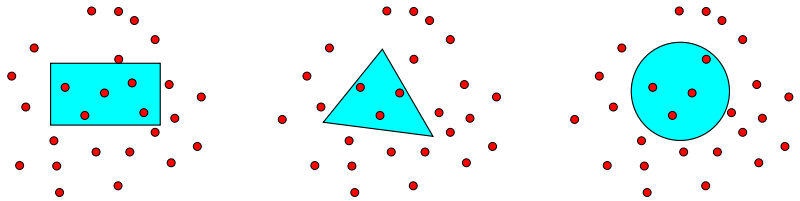
Geometric Range Searching



1. $S \subset \mathbb{R}^d$ (n points) et $\mathcal{R} \subset \mathcal{P}(\mathbb{R}^d)$ (*domaines*– ranges).
2. Problème :

Compter (reporter) les points de S situés dans le domaine R.

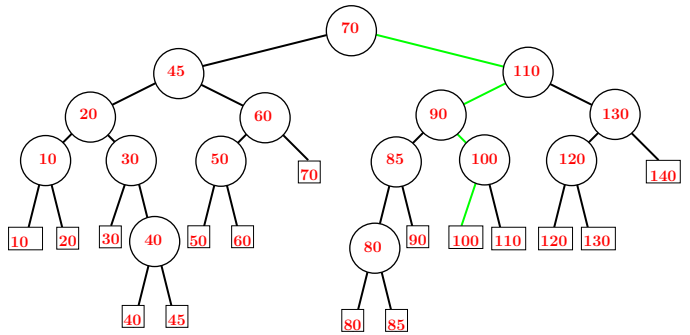
RECHERCHE MULTIDIMENSIONNELLE

1. $S \subset \mathbb{R}^d$ (n points) et $\mathcal{R} \subset \mathcal{P}(\mathbb{R}^d)$ (*domaines*– ranges).
2. Le problème est celui d'organiser S en une structure de données de telle sorte à pouvoir répondre efficacement à une requête du type:

Compter (reporter) les points de S situés dans le domaine R.

3. En général la solution consiste à définir une famille de sous-ensembles "canoniques" de S (phase de prétraitement), puis à chercher dans cette famille une partition de l'ensemble des points à reporter $R \cap S$, dont on extrait l'information requise (phase de réponse à la requête).
4. Les ensembles canoniques sont en général organisés en une structure d'arbre binaire (à plusieurs niveaux) dans laquelle on effectue une recherche guidée par des informations auxilliaires.

ARBRES BINAIRES DE RECHERCHE



Soit S un ensemble fini de n clés d'un univers totalement ordonné. Un arbre binaire de recherche sur S est un arbre binaire de taille $n-1$ dont les noeuds externes et internes sont étiquetés par les clés de S de telle sorte que (1) la clé d'un noeud interne coincide avec la clé de la feuille la plus à droite de son sous-arbre gauche et (2) un parcours infixe de l'arbre énumére les clés des noeuds internes (externes) dans l'ordre croissant.

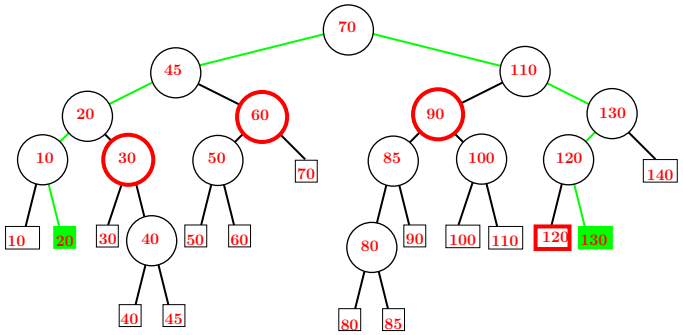
CHEMIN DE RECHERCHE D'UNE CLÉ

procedure $search(key : typekey;$ **var** $t : tree);$
0. **begin**
1. **if** $t =$ **nil then** $notfound(key)$
2. **else if** $t\uparrow .k = key$ **then** $found(t\uparrow)$
3. **else if** $t\uparrow .k < key$ **then** $search(key, t\uparrow .right)$
4. **else** $search(key, t\uparrow .left)$
5. **end;**

