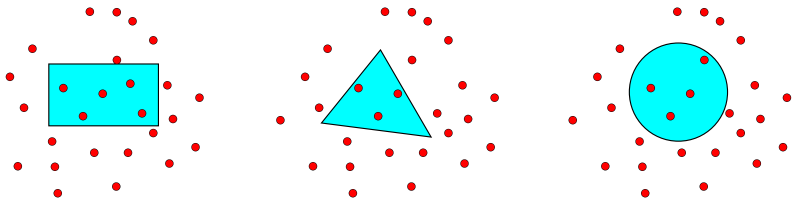
Geometric Range Searching



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.

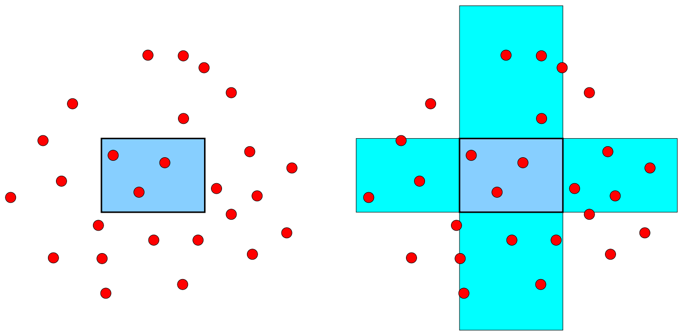
RECHERCHE MULTIDIMENSIONNELLE

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.

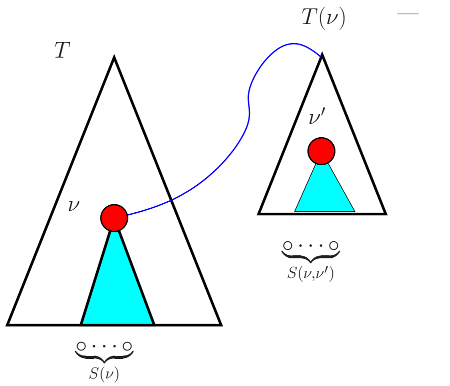
3. En général la solution consiste à définir une famille de sous-ensembles "canoniques" de S (phase de prétraitement), puis à chercher dans cette famille une partition de l'ensemble des points à reporter $R \cap S$, dont on extrait l'information requise (phase de réponse à la requête).
4. Les ensembles 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.

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

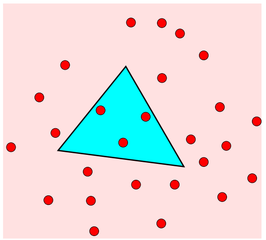


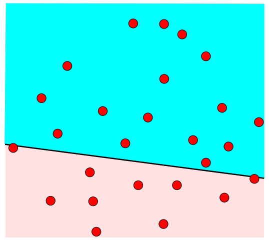
$$\begin{aligned}\sum_{\nu \in T} \#T(\nu) &= \sum_{\nu \in T} \#S(\nu) \\ &= \sum_{p \in S} \#\mathrm{CheRech}(T, p) \\ &= O(n \log n).\end{aligned}$$

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

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

RECHERCHE SIMPLICIALE





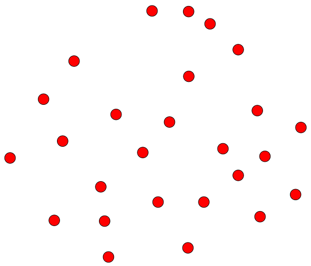
– via les méthodes de grille

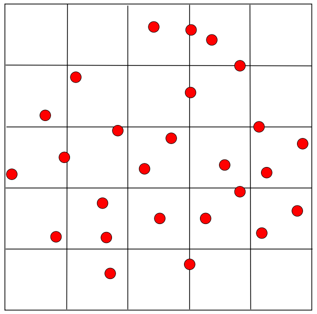
– via les équipartitions

– via les partitions aléatoires (théorie des ϵ-nets)

– via les partitions simpliciales (cuttings)

