and Definition 4.3 in [**22**]).

We note the following:

Proposition 4.10. — *Consider an asymptotically flat stationary space-time which is vacuum at large distances, recall that $\mathscr{E}^+ = \dot{I}^-(\mathscr{M}_{\rm ext}) \cap I^+(\mathscr{M}_{\rm ext})$. If $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$ is globally hyperbolic, then $(\mathscr{M}, \mathfrak{g})$ admits an $\mathscr{E}^+$–regular conformal completion.*

Proof. — Let $\widetilde{\mathscr{M}}$ be obtained by adding to $\mathscr{M}_{\rm ext}$ the surface $\tilde{r} = 0$ in the coordinate system $(u, \tilde{r}, \theta, \varphi)$ of [**34**, Appendix A] (see also [**32**], where the construction of [**34**] is corrected; those results generalize without difficulty to higher dimensions). Let t be any time function on $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$ which tends to infinity when $\mathscr{E}^+$ is approached, which tends to $-\infty$ when $\dot{I}^+(\mathscr{M}_{\rm ext})$ is approached, and which coincides with the coordinate t in $\mathscr{M}_{\rm ext}$ as in [**34**, Appendix A]. Let

$$\mathscr{O} = \{p \mid t(p) > 0\} \cup I^+(\mathscr{E}^+) \cup \mathscr{E}^+ \; ;$$

then $\mathscr{O}$ forms an open neighborhood of $\mathscr{E}^+$. Let C be any compact subset of $\mathscr{O}$ such that $I^+(C; \widetilde{\mathscr{M}}) \cap \mathscr{I}^+ \neq \varnothing$; then $\varnothing \neq C \cap \langle\langle \mathscr{M}_{\rm ext}\rangle\rangle \subset \{t > 0\}$. Let γ be any future directed causal curve from C to $\mathscr{I}^+$, then γ is entirely contained in $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$, with $t \circ \gamma > 0$. In particular any intersection of γ with $\partial\mathscr{M}_{\rm ext}$ belongs to the set $\{t > 0\}$, so that at each intersection point

$$u \circ \gamma > \inf u|_{\{t=0\}\cap\partial\mathscr{M}_{\rm ext}} =: c > -\infty.$$

The coordinate u of [**34**, Appendix A] is null, hence non-increasing along causal curves, so $u \circ \gamma > c$, which implies the regularity condition. □

We are ready to prove now:

Theorem 4.11. — *Let $(\mathscr{M}, \mathfrak{g})$ be a smooth, asymptotically flat, $(n+1)$–dimensional space-time with stationary Killing vector $K_{(0)}$, the orbits of which are complete. Suppose that $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$ is globally hyperbolic, vacuum at large distances in the asymptotic* *region, and assume that the* null energy condition (2.16) *holds. Assume that a connected component* $\mathscr{H}_0$ *of*

$$\mathscr{H} := \mathscr{H}^- \cup \mathscr{H}^+$$

admits a compact cross-section satisfying $S \subset I^+(\mathscr{M}_{\mathrm{ext}})$. *If*

1. *either*
$$\overline{\langle\langle \mathscr{M}_{\mathrm{ext}}\rangle\rangle} \cap I^+(\mathscr{M}_{\mathrm{ext}}) \textit{ is strongly causal,}$$
2. *or there exists in* $\langle\langle \mathscr{M}_{\mathrm{ext}}\rangle\rangle$ *a spacelike hypersurface* $\mathscr{S} \supset \mathscr{S}_{\mathrm{ext}}$, *achronal in* $\langle\langle \mathscr{M}_{\mathrm{ext}}\rangle\rangle$, *so that* S *as above coincides with the boundary of* $\overline{\mathscr{S}}$:
$$S = \partial \overline{\mathscr{S}} \subset \mathscr{E}^+,$$

then

$$\cup_t \phi_t[K_{(0)}](S) \subset \mathscr{H}_0$$

is a smooth null hypersurface, which is analytic if the metric is.

Remark 4.12. — The condition that the space-time is vacuum at large distances can be replaced by the requirement of existence of an $\mathscr{E}^+$–regular conformal completion at null infinity.

Proof. — Let Σ be a Cauchy surface for $\langle\langle \mathscr{M}_{\mathrm{ext}}\rangle\rangle$, and let $\widetilde{\mathscr{M}}$ be the conformal completion of $\mathscr{M}$ provided by Proposition 4.10. By [**22**, Proposition 4.8] the hypotheses of [**22**, Proposition 4.1] are satisfied, so that the Aleksandrov divergence $\theta_{\mathscr{A}l}$ of $\mathscr{E}^+$, as defined in [**22**], is nonnegative. Let S_1 be given by Proposition 4.1. Since isometries preserve area we have $\theta_{\mathscr{A}l} = 0$ almost everywhere on $\cup_t\phi_t(S_1) = \cup_t\phi_t(S)$. The result follows now from [**22**, Theorem 6.18]. □

4.4. Event horizons vs Killing horizons in analytic vacuum space-times. — We have the following result, first proved by Hawking for $n = 3$ [**49**] (compare [**38**] or [**16**, Theorem 5.1]), while the result for $n \geq 4$ in the mean-non-degenerate case is due to Hollands, Ishibashi and Wald [**55**], see also [**54, 60, 68**]:

Theorem 4.13. — *Let* $(\mathscr{M}, \mathfrak{g})$ *be an analytic,* $(n+1)$*–dimensional, vacuum space-time with complete Killing vector* $K_{(0)}$. *Assume that* $\mathscr{M}$ *contains an analytic null hypersurface* $\mathscr{E}$ *with a compact cross-section* S *transverse both to* $K_{(0)}$ *and to the generators of* $\mathscr{E}$. *Suppose that*

1. *either* $\langle\kappa\rangle_S \neq 0$, *where* $\langle\kappa\rangle_S$ *is defined in* (2.13),
2. *or* $n = 3$.

Then there exists a neighborhood $\mathscr{U}$ *of* $\mathscr{E}$ *and a Killing vector defined on* $\mathscr{U}$ *which is null on* $\mathscr{E}$.

In fact, if $K_{(0)}$ *is not tangent to the generators of* $\mathscr{E}$, *then there exist, near* $\mathscr{E}$, N *commuting linearly independent Killing vector fields* $K_{(1)}, \ldots, K_{(N)}$, $N \geq 1$, *(not* *necessarily complete but) with* 2π*-periodic orbits near* $\mathscr{E}$*, and numbers* $\Omega_{(1)},\dots,\Omega_{(N)}$*, such that*

$$K_{(0)}+\Omega_{(1)}K_{(1)}+\dots+\Omega_{(N)}K_{(N)}$$

is null on $\mathscr{E}$*.*

In the black hole context, Theorem 4.13 implies:

Theorem 4.14. — *Let* $(\mathscr{M},\mathfrak{g})$ *be an analytic, asymptotically flat, strongly causal, vacuum,* $(n+1)$*-dimensional space-time with stationary Killing vector* $K_{(0)}$*, the orbits of which are complete. Assume that* $\langle\langle\mathscr{M}_{\mathrm{ext}}\rangle\rangle$ *is globally hyperbolic, that a connected component* $\mathscr{H}_0^+$ *of* $\mathscr{H}^+$ *contains a compact cross-section* S *satisfying*

$$S\subset I^+(\mathscr{M}_{\mathrm{ext}}),$$

and that

1. *either* $\langle\kappa\rangle_S\neq 0$*,*
2. *or the flow defined by* $K_{(0)}$ *on the space of the generators of* $\mathscr{H}_0^+$ *is periodic.*

Suppose moreover that

a) *either*

$$\overline{\langle\langle\mathscr{M}_{\mathrm{ext}}\rangle\rangle}\cap I^+(\mathscr{M}_{\mathrm{ext}}) \text{ is strongly causal,}$$

b) *or there exists in* $\langle\langle\mathscr{M}_{\mathrm{ext}}\rangle\rangle$ *an asymptotically flat spacelike hypersurface* $\mathscr{S}$*, achronal in* $\langle\langle\mathscr{M}_{\mathrm{ext}}\rangle\rangle$*, so that* S *as above coincides with the boundary of* $\overline{\mathscr{S}}$*:*

$$S=\partial\overline{\mathscr{S}}\subset\mathscr{E}^+.$$

If $K_{(0)}$ *is* not *tangent to the generators of* $\mathscr{H}$*, then there exist, on* $\langle\langle\mathscr{M}_{\mathrm{ext}}\rangle\rangle\cup\mathscr{H}_0^+$*,* N *complete, commuting, linearly independent Killing vector fields* $K_{(1)},\dots,K_{(N)}$*,* $N\geq 1$*, with* 2π*-periodic orbits, and numbers* $\Omega_{(1)},\dots,\Omega_{(N)}$*, such that the Killing vector field*

$$K_{(0)}+\Omega_{(1)}K_{(1)}+\dots+\Omega_{(N)}K_{(N)}$$

is null on $\mathscr{H}_0$*.*

Remark 4.15. — For I^+-regular four-dimensional black holes S is a two-dimensional sphere (see Corollary 2.5), and then every Killing vector field acts periodically on the generators of $\mathscr{H}_0^+$.

Proof. — Theorem 4.11 shows that $\mathscr{E}_0^+ := \cup_t\phi_t[K_{(0)}](S)$ is an analytic null hypersurface. By Proposition 4.3 there exists a smooth compact section of $\mathscr{E}_0^+$ which is transverse both to its generators and to the stationary Killing vector. (11) We can thus invoke Theorem 4.13 to conclude existence of Killing vector fields $K_{(i)}$, $i=1,\dots,N$, defined near $\mathscr{E}_0^+$. By Corollary 2.4 and a theorem of Nomizu [**78**] we infer that the

(11) The hypothesis of existence of such a section needs to be added to those of [**55**, Theorem 2.1].

$K_{(i)}$'s extend globally to $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$. It remains to prove that the orbits of all Killing vector fields are complete. In order to see that, we note that by the asymptotic analysis of Killing vectors of [**5, 25**] there exists R large enough so that the flows of all $K_{(i)}$'s through points in the asymptotically flat region with $r \geq R$ are defined for all parameter values $t \in [0, 2\pi]$. The arguments in the proof of Theorem 1.2 of [**17**] then show that the flows $\phi_t[K_{(i)}]$'s are defined for $t \in [0, 2\pi]$ throughout $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$. But $\phi_{2\pi}[K_{(i)}]$ is an isometry which is the identity on an open set near $\mathscr{E}_0^+$, hence everywhere, and completeness of the orbits follows. □

5. Stationary axisymmetric black hole space-times: the area function

As will be explained in detail below, it follows from Theorem 4.14 together with the results on Killing vectors in [**6, 17**], that I^+–regular, 3+1 dimensional, asymptotically flat, rotating black holes have to be axisymmetric. The next step of the analysis of such space-times is the study of the area function

$$W := -\det\Big(\mathfrak{g}(K_{(\mu)}, K_{(\nu)})\Big)_{\mu,\nu=0,1}, \tag{5.1}$$

with $K_{(0)}$ being the asymptotically timelike Killing vector, and $K_{(1)}$ the axial one. Whenever $\sqrt{W}$ can be used as a coordinate, one obtains a dramatic simplification of the field equations, whence the interest thereof.

The function W is clearly positive in a region where $K_{(0)}$ is timelike and $K_{(1)}$ is spacelike, in particular it is non-negative on $\mathscr{M}_{\rm ext}$. As a starting point for further considerations, one then wants to show that W is non-negative on $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$:

Theorem 5.1. — *Let $(\mathscr{M}, \mathfrak{g})$ be a four-dimensional, analytic, asymptotically flat, vacuum space-time with stationary Killing vector $K_{(0)}$ and periodic Killing vector $K_{(1)}$, jointly generating an $\mathbb{R}\times \mathrm{U}(1)$ subgroup of the isometry group of $(\mathscr{M}, \mathfrak{g})$. If $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$ is globally hyperbolic, then the area function* (5.1) *is non-negative on $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$, vanishing precisely on the union of its boundary with the (non-empty) set $\{\mathfrak{g}(K_{(1)}, K_{(1)}) = 0\}$.*

We also have a version of Theorem 5.1, where the hypothesis of analyticity is replaced by that of I^+–regularity:

Theorem 5.2. — *Under the remaining hypotheses of Theorem 5.1, instead of analyticity assume that $(\mathscr{M}, \mathfrak{g})$ is I^+–regular. Then the conclusion of Theorem 5.1 holds.*

Keeping in mind our discussion above, Theorem 5.1 follows from Proposition 5.3 and Theorem 5.4 below. Similarly, Theorem 5.2 is a corollary of Theorem 5.6.

5.1. Integrability. — The first key fact underlying the analysis of the area function W is the following purely local fact, observed independently by Kundt and Trümper [**65**] and by Papapetrou [**80**] in dimension four (for a modern derivation see [**51, 95**]). The result, which does neither require $K_{(0)}$ to be stationary, nor the $K_{(i)}$'s to generate S^1 actions, generalizes to higher dimensions as follows (compare [**11, 35**]):

Proposition 5.3. — *Let $(\mathscr{M}, \mathfrak{g})$ be a vacuum, possibly with a cosmological constant, $(n+1)$-dimensional pseudo-Riemannian manifold with $n-1$ linearly independent commuting Killing vector fields $K_{(\mu)}$, $\mu = 0, \ldots, n-2$. If*

$$\mathscr{Z}_{dgt} := \{p \in \mathscr{M} \mid K_{(0)} \wedge \ldots \wedge K_{(n-2)}|_p = 0\} \neq \varnothing, \tag{5.2}$$

then[(12)]

$$dK_{(\mu)} \wedge K_{(0)} \wedge \cdots \wedge K_{(n-2)} = 0. \tag{5.3}$$

Proof. — To fix conventions, we use a Hodge star defined through the formula

$$\alpha \wedge \beta = \pm \langle *\alpha, \beta \rangle \mathrm{Vol},$$

where the plus sign is taken in the Riemannian case, minus in our Lorentzian one, while Vol is the volume form. The following (well known) identities are useful [**51**];

$$* * \theta = (-1)^{s(n+1-s)-1}\theta, \qquad \forall \theta \in \Lambda^s, \tag{5.4}$$

$$i_K * \theta = *(\theta \wedge K), \qquad \forall \theta \in \Lambda^s, \quad K \in \Lambda^1. \tag{5.5}$$

Further, for any Killing vector K,

$$[\mathscr{L}_K, *] = 0. \tag{5.6}$$

The Leibniz rule for the divergence $\delta := *d*$ reads, for $\theta \in \Lambda^s$,

$$\begin{aligned}
\delta(\theta \wedge K) &= *d * (\theta \wedge K) \overset{(5.5)}{=} * d(i_K * \theta) = *(\mathscr{L}_K * \theta - i_K d * \theta) \\
&\overset{(5.4),(5.6)}{=} * * \mathscr{L}_K \theta - * i_K (-1)^{(n+1-s+1)(n+1-(n+1-s+1))-1} * * d * \theta \\
&= (-1)^{s(n+1-s)-1} \mathscr{L}_K \theta - (-1)^{s(n+1-s)-n+1} * *(\delta\theta \wedge K) \\
&= (-1)^{s(n+1-s)-1} \mathscr{L}_K \theta + (-1)^{n+1} \delta\theta \wedge K.
\end{aligned}$$

Applying this to $\theta = dK$ one obtains

$$\begin{aligned}
*d * (dK \wedge K) &= -\mathscr{L}_K dK + (-1)^{n+1} \delta dK \wedge K \\
&= (-1)^{n+1} \delta dK \wedge K.
\end{aligned}$$

(12) By an abuse of notation, we use the same symbols for vector fields and for the associated 1-forms.

As any Killing vector is divergence free, we see that

$$\delta dK = (-1)^n \Delta K = (-1)^{n+1} i_K \,\mathrm{Ric}\,.$$

Assuming that the Ricci tensor is proportional to the metric, $\mathrm{Ric} = \lambda \mathfrak{g}$, we conclude that

$$* d * (dK \wedge K) = (i_K \lambda \mathfrak{g}) \wedge K = 0.$$

Let $\omega_{(\mu)}$ be the μ'th twist form,

$$\omega_{(\mu)} := *(dK_{(\mu)} \wedge K_{(\mu)}).$$

The identity

$$\begin{aligned}\mathscr{L}_{K_{(\mu)}}\omega_{(\nu)} &= \mathscr{L}_{K_{(\mu)}} * (dK_{(\mu)} \wedge K_{(\nu)})\\ &= *(\mathscr{L}_{K_{(\mu)}} dK_{(\nu)} + dK_{(\nu)} \wedge \mathscr{L}_{K_{(\mu)}} K_{(\nu)}) = 0,\end{aligned}$$

together with

$$\mathscr{L}_{K_{(\mu_1)}}(i_{K_{(\mu_2)}} \dots i_{K_{(\mu_\ell)}} \omega_{(\mu_{\ell+1})}) = i_{K_{(\mu_2)}} \dots i_{K_{(\mu_{n-1})}} \mathscr{L}_{K_{(\mu_\ell)}} \omega_{(\mu_{\ell+1})} = 0,$$

and with Cartan's formula for the Lie derivative, gives

$$d(i_{K_{(\mu_1)}} \dots i_{K_{(\mu_\ell)}} \omega_{(\mu_{\ell+1})}) = (-1)^\ell i_{K_{(\mu_1)}} \dots i_{K_{(\mu_{n-1})}} d\omega_{(\mu_{\ell+1})}\,. \tag{5.7}$$

We thus have

$$\begin{aligned} d * (dK_{(\mu_1)} \wedge K_{(\mu_1)} \wedge \dots \wedge K_{(\mu_{n-1})}) &= d(i_{K_{(\mu_{n-1})}} \dots i_{K_{(\mu_2)}} * (dK_{(\mu_1)} \wedge K_{(\mu_1)}))\\ &= (-1)^{n-2} i_{K_{(\mu_{n-1})}} \dots i_{K_{(\mu_2)}} d\omega_{(\mu_1)} = 0.\end{aligned}$$

So the function $*(dK_{(\mu_1)} \wedge K_{(\mu_1)} \wedge K_{(\mu_2)} \wedge \dots \wedge K_{(\mu_{n-1})})$ is constant, and the result follows from (5.2). □

5.2. The area function for a class of space-times with a commutative group of isometries. — The simplest non-trivial reduction of the Einstein equations by isometries, which does *not* reduce the equations to ODEs, arises when orbits have co-dimension two, and the isometry group is abelian. It is useful to formulate the problem in a general setting, with $1 \le s \le n-1$ commuting Killing vector fields $K_{(\mu)}$, $\mu = 0, \dots, s-1$, satisfying the following orthogonal integrability condition:

$$\forall \mu = 0, \dots, s-1 \qquad dK_{(\mu)} \wedge K_{(0)} \wedge \dots \wedge K_{(s-1)} = 0. \tag{5.8}$$

For the problem at hand, (5.8) will hold when $s = n-1$ by Proposition 5.3. Note further that (5.8) with $s = 1$ is the definition of staticity. So, the analysis that follows covers simultaneously static analytic domains of dependence in all dimensions $n \ge 3$ (filling a gap in previous proofs), or stationary axisymmetric analytic four-dimensional space-times, or five dimensional stationary analytic space-times with two further periodic Killing vectors as in [**56**]. It further covers stationary axisymmetric I^+-regular black holes in $n = 3$, in which case analyticity is not needed.

Similarly to (5.2) we set

$$\mathscr{Z}_{dgt} := \{K_{(0)} \wedge \dots \wedge K_{(s-1)} = 0\}, \tag{5.9}$$

$$\widetilde{\mathscr{Z}} := \{p \in \mathscr{M} : \det\Big(\mathfrak{g}(K_{(i)}, K_{(j)})\Big)_{i,j=1,\dots s-1} = 0\}. \tag{5.10}$$

In the following result, the proof of which builds on key ideas of Carter [**11, 12**], we let $K_{(0)}$ denote the Killing vector associated to the $\mathbb{R}$ factor of $\mathbb{R} \times \mathbb{T}^{s-1}$, and we let $K_{(i)}$ denote the Killing vector field associated with the $i-th$ S^1 factor of $\mathbb{T}^{s-1}$:

Theorem 5.4. — *Let $(\mathscr{M}, \mathfrak{g})$ be an $(n+1)$–dimensional, asymptotically flat, analytic space-time with a metric invariant under an action of the abelian group $G = \mathbb{R} \times \mathbb{T}^{s-1}$ with s–dimensional principal orbits, $1 \le s \le n-1$, and assume that* (5.8) *holds. If $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$ is globally hyperbolic, then the function*

$$W := -\det\Big(\mathfrak{g}(K_{(\mu)}, K_{(\nu)})\Big)_{\mu,\nu=0,\dots,s-1} \tag{5.11}$$

is non-negative on $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$, vanishing on $\partial\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle \cup \widetilde{\mathscr{Z}}$.

Remark 5.5. — Here analyticity could be avoided if, in the proof below, one could show that one can extract out of the degenerate $\hat{S}_p$'s (if any) a closed embedded hypersurface. Alternatively, the hypothesis of analyticity can be replaced by that of non-existence of non-embedded degenerate prehorizons within $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$. Moreover, one also has:

Theorem 5.6. — *Let $n = 3$, $s = 2$ and, under the remaining conditions of Theorem 5.4, instead of analyticity assume that $(\mathscr{M}, \mathfrak{g})$ is I^+–regular. Then the conclusion of Theorem 5.4 holds.*

Before passing to the proof, some preliminary remarks are in order. The fact that $\mathscr{M} \setminus \mathscr{Z}_{dgt}$ is open, where $\mathscr{Z}_{dgt}$ is as in (5.9), together with (5.8), establishes the conditions of the Frobenius theorem (see, e.g., [**52**]). Therefore, for every $p \notin \mathscr{Z}_{dgt}$ there exists a unique, maximal submanifold (not necessarily embedded), passing through p and orthogonal to $\mathrm{Span}\{K_{(0)}, \dots, K_{(s-1)}\}$, that we denote by $\mathscr{O}_p$. Carter builds his further analysis of stationary axisymmetric black holes on the sets $\mathscr{O}_p$. This leads to severe difficulties at the set $\widetilde{\mathscr{Z}}$ of (5.10), which we were not able to resolve using neither Carter's ideas, nor those in [**91**]. There is, fortunately, an alternative which we provide below. In order to continue, some terminology is needed:

Definition 5.7. — *Let K be a Killing vector and set*

$$\mathscr{N}[K] := \{\mathfrak{g}(K,K) = 0,\ K \neq 0\}. \tag{5.12}$$

Every connected, not necessarily embedded, null hypersurface $\mathscr{N}_0 \subset \mathscr{N}[K]$ to which K is tangent will be called a Killing prehorizon.

In this terminology, a Killing horizon is a Killing prehorizon which forms an *embedded* hypersurface which *coincides* with a connected component of $\mathscr{N}[K]$.

The Minkowskian Killing vector $\partial_t - \partial_x$ provides an example where $\mathscr{N}$ is not a hypersurface, with every hyperplane $t + x = \text{const}$ being a prehorizon. The Killing vector $K = \partial_t + Y$ on $\mathbb{R} \times \mathbb{T}^n$, equipped with the flat metric, where $\mathbb{T}^n$ is an n-dimensional torus, and where Y is a unit Killing vector on $\mathbb{T}^n$ with dense orbits, admits prehorizons which are not embedded. This last example is globally hyperbolic, which shows that causality conditions are not sufficient to eliminate this kind of behavior.

Our first step towards the proof of Theorem 5.4 will be Theorem 5.8, inspired again by some key ideas of Carter, together with their variations by Heusler. We will assume that the $K_{(i)}$'s, $i = 1, \dots, s-1$, are spacelike (by this we mean that they are spacelike away from their zero sets), but no periodicity or completeness assumptions are made concerning their orbits. This can always be arranged locally, and therefore does not involve any loss of generality for the local aspects of our claim; but we emphasize that our claims are global when the $K_{(i)}$'s are spacelike everywhere.

In our analysis below we will be mainly interested in what happens in $\langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle$ where, by Corollary 3.8, we have

$$\widetilde{\mathscr{Z}} \cap \langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle = \mathscr{Z}_{dgt} \cap \langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle,$$

in a chronological domain of outer communications. We note that $\mathscr{Z}_{dgt} \subset \{W = 0\}$, but equality does not need to hold for Lorentzian metrics. For example, consider in $\mathbb{R}^{1,2}$, $K_{(0)} = \partial_x + \partial_t$ and $K_{(1)} = \partial_y$; then $K_{(0)} \wedge K_{(1)} = dx \wedge dy - dt \wedge dy \not\equiv 0$ and $W \equiv 0$.

If the $K_{(i)}$'s generate a torus action on a stably causal manifold,[(13)] it is well known that $\widetilde{\mathscr{Z}}$ is a closed, totally geodesic, timelike, stratified, embedded submanifold of $\mathscr{M}$ with codimension of each stratum at least two (this follows from **[63]** or **[2**, Appendix C]). So, under those hypotheses, within $\langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle$, we will have

(5.13) the intersection of $\mathscr{Z}_{dgt}$ with any null hypersurface $\mathscr{N}$ is a stratified submanifold of $\mathscr{N}$, with $\mathscr{N}$–codimension at least two.

This condition will be needed in our subsequent analysis. We expect this property not to be needed, but we have not investigated this question any further.

(13) Let t be a time-function on $(\mathscr{M}, \mathfrak{g})$; averaging t over the orbits of the torus generated by the $K_{(i)}$'s we obtain a new time function such that the $K_{(i)}$'s are tangent to its level sets. This reduces the problem to the analysis of zeros of Riemannian Killing vectors.

Theorem 5.8. — *Let $(\mathscr{M}, \mathfrak{g})$ be an $(n+1)$-dimensional Lorentzian manifold with $s \geq 1$ linearly independent commuting Killing vectors $K_{(\mu)}$, $\mu = 0, \dots, s-1$, satisfying the integrability conditions (5.8), as well as (5.13), with the $K_{(i)}$'s, $i = 1, \dots, s-1$, spacelike. Suppose that $\{W = 0\} \setminus \mathscr{Z}_{dgt}$ is not empty, and for each p in this set consider the Killing vector field l_p defined as*(14)

$$(5.14) \qquad l_p = K_{(0)} - (h^{(i)(j)}\mathfrak{g}(K_{(0)}, K_{(i)}))|_p K_{(j)},$$

where $h^{(i)(j)}$ is the matrix inverse to

$$(5.15) \qquad h_{(i)(j)} := \mathfrak{g}(K_{(i)}, K_{(j)}), \quad i, j \in \{1, ..., s-1\}.$$

Then the distribution $l_p^{\perp} \subset T\mathscr{M}$ of vectors orthogonal to l_p is integrable over the non-empty set

$$(5.16) \qquad \overline{\{q \in \mathscr{M} \setminus \mathscr{Z}_{dgt} \mid \mathfrak{g}(l_p, l_p)|_q = 0,\ W(q) = 0\}} \setminus \{q \in \mathscr{M} \mid l_p(q) = 0\}.$$

If we define $\hat{S}_p$ to be the maximally extended over $\{W = 0\}$, connected, integral leaf of this distribution(15) *passing through p, then all $\hat{S}_p$'s are Killing prehorizons, totally geodesic in $\mathscr{M} \setminus \{l_p = 0\}$.*

In several situations of interest the $\hat{S}_p$'s form embedded hypersurfaces which coincide with connected components of the set defined in (5.16), but this is certainly not known at this stage of the argument:

Remark 5.9. — Null translations in Minkowski space-time, or in *pp*-wave space-times, show that the $\hat{S}_p$'s might be different from connected components of $\mathscr{N}[l_p]$.

Remark 5.10. — It follows from our analysis here that for $q \in \hat{S}_p \setminus \mathscr{Z}_{dgt}$ we have $l_q = l_p$. For $q \in \hat{S}_p \cap \mathscr{Z}_{dgt}$ we can define l_q by setting $l_q := l_p$. We then have $l_p = l_q$ for all $q \in \hat{S}_p$.

Proof. — Let

$$(5.17) \qquad w := K_{(0)} \wedge \cdots \wedge K_{(s-1)}.$$

We need an equation of Carter [**11**]:

Lemma 5.11 **([11])**. — *We have*

$$(5.18) \qquad w \wedge dW = (-1)^s W dw.$$

(14) If $s = 1$ then $\widetilde{\mathscr{Z}} = \varnothing$ and $l_p = K_{(0)}$.

(15) To avoid ambiguities, we emphasize that points at which l_p vanishes do not belong to $\hat{S}_p$.

Proof. — Let $F = \{W = 0\}$. The result is trivial on the interior $\mathring{F}$ of F, if non-empty. By continuity, it then suffices to prove (5.18) on $\mathscr{M} \setminus F$. Let $\mathscr{O}$ be the set of points in $\mathscr{M} \setminus F$ at which the Killing vectors are linearly independent. Consider any point $p \in \mathscr{O}$, and let (x^a, x^A), $a = 0, \dots, s-1$, be local coordinates near p chosen so that $K_{(a)} = \partial_a$ and $\mathrm{Span}\{\partial_a\} \perp \mathrm{Span}\{\partial_A\}$; this is possible by (5.8). Then

$$w = -W dx^0 \wedge \cdots \wedge dx^{s-1},$$

and (5.18) follows near p. Since $\mathscr{O}$ is open and dense, the lemma is proved. □

Returning to the proof of Theorem 5.8, as already said, (5.8) implies that for every $p \notin \mathscr{Z}_{dgt}$ there exists a unique, maximal, $(n+1-s)$–dimensional submanifold (not necessarily embedded), passing through p and orthogonal to $\mathrm{Span}\{K_{(0)}, \dots, K_{(s-1)}\}$, that we denote by $\mathscr{O}_p$. By definition,

$$\mathscr{O}_p \cap \mathscr{Z}_{dgt} = \varnothing, \tag{5.19}$$

and clearly

$$\mathscr{O}_p \cap \mathscr{O}_q \neq \varnothing \quad \Longleftrightarrow \quad \mathscr{O}_p = \mathscr{O}_q. \tag{5.20}$$

Recall that $p \in \{W = 0\} \setminus \mathscr{Z}_{dgt}$; then $K_{(0)} \wedge \cdots \wedge K_{(s-1)} \neq 0$ in $\mathscr{O}_p$ and we may choose vector fields $u_{(\mu)} \in TM$, $\mu = 0, \dots, s-1$, such that

$$K_{(0)} \wedge \cdots \wedge K_{(s-1)}(u_{(0)}, \dots, u_{(s-1)}) = 1$$

in some neighborhood of p. Let γ be a C^k curve, $k \geq 1$, passing through p and contained in $\mathscr{O}_p$. Since $\dot\gamma(s) \in T_{\gamma(s)}\mathscr{O}_p = \mathrm{Span}\{K_{(0)}, \dots, K_{(s-1)}\}^{\perp}|_{\gamma(s)}$, after contracting (5.18) with $(u_0, \dots, u_{s-1}, \dot\gamma)$ we obtain the following Cauchy problem

$$\begin{cases} \frac{d}{ds}(W \circ \gamma)(s) \sim W \circ \gamma(s), \\ \qquad W|_p = 0. \end{cases} \tag{5.21}$$

Uniqueness of solutions of this problem guarantees that $W \circ \gamma(s) \equiv 0$ and therefore W vanishes along the $(n+1-s)$-dimensional submanifold $\mathscr{O}_p$. Since G preserves W, W must vanish on the sets

$$S_p := G_s \cdot \mathscr{O}_p. \tag{5.22}$$

Here $G_s\cdot$ denotes the motion of a set using the group generated by the $K_{(i)}$'s, $i = 1, \dots, s-1$; if the orbits of some of the $K_{(i)}$'s are not complete, by this we mean "the motion along the orbits of all linear combinations of the $K_{(i)}$'s starting in the given set, as far as those orbits exist". Since $T_q\mathscr{O}_p$ is orthogonal to all Killing vectors by definition, and the $K_{(i)}$'s are spacelike, the $K_{(i)}$'s are transverse to $\mathscr{O}_p$, so that the S_p's are smooth (not necessarily embedded) submanifolds of codimension one.

On $\{W = 0\} \setminus \mathscr{Z}_{dgt}$ the metric $\mathfrak{g}$ restricted to $\mathrm{Span}\{K_{(0)}, \dots, K_{(s-1)}\}$ is degenerate, so that $\mathrm{Span}\{K_{(0)}, \dots, K_{(s-1)}\}$ is a null subspace of $T\mathscr{M}$. It follows that for $q \in$

$\{W = 0\} \setminus \mathscr{Z}_{dgt}$ some linear combination of Killing vectors is null and orthogonal to $\mathrm{Span}\{K_{(0)}, \ldots, K_{(s-1)}\}$, thus in $T_q\mathscr{O}_p$. So for $q \in \{W = 0\} \setminus \mathscr{Z}_{dgt}$ the tangent spaces T_qS_p are orthogonal sums of the null spaces $T_q\mathscr{O}_p$ and the spacelike ones $\mathrm{Span}\{K_{(1)}, \ldots, K_{(s-1)}\}$. We conclude that the S_p's form smooth, null, not necessarily embedded, hypersurfaces, with

$$S_p = G \cdot \mathscr{O}_p \subset \{W = 0\} \setminus \mathscr{Z}_{dgt}, \tag{5.23}$$

where the action of G is understood as explained after (5.22).