$S[x, x'] = \{k \in S \mid x \leq k \leq x'\}$

$$S[x, x'] = \bigoplus_{\nu} S(\nu)$$

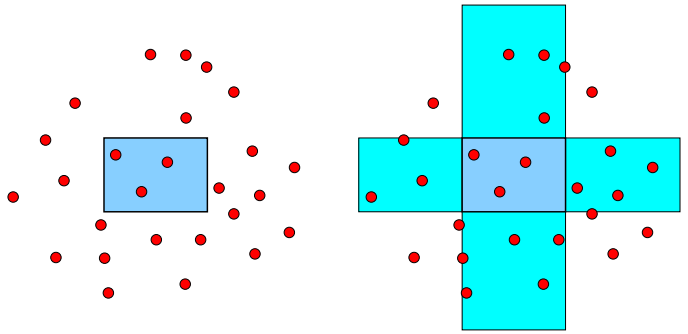
où ν décrit l'ensemble des racines des sous-arbres pendant respectivement à gauche et à droite des chemins de recherche de x et x' à partir de leur premier ancêtre commun.



$$\begin{aligned}\{K \mid 20 < K < 130\} &= S(30) \cup S(60) \cup S(90) \cup S_{\square}(120) \\ &= \{30, 40, 45\} \cup \{50, 60, 70\} \cup \{80, 85, 90, 100, 110\} \cup \{120\}\end{aligned}$$

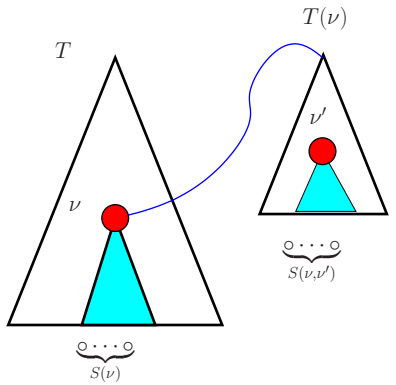
NOTES

ORTHOGONAL RANGE SEARCHING



1. S un ensemble de n points du plan
2. $R = [a_1, a_2] \times [b_1, b_2]$ un domaine requête. $S \cap R =$??
3. $R_1 = [a_1, a_2] \times \mathbb{R}$ et $R_2 = \mathbb{R} \times [b_1, b_2]$.
4. $S \cap R = (S \cap R_1) \cap R_2$

RANGE TREES

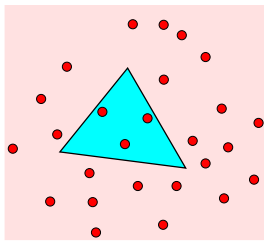


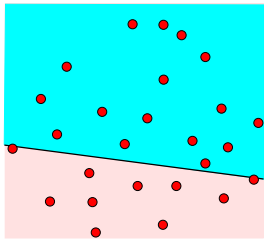
$$\begin{aligned}\sum_{\nu \in T} \#T(\nu) &= \sum_{\nu \in T} \#S(\nu) \\ &= \sum_{p \in S} \#\mathrm{CheRech}(T, p) \\ &= O(n \log n).\end{aligned}$$

$$(S \cap R_1) \cap R_2 = \bigoplus_{\nu} S(\nu) \cap R_2 = \bigoplus_{\nu} \bigoplus_{\nu'} S(\nu, \nu')$$

où ν décrit $\mathrm{CheRech}(T, [a_1, a_2])$ et ν' décrit $\mathrm{CheRech}(T(\nu), [b_1, b_2])$.

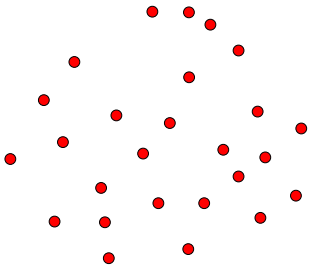
RECHERCHE SIMPLICIALE

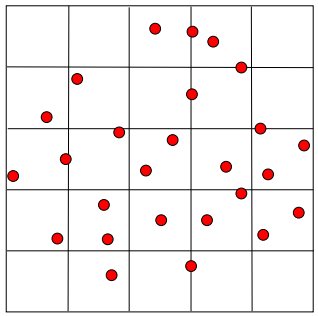




– via les méthodes de grille
– via les équipartitions
– via les partitions aléatoires (théorie des ϵ-nets)
– via les partitions simpliciales (cuttings)