GRID METHODS – MÉTHODES DE GRILLE

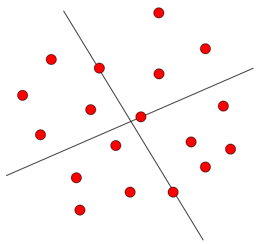


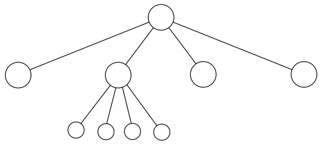


– V. Gaede and O. Günther. Multidimensional Acces Methods. ACM Computing Surveys. pages 170–231, 30(2),1998

– Samet and co

2D EQUIPARTITIONS WITH TWO LINES





– Définition.

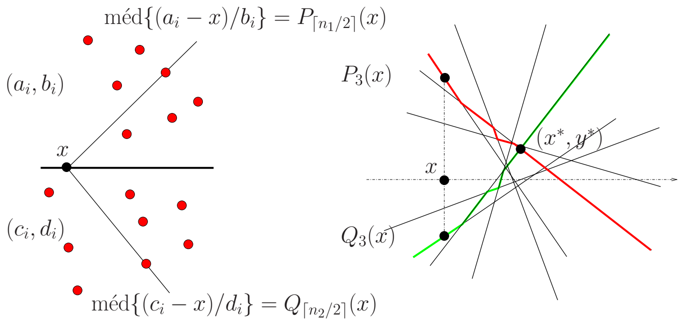
– Temps de requête pour un demi-plan:

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

– Temps de requête pour un triangle.

– Nimrod Meggido. Partitionning with Two Lines in the Plane. *J. Algorithms.* 6: 430–433, 1985.

2D EQUIPARTITIONS (EXISTENCE)

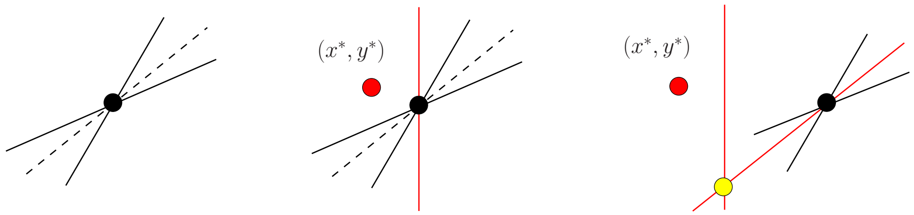


– Soit $P_k(x)$ la k-ième valeur des $\{A_i x + B_i\}$ avec $A_i = -1/b_i < 0$ et $B_i = a_i/b_i$

– 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)

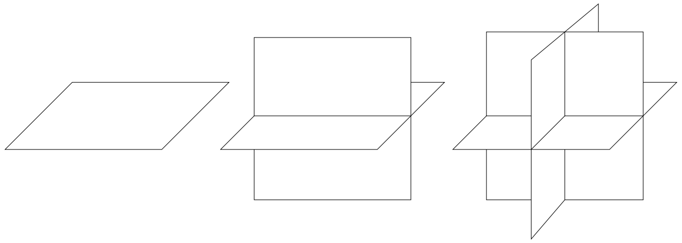


– Droites actives

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

– Chaque itération désactive 1/8 des droites

3D equipartitions with three planes



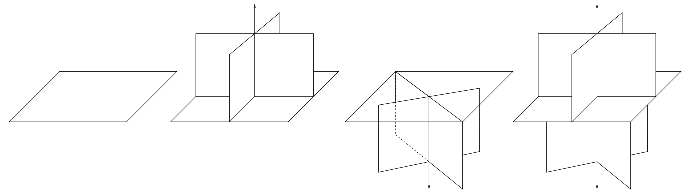
❑ — Est-il possible de partitionner une distribution de points de $\mathbb{R}^d$ en 2^d parts égales avec d hyperplans? Oui en dimension 3 (thanks to the Borsuk-Ulam Theorem), non en dimension ≥ 5, ouvert en dimension 4.

