Réseaux de neurones et traitement de données: la notion de voisinage γ-Observable

Michaël Aupetit[1] Dominique Lepetz[2] Max Nemoz-Gaillard[3]
Pierre Couturier[1] Pierre Massotte[1]

[1]LGI2P - Site EERIE - EMA
Parc Scientifique Georges Besse
F30035 Nîmes, France

[3]CMGD - [2]EMA
6, avenue de Clavières
F30319 Alès, France

e-mail: {Michael.Aupetit, Pierre.Couturier, Pierre.Massotte}@site-eerie.ema.fr
Max.Nemoz-Gaillard@ema.fr

Résumé

Dans cet article, nous présentons un nouveau voisinage appelé Voisinage γ-Observable. Ce voisinage constitue un intermédiaire entre les k-Plus-Proches-Voisins et les Voisins Naturels en améliorant la répartition des représentants sélectionnés autour du point considéré. Un paramètre réel permet de faire varier la taille du voisinage entre un unique représentant (le plus proche voisin) et la totalité des représentants. Ce voisinage se détermine par des calculs de distance et de tri peu coûteux en temps machine et en mémoire. Nous étudions ses premières propriétés. Nous donnons une caractérisation d'un représentant dit γ-Observable et nous établissons un critère de γ-Observabilité global. Nous montrons que le Voisinage 0.5-Observable permet d'approcher les Voisins Naturels en grande dimension et le vérifions par l'expérience. Nous signalons en conclusion quelques applications possibles en Quantification Vectorielle et en interpolation dans des espaces de grande dimension basée sur la triangulation induite de Delaunay.

I. Introduction

La définition d'un voisinage joue un rôle important en géométrie algorithmique pour appliquer des techniques d’interpolation en modélisation de données [1][21]. En statistiques (approches descriptives), le voisinage permet d'associer à un point de l'espace d'entrée, une information de sortie stockée par des représentants (*e.g.* dans les Systèmes d'Information Géographique [9], les Réseaux de Neurones [4][5] et en logique Floue [11]) ou bien de déplacer des représentants (appelés vecteurs «codebook») afin de mieux modéliser la distribution des données *a priori* inconnue (*e.g.* en Quantification Vectorielle [17] ou pour la représentation et le regroupement de données («data clustering») [16]).

Deux voisinages particuliers, les k-Plus-Proches-Voisins (k-PPV) et les Voisins Naturels (VN), nous ont conduits à introduire la notion de voisinage γ-Observable. Ces voisinages sont basés sur une structure fondamentale en géométrie algorithmique connue sous le nom de «diagramme de Voronoï» [6][13][20].

Soient un ensemble $S = (\mathbf{p_1},\ldots,\mathbf{p_n})$ de n points distincts dans un espace euclidien R^D et $\mathbf{q}$ un point de R^D. Le polytope de Voronoï d'un $\mathbf{p_i}$ est la partie convexe non vide de R^D définie par:

Définition 1

$$\boxed{V_S(\mathbf{p_i}) = \left\{\mathbf{q} \in R^D \middle| \forall \mathbf{p_j} \in S, \|\mathbf{q} - \mathbf{p_i}\| \le \|\mathbf{q} - \mathbf{p_j}\|\right\}}$$

Les polytopes de Voronoï fournissent une décomposition cellulaire de R^D.

Les k-PPV d'un point $\mathbf{q}$ de R^D forment le sous-ensemble S_k de S contenant les k points de S les plus proches de $\mathbf{q}$ (Figure 1a). Dans la suite, nous appelons les points de S des «représentants».

Ainsi, les k-PPV définissent un voisinage basé sur la distance. Cette propriété est utilisée dans l'algorithme du «Neural-Gas» menant à une technique de Quantification Vectorielle (QV) efficace [17].