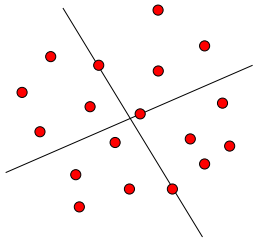
GRID METHODS – MÉTHODES DE GRILLE

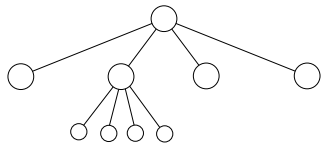




– V. Gaede and O. Günther. Multidimensional Acces Methods. ACM Computing Surveys. pages 170–231, 30(2),1998
– Samet and co

2D EQUIPARTITIONS WITH TWO LINES



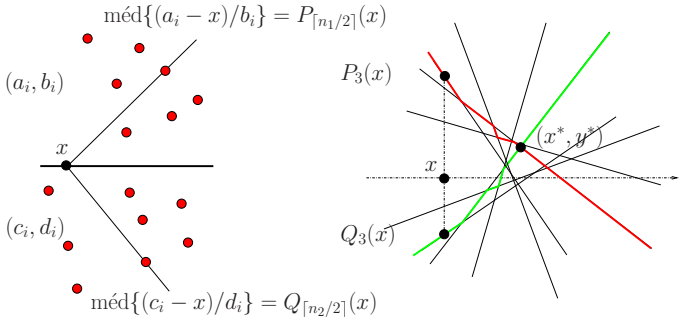


– Définition.
– Temps de requête pour un demi-plan:

$$T(n) = O(1) + 3T(n/4) = O(n^{\log_4 3}) \approx O(n^{0.7924812507}).$$

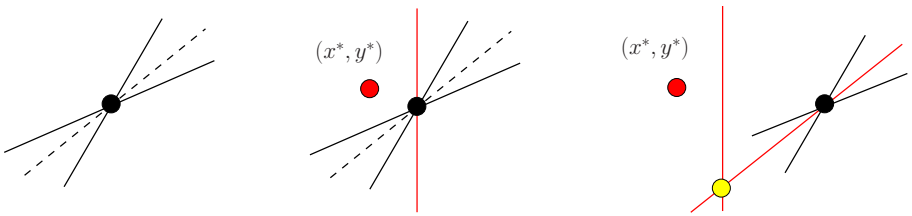
– Temps de requête pour un triangle.
– Nimrod Meggido. Partitionning with Two Lines in the Plane. *J. Algorithms.* 6: 430–433, 1985.

2D EQUIPARTITIONS (EXISTENCE)



– Soit $P_k(x)$ la k-ième valeur des $\{A_i x + B_i\}$ avec $A_i = -1/b_i < 0$ et $B_i = a_i/b_i$
– Soit $Q_{k'}(x)$ la k'-ième valeur des $\{C_i x + D_i\}$ avec $C_i = -1/d_i > 0$ et $D_i = c_i/d_i$
– Requête de position : la position de (x^*, y^*) par rapport à n'importe quelle droite est décidable en temps $O(n)$.

2D EQUIPARTITIONS (CONSTRUCTION)

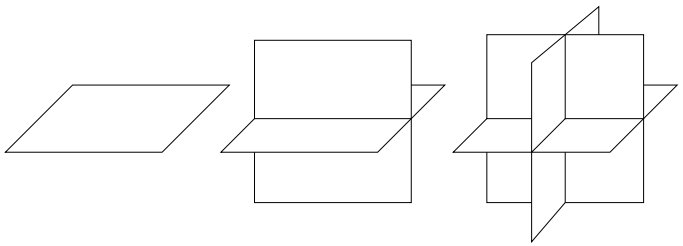


– Droites actives

– Itérer: pente médiane, couplage, abscisse médian, requête, intercept médian, requête, élimination.

– Chaque itération désactive 1/8 des droites

3D EQUIPARTITIONS WITH THREE PLANES



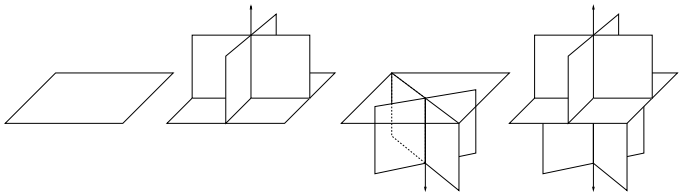
❑ — Est-il possible de partitionner une distribution de points de $\mathbb{R}^d$ en 2^d parts égales avec d hyperplans? Oui en dimension 3 (thanks to the Borsuk-Ulam Theorem), non en dimension ≥ 5, ouvert en dimension 4.

