Énumérer les vecteurs les plus courts d'un réseau Euclidien : l'algorithme de Kannan

Damien STEHLÉ

Rocquencourt, 12/03/2007

Travail en commun avec Guillaume HANROT

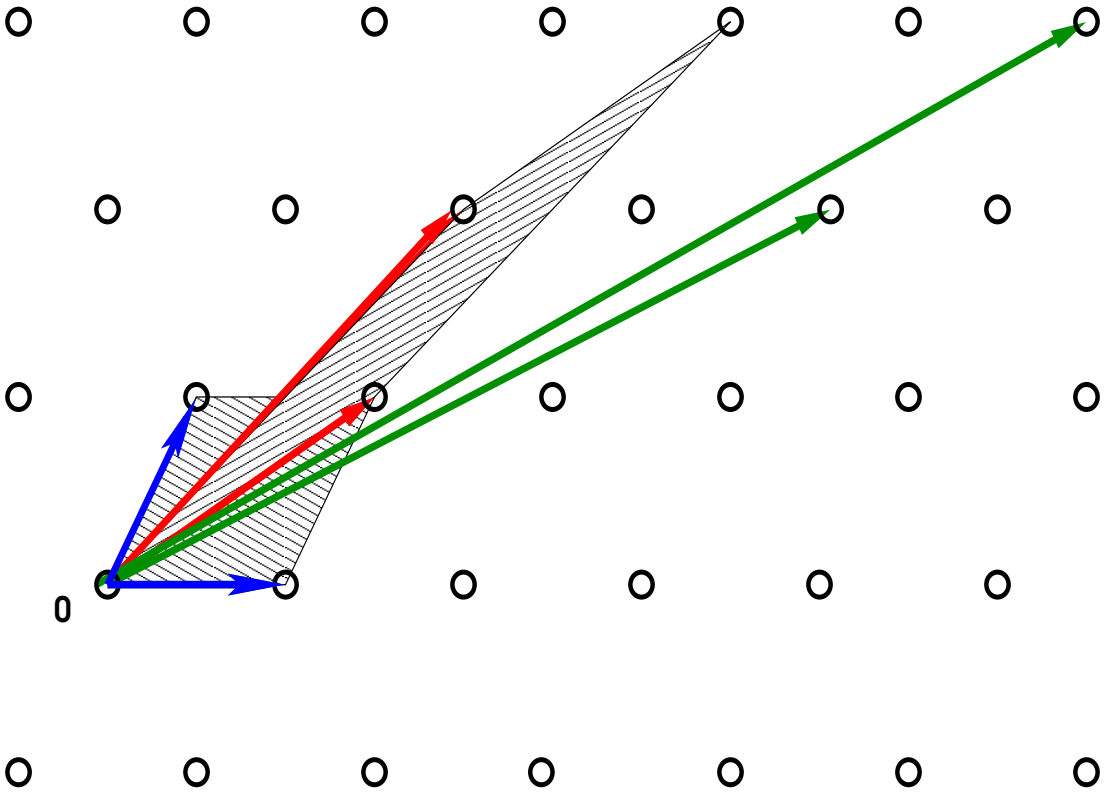
`http://perso.ens-lyon.fr/damien.stehle/`

Plan de l'exposé

1. Quelques rappels sur les réseaux.
2. Les algorithmes d'énumération des vecteurs les plus courts.
3. Une nouvelle borne de complexité.
4. Problèmes connexes.

