Séminaire BOURBAKI Octobre 2018
71e année, 2018–2019, no 1153, p. 109 à 147
doi:10.24033/ast.1132

NORMALITÉ ASYMPTOTIQUE DES VECTEURS PROPRES D'UN GRAPHE RÉGULIER ALÉATOIRE
[d'après Ágnes Backhausz et Balázs Szegedy]

par **Charles BORDENAVE**

INTRODUCTION

Nous présentons ici le résultat principal de Backhausz et Szegedy [**5**] et nous introduisons les définitions les plus importantes.

Graphe. — Un *graphe* $G = (V, E)$ est la paire formée par un ensemble dénombrable V et un ensemble E formé de parties à deux éléments de V. Les éléments de V et E sont appelés respectivement les *sommets* et les *arêtes* du graphe. Cette définition de graphe correspond aux graphes simples (ni boucles ni arêtes multiples). Le *degré* du sommet $x \in V$ est le nombre d'arêtes $e \in E$ telles que $x \in e$. Pour d entier, un graphe est dit *d-régulier* si tous ses sommets ont degré d.

Soient d, n des entiers positifs. Nous noterons $\mathbb{G}(n, d)$, l'ensemble des graphes $G = (V, E)$ d-réguliers tels que $V = \{1, \dots, n\}$. Si dn est pair et $2 \leq d \leq n - 1$ alors $\mathbb{G}(n, d)$ n'est pas vide. Si dn est impair ou $d \geq n$, $\mathbb{G}(n, d)$ est vide.

Soit $G = (V, E)$ un graphe. Pour k entier, $x, y \in V$, un *chemin de longueur k de x à y* est une suite $(x_0, \dots, x_k)$ telle que $x_0 = x$, $x_k = y$ et pour tout $t = 1, \dots, k$, $\{x_{t-1}, x_t\} \in E$. Le graphe G est *connexe* si pour tous $x, y \in V$ il existe un chemin de x à y. La *distance* entre x et y est la longueur du plus court chemin qui les relie. Un *cycle* est un chemin $(x_0, \dots, x_k)$ tel que $x_0 = x_k$ et les sommets $(x_1, \dots, x_k)$ sont tous distincts. Enfin, un *arbre* est un graphe connexe et sans cycle.

Spectre. — On peut associer plusieurs opérateurs à un graphe $G = (V, E)$. Le plus simple est *l'opérateur d'adjacence*, défini dans $\ell^2(V)$ par, pour tous $f \in \ell^2(V)$ et $x \in V$,

$$Af(x) = \sum_{y:\{x,y\} \in E} f(y),$$

où la somme porte sur les sommets y qui partagent une arête avec x. Si le degré des sommets est uniformément borné, l'opérateur A est un opérateur borné et auto-adjoint.

Si le graphe G est d-régulier, l'opérateur $d^{-1}A$ est le noyau de transition de la marche au hasard sur G. Si G est un graphe d-régulier et V est fini alors le vecteur constant $\mathbf{1} \in \ell^2(V)$ défini pour tout $x \in V$ par $\mathbf{1}(x) = 1$ est un vecteur propre de A associé à la valeur propre d.

Il existe de nombreuses relations entre la géométrie du graphe et les propriétés spectrales de son opérateur d'adjacence. Ces dernières années un effort de recherche tout particulier vise à décrire le spectre lorsque le graphe sous-jacent est aléatoire ou quasi-aléatoire. Pour ne citer que quelques références récentes relatives aux graphes réguliers, des outils de combinatoire [**19, 12**], de la théorie des matrices aléatoires [**17, 16, 7, 6, 22**], de l'analyse semi-classique [**14, 3, 2**] ont été utilisés dans ce contexte. Nous allons présenter ici une méthode très originale d'analyse du spectre développée dans [**5**]. Elle repose sur des liens fructueux entre la théorie de l'information et la convergence locale des graphes (sur les liens entre convergence locale des graphes et spectre, voir [**11**]). Le résultat principal de Backhausz et Szegedy portera sur les vecteurs propres de A orthogonaux à $\mathbf{1}$ lorsque G est distribué suivant la mesure uniforme sur $\mathbb{G}(n, d)$. Dans la suite de cette introduction, nous allons d'abord énoncer ce résultat sous sa forme la plus simple pour ensuite donner l'énoncé général qui exige une certaine préparation.