Let the vector $\ell = \Omega^{(\mu)}K_{(\mu)}$, $\Omega^{(\mu)} \in \mathbb{R}$ be tangent to the null generators of S_p, thus

$$\Omega^{(\mu)}\mathfrak{g}(K_{(\mu)}, K_{(\nu)})\Omega^{(\nu)} = 0. \tag{5.24}$$

Since $\det(\mathfrak{g}(K_{(\mu)}, K_{(\nu)})) = 0$ with one-dimensional null space on $\{W = 0\} \setminus \mathscr{Z}_{dgt}$, (5.24) is equivalent there to

$$\mathfrak{g}(K_{(\mu)}, K_{(\nu)})\Omega^{(\nu)} = 0. \tag{5.25}$$

Since the $K_{(i)}$'s are spacelike we must have $\Omega^{(0)} \neq 0$, and it is convenient to normalize ℓ so that $\Omega^{(0)} = 1$. Assuming $p \notin \widetilde{\mathscr{Z}}$, from (5.25) one then immediately finds

$$\ell = K_{(0)} + \Omega^{(i)}K_{(i)} = K_{(0)} - h^{(i)(j)}\mathfrak{g}(K_{(0)}, K_{(j)})K_{(i)}, \tag{5.26}$$

where $h^{(i)(j)}$ is the matrix inverse to

$$h_{(i)(j)} = \mathfrak{g}(K_{(i)}, K_{(j)}),\ i, j \in \{1, \ldots, s-1\}. \tag{5.27}$$

To continue, we show that:

Proposition 5.12. — *For each $j = 1, \ldots, n$, the function*

$$S_p \ni q \mapsto \Omega^{(j)}(q) := -h^{(i)(j)}(q)\mathfrak{g}(K_{(0)}, K_{(i)})(q)$$

is constant over S_p.

Proof. — The calculations here are inspired by, and generalize those of [**51**, pp. 93-94]. As is well known,

$$dh^{(i)(j)} = -h^{(i)(m)}h^{(j)(s)}dh_{(m)(s)}. \tag{5.28}$$

From (5.4)-(5.5) together with $\mathscr{L}_{K_{(i)}}K_{(j)} = 0$ we have

$$\begin{aligned} dh_{(i)(j)} &= d[\mathfrak{g}(K_{(i)}, K_{(j)})] = di_{K_{(i)}}K_{(j)} = -i_{K_{(i)}}dK_{(j)} \\ &= -i_{K_{(i)}}(-1)^{2(n+1-2)-1} * *dK_{(j)} = (-1)^n * (K_{(i)} \wedge *dK_{(j)}), \end{aligned}$$

with a similar formula for $d[\mathfrak{g}(K_{(0)}, K_{(j)})]$. Next,

$$\begin{aligned}
d\Omega^{(i)} &= d(-h^{(i)(j)}\mathfrak{g}(K_{(0)}, K_{(j)}))\\
&= -[\mathfrak{g}(K_{(0)}, K_{(j)})dh^{(i)(j)} + h^{(i)(j)}d[\mathfrak{g}(K_{(0)}, K_{(j)})]]\\
&= -[-\mathfrak{g}(K_{(0)}, K_{(j)})h^{(i)(m)}h^{(j)(s)}dh_{(s)(m)} + h^{(i)(m)}d[\mathfrak{g}(K_{(0)}, K_{(m)})]]\\
&= -h^{(i)(m)}[-(-1)^n\mathfrak{g}(K_{(0)}, K_{(j)})h^{(j)(s)} * (K_{(s)} \wedge *dK_{(m)})\\
&+ (-1)^n * (K_{(0)} \wedge *dK_{(m)})]\\
&= (-1)^{n+1}h^{(i)(m)} * [(\Omega^{(s)}K_{(s)} + K_{(0)}) \wedge *dK_{(m)}]\\
&= (-1)^{n+1}h^{(i)(m)} * (\ell \wedge *dK_{(m)}),
\end{aligned}$$

and

$$\begin{aligned}
i_{K_{(0)}}\dots i_{K_{(s-1)}} * d\Omega^{(i)} &= (-1)^{n+1}i_{K_{(0)}}\dots i_{K_{(s-1)}}h^{(i)(m)} * *(\ell \wedge *dK_{(m)})\\
&= h^{(i)(m)}i_{K_{(0)}}\dots i_{K_{(s-1)}}(\ell \wedge *dK_{(m)}).
\end{aligned}$$

Since $i_{K_{(i)}}\ell|_{S_p} = \mathfrak{g}(\ell, K_{(i)})|_{S_p} = 0$, we obtain

$$\begin{aligned}
i_{K_{(0)}}\dots i_{K_{(s-1)}}(\ell \wedge *dK_{(m)})|_{S_p} &= i_{K_{(0)}}\dots i_{K_{(s-2)}}[i_{K_{(s-1)}}\ell \wedge *dK_{(m)}\\
&+ (-1)^1\ell \wedge i_{K_{(s-1)}} * dK_{(m)}]|_{S_p}\\
&= -i_{K_{(0)}}\dots i_{K_{(s-2)}}(\ell \wedge i_{K_{(s-1)}} * dK_{(m)})|_{S_p} = \dots\\
&= (-1)^s\ell \wedge i_{K_{(0)}}\dots i_{K_{(s-1)}} * dK_{(m)}\,|_{S_p}\\
&= (-1)^s\ell \wedge *(dK_{(m)} \wedge K_{(s-1)} \wedge \dots \wedge K_{(0)})|_{S_p}\\
&\overset{(5.3)}{=} 0,
\end{aligned}$$

and therefore

$$i_{K_{(0)}}\dots i_{K_{(s-1)}} * d\Omega^{(i)}|_{S_p} = 0. \tag{5.29}$$

This last result says that $d\Omega^{(i)}|_{S_p}$ is a linear combination of the $K_{(\mu)}$'s, so for each i there exist numbers $\alpha^{(\mu)} \in \mathbb{R}$ such that

$$d\Omega^{(i)}|_{S_p} = \alpha^{(\mu)}K_{(\mu)}. \tag{5.30}$$

Now, the $\Omega^{(i)}$'s are clearly invariant under the action of the group generated by the $K_{(\mu)}$'s, which implies

$$0 = i_{K_{(\mu)}}d\Omega^{(i)} = \mathfrak{g}(K_{(\mu)}, \alpha^{(\nu)}K_{(\nu)}).$$

This shows that $\alpha^{(\mu)}K_{(\mu)}$ is orthogonal to all Killing vectors, so it must be proportional to ℓ. Since $T_qS_p = \ell^\perp$, we are done. □

Returning to the proof of Theorem 5.8, we have shown so far that S_p is a null hypersurface in $\{W = 0\} \setminus \mathscr{Z}_{dgt}$, with the Killing vector $l_p := \ell$ as in (5.14) tangent to the generators of S_p. In other words, S_p is a prehorizon. Furthermore,

$$\begin{aligned} T_q\mathscr{M} \ni Y \in T_qS_p \ \text{ for some } p \quad &\Longleftrightarrow \\ & W(q) = 0,\ K_{(0)} \wedge \cdots \wedge K_{(s-1)}|_q \neq 0, \quad Y \perp l_p. \end{aligned} \tag{5.31}$$

For further purposes it is necessary to extend this result to the hypersurface $\hat{S}_p$ defined in the statement of Theorem 5.8. This proceeds as follows:

It is well known [**43**] that Killing horizons are *locally totally geodesic*, by which we mean that geodesics initially tangent to the horizon remain on the horizon for some open interval of parameters. This remains true for prehorizons:

Corollary 5.13. — *S_p is locally totally geodesic. Furthermore, if $\gamma : [0,1) \to S_p$ is a geodesic such that $\gamma(1) \notin S_p$, then $\gamma(1) \in \mathscr{Z}_{dgt}$.*

Proof. — Let $\gamma : I \to \mathscr{M}$ be an affinely-parameterized geodesic satisfying $\gamma(0) = q \in S_p$ and $\dot\gamma(0) \in T_qS_p \Longleftrightarrow \mathfrak{g}(\dot\gamma(0), l_p) = 0$. Then

$$\frac{d}{dt}\mathfrak{g}(\dot\gamma(t), l_p) = \mathfrak{g}(\nabla_{\dot\gamma(t)}\dot\gamma(t), l_p) + \mathfrak{g}(\dot\gamma(t), \nabla_{\dot\gamma(t)} l_p) = 0, \tag{5.32}$$

where the first term vanishes because γ is an affinely parameterized geodesic, while the second is zero by the Killing equation. Since $\mathfrak{g}(\dot\gamma(0), l_p) = 0$, we get

$$\mathfrak{g}(\dot\gamma(t), l_p) = 0,\ \forall t \in I. \tag{5.33}$$

We conclude that $\dot\gamma$ remains perpendicular to l_p, hence remains within S_p as long as a zero of $K_{(0)} \wedge \cdots \wedge K_{(s-1)}$ is not reached, compare (5.31). □

Consider, now, the following set of points which can be reached by geodesics initially tangent to S_p:

$$\begin{aligned} \tilde{S}_p \quad := \quad & \{q :\ \exists \text{ a geodesic segment } \gamma : [0,1] \to \mathscr{M} \text{ such} \\ & \quad \text{that } \gamma(1) = q \text{ and } \gamma(s) \in S_p \text{ for } s \in [0,1)\} \setminus \{q : l_p(q) = 0\}. \end{aligned} \tag{5.34}$$

Then $S_p \subset \tilde{S}_p$, and if $q \in \tilde{S}_p \setminus S_p$ then $q \in \mathscr{Z}_{dgt}$ by Corollary 5.13. We wish to show that $\tilde{S}_p$ is a smooth hypersurface, included and maximally extended in the set (5.16); equivalently

$$\tilde{S}_p = \hat{S}_p. \tag{5.35}$$

For this, let $q \in \tilde{S}_p$, let $\mathscr{O}$ be a geodesically convex neighborhood of q not containing zeros of l_p, and for $r \in \mathscr{O}$ define

$$R_r = \exp_{\mathscr{O},r}(l_q(r)^\perp), \tag{5.36}$$

here $\exp_{\mathscr{O},r}$ is the exponential map at the point $r \in \mathscr{O}$ in the space-time $(\mathscr{O}, \mathfrak{g}|_{\mathscr{O}})$. It is convenient to require that $\mathscr{O}$ is included within the radius of injectivity of all its points (see [**64**, Theorem 8.7]). Let γ be as in the definition of $\tilde{S}_p$. Without loss of generality we can assume that $\gamma(0) \in \mathscr{O}$. We have $\dot\gamma(s) \perp l_p$ for all $s \in [0,1)$, and by continuity also at $s = 1$. This shows that $\gamma([0,1]) \subset R_q$.

Now, $R_{\gamma(0)}$ is a smooth hypersurface in $\mathscr{O}$. It coincides with S_p near $\gamma(0)$, and every null geodesic starting at $\gamma(0)$ and normal to l_p there belongs both to $R_{\gamma(0)}$ and S_p until a point in $\mathscr{Z}_{dgt}$ is reached. This shows that $R_{\gamma(0)}$ is null near every such geodesic until, and including, the first point on that geodesic at which $\mathscr{Z}_{dgt}$ is reached (if any). By (5.13) $R_{\gamma(0)} \cap S_p$ is open and dense in $R_{\gamma(0)}$. Thus the tangent space to $R_{\gamma(0)}$ coincides with $l_p^\perp$ at the open dense set of points $R_{\gamma(0)} \cap S_p$, with that intersection being a null, locally totally geodesic (not necessarily embedded) hypersurface. By continuity $R_{\gamma(0)}$ is a subset of (5.16), with $TR_{\gamma(0)} = l_p^\perp$ everywhere. Since $R_{\gamma(0)} \subset \tilde S_p$, Equation (5.35) follows.

The construction of the $\tilde S_p$'s shows that every integral manifold of the distribution $l_p^\perp$ over the set

$$\Omega := \{q \in \mathscr{M} \setminus \mathscr{Z}_{dgt} \mid \mathfrak{g}(l_p, l_p)|_q = 0,\ W(q) = 0\} \tag{5.37}$$

can be extended to a maximal leaf contained in $\overline{\Omega} \setminus \{q | l_p(q) = 0\}$, compare (5.16). To finish the proof of Theorem 5.8 it thus remains to show that there exists a leaf through every point in $\overline{\Omega} \setminus \{q|l_p(q) = 0\}$. Since this last set is contained in the closure of Ω, we need to analyze what happens when a sequence of null leaves $\hat S_{p_n}$, all normal to a fixed Killing vector field l_q, has an accumulation point. We show in Lemma 5.14 below that such sequences accumulate to an integral leaf through the limit point, which completes the proof of the theorem. □

We shall say that S is an *accumulation set* of a sequence of sets S_n if S is the collection of limits, as i tends to infinity, of sequences $q_{n_i} \in S_{n_i}$.

Lemma 5.14. — *Let $\hat S_{p_n}$ be a sequence of leaves such that $l_{p_n} = l_q$, for some fixed q, and suppose that $p_n \to p$. If $l_q(p) \neq 0$, then p belongs to a leaf $\hat S_p$ with $l_p = l_q$. Furthermore there exists a neighborhood $\mathscr{U}$ of p such that $exp_{\mathscr{U},p}(l_q(p)^\perp) \subseteq \hat S_p \cap \mathscr{U}$ is the accumulation set of the sequence $exp_{\mathscr{U},p_n}(l_q(p_n)^\perp) \subseteq \hat S_{p_n} \cap \mathscr{U}$, $n \in \mathbb{N}$.*

Proof. — Let $\mathscr{U}$ be a small, open, conditionally compact, geodesically convex neighborhood of p which does not contain zeros of l_q. Let $\hat S_{p_n}$ be that leaf, within $\mathscr{U}$, of the distribution $l_q^\perp$ which contains p_n. The $\hat S_{p_n}$'s are totally geodesic submanifolds of $\mathscr{U}$ by Corollary 5.15, and therefore are uniquely determined by prescribing $T_{p_n}\hat S_{p_n}$. Now, the subspaces $T_{p_n}\hat S_{p_n} = l_q(p_n)^\perp$ obviously converge to $l_q(p)^\perp$ in the sense of accumulation sets. Smooth dependence of geodesics upon initial values implies that $\exp_{\mathscr{U},p_n}(l_q(p_n)^\perp)$ converges in C^k, for any k, to $\exp_{\mathscr{U},p}(l_q(p)^\perp)$. Since W vanishes on $\exp_{\mathscr{U},p_n}(l_q(p_n)^\perp)$, we obtain that W vanishes on $\exp_{\mathscr{U},p}(l_q(p)^\perp)$. Since $T_{q_n} \exp_{\mathscr{U},p_n}(l_q(p_n)^\perp) = l_p^\perp(q_n)$ for any $q_n \in \exp_{\mathscr{U},p_n}(l_q(p_n)^\perp)$ we conclude that

$T_r \exp_{\mathscr{U},p}(l_q(p)^\perp) = l_p^\perp(r)$ for any $r \in \exp_{\mathscr{U},p}(l_q(p)^\perp)$. So $\exp_{\mathscr{U},p}(l_q(p)^\perp)$ is a leaf, within $\mathscr{U}$, through p of the distribution $l_q^\perp$ over the set (5.16), and $\exp_{\mathscr{U},p}(l_q(p)^\perp) = \hat{S}_p \cap \mathscr{U}$ is the accumulation set of the totally geodesic submanifolds $\hat{S}_{p_n} \cap \mathscr{U}$'s. □

The remainder of the proof of Theorem 5.4 consists in showing that the $\hat{S}_p$'s cannot intersect $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$. We start with an equivalent of Corollary 5.13, with identical proof:

Corollary 5.15. — *$\hat{S}_p$ is locally totally geodesic. Furthermore, if $\gamma : [0,1) \to \hat{S}_p$ is a geodesic segment such that $\gamma(1) \notin \hat{S}_p$, then l_p vanishes at $\gamma(1)$.* □

Corollary 3.8 shows that Killing vectors as described there have no zeros in $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$, and Corollary 5.15 implies now:

Corollary 5.16. — *$\hat{S}_p \cap \langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$ is totally geodesic in $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$ (possibly empty).* □

To continue, we want to extract, out of the $\hat{S}_p$'s, a closed, embedded, Killing horizon S_0^+. Now, e.g. the analysis in [**55**] shows that the gradient of $\mathfrak{g}(l_p, l_p)$ is either everywhere zero on $\hat{S}_p$ (we then say that $\hat{S}_p$ is degenerate), or nowhere vanishing there. One immediately concludes that non-degenerate $\hat{S}_p$'s, if non-empty, are embedded, closed hypersurfaces in $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$. Then, if there exists non-empty non-degenerate $\hat{S}_p$'s, we choose one and we set

$$S_0^+ = \hat{S}_p. \tag{5.38}$$

Otherwise, all non-empty $\hat{S}_p$'s are degenerate; to show that such prehorizons, if non-empty, are embedded, we will invoke analyticity (which has not been used so far). So, consider a degenerate component $\hat{S}_p$, and note that $\hat{S}_p$ does not self-intersect, being a subset of the union of integral manifolds of a smooth distribution of hyperplanes. Suppose that $\hat{S}_p$ is not embedded. Then there exists a point $q \in \hat{S}_p$, a conditionally compact neighborhood $\mathscr{O}$ of q, and a sequence of points $p_n \in \hat{S}_p$ lying on pairwise disjoint components of $\mathscr{O} \cap \hat{S}_p$, with p_n converging to q. Now, Killing vectors are solutions of the overdetermined set of PDEs

$$\nabla_\mu \nabla_\nu X_\rho = R^\alpha{}_{\mu\nu\rho} X_\alpha ,$$

which imply that they are analytic if the metric is. So $\mathfrak{g}(l_p, l_p)$ is an analytic function that vanishes on an accumulating family of hypersurfaces. Consequently $\mathfrak{g}(l_p, l_p)$ vanishes everywhere, which is not compatible with asymptotic flatness. Hence the $\hat{S}_p$'s are embedded, coinciding with connected components of the set $\{\mathfrak{g}(l_p, l_p) = 0 = W\} \setminus \{l_p = 0\}$; it should be clear now that they are closed in $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$. We define S_0^+ again using (5.38), choosing one non-empty $\hat{S}_p$,

We can finish the proof of Theorem 5.4. Suppose that W changes sign within $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$. Then S_0^+ is a non-empty, closed, connected, embedded null hypersurface within $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$. Now, any embedded null hypersurface S_0^+ is locally two-sided, and we can assign an intersection number one to every intersection point of S_0^+ with a curve that crosses S_0^+ from its local past to its local future, and minus one for the remaining ones (this coincides with the oriented intersection number as in [**45**, Chapter 3]). Let $p \in S_0^+$, there exists a smooth timelike future directed curve γ_1 from some point $q \in \mathscr{M}_{\rm ext}$ to p. By definition there exists a future directed null geodesic segment γ_2 from p to some point $r \in \mathscr{M}_{\rm ext}$ intersecting S precisely at p. Since $\mathscr{M}_{\rm ext}$ is connected there exists a curve $\gamma_3 \subset \mathscr{M}_{\rm ext}$ (which, in fact, cannot be causal future directed, but this is irrelevant for our purposes) from r to q. Then the path γ obtained by following γ_1, then γ_2, and then γ_3 is closed. Since S_0^+ does not extend into $\mathscr{M}_{\rm ext}$, γ intersects S_0^+ only along its timelike future directed part, where every intersection has intersection number one, and γ intersects S_0^+ at least once at p, hence the intersection number of γ with S_0^+ is strictly positive. Now, Corollary 2.4 shows that $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$ is simply connected. But, by standard intersection theory [**45**, Chapter 3], the intersection number of a closed curve with a closed, externally orientable, embedded hypersurface in a simply connected manifold vanishes, which gives a contradiction and proves that W cannot change sign on $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$.

It remains to show that W vanishes at the boundary of $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$. For this, note that, by definition of W, in the region $\{W > 0\}$ the subspace of $T\mathscr{M}$ spanned by the Killing vectors $K_{(\mu)}$ is timelike. Hence at every p such that $W(p) > 0$ there exist vectors of the form $K_{(0)} + \sum \alpha_i K_{(i)}$ which are timelike. But $\partial\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle \subset \dot I^-(\mathscr{M}_{\rm ext}) \cup \dot I^+(\mathscr{M}_{\rm ext})$, and each of the boundaries $\dot I^-(\mathscr{M}_{\rm ext})$ and $\dot I^+(\mathscr{M}_{\rm ext})$ is invariant under the flow of any linear combination of $K_{(\mu)}$'s, and each is achronal, hence $W \le 0$ on $\partial\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$, whence the result. □

In view of what has been said, the reader will conclude:

***Corollary 5.17* (Killing horizon theorem).** — *Under the conditions of Theorem 5.4, away from the set $\mathscr{Z}_{dgt}$ as defined in* (5.9)*, the boundary $\overline{\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle} \setminus \langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$ is a union of embedded Killing horizons.* □

Let us pass now to the

PROOF OF THEOREM 5.6: Let

$$\pi : \langle\langle \mathscr{M}_{\rm ext}\rangle\rangle \cup \mathscr{E}^+ \to \underbrace{\Big(\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle \cup \mathscr{E}^+\Big)/\Big(\mathbb{R}\times \mathrm{U}(1)\Big)}_{=:\mathscr{Q}}$$

denote the quotient map. As discussed in more detail in Sections 6.1 and 6.2 (keeping in mind that, by topological censorship, $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$ has only one asymptotically flat end), the orbit space $\mathscr{Q}$ is diffeomorphic to the half-plane $\{(x,y) \mid x \ge 0\}$ from which a finite number $\mathring{n} \ge 0$ of open half-discs, centred at the axis $\{x = 0\}$, have been removed. As explained at the beginning of Section 7, the case $\mathring{n} = 0$ leads to Minkowski space-time, in which case the result is clear, so from now on we assume $\mathring{n} \geq 1$.

Suppose that $\{W = 0\} \cap \langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$ is non-empty. Let p_0 be an element of this set, with corresponding Killing vector field $l_0 := l_{p_0}$. Let W_0 be the norm squared of l_0:

$$W_0 := \mathfrak{g}(l_0, l_0).$$

In the remainder of the proof of Theorem 5.2 we consider only those $\hat{S}_p$'s for which $l_p = l_0$:

$$\hat{S}_p \subset \{W = 0\} \cap \{W_0 = 0\}.$$

We denote by $C_{\pi(p)}$ the image in $\mathcal{Q}$, under the projection map π, of $\hat{S}_p \cap (\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle \cup \mathscr{E}^+)$. Define

$$\mathring{\mathcal{Q}} = \langle\langle \mathscr{M}_{\rm ext}\rangle\rangle / \Big(\mathbb{R} \times \mathrm{U}(1)\Big),$$

$$\mathscr{W}_0^\flat := \Big(\{W_0 = 0\} \cap \{W = 0\} \cap (\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle \cup \mathscr{E}^+)\Big) / \Big(\mathbb{R} \times \mathrm{U}(1)\Big).$$

Then $\mathscr{W}_0^\flat$ is a closed subset of $\mathcal{Q}$, with the following property: through every point q of $\mathscr{W}_0^\flat$ there exists a smooth maximally extended curve C_q, which will be called *orbit*, entirely contained in $\mathscr{W}_0^\flat$. The C_q's are pairwise disjoint, or coincide. Their union forms a closed set, and locally they look like a subcollection of leaves of a foliation. (Such structures are called laminations; see, e.g., [**39**].)

An orbit will be called a *Jordan orbit* if C_q forms a Jordan curve.

We need to consider several possibilities; we start with the simplest one:

Case I: If an orbit C_q forms a Jordan curve entirely contained in $\mathring{\mathcal{Q}}$, then the corresponding $\hat{S}_p = \pi^{-1}(C_q)$ forms a closed embedded hypersurface in $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$, and a contradiction arises as at the end of the proof of Theorem 5.4.

Case II: Consider, next, an orbit C_q which meets the boundary of $\mathcal{Q}$ at two or more points which belong to $\pi(\mathscr{A})$, and only at such points. Let $I_q \subset C_q$ denote that part of C_q which connects any two subsequent such points, in the sense that I_q meets $\partial\mathcal{Q}$ at its end points only. Now, every $\hat{S}_p$ is a smooth hypersurface in $\mathscr{M}$ invariant under $\mathbb{R} \times \mathrm{U}(1)$, and therefore meets the rotation axis $\mathscr{A}$ orthogonally. This implies that $\pi^{-1}(I_q)$ is a closed, smooth, embedded hypersurface in $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$, providing again a contradiction.

To handle the remaining cases, some preliminary work is needed. It is convenient to double $\mathcal{Q}$ across $\{x = 0\}$ to obtain a manifold $\widehat{\mathcal{Q}}$ diffeomorphic to $\mathbb{R}^2$ from which a finite number of open discs, centered at the axis $\{x = 0\}$, have been removed, see Figure 5.1. Connected components of the event horizon $\mathscr{E}^+$ correspond to smooth circles forming the boundary of $\widehat{\mathcal{Q}}$, regardless of whether or not they are degenerate.

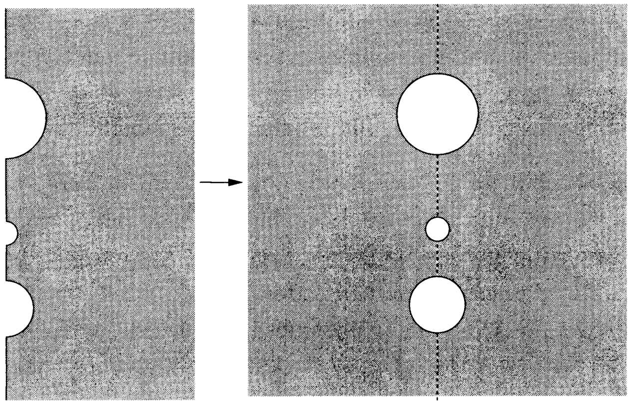

FIGURE 5.1. The quotient space $\mathcal{Q}$ and its double $\widehat{\mathcal{Q}}$.

From what has been said, every C_q which has an end point at $\pi(\mathscr{A})$ is smoothly extended in $\widehat{\mathcal{Q}}$ across $\{x = 0\}$ by its image under the map $(x, y) \mapsto (-x, y)$. We will continue to denote by C_q the orbits so extended in $\widehat{\mathcal{Q}}$.

The analysis of CASES I and II also shows:

Lemma 5.18. — *An orbit C_q which does not meet $\partial\widehat{\mathcal{Q}}$ can cross the axis $\{x = 0\}$ at most once.* □

An orbit C_q will be called an *accumulation orbit* of an orbit C_r if there exists a sequence $q_n \in C_r$ such that $q_n \to q$. Every orbit is its own accumulation orbit. It is a simple consequence of the accumulation Lemma 5.14 that:

Lemma 5.19. — *Let C_q be an accumulation orbit of C_r. Then for every $p \in C_q$ there exists a sequence $p_n \in C_r$ such that $p_n \to p$.* □

We will need the following:

Lemma 5.20. — *Let $r_n \in C_r$ be a sequence accumulating at $p \in \pi(\mathscr{A}) \setminus \partial\widehat{\mathcal{Q}}$. Then $p \in C_r$, and C_r continues smoothly across $\{x = 0\}$ at p.*

Proof. — By Lemma 5.14 there exists an orbit C_p crossing the axis $\{x = 0\}$ transversally at p. Lemma 5.19 shows that C_r crosses the axis. But, by Lemma 5.18, C_r can cross the axis only once. It follows that $C_r = C_p$ and that $p \in C_r$. □

Abusing notation, we still denote by W and W_0 the functions $W \circ \pi$ and $W_0 \circ \pi$. If W and W_0 vanish at a point lying at the boundary $\partial\widehat{\mathcal{Q}}$, then the corresponding circle forms a Jordan orbit. We have:

Lemma 5.21. — *The only orbits accumulating at $\partial\widehat{\mathcal{Q}}$ are the boundary circles.*

Proof. — Suppose that $r_n \in C_q$ accumulates at $p \in \partial\widehat{\mathcal{Q}}$. Then, by continuity, $W(p) = W_0(p) = 0$, which implies that the boundary component through p is a Jordan orbit. But it follows from Lemma 5.19 that any orbit accumulating at $\partial\widehat{\mathcal{Q}}$ has to cross the axis more than once, and the result follows from Lemma 5.18. □

The remaining possibilities will be excluded by a lamination version of the Poincaré-Bendixson theorem. We will make use of a smooth transverse orientation of all the $\hat{S}_p$'s; such a structure is not available for a general lamination, but exists in the problem at hand. More precisely, we will endow $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle \cup \mathscr{E}^+$ with a smooth vector field Z transverse to all $\hat{S}_p$'s. The construction proceeds as follows: Choose any decomposition of $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle \cup \mathscr{E}^+$ as $\mathbb{R} \times \overline{\mathring{\mathscr{S}}}$, as in Theorem 4.5: thus each level set $\overline{\mathring{\mathscr{S}}_t}$ of the time function t is transverse to the stationary Killing vector field K_0, with the periodic Killing vector K_1 tangent to $\overline{\mathring{\mathscr{S}}_t}$. Let $q \in \hat{S}_p \cap \mathring{\mathscr{S}}_0$; as the null leaf $\hat{S}_p$ is transversal to $\mathring{\mathscr{S}}_0$, the intersection $\mathring{\mathscr{S}}_0 \cap \hat{S}_p$ is a hypersurface in $\mathscr{M}$ of co-dimension two. There exist precisely two null directions at q which are normal to $\mathring{\mathscr{S}}_0 \cap \hat{S}_p$, one of them is spanned by $l_0(q)$; we denote by $\mathring{Z}_q$ the unique future directed null vector spanning the other direction and satisfying $\mathring{Z}_q = T_q + \tilde{Z}_q$, where T_q is the unit timelike future directed normal to $\mathring{\mathscr{S}}_0$ at q, and $\tilde{Z}_q$ is tangent to $\mathring{\mathscr{S}}$.

The above definition of $\tilde{Z}_q$ extends by continuity to $q \in \hat{S}_p \cap \overline{\mathring{\mathscr{S}}_0}$.

Transversality and smoothness of l_0 imply that there exists a neighborhood $\mathscr{O}_q$ of q and an extension $\hat{Z}_q$ of $\tilde{Z}_q$ to $\mathscr{O}_q$ with the property that $\hat{Z}_q(r)$ is transverse to $\hat{S}_r$ for every $r \in \mathscr{O}_q$ satisfying $W_0(r) = W(r) = 0$. The neighborhood $\mathscr{O}_q$ can, and will, be chosen to be invariant under $\mathbb{R} \times \mathrm{U}(1)$; similarly for $\hat{Z}_q(r)$.

Consider the covering of $\overline{\mathring{\mathscr{S}}_0} \cap \{W_0 = 0\} \cap \{W = 0\}$ by sets of the form $\mathscr{O}_q \cap \overline{\mathring{\mathscr{S}}_0}$. Asymptotic flatness implies that $\overline{\mathring{\mathscr{S}}_0} \cap \{W_0 = 0\} \cap \{W = 0\}$ is compact, which in turn implies that a finite subcovering $\mathscr{O}_i := \mathscr{O}_{q_i}$ can be chosen. Let φ_i be a partition of unity subordinated to the covering of $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle \cup \mathscr{E}^+$ by the $\mathscr{O}_i$'s together with

$$\mathscr{O}_0 := \Big(\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle \cup \mathscr{E}^+\Big) \setminus \Big(\{W = 0\} \cap \{W_0 = 0\}\Big).$$

The φ_i's can, and will, be chosen to be $\mathbb{R} \times \mathrm{U}(1)$–invariant. Set

$$Z := \sum_{i \geq 1} \varphi_i \hat{Z}_{q_i}.$$

Then Z is smooth, tangent to $\mathring{\mathscr{S}}_0$, and transverse to all $\hat{S}_p$'s.

Choose an orientation of $\widehat{\mathcal{Q}}$. The vector field Z projects under π to a vector field $Z^\flat$ on $\widehat{\mathcal{Q}}$ transverse to each C_q. For each $r \in C_q$ we define a vector $V_q(r)$ by requiring $V_q(r)$ to be tangent to C_1 at r, with $\{V_q, Z^\flat\}$ positively oriented, and with V_q having length one with respect to some auxiliary Riemannian metric on $\widehat{\mathcal{Q}}$. Then V_q varies smoothly along C_q, and each C_q is in fact a complete integral curve of its own V_q.

The vector field V_p along C_p defines an order, and diverging sequences, on C_p in the obvious way: we say that a point $r' \in C_p$ is subsequent to $r \in C_p$ if one flows from r to r' along V_p in the forward direction; a sequence $r_n \in V_p$ is diverging if $r_n = \phi(s_n)(p)$, where $\phi(s)$ is the flow of V_p along C_p, with $s_n \nearrow \infty$ or $s_n \searrow -\infty$.

By Lemma 5.14, if a sequence $r_n \in C_{q_n}$ tends to $r \in C_q$, then the tangent spaces TC_{q_n} accumulate on TC_q. This implies that there exist numbers $\epsilon_n \in \{\pm 1\}$ such that $\epsilon_n V_{q_n}(r_n) \to V_q(r)$, and this is the best one can say in general. However, the existence of Z guarantees that $V_{q_n}(r_n) \to V_q(r)$.