❑ —

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;

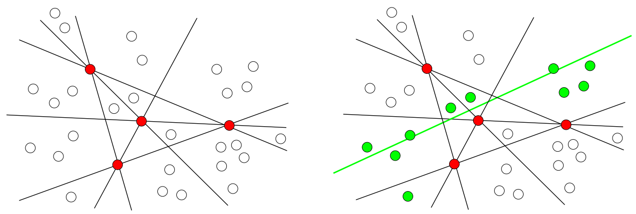
– Temps de requête: $T(n) = O(1) + (2^d - 1)T(n/2^d) = O(n^{\log_{2^d}(2^d-1)})$;

– 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;

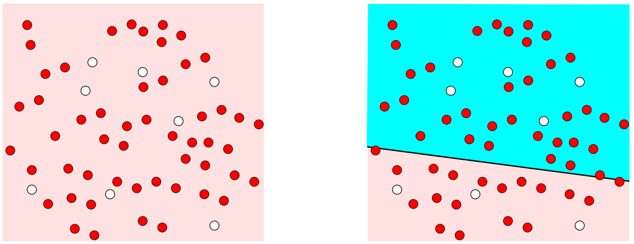
– Let R be a random r-tuple of points of X. Then the maximal number of points of $X \setminus R$ lying in the zone of a line in the arrangement of the $\binom{r}{2}$ lines determined by the points of R is $\leq \epsilon n$ with probability $\geq 1/2$ for $r = c\frac{1}{\epsilon}\log(\frac{1}{\epsilon})$.

– 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

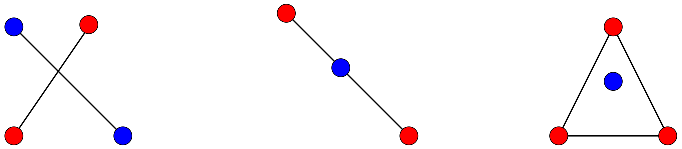


Avec probabilité $\geq 1/2$ (?) un échantillon aléatoire de 5355 points d'un ensemble de n points est tel que tout demi-plan contenant 1% des n points contient un point de l'échantillon. Un tel échantillon est appelé un 0.01-net de l'ensemble des n points pour l'hypergraphe des demi-plans.

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

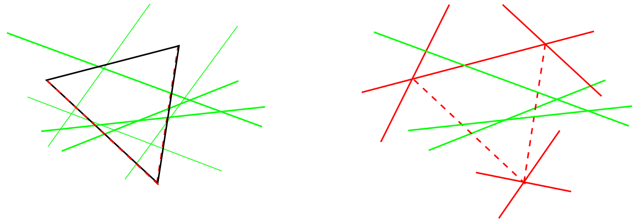
1. Soit (X, R) un hypergraphe fini de dimension au plus d tel que $|r| \geq \epsilon|X|$ pour tout $r \in R$ où ϵ est un réel strictement positif fixé.
2. Pour $z = z_1z_2 \ldots z_{2m} \in X^{2m}$, posons $N(z) = \{z_1, z_2, \ldots, z_m\}$ et considérons les ensembles

$$\begin{aligned} Q &= \{\, z \in X^{2m} \mid \ N(z) \text{ n'est pas un } \epsilon\text{-net, i.e. } \exists r \in R,\ N(z) \cap r = \emptyset \,\} \\ J &= \{\, z \in X^{2m} \mid \exists r \in R,\ N(z) \cap r = \emptyset,\ |\{i > m \mid z_i \in r\}| \geq \epsilon m/2 \,\} \end{aligned}$$

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

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

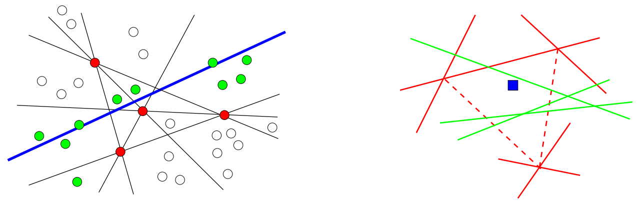
HYPERGRAPH OF SETS OF LINES CUTTING A TRIANGLE



- 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