❑ —

Théorème 1 (Borsuk-Ulam Theorem)**.** *For every continuous mapping $f : \mathbb{S}^d \to \mathbb{R}^d$ there exists a point $x \in \mathbb{S}^d$ with $f(x) = f(-x)$.* □

❑ — F. F. Yao, D. P. Dobkin, H. Edelsbrunner, and M. S. Paterson. Partitioning space for range queries. *SIAM J. Comput.*, 18(2):371–384, 1989.

dD EQUIPARTITIONS VIA PROJECTIVE PARTITIONS



– Definition;

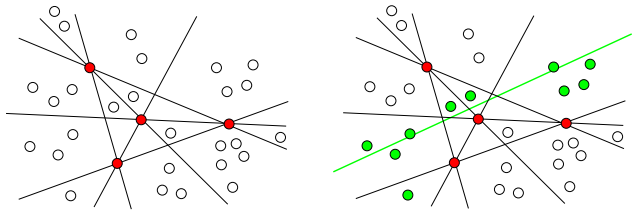
– Temps de requête: $T(n) = O(1) + (2^d - 1)T(n/2^d) = O(n^{\log_{2^d}(2^d-1)})$;

– Existence (thanks to the Borsuk-Ulam Theorem)

– A. C. Yao and F. F. Yao. A general approach to D-dimensional geometric queries. In *Proc. 17th Annu. ACM Sympos. Theory Comput.*, pages 163–168, 1985.

– R. Cole. *Partitionning point sets in arbitrary dimensions.* In Theoret. Comput. Sci. 49 (1987), pp 239–266.

RANDOMIZED PARTITIONS



– $X \subset \mathbb{R}^2$, $|X| = n$, generic position;

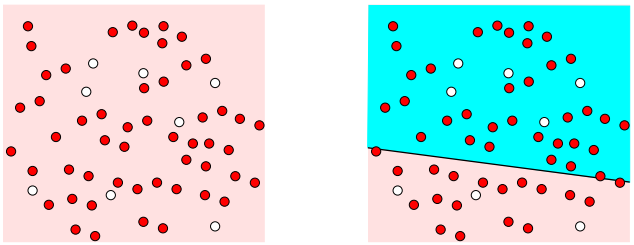
– Let R be a random r-tuple of points of X. Then the maximal number of points of $X \setminus R$ lying in the zone of a line in the arrangement of the $\binom{r}{2}$ lines determined by the points of R is $\leq \epsilon n$ with probability $\geq 1/2$ for $r = c\frac{1}{\epsilon}\log(\frac{1}{\epsilon})$.

– Partition tree with $r = O(1)$, query time $O(n^{2/3+\gamma})$

– Generalize in dimension d with query time $O(n^\alpha)$ with $\alpha = \frac{d(d-1)}{d(d-1)+1} + \gamma$.

– Haussler and Welzl. Epsilon-nets and simplex ranges queries. *Discrete Comput. Geom. 2*, 127-151, 1987.

EPSILON-NETS

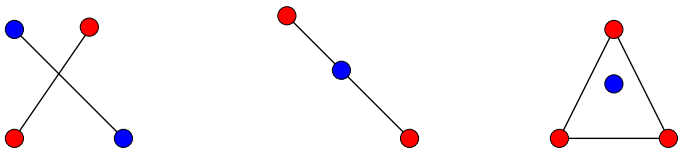


Avec probabilité $\geq 1/2$ (?) un échantillon aléatoire de 5355 points d'un ensemble de n points est tel que tout demi-plan contenant 1% des n points contient un point de l'échantillon. Un tel échantillon est appelé un 0.01-net de l'ensemble des n points pour l'hypergraphe des demi-plans.