We are ready now to pass to the analysis of

CASE III: In view of Lemmata 5.18 and 5.21, it remains to exclude the existence of orbits C_q which are entirely contained within $\widehat{\mathcal{Q}} \setminus \partial\widehat{\mathcal{Q}}$, and which do not intersect $\pi(\mathscr{A})$, or which intersect $\pi(\mathscr{A})$ only once, and which do *not* form Jordan curves in $\mathring{\mathcal{Q}}$. Since $\{W = 0\} \cap \overline{\mathring{\mathscr{S}}_0}$ is compact, there exists $p \in \widehat{\mathcal{Q}}$ and a diverging sequence $q_n \in C_q$ such that $q_n \to p$. Again by Lemmata 5.18 and 5.21, $p \notin \partial\widehat{\mathcal{Q}}$. The fact that C_p is a closed embedded curve follows now by the standard arguments of the proof of the Poincaré–Bendixson theorem, as e.g. in [**53**]. The orbit C_p does not meet $\partial\widehat{\mathcal{Q}}$ by Lemma 5.21. If C_p met $\pi(\mathscr{A})$, it would have an intersection number with $\{x = 0\}$ equal to one by Lemma 5.18, which is impossible for a Jordan curve in the plane. Thus C_p is entirely contained in $\mathring{\mathcal{Q}}$, which has already been shown to be impossible in CASE I, and the result is established. □

Similarly to Corollary 5.17, we have the following Corollary of Theorem 5.6, which is essentially a rewording of Lemma 5.21:

***Corollary 5.22* (Embedded prehorizons theorem).** — *Under the conditions of Theorem 5.2, away from the set $\mathscr{Z}_{dgt}$ as defined in* (5.9), *the boundary $\overline{\langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle} \setminus \langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle$ is a union of embedded Killing prehorizons.* □

5.3. The ergoset in space-time dimension four. — The *ergoset* E is defined as the set where the stationary Killing vector field $K_{(0)}$ is spacelike or null:

$$E := \{p \mid \mathfrak{g}(K_{(0)}, K_{(0)})|_p \geq 0\}. \tag{5.39}$$

In this section we wish to show that, in vacuum, the ergoset cannot intersect the rotation axis within $\langle\langle \mathscr{M}_{\text{ext}} \rangle\rangle$, if we assume the latter to be chronological.

The first part of the argument is purely local. For this we will assume that the space-time dimension is four, that $K_{(0)} \equiv X$ has no zeros near a point p, that $K_{(1)} \equiv Y$ has 2π–periodic orbits and vanishes at p, and that X and Y commute.

Let $\hat{T}$ be any timelike vector at p, set

$$T := \int_0^{2\pi} \phi_t[Y]_* \hat{T} dt, \tag{5.40}$$

then T is invariant under the flow of Y. Hence $T^\perp$ is also invariant under Y. Let $\mathscr{S}_{\mathscr{O}}$ denote $\exp_p(T^\perp) \cap \mathscr{O}$, where $\mathscr{O}$ is any neighborhood of p lying within the injectivity radius of $\exp_p$, sufficiently small so that $\mathscr{S}_{\mathscr{O}}$ is spacelike; note that $\mathscr{S}_{\mathscr{O}}$ is invariant under the flow of Y. A standard argument (see, e.g., [**2**] Appendix C) shows that Y vanishes on

$$\mathscr{A}_p := \exp_p(\operatorname{Ker} \nabla Y),$$

and that $\mathscr{A}_p$ is totally geodesic. Note that $T \in \operatorname{Ker} \nabla Y$, which implies that $\mathscr{A}_p$ is timelike.

We are interested in the behavior of the area function W near $\mathscr{A}$, the set of points where Y vanishes. We have $\nabla W|_{\mathscr{A}} = 0$ and

$$\begin{aligned}(5.41)\quad \nabla_\mu \nabla_\nu W|_{\mathscr{A}} &= -\nabla_\mu \nabla_\nu \left(\mathfrak{g}(X,X)\mathfrak{g}(Y,Y) - \mathfrak{g}(X,Y)^2\right)\\ &= -2\left(\mathfrak{g}(X,X)\mathfrak{g}(\nabla_\mu Y, \nabla_\nu Y) - \mathfrak{g}(X,\nabla_\mu Y)\mathfrak{g}(X,\nabla_\nu Y)\right).\end{aligned}$$

The second term vanishes because $[X,Y]=0$, with Y vanishing on $\mathscr{A}$:

$$X^\alpha \nabla_\nu Y_\alpha|_{\mathscr{A}} = -X^\alpha \nabla_\alpha Y_\nu = -X^\alpha \nabla_\alpha Y_\nu + \underbrace{Y^\alpha}_{=0} \nabla_\alpha X_\nu = -[X,Y]_\nu = 0.$$

Now, the axis $\mathscr{A}$ is timelike, and the only non-vanishing components of the tensor $\nabla_\mu Y_\nu$ have a spacelike character on $\mathscr{A}$. This implies that the quadratic form $\nabla_\mu Y^\alpha \nabla_\nu Y_\alpha$ is semi-positive definite. We have therefore shown

Lemma 5.23. — *If X is spacelike at $p \in \mathscr{A}$, then $W < 0$ in a neighborhood of p away from $\mathscr{A}$.*

Under the conditions of Theorem 5.1, we conclude that X cannot be spacelike on $\mathscr{A} \cap \langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$. To exclude the possibility that $g(X,X)=0$ there, [16] let w be defined as in (5.17),

$$w = X^\flat \wedge Y^\flat\ ;$$

here, and throughout this section, we explicitly distinguish between a vector Z and its dual $Z^\flat := \mathfrak{g}(Z,\cdot)$. We will further assume that X is causal at p, and that the conclusion of Lemma 5.11 holds:

$$dW \wedge w = W dw. \tag{5.42}$$

Let T denote the field of vectors normal to $\mathscr{S}_{\mathscr{O}}$ normalized so that $\mathfrak{g}(T,X)=1$; note that T_p is, up to a multiplicative factor, as in (5.40). Let γ be any affinely

[16] The analysis in Section 6 shows that X cannot become null on $\mathscr{A} \cap \langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$ when the vacuum equations hold *and* the axis can be identified with a smooth boundary for the metric q; this can be traced to the "boundary point Lemma", which guarantees that the gradient of the harmonic function ρ has no zeros at the boundary $\{\rho = 0\}$. But the behavior of q at those axis points which are not on a non-degenerate horizon and on which X is null is not clear.

parameterized geodesic such that $\gamma(0) = p$, $\dot\gamma(0) \perp T_p$ and $\dot\gamma(0) \perp X_p$; a calculation as in (5.32) shows that

$$\mathfrak{g}(Y,\dot\gamma) = \mathfrak{g}(X,\dot\gamma) = 0$$

along γ. As Y is tangent to $\mathscr{S}_{\mathscr{O}}$, from (5.42) we obtain

$$\underbrace{\frac{dW}{ds}\mathfrak{g}(Y,Y)}_{=dW\wedge X^\flat\wedge Y^\flat(\dot\gamma,T,Y)} = W dw(\dot\gamma,T,Y). \tag{5.43}$$

Now, $i_Y dw = \mathscr{L}_Y w - d(i_Y w) = -d(i_Y w)$, so that

$$\begin{aligned} dw(\dot\gamma,T,Y) &= -d\Big(i_Y(X^\flat\wedge Y^\flat)\Big)(\dot\gamma,T)\\ &= d\Big(-\mathfrak{g}(Y,X)Y^\flat + \mathfrak{g}(Y,Y)X^\flat\Big)(\dot\gamma,T)\\ &= \Big(-\mathfrak{g}(Y,X)dY^\flat + \mathfrak{g}(Y,Y)dX^\flat\Big)(\dot\gamma,T) + \frac{d(\mathfrak{g}(Y,Y))}{ds}. \end{aligned}$$

Inserting this in (5.43), we conclude that

$$\frac{d}{ds}\left(\frac{W}{\mathfrak{g}(Y,Y)}\right) = \underbrace{\Big(-\frac{\mathfrak{g}(Y,X)}{\mathfrak{g}(Y,Y)}dY^\flat + dX^\flat\Big)(\dot\gamma,T)}_{=:f}\times\frac{W}{\mathfrak{g}(Y,Y)}. \tag{5.44}$$

Let h be the metric induced on $\mathscr{S}_{\mathscr{O}}$ by $\mathfrak{g}$. Then h is a Riemannian metric invariant under the flow of Y. As is well known (compare [**19**]) we have $c^{-1}s^2 \le \mathfrak{g}(Y,Y) = h(Y,Y) \le cs^2$. Since $T \in \mathrm{Ker}\nabla Y$ we have $dY^\flat(T,\cdot) = 0$ at p. It follows that the function f defined in (5.44) is bounded along γ near p. If $\mathfrak{g}(X,X) = 0$ at p, then the limit at p of $W/\mathfrak{g}(Y,Y)$ along γ vanishes by (5.41). Using uniqueness of solutions of ODE's, it follows from (5.44) that W vanishes along γ. But this is not possible in $\langle\langle\mathscr{M}_{\mathrm{ext}}\rangle\rangle$ away from $\mathscr{A}$ by Theorem 5.1. We have therefore proved that the ergoset does *not* intersect the axis within $\langle\langle\mathscr{M}_{\mathrm{ext}}\rangle\rangle$:

***Theorem 5.24* (Ergoset theorem).** — *In space-time dimension four, and under the conditions of Theorem 5.1, $K_{(0)}$ is timelike on $\langle\langle\mathscr{M}_{\mathrm{ext}}\rangle\rangle\cap\mathscr{A}$.* □

A higher dimensional version of Theorem 5.24 can be found in [**20**].

A corollary of Theorem 5.24 is that, under the conditions there, the existence of an ergoset implies that of an event horizon. Here one should keep in mind a similar result of Hajiček [**46**], under conditions that include the hypothesis of smoothness of ∂E (which does not hold e.g. in Kerr [**81**]), and affine completeness of those Killing orbits which are geodesics, and non-existence of degenerate Killing horizons. On the other hand, Hajiček assumes the existence of only one Killing vector, while in our work two Killing vectors are required.

6. The reduction to a harmonic map problem

6.1. The orbit space in space-time dimension four. — Let $(\mathscr{M}, \mathfrak{g})$ be a chronological, four-dimensional, asymptotically flat space-time invariant under a $\mathbb{R} \times \mathrm{U}(1)$ action, with stationary Killing vector field $K_{(0)} \equiv X$ and 2π–periodic Killing vector field $K_{(1)} \equiv Y$. Throughout this section we shall assume that

(6.1)

$\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle = \mathbb{R} \times M$, where M is a three dimensional, simply connected manifold with boundary, invariant under the flow of Y, with the flow of X consisting of translations along the $\mathbb{R}$ factor. Moreover the closure $\bar{M}$ of M is the union of a compact set and of a finite number of asymptotically flat ends.

Recall that (6.1) follows from Corollary 2.4 and Theorem 4.5 under appropriate conditions.

Because X and Y commute, the periodic flow of Y on $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$ defines naturally a periodic flow on M; in our context this flow consists of rotations around an axis in the asymptotically flat regions. Now, every asymptotic end can be compactified by adding a point, with the action of U(1) extending to the compactified manifold by fixing the point at infinity. Similarly every boundary component has to be a sphere [**50**, Lemma 4.9], which can be filled in by a ball, with the (unique) action of U(1) on S^2 extending to the interior as the associated rotation of a ball in $\mathbb{R}^3$, reducing the analysis of the group action to the boundaryless case. Existence of asymptotically flat regions, or of boundary spheres, implies that the set of fixed points of the action is non-empty (see, e.g., [**6**, Proposition 2.4]). Assuming, for notational simplicity, that there is only one asymptotically flat end, it then follows from [**83**] (see the italicized paragraph on p.52 there) that, after the addition of a ball B_i to every boundary component, and after the addition of a point i_0 at infinity to the asymptotic region, the new manifold $M \cup B_i \cup \{i_0\}$ is homeomorphic to S^3, with the action of U(1) conjugate, by a homeomorphism, to the usual rotations of S^3. On the other hand, it is shown in [**79**, Theorem 1.10] that the actions are classified, up to smooth conjugation, by topological invariants, so that the action of U(1) is smoothly conjugate to the usual rotations of S^3. It follows that the manifold $M \cup B_i$ is diffeomorphic to $\mathbb{R}^3$, with the U(1) action smoothly conjugate to the usual rotations of $\mathbb{R}^3$. In particular: a) there exists a global cross-section $\mathring{M}^2$ for the action of U(1) on $M \cup B_i$ away from the set of fixed points $\mathscr{A}$,[17] with $\mathring{M}^2$ diffeomorphic to an open half-plane; b) all isotropy groups are trivial or equal to U(1); c) $\mathscr{A}$ is diffeomorphic to $\mathbb{R}$.[18]

(17) We will use the symbol $\mathscr{A}$ to denote the set of fixed points of the Killing vector Y in M *or* in $\mathscr{M}$, as should be clear from the context.

(18) We are grateful to Allen Hatcher for clarifying comments on the classification of U(1) actions.

Somewhat more generally, the above analysis applies whenever M can be compactified by adding a finite number of points or balls. A nontrivial example is provided by manifolds with a finite number of asymptotically flat and asymptotically cylindrical ends, as is the case for the Cauchy surfaces for the domain of outer communication of the extreme Kerr solution.

Summarizing, under (6.1) there exists in $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$ an embedded two-dimensional manifold $\bar{M}^2$, diffeomorphic to $\hat{M}^2 \approx [0,\infty) \times \mathbb{R}$ minus a finite number of points (corresponding to the remaining asymptotic ends), and minus a finite number of open half-discs (the boundary of each corresponding to a connected component of the horizon). We denote by M^2 the manifold obtained by removing from $\bar{M}^2$ all its boundaries.

6.2. Global coordinates on the orbit space. — We turn our attention now to the construction of a convenient coordinate system on a four-dimensional, globally hyperbolic, $\mathbb{R} \times \mathrm{U}(1)$ invariant, simply connected domain of outer communications $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$. Let $\bar{M}^2$ and $\mathring{M}^2$ be as in Section 6.1. We will invoke the uniformization theorem to understand the geometry of $\bar{M}^2$; however, some preparatory work is useful, which will allow us to control both the asymptotic behavior of the fields involved, as well as the boundary conditions at various boundaries.

For simplicity we assume that $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$ contains only one asymptotically flat region, which is necessarily the case under the hypotheses of Theorem 2.3. On M^2 there is a naturally defined orbit space-metric which, away from the rotation axis $\{Y = 0\}$, is defined as follows. Let us denote by $\mathfrak{g}$ the metric on space-time, let $X_1 = X$, $X_2 = Y$, set $h_{ij} = \mathfrak{g}(X_i, X_j)$, let h^{ij} denote the matrix inverse to h_{ij} wherever defined, and on that last set for $Z_1, Z_2 \in T_p\mathring{M}^2$ set

$$q(Z_1, Z_2) = \mathfrak{g}(Z_1, Z_2) - h^{ij}\mathfrak{g}(Z_1, X_i)\mathfrak{g}(Z_2, X_j). \tag{6.2}$$

Note that if Z_1 and Z_2 are orthogonal to the Killing vectors, then $q(Z_1, Z_2) = \mathfrak{g}(Z_1.Z_2)$. This implies that if the linear span of the Killing vectors is timelike (which, under our hypotheses below, is the case away from the axis $\{Y = 0\}$ in the domain of outer communications), then q is positive definite on the space orthogonal to the Killing vectors. Also note that q is independent of the choice of the basis of the space of Killing vectors.

To take advantage of the asymptotic analysis in [**19**], a straightforward calculation shows that q equals

$$q(Z_1, Z_2) = \gamma(Z_1, Z_2) - \frac{\gamma(Y, Z_1)\gamma(Y, Z_2)}{\gamma(Y, Y)}, \tag{6.3}$$

where γ is the (obviously $\mathrm{U}(1)$–invariant) metric on the level sets of t (where t is any time function as in Section 6.1) obtained from the space-time metric by a formula similar to (6.2):

$$\gamma(Z_1, Z_2) = \mathfrak{g}(Z_1, Z_2) - \frac{\mathfrak{g}(Z_1, X)\mathfrak{g}(Z_2, X)}{\mathfrak{g}(X, X)}. \tag{6.4}$$

(So γ is *not* the metric induced on the level sets of t by $\mathfrak{g}$.) The right-hand-side is manifestly well-behaved in the region where X is timelike; this is the case in the asymptotic region, and near the axis on $\langle\langle \mathscr{M}_{\rm ext} \rangle\rangle$ under the conditions of Theorem 5.24.

In any case, the asymptotic analysis of [**19**] can be invoked directly to obtain information about the metric q at large distances. Recall that if the asymptotic flatness conditions (2.1) hold with $k \geq 1$, then by the field equations (2.1) holds with k arbitrarily large. We can thus use [**19**] to conclude that there exist coordinates x^A, covering the complement of a compact set in $\mathbb{R}^2$ after the quotient space has been doubled across the rotation axis, in which q is manifestly asymptotically flat as well (see Proposition 2.2 and Remark 2.8 in [**19**]):

$$q_{AB} - \delta_{AB} = o_{k-3}(r^{-1}). \tag{6.5}$$

To gain insight into the geometry of q near the horizons, one can use (6.4) with X being instead the Killing vector which is null on the horizon. It is then shown in [**18**] that each non-degenerate component of the horizon corresponds to a smooth totally geodesic boundary for γ. (It is also shown there that every degenerate component corresponds to a metrically complete end of infinite extent *provided* that the Killing vector tangent to the generators of the horizon is timelike on $\langle\langle \mathscr{M}_{\rm ext} \rangle\rangle$ near the horizon, but it is not clear that this property holds.) Some information on the asymptotic geometry of γ in the degenerate case can be obtained from [**47, 66**]; whether or not the information there suffices to extend our analysis below to the non-degenerate case remains to be seen.

6.3. All horizons non-degenerate. — Assuming that all horizons are non-degenerate, we proceed as follows: Every non-degenerate component of the boundary ∂M is a smooth sphere S^2 invariant under U(1). As is well known, every isometry of S^2 is smoothly conjugate to the action of rotations around the z axis in a flat $\mathbb{R}^3$, with the rotation axis meeting S^2 at exactly two points. Thus, as already mentioned in Section 6.1, we can fill each component of the boundary ∂M by a smooth ball B^3, with a rotation-invariant metric there. We denote by γ any rotation-invariant smooth Riemannian metric on $\mathbb{R}^3$ which extends the original metric γ, and by q the associated two-dimensional metric as in (6.3). From what has been said we conclude that every non-degenerate component of the horizon corresponds to a smooth boundary $\partial M/\mathrm{U}(1)$ for the metric q, consisting of a segment which meets the rotation axis at precisely two points. The filling-in just described is equivalent to filling in a half-disc in the quotient manifold. Since the boundary ∂M is a smooth

U(1) invariant surface for γ, it meets the rotation axis orthogonally. This implies that each one-dimensional boundary segment of $\partial M/\mathrm{U}(1)$ meets the rotation axis orthogonally in the metric q.

Consider, then, a black hole space-time which contains one asymptotically flat end and N non-degenerate spherical horizons. After adding N half-discs as described above, the quotient space, denoted by $\hat{M}^2$, is then a two-dimensional non-compact asymptotically flat manifold diffeomorphic to a half-plane. Recall that we are assuming (6.1), and that there is only one asymptotically flat region. We will also suppose that

(6.6) $\qquad W > 0$ on $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle \setminus \mathscr{A}$, and

(6.7) $\qquad$ on $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle \cap \mathscr{A}$ the stationary Killing vector field X is timelike.

Note that those conditions necessarily hold under the hypotheses of Theorem 5.1, compare Theorem 5.24.

By (6.6) the metric q is positive definite away from $\mathscr{A}$. Near $\mathscr{A}$ the metric γ defined in (6.4) is Riemannian and smooth by (6.7), and the analysis in [**19**] shows that $\mathscr{A}$ is a smooth boundary for q. After doubling across the boundary, one obtains an asymptotically flat metric on $\mathbb{R}^2$. By [**19**, Proposition 2.3], for $k \geq 5$ in (2.1) there exist global isothermal coordinates for q:

$$q = e^{2u}(dx^2 + dy^2), \quad \text{with } u \longrightarrow_{\sqrt{x^2+y^2}\to\infty} 0. \tag{6.8}$$

In fact, $u = o_{k-4}(r^{-1})$. The existence of such coordinates also follows from the uniformization theorem (see, e.g., [**1**]), but this theorem does not seem to provide the information about the asymptotic behavior in various regimes, needed here, in any obvious way. As explained in the proof of [**19**, Theorem 2.7], the coordinates (x, y) can be chosen so that the rotation axis corresponds to $x = 0$, with $\hat{M}^2 = \{x \geq 0\}$.

The next step of the construction is to modify the coordinates (x, y) of (6.8) to a coordinate system (ρ, z) on the quotient manifold $\bar{M}^2$, covering $[0, \infty) \times \mathbb{R}$, so that ρ vanishes on the rotation axis *and* the event horizons. This is done by first solving the equation

$$\Delta_q \rho_R = 0,$$

on $\Omega_R := \bar{M}^2 \cap \{x^2 + y^2 \leq R^2\}$, with zero boundary value on $\partial\bar{M}^2$, and with $\rho_R = x$ on $\{x^2 + y^2 = R^2\}$. Note that

$$C = \sup_{\partial\Omega_R \setminus \mathscr{A}} x - \rho_R,$$

is independent of R, for R large, since x and ρ_R differ only on the event horizons. Since $\Delta_q x = 0$, the maximum principle implies

$$x - C \leq \rho_R \leq x \ \text{ on } \Omega_R.$$

By usual arguments there exists a subsequence ρ_{R_i} which converges, as i tends to infinity, to a q–harmonic function ρ on $\bar{M}^2$, satisfying the desired boundary values. By standard asymptotic expansions (see, e.g., [**15**]) we find that $\nabla\rho$ approaches ∇x as $\sqrt{x^2+y^2}\to\infty$. In fact, for any $j\in\mathbb{N}$ we have

$$\rho - x = \sum_{i=0}^{j} \frac{\alpha_i(\varphi)}{(x^2+y^2)^{i/2}} + O((x^2+y^2)^{-(j+1)/2}) \,, \tag{6.9}$$

where φ denotes an angular coordinate in the (x,y) plane, with α_i being linear combinations of $\cos(i\varphi)$ and $\sin(i\varphi)$, with the expansion being preserved under differentiation in the obvious way. In particular $\nabla\rho$ does not vanish for large x, so that for R sufficiently large the level sets $\{\rho = R\}$ are smooth submanifolds. The strips $0<\rho<R$ are simply connected so, by the uniformization theorem, there exists a holomorphic diffeomorphism

$$(x,y)\mapsto(\alpha(x,y),\beta(x,y))$$

from that strip to the set $\{0<\alpha<R,\ \beta\in\mathbb{R}\}$. By composing with a Möbius map we can further arrange so that the point at infinity of the (x,y)–variables is mapped to the point at infinity of the (α,β)–variables. As the map is holomorphic, the function $\alpha(x,y)$ is harmonic, with the same boundary values and boundary and asymptotic conditions as ρ, hence $\alpha(x,y)=\rho(x,y)$ wherever both are defined. If we denote by z a harmonic conjugate to ρ, we similarly obtain that $z-\beta$ is a constant, so that the map

$$(x,y)\mapsto(\rho,z) \tag{6.10}$$

is a holomorphic diffeomorphism between the strips described above. Since the constant R was arbitrarily large, we conclude that the map (6.10) provides a holomorphic diffeomorphism from the interior of $\bar{M}^2$ to $\{\rho>0,\ z\in\mathbb{R}\}$, and provides the desired coordinate system in which q takes the form

$$q=e^{2\hat{u}}(d\rho^2+dz^2). \tag{6.11}$$

From (6.9) and its equivalent for z (which is immediately obtained from the defining equations $\partial_x\rho=\partial_y z$, $\partial_y\rho=-\partial_x z$) we infer that $\hat{u}\to 0$ as $\sqrt{\rho^2+z^2}$ goes to infinity, with the decay rate $\hat{u}=o_{k-4}(r^{-1})$ remaining valid in the new coordinates.

In vacuum the area function W satisfies $\Delta_q\sqrt{W}=0$ (see, e.g., [**91**]). If we assume that W vanishes on $\partial\langle\langle\mathscr{M}_{\rm ext}\rangle\rangle\cup\mathscr{A}$ (which is the case under the hypotheses of Theorem 5.1), then $W=\rho$ on $\partial\langle\langle\mathscr{M}_{\rm ext}\rangle\rangle\cup\mathscr{A}$. Since $\Delta_q\rho=0$ as well, we have $\Delta_q(\sqrt{W}-\rho)=0$, with $W-\rho$ going to zero as one tends to infinity by [**19**], and the maximum principle gives

$$\sqrt{W}=\rho. \tag{6.12}$$

6.4. Global coordinates on $\langle\langle\mathscr{M}_{\mathrm{ext}}\rangle\rangle$. — According to Section 6.1 we have

$$\langle\langle\mathscr{M}_{\mathrm{ext}}\rangle\rangle \setminus \mathscr{A} \approx \mathbb{R} \times S^1 \times \mathbb{R}^*_+ \times \mathbb{R},$$

and this diffeomorphism defines a global coordinate system (t, φ, ρ, z) on $\langle\langle\mathscr{M}_{\mathrm{ext}}\rangle\rangle \setminus \mathscr{A}$, with $X = \partial_t$ and $Y = \partial_\varphi$. Letting $(x^A) = (\rho, z)$ and $(x^a) = (t, \varphi)$, we can write the metric in the form

$$\mathfrak{g} = \mathfrak{g}_{ab}(dx^a + \underbrace{\theta^a{}_A dx^A}_{=:\theta^a})(dx^b + \theta^b{}_B dx^B) + q_{AB} dx^A dx^B,$$

with all functions independent of t and φ. The orthogonal integrability condition of Proposition 5.3 gives

$$d\theta^a = 0,$$

so that, by simple connectedness of $\mathbb{R}^*_+ \times \mathbb{R}$, there exist functions f^a such that $\theta^a = df^a$. Redefining the x^a's to $x^a + f^a$, and keeping the same symbols for the new coordinates, we conclude that the metric on $\langle\langle\mathscr{M}_{\mathrm{ext}}\rangle\rangle \setminus \mathscr{A}$ has a global coordinate representation as

$$\mathfrak{g} = -\rho^2 e^{2\lambda} dt^2 + e^{-2\lambda}(d\varphi - v dt)^2 + e^{2\hat{u}}(d\rho^2 + dz^2) \tag{6.13}$$

for some functions $v(\rho, z)$, $\lambda(\rho, z)$, with ρ, z and $\hat{u}$ as in Section 6.3, see in particular (6.12). We set

$$U = \lambda + \ln \rho, \quad \text{so that} \quad \mathfrak{g}(\partial_\varphi, \partial_\varphi) = \rho^2 e^{-2U} = e^{-2\lambda}. \tag{6.14}$$

Let ω be the *twist potential* defined by the equation

$$d\omega = *(dY \wedge Y), \tag{6.15}$$

its existence follows from simple-connectedness of $\langle\langle\mathscr{M}_{\mathrm{ext}}\rangle\rangle$ and from $d * (dY \wedge Y) = 0$ (see, e.g.,[**91**]). As discussed in more detail in Section 6.7 below (compare [**91**, Proposition 2]), the space-time metric is uniquely determined by the axisymmetric map

$$\Phi = (\lambda, \omega) : \mathbb{R}^3 \setminus \mathscr{A} \to \mathbb{H}^2, \tag{6.16}$$

where $\mathbb{H}^2$ is the hyperbolic space with metric

$$b := d\lambda^2 + e^{4\lambda} d\omega^2, \tag{6.17}$$

and $\mathscr{A}$ is the rotation axis $\mathscr{A} := \{(0,0,z), z \in \mathbb{R}\} \subset \mathbb{R}^3$. The metric coefficients can be determined from Φ by solving equations (6.45)-(6.47) below. The map Φ solves the harmonic map equations [**36, 88**]:

$$|T|^2_b := (\Delta\lambda - 2e^{4\lambda}|D\omega|^2)^2 + e^{4\lambda}(\Delta\omega + 4D\lambda \cdot D\omega)^2 = 0, \tag{6.18}$$

where both D and Δ refer to the flat metric on $\mathbb{R}^3$, together with a set of asymptotic conditions depending upon the configuration at hand.

We continue with the derivation of those boundary conditions.

6.5. Boundary conditions at non-degenerate horizons. — Near the points at which the boundary is analytic (so, e.g., at those points of the axis at which X is timelike), the map defined by (6.10) extends to a holomorphic map across the boundary (see, e.g., [**30**]). This implies that $\hat{u}$ extends across the axis as a smooth function of ρ^2 and z *away* from the set of points $\{\mathfrak{g}(X,X)=0\}$.

Let us now analyze the behavior of $\hat{u}$ near the points $z_i \in \mathscr{A}$ where non-degenerate horizons meet the axis. As described above, after performing a constant shift in the y coordinate, any component of a non-degenerate horizon can locally be described by a smooth curve in the $\zeta := x + iy$ plane of the form

$$y = \gamma(x), \ \gamma(0) = 0, \ \gamma(x) = \gamma(-x). \tag{6.19}$$

Near the origin, the points lying in the domain of outer communications correspond then to the values of $x+iy$ lying in a region, say Ω, bounded by the half-axis $\{x = 0, y \geq 0\}$ and by the curve $x + i\gamma(x)$, with $x \geq 0$.

To get rid of the right-angle-corner where the curve $x+i\gamma(x)$ meets the axis, the obvious first attempt is to introduce a new complex coordinate

$$w := \alpha + i\beta = -i\zeta^2. \tag{6.20}$$

If we write $\gamma(x) = a_2x^2 + O(x^4)$, then the image of $\{x+i\gamma(x), \ x \geq 0\}$ under (6.20) becomes

$$\begin{aligned} f_1(x+i\gamma(x)) &= 2a_2x^3 + O(x^5) - i\underbrace{(x^2 - a_2^2x^4 + O(x^6))}_{=:-t} \\ &= it + 2a_2|t|^{3/2} + O(|t|^{5/2}). \end{aligned} \tag{6.21}$$

The remaining part $\{iy, \ y \in \mathbb{R}^+\}$, of the boundary of Ω, is mapped to itself. It follows that the boundary of the image of Ω by the map (6.20) is a $C^{1,1/2}$ curve. Here $C^{k,\lambda}$ denotes the space of k-times differentiable functions, the k'th derivatives of which satisfy a Hölder condition with index λ.

To improve the regularity we replace $-i\zeta^2$ by $f_2(\zeta) = -i\zeta^2 + \sigma_3\zeta^3$ for some constant σ_3. Then (6.21) becomes

$$\begin{aligned} f_2(x+i\gamma(x)) &= (2a_2 + \Re\sigma_3)x^3 + O(x^5) - i\underbrace{(x^2 + O(x^4))}_{=:-t} - \Im(\sigma_3)O(x^4) \\ &= it + (2a_2 + \Re\sigma_3)|t|^{3/2} + O(|t|^{5/2}). \end{aligned} \tag{6.22}$$

The remaining part of the boundary of Ω is mapped to the curve $f_2(iy)$, with $y \geq 0$:

$$\begin{aligned} f_2(iy) &= \Im\sigma_3 y^3 + i\underbrace{(y^2 - \Re\sigma_3 y^3)}_{=:t} \\ &= it + \Im\sigma_3(|t|^{3/2} + O(|t|^2)). \end{aligned} \tag{6.23}$$

and is thus mapped to itself if σ_3 is real. Choosing $\sigma_3 = -2a_2 \in \mathbb{R}$ one gets rid of the offending $|t|^{3/2}$ terms in (6.22)-(6.23), resulting in the boundary of $f_2(\Omega)$ of $C^{2,1/2}$ differentiability class.

More generally, suppose that the image of $x + i\gamma(x)$ by the polynomial map $\zeta \mapsto w = f_{k-1}(\zeta) = -i\zeta^2 + \dots$ has a real part equal to $\beta_{2k-1}x^{2k-1} + O(x^{2k+1})$; then the substraction from f_{k-1} of a term $\beta_{2k-1}\zeta^{2k-1}$ leads to a new polynomial map $\zeta \to w = f_k(\zeta)$ which has real part $\beta_{2k+1}x^{2k+1} + O(x^{2k+3})$, and the differentiability of the image has been improved by one. Since all the coefficients β_{2k+1} are real, the maps f_k map the imaginary axis to itself. One should note that this argument wouldn't work if γ had odd powers of x in its Taylor expansion.

Summarizing, for any k we can choose a finite polynomial $f_k(\zeta)$, with lowest order term $-i\zeta^2$, and with the remaining coefficients real and involving only odd powers of ζ, which maps the boundary of Ω to a curve

$$(-\epsilon, \epsilon) \ni t \mapsto (\mu(t), \nu(t)) := \begin{cases} (0, t), & t \geq 0; \\ (O(t^{k+1/2}), t), & t \leq 0, \end{cases} \tag{6.24}$$

which is $C^{k,1/2}$.