Soit $A \in M_n(\mathbb{R})$ une matrice symétrique de taille n. Pour tous $\delta \in [0, 1[$ et $\lambda \in \mathbb{R}$, un *quasi-vecteur propre de déviation δ et de valeur propre λ* est un vecteur non nul $f \in \mathbb{R}^n$ tel que $\|Af - \lambda f\|_2 \leq \delta \|f\|_2$ où $\|\cdot\|_2$ est la norme euclidienne. Un *quasi-vecteur propre de déviation δ* est un quasi-vecteur propre de déviation δ et de valeur propre λ pour un certain $\lambda \in \mathbb{R}$. Évidemment, un vecteur propre est un quasi-vecteur propre de déviation nulle.

Distribution d'un vecteur. — À tout vecteur $f \in \mathbb{R}^n$, la *mesure empirique de ses entrées* est la mesure de probabilité $\mathrm{distr}(f)$ sur $\mathbb{R}$:

$$\mathrm{distr}(f) = \frac{1}{n} \sum_{x=1}^{n} \delta_{f(x)},$$

où δ est la distribution de Dirac. En termes probabilistes, $\mathrm{distr}(f)$ est la loi de $f(o)$ où o est uniformément distribué sur $\{1, \dots, n\}$. La loi $\mathrm{distr}(f)$ est donc la *loi d'une entrée typique*. Le second moment de $\mathrm{distr}(f)$ est égal à $\|f\|_2^2/n$ de telle sorte que si f est non nul, $\mathrm{distr}(\sqrt{n} f / \|f\|_2)$ a un second moment égal à 1.

Gaussianité d'un vecteur. — Pour tout $\sigma \in \mathbb{R}_+$, on note traditionnellement $N(0, \sigma^2)$ la mesure de probabilité gaussienne sur $\mathbb{R}$ centrée et de variance σ^2. Pour $\sigma = 0$, on pose $N(0,0) = \delta_0$. La *proximité d'un vecteur* $f \in \mathbb{R}^n$ *non nul à un vecteur gaussien centré* est définie comme

$$D(f) = \inf_{\sigma \in [0,1]} \mathrm{d_L}\big(\mathrm{distr}(\sqrt{n} f/\|f\|_2), N(0,\sigma^2)\big),$$

où $\mathrm{d_L}$ est la *distance de Lévy-Prohorov* (toute autre distance qui définit la topologie faible conviendrait également). Rappelons que si μ et ν sont deux mesures de probabilités sur un espace métrique (M,d) alors, en notant $B(M)$ sa tribu borélienne, on a

$$\mathrm{d_L}(\mu,\nu) = \inf\{\varepsilon > 0 : \forall A \in B(M),\ \mu(A) \leq \nu(A^\varepsilon) + \varepsilon \text{ and } \nu(A) \leq \mu(A^\varepsilon) + \varepsilon\},$$

où pour tous $A \in B(M)$ et $\varepsilon > 0$, on note A^ε l'ensemble des éléments de M à distance au plus ε de A.

Considérons deux cas extrêmes de vecteurs $f \in \mathbb{R}^n$ « proches » d'un vecteur gaussien au sens que $D(f)$ est petit. Si f est un vecteur de la base canonique alors $\mathrm{distr}(\sqrt{n} f) = 1/n \cdot \delta_{\sqrt{n}} + (1 - 1/n) \cdot \delta_0$. Il est donc immédiat que lorsque n tend vers l'infini, $\mathrm{distr}(\sqrt{n} f)$ converge faiblement vers $\delta_0 = N(0,0)$. Si f est un vecteur aléatoire gaussien standard dans $\mathbb{R}^n$ (les coordonnées de f dans une base orthogonale sont des variables gaussiennes $N(0,1)$ indépendantes) alors la loi faible des grands nombres implique que lorsque n tend vers l'infini, la variable aléatoire $\mathrm{d_L}(\mathrm{distr}(f), N(0,1))$ converge en probabilité vers 0 et $\|f\|_2/\sqrt{n}$ vers 1. Dans ces deux cas, $D(f)$ tend vers 0 mais pour des raisons différentes. Dans le premier cas, la norme ℓ^2 du vecteur f est localisée sur $o(n)$ coordonnées (ici une seule). Cela se traduit par une perte de masse ℓ^2 dans le sens que $\mathrm{distr}(\sqrt{n} f/\|f\|_2)$, dont le second moment est égal à 1, converge faiblement vers δ_0, dont le second moment est nul. Dans le second cas, en probabilité, il n'y a pas de perte de masse et la mesure empirique $\mathrm{distr}(\sqrt{n} f/\|f\|_2)$ converge vers la loi d'une coordonnée de f.

