Application des estimées sommes-produits aux extracteurs, d'après Barak, Impagliazzo et Widgerson

Pierre Pansu

12 février 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 d'extracteurs par Barak, Impagliazzo et Widgerson. Pour plus tard, une construction explicite de sous-espaces bien étalés dans $\mathbb{R}^N$ par Guruswami, Lee et Razborov.

2 Extracteurs

Le problème suivant remonte à Von Neumann. Etant donnée une source physique (compteur Geiger...) produisant des chaînes de bits aléatoires, dont la distribution est inconnue (mais pas trop exotique), comment, en leur appliquant un algorithme déterministe, produire des chaînes de bits équiréparties, avec une précision prescrite ?

L'algorithme déterministe, c'est une application $\texttt{Ext} : \{0,1\}^m \to \{0,1\}^n$. La source, c'est une variable aléatoire X à valeurs dans $\{0,1\}^m$, dont seule la loi, une mesure de probabilité μ sur $\{0,1\}^m$, joue un rôle. Le générateur de bits aléatoires obtenu est la variable aléatoire $\texttt{Ext} \circ X$. Ce qui nous intéresse, c'est sa loi, i.e. la mesure image $\texttt{Ext}_*\mu$. Traditionnellement, on mesure le défaut d'équirépartition par la distance ℓ^1 à la distribution uniforme U.

Définition 1 *La distance entre deux mesures de probabilité μ et ν sur un ensemble fini est*

$$dist(\mu,\nu) = \sum_x |\mu(\{x\}) - \nu(\{x\})|.$$

Autrement dit, le problème consiste à trouver $\texttt{Ext}$ de sorte que $dist(\texttt{Ext}_*\mu, U)$ soit petit pour toutes les mesures μ d'une famille, la plus grande possible.

Si μ est fortement concentrée en un point, il en est de même de $\texttt{Ext}_*\mu$, qui est donc très mal répartie. Cela se mesure quantitativement au moyen de la min-entropie.

Définition 2 *La* min-entropie *d'une mesure de probabilité μ sur un ensemble fini est*

$$H^\infty(\mu) = \min\{-\log_2 \mu(\{x\})\,;\, \mu(\{x\}) \neq 0\}.$$

Exemple 3 *La min-entropie de toute mesure de probabilité sur $\{0,1\}^n$ est $\geq n$, avec égalité exactement pour la distribution uniforme.*

Clairement, la min-entropie diminue quand on remplace une mesure par son image par une application. D'autre part, la min-entropie est continue pour la distance ℓ^1. Par conséquent, si l'image de μ est 2^{-N}-proche de la distribution uniforme sur $\{0,1\}^n$, la min-entropie de μ vaut au moins $m - 2^{n-N}$. Il est donc impossible d'améliorer sensiblement une distribution de min-entropie inférieure à n.

3 Extracteurs multi-sources

On peut y arriver à condition d'utiliser plusieurs sources indépendantes. En termes de mesures de probabilités, la loi jointe de plusieurs variables aléatoires indépendantes est un produit tensoriel de mesures. La famille de mesures considérée sur $\{0,1\}^m$ sera donc celle des produits tensoriels de ℓ mesures sur $\{0,1\}^{m'}$, où $m = \ell m'$.

Comment mesurer la qualité d'un extracteur ? Le lemme élémentaire suivant montre que, si on s'autorise au moins un nombre quelconque (mais borné) de sources, on peut réduire facilement l'écart à l'équirépartition.

Lemme 4 *Soient μ et ν des mesures sur un groupe. Alors*

$$dist(\mu \star \nu, U) \leq dist(\mu, U) dist(\nu, U).$$

Par conséquent, si $dist(\mu, U) \leq 2^{-N}$, alors l'image $\mu^{\star\ell}$ de $\mu^{\otimes\ell}$ par l'extracteur $\sum : \{0,1\}^{\ell m'} \to \{0,1\}^{m'}$, $(x_1, \ldots, x_\ell) \mapsto x_1 + \cdots + x_\ell$ satisfait $dist(\mu^{\star\ell}, U) \leq 2^{-N\ell}$.

Preuve En effet, si $\mu = U + f$, $\nu = U + g$, $\mu \star \nu = U + f \star g$, et $\| f \star g \|_1 \leq \| f \|_1 \| g \|_1$. ∎

Ce qui compte donc, c'est plutôt le nombre de sources, et comment il dépend de la borne sur la min-entropie des sources. Il est communément admis que c'est une application tirée au hasard parmi toutes les applications possibles $\sum : \{0,1\}^{\ell m'} \to \{0,1\}^n$ qui donne le meilleur résultat comme extracteur multisource. Voici un exemple de résultat sur la performance de tels extracteurs. Deux sources suffisent.