Note that

$$\psi_k(\zeta) := \sqrt{if_k(\zeta)} = \zeta\Big(1 + O(|\zeta|)\Big), \tag{6.25}$$

where $\sqrt{\cdot}$ denotes the principal branch of the square root, is a holomorphic diffeomorphism near the origin. So

$$w = f_k(\zeta) = -i\psi_k^2(\zeta) \tag{6.26}$$

and we have

$$dw\, d\bar{w} = 4|\psi_k \psi_k'|^2 d\zeta\, d\bar{\zeta} = 4|w||\psi_k'|^2 d\zeta\, d\bar{\zeta}. \tag{6.27}$$

We claim that the map

$$w \mapsto \eta := \rho + iz$$

extends across $\rho = 0$ to a C^k diffeomorphism near the origin. To see this, note that we have again $\Delta\rho = 0$ with respect to the metric $dwd\bar{w}$, with ρ vanishing on a $C^{k,1/2}$ boundary. We can straighten the boundary using the transformation

$$w = (\alpha, \beta) \mapsto (\alpha - \mu(\beta), \beta) = w + (O(|\beta|^{k+1/2}), 0) = w + O(|w|^{k+1/2}), \tag{6.28}$$

where μ is as (6.24), and $O(\cdot)$ is understood for small $|w|$. Extending ρ with $-\rho$ across the new boundary, one can use the standard interior Schauder estimates on the extended function to conclude that $w \mapsto \rho(w)$ is $C^{k,1/2}$ up-to-boundary. Now, the condition $dz = \star d\rho$, where $\star$ is the Hodge dual of the metric q, is conformally invariant and therefore holds in the metric $dw\, d\bar{w}$, so z is a $C^{k,1/2}$ function of w. By the boundary version of the maximum principle we have $d\rho \neq 0$ at the boundary (when understood as a function of w), and hence near the boundary, so dz is non-vanishing near the boundary and orthogonal to $d\rho$. The implicit function theorem allows us to conclude that the map $w \mapsto \eta$ is a $C^{k,1/2}$ diffeomorphism near $w = 0$.

Comparing (6.8) and (6.11) we have

$$e^{2\hat{u}} d\eta\, d\bar{\eta} = q = e^{2u} d\zeta\, d\bar{\zeta} = \frac{e^{2u}}{4|w||\psi_k'|^2} dw\, d\bar{w}, \tag{6.29}$$

in particular $dw\, d\bar{w} = e^{2\tilde{u}_k} d\eta\, d\bar{\eta}$, and from what has been said the function $\tilde{u}_k$ is $C^{k-1,1/2}$ up to boundary. Hence

$$e^{2\hat{u}} = \frac{e^{2u+2\tilde{u}_k}}{4|w||\psi_k'|^2} \tag{6.30}$$

where u is a smooth function of (x^2, y), while ψ_k' is a non-vanishing holomorphic function of $\zeta = x + iy$, $\tilde{u}_k$ is a C^{k-1} function of $\eta = \rho + iz$, and $\eta \mapsto w$ is a C^k diffeomorphism, with w having a zero of order one where the horizon meets the axis. Finally $x + iy$ is a holomorphic function of $\sqrt{iw}$, compare (6.26).

Choosing $k = 2$ we obtain

$$\hat{u} = -\frac{1}{2} \ln |w| + \hat{u}_1 + \hat{u}_2, \tag{6.31}$$

where w is a smooth complex coordinate which vanishes where the horizon meets the axis, $\hat{u}_2 = -\ln |\psi_2'|^2/2$ is a smooth function of (x, y), and $\hat{u}_1$ is a C^1 function of (ρ, z).

Taylor expanding at the origin, from what has been said (recall that $\eta \mapsto w$ is conformal and that, near the origin, $\{\rho = 0\}$ coincides with $\{\alpha - \mu(\beta) = 0\}$) it follows that there exists a real number $a > 0$ such that

$$(\rho, z) = \big(a^{-2}(\alpha - \mu(\beta)), a^{-2}\beta\big) + O((\alpha - \mu(\beta))^2 + \beta^2),$$

which implies

$$(\alpha, \beta) = (a^2\rho, a^2 z) + O(\rho^2 + z^2). \tag{6.32}$$

Here we have assumed that z has been shifted by a constant so that it vanishes at the chosen intersection point of the axis and of the event horizon.

We conclude that there exists a constant C such that

$$|\hat{u} + \frac{1}{2} \ln \sqrt{\rho^2 + z^2}| \leq C \quad \text{near } (0,0). \tag{6.33}$$

This is the desired equation describing the leading order behavior of $\hat{u}$ near the meeting point of the axis and a non-degenerate horizon.

6.5.1. *The Ernst potential.* — We continue by deriving the boundary conditions satisfied by the Ernst potential (U, ω) near the point where the horizon meets the axis. Here U is as in (6.13)-(6.14), and ω is obtained from the function v appearing in the metric by solving (6.45) below.

Our analysis so far can be summarized as:

$$x+iy=\zeta \;\mapsto\; \psi_k(\zeta)=\sqrt{if_k(\zeta)} \;\mapsto\; -i(\psi_k(\zeta))^2=w \;\mapsto\; \rho+iz. \tag{6.34}$$

Each map is invertible on the sets under consideration; and each is a C^k diffeomorphism up-to-boundary except for the middle one, which involves the squaring of a complex number.

Using $\zeta=\psi_k^{-1}(\sqrt{iw})$, the expansion

$$\psi_k^{-1}(c+id)=(c+id)\Big(1+O(\sqrt{c^2+d^2})\Big),$$

which follows from (6.25), together with (6.32), we obtain

$$x+iy=a\sqrt{-z+i\rho}+O(\rho^2+z^2).$$

Equivalently,

$$x=\frac{a\rho}{\sqrt{2(z+\sqrt{z^2+\rho^2})}}+O(\rho^2+z^2),\quad y=a\sqrt{\frac{z+\sqrt{z^2+\rho^2}}{2}}+O(\rho^2+z^2). \tag{6.35}$$

To continue, in addition to (6.1), (6.6) and (6.7) we assume that

(6.36) The level sets of the function t, defined as the projection on the $\mathbb{R}$ factor in (6.1), are spacelike, with $\partial_\varphi t=0$;

this is justified for our purposes by Theorem 4.5. Thus, the Killing vector ∂_φ is tangent to the level sets of t, so that

$$\mathfrak{g}(\partial_\varphi,\partial_\varphi)=h(\partial_\varphi,\partial_\varphi),$$

where h is the Riemannian metric induced on the level sets of t. As shown in [**19**], we have

$$h(\partial_\varphi,\partial_\varphi)=f(x,y)x^2, \tag{6.37}$$

where the function $f(x,y)$ is uniformly bounded above and below on compact sets.

Recall that U has been defined as $-\frac{1}{2}\ln(\mathfrak{g}_{\varphi\varphi}\rho^{-2})$, and that (ρ,z) have been normalized so that $(0,0)$ corresponds to a point where a non-degenerate horizon meets the axis. We want to show that

$$U=\ln\sqrt{z+\sqrt{z^2+\rho^2}}+O(1)\;\text{ near }(0,0). \tag{6.38}$$

(This formula can be checked for the Kerr metrics by a direct calculation, but we emphasize that we are considering a general non-degenerate horizon.) To see that, we use (6.37) to obtain

$$\ln(\mathfrak{g}_{\varphi\varphi}\rho^{-2})=\ln(x^2\rho^{-2})+\ln(\mathfrak{g}_{\varphi\varphi}x^{-2})=2\ln(x\rho^{-1})+O(1).$$

We assume that ρ^2+z^2 is sufficiently small, as required by the calculations that follow. In the region $0 \le |z| \le 2\rho$ we use (6.35) as follows:

$$\begin{aligned}\ln(x\,\rho^{\,-1}) &= \ln\left(\frac{a+\sqrt{2\left(\frac{z}{\rho}+\sqrt{\frac{z^2}{\rho^2}+1}\right)}O(\rho^{3/2}+\frac{z^2}{\rho^{1/2}})}{\sqrt{2(z+\sqrt{z^2+\rho^2})}}\right)\\ &= -\ln\left(\sqrt{2(z+\sqrt{z^2+\rho^2})}\right)+O(1).\end{aligned}$$

In the region $z \le 0$ we note that

$$\begin{aligned}\frac{1}{\rho}\sqrt{2(z+\sqrt{z^2+\rho^2})} &= \frac{\sqrt{2(z+\sqrt{z^2+\rho^2})}\sqrt{2(-z+\sqrt{z^2+\rho^2})}}{\rho\sqrt{2(-z+\sqrt{z^2+\rho^2})}}\\ &= \frac{2}{\sqrt{2(\,-z+\sqrt{z^2+\rho^2})}} \le \frac{\sqrt{2}}{(z^2+\rho^2)^{1/4}}.\end{aligned}$$

Hence, again by (6.35),

$$\begin{aligned}\ln(x\,\rho^{\,-1}) &= \ln\left(\frac{a\;+\frac{1}{\rho}\sqrt{2(z+\sqrt{z^2+\rho^2})}O(\rho^2+z^2)}{\sqrt{2(z+\sqrt{z^2+\rho^2})}}\right)\\ &= \ln\left(\frac{a\;+O\big((\rho^2+z^2)^{3/4}\big)}{\sqrt{2(z+\sqrt{z^2+\rho^2})}}\right) = -\ln\left(\sqrt{2(z+\sqrt{z^2+\rho^2})}\right)+O(1).\end{aligned}$$

In the region $0 \le \rho \le z/2$ some more work is needed. Instead of (6.35), we want to use a Taylor expansion of ρ around the axis $\alpha = 0$, where α is as in (6.20). To simplify the calculations, note that there is no loss of generality in assuming that the map ψ_k of (6.25) is the identity, by redefining the original (x, y) coordinates to the new ones obtained from ψ_k. Since in the region $0 \le \rho \le z/2$ we have $\beta \ge 0$, the function $\mu(\beta)$ in (6.28) vanishes, so

$$\alpha(\rho,z) = \underbrace{\alpha(0,z)}_{=\mu\left(\beta(0,z)\right)=0} + \partial_\rho\alpha(0,z)\rho + O(\rho^2) \;=\; \partial_\rho\alpha(0,z)\rho + O(\rho^2).$$

Note that $\partial_\rho\alpha(0,z)$ tends to a^2 as z tends to zero, so is strictly positive for z small enough. Instead of (6.35) we now have directly

$$x = \frac{\alpha}{\sqrt{2(\beta+\sqrt{\beta^2+\alpha^2})}} \qquad \Longrightarrow \qquad \frac{x}{\rho} = \frac{\partial_\rho\alpha(0,z)\;+O(\rho)}{\sqrt{2(\beta+\sqrt{\beta^2+\alpha^2})}}.$$

In the current region α is equivalent to ρ, β is equivalent to z, $\sqrt{\beta^2+\alpha^2}$ is equivalent to z, and z is equivalent to $2(z+\sqrt{z^2+\rho^2})$, which leads to the desired formula:

$$\begin{aligned}\ln(x\rho^{-1}) &= -\ln\left(\sqrt{2(\beta+\sqrt{\beta^2+\alpha^2})}\right)+O(1)\\ &= -\ln\left(\sqrt{2(z+\sqrt{z^2+\rho^2})}\frac{\sqrt{2(\beta+\sqrt{\beta^2+\alpha^2})}}{\sqrt{2(z+\sqrt{z^2+\rho^2})}}\right)+O(1)\\ &= -\ln\left(\sqrt{2(z+\sqrt{z^2+\rho^2})}\right)+O(1)\,.\end{aligned}$$

This finishes the proof of (6.38).

Let us turn our attention now to the twist potential ω: as is well known, or from [**24**, Equation (2.6)] together with the analysis in [**19**], ω is a smooth function of (x, y), constant on the axis $\{x = 0\}$, with odd x–derivatives vanishing there. So, Taylor expanding in x, there exists a constant ω_0 and a bounded function $\mathring{\omega}$ such that

$$\begin{aligned}\omega &= \omega_0+\mathring{\omega}(x,y)x^2\\ &= \omega_0+\frac{\mathring{\omega}(x,y)\left(a\rho+\sqrt{2(z+\sqrt{z^2+\rho^2})}O(\rho^2+z^2)\right)^2}{2(z+\sqrt{z^2+\rho^2})}.\end{aligned}\tag{6.39}$$

In our approach below, the proof of black hole uniqueness requires a uniform bound on the distance between the relevant harmonic maps. Now, using the coordinates (λ, ω) on hyperbolic space as in (6.17), the distance d_b between two points (x_1, ω_1) and (x_2, ω_2) is implicitly defined by the formula [**3**, Theorem 7.2.1]:

$$\cosh(d_b)-1 = \frac{(e^{-2x_1}-e^{-2x_2})^2+4(\omega_1-\omega_2)^2}{2e^{-2x_1-2x_2}}.$$

Using the (U, ω) parameterization of the maps, with U as in (6.14), the distance measured in the hyperbolic plane between two such maps is the supremum of the function d_b:

$$\begin{aligned}\cosh(d_b)-1 &= \frac{\rho^4(e^{-2U_1}-e^{-2U_2})^2+4(\omega_1-\omega_2)^2}{2\rho^4e^{-2U_1-2U_2}}\\ &= \frac{1}{2}\underbrace{(e^{2(U_1-U_2)}+e^{2(U_2-U_1)}-2)}_{(a)}+2\underbrace{\rho^{-4}e^{2(U_1+U_2)}(\omega_1-\omega_2)^2}_{(b)}.\end{aligned}$$

Inserting (6.38) and the analogous expansion for the Ernst potential of a second metric into (a) above we obviously obtain a bounded contribution. Finally, assuming $\omega_1(0,0) = \omega_2(0,0)$, up to a multiplicative factor which is uniformly bounded above and bounded away from zero, (b) can be rewritten as a square of the difference of two terms of the form

$$f_i := \mathring{\omega}_i\left(a_i+\rho^{-1}\sqrt{2(z+\sqrt{z^2+\rho^2})}O(\rho^2+z^2)\right)^2,\tag{6.40}$$

with $i = 1, 2$. We have the following, for all $z^2 + \rho^2 \le 1$:

1. The functions f_i in (6.40) are uniformly bounded in the sector $|z| \le \rho$:
$$|f_i| \le C\Big(a_i + \sqrt{2(z + \sqrt{z^2+\rho^2})}O(\rho + z^2/\rho)\Big)^2 \le C'.$$
2. For $0 \le \rho \le -z$ we write
$$0 \le z + \sqrt{z^2+\rho^2} = |z|(\sqrt{1+\frac{\rho^2}{z^2}} - 1) \le C\frac{\rho^2}{|z|},$$
so that
$$|f_i| \le C\Big(a_i + \frac{1}{|z|^{1/2}}O(\rho^2 + z^2)\Big)^2 = C(a_i + O(|z|^{3/2}))^2 \le C'.$$
3. For $0 \le \rho \le z$ one can proceed as follows: by (6.37), together with the analysis of ω in [**19**], there exists a constant C such that near the axis we have
$$C^{-1}x^2 \le \mathfrak{g}(\partial_\varphi, \partial_\varphi) = h(\partial_\varphi, \partial_\varphi) \le Cx^2, \quad \Big|\omega - \underbrace{\omega|_{x=0}}_{=:\omega_0}\Big| \le Cx^2 \tag{6.41}$$
(recall that h denotes the metric induced by $\mathfrak{g}$ on the slices $t =$ const, where t is a time function invariant under the flow of ∂_φ). But

(6.42)
$$\begin{aligned}\frac{(\omega_1-\omega_2)^2}{\rho^4 e^{-2U_1-2U_2}} &= \frac{(\omega_1-\omega_2)^2}{\mathfrak{g}_1(\partial_\varphi,\partial_\varphi)\mathfrak{g}_2(\partial_\varphi,\partial_\varphi)} \le 2\frac{(\omega_1-\omega_0)^2 + (\omega_2-\omega_0)^2}{\mathfrak{g}_1(\partial_\varphi,\partial_\varphi)\mathfrak{g}_2(\partial_\varphi,\partial_\varphi)}\\ &= 2\underbrace{\left(\frac{\omega_1-\omega_0}{\mathfrak{g}_1(\partial_\varphi,\partial_\varphi)}\right)^2}_{\le C^2}\underbrace{\frac{\mathfrak{g}_1(\partial_\varphi,\partial_\varphi)}{\mathfrak{g}_2(\partial_\varphi,\partial_\varphi)}}_{=e^{2(U_2-U_1)}} + 2\underbrace{\left(\frac{\omega_2-\omega_0}{\mathfrak{g}_2(\partial_\varphi,\partial_\varphi)}\right)^2}_{\le C^2}\underbrace{\frac{\mathfrak{g}_2(\partial_\varphi,\partial_\varphi)}{\mathfrak{g}_1(\partial_\varphi,\partial_\varphi)}}_{=e^{2(U_1-U_2)}},\end{aligned}$$
where $\mathfrak{g}_i$ denotes the respective space-time metric, while x_i denotes the respective x coordinate. Uniform boundedness of this expression, in a neighborhood of the intersection point, follows now from (6.38).

We are ready now to prove one of the significant missing elements of all previous uniqueness claims for the Kerr metric:

Theorem 6.1. — *Suppose that* (6.1), (6.6)-(6.7) *and* (6.36) *hold. Let* (U_i, ω_i), $i = 1, 2$, *be the Ernst potentials associated with two vacuum, stationary, asymptotically flat axisymmetric metrics with smooth non-degenerate event horizons. If* $\omega_1 = \omega_2$ *on the rotation axis, then the hyperbolic-space distance between* (U_1, ω_1) *and* (U_2, ω_2) *is bounded, going to zero as* r *tends to infinity in the asymptotic region.*

Proof. — We have just proved that the distance between two different Ernst potentials is bounded near the intersection points of the horizon and of the axis. In view of (6.7), the distance is bounded on bounded subsets of the axis away from the horizon intersection points by the analysis in [**19**]. Next, both ω_a's are bounded on the horizon, and both functions $\rho^2 e^{-2U_a}$'s are bounded on the horizon away from its end points. Finally, both ω_a's approach the Kerr twist potential at infinity by the results in [**87**] (the asymptotic Poincaré Lemma 8.7 in [**21**] is useful in this context), so the distance approaches zero as one recedes to infinity by a calculation as in (6.42), together with the asymptotic analysis of [**19**]; a more detailed exposition can be found in [**31**]. □

6.6. The harmonic map problem: existence and uniqueness. — In this section we consider Ernst maps satisfying the following conditions, modeled on the local behavior of the Kerr solutions:

1. There exists $N_{\mathrm{dh}} \geq 0$ degenerate event horizons, which are represented by punctures $(\rho = 0, z = b_i)$, together with a mass parameter $m_i > 0$ and angular momentum parameter $a_i = \pm m_i$, with the following behavior for small $r_i := \sqrt{\rho^2 + (z - b_i)^2}$,

$$U = \ln\left(\frac{r_i}{2m_i}\right) + \frac{1}{2}\ln\left(1 + \frac{(z-b_i)^2}{r_i^2}\right) + O(r_i). \tag{6.43}$$

   The twist potential ω is a bounded, angle-dependent function which jumps by $-4J_i = -4a_i m_i$ when crossing b_i from $z < b_i$ to $z > b_i$, where J_i is the "angular momentum of the puncture".
2. There exists $N_{\mathrm{ndh}} \geq 0$ non-degenerate horizons, which are represented by bounded open intervals $(c_i^-, c_i^+) = I_i \subset \mathscr{A}$, with none of the previous b_j's belonging to the union of the closures of the I_i. The functions $U - 2\ln\rho$ and ω extend smoothly across each interval I_i, with the following behavior near the end points, for some constant C, as derived in (6.38):

$$|U - \frac{1}{2}\ln(\sqrt{\rho^2 + (z - c_i^{\pm})^2} + z - c_i^{\pm})| \leq C \quad \text{near } (0, c_i^{\pm}). \tag{6.44}$$

   The function ω is assumed to be locally constant on $\mathscr{A} \setminus (\cup_i\{b_i\} \cup_j I_j)$, with expansions as in (6.39) nearby.
3. The functions U and ω are smooth across $\mathscr{A} \setminus (\cup_i\{b_i\} \cup_j I_j)$.

A collection $\{b_i, m_i\}_{i=1}^{N_{\mathrm{dh}}}$, I_j, $j = 1, \ldots, N_{\mathrm{ndh}}$, and $\{\omega_k\}$, where the ω_k's are the values of ω_i on the connected components of $\mathscr{A} \setminus (\cup_i\{b_i\} \cup_j I_j)$, will be called *"axis data"*.

We have the following [**24**, Appendix C] (compare [**33, 93**] and references therein for previous related results):

Theorem 6.2. — *For any set of axis data there exists a unique harmonic map $\Phi : \mathbb{R}^3 \setminus \mathscr{A} \to \mathbb{H}^2$ which lies a finite distance from a solution with the properties 1.–3. above, and such that $\omega = 0$ on $\mathscr{A}$ for large positive z.* □

Here the distance between two maps Φ_1 and Φ_2 is defined as

$$d(\Phi_1, \Phi_2) = \sup_{p \in \mathbb{R}^3 \setminus \mathscr{A}} d_b(\Phi_1(p), \Phi_2(p)),$$

where the distance d_b is taken with respect to the hyperbolic metric (6.17).

We emphasize the following corollary, first established by Robinson [**84**] using different methods (and assuming $|a| < m$, which Weinstein [**91**] does not); the approach presented here is due to Weinstein [**91**]: (19)

Corollary 6.3. — *For each mass parameter m and angular momentum parameter $a \in (-m, m)$ there exists only one map Φ with the behavior at the axis corresponding to an I^+-regular axisymmetric vacuum black hole with a connected non-degenerate horizon centered at the origin and with ω vanishing on $\mathscr{A}$ for large positive z. Furthermore, no I^+-regular non-degenerate axisymmetric vacuum black holes with $|a| \geq m$ exist.*

Proof. — Theorem 4.5 shows that (6.1) and (6.36) hold, (6.6) follows from Theorem 5.1, while (6.7) holds by the Ergoset Theorem 5.24. One can thus introduce (ρ, z) coordinates on the orbit space as in Section 6.2, then the event horizon corresponds to a connected interval of the axis of length ℓ, for some $\ell > 0$. Let (U, ω) be the Ernst potential corresponding to the black hole under consideration, with ω normalized to vanish on $\mathscr{A}$ for large positive z. Let J be the total angular momentum of the black hole, there exists a Kerr solution (U_K, ω_K), with ω_K normalized to vanish on $\mathscr{A}$ for large positive z, and such that the corresponding "horizon interval" has the same length ℓ. We can adjust the z coordinate so that the horizon intervals coincide. The value of ω on the axis for large negative z equals $4J$, similarly for ω_K, hence $\omega = \omega_K$ on the axis except possibly on the horizon interval. Theorem 6.1 shows that (U, ω) lies at a finite distance from (U_K, ω_K). By the uniqueness part of Theorem 6.2 we find $(U, \omega) = (U_K, \omega_K)$, thus the ADM mass of the black hole equals the mass of the comparison Kerr solution, and $|a| < m$ follows. □

6.7. Candidate solutions. — Each harmonic map (λ, ω) of Theorem 6.2 with $N_{\mathrm{dh}} + N_{\mathrm{ndh}} \geq 1$ provides a candidate for a solution with $N_{\mathrm{dh}} + N_{\mathrm{ndh}}$ components of the event horizon, as follows: let the functions v and $\hat{u}$ be the unique solutions of the

(19) Yet another approach can be found in [**77**]; compare [**72**, Section 2.4]. In order to become complete, the proof there needs to be complemented by a justification of the assumed behavior of their potential Φ (not to be confused with the map Φ here) on the set $\{\rho = 0\}$. More precisely, one needs to justify differentiability of Φ on $\{\rho = 0\}$ away from the horizons, continuity of Φ and Φ' at the points where the horizon meets the rotation axis, as well as the detailed differentiability properties of Φ near degenerate horizons as implicitly assumed in [**72**, Section 2.4].

set of equations

$$\partial_\rho v = -e^{4\lambda}\rho\,\partial_z\omega, \qquad \partial_z v = e^{4\lambda}\rho\,\partial_\rho\omega, \tag{6.45}$$

$$\partial_\rho \hat{u} = \rho\left[(\partial_\rho\lambda)^2 - (\partial_z\lambda)^2 + \tfrac{1}{4}e^{4\lambda}((\partial_\rho\omega)^2 - (\partial_z\omega)^2)\right] + \partial_\rho\lambda, \tag{6.46}$$

$$\partial_z \hat{u} = 2\,\rho\left[\partial_\rho\lambda\,\partial_z\lambda + \tfrac{1}{4}e^{4\lambda}\partial_\rho\omega\,\partial_z\omega\right] + \partial_z\lambda, \tag{6.47}$$

which go to zero at infinity. (Those equations are compatible whenever (λ, ω) satisfy the harmonic map equations.) Then the metric (6.13) satisfies the vacuum Einstein equations (see, e.g., [**95**, Eqs. (2.19)-(2.22)]). Every such solution provides a candidate for a regular, vacuum, stationary, axisymmetric black hole with several components of the event horizon. If $N_{\mathrm{dh}} + N_{\mathrm{ndh}} = 1$ the resulting metrics are of course the Kerr ones.

At the time of writing of this work, it is not known whether any such candidate solution other than Kerr itself describes an I^+–regular black hole. It should be emphasized that there are two separate issues here: The first is that of uniqueness, which is settled by the uniqueness part of Theorem 6.2 together with the remaining analysis in this section: *if* there exist stationary axisymmetric multi-black hole solutions, with all components of the horizon non-degenerate, *then* they belong to the family described by the harmonic maps of Theorem 6.2. Note that Theorem 6.2 extends to those solutions with degenerate horizons with the behavior described in (6.43). Conceivably this covers all degenerate horizons, but this remains to be established.

Another question is that of the global properties of the candidate solutions: for this one needs, first, to study the behavior of the harmonic maps of Theorem 6.2 near the singular set in much more detail in order to establish e.g. existence of a smooth event horizon; an analysis of this issue has only been done so far [**69, 91**] if $N_{\mathrm{dh}} = 0$ away from the points where the axis meets the horizon, and the question of space-time regularity at those points is wide open. Regardless of this, one expects that for all such solutions the integration of the remaining equations (6.45)-(6.47) will lead to singular "struts" in the space-time metric (6.13) somewhere on $\mathscr{A}$.

7. Proof of Theorem 1.3

If $\mathscr{E}^+$ is empty, the conclusion follows from the Komar identity and the rigid positive energy theorem (see, e.g. [**18**, Section 4]). Otherwise the proof splits into two cases, according to whether or not X is tangent to the generators of $\mathscr{E}^+$, to be covered separately in Sections 7.1 and 7.2.

7.1. Rotating horizons. — Suppose, first, that the Killing vector is not tangent to the generators of some connected component $\mathscr{E}^+_0$ of $\mathscr{E}^+ = \mathscr{H}^+ \cap I^+(\mathscr{M}_{\mathrm{ext}})$. Theorem 4.14 shows that the isometry group of $(\mathscr{M}, \mathfrak{g})$ contains $\mathbb{R}\times \mathrm{U}(1)$. By Corollary 2.4

$\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$ is simply connected so that, in view of Theorem 4.5, the analysis of Section 6 applies, leading to the global representation (6.13) of the metric. The analysis of the behavior near the symmetry axis of the harmonic map Φ of Section 6.5 shows that Φ lies a finite distance from one of the solutions of Theorem 6.2, and the uniqueness part of that last theorem allows us to conclude; compare Corollary 6.3 in the connected case.

7.2. Non-rotating case. — The case where the stationary Killing vector X is tangent to the generators of every component of $\mathscr{H}^+$ will be referred to as the *non-rotating one.* By hypothesis $\nabla(\mathfrak{g}(X, X))$ has no zeros on $\mathscr{E}^+$, so all components of the future event horizon are non-degenerate.

Deforming $\mathscr{S}$ near $\partial\mathscr{S}$ if necessary, we may without loss of generality assume that $\mathscr{S}$ can be extended across $\mathscr{E}^+$ to a smooth spacelike hypersurface there.

For the proof we need a new hypersurface $\mathscr{S}''$ which is maximal, Cauchy for $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$, with X vanishing on $\partial\mathscr{S}''$. Under our hypotheses such a hypersurface will not exist in general, so we start by replacing $(\mathscr{M}, \mathfrak{g})$ by a new space-time $(\mathscr{M}', \mathfrak{g}')$ with the following properties:

1. $(\mathscr{M}', \mathfrak{g}')$ contains a region $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle'$ isometric to $(\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle, \mathfrak{g})$;
2. $(\mathscr{M}', \mathfrak{g}')$ is invariant under the flow of a Killing vector X' which coincides with X on $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$;
3. Each connected component of the horizon $\mathscr{E}_0^{+\prime}$ is contained in a bifurcate Killing horizon, which contains a "bifurcation surface" where X' vanishes. We will denote by S the union of these bifurcation surfaces.

This is done by attaching to $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$ a bifurcate horizon near each connected component of $\mathscr{E}^+$ as in [**82**], invoking Corollary 5.17.

We wish, now to construct a Cauchy surface $\mathscr{S}'$ for $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle'$ such that $\partial\mathscr{S}' = S$. To do that, for $\epsilon > 0$ let $\mathfrak{g}_\epsilon$ denote a family of metrics such that $\mathfrak{g}_\epsilon$ tends to $\mathfrak{g}$, as ϵ goes to zero, uniformly on compact sets, with the property that null directions for $\mathfrak{g}_\epsilon$ are spacelike for $\mathfrak{g}$. Consider the family of $\mathfrak{g}_\epsilon$–null Lipschitz hypersurfaces

$$\mathscr{N}_\epsilon := \dot{J}_\epsilon^+(S) \cap \mathscr{M},$$

where $\dot{J}_\epsilon^+$ denotes the boundary of the causal future with respect to the metric $\mathfrak{g}_\epsilon$. The $\mathscr{N}_\epsilon$'s are threaded with $\mathfrak{g}_\epsilon$–null geodesics, with initial points on S, which converge uniformly to $\mathfrak{g}$-null geodesics starting from S, hence to the generators of $\mathscr{E}^+$ (within $\mathscr{M}'$). It follows that, for all ϵ small enough, $\mathscr{N}_\epsilon$ intersects $\mathscr{S}$ transversally. Furthermore, since $\mathscr{E}^+$ is smooth, decreasing ϵ if necessary, continuity of Jacobi fields with respect to ϵ implies that the $\mathscr{N}_\epsilon$'s remain smooth in the portion between S and their intersection with $\mathscr{S}$. Choosing ϵ small enough, one obtains a smooth $\mathfrak{g}$-spacelike hypersurface $\mathscr{S}'$, with boundary at S, by taking the union of the portion of $\mathscr{N}_\epsilon$ between S and where it meets $\mathscr{S}$, with that portion of $\mathscr{S}$ which extends to infinity and which is bounded by the intersection with $\mathscr{N}_\epsilon$, and smoothing out the intersection. The hypersurface $\mathscr{S}'$ can be shown to be Cauchy by the usual arguments [**9, 40**].

By [**27**] there exists an asymptotically flat Cauchy hypersurface $\mathscr{S}''$ for $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$, with boundary on S, which is maximal.

We wish to show, now, that $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle'$, and hence $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$, are static; this has been first proved in [**89**], but a rather simple proof proceeds as follows: Let us decompose X' as $Nn+Z$, where n is the future-directed normal to $\mathscr{S}''$, while Z is tangent. The space-time Killing equations imply

$$D_iZ_j + D_jZ_i = -2NK_{ij}, \tag{7.1}$$

where g_{ij} is the metric induced on $\mathscr{S}''$, K_{ij} is its extrinsic curvature tensor, and D is the covariant derivative operator of g_{ij}. Since $\mathscr{S}''$ is maximal, the (vacuum) momentum constraint reads

$$D_iK^{ij} = 0. \tag{7.2}$$

From (7.1)-(7.2) one obtains

$$D_i(K^{ij}Z_j) = -NK^{ij}K_{ij}. \tag{7.3}$$

Integrating (7.3) over $\mathscr{S}''$, the boundary integral in the asymptotically flat regions gives no contribution because K_{ij} approaches zero there as $O(1/r^{n-1})$, while Z approaches zero there as $O(1/r^{n-2})$ [**25**]. The boundary integral at the horizons vanishes since Z and N vanish on $S=\partial\mathscr{S}''$ by construction. Hence

$$\int_{\mathscr{S}''} NK^{ij}K_{ij} = 0. \tag{7.4}$$

On a maximal hypersurface the normal component N of a Killing vector satisfies the equation

$$\Delta N = K^{ij}K_{ij}N, \tag{7.5}$$

and the maximum principle shows that N is strictly positive except at $\partial\mathscr{S}''$. Staticity of $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle'$ along $\mathscr{S}''$ follows now from (7.4). Moving the $\mathscr{S}''$'s with the isometry group one covers $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle'$ [**27**], and staticity of $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle'$ follows. Hence $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$ is static as well, and Theorem 1.4 allows us to conclude that $\langle\langle \mathscr{M}_{\rm ext}\rangle\rangle$ is Schwarzschildian. This achieves the proof of Theorem 1.3. □

8. Concluding remarks

To obtain a satisfactory uniqueness theory in four dimensions, the following issues remain to be addressed:

1. The previous versions of the uniqueness theorem required analyticity of *both* the metric *and* the horizon. As shown in Theorem 4.11, the latter follows from the former. This is a worthwhile improvement, as even C^1*–differentiability of the horizon* is not clear a priori. But the hypothesis of analyticity of the metric remains to be removed.

   In this context one should keep in mind the Curzon solution, where analyticity of the metric fails precisely at the horizon. We further note an interesting recent uniqueness theorem for Kerr without analyticity conditions [**59**]. However, the examples constructed at the end of Section 2.3.1 show that further insights are needed to be able to conclude along the lines envisaged there.

