Reconstruction creuse déterministe, d'après Guruswami, Lee et Razborov

Pierre Pansu

17 mars 2008

1 Objectif

Exposer brièvement des résultats récents, trouvés dans la littérature informatique, qui utilisent l'estimée somme-produit de Bourgain, Katz et Tao.

Aujourd'hui, c'est une construction explicite de sous-espaces dont l'intersection avec la boule unité de ℓ^1 est presque ronde, par Guruswami, Lee et Razborov.

2 Distorsion des sous-espaces

2.1 Compression

En compression, la méthode suivante est à la mode, depuis les travaux de Candes/Tao et Donoho (voir [KT]). Il s'agit de coder des vecteurs $u \in \mathbb{R}^N$, en faisant l'hypothèse qu'ils sont k-creux, i.e. que seules k composantes du vecteur sont non nulles (e.g., on a décomposé un signal dans une base d'ondelettes, et on n'a conservé que les k composantes de plus grande intensité). On parvient à construire des matrices $\Phi : \mathbb{R}^N \to \mathbb{R}^n$, $n \ll N$, qui ont la propriété suivante. Si on minimise la norme ℓ^1 des vecteurs v tels que $\Phi(v) = \Phi(u)$, on retombe forcément sur u. On dispose d'un algorithme polynômial pour reconstruire u à partir de $\Phi(u)$, on peut donc coder u par $\Phi(u)$ qui est beaucoup plus petit.

La raison géométrique très grossière est que les vecteurs creux sont dans des coins très aigus de la boule ℓ^1. Un sous-espace affine qui contient un tel point a toutes les chances de toucher la boule ℓ^1 en ce point. Donoho, puis Candes et Tao, ont dégagé des conditions suffisantes sur le noyau de Φ, qui prennent la forme de comparaisons entre normes ℓ^1 et ℓ^2 des projections des vecteurs du noyau sur des paquets de coordonnées. Kashin et Temlyakov ont notablement simplifié ces conditions, en éliminant les projections sur des paquets de coordonnées.

Définition 1 *La* distorsion *d'un sous-espace vectoriel $X \subset \mathbb{R}^N$ est*

$$\Delta(X) = \sqrt{N} \max_{x \in X} \frac{|x|_2}{|x|_1}.$$

Exemple 2 *La distorsion est au plus égale à $\sqrt{N}$, avec égalité (entre autres) pour l'espace $\mathbb{R}^N$ entier.*

Les sous-espaces favorables à la reconnaissance des vecteurs creux sont ceux dont la distorsion est nettement inférieure à $\sqrt{N}$.

Lemme 3 [KT]. *Soit $X \subset \mathbb{R}^N$ un sous-espace tel que $\Delta(X) \leq D$. Alors X reconstruit uniquement les vecteurs k-creux pour $k = \dfrac{N}{4D^2}$.*

Preuve Etant donné $S \subset \{1, \ldots, n\}$, on note x_S le vecteur tronqué dans lequel on a mis à zéro les coordonnées de x d'indices n'appartenant pas à S.

1. Si $Ax = 0$, alors pour tout S de taille inférieure à $k = \dfrac{N}{4D^2}$, $\| x_S \|_1 \leq \frac{1}{2} \| x \|_1$. En effet, par Cauchy-Schwarz,

$$\| x_S \|_1 \leq \sqrt{|S|} \| x_S \|_2 \leq \sqrt{|S|} \| x \|_2 \leq \frac{\sqrt{N}}{2D} \frac{D}{\sqrt{N}} \| x \|_1 .$$

2. Si u est k-creux et si $Av = Au$, alors $\| v \|_1 > \| u \|_1$. En effet, si S est le support de u, et si $x = v - u$,

$$\begin{aligned} \| v \|_1 &= \| u + x \|_1 = \| (u+x)_S \|_1 + \| x_{\bar{S}} \|_1 \\ &\geq \| u_S \|_1 - \| x_S \|_1 + \| x_{\bar{S}} \|_1 \\ &= \| u \|_1 + \| x \|_1 - 2 \| x_{\bar{S}} \|_1 \\ &\geq \| u \|_1 . \end{aligned}$$

∎

Il s'agit donc de construire des sous-espaces de distorsion la plus petite possible.