Plus généralement, le lemme élémentaire suivant décrit la proximité à un vecteur gaussien en terme d'une partie localisée et d'une partie proprement asymptotiquement gaussienne.

LEMME 0.1. — *Soit* $\varepsilon > 0$. *Il existe* $\delta > 0$ *tel que pour tout entier* $n \geq 1$ *si* $f \in \mathbb{R}^n$ *vérifie* $\|f\|_2 = \sqrt{n}$ *et* $\mathrm{d_L}(\mathrm{distr}(f), N(0,\sigma^2)) \leq \delta$ *pour un certain* $\sigma \in [0,1]$ *alors il existe* $\tau \in [0,1]$, $a, b \in \mathbb{R}^n$ *tels que* f *se décompose de la façon suivante*

$$f = \tau a + \sqrt{1-\tau^2}\, b,$$

avec $\langle a, b\rangle = 0$, $\|a\|_2 = \|b\|_2 = \sqrt{n}$, $|\sigma - \tau| \leq \varepsilon$, $\mathrm{d_L}(\mathrm{distr}(a), N(0,1)) \leq \varepsilon$ *et* $\mathrm{d_L}(\mathrm{distr}(b), \delta_0) \leq \varepsilon$.

Théorème principal, forme simplifiée. — La forme simplifiée du théorème principal de [**5**] est la suivante.

THÉORÈME 0.2. — *Soient un entier $d \geq 3$ et $\varepsilon > 0$. Il existe m et δ strictement positifs tels que, pour tout entier $n \geq m$ avec nd pair, si G est un graphe distribué uniformément sur $\mathbb{G}(n,d)$ et A est sa matrice d'adjacence, alors l'événement suivant a une probabilité au moins $1-\varepsilon$: tout quasi-vecteur propre f de déviation δ et orthogonal à $\mathbf{1}$ vérifie $D(f) \leq \varepsilon$.*

En regard du lemme 0.1, le théorème 0.2 affirme donc que si f est un vecteur propre de norme 1 de la matrice d'adjacence d'un graphe d-régulier uniforme avec n sommets alors, lorsque n tend l'infini, f se décompose en une partie de masse ℓ^2 concentrée sur $o(n)$ entrées et le reste est asymptotiquement gaussien. Il est conjecturé que la partie localisée de cette décomposition est triviale :

CONJECTURE 0.3. — *Soient un entier $d \geq 3$ et $\varepsilon > 0$. Il existe m positif tel que, pour tout entier $n \geq m$ avec nd pair, si G est un graphe distribué uniformément sur $\mathbb{G}(n,d)$ et A est sa matrice d'adjacence alors l'événement suivant a une probabilité au moins $1-\varepsilon$: tout vecteur propre f orthogonal à $\mathbf{1}$ vérifie* $\mathrm{d_L}(\mathrm{distr}(\sqrt{n}f/\|f\|_2), N(0,1)) \leq \varepsilon$.

Il est important de remarquer que dans la conjecture ci-dessus, on se restreint aux vecteurs propres. Nous verrons en effet que pour tout $\sigma \in [0,1]$, il existe une suite de quasi-vecteurs propres f_n de déviation δ_n avec $\delta_n \to 0$ telle que, en probabilité, $\mathrm{d_L}(\mathrm{distr}(\sqrt{n}f_n/\|f_n\|_2), N(0,\sigma^2)) \to 0$. Voir remarque 2.2.

Nous allons maintenant énoncer le résultat principal de [**5**]. L'objectif est de mesurer la distance d'un vecteur propre f de la matrice d'adjacence de G uniforme sur $\mathbb{G}(n,d)$ à un processus gaussien qui satisfait, avec probabilité 1, l'équation de vecteur propre sur le revêtement universel de G. Nous allons d'abord expliquer tous ces termes.