   The hypothesis of analyticity is particularly annoying in the static context, being needed there only to exclude non-embedded Killing prehorizons. The nature of that problem seems to be rather different from Hawking's rigidity, with presumably a simpler solution, yet to be found.
2. The question of uniqueness of black holes with degenerate components of the Killing horizon requires further investigations. Recall that non-existence of stationary, vacuum, I^+–regular black holes with *all* components of the event horizon non-rotating and degenerate, follows immediately from the Komar identity and the positive energy theorem [**58**] (compare [**18**, Section 4]). Furthermore, the results here go a long way to prove uniqueness of degenerate, stationary, axisymmetric, rotating configurations: the only element missing is an equivalent of Theorem 6.1. We expect that Theorem 2.2 can be useful for solving this problem, and we hope to return to that question in the near future.

   In any case, the above would not cover solutions with degenerate non-rotating components. One could exclude such solutions by proving existence of maximal hypersurfaces within $\langle\langle \mathscr{M}_{\mathrm{ext}} \rangle\rangle$ with an appropriate asymptotic behavior at the cylindrical ends. The argument presented in Section 7.2 would then apply to give staticity, and non-existence would then follow from [**26**], or from Theorem 1.4.
3. The question of existence of multi-component solutions needs to be settled.

And, of course, the question of classification of higher dimensional stationary black holes is largely unchartered territory.

Acknowledgements. — We are grateful to J. Isenberg for numerous comments on a previous version of the paper. PTC is grateful to R. Wald and G. Weinstein for useful discussions. Similarly JLC wishes to thank J. Natário for many useful discussions.

References

[1] W. Abikoff – "The uniformization theorem", *Amer. Math. Monthly* **88** (1981), p. 574–592.

[2] M. T. ANDERSON, P. T. CHRUŚCIEL & E. DELAY – "Non-trivial, static, geodesically complete space-times with a negative cosmological constant. II. $n \geq 5$", in *AdS/CFT correspondence: Einstein metrics and their conformal boundaries*, IRMA Lect. Math. Theor. Phys., vol. 8, Eur. Math. Soc., Zürich, 2005, p. 165–204.

[3] A. F. BEARDON – *The geometry of discrete groups*, Graduate Texts in Math., vol. 91, Springer, 1983.

[4] R. BEIG & P. T. CHRUŚCIEL – "Killing initial data", *Class. Quantum Grav.* **14** (1996), p. A83–A92, a special issue in honour of Andrzej Trautman on the occasion of his 64^{th} Birthday.

[5] ______, "Killing vectors in asymptotically flat space-times. I. Asymptotically translational Killing vectors and the rigid positive energy theorem", *J. Math. Phys.* **37** (1996), p. 1939–1961.

[6] ______, "The isometry groups of asymptotically flat, asymptotically empty space-times with timelike ADM four-momentum", *Comm. Math. Phys.* **188** (1997), p. 585–597.

[7] ______, "The asymptotics of stationary electro-vacuum metrics in odd spacetime dimensions", *Classical Quantum Gravity* **24** (2007), p. 867–874.

[8] R. H. BOYER – "Geodesic Killing orbits and bifurcate Killing horizons", *Proc. Roy. Soc. London Ser. A* **311** (1969), p. 245–252.

[9] R. BUDIC, J. ISENBERG, L. LINDBLOM & P. B. YASSKIN – "On determination of Cauchy surfaces from intrinsic properties", *Comm. Math. Phys.* **61** (1978), p. 87–95.

[10] G. L. BUNTING & A. K. M. MASOOD-UL ALAM – "Nonexistence of multiple black holes in asymptotically Euclidean static vacuum space-time", *Gen. Relativity Gravitation* **19** (1987), p. 147–154.

[11] B. CARTER – "Killing horizons and orthogonally transitive groups in space-time", *J. Mathematical Phys.* **10** (1969), p. 70–81.

[12] ______, "Black hole equilibrium states", in *Black holes/Les astres occlus (École d'Été Phys. Théor., Les Houches, 1972)*, Gordon and Breach, 1973, p. 57–214.

[13] ______, "Bunting identity and Mazur identity for nonlinear elliptic systems including the black hole equilibrium problem", *Comm. Math. Phys.* **99** (1985), p. 563–591.

[14] U. CHRIST & J. LOHKAMP – "Singular minimal hypersurfaces and scalar curvature", preprint arXiv:math.DG/0609338, 2006.

[15] P. T. CHRUŚCIEL – "Asymptotic estimates in weighted Hölder spaces for a class of elliptic scale-covariant second order operators", *Ann. Fac. Sci. Toulouse Math.* **11** (1990), p. 21–37.

[16] ______, "Uniqueness of stationary, electro-vacuum black holes revisited", *Helv. Phys. Acta* **69** (1996), p. 529–552, Journées Relativistes 96, Part II (Ascona, 1996).

[17] ______, "On rigidity of analytic black holes", *Comm. Math. Phys.* **189** (1997), p. 1–7.

[18] ______, "The classification of static vacuum spacetimes containing an asymptotically flat spacelike hypersurface with compact interior", *Classical Quantum Gravity* **16** (1999), p. 661–687, corrigendum in arXiv:gr-qc/9809088.

[19] ______, "Mass and angular-momentum inequalities for axi-symmetric initial data sets. I. Positivity of mass", *Annals Phys.* **323** (2008), p. 2566–2590.

[20] ______, "On higher dimensional black holes with abelian isometry group", in preparation, 2008.

[21] P. T. CHRUŚCIEL & E. DELAY – "On mapping properties of the general relativistic constraints operator in weighted function spaces, with applications", *Mém. Soc. Math. Fr. (N.S.)* **94** (2003).

[22] P. T. CHRUŚCIEL, E. DELAY, G. J. GALLOWAY & R. HOWARD – "Regularity of horizons and the area theorem", *Ann. Henri Poincaré* **2** (2001), p. 109–178.

[23] P. T. Chruściel, G. J. Galloway & D. Solis – "On the topology of Kaluza-Klein black holes", preprint arXiv:0808.3233, 2008.

[24] P. T. Chruściel, Y. Y. Li & G. Weinstein – "Mass and angular-momentum inequalities for axi-symmetric initial data sets. II. Angular momentum", *Annals Phys.* **323** (2008), p. 2591–2613.

[25] P. T. Chruściel & D. Maerten – "Killing vectors in asymptotically flat space-times. II. Asymptotically translational Killing vectors and the rigid positive energy theorem in higher dimensions", *J. Math. Phys.* **47** (2006), p. 022502, 10.

[26] P. T. Chruściel, H. S. Reall & P. Tod – "On non-existence of static vacuum black holes with degenerate components of the event horizon", *Classical Quantum Gravity* **23** (2006), p. 549–554.

[27] P. T. Chruściel & R. M. Wald – "Maximal hypersurfaces in stationary asymptotically flat spacetimes", *Comm. Math. Phys.* **163** (1994), p. 561–604.

[28] ______, "On the topology of stationary black holes", *Classical Quantum Gravity* **11** (1994), p. L147–L152.

[29] F. H. Clarke – *Optimization and nonsmooth analysis*, second ed., Classics in Applied Mathematics, vol. 5, Society for Industrial and Applied Mathematics (SIAM), 1990.

[30] H. Cohn – *Conformal mapping on Riemann surfaces*, Dover Publications Inc., 1980, Reprint of the 1967 edition, Dover Books on Advanced Mathematics.

[31] J. Costa – "On black hole uniqueness theorems", Ph.D. Thesis, University of Oxford, in preparation.

[32] S. Dain – "Initial data for stationary spacetimes near spacelike infinity", *Classical Quantum Gravity* **18** (2001), p. 4329–4338.

[33] ______, "Proof of the angular momentum-mass inequality for axisymmetric black holes", *J. Differential Geom.* **79** (2008), p. 33–67.

[34] T. Damour & B. Schmidt – "Reliability of perturbation theory in general relativity", *J. Math. Phys.* **31** (1990), p. 2441–2453.

[35] R. Emparan & H. S. Reall – "Generalized Weyl solutions", *Phys. Rev. D* **65** (2002), p. 084025.

[36] F. Ernst – "New formulation of the axially symmetric gravitational field problem", *Phys. Rev.* **167** (1968), p. 1175–1178.

[37] J. L. Friedman, K. Schleich & D. M. Witt – "Topological censorship", *Phys. Rev. Lett.* **71** (1993), p. 1486–1489, erratum idem **75** (1995) p. 1872.

[38] H. Friedrich, I. Rácz & R. M. Wald – "On the rigidity theorem for spacetimes with a stationary event horizon or a compact Cauchy horizon", *Comm. Math. Phys.* **204** (1999), p. 691–707.

[39] D. Gabai – "3 lectures on foliations and laminations on 3-manifolds", in *Laminations and foliations in dynamics, geometry and topology (Stony Brook, NY, 1998)*, Contemp. Math., vol. 269, Amer. Math. Soc., 2001, p. 87–109.

[40] G. J. Galloway – "Some results on Cauchy surface criteria in Lorentzian geometry", *Illinois J. Math.* **29** (1985), p. 1–10.

[41] ______, "On the topology of the domain of outer communication", *Classical Quantum Gravity* **12** (1995), p. L99–L101.

[42] ______, "A "finite infinity" version of topological censorship", *Classical Quantum Gravity* **13** (1996), p. 1471–1478.

[43] ______, "Maximum principles for null hypersurfaces and null splitting theorems", *Ann. Henri Poincaré* **1** (2000), p. 543–567.

[44] G. J. Galloway, K. Schleich, D. M. Witt & E. Woolgar – "Topological censorship and higher genus black holes", *Phys. Rev. D* **60** (1999), p. 104039.

[45] V. GUILLEMIN & A. POLLACK – *Differential topology*, Prentice-Hall Inc., 1974.

[46] P. HAJIČEK – "General theory of vacuum ergospheres", *Phys. Rev.* **D7** (1973), p. 2311–2316.

[47] P. HÁJÍČEK – "Three remarks on axisymmetric stationary horizons", *Comm. Math. Phys.* **36** (1974), p. 305–320.

[48] S. W. HAWKING – "Black holes in general relativity", *Comm. Math. Phys.* **25** (1972), p. 152–166.

[49] S. W. HAWKING & G. F. R. ELLIS – *The large scale structure of space-time*, Cambridge University Press, 1973, Cambridge Monographs on Mathematical Physics, No. 1.

[50] J. HEMPEL – *3-Manifolds*, Princeton University Press, 1976, Ann. of Math. Studies, No. 86.

[51] M. HEUSLER – *Black hole uniqueness theorems*, Cambridge Lecture Notes in Physics, vol. 6, Cambridge University Press, 1996.

[52] N. J. HICKS – *Notes on differential geometry*, Van Nostrand Mathematical Studies, No. 3, D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto-London, 1965.

[53] M. W. HIRSCH, S. SMALE & R. L. DEVANEY – *Differential equations, dynamical systems, and an introduction to chaos*, second ed., Pure and Applied Mathematics (Amsterdam), vol. 60, Elsevier/Academic Press, Amsterdam, 2004.

[54] S. HOLLANDS & A. ISHIBASHI – "On the 'stationary implies axisymmetric' theorem for extremal black holes in higher dimensions", preprint arXiv:0809.2659, 2008.

[55] S. HOLLANDS, A. ISHIBASHI & R. M. WALD – "A higher dimensional stationary rotating black hole must be axisymmetric", *Comm. Math. Phys.* **271** (2007), p. 699–722.

[56] S. HOLLANDS & S. YAZADJIEV – "Uniqueness theorem for 5-dimensional black holes with two axial Killing fields", preprint arXiv:0707.2775, 2007.

[57] ______, "A uniqueness theorem for five-dimensional Einstein-Maxwell black holes", *Classical Quantum Gravity* **25** (2008), p. 095010.

[58] G. T. HOROWITZ – "The positive energy theorem and its extensions", in *Asymptotic behavior of mass and spacetime geometry (Corvallis, Ore., 1983)*, Lecture Notes in Phys., vol. 202, Springer, 1984, p. 1–21.

[59] A. IONESCU & S. KLAINERMAN – "On the uniqueness of smooth, stationary black holes in vacuum", preprint arXiv:0711.0040, 2007.

[60] J. ISENBERG & V. MONCRIEF – "Symmetries of higher dimensional black holes", *Class. Quantum Grav.* **25** (2008), p. 195015.

[61] W. ISRAEL – "Event horizons in static vacuum space-times", *Phys. Rev.* **164** (1967), p. 1776–1779.

[62] B. S. KAY & R. M. WALD – "Theorems on the uniqueness and thermal properties of stationary, nonsingular, quasifree states on spacetimes with a bifurcate Killing horizon", *Phys. Rep.* **207** (1991), p. 49–136.

[63] S. KOBAYASHI – "Fixed points of isometries", *Nagoya Math. J.* **13** (1958), p. 63–68.

[64] S. KOBAYASHI & K. NOMIZU – *Foundations of differential geometry. Vol I*, Interscience Publishers, a division of John Wiley & Sons, New York-London, 1963.

[65] W. KUNDT & M. TRÜMPER – "Orthogonal decomposition of axi-symmetric stationary spacetimes", *Z. Physik* **192** (1966), p. 419–422.

[66] J. LEWANDOWSKI & T. PAWLOWSKI – "Extremal isolated horizons: a local uniqueness theorem", *Classical Quantum Gravity* **20** (2003), p. 587–606.

[67] ______, "Quasi-local rotating black holes in higher dimension: geometry", *Classical Quantum Gravity* **22** (2005), p. 1573–1598.

[68] ______, "Symmetric non-expanding horizons", *Classical Quantum Gravity* **23** (2006), p. 6031–6058.

[69] Y. Y. Li & G. Tian – "Regularity of harmonic maps with prescribed singularities", *Comm. Math. Phys.* **149** (1992), p. 1–30.

[70] J. Lohkamp – "The higher dimensional positive mass theorem. I", preprint arXiv:math.DG/0608795, 2006.

[71] P. Mazur – "Proof of uniqueness of the Kerr-Newman black hole solution", *J. Phys. A* **15** (1982), p. 3173–3180.

[72] R. Meinel, M. Ansorg, A. Kleinwächter, G. Neugebauer & D. Petroff – *Relativistic figures of equilibrium*, Cambridge University Press, 2008.

[73] V. Moncrief – "Neighborhoods of Cauchy horizons in cosmological spacetimes with one Killing field", *Ann. Physics* **141** (1982), p. 83–103.

[74] V. Moncrief & J. Isenberg – "Symmetries of cosmological Cauchy horizons", *Comm. Math. Phys.* **89** (1983), p. 387–413.

[75] C. B. J. Morrey – *Multiple integrals in the calculus of variations*, Die Grund. Math. Wiss., Band 130, Springer New York, Inc., New York, 1966.

[76] J. R. Munkres – *Topology: a first course*, Prentice-Hall Inc., 1975.

[77] G. Neugebauer & R. Meinel – "Progress in relativistic gravitational theory using the inverse scattering method", *J. Math. Phys.* **44** (2003), p. 3407–3429.

[78] K. Nomizu – "On local and global existence of Killing vector fields", *Ann. of Math.* **72** (1960), p. 105–120.

[79] P. Orlik – *Seifert manifolds*, Lecture Notes in Math., Vol. 291, Springer, 1972.

[80] A. Papapetrou – "Champs gravitationnels stationnaires à symmétrie axiale", *Ann. Inst. H. Poincaré* **4** (1966), p. 83–105.

[81] N. Pelavas, N. Neary & K. Lake – "Properties of the instantaneous ergo surface of a Kerr black hole", *Classical Quantum Gravity* **18** (2001), p. 1319–1331.

[82] I. Rácz & R. M. Wald – "Global extensions of spacetimes describing asymptotic final states of black holes", *Classical Quantum Gravity* **13** (1996), p. 539–552.

[83] F. Raymond – "Classification of the actions of the circle on 3-manifolds", *Trans. Amer. Math. Soc.* **131** (1968), p. 51–78.

[84] D. Robinson – "Uniqueness of the Kerr black hole", *Phys. Rev. Lett.* **34** (1975), p. 905–906.

[85] ______, "A simple proof of the generalization of Israel's theorem", *Gen. Rel. Grav.* **8** (1977), p. 695–698.

[86] R. M. Schoen – "Variational theory for the total scalar curvature functional for Riemannian metrics and related topics", in *Topics in calculus of variations (Montecatini Terme, 1987)*, Lecture Notes in Math., vol. 1365, Springer, 1989, p. 120–154.

[87] W. Simon & R. Beig – "The multipole structure of stationary space-times", *J. Math. Phys.* **24** (1983), p. 1163–1171.

[88] H. Stephani, D. Kramer, M. MacCallum, C. Hoenselaers & E. Herlt – *Exact solutions of Einstein's field equations*, second ed., Cambridge Monographs on Mathematical Physics, Cambridge University Press, 2003.

[89] D. Sudarsky & R. M. Wald – "Extrema of mass, stationarity, and staticity, and solutions to the Einstein-Yang-Mills equations", *Phys. Rev. D* **46** (1992), p. 1453–1474.

[90] R. M. Wald – *General relativity*, University of Chicago Press, 1984.

[91] G. Weinstein – "On rotating black holes in equilibrium in general relativity", *Comm. Pure Appl. Math.* **43** (1990), p. 903–948.

[92] ______, "On the force between rotating co-axial black holes", *Trans. Amer. Math. Soc.* **343** (1994), p. 899–906.

[93] ______, "Harmonic maps with prescribed singularities into Hadamard manifolds", *Math. Res. Lett.* **3** (1996), p. 835–844.

[94] ———, "Harmonic maps with prescribed singularities on unbounded domains", *Amer. J. Math.* **118** (1996), p. 689–700.

[95] ———, "N-black hole stationary and axially symmetric solutions of the Einstein/Maxwell equations", *Comm. Partial Differential Equations* **21** (1996), p. 1389–1430.

Index

P. T. Chruściel, LMPT, Fédération Denis Poisson, Tours; Mathematical Institute and Hertford College, Oxford • *E-mail :* chrusciel@maths.ox.ac.uk • *Url :* www.phys.univ-tours.fr/~piotr

J.L. Costa, Lisbon University Institute (ISCTE); Mathematical Institute and Magdalen College, Oxford • *E-mail :* jlca@iscte.pt

Astérisque

Hideki Omori
Yoshiaki Maeda
Naoya Miyazaki
Akira Yoshioka

A new nonformal noncommutative calculus: associativity and finite part regularization

Astérisque, tome 321 (2008), p. 267-297
<http://www.numdam.org/item?id=AST_2008__321__267_0>

© Société mathématique de France, 2008, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (http://smf4.emath.fr/Publications/Asterisque/) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

Numdam
Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

Astérisque
321, 2008, p. 267–297

A NEW NONFORMAL NONCOMMUTATIVE CALCULUS: ASSOCIATIVITY AND FINITE PART REGULARIZATION

by

Hideki Omori, Yoshiaki Maeda, Naoya Miyazaki & Akira Yoshioka

Abstract. — We interpret the element $\frac{1}{2i\hbar}(u*v+v*u)$ in the generators u, v of the Weyl algebra W_2 as an indeterminate in $\mathbb{N}+\frac{1}{2}$ or $-(\mathbb{N}+\frac{1}{2})$, using methods of the transcendental calculus outlined in the announcement [**13**]. The main purpose of this paper is to give a rigorous proof for the part of [**13**] which introduces this indeterminate phenomenon. Namely, we discuss how to obtain associativity in the transcendental calculus and show how the Hadamard finite part procedure can be implemented in our context.

Résumé **(Un nouveau calcul non-formel et non-commutatif : associativité et régularisation des parties finies)**

Nous interprêtons l'élément $\frac{1}{2i\hbar}(u*v+v*u)$ dans les générateurs u, v de l'algèbre de Weyl W_2 en tant qu'indéterminés dans $\mathbb{N}+\frac{1}{2}$ ou $-(\mathbb{N}+\frac{1}{2})$, en utilisant des méthodes du calcul transcendental décrit dans l'annonce [**13**]. Le but principal de cet article est de donner une preuve rigoureuse de la partie de [**13**] qui introduit ce phénomène indéterminé. À savoir, nous discutons la manière d'obtenir l'associativité dans le calcul transcendental et de montrer comment la procédure de parties finies de Hadamard peut être implémentée dans notre contexte.

1. Introduction

Deformation quantization, first proposed in [**3**], is a fruitful approach to developing quantum theory in a purely algebraic framework, and was a prototype for noncommutative calculus on noncommutative spaces. It was first treated as a formal noncommutative calculus, with the Planck constant $\hbar$ regarded as a formal parameter, but has been extended to nonformal cases, as in the studies of noncommutative tori

2000 Mathematics Subject Classification. — 53D55, 53D10; 46L65.
Key words and phrases. — Weyl algebra, transcendental calculus.

Partially supported by Grant-in-Aid for Scientific Research #18204006 (Y. Maeda), #18540093 (N. Miyazaki), #19540103 (A. Yoshioka) by the Ministry of Education, Science and Culture, Japan.

© Astérisque 321, SMF 2008

[18] and quantum groups [20]. In fact, the formal and nonformal noncommutative calculus have quite different features.

In [12], we analyzed star exponential functions of quadratic forms in the Weyl algebra and uncovered several mysterious phenomena unanticipated from the formal case. These mysterious phenomena reflect the fact that star exponential functions of quadratic forms (see [11] and [15]) lie outside of the Weyl algebra. These new features suggest a new approach to noncommutative nonformal calculus. In this paper, we show that this new calculus is necessary to treat transcendental elements of the Weyl algebra.

From the papers [12]-[13], we know that the Moyal product, the most typical product on the Weyl algebra, is not sufficient to treat transcendental elements such as star exponential functions. For this reason, we introduced a family of $*_K$-products on the Weyl algebra depending on a complex symmetric matrix K and developed a transcendental nonformal noncommutative calculus specifically formulated to treat star exponential functions of quadratic forms. The transcendental elements of the Weyl algebra have a realization depending on the $*_K$-product, which we called the *K-ordered expression*. Thus, properties of (transcendental) elements of the Weyl algebra depend on the choice of product $*_K$,

We now propose as a principle, called the *Independence of Ordering Principle* (IOP), that the relevant properties of transcendental elements of the Weyl algebra do not depend on the choice of ordered expression, just as properties and objects in differential geometry do not depend on the choice of coordinate expression. Following this principle, in [12] we proposed the notion of a group-like object of star exponential functions of quadratic forms on the Weyl algebra. The IOP seems to be a new outlook on interpreting physical phenomena/mathematical phenomena, especially for treating quantum objects and phenomena from an algebraic point of view.

As a test case, we examine this principle on the nonformal noncommutative calculus for transcendental elements of the Weyl algebra. As part of this approach, we interpret an element as an indeterminate in a discrete set in the case of the Weyl algebra with two generators.

Let W_2 be the Weyl algebra with generators u, v obeying the commutation relation

$$[u, v] = -i\hbar. \tag{1}$$

We consider the element $\frac{1}{i\hbar}u \circ v = \frac{1}{2i\hbar}(u * v + v * u)$ of W_2. We show that $\frac{1}{i\hbar}u \circ v$ can be interpreted as an indeterminate in $\mathbb{N} + \frac{1}{2}$ or $-(\mathbb{N} + \frac{1}{2})$, not from a more standard operator theoretic point of view but from a purely algebraic approach, using the IOP that a physical/mathematical object should be independent of its various ordered expressions.

In our approach, we interpret $\frac{1}{i\hbar}u\circ v$ in two ways: 1) via the analytic continuation of inverses of $z+\frac{1}{i\hbar}u\circ v$ and 2) via the $*$-product of the $*$-sin function and the $*$-gamma function using ordered expressions. These results have been already announced in [**13**] with outlines of proofs. The main purpose of this paper is to give a rigorous description of method 1) and therefore to realize $\frac{1}{i\hbar}u\circ v$ as an indeterminate in the discrete set. The main ingredients of the proof are dealing with associativity in the framework of the transcendental calculus of [**13**] and applying the Hadamard finite part procedure in this context. For a family of $*_K$-products on the Weyl algebra W_2, we provide rules for the associativity of the extended products $*_K$, and in preparation for the definition of the inverse of $z+\frac{1}{i\hbar}u\circ v$, we investigate star exponentials $e_*^{z+\frac{1}{i\hbar}u\circ v}$ and their ordered versions.

We leave the the finite part regularization method for Fréchet algebra valued functions in the subsection 6.1. For a holomorphic function $f(z)$ with a pole at $z=z_0$, we define the finite part of $f(z)$ as

$$\mathrm{FP}(f(z))=\begin{cases} f(z) & z\neq z_0\\ \mathrm{Res}_{w=0}\frac{1}{w}(f(z_0+w)) & z=z_0.\end{cases}$$

We first construct the inverses of $z+\frac{1}{i\hbar}u\circ v$ by using the star exponential function $e_*^{z+\frac{1}{i\hbar}u\circ v}$ and a K-ordered expression. We can construct two inverses of $z+\frac{1}{i\hbar}u\circ v$ as follows:

$$(z+\frac{1}{i\hbar}u\circ v)_{*+}^{-1}=\int_{-\infty}^{0}e_*^{t(z+\frac{1}{i\hbar}u\circ v)}dt$$

and

$$(z+\frac{1}{i\hbar}u\circ v)_{*-}^{-1}=-\int_{0}^{\infty}e_*^{t(z+\frac{1}{i\hbar}u\circ v)}dt$$

(see [**7**] and [**10**] for more details). Both inverses have analytic continuations for generic ordered expression. In §6, we mainly study the inverse $(z+\frac{1}{i\hbar}u\circ v)_{*+}^{-1}$, as the other inverse has similar properties.

In §6, we show the following:

Theorem 1.1. — *For generic ordered expressions, the inverses* $(z+\frac{1}{i\hbar}u\circ v)_{*+}^{-1}$, $(z-\frac{1}{i\hbar}u\circ v)_{*-}^{-1}$ *extend to* $\mathcal{E}_{2+}(\mathbb{C}^2)$*-valued holomorphic functions of* z *on* $\mathbb{C}-\{-(\mathbb{N}+\frac{1}{2})\}$.

Here, we refer the class $\mathcal{E}_{2+}(\mathbb{C}^2)$ in the subsection 2.2.

Employing the Hadamard technique of extracting finite parts of divergent integrals, we now extend the definition of the $*$-product using finite part regularization. We define the new product of $(z+\frac{1}{i\hbar}u\circ v)_{*\pm}^{-1}$ with either the polynomial $q(u,v)$ or $q(u,v)=e_*^{s\frac{1}{i\hbar}u\circ v}$ by

$$(z+\tfrac{1}{i\hbar}u\circ v)_{*\pm}^{-1}\tilde{*}q(u,v)=(\mathrm{FP}(z+\tfrac{1}{i\hbar}u\circ v)_{*\pm}^{-1})*q(u,v). \tag{2}$$

Note that the result may not be continuous in z.

The following is an description of the discrete phenomena for $\frac{1}{i\hbar}u \circ v$ via method 1):

Theorem 1.2. — *Using definition* (2) *for the* $\tilde{*}$*-product, we have*

$$(3)\qquad (z+\tfrac{1}{i\hbar}u\circ v)\tilde{*}(z+\tfrac{1}{i\hbar}u\circ v)^{-1}_{*+} = \begin{cases} 1 & z\notin -(\mathbb{N}+\frac{1}{2}) \\ 1-\frac{1}{n!}(\frac{1}{i\hbar}u)^n * \varpi_{00} * v^n & z=-(n+\frac{1}{2}) \end{cases},$$

$$(4)\qquad (z-\tfrac{1}{i\hbar}u\circ v)\tilde{*}(z-\tfrac{1}{i\hbar}u\circ v)^{-1}_{*-} = \begin{cases} 1 & z\notin -(\mathbb{N}+\frac{1}{2}) \\ 1-\frac{1}{n!}(\frac{1}{i\hbar}v)^n * \overline{\varpi}_{00} * u^n & z=-(n+\frac{1}{2}) \end{cases}.$$

for generic ordered expressions.

We will interpret this discrete phenomena for $\frac{1}{i\hbar}u\circ v$ via method 2) in a forthcoming paper.

We would like to thank Steven Rosenberg and Sylvie Paycha for their suggestions about regularization methods.

Finally, we are honored to contribute our paper to the Proceedings for the 60th birthday celebration of Jean Pierre Bourguignon, whose friendship with us for over 20 years we warmly acknowledge.

2. General ordered expressions and IOP

We introduce a method to realize the Weyl algebra via a family of ordered expressions. This leads to a transcendental calculus for the Weyl algebra.

2.1. Fundamental product formulas and intertwiners.

— Let $\mathfrak{S}(n)$ and $\mathfrak{A}(n)$ be the spaces of complex symmetric matrices and skew-symmetric matrices respectively, and set $\mathfrak{M}(n) = \mathfrak{S}(n) \oplus \mathfrak{A}(n)$. We denote by $\boldsymbol{u}$ the set of generators $\boldsymbol{u} = (u_1, \dots, u_{2m})$. For an arbitrary fixed $n\times n$-complex matrix $\Lambda\in\mathfrak{M}(n)$, we define a product $*_{\Lambda}$ on the space of polynomials $\mathbb{C}[\boldsymbol{u}]$ by the formula

$$(5)\qquad f *_{\Lambda} g = f e^{\frac{i\hbar}{2}(\sum \overleftarrow{\partial_{u_i}}\Lambda^{ij}\overrightarrow{\partial_{u_j}})} g = \sum_k \frac{(i\hbar)^k}{k!2^k}\Lambda^{i_1j_1}\cdots\Lambda^{i_kj_k}\partial_{u_{i_1}}\cdots\partial_{u_{i_k}} f\; \partial_{u_{j_1}}\cdots\partial_{u_{j_k}} g.$$

It is known and not hard to prove that $(\mathbb{C}[\boldsymbol{u}], *_{\Lambda})$ is an associative algebra.

The algebraic structure of $(\mathbb{C}[\boldsymbol{u}], *_{\Lambda})$ is determined by the skew-symmetric part of Λ, if the generators are fixed. In particular, if Λ is a symmetric matrix, $(\mathbb{C}[\boldsymbol{u}], *_{\Lambda})$ is isomorphic to the usual polynomial algebra.

For every symmetric matrix $K \in \mathfrak{S}(n)$, the operator

$$I_0^K(f) = \exp\Big(\frac{i\hbar}{4}\sum_{i,j} K^{ij}\partial_{u_i}\partial_{u_j}\Big)f \tag{6}$$

gives an isomorphism $I_0^K : (\mathbb{C}[\boldsymbol{u}], *_\Lambda) \to (\mathbb{C}[\boldsymbol{u}], *_{\Lambda+K})$. Namely, for any $f, g \in \mathbb{C}[\boldsymbol{u}]$:

$$I_0^K(f *_\Lambda g) = I_0^K(f) *_{\Lambda+K} I_0^K(g). \tag{7}$$

Let $\Lambda = K + J$ be the symmetric/skew symmetric parts of Λ, $K \in \mathfrak{S}(n)$, $J \in \mathfrak{A}(n)$. Changing K while leaving J fixed will be called a *deformation* of the expression of elements, as the algebra remains in the same isomorphism class.

We view these expressions of algebra elements as analogous to the "local coordinate expression" of functions on a manifold. Changing K corresponds to a local coordinate transformation on a manifold. In this context, we call the product formula (5) the *K-ordered expression*, i.e. ignoring the fixed skew part J, and $*_K$ stands sometimes for $*_\Lambda$ with J understood.

The big difference from local coordinate expressions for functions on a manifold is precisely that in our context there is no "underlying topological space".

In the following we set $n = 2m$ and $J = \begin{bmatrix} 0 & -I_m \\ I_m & 0 \end{bmatrix}$. $(\mathbb{C}[\boldsymbol{u}], *_\Lambda)$ is called the Weyl algebra, with isomorphism class denoted by W_{2m}.

According to the choice of $K = 0, \begin{bmatrix} 0 & I_m \\ I_m & 0 \end{bmatrix}, \begin{bmatrix} 0 & -I_m \\ -I_m & 0 \end{bmatrix}$, the K-ordered expression is called the Weyl ordered, the normal ordered and the anti-normal ordered expressions, respectively. The intertwiner between a K-ordered expression and a K'-ordered expression is given by

$$I_K^{K'}(f) = \exp\Big(\frac{i\hbar}{4}\sum_{i,j}(K'^{ij} - K^{ij})\partial_{u_i}\partial_{u_j}\Big)f \ (= I_0^{K'}(I_0^K)^{-1}(f)), \tag{8}$$

giving an isomorphism $I_K^{K'} : (\mathbb{C}[\boldsymbol{u}]; *_{K+J}) \to (\mathbb{C}[\boldsymbol{u}]; *_{K'+J})$ between algebras.

2.2. Extension of products and intertwiners. —

In what follows we write $*_K$ for $*_{K+J}$ for simplicity. Let $\mathbb{C}[\boldsymbol{u}][[\hbar]]$ be the space of all formal power series in $\hbar$ with polynomials in $\boldsymbol{u}$ as coefficients. Obviously, the $*_K$-product and the intertwiners extend naturally to $\mathbb{C}[\boldsymbol{u}][[\hbar]]$ by the same formulas. $(\mathbb{C}[\boldsymbol{u}][[\hbar]], *_K)$ is an associative algebra and $I_K^{K'}$ is an algebra isomorphism from $(\mathbb{C}[\boldsymbol{u}][[\hbar]], *_K)$ to $(\mathbb{C}[\boldsymbol{u}][[\hbar]], *_{K'})$.