Définition 1. *Soient $H = (X, R)$ un hypergraphe, A un sous-ensemble fini de X et $0 \leq \epsilon \leq 1$. Un sous-ensemble N de A est appelé un ϵ-net de A pour H si*

$$\forall r \in R, \qquad |r \cap A| \geq \epsilon |A| \implies r \cap N \neq \emptyset.$$

DIMENSION DE VAPNIK-CHERVONENKIS D'UN HYPERGRAPHE



Définition 2. *Soit $H = (X, R)$ un hypergraphe. Un sous-ensemble fini A de X est dit H-séparable si $R_{|A} = 2^A$ où $R_{|A} = \{r \cap A \mid r \in R\}$.* □

Définition 3. *La dimension de Vapnik-Chervonenkis d'un hypergraphe $H = (X, R)$ est la borne supérieure de l'ensemble des cardinaux des sous-ensembles finis H-séparables de X.* □

Théorème 2. *La dimension de l'hypergraphe des demi-espaces (fermés) de $\mathbb{R}^d$ est $d+1$.* □

Théorème 3. *La dimension de l'hypergraphe des $\{x \in \mathbb{R}^d \mid p(x) \geq 0\}$ où p décrit l'ensemble des polynomes à d variables de degré au plus D est de dimension bornée par $\binom{d+D}{d}$.* □

L'hypergraphe des droites sécantes à une demi-droite est de VC-dimension finie.

DIMENSION D'UN HYPERGRAPHE

Théorème 4. *Soit (X, R) un hypergraphe de dimension au plus d et soit $A \subseteq X$ de cardinalité n. Alors*

$$|R_{|A}| \leq \Phi_d(n) = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{d} \leq (\frac{en}{d})^d.$$

□

Théorème 5. *Soit (X, R) un hypergraphe de dimension d et soit $R' = \{F(r_1, r_2, \ldots, r_k) \mid r_i \in R\}$ où F est une expression ensembliste utilisant les opérations d'union, intersection et différence. Alors la dimension de l'hypergraphe (X, R') est un $O(kd \ln k)$.* □

EPSILON-NETS

Théorème 6 (Epsilon-Net Theorem)**.** *Soit $H = (X, R)$ un hypergraphe de dimension fini d, A un sous-ensemble fini de X, $0 < \epsilon, \delta < 1$, et $N \subseteq A$ un échantillon aléatoire de A obtenu à la suite de $m \geq 8/\epsilon$ tirages uniformes avec remise d'un élément de A. Alors N est un ϵ-net de A avec probabilité supérieure à*

$$1 - 2\Phi_d(2m)2^{-\epsilon m/2}. \tag{1}$$

En particulier si

$$m \geq \max\left(\frac{4}{\epsilon}\log\frac{2}{\delta}, \frac{8d}{\epsilon}\log\frac{8d}{\epsilon}\right), \tag{2}$$

alors N est un ϵ-net de A avec probabilité $> 1 - \delta$. □

Un simple calcul (—) montre que N est un ϵ-net avec probabilité $\geq 1 - \Phi_d(|A|)e^{-\epsilon m}$; par suite pour $m = \Omega(\frac{\log|A|}{\epsilon})$ l'échantillon N est un ϵ-net avec probabilité $\neq 0$. La difficulté est d'obtenir une borne indépendante de la taille de A.

PREUVE

Soit (X, R) un hypergraphe fini de dimension au plus d tel que $|r| \geq \epsilon|X|$ pour tout $r \in R$ où ϵ est un réel strictement positif fixé. Pour $z = z_1 z_2 \dots z_{2m} \in X^{2m}$, $r \in R$, posons

$$\begin{aligned} l_1(z,r) &= \#\{i \leq m \mid z_i \in r\}, \\ l_2(z,r) &= \#\{i \leq 2m \mid z_i \in r\} \end{aligned}$$

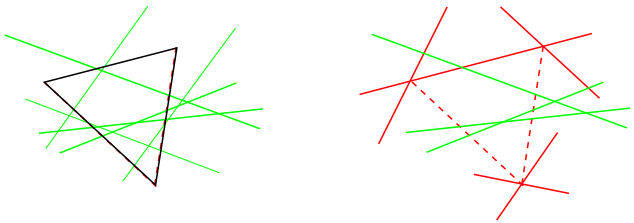
et considérons les ensembles

$$\begin{aligned} N(z) &= \{z_1, z_2, \dots, z_m\} \\ Q &= \{\, z \in X^{2m} \mid \ N(z) \text{ n'est pas un } \epsilon\text{-net, i.e. } \exists r \in R,\, l_1(z,r) = 0 \,\} \\ J(r) &= \{\, z \in X^{2m} \mid l_1(z,r) = 0, l_2(z,r) \geq \epsilon m/2\} \\ J &= \bigcup_{r \in R} J(r). \end{aligned}$$