Arbre régulier infini. — Pour $d \geq 2$ entier, nous noterons $T_d = (V_d, E_d)$ l'*arbre d-régulier infini*. Plus précisément, il sera commode d'utiliser la notation généalogique suivante. L'ensemble des sommets de T_d est le sous-ensemble suivant des suites finies d'entiers, $\varnothing$ représentant la suite vide,

$$V_d = \{\varnothing\} \cup \{1,\ldots,d\} \cup \bigcup_{k=1}^{\infty} \{1,\ldots,d\} \times \{1,\ldots,d-1\}^k.$$

On définit le *parent* de $x \in \{1,\ldots,d\}$ comme $\varnothing$. Pour $k \geq 1$ entier, le parent de $x = (x_1,\ldots,x_{k+1}) \in \mathbb{N}^{k+1} \cap V_d$ est $(x_1,\ldots,x_k) \in \mathbb{N}^k \cap V_d$. Enfin $\varnothing$ n'a pas de parent. L'ensemble des arêtes de T_d est

$$E_d = \{\{x,y\} : x \text{ est parent de } y\}.$$

Nous dirons que y est un *enfant* de x si x est le parent de y. Nous dirons que y est un *descendant* de x si il existe une suite $(x_0, \dots, x_k)$ avec $x_0 = x$, $x_k = y$ et x_{t-1} est le parent de x_t pour tout entier $1 \leq t \leq k$ (en d'autres termes, x est un suffixe de y).

Pour d pair, l'arbre T_d est isomorphe au graphe de Cayley du groupe libre engendré par $d/2$ générateurs libres et leurs inverses. Pour tout d entier, T_d est aussi le graphe de Cayley du groupe librement engendré par d copies de $\mathbb{Z}/2\mathbb{Z}$ et ses générateurs libres.

Revêtement aléatoire. — Soit $G = (V, E)$ un graphe d-régulier. Nous dirons qu'une application $\pi\colon V_d \to V$ est un *revêtement de G* si π est un homéomorphisme local au sens où pour tout $x \in V_d$, l'image par π des d voisins de x dans T_d sont les d voisins de $\pi(x)$ dans G. Remarquons que $\pi(V_d)$ est alors une composante connexe de G.

Tout graphe d-régulier admet un revêtement. Si V est fini, alors il existe une mesure de probabilité naturelle sur l'ensemble de revêtements de G. Si G est connexe, ce *revêtement aléatoire* π est caractérisé par la propriété que pour tout automorphisme ϕ de T_d, $\pi \circ \phi$ et π ont même loi. Si G n'est pas connexe, pour caractériser de façon unique cette mesure, nous supposons également que $\pi(\varnothing)$ est uniformément distribuée sur V.

Ce revêtement aléatoire peut être construit explicitement de la façon suivante. Soit $(\sigma_x)_{x \in V_d}$ une famille de permutations aléatoires indépendantes, $\sigma_\varnothing$ permutation uniformément distribuée dans S_d (le groupe symétrique à d éléments) et pour tout $x \in V_d \backslash \{\varnothing\}$, σ_x uniforme dans S_{d-1}. Le revêtement aléatoire est défini itérativement de la façon suivante. Tout d'abord, $\pi(\varnothing)$ est distribué uniformément sur V et pour tout $i \in \{1, \dots, d\}$, $\pi(i) = v_{\sigma_\varnothing(i)}$ où $\{v_1, \dots, v_d\}$ sont les voisins de $\pi(\varnothing)$ dans G. Ensuite, par récurrence sur $k \geq 1$, pour tout $x \in \mathbb{N}^k \cap V_d$, on étend π en posant pour tout $i \in \{1, \dots, d-1\}$, $\pi((x, i)) = v_{\sigma_x(i)}$ où $\{v_1, \dots, v_{d-1}\}$ sont les $d-1$ voisins de $\pi(x)$ différents de $\pi(y)$, y étant le parent de x.

Distribution d'un vecteur sur le graphe. — La preuve du théorème 0.2 repose sur la possibilité d'associer à la paire formée par un graphe d-régulier $G = (V, E)$ et un vecteur $f \in \mathbb{R}^V$ un processus aléatoire sur T_d qui est invariant par les automorphismes de l'arbre. Cette idée à la fois simple et profonde remonte au moins à Benjamini et Schramm [**9**]. Plus précisément, soit $G = (V, E) \in \mathbb{G}(n, d)$ un graphe d-régulier avec sommets $V = \{1, \dots, n\}$. On peut définir la *loi d'un vecteur vue d'un sommet typique du graphe* de la façon suivante. Soient M un espace métrique et $f \in M^V$. On munit M^{V_d} d'une distance qui induit la topologie produit. On définit $\mathrm{distr}_G(f)$ comme la mesure de probabilité sur M^{V_d}

$$\mathrm{distr}_G(f) = \text{loi de } f \circ \pi,$$

où $\pi\colon V_d \to V$ est le revêtement aléatoire de G défini ci-dessus.