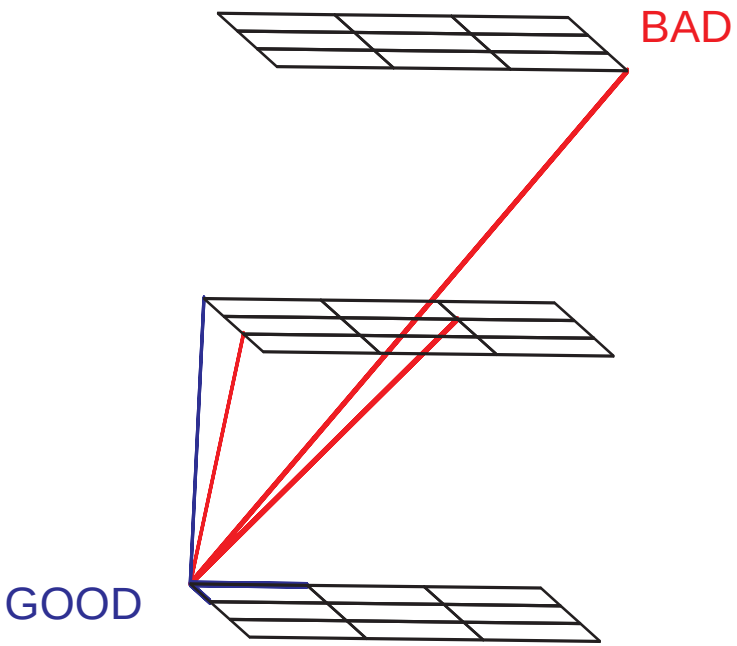
1) Rappels sur les réseaux

Un réseau de dimension 2 avec 3 bases



But : des vecteurs de base les plus orthogonaux possible.

Dimensions 3 et 5



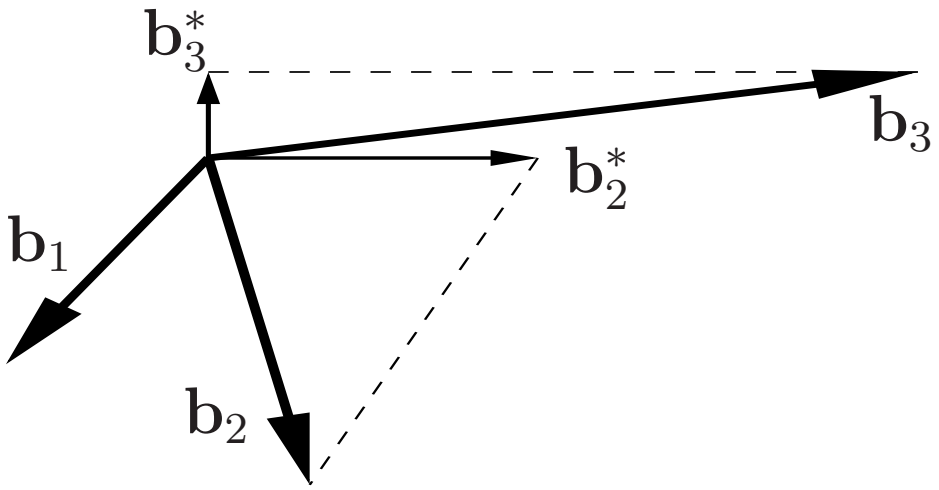
$$\begin{bmatrix} 853539607829 & 0 & 0 & 0 & 0 \\ 512469270672 & 1 & 0 & 0 & 0 \\ 487596978484 & 0 & 1 & 0 & 0 \\ 112511841846 & 0 & 0 & 1 & 0 \\ 24050211137 & 0 & 0 & 0 & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 62 & -115 & -56 & -26 & -19 \\ 27 & -37 & 163 & -156 & 59 \\ 4 & 82 & 65 & -97 & -198 \\ -256 & -133 & -75 & -82 & 89 \\ 91 & 151 & -174 & -255 & 48 \end{bmatrix}$$

Quelques définitions

- Un réseau est un sous-groupe discret d'un certain $\mathbb{R}^n$.
- Soient $\mathbf{b}_1, \ldots, \mathbf{b}_d \in \mathbb{R}^n$ linéairement indépendants.
$$L[\mathbf{b}_1, \ldots, \mathbf{b}_d] = \left\{ \sum_{i=1}^{d} x_i \mathbf{b}_i, x_i \in \mathbb{Z} \right\}.$$
- Les $\mathbf{b}_i$ forment une base.
- Il existe une infinité de bases, liées entre elles par des relations unimodulaires (matrices entières carrées de déterminants ± 1).
- Pour simplifier, on prend des sous-groupes de $\mathbb{Z}^n$ avec $d = n$, donnés par d vecteurs indépendants. La complexité est évaluée par rapport à la dimension d.