Lemme 6.1. $\mathrm{Prob}(z \in Q) \leq 2\,\mathrm{Prob}(z \in J)$ *pour* $m \geq 8/\epsilon$. □

Lemme 6.2. $\mathrm{Prob}(z \in J) \leq \Phi_d(2m) 2^{-\epsilon m/2}$. □

HYPERGRAPH OF SETS OF LINES CUTTING A TRIANGLE

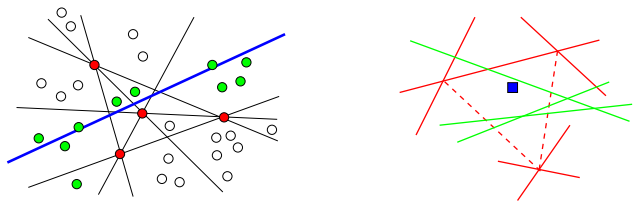


– L'hypergraphe des ensembles de droites sécantes à un triangle est de dimension finie.

–

Théorème 7. *Soient X un ensemble fini de n droites du plan en position générale, ϵ un réel de l'intervalle ouvert $(0,1)$, et N un échantillon aléatoire de X obtenu à la suite de m tirages uniformes avec remise d'un élément de X. Alors le nombre de droites de X sécantes à l'intérieur d'un triangle quelconque de la triangulation "canonique" de l'arrangement des droites de N est $\leq \epsilon n$ avec probabilité $\geq 1/2$ dès que $m \geq c\frac{1}{\epsilon}\log\frac{1}{\epsilon}$.* □

RANDOMIZED PARTITIONS : A JUSTIFICATION



– Les droites duales des points de $X \setminus R$ appartenant à la zone de la droite h dans l'arrangement des $\binom{r}{2}$ droites passant par deux points de R sont sécantes au triangle de la triangulation canonique de l'arrangement des droites duales des points de R contenant le point dual de la droite h.

BIBLIOGRAPHIE

[1] J. Matoušek. Geometric range searching. *ACM Comput. Surv.*, 26:421–461, 1994.

[2] K. Mulmuley. *Computational Geometry : An Introduction Through Randomized Algorithms*. Prentice-Hall, 1994.

[3] J. Matoušek. *Lectures on Discrete Geometry*. Number 212 in Graduate texts in Mathematics. Springer-Verlag, 2002.

[4] J. Matousek. *Using the Borsuk-Ulam Theorem - Lectures on Topological Methods in Combinatorics and Geometry*. Universitext. Springer-Verlag, 2003.

[5] J. Pach and P. K. Agarwal. *Combinatorial Geometry*. John Wiley & Sons, New York, NY, 1995.

[6] K. L. Clarkson and P. W. Shor. Applications of random sampling in computational geometry, II. *Discrete Comput. Geom.*, 4:387–421, 1989.

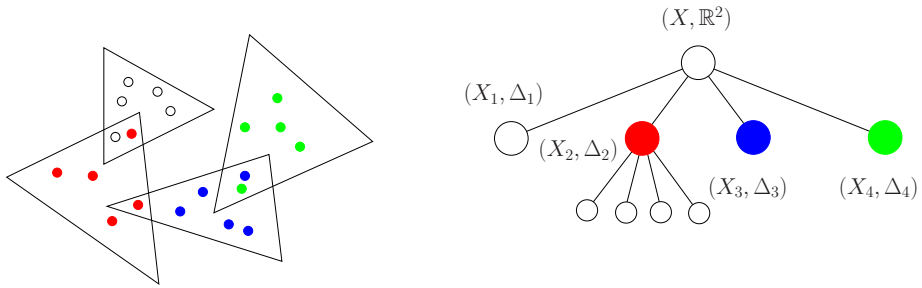
NOTES

NOTES

NOTES

NOTES

PARTITION SIMPLICIALE – ARBRE DE PARTITION



1. $\{X_1, X_2, \ldots, X_r\}$ est une partition de X et Δ_i est un simplexe contenant X_i
2. module = le nombre maximal de simplexes sécants à une droite
3. taille $= r$
4. ratio $= \max_i |X_i|$

$$T_{\text{ime}}(n) \;=\; O(\text{taille}) + O(\text{module}) T_{\text{ime}}(\text{ratio})$$