2.2 Résultats

Un sous-espace choisi au hasard est ce qui se fait de mieux.

Remarque 4 *Le théorème de Dvoretzky affirme que dans tout espace de Banach de dimension N, il existe un sous-espace vectoriel de dimension $k = const(\delta)\dfrac{\log(k)}{k}$ tel que la norme induite soit à distance (de Banach-Mazur) au plus $1 + \delta$ d'un espace euclidien. Ici, on se pose une question differente, car il s'agit de comparer directement des normes, et non des normes à changement de coordonnées linéaire près. Néanmoins, on trouve quand même des sous-espaces de distorsion proche de 1. Comme il ne s'agit pas de n'importe quelle norme, mais de la norme ℓ_1, les sections presque euclidiennes ont des dimensions linéaires en N.*

Théorème 5 [FLM]. *Pour tout $\delta > 0$, il existe c_δ tel que si $c_\delta N$ vecteurs sont tirés au hasard dans S^{N-1} et engendrent un sous-espace X, alors avec probabilité tendant vers 1 quand N tend vers l'infini,*

$$\Delta(X) \leq 1 + \delta.$$

L'énoncé suivant couvre un spectre plus large de situations.

Théorème 6 [K], [GG]. *Soit A une matrice de taille $n \times N$ dont les coefficients sont tirés au hasard dans $\{\pm 1\}$. Avec probabilité tendant vers 1 quand N et n tendent vers l'infini,*

$$\Delta(\mathrm{ker} A) \leq \sqrt{\frac{N}{n}(1+\log(\frac{N}{n}))}.$$

Par conséquent, les sous-espaces tirés au hasard reconstruisent uniquement les vecteurs k-creux pour $k = \dfrac{n}{4\log(N/n)}$.

On voudrait éliminer le hasard. Un gadget favori en combinatoire, les matrices d'Hadamard, semble un bon candidat. Malheureusement, la matrice formée des $N/2$ premières lignes d'une matrice d'Hadamard a un noyau dont la distorsion est en $N^{1/4}$.

Le théorème exposé aujourd'hui fournit de façon déterministe des sous-espaces satisfaisants lorsque le rapport N/n tend vers l'infini assez lentement.

Théorème 7 [GLR]. *Pour toute fonction positive $N \mapsto \eta(N)$, il existe un algorithme déterministe qui produit, en un temps polynômial, des sous-espaces vectoriels $X \subset \mathbb{R}^N$ de dimensions $\geq (1-\eta)N$ et de distorsions*

$$\Delta(X) \leq (\eta^{-1} \log\log N)^{C \log\log N}.$$

Lorsque N/n croît comme une puissance de $\log N$, la distorsion obtenue est de la forme $(\frac{N}{n}(1+\log(\frac{N}{n})))^{C\log\log N}$, ce qui est moins bon que le tirage aléatoire, mais permet tout de même de reconstruire les vecteurs k-creux avec $k = N/(\log N)^{C \log\log\log(N)}$.

3 Stratégie

Etant donné un sous-espace vectoriel L de $\mathbb{R}^d$ et un graphe biparti G qui possède N sommets de gauche, et dont tous les sommets de droite sont de même degré d, considérons le sous-espace $X(G, L)$ de $\mathbb{R}^N$ formé des vecteurs dont la restriction aux premiers voisins de chaque sommet de droite appartient à L. Si le graphe G a un bon profil asymétrique (e.g. linéaire), alors $X(G, L)$ a une faible distorsion. Les graphes aléatoires ont un profil linéaire, mais aucun graphe explicite (et constructible en un temps polynômial en N) n'a un aussi bon profil. L'estimée somme-produit garantit que le graphe $ab + c$ a un profil suffisant pour les sous-ensembles de petite taille. Pour les grandes tailles, il vaut mieux utiliser les graphes de Ramanujan ([LPS]).

4 Profil des graphes bipartis

4.1 Définition

Un graphe biparti a son ensemble de sommets divisé en deux parties V_L et V_R, les sommets de gauche et ceux de droite. Un sous-ensemble $S \subset V_L$ a un bord ∂S contenu dans V_R, c'est l'ensemble des voisins des éléments de S.