Orthogonalisation de Gram-Schmidt

- Procédé pour orthogonaliser $(\mathbf{b}_1, \ldots, \mathbf{b}_d)$.
- $\mathbf{b}_1^* = \mathbf{b}_1$, $\mathbf{b}_i^* = \mathbf{b}_i - \sum_{j=1}^{i-1} \mu_{i,j} \mathbf{b}_j^*$, $\mu_{i,j} = \frac{\langle \mathbf{b}_i, \mathbf{b}_j^* \rangle}{\|\mathbf{b}_j^*\|^2}$.
- $\det(L[\mathbf{b}_i]) = \prod \|\mathbf{b}_i^*\|$ dépend uniquement du réseau.



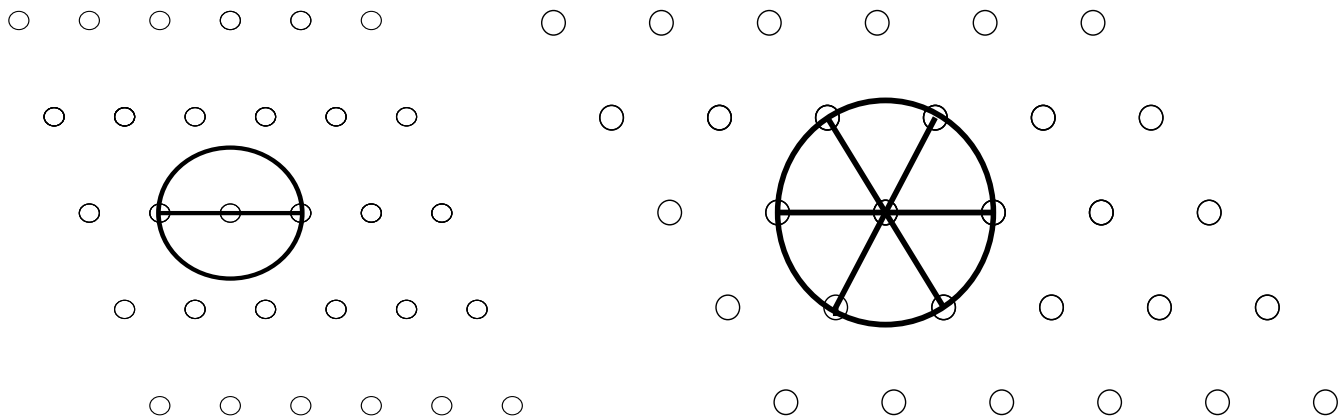
Dans ce qui suit, $\mathbf{b}_i'$ est la projection de $\mathbf{b}_i$ orthogonalement à $\mathbf{b}_1$.

SVP

Étant donnée une base, trouver un vecteur non-nul le plus court.

- La longueur des solutions est le minimum du réseau : λ.
- Le nombre de solutions est le kissing number.

  Il peut croître exponentiellement avec la dimension.



SVP est un problème difficile

- 1981 : Conjecturé NP-difficile (Van Emde Boas).
- 1996 : Prouvé NP-difficile sous des réductions probabilistes (Ajtai).
- 2007 : Toujours NP-difficile sous des réductions probabilistes avec un facteur d'affaiblissement $2^{(\log d)^{1-\varepsilon}}$ (Haviv et Regev).

SVP est un problème très étudié

- 1996 : SVP n'est vraisemblablement plus NP-difficile avec un affaiblissement $\frac{\sqrt{d}}{\log d}$ (Goldreich et Godwasser).
- 1996 : SVP est équivalent dans les cas moyen et le pire à partir d'un certain affaiblissement polynomial (Ajtai).
- 1988 et 2001 : SVP devient polynomial avec un affaiblissement $2^{\frac{\log\log d}{\log d}d}$ (Schnorr; Ajtai, Kumar et Sivakumar).

Pourquoi tant d'efforts autour de SVP?