Théorème 5 *Pour tout $\delta > 0$ il existe $C > 0$ tel que, pour tout m', tout $k \geq \delta \log(m)$ et tout $n \leq 2k - C\log(m')$, une application $\mathtt{Ext} : \{0,1\}^{2m'} \to \{0,1\}^n$ tirée au hasard possède, avec forte probabilité, la propriété suivante : si μ_1 et μ_2 sont des mesures de probabilité de min-entropies $\geq k$ sur $\{0,1\}^{m'}$, alors*

$$dist(\mathtt{Ext}_*(\mu_1 \otimes \mu_2), U) < 2^{-n/C}.$$

En d'autres termes, l'extracteur aléatoire extrait effectivement la quasi-totalité de l'entropie des sources. Le théorème exposé aujourd'hui fait les choix suivants : ℓ indépendant de n, $m' = n$, ce qui implique une min-entropie $\geq \delta n$ où $0 < \delta < 1$ est indépendant de n. Noter que ℓ est forcément $\geq 1/\delta$.

Théorème 6 [BIW]. *Il existe une constante absolue C telle que, pour tout $\delta > 0$, il existe $\ell \leq \delta^{-C}$ et une application $\mathbf{Ext} : \{0,1\}^{n\ell} \to \{0,1\}^n$ calculable en temps polynômial tels que si $\mu_1, \ldots, \mu_\ell$ sont des mesures de probabilité de min-entropies $\geq \delta n$, alors*

$$dist(\mathbf{Ext}_*(\mu_1 \otimes \ldots \otimes \mu_\ell), U) < 2^{-n/C}.$$

Ce résultat semble assez faible, en comparaison du résultat randomisé. En effet, ℓ est grand par rapport à $1/\delta$, donc l'entropie est inutilement gaspillée. Néanmoins, la preuve est intéressante.

La référence [R] comporte, outre des améliorations de ce résultat obtenues par des méthodes différentes, un survol des constructions d'extracteurs multi-sources.

4 Idée de la preuve

L'idée de départ est d'identifier $\{0,1\}^n$ au corps premier $\mathbb{F}_p$ et de considérer l'application

$$\mathtt{Ext} : \mathbb{F}_p^3 \to \mathbb{F}_p, \quad (a,b,c) \mapsto ab + c.$$

De l'estimée somme-produit, il résulte que pour tout $A \in \mathbb{F}_p$ pas trop petit ni trop grand, $|\mathtt{Ext}(A \times A \times A)| > |A|^{1+\epsilon}$. En itérant $\mathtt{Ext}$ un nombre $\ell = O(\log(1/\delta))$ fois, on obtient $\mathtt{Ext}^\ell : \mathbb{F}_p^{3^\ell} \to \mathbb{F}_p$ telle que $\mathtt{Ext}^\ell(A^{3^\ell}) = \mathbb{F}_p$ pour A pas trop petit. C'est ce qu'on voulait, pour les mesures particulières proportionnelles aux fonctions caractéristiques des sous-ensembles.

Il reste à passer des sous-ensembles aux mesures. Attention, il n'y a pas d'estimée somme-produit pour les mesures. En effet, la somme des fonctions caractéristiques d'une progression arithmétique et d'une progression géométrique donne une mesure de probabilité μ dont les convolées additive et multiplicative ont des min-entropies très petites. En revanche, la version plus faible suivante est vraie : la min-entropie de toute mesure augmente significativement quand on prend son image par $\mathtt{Ext}$.

Lemme 7 *Il existe une constante absolue $\epsilon > 0$ telle que pour tout p, si μ, ν et π sont des mesures de probabilité sur $\mathbb{F}_p$ de min-entropie au moins m, alors $\mathbf{Ext}(\mu \otimes \nu \otimes \pi)$ est $2^{-\epsilon m}$-proche d'une mesure de min-entropie au moins égale à $\min\{(1+\epsilon)m, 0.9 \log |\mathbb{F}_p|\}$.*

Pour compléter la preuve du théorème 6, il suffit d'établir le lemme suivant.