Définition 8 *Le* profil asymétrique *d'un graphe biparti est la fonction qui à un entier m associe $\Lambda(m)$, le minimum des tailles des bords des sous-ensembles de V_L de taille au moins m.*

Exemple 9 *Pour le graphe biparti complet, alors $\Lambda(m) = |V_R|$ dès que $m > 0$.*

Exemple 10 *Pour le graphe biparti aléatoire de degré d, $\Lambda(m) = c(d)m$.*

4.2 Graphes d'incidence des graphes de Ramanujan

Ce qui vient en premier à l'esprit, c'est d'utiliser les graphes de Ramanujan, graphes régulier de petit λ_2 construits dans [LPS]. On les transforme en graphes bipartis en prenant leurs graphes d'incidence.

Définition 11 *Soit Γ un graphe. Son* graphe d'incidence *est le graphe biparti dont les sommets gauche sont les arêtes de Γ, les sommets droits les sommets de Γ, et on relie chaque arête à ses extrémités.*

Proposition 12 *Les graphes d'incidence des graphes de Ramanujan permettent d'obtenir, pour tout $d \geq 5$ et tout $N \geq d$ un graphe à N sommets gauche, n sommets droits, de degré gauche égal à 2, de degré droit égal à d, et de profil asymétrique $\Lambda(m) \geq \min\{\frac{m}{2\sqrt{d}}, \frac{2mN}{d}\}$.*

Preuve Le lien entre λ_2 et Γ et profil du graphe d'incidence provient d'une majoration du nombre d'arêtes des sous-graphes des expanseurs, due à N. Alon et F.R.K. Chung, [AC]. Cette estimation conduit à une perte : le profil n'est pas linéaire pour m petit. ∎

C'est pour compenser cette faiblesse des graphes d'incidence qu'on utilise les graphes somme-produit.

4.3 Graphes somme-produit

Théorème 13 [GLR], interprétation de [BKSSW]. *Il existe une constante absolue $\epsilon > 0$ telle que, pour tout premier p, le graphe biparti tel que*

$$V_L = \mathbb{F}_p^3, \quad V_R = \{1, 2, 3, 4\} \times \mathbb{F}_p,$$

et où $(a, b, c) \in \mathbb{F}_p^3$ est relié à $(1, a)$, $(2, b)$, $(3, c)$ et $(4, ab+c)$, a un profil asymétrique $\Lambda(m) \geq \min\{p^{0.9}, m^{1/3+\epsilon}\}$.

Remarque 14 *Trivialement, si $S \subset V_L$, $|\partial S| \geq |S|^{1/3}$.*

En effet, si $S_1 = \{(1, a) \,|;\, (a, b, c) \in S\}$, etc..., alors $S \to S_1 \times S_2 \times S_3$, $(a, b, c) \mapsto ((1, a), (2, b), (3, c))$ est injective, donc $|S| \leq |S_1| \times |S_2| \times |S_3|$, d'où $|\partial S| \geq |S_1| + |S_2| + |S_3| \geq |S|^{1/3}$.

Autrement dit, l'estimée somme-produit permet d'améliorer un petit peu cette inégalité triviale, de façon uniforme en p.

Ce profil peut paraître très mauvais (bien moins que linéaire). Attention, le nombre de sommets de gauche est le cube du nombre de sommets de droite, $|V_L| = |V_R|^3$, donc il y a forcément peu de premiers voisins. En revanche, il y a beaucoup de seconds voisins, ce qui donne la propriété d'expansion souhaitée.

Preuve Soit $S \subset V_L = \mathbb{F}_p^3$. De deux choses l'une.
- Ou bien l'une des coordonnées a une influence prépondérante dans S (i.e. S contient un gros sous-ensemble proche d'un produit $A \times \{b\} \times \{c\}$). Dans ce cas, le sous-ensemble $Ab + c \subset \mathbb{F}_p = V_R$ est contenu dans le bord de S, et $|Ab + c| = |A|$ est de l'ordre de $|S|$.
- Ou bien les trois coordonnées jouent des rôles équivalents. Dans ce cas, S contient un gros sous-ensemble proche d'un produit $A \times B \times C$. Le sous-ensemble $AB+C \subset \mathbb{F}_p = V_R$ est contenu dans le bord de S, et $|AB+C| > |A|^{1+\epsilon} > |S|^{1/3+\epsilon}$, d'après l'estimée somme-produit, à condition que $|A| < |\mathbb{F}_p|^{1-\epsilon}$.