- On sait dire beaucoup de choses :
  structures algébrique et topologique très riches.
- Des cryptosystèmes reposent sur des modifications de SVP :
  Ajtai-Dwork, GGH, NTRU.
- Les meilleures attaques contre ceux-ci utilisent l'algorithme de réduction de Schnorr par blocs. C'est un mélange entre LLL et l'énumération de vecteurs les plus courts.
- De nombreux autres problèmes de réseaux reposent sur l'énumération : réduction HKZ, kissing number, séries théta, problème du vecteur le plus proche, etc.

Le théorème de Minkowski

Si L est un réseau de dimension d, il existe un vecteur $\mathbf{b}$ non nul de L tel que:

$$\|\mathbf{b}\| \leq \sqrt{d} \cdot (\det L)^{1/d}.$$

La preuve classique utilise le principe des tiroirs.

La réduction de réseaux

- Problème de représentation.
- LLL : résout SVP affaibli d'un facteur exponentiel.
- Minkowski : chaque $\mathbf{b}_i$ est un vecteur plus court parmi ceux pour lesquels $(\mathbf{b}_1, \ldots, \mathbf{b}_i)$ peut être étendu en une base.
- HKZ (Hermite-Korkine-Zolotarev) :
  $\mathbf{b}_1$ est un plus court vecteur et les $\mathbf{b}'_i$ sont HKZ-réduits.

HKZ est nettement plus populaire que celle de Minkowski.

Les algorithmes résolvant SVP

- Fincke-Pohst (’83) : énumération de points entiers dans des ellipsoïdes, après une réduction LLL.
- Kannan (’83), Helfrich (’85) : énumération de points entiers dans des parallélépipèdes, après une réduction quasi-HKZ.
- Ajtai-Kumar-Sivakumar (’01) : interprétation algorithmique de la preuve classique du théorème de Minkowski.

Les algorithmes résolvant SVP

	FP	KH	AKS
	déterministe	déterministe	probabiliste
Temps	$\left(2^{O(d^2)}\right)$	$d^{d/2}$	$2^{O(d)}$
Espace	polynomial	polynomial	$2^{O(d)}$

L'algorithme AKS est voué à perdre en pratique.

Contribution

- Meilleure borne de complexité pour l'algorithme FPKH:

$$d^{\frac{d}{2}+o(d)} \longrightarrow d^{\frac{d}{2e}+o(d)} \approx d^{0.184 \cdot d}$$

- L'analyse prend en compte le fait que l'on énumère des points entiers dans des ellipsoïdes.
- Meilleure compréhension de l'algorithme?

2) Algorithmes d’énumération

Le principe de base

On a $\mathbf{b}_1, \ldots, \mathbf{b}_d, A$ et on veut tous les entiers x_i's tels que :

$$\|x_1\mathbf{b}_1 + \ldots + x_d\mathbf{b}_d\|^2 \leq A^2$$

Si on regarde les composantes sur les vecteurs de Gram-Schmidt :

$$\begin{aligned}
\|x_d\mathbf{b}_d^*\|^2 &\leq A^2 \\
\|(x_{d-1} + \mu_{d,d-1}x_d)\mathbf{b}_{d-1}^*\|^2 + \|x_d\mathbf{b}_d^*\|^2 &\leq A^2 \\
&\ldots \\
\sum_{j\geq i} \|(x_j + \sum_{k>j} \mu_{k,j}x_k)\mathbf{b}_j^*\|^2 &\leq A^2
\end{aligned}$$

L'algorithme de Fincke-Pohst

Les équations précédentes peuvent être réécrites :

$$\begin{aligned}
|x_d|^2 \|\mathbf{b}_d^*\|^2 &\leq A^2 \\
|x_{d-1} + \mu_{d,d-1} x_d|^2 \|\mathbf{b}_{d-1}^*\|^2 &\leq A^2 - |x_d|^2 \|\mathbf{b}_d^*\|^2 \\
&\cdots \\
|x_i + \sum_{k>i} \mu_{k,i} x_k|^2 \|\mathbf{b}_i^*\|^2 &\leq A^2 - \sum_{j>i} (x_j + \sum_{k>j} \mu_{k,j} x_k)^2 \|\mathbf{b}_j^*\|^2
\end{aligned}$$