Lemme 8 *Soient $\mu_1, \ldots, \mu_9$ neuf mesures de probabilité sur $\mathbb{F}_p$ de min-entropies au moins égales à $0.9 \log |\mathbb{F}_p|$. Alors l'image de la mesure produit par $\mathbf{Ext}^2$ est $|\mathbb{F}_p|^{-0.1}$-proche de la distribution uniforme.*

On a triché en identifiant $\{0,1\}^n$ au corps premier $\mathbb{F}_p$. En effet, actuellement, il n'existe aucun algorithme qui, en un temps polynômial en n, retourne un nombre premier compris entre 2^n et 2^{n+1}. Néanmoins, en détournant une faible partie du hasard contenu dans les sources, et en utilisant le test de primalité polynômial d'Agrawal, Kayal et Saxena, on peut le faire.

5 Preuve du lemme 7

Etant donnée une mesure de probabilité μ, on note $cp(\mu)$ la *probabilité de collision*, i.e. la probabilité que deux éléments tirés indépendamment suivant μ coïncident. En d'autres termes, $cp(\mu)$ est le carré de la norme ℓ^2 de μ vue comme fonction positive. Pour la mesure uniforme sur un sous-ensemble A, $cp(\mu_A) = |A|^{-1}$.

Pour qu'une mesure soit proche d'une mesure de grande min-entropie, il suffit que sa probabilité de collision soit petite. De même, sur un grand ensemble, petite probabilité de collision équivaut à équirépartition.

Lemme 9 *1. Si $cp(\mu) \leq 1/KL$, alors il existe une mesure de probabilité ν telle que $dist(\mu, \nu) \leq 1/\sqrt{L}$ et $H^\infty(\nu) \geq \log K$.*

2. Pour toute mesure de probabilité μ sur un ensemble à N éléments, $\frac{1}{N} dist(\mu, U)^2 \leq cp(\mu) \leq \frac{1}{N} + dist(\mu, U)^2$.

Preuve

1. ν est la restriction de μ aux points de masse $\leq 1/K$.

2. On écrit $\mu = U + f$. Alors $\| f \|_1 = dist(\mu, U)$ et $cp(\mu) - \frac{1}{N} = \| f \|_2^2$ est minimale quand $|f|$ est constante, maximale quand f est concentrée en un point. ∎

Lemme 10 *Il existe une constante absolue $\epsilon > 0$ telle que pour tout p, et tout entier M, on a la propriété suivante. Soit $A \subset \mathbb{F}_p$ un ensemble de taille $\geq M^{1-10\epsilon}$. On suppose qu'il existe un autre ensemble $B \subset \mathbb{F}_p$ de taille $\geq M^{1-10\epsilon}$ tel que $A \cdot B \leq M^{1+\epsilon}$. Alors pour tout ensemble C de taille M,*

$$cp(\mu_A + \mu_C) \leq M^{-1-10\epsilon}.$$

Preuve du lemme 10. On combine l'estimée somme-produit avec le lemme de Balog et Szemerédi revu par Gowers (qui dit que si $A +_E B$ est petit, on peut extraire de grands sous-ensembles $A' \subset A$ et $B' \subset B$ tels que $A' + B'$ est petit, le lien étant que $cp(\mu_A + \mu_B)$ grand entraîne que $A +_E B$ est petit pour un E très grand) et un lemme de Rusza (qui dit que si A et B sont de tailles égales et $A + B$ est petit, alors $A + A$ est petit).

Retour à la preuve du lemme 7. Appelons $+$-*amicaux* les ensembles qui satisfont à la conclusion du lemme 2. Il y a un lemme 3 qui consiste à échanger addition et multiplication dans le lemme 2. Appelons $\times$-*amicaux* ceux qui satisfont à la conclusion du lemme 3. Alors, en utilisant Balog-Szemerédi-Gowers, on montre que toute partie de $\mathbb{F}_p$ est, à un sous-ensemble très petit près, partitionné en un ensemble $+$-amical A_+ et un ensemble $\times$-amical $A_\times$. Puis on observe que $\mu_A \cdot \mu_B + \mu_C$ est proche d'une combinaison convexe de $\mu_{A_+} \cdot \mu_B + \mu_C$ et de $\mu_{A_\times} \cdot \mu_B + \mu_C$, qui ont chacune une faible probabilité de collision. ∎

6 Preuve du lemme 8

Attention, on va utiliser des notations suggestives mais qui peuvent induire en erreur.

Notation. *Si μ et ν sont des mesures de probabilité sur $\mathbb{F}_p$, on note $\mu+\nu$ (resp. $\mu\cdot\nu$) la convolution additive (resp. multiplicative).*

Avec ces notations, $\texttt{Ext}_*(\mu\otimes\nu\otimes\pi)=\mu\cdot\nu+\pi$.

