348

ASTÉRISQUE

2012

SÉMINAIRE BOURBAKI
VOLUME 2010/2011
EXPOSÉS 1027-1042

(1033) *Sparse quadratic forms and their geometric applications*

Assaf NAOR

SOCIÉTÉ MATHÉMATIQUE DE FRANCE
Publié avec le concours du CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Séminaire BOURBAKI Janvier 2011
63e année, 2010-2011, n° 1033, p. 189 à 217

SPARSE QUADRATIC FORMS AND THEIR GEOMETRIC APPLICATIONS [following Batson, Spielman and Srivastava]

by **Assaf NAOR**

1. INTRODUCTION

In what follows all matrices are assumed to have real entries, and square matrices are always assumed to be symmetric unless stated otherwise. The support of a $k \times n$ matrix $A = (a_{ij})$ will be denoted below by

$$\mathrm{supp}(A) = \big\{(i,j) \in \{1,\dots,k\} \times \{1,\dots,n\} :\ a_{ij} \neq 0\big\}.$$

If A is an $n \times n$ matrix, we denote the decreasing rearrangement of its eigenvalues by

$$\lambda_1(A) \geqslant \lambda_2(A) \geqslant \cdots \geqslant \lambda_n(A).$$

$\mathbb{R}^n$ will always be assumed to be equipped with the standard scalar product $\langle \cdot,\cdot \rangle$. Given a vector $v \in \mathbb{R}^n$ and $i \in \{1,\dots,n\}$, we denote by v_i the ith coordinate of v. Thus for $u, v \in \mathbb{R}^n$ we have $\langle u, v\rangle = \sum_{i=1}^n u_i v_i$.

Our goal here is to describe the following theorem of Batson, Spielman and Srivastava [**5**], and to explain some of its recently discovered geometric applications. We expect that there exist many more applications of this fundamental fact in matrix theory.

Theorem 1.1. — *For every $\varepsilon \in (0,1)$ there exists $c(\varepsilon) = O(1/\varepsilon^2)$ with the following properties. Let $G = (g_{ij})$ be an $n \times n$ matrix with nonnegative entries. Then there exists an $n \times n$ matrix $H = (h_{ij})$ with nonnegative entries that satisfies the following conditions:*

1. $\mathrm{supp}(H) \subseteq \mathrm{supp}(G)$.
2. *The cardinality of the support of H satisfies $|\mathrm{supp}(H)| \leqslant c(\varepsilon) n$.*

(*) Supported in part by NSF grant CCF-0635078, BSF grant 2006009, and the Packard Foundation.

3. *For every $x \in \mathbb{R}^n$ we have*

$$\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}(x_i - x_j)^2 \leqslant \sum_{i=1}^{n}\sum_{j=1}^{n} h_{ij}(x_i - x_j)^2 \leqslant (1+\varepsilon)\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}(x_i - x_j)^2. \tag{1}$$

The second assertion of Theorem 1.1 is that the matrix H is *sparse*, yet due to the third assertion of Theorem 1.1 the quadratic form $\sum_{i=1}^{n}\sum_{j=1}^{n} h_{ij}(x_i - x_j)^2$ is nevertheless a good approximation of the quadratic form $\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}(x_i - x_j)^2$. For this reason Theorem 1.1 is called in the literature a *sparsification theorem.*

The bound on $|\mathrm{supp}(H)|$ obtained in [**5**] is

$$|\mathrm{supp}(H)| \leqslant 2\left\lceil \frac{(\sqrt{1+\varepsilon}+1)^4}{\varepsilon^2} n \right\rceil. \tag{2}$$

Thus $c(\varepsilon) \leqslant 32/\varepsilon^2 + O(1/\varepsilon)$. There is no reason to expect that (2) is best possible, but a simple argument [**5**, Section 4] shows that necessarily $c(\varepsilon) \geqslant 8/\varepsilon^2$.