Cela revient à énumérer des points entiers dans un ellipsoïde défini par la forme quadratique :

$$(y_1, \ldots, y_d) \rightarrow \| \sum y_i \mathbf{b}_i \|^2.$$

L'analyse de Kannan-Helfrich

Simplifions l'énumération :

$$\begin{aligned} |x_d| &\leq A/\|\mathbf{b}_d^*\| \\ |x_{d-1} + \mu_{d,d-1} x_d| &\leq A/\|\mathbf{b}_{d-1}^*\| \\ &\cdots \\ |x_i + \sum_{k>i} \mu_{k,i} x_k| &\leq A/\|\mathbf{b}_i^*\| \end{aligned}$$

Désormais, nous énumérons dans un parallélépipède.

Le nombre de points considérés est : $2^{O(d)} \frac{A^d}{\prod \|\mathbf{b}_i^*\|} = 2^{O(d)} \frac{A^d}{\det(L)}$.

Avec LLL-réduite et $A = \|\mathbf{b}_1\|$, cela donne $2^{O(d^2)}$.

L'algorithme de Kannan

L'énumération est tellement coûteuse qu'on peut se permettre de pré-calculer énormément.

Entrée : Une base $(\mathbf{b}_1, \ldots, \mathbf{b}_d)$ d'un réseau L.

Sortie : Une base HKZ-réduite de L.

1. HKZ-reduire les $\mathbf{b}'_i$.
2. Étendre les $\mathbf{b}'_i$ en des $\mathbf{b}_i$.
3. Si $\|\mathbf{b}_1^*\| > 2\|\mathbf{b}_2^*\|$, échanger $\mathbf{b}_1$ et $\mathbf{b}_2$ et retourner à l'étape 1.
4. Énumérer tous les vecteurs plus courts que $\mathbf{b}_1$, et garder le plus court d'entre eux.
5. Étendre ce vecteur en une base $(\mathbf{b}_1, \ldots, \mathbf{b}_d)$, et HKZ-réduire les $\mathbf{b}'_i$.

Analyse de Helfrich

- L'énumération est si chère qu'elle domine tout le reste.
- Le nombre de points considérés dans l'énumération donne la complexité.
- L'énumération est effectuée sur une base quasi-HKZ-réduite : $\|\mathbf{b}_1^*\| \leq 2\|\mathbf{b}_2^*\|$ et les $\mathbf{b}'_i$ sont HKZ-réduits.
- Pour une telle base : $\frac{\|\mathbf{b}_1\|^d}{\det(L)} \leq 2^{O(d)} \frac{\|\mathbf{b}_2^*\|^{d-1}}{\det(L')} \leq (\sqrt{d})^{d+o(d)}$.

Et en pratique? Faut-il faire confiance à l'asymptotique?

NTL et MAGMA implantent la variante de Fincke-Pohst (ou plutôt Schnorr-Euchner '94) en ne partant que d'une base LLL-réduite.

Ce n'est certainement pas optimal :

d	40	45	50
Le `Minimum` de MAGMA	19.630	130.890	_
Énumération personnelle après LLL	2.290	6.660	399.110
Énumération personnelle après HKZ	0.150	0.860	6.130

3) Une borne de complexité améliorée

Points entiers dans des ellipsoïdes

À un facteur exponentiel près, le coût de l'énumération est borné par :

$$\sum_i \left| \left\{ x_i, \ldots, x_d \in \mathbb{Z}, \sum_{j \geq i} (x_j + \sum_{k>j} \mu_{k,j} x_k)^2 \|\mathbf{b}_j^*\|^2 \leq A^2 \right\} \right|.$$

Avec un changement de variable et l'inégalité $(x+\varepsilon)^2 \geq x^2/4$ pour $x \in \mathbb{Z}$ et $|\varepsilon| \leq 1/2$, on obtient la borne :

$$\sum_i \left| \left\{ x_i, \ldots, x_d \in \mathbb{Z}, \sum_{j \geq i} x_j^2 \|\mathbf{b}_j^*\|^2 \leq 4A^2 \right\} \right|.$$

Des ellipsoïdes aux fonctions théta

Soit $N_i = \left|\left\{x_i, \ldots, x_d \in \mathbb{Z}, \sum_{j \geq i} x_j^2 \|\mathbf{b}_j^*\|^2 \leq A^2\right\}\right|$.