Soient $\mu_1,\ldots,\mu_9$ des mesures de probabilité sur $\mathbb{F}_p$ de min-entropies $\geq 0.9\log|\mathbb{F}_p|$, i.e. pour tout $x\in\mathbb{F}_p$, $\mu_i(x)\leq|\mathbb{F}_p|^{-0.9}$. Alors $cp(\mu_i)=\sum_x\mu_i(x)^2\leq|\mathbb{F}_p|^{-1.8}|\mathbb{F}_p|=|\mathbb{F}_p|^{-0.8}$.

Sachant que pour toutes mesures μ et ν, $cp(\mu+\nu)\leq cp(\mu)$ (moyenner une fonction n'augmente pas la norme ℓ^2),

$$cp(\texttt{Ext}^2(\mu_1,\ldots,\mu_9))\leq cp((\mu_1\cdot\mu_2+\mu_3)\cdot(\mu_4\cdot\mu_5+\mu_6)).$$

Pour toutes mesures μ, ν et π et toute application $f:\mathbb{F}_p^2\to\mathbb{F}_p$, $f_*((\mu\cdot\nu)\otimes\pi)$, vue comme une fonction, est une moyenne des fonctions $f_*((x\cdot\nu)\otimes\pi$ lorsque x décrit $\mathbb{F}_p$, donc $cp(f_*((\mu\cdot\nu)\otimes\pi)\leq\max\{cp(f_*((x\cdot\nu)\otimes\pi)\,;\,x\in\mathbb{F}_p\}$. En particulier,

$$cp((\mu_1\cdot\mu_2+\mu_3)\cdot(\mu_4\cdot\mu_5+\mu_6))\leq\max\{cp((\mu_1\cdot x+\mu_3)\cdot(\mu_4\cdot y+\mu_6))\,;\,x,\,y\in\mathbb{F}_p\}.$$

La preuve du lemme 8 se ramène donc au lemme suivant.

Lemme 11 *Soient $\mu_1,\ldots,\mu_4$ quatre mesures de probabilité sur $\mathbb{F}_p$ de probabilités de collisions $\leq c$. Alors*

$$dist((\mu_1+\mu_2)\cdot(\mu_3+\mu_4),U)\leq O(c^2|\mathbb{F}_p|^{3/2}).$$

Preuve du lemme 11.

On va montrer que $dist((\mu_1+\mu_2)\cdot(\mu_3+\mu_4)^{-1},U)\leq O(c^2|\mathbb{F}_p|)$. Ensuite, il suffit de passer aux probabilités de collision, car celles-ci satisfont $cp(\mu\cdot\nu)=cp(\mu\cdot\nu^{-1})$ (la probabilité que $xy=x'y'$ est égale à la probabilité que $xy'^{-1}=x'y^{-1}$), puis revenir à la distance à U, grâce à la deuxième partie du lemme 9, le prix à payer est un facteur $|\mathbb{F}_p|^{1/2}$.

Quitte à transporter μ_2 et μ_4 par $x\mapsto -x$, on étudie $dist((\mu_1-\mu_2)\cdot(\mu_3-\mu_4)^{-1},U)$. On démontre le cas particulier où $\mu_2=\mu_1$ et $\mu_3=\mu_4$ (le cas général s'en déduit assez rapidement).

Soient X_i des variables indépendantes de lois μ_i. Comme une probabilité de collision est une norme ℓ^2, elle est toujours $\geq 1/|\mathbb{F}_p|$. Par conséquent, pour tout $s\in\mathbb{F}_p\setminus\{0\}$,

$$p(X_1+sX_3=X_2+sX_4)\geq\frac{1}{|\mathbb{F}_p|}.$$

Or cet évènement se produit seulement si $X_3=X_4$ (ce qui entraîne que $X_1=X_2$), ou bien si $s=(X_2-X_1)/(X_3-X_4)$. Par conséquent,

$$p(s=(X_2-X_1)/(X_3-X_4))+p(X_1=X_2\wedge X_3=X_4)\geq\frac{1}{|\mathbb{F}_p|}.$$

Donc, pour $s\neq 0$, $p(s=(X_2-X_1)/(X_3-X_4))\geq\frac{1}{|\mathbb{F}_p|}-c^2$, ce qui entraîne que $dist((\mu_1-\mu_2)\cdot(\mu_3-\mu_4)^{-1},U)\leq\frac{2}{|\mathbb{F}_p|}+2|\mathbb{F}_p|c^2$. ∎

Références

[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.

[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

[R] Rao, Anup, *Extractors for a constant number of polynomially small min-entropy independent sources.* http ://www.math.ias.edu/∼arao/pubs/indep.pdf