1.1. Historical discussion

The sparsification problem that is solved (up to constant factors) by Theorem 1.1 has been studied for some time in the theoretical computer science literature. The motivations for these investigations were algorithmic, and therefore there was emphasis on constructing the matrix H quickly. We will focus here on geometric applications of Theorem 1.1 for which the existential statement suffices, but we do wish to state that [**5**] shows that H can be constructed in time $O(n^3|\mathrm{supp}(G)|/\varepsilon^2) = O(n^5/\varepsilon^2)$. For certain algorithmic applications this running time is too slow, and the literature contains works that yield weaker asymptotic bounds on $|\mathrm{supp}(H)|$ but have a faster construction time. While such tradeoffs are important variants of Theorem 1.1, they are not directly relevant to our discussion and we will not explain them here. For the applications described below, even a weaker bound of, say, $|\mathrm{supp}(H)| \leqslant c(\varepsilon) n \log n$ is insufficient.

Benczúr and Karger [**6**] were the first to study the sparsification problem. They proved the existence of a matrix H with $|\mathrm{supp}(H)| \leqslant c(\varepsilon) n \log n$, that satisfies the conclusion (1) only for *Boolean* vectors $x \in \{0,1\}^n$. In their series of works on fast solvers for certain linear systems [**43, 46, 45, 44**], Spielman and Teng studied the sparsification problem as stated in Theorem 1.1, i.e., with the conclusion (1) holding for *every* $x \in \mathbb{R}^n$. Specifically, in [**44**], Spielman and Teng proved Theorem 1.1 with the weaker estimate $|\mathrm{supp}(H)| = O\left(n(\log n)^7/\varepsilon^2\right)$. Spielman and Srivastava [**41**] improved this estimate on the size of the support of H to $|\mathrm{supp}(H)| = O(n(\log n)/\varepsilon^2)$. As we stated above, Theorem 1.1, which answers positively a conjecture of Spielman-Srivastava [**41**], is due to Batson-Spielman-Srivastava [**5**], who proved this sharp result via a new deterministic iterative technique (unlike the previous probabilistic arguments) that we will describe below. This beautiful new approach does not only yield an asymptotically sharp bound on $|\mathrm{supp}(H)|$: it gives for the first time a deterministic algorithm for constructing H (unlike the previous randomized algorithms), and it also gives additional results that will be described later. We refer to Srivastava's dissertation [**48**] for a very nice and more complete exposition of these ideas. See also the work of Kolla-Makarychev-Saberi-Teng [**23**] for additional results along these lines.