Une astuce dûe à Mazo et Odlyzko ('90) donne :

$$N_i e^{-d} \leq \sum_{x_i,\ldots,x_d \in \mathbb{Z}} \exp\left(-\sum_{j \geq i} x_j^2 \frac{d\|\mathbf{b}_j^*\|^2}{A^2}\right)$$

$$\leq \prod_{j \geq i} \sum_{x \in \mathbb{Z}} \exp\left(-x^2 \frac{d\|\mathbf{b}_j^*\|^2}{A^2}\right) = \prod_{j \geq i} \Theta\left(\frac{d\|\mathbf{b}_j^*\|^2}{A^2}\right),$$

avec $\Theta(t) = \sum_{x \in \mathbb{Z}} \exp(-tx^2)$.

Des fonctions théta à des sous-produits du déterminant

Puisque $\Theta(t) = O\left(\max\left(1, \frac{1}{\sqrt{t}}\right)\right)$, le coût est borné par :

$$2^{O(d)} \cdot \sum_{i \leq d} \prod_{j, \|\mathbf{b}_j^*\| \leq \frac{A}{\sqrt{d}}} \left(\frac{A}{\sqrt{d}\|\mathbf{b}_j^*\|}\right).$$

On montre que pour toute base HKZ-réduite $(\mathbf{b}_1, \ldots, \mathbf{b}_d)$, et tout $I \subset [1, d]$,

$$\frac{\|\mathbf{b}_1\|^{|I|}}{\sqrt{d}^{|I|} \cdot \prod_{i \in I} \|\mathbf{b}_i^*\|} \leq d^{\frac{d}{2e}}.$$

Utilisation exclusive du théorème de Minkowski.

Un cas simple : le lemme de Schnorr

- Supposons que les $\|\mathbf{b}_i^*\|$'s soient décroissants. Alors I est de la forme $[k+1, d]$.
- Schnorr ('87) a montré :

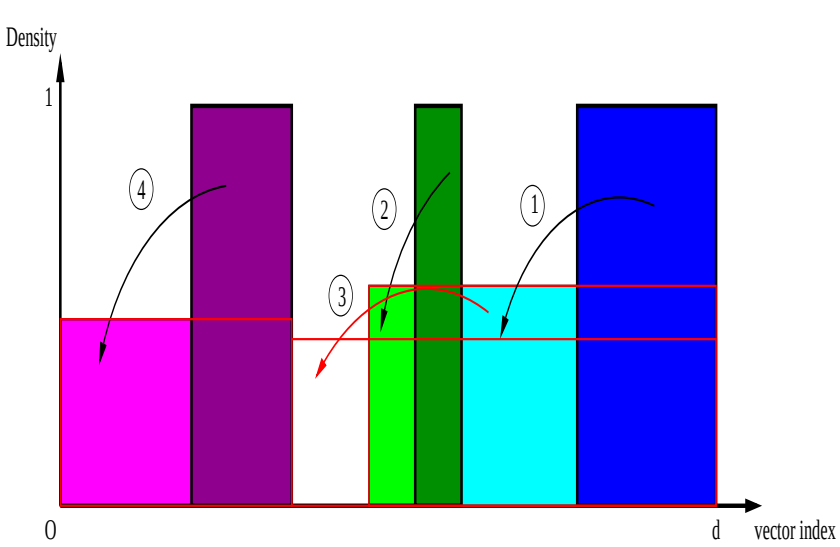
$$\prod_{i>k} \|\mathbf{b}_i^*\| \geq \sqrt{d}^{(d-k)\log\frac{d-k}{d}} \cdot (\det(L))^{\frac{d-k}{d}}$$