En fait, on a besoin de la version mesures de l'estimée somme-produit, qui apparaît initialement dans [BIW]. ■

5 Mécanisme améliorant la distorsion

Voici une formulation technique de la distorsion.

Définition 15 *Un sous-espace vectoriel de $\mathbb{R}^d$ est* (t, T, ϵ)-étalé *si tout vecteur de X satisfait*

$$\min_{S\,;\,|S|\leq T} |x_{\bar{S}}|_2 \geq \epsilon \min_{S\,;\,|S|\leq t} |x_{\bar{S}}|_2$$

On montre aisément qu'un espace (T, t, ϵ)-étalé a une distorsion petite.

Lemme 16 *Soit L un sous-espace vectoriel de $\mathbb{R}^d$ qui est $(0, t, \epsilon)$-étalé. Soit G un graphe biparti de profil asymétrique Λ, dont les sommets gauches sont de degré au plus D et les sommet droits de degré exactement d. Alors pour tout T, le sous-espace $X(G, L)$ des vecteurs de $\mathbb{R}^{V_R}$ dont la restriction aux voisins d'un sommet droit appartient toujours à L est $(T, \frac{t}{T}\Lambda(T), \frac{\epsilon}{\sqrt{2D}})$-étalé.*

On applique ce lemme d'abord à des graphes somme-produits, avec un L initial construit à l'aide de matrices d'Hadamard. Par récurrence, on obtient des espaces vectoriels X de dimension de plus en plus petite. A un moment donné, les graphes d'incidence prennent le relai.

Références

[AC] Alon, N. ; Chung, F. R. K. *Explicit construction of linear sized tolerant networks.* Proceedings of the First Japan Conference on Graph Theory and Applications (Hakone, 1986). Discrete Math. **72** (1988), no. 1-3, 15–19.

[BKT] Bourgain, Jean ; Katz, Netz ; Tao, Terry. *A sum-product estimate in finite fields, and applications.* Geom. Funct. Anal. **14** (2004), no. 1, 27–57.

[BKSSW] Barak, Boaz ; Kindler, Guy ; Shaltiel, Ronen, Sudakov, Benny ; Widgerson, Avi. *Simulating independance : new constructions of condensers, Ramsey graphs, dispersers, and extractors.* STOC 2005 et http ://www.math.ias.edu/∼avi/PUBLICATIONS/MYPAPERS/bkssw04/bkssw04.pdf

[BIW] Barak, Boaz ; Impagliazzo, Russel ; Widgerson, Avi *Extracting randomness using few independant sources.* SIAM J. Comput. **36** (2006), 1095-1118. Aussi : FOCS 2004 et http ://www.math.ias.edu/∼avi/PUBLICATIONS/MYPAPERS/BIW04/BIW.pdf

[FLM] Figiel, T. ; Lindenstrauss, J. ; Milman, V. D. *The dimension of almost spherical sections of convex bodies.* Acta Math. **139** (1977), no. 1-2, 53–94.

[GG] Garnaev, A. Yu. ; Gluskin, E. D. *The widths of a Euclidean ball.* Dokl. Akad. Nauk SSSR **277** (1984), no. 5, 1048–1052.

[GLR] Guruswami, Venkatesan ; Lee, James ; Razborov, Alexander. *Almost Euclidean subspaces of ℓ_1^N via expander codes.* SODA 2008 et http ://www.cs.washington.edu/homes/jrl/papers/spread.pdf

[K] Kašin, B. S. *The widths of certain finite-dimensional sets and classes of smooth functions.* Izv. Akad. Nauk SSSR Ser. Mat. **41** (1977), no. 2, 334–351.

[KT] Kashin, B.S., Temlyakov, V.N. *A remark on compressed sensing.* http ://www.dsp.ece.rice.edu/cs/KT2007.pdf

[LPS] Lubotzky, A. ; Phillips, R. ; Sarnak, P. *Ramanujan graphs.* Combinatorica **8** (1988), no. 3, 261–277.