1.2. Combinatorial interpretation

Suppose that G is the adjacency matrix of the complete graph, i.e., the diagonal entries of G vanish and $g_{ij} = 1$ if $i \neq j$. Assume also that the matrix H of Theorem 1.1 happens to be a multiple of the adjacency matrix of a d-regular graph $\Gamma = (\{1, \dots, n\}, E)$, i.e., for some $\gamma > 0$ and all $i, j \in \{1, \dots, n\}$ we have $h_{ij} = \gamma$ if $\{i, j\} \in E$ and $h_{ij} = 0$ otherwise. Thus $|\mathrm{supp}(H)| = dn$. By expanding the squares in (1) and some straightforward linear algebra, we see that (1) is equivalent to the bound $(\lambda_1(H) - \lambda_n(H))/(\lambda_1(H) - \lambda_2(H)) \leqslant 1 + \varepsilon$. Thus if ε is small then the graph Γ is a good expander (see [**18**] for background on this topic). The Alon-Boppana bound [**30**] implies that H satisfies $(\lambda_1(H) - \lambda_n(H))/(\lambda_1(H) - \lambda_2(H)) \geqslant 1 + 4(1 - o(1))\sqrt{d}$ as $n, d \to \infty$. This lower bound can be asymptotically attained since if Γ is a Ramanujan graph of Lubotzky-Phillips-Sarnak [**24**] then $\lambda_1(H)/\gamma, \lambda_n(H)/\gamma \in \left[-2\sqrt{d-1}, 2\sqrt{d-1}\right]$. Writing $1 + \varepsilon = \left(d + 2\sqrt{d-1}\right) / \left(d - 2\sqrt{d-1}\right) = 1 + 4(1 + o(1))/\sqrt{d}$, we see that the existence of Ramanujan graphs means that (in this special case of the complete graph) there exists a matrix H satisfying (1) with $|\mathrm{supp}(H)| = dn = 16n(1 + o(1))/\varepsilon^2$. The bound on $|\mathrm{supp}(H)|$ in (2) shows that Thereom 1.1 achieves the optimal Ramanujan bound up to a factor of 2. For this reason Batson-Spielman-Srivastava call the matrices produced by Theorem 1.1 "twice-Ramanujan sparsifiers". Of course, this analogy is incomplete since while the matrix H is sparse, it need not be a multiple of the adjacency matrix of a graph, but rather an adjacency matrix of a weighted graph. Moreover, this graph has bounded average degree, rather than being a regular graph of bounded degree. Such weighted sparse (though non-regular) graphs still have useful pseudorandom properties (see [**5**, Lemma 4.1]). Theorem 1.1 can be therefore viewed as a new deterministic construction of "expander-like" weighted graphs, with very good spectral gap. Moreover, it extends the notion of expander graphs since one can start with an arbitrary matrix G before applying the sparsification procedure, with the quality of the resulting expander (measured in terms of absolute spectral gap) being essentially the same as the quality of G as an expander.

1.3. Structure of this paper

In Section 2 we state a stronger theorem (Theorem 2.1) of Batson-Spielman-Srivastava [**5**], and prove that it implies Theorem 1.1. Section 3 contains the Batson-Spielman-Srivastava proof of this theorem, which is based on a highly original iterative argument. Section 4 contains an application of Theorem 2.1, due to Srivastava [**47**], to approximate John decompositions. In section 5 we describe two applications of Theorem 2.1, due to Newman-Rabinovich [**29**] and Schechtman [**38**], to dimensionality reduction problems. Section 6 describes the work of Spielman-Srivastava [**42**] that shows how their proof technique for Theorem 2.1 can be used to prove a sharper version of the Bourgain-Tzafriri restricted invertibility principle. Section 7 contains concluding comments and some open problems.

2. A STRONGER THEOREM

Batson-Spielman-Srivastava actually proved a stronger theorem that implies Theorem 1.1. The statement below is not identical to the statement in [**5**], though it easily follows from it. This formulation is stated explicitly as Theorem 1.6 in Srivastava's dissertation [**48**].

THEOREM 2.1. — *Fix $\varepsilon \in (0,1)$ and $m,n \in \mathbb{N}$. For every $x_1,\ldots,x_m \in \mathbb{R}^n$ there exist $s_1,\ldots,s_m \in [0,\infty)$ such that*

$$\left|\left\{i \in \{1,\ldots,m\} : s_i \neq 0\right\}\right| \leqslant \left\lceil \frac{n}{\varepsilon^2} \right\rceil, \tag{3}$$

and for all $y \in \mathbb{R}^n$ we have

$$(1-\varepsilon)^2 \sum_{i=1}^m \langle x_i, y\rangle^2 \leqslant \sum_{i=1}^m s_i \langle x_i, y\rangle^2 \leqslant (1+\varepsilon)^2 \sum_{i=1}^m \langle x_i, y\rangle^2. \tag{4}$$

2.1. Deduction of Theorem 1.1 from Theorem 2.1

Let $G = (g_{ij})$ be an $n \times n$ matrix with nonnegative entries. Note that the diagonal entries of G play no role in the conclusion of Theorem 1.1, so we may assume in what follows that $g_{ii} = 0$ for all $i \in \{1,\ldots,n\}$.