Let $Hol(\mathbb{C}^n)$ be the space of all holomorphic functions on $\mathbb{C}^n$ with the topology of uniform convergence topology on compact domains. The following fundamental lemma follows easily from the product formula (5) together with Taylor's formula:

Lemma 2.1. — *Let $p(\boldsymbol{u})$ be either a polynomial of $\boldsymbol{u}$ or an exponential function of a linear combination of generators $p(\boldsymbol{u}) = e^{\sum a_i u_i}$. Then the left multiplication $p(\boldsymbol{u})*_{_K}$ (resp. the right multiplication $*_{_K}p(\boldsymbol{u})$) is a continuous linear mapping from $Hol(\mathbb{C}^n)$ to itself. Associativity $(f*_{_K}g)*_{_K}h = f*_{_K}(g*_{_K}h)$ holds if two of f, g, h are polynomials.*

For every positive real number p, we set

$$\mathcal{E}_p(\mathbb{C}^n) = \{f \in \mathrm{Hol}(\mathbb{C}^n) \mid \|f\|_{p,s} = \sup |f|\, e^{-s|\xi|^p} < \infty,\ \forall s > 0\} \tag{9}$$

where $|\xi| = (\sum_i |u_i|^2)^{1/2}$. The family of seminorms $\{\|\cdot\|_{p,s}\}_{s>0}$ induces a topology on $\mathcal{E}_p(\mathbb{C}^n)$ and $(\mathcal{E}_p(\mathbb{C}^n), \cdot)$ is an associative commutative Fréchet algebra, where the dot $\cdot$ is the ordinary product for functions in $\mathcal{E}_p(\mathbb{C}^n)$.

Let H be a polynomial of order p. Then, $e^H \in \mathcal{E}_{p'}(\mathbb{C}^n)$ for every $p' > p$, but $e^H \notin \mathcal{E}_p(\mathbb{C}^n)$. Note also that $exp\sqrt[q]{H} \in \mathcal{E}_{p'/q}$ for every $p' > p$ on a suitable Riemann surface.

It is easily seen that for $0 < p < p'$, there is a continuous embedding

$$\mathcal{E}_p(\mathbb{C}^n) \subset \mathcal{E}_{p'}(\mathbb{C}^n) \tag{10}$$

as commutative Fréchet algebras (cf. [**4**],[**15**]), and that $\mathcal{E}_p(\mathbb{C}^n)$ is $\mathrm{Sp}(m, \mathbb{C})$-invariant.

It is obvious that every polynomial is contained in $\mathcal{E}_p(\mathbb{C}^n)$ and that $\mathbb{C}[\boldsymbol{u}]$ is dense in $\mathcal{E}_p(\mathbb{C}^n)$ for any $p > 0$ in the Fréchet topology defined by the family of seminorms $\{\|\ \|_{p,s}\}_{s>0}$.

Theorem 2.1. — *Assume $0 < p \leq 2$. The product formula* (5) *extends in the following way:*
(a) The space $(\mathcal{E}_p(\mathbb{C}^n), *_{_K})$ forms a complete noncommutative topological associative algebra over $\mathbb{C}$ (cf. [**11**]).
(b) The intertwiner $I_{_K}^{^{K'}}$ extends to an isomorphism of $(\mathcal{E}_p(\mathbb{C}^n), *_{_K})$ onto itself (cf.[**12**]).

See also [**15**] for the general case with precise proofs and several comments.

It is easily seen that the following identities hold on $\mathcal{E}_p(\mathbb{C}^n)$, $p \leq 2$:

$$I_{_{K'}}^{^K} I_{_K}^{^{K'}} = 1, \quad I_{_{K'}}^{^{K''}} I_{_K}^{^{K'}} = I_{_K}^{^{K''}}. \tag{11}$$

For every $f \in \mathcal{E}_p(\mathbb{C}^n)$ such that $p \leq 2$, $f(K) = I_0^{^K}(f)$ is globally defined on $\mathfrak{S}(n)$.

Thus, we naturally extend our object f to the space of all mutually intertwined sections $\{f(K); K \in \mathfrak{S}(n)\}$ of the trivial bundle $\coprod_{K\in\mathfrak{S}(n)}(\mathcal{E}_p(\mathbb{C}^n), *_{_K})$, $0<p\leq 2$. However, several anomalous phenomena occur in the space $(\mathcal{E}_{2+}(\mathbb{C}^n), *_{_K}) = \bigcap_{p>2}(\mathcal{E}_p(\mathbb{C}^n), *_{_K})$. See [**7**]-[**10**], [**12**]-[**14**].

Theorem 2.2. — *For every pair (p, p') such that $\frac{1}{p} + \frac{1}{p'} \geq 1$ the product* (5) *extends to a continuous bilinear mapping $\mathcal{E}_p(\mathbb{C}^n) \times \mathcal{E}_{p'}(\mathbb{C}^n) \to \mathcal{E}_{p\vee p'}(\mathbb{C}^n)$.*

By Theorems 2.1 and 2.2, associativity $f*(g*h)=(f*g)*h$ holds for $f,g,h \in \mathcal{E}_2(\mathbb{C}^n)$. Moreover if one of f,g,h is in $\mathcal{E}_p(\mathbb{C}^n)$, $p>2$, then by using the polynomial approximation theorem, we have that associativity holds if the two others are in $\mathcal{E}_{p'}(\mathbb{C}^n)$ such that $\frac{1}{p}+\frac{1}{p'}\geq 1$.

Since $\mathcal{E}_p(\mathbb{C}^n)$ is a Fréchet space, we have:

Lemma 2.2. — *Let M be a compact domain in $\mathbb{R}^m$, and let $x \mapsto f_x \in \mathcal{E}_p(\mathbb{C}^n)$ be a continuous mapping of M into $\mathcal{E}_p(\mathbb{C}^n)$. Then the integral $\int_M f_x dV_x$ of f_x on M is an element of $\mathcal{E}_p(\mathbb{C}^n)$.*

3. Star exponential functions

In differential geometry, it is widely accepted that geometrical notions should have coordinate free expressions. Obviously, the algebraic structure of $(\mathbb{C}[\boldsymbol{u}], *_\Lambda)$ depends only on the skew part of Λ. This analogy with geometry makes it plausible to introduce the Independence of Ordering Principle (IOP), namely that the algebraic interpretation of physical phenomena should be independent of the choice of ordered expression (cf. [**1**]).

In fact, this principle for the class $\mathcal{E}_2(\mathbb{C}^n)$ is reflected in Theorem 2.1. However, as will be seen below, we have to think carefully about the true meaning of IOP, since there are many delicate anomalous phenomena in the transcendental calculus of star-exponential functions. In spite of these difficulties, we believe that properties which appear in generic (i.e. almost all/open dense) ordered expressions are fundamental features of this theory. In the end, IOP provides deeper insight into the extended Weyl algebra.

For an element H_* of the Weyl algebra, we define the $*$-exponential function $e_*^{tH_*}$ as the family $\{f_t(K)\}$ of real analytic solutions of the evolution equation

$$\frac{d}{dt}f_t(K) = {:}H_*{:}_K *_K f_t(K), \tag{12}$$

with the initial condition $f_0(K)=1$. We think of $f_t(K)$ as the K-ordered expression of $e_*^{tH_*}$, and denote it by ${:}e_*^{tH_*}{:}_K = f_t(K)$.

Provided ${:}e_*^{sH_*}{:}_K$ exists for every $s \in \mathbb{C}$, they form a complex one parameter subgroup, for the exponential law holds by the uniqueness of real analytic solutions. If ${:}e_*^{sH_*}{:}_K$ exists for every $s \in \mathbb{R}$, it is a real one parameter subgroup.

If we have the real analytic solution of (12) with initial condition $f_0(K)=g$, then it is natural to denote the solution by ${:}e_*^{tH_*}{:}_K *_K g$. This definition works for $g \in \mathcal{E}_p(\mathbb{C}^{2m})$, $p>2$.

Warning In general, (12) is a misleading definition, as we can expect neither the existence of a solution of (12), nor any continuity in the initial data. For $H_* = \frac{1}{i\hbar}u \circ v$, there are branching singular points in $e_*^{tH_*}$. If H_* is an exponential function such as $e^{au}v$, then (12) is not a partial differential equation, but rather a difference-differential equation (cf.[**7**]).

If H_* is a quadratic form, $:e_*^{sH_*}:_{_K}$ is defined with a certain discrete set of singularities, as we shall see in §3.1. In general, there is no reflection symmetry for the domain of existence of the solution of (12).

3.1. General properties of $*$-exponential functions. — For a given K, suppose that (12) has real analytic solutions in t on some domain $D(K)$ including 0 for the initial functions 1 and g. We denote the solution of (12) with initial function g by

$$:e_*^{tH_*}:_{_K} *_{_K} g, \quad t \in D(K). \tag{13}$$

Proposition 3.1. — *If H_* is a polynomial and $:e_*^{tH_*}:_{_K}$ is defined on a domain $D(K)$, then $:e_*^{tH_*}:_{_K} *_{_K} p(\boldsymbol{u})$ is defined for every polynomial $p(\boldsymbol{u})$ on the same domain $D(K)$.*

If $p(\boldsymbol{u}) = \sum A_\alpha(s)\boldsymbol{u}^\alpha$ is a polynomial whose coefficients depend smoothly on s, then the formula

$$\partial_s^\ell :e_*^{tH_*}:_{_K} *_{_K} p(\boldsymbol{u}) = :e_*^{tH_*}:_{_K} *_{_K} \partial_s^\ell p(\boldsymbol{u})$$

holds for every ℓ.

Proof Multiplying the defining equation (12) by $*p(\boldsymbol{u})$ and applying the associativity in Lemma 2.1, we have

$$\frac{d}{dt} f_t(K) * p(\boldsymbol{u}) = :H_*:_{_K} *_{_K} (f_t(K) * p(\boldsymbol{u})), \quad f_0(K) = 1. \tag{14}$$

Since $f_t(K) * p(\boldsymbol{u})$ is a real analytic solution, this is written in our notation as $e_*^{tH_*} * p(\boldsymbol{u})$. Applying ∂_s^ℓ to (14) gives the second assertion by a similar argument. □

Let P_n be the space of polynomials of degree at most n. Then there are natural inclusions $P_n \subset P_{n+1}$. We view $\mathbb{C}[\boldsymbol{u}]$ as the inductive limit $lim_{\rightarrow} P_n$ with the inductive limit topology. The second assertion of Proposition 3.1 then yields continuity with respect to the initial condition in the inductive limit topology. We use this topology in calculations with inverse elements. However, we should remark that $\mathbb{C}[\boldsymbol{u}]$ is *not* a Fréchet space in this topology, as the first axiom of countability fails.

Remark Although $:e_*^{tH_*}:_{_K} *_{_K} 0 = 0$, since (12) is linear, it does not necessary follow that
$\lim_k :e_*^{tH_*}:_{_K} *_{_K} p_k(\boldsymbol{u}) = 0$, when $\lim_k p_k(\boldsymbol{u}) = 0$ in the uniform convergence topology. Suppose $e_*^{tH_*}$ is singular at $t = t_0$. Since $:e_*^{tH_*}:_{_K} *_{_K} 0 = 0$ on an open dense domain, the zero function is the real analytic solution of (12), but for a series $c_n \in \mathbb{C}$ such that $\lim_{n\to\infty} c_n = 0$, $\lim_{n\to\infty} :e_*^{tH_*}:_{_K} *_{_K} c_n$ does not converge to 0 in this topology.

In the following, we often omit the subscript K, and so denote $*_{_K}$, $:g_*:_{_K}$ simply by $*$, g_* when the context is clear.

Suppose H_* is a polynomial and $G(t;K) = :e_*^{tH_*}:_K *_K :g_*:_K$ is defined. Then for every polynomial $p(\boldsymbol{u})$, $G(t;K)$ satisfies

$$\frac{d}{dt}G(t,K)*_K p(\boldsymbol{u}) = (:H_*:_K *_K G(t,K))*_K p(\boldsymbol{u}) = :H_*:_K *_K (G(t,K)*_K p(\boldsymbol{u})),$$
$$G(0,K)*_K p(\boldsymbol{u}) = :g_*:_K *_K p(\boldsymbol{u}).$$

Since $G(t,K)*_K p(\boldsymbol{u})$ is real analytic in t, we have the following associativity:

Proposition 3.2. — *If $e_*^{tH_*} * g_*$ is defined for some K, then $e_*^{tH_*} * (g_* * p(\boldsymbol{u}))$ is defined for K and*

$$e_*^{tH_*} * (g_* * p(\boldsymbol{u})) = (e_*^{tH_*} * g_*) * p(\boldsymbol{u}) \quad \textit{for every } p(\boldsymbol{u}) \in \mathbb{C}[\boldsymbol{u}].$$

Let H_* be a polynomial. Since $:e_*^{tH_*} * H_*:_K$ and $:H_* * e_*^{tH_*}:_K$ satisfy the same differential equation with the same initial data, the uniqueness of real analytic solutions gives $:e_*^{tH_*} * H_*:_K = :H_* * e_*^{tH_*}:_K$.

Using this, we also have

Proposition 3.3. — *If H_* is a polynomial such that $e_*^{tH_*}$ is defined, then $:e_*^{tH_*}:_K$ is the real analytic solution $h_t(K)$ of the equation*

$$\frac{d}{dt}h_t(K) = h_t(K)*_K :H_*:_K \tag{15}$$

with the initial condition $h_0(K) = 1$.

From this fact, we see that $p(\boldsymbol{u}) * e_*^{tH_*}$ is the solution of (15) with the initial condition $h_0 = p(\boldsymbol{u})$. Hence the exponential law and the uniqueness of solutions give

$$e_*^{sH_*} * (e_*^{tH_*} * p(\boldsymbol{u})) = e_*^{(s+t)H_*} * p(\boldsymbol{u}), \quad (p(\boldsymbol{u}) * e_*^{sH_*}) * e_*^{tH_*} = p(\boldsymbol{u}) * e_*^{(s+t)H_*}. \tag{16}$$

Let $\mathrm{ad}(H_*)(h) = [H_*, h] = H_* * h - h * H$. If H_* is a quadratic form, then $\mathrm{ad}(H_*)$ defines a linear transformation on the linear hull of the generators. By exponentiation, $\exp s\mathrm{ad}(H_*)$ is a degree preserving linear transformation on the space $\mathbb{C}[\boldsymbol{u}]$ of polynomials such that

$$[(\exp s\mathrm{ad}(H_*))f, (\exp s\mathrm{ad}(H_*))g] = (\exp s\mathrm{ad}(H_*))[f,g].$$

Note also that $(\exp t\mathrm{ad}(H_*))(p(\boldsymbol{u}))$ is the solution f_t of

$$\frac{d}{dt}f_t = [H_*, f_t], \quad f_0 = p(\boldsymbol{u}).$$

Since $p_s(\boldsymbol{u}) = (\exp s\mathrm{ad}(H_*))(p(\boldsymbol{u}))$ is a polynomial, we see by Proposition 3.1

$$\frac{d}{ds}e_*^{-sH_*} * p_s(\boldsymbol{u}) = e_*^{-sH_*}(-H_* * p_s(\boldsymbol{u}) + [H_*, p_s(\boldsymbol{u})]) = e_*^{-sH_*} * p_s(\boldsymbol{u}) * (-H_*).$$

Since $e_*^{sH_*} * p_s(\boldsymbol{u})|_{s=0} = p(\boldsymbol{u})$, we have

$$e^{-sH_*} * p_s(\boldsymbol{u}) = p(\boldsymbol{u}) * e_*^{-sH_*}. \tag{17}$$

Combining (17) with (16), we get the following associativity:

$$e_*^{sH_*} * (p(\boldsymbol{u}) * e_*^{tH_*}) = (e_*^{sH_*} * p(\boldsymbol{u})) * e_*^{tH_*}. \tag{18}$$

It also follows that

$$e_*^{sH_*} * (p(\boldsymbol{u}) * e_*^{-sH_*}) = (e_*^{sH_*} * p(\boldsymbol{u})) * e_*^{-sH_*} = (\exp s\mathrm{ad}(H_*))(p(\boldsymbol{u})). \tag{19}$$

3.2. Star-exponentials of quadratic forms in the normal ordered expression. — In this section, we set $n = 2m$ and $u = (u_1, \cdots, u_m)$, $v = (v_1, \cdots, v_m) = (u_{m+1}, \cdots, u_{2m})$. For every $C = (C_{ij}) \in \mathfrak{M}(m)$, we consider $C(u, v) = \sum C_{ij} u_i v_j$. The star exponential function of this special quadratic form is easily obtained in the normal ordered expression, since no anomalous phenomena occur. By setting $\Lambda = K_0 + J$, $K_0 = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}$ in the product formula (5), a direct calculation gives

$$g e^{\frac{2}{i\hbar} \sum A_{kl} u_k v_l} *_{K_0} g' e^{\frac{2}{i\hbar} \sum B_{st} u_s v_t} = g g' e^{\frac{2}{i\hbar} \sum C_{ij} u_i v_j}, \tag{20}$$

where $C = A + B + 2AB$. For $(g; A) = g e^{\frac{2}{i\hbar} \sum A_{kl} u_k v_l}$, this product formula becomes

$$(g; A) *_{K_0} (g'; B) = (gg'; A + B + 2AB), \quad (g; A), (g'; B) \in \mathbb{C} \times \mathfrak{M}(m). \tag{21}$$

Note that

$$(I + 2A)(I + 2B) = I + 2(A + B + 2AB).$$

Under the correspondence $A \leftrightarrow I + 2A$, the structure of the usual matrix algebra $\mathfrak{M}(m)$ is carried over to the space $\{e^{\frac{2}{i\hbar} C(u,v)}; C \in \mathfrak{M}(m)\}$. However, note here that 0 corresponds to I. In (21) we see that $(-I) + (-I) + 2(-I)(-I) = 0$, and $-\frac{1}{2} I + C + 2(-\frac{1}{2} I) C = -\frac{1}{2} I$ for every C.

Although these elements are in $\mathcal{E}_{2+}(\mathbb{C}^n)$, associativity still holds for the products

$$(g; A) *_{K_0} ((g'; B) *_{K_0} (g''; C)) = ((g; A) *_{K_0} (g'; B)) *_{K_0} (g''; C), \tag{22}$$

and

$$e^{\frac{2}{i\hbar} C(u,v)} *_{K_0} e^{\frac{i}{\hbar} I(u,v)} = e^{\frac{i}{\hbar} I(u,v)} *_{K_0} e^{\frac{2}{i\hbar} C(u,v)} = e^{\frac{i}{\hbar} I(u,v)}.$$

By (21), we see that

$$e^{\frac{1}{i\hbar}(e^{isC} - I)(u,v)} *_{K_0} e^{\frac{1}{i\hbar}(e^{itC} - I)(u,v)} = e^{\frac{1}{i\hbar}(e^{i(s+t)C} - I)(u,v)}. \tag{23}$$

Differentiating the exponential law (23) to obtain the K_0-expression (the normal ordered expression) of the $*$-exponential function, we have

$$:e_*^{\frac{it}{i\hbar} \sum C_{kl} u_k * v_l}:_{K_0} = e^{\frac{1}{i\hbar} \sum (e^{itC} - I)_{kl} u_k v_l}. \tag{24}$$

This is holomorphic in $t \in \mathbb{C}$ and the r.h.s of (24) is contained in $\mathcal{E}_{2+}(\mathbb{C}^2)$.

Set

$$a \circ b = \frac{1}{2}(a * b + b * a).$$

By the exponential law for scalar exponential functions, (24) becomes

$$:e_*^{\frac{it}{i\hbar}\sum C_{kl}u_k\circ v_l}:_{K_0} = e^{\frac{it}{2}\mathrm{Tr}(C)}e^{\frac{1}{i\hbar}\sum(e^{itC}-I)_{kl}u_kv_l}. \tag{25}$$

This is also a holomorphic one parameter group contained in $(\mathcal{E}_{2+}(\mathbb{C}^2); *_{K_0})$. However, this property of $e_*^{\frac{it}{i\hbar}\sum C_{kl}u_k\circ v_l}$ is not generic, as we see in §3.3. Indeed, a generic element has branching singular points periodically distributed in $\mathbb{C}$. On the other hand, for the special case $C = I$, we see that $:e_*^{\frac{2\pi i}{i\hbar}\sum u_k\circ v_k}:_{K_0} = (-1)^m$. Intertwiners map scalars to scalars, but may change the sign of the scalar $(-1)^m$ in this equation. The property that $:e_*^{\frac{2\pi i}{i\hbar}\sum u_k\circ v_k}:_K = \pm 1$ is generic.

We note that $\lim_{t\to-\infty} :e_*^{t\frac{1}{i\hbar}\sum u_k\circ v_k}:_{K_0} = 0$ but $\lim_{t\to\infty} :e_*^{t\frac{1}{i\hbar}\sum u_k\circ v_k}:_{K_0} = \infty$. It is rather surprising that the finiteness of the integral

$$:\int_{-\infty}^{\infty} e_*^{it\frac{1}{i\hbar}u\circ v}dt:_K \in \mathcal{E}_{2+}(\mathbb{C}^2)$$

is a generic property, as we will see in §3.4.

3.3. Intertwiners for exponential functions of quadratic forms. — In this section we extend intertwiners to the space $\mathbb{C}e^{\mathfrak{S}(2m)}$ of exponential functions of quadratic forms $ge^{\langle uQ,u\rangle}$, where $g \in \mathbb{C}, Q \in \mathfrak{S}(2m)$. This will be used to obtain K-ordered expressions of star-exponential functions of quadratic forms.

The exact formula for intertwiners is obtained by solving the evolution equation

$$\frac{d}{dt}g(t)e^{\frac{1}{i\hbar}\langle uQ(t),u\rangle} = \sum_{ij} K^{ij}\partial_{u_i}\partial_{u_j}\big(g(t)e^{\frac{1}{i\hbar}\langle uQ(t),u\rangle}\big), \quad Q(0) = A, \quad g(0) = g$$

by setting

$$e^{t\sum_{ij}K^{ij}\partial_{u_i}\partial_{u_j}}(ge^{\frac{1}{i\hbar}\langle uA,u\rangle}) = g(t)e^{\frac{1}{i\hbar}\langle uQ(t),u\rangle}. \tag{26}$$

A direct calculation gives

$$\textstyle\sum_i K^{ij}\partial_{u_i}\partial_{u_j}\big(g(t)e^{\frac{1}{i\hbar}\langle uQ(t),u\rangle}\big) = g(t)\Big(2\mathrm{Tr}K\frac{1}{i\hbar}Q(t) + 4\frac{1}{(i\hbar)^2}(QKQ)_{ij}u_iu_j\Big)e^{\frac{1}{i\hbar}\langle uQ(t),u\rangle}.$$

To find the intertwiner, we solve the ODE system:

$$\begin{cases} \dfrac{d}{dt}Q(t) = \dfrac{4}{i\hbar}Q(t)KQ(t) \\ \dfrac{d}{dt}g(t) = g(t)(\dfrac{2}{i\hbar}\mathrm{Tr}KQ(t)) \end{cases} \qquad Q(0) = A, \quad g(0) = g.$$

Then $Q(t) = \frac{1}{I-\frac{4t}{i\hbar}AK}A$, $g(t) = g(\det(I - \frac{4t}{i\hbar}AK))^{-1/2}$ is the solution of the ODE system by the uniqueness of real analytic solutions.

Here the inverse matrix of X is denoted by $\frac{1}{X}$. Note also that $\frac{1}{X}\frac{1}{Y} = \frac{1}{YX}$. It is easy to check that $\frac{1}{I-AK}A$ is a symmetric matrix by the identity:

$$\frac{1}{I-AK}A = A\frac{1}{I-KA}. \tag{27}$$

Setting $t = \frac{\hbar i}{4}$, we can build the intertwiner I_0^K from

$$Q(\frac{\hbar i}{4}) = \frac{1}{I - AK}A, \quad g(\frac{\hbar i}{4}) = g(\det(I - AK))^{-\frac{1}{2}}. \tag{28}$$

as follows. For $ge^{\frac{1}{i\hbar}\langle uA,u\rangle} = (g; A)$ as before, we call g and A the *amplitude* and *phase* part of $(g; A)$, respectively. In this notation, we see that

$$I_0^K(g; A) = \Big(g \det(I - AK)^{-\frac{1}{2}}; T_K(A)\Big),$$

where $T_K : \mathfrak{S}(2m) \to \mathfrak{S}(2m)$, $\quad T_K(A) = \frac{1}{I-AK}A$ is the phase part of the intertwiner I_0^K.

Computing the inverse $I_K^0 = (I_0^K)^{-1}$ and taking the composition $I_0^{K'} I_K^0$, we easily obtain

$$I_K^{K'}(g; A) = \Big(g \det(I - A(K' - K))^{-\frac{1}{2}}; \frac{1}{I - A(K' - K)}A\Big). \tag{29}$$

The mapping (29) is singular at those A where either $\det(I - A(K' - K)) = 0$ or the sign ambiguity in the square root cannot be removed. We denote the phase part of the intertwiner $I_K^{K'}$ by $T_K^{K'}(A) = \frac{1}{I-A(K'-K)}A$.

Note that the identities

$$T_K^{K'} \sim T_{K'}(T_K)^{-1}, \quad I_K^{K'} \sim I_0^{K'} I_K^0$$

hold in the same sense as the algebraic identities $x/x = 1$, $\sqrt{1+x}/\sqrt{1+x} = 1$, i.e. whenever the denominator is nonzero. Here we use the notation $\sim$ to distinguish such an algebraic calculation. Singularities are moving by this algebraic trick.

By setting $B = \frac{1}{I-A(K'-K)}A$, the r.h.s of (29) is $\Big(g \det(I + B(K' - K))^{\frac{1}{2}}; B\Big)$. Moving branching singularities are a remarkable feature of this calculus.

For every A, $ge^{\frac{1}{i\hbar}\langle uA,u\rangle}$ is an element of $\mathcal{E}_{2+}(\mathbb{C}^n)$. If $(g(\cdot), A(\cdot))$ is a a continuous mapping from a compact manifold M into $\mathbb{C}\times\mathfrak{S}(n)$, then Lemma 2.2 shows that

$$\int_M g(x)e^{\frac{1}{i\hbar}\langle uA(x),u\rangle}dV_x \in \mathcal{E}_{2+}(\mathbb{C}^n).$$

Suppose further that M is simply connected. Since the intertwiner $I_K^{K'}$ is given in a concrete form, we see the following:

Lemma 3.1. — *For every $K, K' \in \mathfrak{S}(n)$,*

$$I_K^{K'}\Big(\int_M g(x)e^{\frac{1}{i\hbar}\langle uA(x),u\rangle}dV_x\Big) = \int_M g(x)I_K^{K'}\Big(e^{\frac{1}{i\hbar}\langle uA(x),u\rangle}\Big)dV_x$$

is also an element of $\mathcal{E}_{2+}(\mathbb{C}^n)$ whenever $\det(I - A(x)(K' - K))$ is nowhere zero on M.

3.4. The general ordered expression of $e_*^{t(z+\frac{1}{i\hbar}u\circ v)}$. — From here on, we set $n = 2m = 2$, and $(u_1, u_2) = (u, v)$. We are mainly concerned with functions of $u \circ v = \frac{1}{2}(u * v + v * u)$ alone. The general ordered expression $:e_*^{t(z+\frac{1}{i\hbar}u\circ v)}:_{_K}$ will be given by applying intertwiners to the normal ordered expression.

For this purpose, we set $2u \circ v = \langle \boldsymbol{u}A, \boldsymbol{u}\rangle$, where $\boldsymbol{u} = (u, v)$, $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. The intertwiner $I_{_K}^{^{K'}}$ is given by (29).

We determine the formula of a general ordering expression $:e_*^{t\frac{1}{i\hbar}2u\circ v}:_{_K}$, $K = \begin{bmatrix} \delta' & \lambda \\ \lambda & \delta \end{bmatrix}$, $\lambda, \delta, \delta' \in \mathbb{C}$.

Setting $B = \begin{bmatrix} 0 & \beta \\ \beta & 0 \end{bmatrix}$, we note that

$$(I - B(K - K_0))^{-1}B = \frac{1}{(1-\beta(\lambda-1))^2 - \beta^2\delta\delta'} \begin{bmatrix} \beta^2\delta & (1-\beta(\lambda-1))\beta \\ (1-\beta(\lambda-1)\beta & \beta^2\delta' \end{bmatrix}$$

Recalling the formulas (24) and (29), we have

$$\begin{aligned} :e_*^{t\frac{1}{i\hbar}2u\circ v}:_{_K} =& \frac{2}{\sqrt{\Delta^2 - (e^t - e^{-t})^2\delta\delta'}} \\ & \times \exp \frac{1}{i\hbar} \frac{e^t - e^{-t}}{\Delta^2 - (e^t - e^{-t})^2\delta\delta'} \Big((e^t - e^{-t})\delta u^2 + \Delta 2uv + (e^t - e^{-t})\delta' v^2\Big) \end{aligned}$$

where $\Delta = (e^t + e^{-t}) - \lambda(e^t - e^{-t})$. Here we note that the sign ambiguity of the square root is removed by choosing a path from the $t = 0$ to t on which no singular point appears, and by choosing the initial condition $e_*^{0\frac{1}{i\hbar}2u\circ v} = 1$ at $t = 0$.

Replacing t by it, we see that

(30)
$$:e_*^{\frac{it}{i\hbar}2u\circ v}:_{_K} = \frac{1}{\sqrt{\Delta_{(K)}(t)}} \exp \frac{1}{\hbar} \frac{\sin t}{\Delta_{(K)}(t)} \langle \boldsymbol{u} \begin{bmatrix} i\delta \sin t & \cos t - i\lambda \sin t \\ \cos t - i\lambda \sin t & i\delta' \sin t \end{bmatrix}, \ \boldsymbol{u}\rangle,$$

where

$$\Delta_{(K)}(t) = \big(\cos t - i(\lambda + \sqrt{\delta\delta'})\sin t\big)\big(\cos t - i(\lambda - \sqrt{\delta\delta'})\sin t\big) \tag{31}$$

Note that $\lambda + \sqrt{\delta\delta'}$ and $\lambda - \sqrt{\delta\delta'}$ can be arbitrary complex numbers. Both (30) and (31) are π-periodic. Here we note that the sign of $\sqrt{\Delta_{(K)}(t)}$ depends on the ordered expression parameter K. It follows that $:e_*^{\pi i\frac{1}{i\hbar}2u\circ v}:_{_K} = \frac{2}{\sqrt{(-2)^2}}$, which is ± 1 depending on K and the path from 0 to πi.

In the remainder of this section, we comment on the appearance of these singular points. The sign ambiguity of $\sqrt{\ }$ cannot be removed on the whole complex plane. Thus these $*$-exponentials are double valued functions of $t \in \mathbb{C}$ in general (cf. [**12**], [**6**]). The sign ambiguity is removed only when $\delta\delta' = 0$ by choosing the initial condition $e_*^{0\frac{1}{i\hbar}2u\circ v} = 1$ at $t = 0$. In this case, cusp singular points appear π (and not 2π)-periodically along a line parallel to the real axis. However, singular points are not stable under general intertwiners, as intertwiners are double valued in general (cf. [**12**]).

From these observations we see that in generic ordered expressions the singular points of $:e_*^{it\frac{1}{i\hbar}2u\circ v}:_K$ appear π-periodically on two lines parallel to the real axis and the ordered expression has $e^{-|t|}$-decay on any line parallel to the imaginary axis. Moreover, the generic ordered expression does not have singular points, and the existence of $\int_{\mathbb{R}} e_*^{t\frac{1}{i\hbar}u\circ v}dt$ is a generic property. However, we see there are several categories for the behavior of expression parameters.

To fix the notation, we denote by $\mathfrak{D}$ the open dense domain of expression parameters K such that $:e_*^{\frac{it}{i\hbar}2u\circ v}:_K$ has no singular point on either the real or imaginary axis. Generic patterns of the properties for $:e_*^{\frac{it}{i\hbar}2u\circ v}:_K$ are as follows:

(1) On a domain $\mathfrak{D}_+$ (resp. $\mathfrak{D}_-$) for the parameter K, the singular set of $:e_*^{\frac{it}{i\hbar}2u\circ v}:_K$ appears only in the open lower (resp. upper) half plane, and the $*$-exponential functions form a complex semi-group over the upper (resp. lower) half plane without sign ambiguity by demanding the value 1 at $t = 0$. $:e_*^{\pm\frac{it}{i\hbar}2u\circ v}:_K$, is alternating π-periodic on the real axis (we call $f(z)$ *alternating π periodic* if $f(z+n\pi) = (-1)^n f(z)$ for any integer n).

(2) On a domain $\mathfrak{D}_0$ for the parameter K, the singular set occurs in both the upper and lower half-planes, but not on the real axis. In this domain, $:e_*^{\pm\frac{it}{i\hbar}2u\circ v}:_K$, is π-periodic on the real axis by demanding the value 1 at $t = 0$.

Note Some delicate arguments about the winding number are required to determine the periodicity of $:e_*^{\pm\frac{it}{i\hbar}2u\circ v}:_K$, as will be discussed in a forthcoming paper.