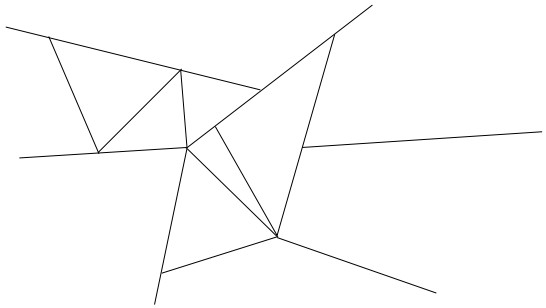
EXISTENCE & CONSTRUCTION

Théorème 8. *Soit X un ensemble de n points du plan en position générale et soit un paramètre $1 \leq r \leq n$. Alors il existe une partition simpliciale de X de taille r, de ratio $2n/r$ et de module $O(\sqrt{r})$. De plus une telle partition est constructible en temps $O(n \log r)$.*

- Chazelle, Matousek, Welzl, etc
- S. Har-Peled. Constructing planar cuttings in theory and practice. In *SIAM J. Comput.*, 29: 2016–2039, 2000.

$$\begin{aligned} T_{\text{ime}}(n) \;&=\; O(r) + O(\sqrt{r}) T_{\text{ime}}(\frac{2n}{r}) \qquad\qquad r = \sqrt{n} \\ &=\; O\left(\sqrt{n}\, 2^{O(\log\log n)}\right) \\ &=\; O\left(\sqrt{n}\, \text{polylog}\; n\right) \end{aligned}$$

CUTTINGS



Théorème 9 ($1/r$**-Cutting de taille** $\ll r^2 \log^2 r$ **).** *Soit X un ensemble de n droites du plan en position générique et soit r un paramètre, $1 \leq r \leq n$. Il existe une partition du plan en $t \ll r^2 \log^2(r)$ "triangles" $\Delta_1, \ldots, \Delta_t$ telle que pour tout Δ_i le nombre de droites de X sécantes à l'intérieur de Δ_i est au plus n/r. Une telle partition du plan est constructible en temps $O(nr \log r)$.* □

CUTTINGS - VERSION RAFFINÉE ET PONDÉRÉE

Théorème 10 ($1/r$**-Cutting de taille** $\ll r^2$ **).** *Soit X un ensemble de n droites du plan en position générique et soit r un paramètre, $1 \leq r \leq n$. Il existe une partition du plan en $t \ll r^2$ "triangles" $\Delta_1, \ldots, \Delta_t$ telle que pour tout Δ_i le nombre de droites de X sécantes à l'intérieur de Δ_i est au plus n/r. Une telle partition du plan est constructible en temps $O(nr)$.* □

Théorème 11 ($1/r$**-Cutting de taille** $\ll r^2$ **(Version pondérée)).** *Soit X un ensemble de n droites du plan en position générique chacune munie d'un poids entier positif $w(h)$, et soit r un paramètre, $1 \leq r \leq n$. Il existe une partition du plan en $t \ll r^2$ triangles $\Delta_1, \ldots, \Delta_t$ telle que*

$$w(\{h \in X \mid h \cap \text{Interior}(\Delta_i) \neq \emptyset\}) \leq \frac{1}{r} w(X)$$

où $w(\mathcal{H}) = \sum_{h \in \mathcal{H}} w(h)$ pour $\mathcal{H} \subseteq X$. □

PREUVE DU THÉO. EXISTENCE DES PARTITIONS SIMPLICIALES

– $X \subset \mathbb{R}^2$ en position générique, $|X| = n$, $1 \leq r \leq n$ un paramètre;

– Si $r = 1$ je pose $\mathbb{P}_r(X) = (X, \mathbb{R}^2)$ et $X' = X$; sinon il existe et je choisis une partition simpliciale $\mathbb{P}_r(X) = \{(X_1, \Delta_1), \ldots, (X_m, \Delta_m)\}$ de taille $m \leq r$, de module $\ll \sqrt{r}$ telle que

(a) $X_i \subseteq X$;

(b) $|X_i| = n/r$;

(c) $|X_1 \cup X_2 \cup \cdots \cup X_m| \geq |X|/2$;

et je pose $X' = X \setminus \bigcup_{i=1}^m X_i$, $n' = |X'|$ et $r'/n' = r/n$.