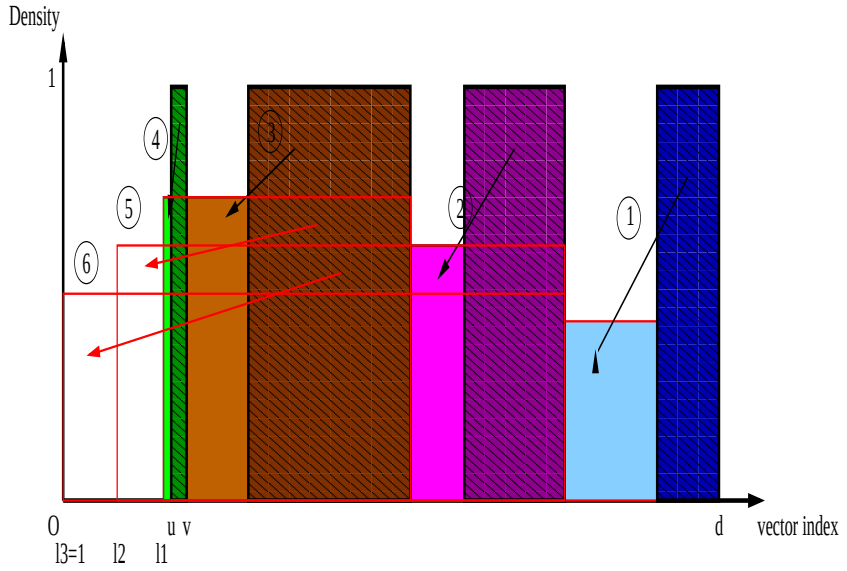
- La complexité est ainsi bornée par :

$$\left(\frac{\|\mathbf{b}_1\|}{\sqrt{d}\cdot\det(L)}\right)^{d-k} \cdot (\sqrt{d})^{-(d-k)\log\frac{d-k}{d}} \leq d^{d/2e}$$

Les difficultés du cas général

- La base à partir de laquelle on part n'est pas HKZ-réduite mais seulement quasi-HKZ-réduite.
- L'ensemble I peut être significativement plus complexe.





4) Problèmes connexes et questions ouvertes

Borne inférieure (en cours)

À l'aide d'une technique introduite par Ajtai ('03), nous pouvons construire une base aléatoire telle que :

- elle est HKZ-réduite avec probabilité proche de 1,
- l'énumération nécessite au moins $d^{a \cdot d + o(d)}$ opérations, pour un certain $a > 0$.

$$\begin{aligned} \log r_{i,i} &= c \sum_{j=1}^{d-i+1} \frac{\log j}{j} \\ \mu_{i,j} &\in_{r,u,i} [-1/2, 1/2) \end{aligned}$$

Problème connexe : CVP

- CVP : étant donnés une base et un vecteur cible, trouver un vecteur du réseau le plus proche de la cible.
- Notre technique s'étend facilement pour faire décroître la complexité de Kannan-Helfrich de $d^{d+o(d)}$ à $d^{\frac{d}{2}+o(d)}$.

Problème connexe : BKZ

- En 1987, Schnorr décrit une famille d'algorithmes de réduction par blocs : chaque bloc est HKZ-réduit, et la gestion des blocs entre eux est similaire à LLL.
- Ça permet d'obtenir des bases de bonne qualité en de très grandes dimensions.
- Coût et qualité dépendent de la taille des blocs.

À coût donné, nous améliorons la borne sur la qualité atteignable, dans le cas le pire.

Arithmétique sous-jacente (en cours)

- Pour le moment, pour garantir la correction, on utilise une orthogonalisation de Gram-Schmidt rationnelle.
- En pratique, on utilise des approximations flottantes ...
- But : prouver qu'en effet on a le droit de le faire (quitte à modifier un peu l'algorithme).

Autres problèmes ouverts

- HKZ est-elle la bonne réduction à utiliser avant d'énumérer?
- Peut-on résoudre SVP en temps déterministe $2^{O(d)}$?
- Comment faut-il paralléliser l'énumération?
- Comment minimiser le nombre d'énumérations dans BKZ?