3.5. Star exponential functions of general quadratic forms. — In this section we give without proof formulas for K-ordered expressions of star exponential functions of general quadratic form, with details in [**6**].

As in [**17**], star exponential functions $e_*^{\frac{1}{i\hbar}\langle\boldsymbol{\xi},\boldsymbol{u}\rangle}$ for a linear form $\langle\boldsymbol{\xi},\boldsymbol{u}\rangle$ are well defined as the family $\{e^{\frac{1}{4i\hbar}\langle\boldsymbol{\xi}K,\boldsymbol{\xi}\rangle}e^{\frac{1}{i\hbar}\langle\boldsymbol{\xi},\boldsymbol{u}\rangle}, K \in \mathfrak{S}(n)\}$. However, for a quadratic form $\langle\boldsymbol{u}A,\boldsymbol{u}\rangle_* = \sum A_{kl}u_k \circ u_l$, the star exponential function $e_*^{\frac{t}{i\hbar}\langle\boldsymbol{u}A,\boldsymbol{u}\rangle_*}$ will be defined only on a dense domain of K-ordered expressions, and is in general a double valued function of $t \in \mathbb{C}$ (cf. [**6**]).

For every $\alpha \in \mathrm{sp}(m,\mathbb{C})$, we first consider the one parameter subgroup $e^{-2t\alpha}$ of $\mathrm{Sp}(m,\mathbb{C})$, and consider the inverse image of the twisted Cayley transform $C_\kappa^{-1}(e^{-2t\alpha})$: For $\kappa \in \mathrm{sp}(m,\mathbb{C})$, we set

(32)
$$C_\kappa^{-1}(e^{-2t\alpha}) = \frac{1}{(I-\kappa)+e^{-2t\alpha}(I+\kappa)}(I-e^{-2t\alpha}) = \frac{1}{\cosh t\alpha - (\sinh t\alpha)\kappa}\sinh t\alpha.$$

The exponential function must lie in a certain submanifold $\widetilde{\mathcal{D}}_\kappa$ through $(1;0)$, and points of this manifold are determined by their phases. Setting $\kappa = JK$, we have

$$:e_*^{s\frac{1}{i\hbar}\langle u(\alpha J),u\rangle_*}:_K = \left(\det(I+C_\kappa^{-1}(e^{-2s\alpha})(I+\kappa))\right)^{\frac{1}{2}} e^{\frac{1}{i\hbar}\langle u(C_\kappa^{-1}(e^{-2s\alpha})J),u\rangle}. \tag{33}$$

More precisely, for every $\alpha \in \mathrm{sp}(m,\mathbb{C})$, the K-ordered expression of the $*$-exponential function is given as follows (see [**12**]-[**13**] for special cases):

$$:e_*^{\frac{t}{i\hbar}\langle u(\alpha J),u\rangle_*}:_K = \frac{2^m}{\sqrt{\det(I-\kappa+e^{-2t\alpha}(I+\kappa))}} e^{\frac{t}{i\hbar}\langle u\frac{1}{I-\kappa+e^{-2t\alpha}(I+\kappa)}(I-e^{-2t\alpha})J,u\rangle}. \tag{34}$$

It is not hard to see that (34) is the real analytic solution of (12). Note that $\det e^{t\alpha}I = 1$ for every $\alpha \in \mathrm{sp}(m,\mathbb{C})$. Thus (34) can be rewritten as

(35)
$$:e_*^{\frac{t}{i\hbar}\langle u(\alpha J),u\rangle_*}:_K = \frac{2^m}{\sqrt{\det(e^{t\alpha}(I-\kappa)+e^{-t\alpha}(I+\kappa))}} e^{\frac{t}{i\hbar}\langle u\frac{1}{e^{t\alpha}(I-\kappa)+e^{-t\alpha}(I+\kappa)}(e^{t\alpha}-e^{-t\alpha})J,u\rangle}$$

In spite of the sign ambiguity of the square root, the exponential law

$$:e_*^{s\frac{1}{i\hbar}\langle u(\alpha J),u\rangle_*}:_K *_K :e_*^{t\frac{1}{i\hbar}\langle u(\alpha J),u\rangle_*}:_K = :e_*^{(s+t)\frac{1}{i\hbar}\langle u(\alpha J),u\rangle_*}:_K \tag{36}$$

holds using $\sqrt{a}\sqrt{b} = \sqrt{ab}$ without regard to sign ambiguities, as the exponential law and associativity hold on the group $\mathrm{Sp}(m,\mathbb{C})$. Note however that we allow $\sqrt{1} = \pm 1$.

To treat these formulas without sign ambiguity, we always have to specify a path with no singular points from $t = 0$ to the considered point.

From (34) we derive the following:

Proposition 3.4. — *If $e^{2\pi\alpha} = I$ (e.g. $\alpha = J$), then $:e_*^{\pi\frac{1}{i\hbar}\langle u\alpha J,u\rangle_*}:_K = \sqrt{1}$ independent of K.*

The sign of $\sqrt{1}$ depends on the K-ordered expression and also on a path from 0 to π as above.

Hence, even though $(\sqrt{1})^2 = 1$ is trivial, the strict exponential law may fail, that is, $:e_*^{2\pi\frac{1}{i\hbar}\langle u\alpha J,u\rangle_*}:_K = 1$ or $:e_*^{\pi\frac{1}{i\hbar}\langle u\alpha J,u\rangle_*}:_K *_K :e_*^{\pi\frac{1}{i\hbar}\langle u\alpha J,u\rangle_*}:_K = 1$ may not hold automatically. If $:e_*^{t\frac{1}{i\hbar}\langle u\alpha J,u\rangle_*}:_K$ has a singular point on the interval $[0,2\pi]$, then it may happen that

$$(e_*^{\pi\frac{1}{i\hbar}\langle u\alpha J,u\rangle_*})^2 \neq e_*^{2\pi\frac{1}{i\hbar}\langle u\alpha J,u\rangle_*},$$

although equality holds up to sign. In spite of this, we have

Proposition 3.5. — *If* $e^{2\pi\alpha} = I$, *then* $(e_*^{\pi\frac{1}{i\hbar}\langle u\alpha J,u\rangle_*})^2 = 1$ *for every K-ordered expression such that* $:e_*^{t\frac{1}{i\hbar}\langle u\alpha J,u\rangle_*}:_{_K}$ *has no singular point on the interval* $[0,\pi]$.

Proof Note first that this is by no means trivial, because $:e_*^{t\frac{1}{i\hbar}\langle u\alpha J,u\rangle_*}:_{_K}$ may have a singular point on the interval $[\pi, 2\pi]$. Since $:e_*^{\pi\frac{1}{i\hbar}\langle u\alpha J,u\rangle_*}:_{_K} = \pm 1$, one can define

$$:e_*^{t\frac{1}{i\hbar}\langle u\alpha J,u\rangle_*}:_{_K} *_{_K} :1:_{_K}, \quad \text{or} \quad :e_*^{t\frac{1}{i\hbar}\langle u\alpha J,u\rangle_*}:_{_K} *_{_K} :(-1):_{_K}$$

by the solution of the evolution equation (12). By Proposition 3.2,

$$:e_*^{t\frac{1}{i\hbar}\langle u\alpha J,u\rangle_*}:_{_K} *_{_K} :e_*^{\pi\frac{1}{i\hbar}\langle u\alpha J,u\rangle_*}:_{_K}$$

is the solution of (12). This gives the result. □

By (34), we also see that $:e_*^{s\frac{1}{i\hbar}\langle u(\alpha J),u\rangle_*}:_{_K}$ has in general discrete branching singularities in $\mathbb{C}$ with some periodicity depending on the parameter $\kappa = JK$.

4. Criteria for associativity

In this section, we give several criteria which imply associativity for the extended product $*_{_K}$. However, we note that there is no generally applicable lemma guaranteeing associativity. For simplifying notation, we often omit the subscript K of the product $*_{_K}$ and the expression $: :_{_K}$ if it contains no confusion.

4.1. Remarks on star exponential functions. — We first note how to define rigorously the product of star exponential functions and a general function $e_*^{z\frac{1}{i\hbar}u\circ v} * f(u,v)$. There are essentially two approaches. The first is, as mentioned in §3.1, to use the real analytic solution f_t of

$$\frac{d}{dt}f_t = \frac{1}{i\hbar}u\circ v * f_t,$$

with the initial condition $f_0 = f(u,v)$ provided such a solution exists. The second approach is to define

$$e_*^{z\frac{1}{i\hbar}u\circ v} * f(u,v) = \lim_{n\to\infty} e_*^{z\frac{1}{i\hbar}u\circ v} * f_n(u,v), \quad \text{if} \ \ f(u,v) = \lim_{n\to\infty} f_n(u,v),$$

where f_n are polynomials. These two definitions do not coincide in general, since multiplication by $e_*^{z\frac{1}{i\hbar}u\circ v}*$ is not a continuous linear mapping of $\mathit{Hol}(\mathbb{C}^2)$ to itself (cf. (42), (43)). Note that $e_*^{z\frac{1}{i\hbar}u\circ v} \in \mathcal{E}_{2+}(\mathbb{C}^2)$. If $f(u,v) \in \mathcal{E}_{2-}(\mathbb{C}^2) = \bigcup_{p<2}\mathcal{E}_p(\mathbb{C}^2)$ with the inductive limit topology, then the two definitions coincide.

Since star exponential functions of quadratic forms are elements of $\mathcal{E}_{2+}(\mathbb{C}^n)$, their product may not be defined, and even if the product is defined associativity may not hold.

We show that $\int_{-\infty}^{\infty} e_*^{t\frac{1}{i\hbar}u\circ v}dt \in \mathcal{E}_{2+}(\mathbb{C}^2)$ in the Weyl ordered expression. In the Weyl ordered expression, we have $:e_*^{t\frac{1}{i\hbar}u\circ v}:_0 = \frac{1}{\cosh\frac{t}{2}}e^{(\tanh\frac{t}{2})\frac{1}{i\hbar}2uv}$. Thus,

$$:\int_{\mathbb{R}} e_*^{t\frac{1}{i\hbar}u\circ v}dt:_0 = \int_{-\infty}^{\infty}\frac{1}{\cosh\frac{t}{2}}e^{(\tanh\frac{t}{2})\frac{1}{i\hbar}2uv}dt.$$

For $\cos s = \tanh\frac{t}{2}$, $-2\sin s ds = \sin^2 s dt$, the integral on the r.h.s. of the last equation becomes

$$2\int_{-\pi}^{0} e^{(\cos s)\frac{1}{i\hbar}2uv}ds = \int_{-\pi}^{\pi} e^{(\cos s)\frac{1}{i\hbar}2uv}ds.$$

Since $g(s) = e^{(\cos s)\frac{1}{i\hbar}uv}$ is a continuous curve in $\mathcal{E}_{2+}(\mathbb{C}^2)$, Lemma 2.2 implies that the last integral belongs to $\mathcal{E}_{2+}(\mathbb{C}^2)$. Hence, by Lemma 3.1 this property is generic.

Using the intertwiner I_0^{K}, we see that $:\int_{\mathbb{R}} e_*^{t\frac{1}{i\hbar}u\circ v}dt:_K = \int_{-\pi}^{\pi} :e^{(\cos s)\frac{1}{i\hbar}2u\circ v}:_K ds \in \mathcal{E}_{2+}(\mathbb{C}^2)$. Then we have

Proposition 4.1. — *In generic ordered expressions, the integral* $:\int_{-\infty}^{\infty} e_*^{t\frac{1}{i\hbar}u\circ v}dt:_K$ *is in* $\mathcal{E}_{2+}(\mathbb{C}^2)$. *Furthermore, integration by parts gives* $\frac{d}{d\theta}\int_{-\infty}^{\infty} e_*^{e^{i\theta}t\frac{1}{i\hbar}u\circ v}e^{i\theta}dt = 0$ *whenever the integral is defined.*

We have seen that in generic ordered expressions, $\frac{1}{i\hbar}u\circ v$ has two different inverses,

$$(\frac{1}{i\hbar}u\circ v)_{*+}^{-1} = \int_{-\infty}^{0} e_*^{t\frac{1}{i\hbar}u\circ v}dt, \quad (\frac{1}{i\hbar}u\circ v)_{*-}^{-1} = -\int_{0}^{\infty} e_*^{t\frac{1}{i\hbar}u\circ v}dt$$

in the space $\mathcal{E}_{2+}(\mathbb{C}^n)$, which implies the failure of associativity in general:

$$(37)\quad \Big((\frac{1}{i\hbar}u\circ v)_{*+}^{-1} * (\frac{1}{i\hbar}u\circ v)\Big) * (\frac{1}{i\hbar}u\circ v)_{*-}^{-1} \neq (\frac{1}{i\hbar}u\circ v)_{*+}^{-1} * \Big((\frac{1}{i\hbar}u\circ v) * (\frac{1}{i\hbar}u\circ v)_{*-}^{-1}\Big).$$

Indeed $(\frac{1}{i\hbar}u\circ v)_{*+}^{-1} * (\frac{1}{i\hbar}u\circ v)_{*-}^{-1}$ diverges in any ordered expression. This gives an example where $(f*g)*h = f*(g*h)$ does not hold even if g is a polynomial.

4.2. Basic criteria for associativity for the extended product. — Suppose $f, g \in \mathit{Hol}(\mathbb{C}^n)$ are given by $f = \lim f_k, g = \lim g_\ell$ in the topology of $\mathit{Hol}(\mathbb{C}^n)$ for sequences $\{f_k\}, \{g_\ell\} \subset \mathrm{Hol}(\mathbb{C}^n)$. Even if $f*g$ and $\lim_\ell f*g_\ell$ exist, $f*g$ may not equal $\lim_\ell f*g_\ell$, since $f*$ is not continuous in general. Moreover, it may happen that even though $\lim g_\ell$ diverges, $\lim_\ell f*g_\ell$ exists.

If $f = \lim f_k, g = \lim g_\ell$, we have

$$\lim_k f_k * p(\boldsymbol{u}) = f * p(\boldsymbol{u}), \quad \lim_\ell p(\boldsymbol{u}) * g_\ell = p(\boldsymbol{u}) * g$$

for every polynomial. However, as we saw in (37), we may have

$$\lim_k(\lim_\ell f_k * (p(\boldsymbol{u}) * g_\ell) \neq \lim_\ell(\lim_k f_k * p(\boldsymbol{u})) * g_\ell),$$

even if both sides exist. In this case, $\lim_{(k,\ell)} f_k * p(\boldsymbol{u}) * g_\ell$ does not converge.

Suppose $f_k * g_\ell$ converges to an element h in $Hol(\mathbb{C}^n)$. Then we define $f * g = h$, i.e.

$$f * g \;=\; \lim_{(k,\ell)\to\infty} f_k * g_\ell = h. \tag{38}$$

where in the limit $k^2 + \ell^2 \to \infty$. The same definition is also employed for the product
(39)

$$\int_{-\infty}^{0} f(s) e_*^{s(z+\frac{1}{i\hbar}u\circ v)} ds * \int_{-\infty}^{0} g(t) e_*^{t(z+\frac{1}{i\hbar}u\circ v)} dt = \lim_{(S,T)} \iint_{-(S,T)}^{(0,0)} f(s)g(t) e_*^{(s+t)(z+\frac{1}{i\hbar}u\circ v)} ds dt$$

although these integrals are not in $\mathcal{E}_2(\mathbb{C}^n)$.

Suppose $f, g \in Hol(\mathbb{C}^n)$ are given as $f = \lim f_k, g = \lim g_\ell$ in the topology of $Hol(\mathbb{C}^n)$ as above. For polynomials $p(\boldsymbol{u})$, $q(\boldsymbol{u})$, Lemma 2.1 gives that $\lim_k p(\boldsymbol{u}) * f_k = p(\boldsymbol{u}) * f$, $\lim_k q(\boldsymbol{u}) * g_k = q(\boldsymbol{u}) * g$.

Lemma 4.1. — *Suppose that associativity holds for the approximating series:*

$$(p(\boldsymbol{u}) * f_k) * (q(\boldsymbol{u}) * g_\ell) = ((p(\boldsymbol{u}) * f_k) * q(\boldsymbol{u})) * g_\ell),$$

and $\lim_{(k,\ell)}(p(\boldsymbol{u}) * f_k) * (q(\boldsymbol{u}) * g_\ell)$ *converges to an element* h *in* $Hol(\mathbb{C}^n)$. *Then* $(p(\boldsymbol{u}) * f) * (q(\boldsymbol{u}) * g)$ *equals* h, *and the following associativity holds:*

$$(p(\boldsymbol{u}) * f) * (q(\boldsymbol{u}) * g) = (p(\boldsymbol{u}) * f * q(\boldsymbol{u})) * g.$$

Proof. — By definition, we have $(p(\boldsymbol{u})*f)*(q(\boldsymbol{u})*g) = \lim_{(k,\ell)}(p(\boldsymbol{u})*f_k)*(q(\boldsymbol{u})*g_\ell)$. Using the associativity of the inside of the r.h.s. of the last equation in Lemma 4.1, we have

$$\lim_{(k,\ell)} (p(\boldsymbol{u}) * f_k) * (q(\boldsymbol{u}) * g_\ell) = \lim_{(k,\ell)} (p(\boldsymbol{u}) * f_k * q(\boldsymbol{u})) * g_\ell.$$

From Lemma 2.1, we see that $\lim_k p(\boldsymbol{u}) * (f_k * q(\boldsymbol{u})) = p(\boldsymbol{u}) * f * q(\boldsymbol{u})$. It follows that

$$(p(\boldsymbol{u}) * f) * (q(\boldsymbol{u}) * g) = \lim_{(k,\ell)} (p(\boldsymbol{u}) * (f_k * q(\boldsymbol{u}))) * g_\ell = (p(\boldsymbol{u}) * f * q(\boldsymbol{u})) * g. \qquad \square$$

Note that if the approximating series are in $\mathcal{E}_2(\mathbb{C}^n)$, then associativity holds before the limiting procedure.

Lemma 4.2. — *Suppose* $f, g, f * g$ *are given as in* (38). *Then, for any polynomials* $p(\boldsymbol{u}), q(\boldsymbol{u})$, *the product* $(p(\boldsymbol{u}) * f) * (g * q(\boldsymbol{u}))$ *is defined and associativity holds:*

$$(p(\boldsymbol{u}) * f) * (g * q(\boldsymbol{u})) = p(\boldsymbol{u}) * (f * g) * q(\boldsymbol{u}).$$

Proof. — By Lemma 2.1, we see that $p(\boldsymbol{u}) * f = \lim_k p(\boldsymbol{u}) * f_k$, $g * q(\boldsymbol{u}) = \lim_\ell (g_\ell * q(\boldsymbol{u}))$, and the product is defined by

$$(p(\boldsymbol{u}) * f) * (g * q(\boldsymbol{u})) = \lim_{(k,\ell)} (p(\boldsymbol{u}) * f_k) * (g_\ell * q(\boldsymbol{u})) = \lim_{(k,\ell)} p(\boldsymbol{u}) * (f_k * g_\ell) * q(\boldsymbol{u}).$$

Hence Lemma 2.1 gives $(p(\boldsymbol{u}) * f) * (g * q(\boldsymbol{u})) = p(\boldsymbol{u}) * (f * g) * q(\boldsymbol{u})$. $\square$

It does not seem that the existence of $\lim_{(k,\ell)} f_k * g_\ell$ yields that of $\lim_{(k,\ell)} f_k * (p(\boldsymbol{u}) * g_\ell)$ or $\lim_{(k,\ell)}(f_k * p(\boldsymbol{u})) * g_\ell$ for every polynomial $p(\boldsymbol{u})$. In spite of this, we have the following for the special element $u \circ v$:

Lemma 4.3. — *If* (39) *is defined, then*

$$\int_{-\infty}^{0} f(s)e_*^{s(z+\frac{1}{i\hbar}u\circ v)}ds * p(\boldsymbol{u}) * \int_{-\infty}^{0} g(t)e_*^{t(z+\frac{1}{i\hbar}u\circ v)}dt$$

is defined for every polynomial $p(\boldsymbol{u})$.

Proof. — Using the "bumping identity" :

$$v * f(u * v) = f(v * u) * v$$

several times, we find a polynomial $\tilde{p}(\boldsymbol{u})$ such that:

$$p(\boldsymbol{u}) * \int_{-\infty}^{0} g(t)e_*^{t(z+\frac{1}{i\hbar}u\circ v)}dt = \int_{-\infty}^{0} g(t)p(\boldsymbol{u}) * e_*^{t(z+\frac{1}{i\hbar}u\circ v)}dt = \int_{-\infty}^{0} g(t)e_*^{t(z+\frac{1}{i\hbar}u\circ v)}dt * \tilde{p}(\boldsymbol{u}).$$

Hence Lemma 4.2 gives the result. □

In the general setting, suppose the limits $f * g = \lim_{(k,\ell)} f_k * g_\ell$ in (38) and $\lim_{(k,\ell)} \partial^\alpha f_k * \partial^\beta g_\ell$ exist for every α, β. Then it is not hard to show the existence of $\lim_{(k,\ell)} f_k * (p(\boldsymbol{u}) * g_\ell)$ and $\lim_{(k,\ell)}(f_k * p(\boldsymbol{u})) * g_\ell$ for every polynomial $p(\boldsymbol{u})$.

The following is useful in concrete computations. Note that for $(\mathbb{C}[\boldsymbol{u}][[\hbar]], *_K)$, the space of formal power series in $\hbar$, the $*_K$-product is always defined by the product formula (5) and associativity holds. The elements of $\mathcal{E}_{2+}(\mathbb{C}^n)$ are often given as a real analytic function of $\hbar$ defined on a certain interval containing $\hbar = 0$.

The following is easy to see:

Lemma 4.4. — *Suppose $f(\hbar, \boldsymbol{u})$, $g(\hbar, \boldsymbol{u})$ and $h(\hbar, \boldsymbol{u})$ are given as real analytic functions of $\hbar$ in some interval* $[0, H]$.

If $f(\hbar, \boldsymbol{u}) * g(\hbar, \boldsymbol{u})$, $(f(\hbar, \boldsymbol{u}) * g(\hbar, \boldsymbol{u})) * h(\hbar, \boldsymbol{u})$, $g(\hbar, \boldsymbol{u}) * h(\hbar, \boldsymbol{u})$ and $f(\hbar, \boldsymbol{u}) * (g(\hbar, \boldsymbol{u}) * h(\hbar, \boldsymbol{u}))$ are defined as real analytic functions on $[0, H]$, then the following associativity holds:

$$\big(f(\hbar, \boldsymbol{u}) * g(\hbar, \boldsymbol{u})\big) * h(\hbar, \boldsymbol{u}) = f(\hbar, \boldsymbol{u}) * \big(g(\hbar, \boldsymbol{u}) * h(\hbar, \boldsymbol{u})\big)$$

Remark In the following, elements are often given in the form $f(\frac{1}{i\hbar}\varphi(t), \boldsymbol{u})$ for a real analytic function $f(t, \boldsymbol{u})$ in $t \in [0, T]$, where $\varphi(t)$ is a real analytic function such that $\varphi(0) = 0$ (cf. (24)). In such a case, replacing t by $s\hbar$ gives a real analytic function of $\hbar$, and such an element lies in $(\mathbb{C}[\boldsymbol{u}][[\hbar]], *_K)$. Thus, we can apply Lemma 4.4.

However, there are many elements in $\mathcal{E}_{2+}(\mathbb{C}^n)$ of the form $f(\frac{1}{i\hbar}\varphi(t), \boldsymbol{u})$ such that $\varphi(0) \neq 0$. For these elements we have to use Lemmas 4.1 and 4.2 carefully.

As mentioned before, we know that $:e_*^{t\frac{1}{i\hbar}u\circ v}:_K \in Hol(\mathbb{C}^2)$ for every fixed t whenever defined. We also see that $:e_*^{t\frac{1}{i\hbar}u\circ v}:_K$ is rapidly decreasing with respect to t in a generic ordered expression.

5. Vacuums and their matrix element expressions

In this section, we give properties of vacuums which we can compare to similar properties in operator theory.

Noting that $v * u = u \circ v + \frac{1}{2}i\hbar$, we begin with the following:

Proposition 5.1. — *In generic ordered expressions with no singular points on the real axis, we have*

$$\lim_{t\to-\infty} e_*^{t\frac{1}{i\hbar}2v*u} = 0, \quad \lim_{t\to\infty} e_*^{t\frac{1}{i\hbar}2u*v} = 0,$$

and the following limits exist:

$$\lim_{t\to\infty} e_*^{t\frac{1}{i\hbar}2v*u} = \varpi_{00}, \quad \lim_{t\to-\infty} e_*^{t\frac{1}{i\hbar}2u*v} = \overline{\varpi}_{00}.$$

We call ϖ_{00} and $\overline{\varpi}_{00}$ the **vacuum** and **bar-vacuum** respectively. Strictly speaking, such vacuums should be defined together with the one parameter semigroups $e_*^{t\frac{1}{i\hbar}2v*u}$, $e_*^{-t\frac{1}{i\hbar}2v*u}$, $t\geq 0$, for they depend on the K-ordered expression and may change sign if there are singular points on $t\geq 0$. When the ordered expressions $K(s)$, $s \in I$, move along a curve, we require that $:e_*^{t\frac{1}{i\hbar}2u*v}:_{K(s)}$ has no singular point on $[0,\infty)\times I$. Since the $*$-exponential function $e_*^{t\frac{1}{i\hbar}2u*v}$ can be defined as a single valued element by requiring it equal 1 at $t = 0$, the sign ambiguity does not occur in the K-ordered expression. Thus, we have

$$\begin{aligned}
&\lim_{t\to\infty} :e_*^{t\frac{1}{i\hbar}2v*u}:_K = \frac{2}{\sqrt{(1-\lambda)^2+\delta\delta'}} e^{\frac{1}{i\hbar}\frac{1}{(1-\lambda)^2-\delta\delta'}(\delta u^2+(1-\lambda)2uv+\delta' v^2)},\\
(40)\qquad &\lim_{t\to-\infty} :e_*^{t\frac{1}{i\hbar}2u*v}:_K = \frac{2}{\sqrt{(1+\lambda)^2+\delta\delta'}} e^{-\frac{1}{i\hbar}\frac{1}{(1+\lambda)^2-\delta\delta'}(\delta u^2+(1+\lambda)2uv+\delta' v^2)},\\
&\lim_{t\to-\infty} :e_*^{t\frac{1}{i\hbar}2v*u}:_K = 0, \quad \lim_{t\to\infty} :e_*^{t\frac{1}{i\hbar}2u*v}:_K = 0.
\end{aligned}$$

The exponential law gives

$$\varpi_{00} *_0 \varpi_{00} = \varpi_{00}, \quad \overline{\varpi}_{00} *_0 \overline{\varpi}_{00} = \overline{\varpi}_{00}.$$

However, we easily see

Theorem 5.1. — *The product $\varpi_{00} *_0 \overline{\varpi}_{00}$ diverges in any ordered expression.*

The existence of the limits (40) also gives

$$u * v * \varpi_{00} = 0 = \varpi_{00} * u * v,$$

but the bumping identity $v * f(u * v) = f(v * u) * v$ gives the following:

Lemma 5.1. — *$v * \varpi_{00} = 0 = \varpi_{00} * u$ in generic ordered expressions.*

Proof. — Using the continuity of $v*$, we see that $v * \lim_{t \to -\infty} e_*^{t\frac{1}{i\hbar}2u*v} = \lim_{t \to -\infty} v * e_*^{t\frac{1}{i\hbar}2u*v}$. Hence, the bumping identity proved by the uniqueness of the real analytic solution for linear ODE and (40) give $\lim_{t \to -\infty} e_*^{t\frac{1}{i\hbar}2v*u} * v = 0$. □

The following identities ensure associativity:

Lemma 5.2. — $\varpi_{00} * (u^p * \varpi_{00}) = 0$, *and* $(\varpi_{00} * v^p) * \varpi_{00} = 0$.

Proof. — By the formal power series expansion in $i\hbar$ for e_*^{su*v}, associativity for the equations in Lemma 5.2 holds, and the following computation is justified by the bumping identity:

$$e_*^{su*v} * (u^p * e_*^{tu*v}) = (e_*^{su*v} * u^p) * e_*^{tu*v} = u^p * e_*^{(s+t)u*v + i\hbar ps}.$$

The r.h.s of this equation is continuous in s, t. In particular,

$$\lim_{t \to a} e_*^{su*v} * (u^p * e_*^{tu*v}) = e_*^{su*v} * \lim_{t \to a} (u^p * e_*^{tu*v}).$$

Using the bumping identity, we have

$$\begin{aligned} e_*^{su*v} * (u^p * \lim_{t \to -\infty} e_*^{tu*v}) &= e_*^{su*v} * \lim_{t \to -\infty} u^p * e_*^{tu*v} = \lim_{t \to -\infty} u^p * e_*^{(s+t)u*v + i\hbar ps} \\ &= u^p * \lim_{t \to -\infty} e_*^{(s+t)u*v + i\hbar ps} = u^p e^{i\hbar ps} * \varpi_{00}. \end{aligned}$$

It follows that

$$\varpi_{00} * (u^p * \varpi_{00}) = \lim_{s \to -\infty} e_*^{s\frac{1}{i\hbar}u*v} * (\lim_{t \to -\infty} u^p * e_*^{t\frac{1}{i\hbar}u*v}) = \lim_{s \to -\infty} u^p e^{ps} * \varpi_{00} = 0.$$

Similarly, we also have $(\varpi_{00} * v^p) * \varpi_{00} = 0$. □

Lemma 5.3. — *For every polynomial $f(u, v) = \sum a_{pq} u^p * v^q$,*

$$\varpi_{00} * (f(u, v) * \varpi_{00}) = f(0, 0)\varpi_{00} = (\varpi_{00} * f(u, v)) * \varpi_{00}.$$

Consequently, associativity holds for $\varpi_{00} * f(u, v) * \varpi_{00}$ for all polynomials $f(u, v)$.

Reasoning as above, we see that

$$(e_*^{su*v} * v^q) * (u^p * e_*^{tu*v}) = e_*^{su*v} * (v^q * u^p * e_*^{tu*v}) = e^{(q-p)ti\hbar} e_*^{(s+t)u*v} * v^q * u^p \quad \text{for } q \geq p,$$

$$(e_*^{su*v} * v^q) * (u^p * e_*^{tu*v}) = e_*^{su*v} * (v^q * u^p * e_*^{tu*v}) = v^q * u^p * e_*^{(s+t)u*v} * e^{(p-q)si\hbar} \quad \text{for } q \leq p.$$

Replacing s, t by $\frac{1}{i\hbar}s$, $\frac{1}{i\hbar}t$ and taking the limits $t \to -\infty$ and $s \to \infty$ for the case $p \geq q$ and $q \geq p$ respectively, we have

(41) $(\varpi_{00} * v^q) * (u^p * \varpi_{00}) = \delta_{p,q} p!(i\hbar)^p = \varpi_{00} * (v^q * u^p * \varpi_{00}) = (\varpi_{00} * v^q * u^p) * \varpi_{00}$.

Since $\varpi_{00} * v^q * u^p * \varpi_{00} = \delta_{p,q} p!(i\hbar)^p \varpi_{00}$, we have the following:

Proposition 5.2. — $\frac{1}{\sqrt{p!q!(i\hbar)^{p+q}}} u^p * \varpi_{00} * v^q$ *is the* (p,q)*-matrix element.*

As mentioned in the Remark in §3.4, we have two definitions of $e_*^{z\frac{1}{i\hbar}u\circ v} * f(u,v)$. However, both definitions satisfy

$$e_*^{z\frac{1}{i\hbar}u\circ v} * \varpi_{00} = e^{-\frac{1}{2}z} * \varpi_{00}. \tag{42}$$

Remark In contrast, since $\frac{1}{i\hbar}u \circ v * \delta_*(\frac{1}{\hbar}u \circ v) = 0$, where $\delta_*(\frac{1}{\hbar}u \circ v) = \int_{\infty}^{\infty} e^{s\frac{1}{i\hbar}u\circ v}$, we must set $e_*^{t\frac{1}{i\hbar}u\circ v} * \delta_*(\frac{1}{\hbar}u \circ v) = \delta_*(\frac{1}{\hbar}u \circ v)$ as the real analytic solution of $\frac{d}{dt}f_t = \frac{1}{i\hbar}u \circ v * f_t$. However, computing

$$\lim_{N\to\infty} e_*^{t\frac{1}{i\hbar}u\circ v} * \int_{-N}^{N} e_*^{s\frac{1}{i\hbar}u\circ v} ds = \lim_{N\to\infty} \int_{-N}^{N} e_*^{(t+s)\frac{1}{i\hbar}u\circ v} ds$$

gives the following:

$$e_*^{(x+iy)\frac{1}{i\hbar}u\circ v} * \delta_*(\frac{1}{i\hbar}u \circ v) = e_*^{iy\frac{1}{i\hbar}u\circ v} * \delta_*(\frac{1}{\hbar}u \circ v). \tag{43}$$

Note that $e_*^{i\pi\frac{1}{i\hbar}u\circ v} = -1$ in the Weyl ordered expression. Thus, (42) is holomorphic with respect to z, while (43) is only continuous and not real analytic with respect to $z = x + iy$.