– $\mathbb{P}_r(X) \cup \mathbb{P}_{r'}(X') \cup \mathbb{P}_{r''}(X'') \cup \cdots$ est une partition simpliciale de X, de taille $\leq 2r$, de module $\ll \sqrt{r}$ et dont toutes les classes contiennent n/r éléments sauf éventuellement la dernière classe qui contient elle au plus n/r éléments.

PREUVE

– $X \subset \mathbb{R}^2$ en position générique, $|X| = n$, $1 \leq r \leq n$ un paramètre; $R \in X^p$ où $p = \Theta(\sqrt{r} \log \sqrt{r})$; $\mathcal{H}$ est l'ensemble des droites passant par deux points de R, $|\mathcal{H}| = O(r \log^2 r)$;

– La suite des (X_i, Δ_i) $(1 \leq i \leq m)$ est définie par induction à partir d'une suite Y_i de sous-ensembles de X de taille $\geq n/2$, d'une suite de fonctions poids $w_i : \mathcal{H} \to \mathbb{N}$ et d'une suite de ϵ-cuttings $\Delta_1^i, \Delta_2^i, \ldots, \Delta_{k(i)}^i$ des $(\mathcal{H}, w_i)$ de tailles $k(i) \leq r/2$ où $1/\epsilon = \Theta(\sqrt{r})$ comme suit :

– $w_1(\ell) = 1$, $Y_1 = X$

– $\Delta_i = \Delta_{j(i)}^i$ où $j(i)$ est l'un des indices $j \in \{1, 2, \ldots, k(i)\}$ tel que $|\Delta_j^i \cap Y_i| \geq n/r$, et X_i est un sous-ensmble quelconque de Y_i de cardinal n/r inclus dans Δ_i;

– Si $Y_i \setminus X_i$ est de taille $\leq n/2$ on pose $m = i$ et $\mathbb{P}_r(X) = \{(X_1, \Delta_1), \ldots, (X_m, \Delta_m)\}$; sinon on pose $Y_{i+1} = Y_i \setminus X_i$ et on définit w_{i+1} à partir de w_i en doublant le poids des droites de $\mathcal{H}$ intersectant l'intérieur de Δ_i.

Théorème 12. *$\mathbb{P}_r(X)$ est une partition simpliciale d'au moins la moitié des points de X, de taille $\leq r$, de module $\ll \sqrt{r}$, et dont chaque classe contient n/r éléments.* □

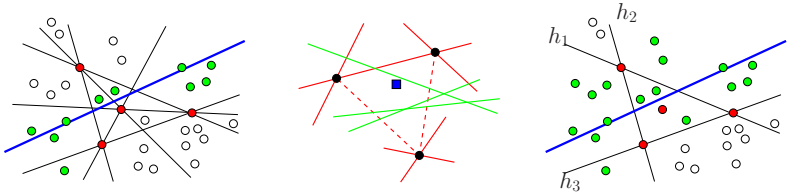
PREUVE

- $h \in \mathcal{H}$, $\rho = \#\{i \mid h \cap \text{Interior}(\Delta_i) \neq \emptyset\}$

$$\begin{aligned} w_{i+1}(\mathcal{H}) &= w_i(\mathcal{H}) + w_i(\{\ell \in \mathcal{H} \mid \ell \cap \text{Interior}(\Delta_i) \neq \emptyset\}) \\ &\leq (1 + \frac{c}{\sqrt{r}}) w_i(\mathcal{H}), \\ w_m(\mathcal{H}) &\leq (1 + \frac{c}{\sqrt{r}})^m w_1(\mathcal{H}). \end{aligned}$$

$$\begin{aligned} w_m(h) = 2^\rho &\leq w_m(\mathcal{H}) \\ &\leq w_1(\mathcal{H})(1 + \frac{c}{\sqrt{r}})^m \end{aligned}$$

PREUVE



– $X \subset \mathbb{R}^2$, $|X| = n$, position générique; et h une droite

– Soit $R \in X^p$ un vecteur aléatoire de $p = \Theta(\sqrt{r} \log \sqrt{r})$ points de X. Nous savons que le nombre de points de $X \setminus R$ situés dans la zone de h dans l'arrangement des droites h_1, h_2, h_3 est $\leq n/\sqrt{r}$ avec probabilité $\geq 1/2$.