6. Inverses and their analytic continuation

6.1. The Hadamard finite part procedure. — We first recall the Hadamard finite part procedure, a well known technique in distribution theory to extract a finite quantity from a divergent expression. (cf. [**19**]). We reformulate this procedure on abstract Fréchet algebra in order to extract information on the eigenspaces of a given matrix via its inverse. We conclude that the element $\frac{1}{i\hbar}u \circ v$ is an indeterminate lying in a discrete set. Let $(\mathcal{A}; *)$ be a complex, complete, topological associative Fréchet algebra with 1 and $\tilde{\mathcal{A}}$ a Fréchet space with a $(\mathcal{A}; *)$-bimodule structure (i.e. a continuous bilinear product $*$ is defined for $\mathcal{A}\times\tilde{\mathcal{A}}, \tilde{\mathcal{A}} \times \mathcal{A}$ into $\tilde{\mathcal{A}}$ with the natural associativity). We call $\lambda \in \mathbb{C}$ a resolvent of $X \in \mathcal{A}$ if $\lambda - X$ has inverse $(\lambda - X)^{-1}$ in $\tilde{\mathcal{A}}$.

Suppose the resolvent set $\rho(X)$ of $X \in \mathcal{A}$ is open and dense in $\mathbb{C}$, and $(\zeta - X)^{-1}$ is holomorphic in $\zeta \in \rho(X)$. Since $(\zeta - X) * (\zeta - X)^{-1} = 1$ on the open dense domain

$\rho(X)$, the singularities of this equation are all removable in the usual complex analysis sense.

An isolated singular point z_0 of $(\zeta - X)^{-1}$ is a pole, if $(\zeta - X)^{-1}$ can be expressed in the form

$$(\zeta - X)^{-1} = \frac{A_{-d}}{(\zeta - z_0)^d} + \cdots + \frac{A_{-1}}{\zeta - z_0} + A_0 + \cdots$$

on a neighborhood of z_0. We call A_0 the finite part of $(\zeta - X)^{-1}$ and denote the finite part by $\mathrm{FP}((\zeta - X)^{-1})$.

In general, for an $\tilde{\mathcal{A}}$-valued holomorphic function $f(z)$ with a pole at $z = z_0$ the finite part $\mathrm{FP}(f(z))$ is defined as follows:

$$\mathrm{FP}(f(z)) = \begin{cases} f(z) & z \neq z_0 \\ \mathrm{Res}_{w=0} \frac{1}{w}(f(z_0 + w)) & z = z_0. \end{cases}$$

This definition is valid for z in a neighborhood of z_0 containing no other pole. Although $(\zeta - X) * (\zeta - X)^{-1} = 1$ for $\zeta \neq z_0$, we have

$$(\zeta - X) * \mathrm{FP}(\zeta - X)^{-1} = \begin{cases} 1 & \zeta \neq z_0 \\ 1 - A_{-1} & \zeta = z_0 \end{cases}$$

where we use $(z_0 - X)A_0 + A_{-1} = 1$, which follows easily from the identity $(\zeta - X) * (\zeta - X)^{-1} = 1$. We will employ this trick to analyze singularities of $(\zeta - X)^{-1}$ in calculations in extensions of star algebras. In particular, we use this procedure to define a new product by

$$(\zeta - X)\tilde{*}(\zeta - X)^{-1} = (\zeta - X) * \mathrm{FP}(\zeta - X)^{-1}.$$

Note that this trick applied to the usual matrix algebra naturally relates to generalized eigenspaces. For a matrix X of finite rank with the eigenvalues $\lambda_1, \ldots, \lambda_n$, we have

$$(zI - X)\tilde{*}(zI - X)^{-1} = \begin{cases} I & z \neq \lambda_1, \ldots, \lambda_n \\ I - P_i & z = \lambda_i \end{cases}$$

where P_i is the projection to the generalized eigenspace corresponding to the eigenvalue λ_i.

Since the inverse $(\zeta - X)^{-1}$ is given very often via the Laplace transform $\int_{-\infty}^{0} e_*^{t(\zeta - X)} dt$, we have the following theorem:

Theorem 6.1. — *Let $\mathcal{A}$ and $\tilde{\mathcal{A}}$ be as above. Suppose $X \in \mathcal{A}$ is an element such that the equation*

$$\frac{d}{dz} f(z) = X * f(z), \quad f(0) = 1 \tag{44}$$

has a complex analytic solution in the Fréchet space $\tilde{\mathcal{A}}$ defined on a connected open domain D. If $\lambda+X$ has an inverse in the Fréchet algebra $\mathcal{A}$ for some $\lambda\in\mathbb{C}$, then D is simply connected and $f(z)=\sum\frac{z^n}{n!}X_*^n$.

Proof. — The proof is elementary. Denote the solution by e_*^{zX}. Let $\Sigma(X)$ be the set of singular points of e_*^{zX} in $\mathbb{C}$. If $\mathbb{C}\setminus\Sigma(X)$ is not simply connected, there is a closed curve C in D surrounding a singular point z_0. □

By the uniqueness of real analytic solutions, the exponential law $e_*^{zX}*e_*^{wX}=e_*^{(z+w)X}$ holds, provided all three terms exist. Suppose there is a $\lambda\in\mathbb{C}$ such that $(\lambda+X)^{-1}\in\mathcal{A}$. Since $e^{z\lambda}e_*^{zX}$ is the solution of the equation $\frac{d}{dz}f_z=(\lambda+X)*f_z$, we derive a second exponential law $e^{z\lambda}e_*^{zX}=e_*^{z(\lambda+X)}$. It follows that $\Sigma(\lambda+X)=\Sigma(X)$.

Obviously, for every integer $k\geq 0$ the contour integral $\int_C(z-z_0)^k e_*^{z(\lambda+X)}dz$ gives an element of $\tilde{\mathcal{A}}$. It follows that

$$(\lambda+X)^{k+1}*\int_C(z-z_0)^k e_*^{z(\lambda+X)}dz=\int_C(z-z_0)^k(\lambda+X)^{k+1}*e_*^{z(\lambda+X)}dz$$
$$=\int_C(z-z_0)^k\frac{d^{k+1}}{dz^{k+1}}e_*^{z(\lambda+X)}dz=(-1)^k\int_C\frac{d}{dz}e_*^{z(\lambda+X)}dz=0.$$

The existence of $(\lambda+X)^{-1}$ gives $\int_C(z-z_0)^k e_*^{z(\lambda+X)}dz=0$ for every integer k, which implies that z_0 is not a singular point. Thus, D is an open simply connected neighborhood of the origin. Standard Taylor series methods yield $f(z)=e_*^{zX}=\sum\frac{1}{n!}(zX)^n$. □

This theorem suggests that we have to go beyond the category of Fréchet algebra valued meromorphic functions to treat the inverse of $z+\frac{1}{i\hbar}u\circ v$, as $e_*^{t\frac{1}{i\hbar}u\circ v}$ has discrete singular points in general ordered expressions. The regularized product $(\zeta-X)\tilde{*}(\zeta-X)^{-1}$ seems to be a good method to treat singularities.

6.2. Basic properties of the inverse of $z+\frac{1}{i\hbar}u\circ v$. — We first study basic properties of the inverse of $z+\frac{1}{i\hbar}u\circ v$.

By the results of §4, the integrals

$$:\int_{-\infty}^{0}e^{tz}e_*^{t\frac{1}{i\hbar}u\circ v}dt:_0=\int_{-\infty}^{0}\frac{e^{tz}}{\cosh\frac{1}{2}t}e^{\frac{1}{i\hbar}2uv\tanh\frac{1}{2}t}dt,\quad \operatorname{Re}z>-\frac{1}{2},\tag{45}$$

$$:-\int_{0}^{\infty}e^{tz}e_*^{t\frac{1}{i\hbar}u\circ v}dt:_0=-\int_{0}^{\infty}\frac{e^{tz}}{\cosh\frac{1}{2}t}e^{\frac{1}{i\hbar}2uv\tanh\frac{1}{2}t}dt,\quad \operatorname{Re}z<\frac{1}{2}\tag{46}$$

converge in the Weyl ordered expression.

One can analyze the r.h.s. of (45) and (46) more closely via a change of variables as in Proposition 4.1. For $-\frac{1}{2} < \mathrm{Re}\, z \leq 0$, the change of variables $\tanh \frac{1}{2}t = \cos s$ transforms the r.h.s of (45) into

$$2\int_{-\pi}^{0} (\frac{1+\cos s}{1-\cos s})^{z} e^{(\cos s)\frac{1}{i\hbar}2uv} ds.$$

For $0 \leq \mathrm{Re}\, z < \frac{1}{2}$ and for $-\cos s = \tanh \frac{t}{2}$, $2\sin s ds = \sin^2 s dt$, the r.h.s. of (46) transforms into

$$2\int_{0}^{\pi} (\frac{1+\cos s}{1-\cos s})^{-z} e^{(\cos s)\frac{1}{i\hbar}2uv} ds.$$

Hence, Lemmas 2.2, 3.1 give that $\int_{-\infty}^{\infty} e_*^{t(z+\frac{1}{i\hbar}u\circ v)} dt$ is an element of $Hol(\mathbb{C}^2)$ in generic ordered expressions. Thus, both (45) and (46) give inverses of $z + \frac{1}{i\hbar}u \circ v$ for generic ordered expressions, which will be denoted by $(z+\frac{1}{i\hbar}u\circ v)_{*+}^{-1}$, $(z+\frac{1}{i\hbar}u\circ v)_{*-}^{-1}$, respectively.

The following may be viewed as a Sato hyperfunction:

Proposition 6.1. — *If $-\frac{1}{2} < \mathrm{Re}\, z < \frac{1}{2}$, then the difference of the two inverses is given by*

$$(z+\frac{1}{i\hbar}u\circ v)_{*+}^{-1} - (z+\frac{1}{i\hbar}u\circ v)_{*-}^{-1} = \int_{-\infty}^{\infty} e_*^{t(z+\frac{1}{i\hbar}u\circ v)} dt. \tag{47}$$

The difference is holomorphic in this strip for generic ordered expressions.

An elementary change of variables gives

$$((-z)+\frac{1}{i\hbar}u\circ v)_{*-}^{-1} = -\int_{0}^{\infty} e_*^{-t(z-\frac{1}{i\hbar}u\circ v)} dt = -\int_{-\infty}^{0} e_*^{(z-\frac{1}{i\hbar}u\circ v)} dt.$$

Thus, for generic ordered expressions, we see that

$$(z-\frac{1}{i\hbar}u\circ v)_{*-}^{-1} = -((-z)+\frac{1}{i\hbar}u\circ v)_{*-}^{-1}. \tag{48}$$

This is holomorphic on the domain $\mathrm{Re}\, z > -\frac{1}{2}$, on which $(z+\frac{1}{i\hbar}u\circ v)_{*+}^{-1}$ is also holomorphic. All of these results are easily proved for the Weyl ordered expression. However, for generic K-ordered expression, $:e_*^{t\frac{1}{i\hbar}u\circ v}:_K$ is rapidly decreasing in t, and the same computation gives the following:

Proposition 6.2. — *For generic ordered expressions, $(z+\frac{1}{i\hbar}u\circ v)_{*+}^{-1}$ and $(z-\frac{1}{i\hbar}u\circ v)_{*-}^{-1}$ are defined for $\mathrm{Re}\, z > -\frac{1}{2}$.*

The product $(z+\frac{1}{i\hbar}u\circ v)_{*+}^{-1} * (w+\frac{1}{i\hbar}u\circ v)_{*+}^{-1}$ is naturally defined for $z, w \notin -(\mathbb{N}+\frac{1}{2})$ by the usual resolvent identity. $\{(z+\frac{1}{i\hbar}u\circ v)_{*+}^{-1}; z \notin -(\mathbb{N}+\frac{1}{2})\}$ forms an associative algebra. in $\mathcal{E}_{2+}(\mathbb{C}^{2m})$.

Note that $(z+\frac{1}{i\hbar}u\circ v)^{-1}_{*+}*(-z-\frac{1}{i\hbar}u\circ v)^{-1}_{*-}$ diverges for any ordered expression. However, the standard resolvent formula gives the following:

Proposition 6.3. — *If $z+w\neq 0$, then*

$$\frac{1}{z+w}\Big((z+\frac{1}{i\hbar}u\circ v)^{-1}_{*+}+(w-\frac{1}{i\hbar}u\circ v)^{-1}_{*-}\Big)$$

is an inverse of $(z+\frac{1}{i\hbar}u\circ v)*(w-\frac{1}{i\hbar}u\circ v)$. In particular, for every positive integer n, and for every complex number z such that $\mathrm{Re}\,z>-(n+\frac{1}{2})$,

$$\frac{1}{2n}\big((1+\frac{1}{n}(z+\frac{1}{i\hbar}u\circ v))^{-1}_{*+}+(1-\frac{1}{n}(z+\frac{1}{i\hbar}u\circ v))^{-1}_{*-}\big)$$

is an inverse of $1-\frac{1}{n^2}(z+\frac{1}{i\hbar}u\circ v)^2_*$ for generic ordered expressions.

6.3. Analytic continuation of inverses. — Recall that $(z\pm\frac{1}{i\hbar}u\circ v)^{-1}_{\pm *}$ is holomorphic on the domain $\mathrm{Re}\,z>-\frac{1}{2}$ for generic ordered expressions. It is natural to expect that $(z\pm\frac{1}{i\hbar}u\circ v)^{-1}_{\pm *}=C(C(z\pm\frac{1}{i\hbar}u\circ v))^{-1}_{\pm *}$ for any non-zero constant C. To confirm this, we set $C=e^{i\theta}$ and consider the θ-derivative of

$$e^{i\theta}\int_{-\infty}^{0}e_*^{e^{i\theta}t(z\pm\frac{1}{i\hbar}u\circ v)}dt.$$

In generic K-ordered expressions, the phase part of the integrand is bounded in t and the amplitude is given by

$$\frac{2e^{i\theta}tz}{(1-\kappa)e^{e^{i\theta}t/2}+(1+\kappa)e^{-e^{i\theta}t/2}},\quad \kappa\neq 1.$$

The integral converges whenever $\mathrm{Re}\,e^{i\theta}(z\pm\frac{1}{2})>0$, and by integration by parts this convergence is independent of θ. It follows that $(z\pm\frac{1}{i\hbar}u\circ v)^{-1}_{\pm *}$ is holomorphic on the domain $\mathbb{C}-\{z;-\infty<z<-\frac{1}{2}\}$.

Next, it is natural to expect that the bumping identity $(u\circ v)*v=v*(u\circ v-i\hbar)$ gives the following "sliding identities"

$$v^{-1}_{*+}*(z+\frac{1}{i\hbar}u\circ v)^{-1}_{*+}*v=(z-1+\frac{1}{i\hbar}u\circ v)^{-1}_{+*},\quad v^{-1}_{*+}*(z-\frac{1}{i\hbar}u\circ v)^{-1}_{*-}*v=(z+1-\frac{1}{i\hbar}u\circ v)^{-1}_{*-}$$

whenever the inverse of v exists in a particular ordered expression. In this section, analytic continuation will be produced via these sliding identities.

However, the existence of v^{-1}_{*+} is not a generic property. As a result, instead of using v^{-1}_{*+} we will apply the sliding identity to the left inverse v° of v given below.

Remark There is a K-ordered expression such that $:\int_{-\infty}^{0}e_*^{tv}dt:_K$ converges to give an inverse of $:v^{-1}_{*+}:_K$ of v (cf. [17]), but it is easy to see that $:v^{-1}_{*+}*\varpi_{00}:_K$ diverges.

First, we remark that the formula in Proposition 6.1 gives

$$(u*v)^{-1}_{*-}=-\frac{1}{i\hbar}\int_0^{\infty}e_*^{t\frac{1}{i\hbar}u*v}dt,\quad (v*u)^{-1}_{*+}=\frac{1}{i\hbar}\int_{-\infty}^{0}e_*^{t\frac{1}{i\hbar}v*u}dt$$

for generic ordered expressions. Then

$$v^\circ = u*(v*u)^{-1}_{*+}, \quad u^\bullet = v*(u*v)^{-1}_{*-},$$

are left and right inverses of v and u respectively. That is,

$$v*v^\circ = 1, \quad v^\circ * v = 1-\varpi_{00}, \quad u*u^\bullet = 1, \quad u^\bullet * u = 1-\varpi_{00}.$$

The bumping identity gives

$$v*(z+\frac{1}{i\hbar}u\circ v)*v^\circ = z+1+\frac{1}{i\hbar}u\circ v, \quad v^\circ*(z+\frac{1}{i\hbar}u\circ v)*v = (1-\varpi_{00})*(z-1+\frac{1}{i\hbar}u\circ v).$$

Successive applications of the bumping identity give the following useful formula:

$$(u*(v*u)^{-1}_{*+})^n * \varpi_{00} = \frac{1}{n!}(\frac{1}{i\hbar}u)^n * \varpi_{00}. \tag{49}$$

Using v° instead of v^{-1}_{*+}, we can produce the analytic continuation of inverses. However, we have to be careful about the continuity of the $*$-product. We compute

$$\begin{aligned} v^\circ * (z+\frac{1}{i\hbar}u\circ v)^{-1}_{*+} &= \Big(u*\int_{-\infty}^0 e_*^{t(\frac{1}{i\hbar}u\circ v+\frac12)}dt\Big)*\int_{-\infty}^0 e_*^{s(z+\frac{1}{i\hbar}u\circ v)}ds \\ &= u*\int_{-\infty}^0\int_{-\infty}^0 e_*^{t(\frac{1}{i\hbar}u\circ v+\frac12)} * e_*^{s(z+\frac{1}{i\hbar}u\circ v)}dtds \quad (\text{cf.}\,(39)) \\ &= \int_{-\infty}^0\int_{-\infty}^0 e^{t\frac12+sz}u*e_*^{(t+s)\frac{1}{i\hbar}u\circ v}dtds \quad (\text{cf. Lemma 2.1}) \\ &= \int_{-\infty}^0\int_{-\infty}^0 e^{t\frac12+sz-(t+s)}e_*^{(t+s)\frac{1}{i\hbar}u\circ v}*udtds. \end{aligned}$$

Hence, whenever both sides are defined, we obtain

$$\begin{aligned} (v^\circ * (z+\frac{1}{i\hbar}u\circ v)^{-1}_{*+})*v &= \int_{-\infty}^0\int_{-\infty}^0 e^{-t\frac12+s(z-1)}e_*^{(t+s)\frac{1}{i\hbar}u\circ v}*(u*v)dtds \\ &= \int_{-\infty}^0 (u*v)*e_*^{t\frac{1}{i\hbar}u*v}dt*\int_{-\infty}^0 e_*^{s(z-1+\frac{1}{i\hbar}u\circ v)}ds \\ &=(1-\varpi_{00})*(z-1+\frac{1}{i\hbar}u\circ v)^{-1}_{*+}. \end{aligned}$$

Noting that

$$\varpi_{00}*(z-1+\frac{1}{i\hbar}u\circ v)^{-1}_{*+} = (z-1+\frac{1}{i\hbar}u\circ v)^{-1}_{*+}*\varpi_{00} = (z-\frac12)^{-1}\varpi_{00},$$

whenever $(z-1+\frac{1}{i\hbar}u\circ v)^{-1}_{*+}$ is defined, we have

$$\big(v^\circ * (z+\frac{1}{i\hbar}u\circ v)^{-1}_{*+}\big)*v + (z-\frac12)^{-1}\varpi_{00} = \big(z-1+\frac{1}{i\hbar}u\circ v\big)^{-1}_{*+}. \tag{50}$$

Since $(z-\frac12)^{-1}\varpi_{00}$ is always defined, we see that the functional equation (50) gives an analytic continuation for $(z+\frac{1}{i\hbar}u\circ v)^{-1}_{*+}$. Namely, we have the following (see [**7**] and [**10**] for more details):

Theorem 6.2. — *For generic ordered expressions, the inverses* $(z+\frac{1}{i\hbar}u\circ v)^{-1}_{*+}$, $(z-\frac{1}{i\hbar}u\circ v)^{-1}_{*-}$ *extend to* $\mathcal{E}_{2+}(\mathbb{C}^2)$*-valued holomorphic functions of* z *on* $\mathbb{C}-\{-(\mathbb{N}+\frac{1}{2})\}$.

In particular, $(z^2-(\frac{1}{i\hbar}u\circ v)^2)^{-1}_{\pm *}$ *extends to a holomorphic function of* z *on this domain.*

The residue at a singular point z_0 is defined as usual by $\frac{1}{2\pi i}\int_{C_{z_0}}(z+\frac{1}{i\hbar}u\circ v)^{-1}_{*\pm}dz$. The analytic continuation formula gives the following:

Theorem 6.3. — $\mathrm{Res}((z+\frac{1}{i\hbar}u\circ v)^{-1}_{*+}, -(n+\frac{1}{2}))=\frac{1}{(i\hbar)^n n!}u^n * \varpi_{00} * v^n$ *for generic ordered expressions.*

For the proof, we remark that $(z+n+\frac{1}{i\hbar}u\circ v)^{-1}_{*\pm}$ is holomorphic for sufficiently large n, and the contour integral is an integral on a compact set. Note that $(z+\frac{1}{i\hbar}u\circ v)^{-1}_{*+}$ is singular at $z=n+\frac{1}{2}$, but $(z+\frac{1}{i\hbar}u\circ v)*(z+\frac{1}{i\hbar}u\circ v)^{-1}_{*+}=1$ for $z\notin -(\mathbb{N}+\frac{1}{2})$ for generic ordered expressions.

Note also that if we exchange $(z+\frac{1}{i\hbar}u\circ v)*$ and the integration, then

$$\int_{-\infty}^{0}(z+\frac{1}{i\hbar}u\circ v)*e_*^{t(z+\frac{1}{i\hbar}u\circ v)}dt=\begin{cases}1 & \mathrm{Re}\,z>-\frac{1}{2}\\ 1-\varpi_{00} & z=-\frac{1}{2}\end{cases},$$

$$\int_{-\infty}^{0}(z-\frac{1}{i\hbar}u\circ v)*e_*^{t(z-\frac{1}{i\hbar}u\circ v)}dt=\begin{cases}1 & \mathrm{Re}\,z>-\frac{1}{2}\\ 1-\overline{\varpi}_{00} & z=-\frac{1}{2}\end{cases}.$$

As suggested by these formulas and Hadamard's technique of extracting finite parts of divergent integrals, we now extend the definition of the $*$-product using the finite part regularization mentioned in the introduction.

We consider the inductive limit topology on the space $\mathbb{C}[\boldsymbol{u}]$. We define the new product of $(z+\frac{1}{i\hbar}u\circ v)^{-1}_{*\pm}$ with either polynomials $q(u,v)$ or $q(u,v)=e_*^{s\frac{1}{i\hbar}u\circ v}$ by

$$(z+\frac{1}{i\hbar}u\circ v)^{-1}_{*\pm}\tilde{*}q(u,v)=(\mathrm{FP}(z+\frac{1}{i\hbar}u\circ v)^{-1}_{*\pm})*q(u,v), \tag{51}$$

where $\mathrm{FP}f(z)$ denotes its finite part of f at z. The result may not be continuous in z.

For $\mathrm{Re}\,z>-\frac{1}{2}$ we easily see that

$$(z+\frac{1}{i\hbar}u\circ v)^{-1}_{*+}\tilde{*}q(u,v)=\lim_{N\to\infty}\int_{-N}^{0}e_*^{t(z+\frac{1}{i\hbar}u\circ v)}*q(u,v)dt$$

Hence we have the formula

$$(z+\frac{1}{i\hbar}u\circ v)\tilde{*}(z+\frac{1}{i\hbar}u\circ v)^{-1}_{*+}=\begin{cases}1 & \mathrm{Re}\,z>-\frac{1}{2}\\ 1-\varpi_{00} & z=-\frac{1}{2}\end{cases}. \tag{52}$$

Using $(v^\circ)^n*(z+\frac{1}{i\hbar}u\circ v)*(z+\frac{1}{i\hbar}u\circ v)^{-1}_{*+}*v^n=(v^\circ)^n*(z+\frac{1}{i\hbar}u\circ v)*v^n*(v^\circ)^n*(z+\frac{1}{i\hbar}u\circ v)^{-1}_{*+}*v^n$ and (50), we have the following:

Theorem 6.4. — *Using definition* (51) *for the* $\tilde{*}$*-product, we have*

$$(53)\quad (z+\frac{1}{i\hbar}u\circ v)\tilde{*}(z+\frac{1}{i\hbar}u\circ v)_{*+}^{-1}=\begin{cases}1 & z\notin -(\mathbb{N}+\frac{1}{2})\\ 1-\frac{1}{n!}(\frac{1}{i\hbar}u)^n*\varpi_{00}*v^n & z=-(n+\frac{1}{2})\end{cases},$$

$$(54)\quad (z-\frac{1}{i\hbar}u\circ v)\tilde{*}(z-\frac{1}{i\hbar}u\circ v)_{*-}^{-1}=\begin{cases}1 & z\notin -(\mathbb{N}+\frac{1}{2})\\ 1-\frac{1}{n!}(\frac{1}{i\hbar}v)^n*\overline{\varpi}_{00}*u^n & z=-(n+\frac{1}{2})\end{cases}.$$

for generic ordered expressions.

Although $z=-(n+\frac{1}{2}),\ n=0,1,2,\cdots$ are all removable singularities for (53) and (54) as a function of z, it is better to retain these singular points.

In these computations, elements are often given via a limiting procedure. As usual, $*$-products of such elements depend delicately on the limiting procedure. There is no general rule guaranteeing associativity.

Via the identity $(1+\frac{1}{m}(z+\frac{1}{i\hbar}u\circ v))_{*+}^{-1}=m(m+z+\frac{1}{i\hbar}u\circ v)_{*+}^{-1}$, we have, in particular

(55)

$$(1+\tfrac{1}{m}(z+\tfrac{1}{i\hbar}u\circ v))\tilde{*}(1+\tfrac{1}{m}(z+\tfrac{1}{i\hbar}u\circ v))_{*+}^{-1}=\begin{cases}1 & z\notin -(\mathbb{N}+m+\frac{1}{2})\\ 1-\frac{1}{k!}(\frac{1}{i\hbar}u)^k*\varpi_{00}*v^k & z=-(k+m+\frac{1}{2})\end{cases}$$

for every fixed positive integer m and for arbitrary $k\in\mathbb{N}$ for generic ordered expressions.

By the associativity stated in Lemma 2.1, we see the following:

Theorem 6.5. — *We have*

$$(-n-\tfrac{1}{2}+\tfrac{1}{i\hbar}u\circ v)*u^n*\varpi_{00}=u^n*(\tfrac{1}{i\hbar}u*v)\varpi_{00}=0.$$

And thus,

$$(1-\frac{1}{\ell}(z+\tfrac{1}{i\hbar}u\circ v))*\Big((1+\tfrac{1}{m}(z+\tfrac{1}{i\hbar}u\circ v))\tilde{*}((1+\tfrac{1}{m}(z+\tfrac{1}{i\hbar}u\circ v))_{*+}^{-1}\Big)$$
$$=\begin{cases}1-\frac{1}{\ell}(z+\frac{1}{i\hbar}u\circ v) & z\notin -(\mathbb{N}+m+\frac{1}{2})\\ 1-\frac{1}{\ell}(z+\frac{1}{i\hbar}u\circ v) & z=-(\ell+\frac{1}{2})\\ (1-\frac{1}{\ell}(z+\frac{1}{i\hbar}u\circ v))*(1-\frac{1}{k!}(\frac{1}{i\hbar}u)^k*\varpi_{00}*v^k) & z=-(k+\frac{1}{2}),\ z\neq-(\ell+\frac{1}{2}).\end{cases}$$

for generic ordered expressions.

We note here that singularities such as $\frac{1}{\ell!}(\frac{1}{i\hbar}u)^\ell*\varpi_{00}*v^\ell$ disappear from the r.h.s. of the above equality because of the term $1-\frac{1}{\ell}(z+\frac{1}{i\hbar}u\circ v)$.

We also define a $\tilde{*}$-product for a certain class of elements by

$$f(\boldsymbol{u})\tilde{*}(z+\frac{1}{i\hbar}u\circ v)_{*+}^{-1}=f(\boldsymbol{u})*\big(\mathrm{FP}(z+\frac{1}{i\hbar}u\circ v)_{*+}^{-1}\big).$$

These formulas will be applied to the computation of

$$\sin_*(z + \frac{1}{i\hbar} u \circ v) \tilde{*} (1 + \frac{1}{m}(z + \frac{1}{i\hbar} u \circ v))_{*+}^{-1}$$

along with an infinite product formula for $\sin_*(z + \frac{1}{i\hbar} u \circ v)$ in a forthcoming paper.

References

[1] G. S. AGARWAL & E. WOLF – "Calculus for functions of noncommuting operators and general phase-space methods in quantum mechanics. I. Mapping theorems and ordering of functions of noncommuting operators", *Phys. Rev. D* **2** (1970), p. 2161–2186.

[2] G. E. ANDREWS, R. ASKEY & R. ROY – *Special functions*, Encyclopedia of Mathematics and its Applications, vol. 71, Cambridge University Press, 1999.

[3] F. BAYEN, M. FLATO, C. FRONSDAL, A. LICHNEROWICZ & D. STERNHEIMER – "Deformation theory and quantization. I. Deformations of symplectic structures", *Ann. Physics* **111** (1978), p. 61–110.

[4] I. M. GEL'FAND & G. E. SHILOV – *Generalized functions. Vol. 2. Spaces of fundamental and generalized functions*, Translated from the Russian by Morris D. Friedman, Amiel Feinstein and Christian P. Peltzer, Academic Press, 1968.

[5] V. GUILLEMIN & S. STERNBERG – *Geometric asymptotics*, Amer. Math. Soc., 1977, Mathematical Surveys, No. 14.

[6] Y. MAEDA, N. MIYAZAKI, H. OMORI & A. YOSHIOKA – "Star exponential functions as two-valued elements", in *The breadth of symplectic and Poisson geometry*, Progr. Math., vol. 232, Birkhäuser, 2005, p. 483–492.

[7] H. OMORI – *数学のなかの物理学 (Physics in Mathematics)*, Tokyo Univ. Publ., 2004.

[8] ______, "Toward geometric quantum theory", in *From geometry to quantum mechanics*, Progr. Math., vol. 252, Birkhäuser, 2007, p. 213–251.

[9] H. OMORI & T. KOBAYASHI – "Singular star-exponential functions", *SUT J. Math.* **37** (2001), p. 137–152.

[10] H. OMORI & Y. MAEDA – *量子的な微分・積分 (Quantum Theoretic Calculus)*, Springer, Tokyo, 2004.

[11] H. OMORI, Y. MAEDA, N. MIYAZAKI & A. YOSHIOKA – "Deformation quantization of Fréchet-Poisson algebras: convergence of the Moyal product", in *Conférence Moshé Flato 1999, Vol. II (Dijon)*, Math. Phys. Stud., vol. 22, Kluwer Acad. Publ., 2000, p. 233–245.

[12] ______, "Strange phenomena related to ordering problems in quantizations", *J. Lie Theory* **13** (2003), p. 481–510.

[13] ______, "Expressions of algebra elements and transcendental noncommutative calculus", in *Noncommutative geometry and physics 2005*, World Sci. Publ., Hackensack, NJ, 2007, p. 3–30.

[14] ______, "Geometric objects in an approach to quantum geometry", in *From geometry to quantum mechanics*, Progr. Math., vol. 252, Birkhäuser, 2007, p. 303–324.

[15] ______, "Non-formal deformation quantization of Fréchet-Poisson algebras: the Heisenberg and Lie algebra case", in *Geometric and topological methods for quantum field theory*, Contemp. Math., vol. 434, Amer. Math. Soc., 2007, p. 99–123.

[16] ______, "Noncommutative Minkowski space and transcendental calculus", *Progress of Theoretical Physics Suppl.* **171** (2007), p. 184–195.

[17] ______, "Orderings and non-formal deformation quantization", *Lett. Math. Phys.* **82** (2007), p. 153–175.

[18] M. A. RIEFFEL – "Noncommutative tori—a case study of noncommutative differentiable manifolds", in *Geometric and topological invariants of elliptic operators (Brunswick, ME, 1988)*, Contemp. Math., vol. 105, Amer. Math. Soc., 1990, p. 191–211.

[19] L. SCHWARZ – *Theory of distributions*, Academic Press, 1966.

[20] S. L. WORONOWICZ – "Differential calculus on compact matrix pseudogroups (quantum groups)", *Comm. Math. Phys.* **122** (1989), p. 125–170.

H. OMORI, Department of Mathematics, Faculty of Science and Technology, Tokyo University of Science, Noda, Chiba, 278-8510, Japan • *E-mail :* omori@ma.noda.tus.ac.jp

Y. MAEDA, Department of Mathematics, Faculty of Science and Technology, Keio University, Hiyoshi, Yokohama, 223-8825, Japan; Mathematical Research Centre University of Warwick Coventry CV4 7AL, United Kingdom • *E-mail :* maeda@math.keio.ac.jp

N. MIYAZAKI, Department of Mathematics, Faculty of Economics, Keio University, Hiyoshi, Yokohama, 223-8521, Japan • *E-mail :* miyazaki@math.hc.keio.ac.jp

A. YOSHIOKA, Department of Mathematics, Faculty of Science, Tokyo University of Science, Kagurazaka, Tokyo, 102-8601, Japan • *E-mail :* yoshioka@rs.kagu.tus.ac.jp