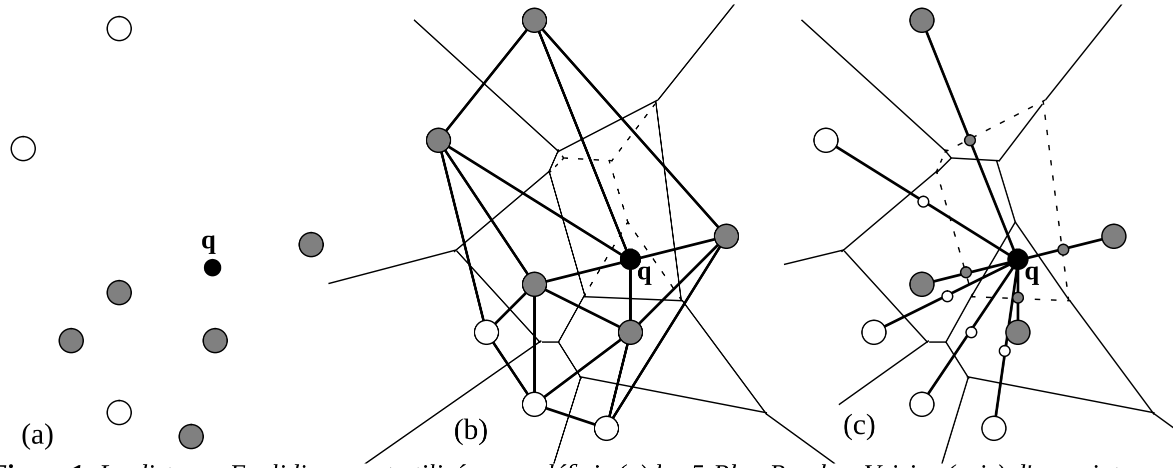

***Figure 1**: La distance Euclidienne est utilisée pour définir (a) les 5-Plus-Proches-Voisins (gris) d'un point **q** (noir), (b) les Voisins Naturels (gris) de **q** et (c) les Voisins 0.5-Observables (grands disques gris) de **q**, pour la même configuration des représentants avec le diagramme de Voronoï (traits continus fins) et la triangulation de Delaunay (traits continus forts). En traits pointillés, le diagramme de Voronoï tel qu'il serait si l'on ne tenait pas compte du point **q** en (b) et si l'on en tenait compte en (c). Dans cet exemple, les Voisins Naturels et les Voisins 0.5-Observables entourent le point **q** tandis que les 5-PPV ne l'entourent pas.*

Cependant, les k-PPV d'un point ne garantissent pas de l'entourer même s'il est situé à l'intérieur de l'enveloppe convexe des représentants (cf. Figure 1a), ce qui peut expliquer que les k-PPV sont moins adaptés à l'interpolation [15].

Un autre voisinage, basé sur les régions de Voronoï des représentants, est utilisé depuis longtemps dans des modèles écologiques et physiques [6][20][24] sans nom spécifique. Dans [22], les représentants appartenant à ce voisinage sont appelés les Voisins Naturels (VN) d'un point **q**.

Soient un ensemble de représentants distincts $S=(\mathbf{p_1},\ldots,\mathbf{p_n})$, et un point **q**, les Voisins Naturels de **q** sont définis comme l'ensemble des représentants $\mathbf{p_i}$ dont la région de Voronoï définie sur l'union de S et de {**q**}, a une frontière commune avec celle du point **q**:

Définition 2

$$\boxed{VN_S(\mathbf{q}) = \{\mathbf{p_i} \in S \mid V_{S \cup \{\mathbf{q}\}}(\mathbf{p_i}) \cap V_{S \cup \{\mathbf{q}\}}(\mathbf{q}) \neq \varnothing\}}$$

Si l'on considère la triangulation de Delaunay, structure duale du diagramme de Voronoï, qui connecte les points dont les régions de Voronoï ont une frontière commune, alors les VN d'un point sont les représentants qui lui sont liés par une arête de la triangulation de Delaunay de l'ensemble des représentants et du point considéré (cf. Figure 1b) [6][13][20].

Les VN ont des propriétés très intéressantes [22]. Ils ont été utilisés en interpolation [8][15][21], pour modéliser les structures biologiques et dynamiques (*e.g.* croissance des arbres; des cellules ou des cristaux [19], biologie moléculaire [14][24]) et comme support pour la Méthode des Eléments Finis appelée Méthode des Eléments Naturels [23].

Cependant, en pratique, trouver les VN d'un point dans des espaces de grande dimension (en pratique, D au-delà de 5 si n est de l'ordre de 25 (cf. Tableau 1)) nécessite des calculs complexes (il faut calculer des volumes de polytopes à D dimensions pour interpoler [22] et résoudre des systèmes linéaires de grande taille si l'on veut seulement trianguler). Ils nécessitent aussi la construction d'une structure de données arborescente pour permettre de diminuer les temps de calcul [7][8][23]. Ces deux faits limitent l'utilisation de VN en QV [17] et en approximation de fonction [4][5] à cause de la grande dimension D des données rencontrées fréquemment dans ces applications. De plus, la taille du voisinage correspondant aux VN ne peut être réglée, on ne peut donc pas l'utiliser en QV [17] où la taille du voisinage doit diminuer au cours du temps pour assurer la convergence de l'algorithme.

Nous présentons ici un nouveau type de voisinage qui cumule les avantages des k-PPV et des VN sans leurs inconvénients. La taille de ce voisinage est réglable grâce à un paramètre; il peut être calculé avec des opérations simples de type calcul de distances et algorithmes de tri et il ressemble au voisinage Naturel (celui défini par les VN) dans le sens où ses voisins entourent le point considéré.

Nous définissons un voisinage qui peut servir à la fois pour l'auto-organisation des représentants dans un espace d'entrée (approche QV) et pour l'interpolation dans un espace de sortie (approche approximation de fonctions).

Ce nouveau voisinage basé sur le diagramme de Voronoï est appelé voisinage γ-Observable (cf. Figure 1c). Il est inspiré par un phénomène d'organisation humaine que nous appelons le problème du «concert de plein-air» décrit ci-après, qui est un cas particulier d'une classe de problèmes plus généraux de «visibilité» [9][12].

II. Le voisinage γ-Observable

Soient un ensemble S=($\mathbf{p_1}$,…,$\mathbf{p_n}$) de représentants distincts dans un espace euclidien R^D, **q** un point de R^D, et γ un nombre réel compris entre 0 et 1. Le voisinage γ-Observable $N\gamma_S$ de **q** est défini sur S par:

Définition 3

$$N\gamma_S(\mathbf{q}) = \{\mathbf{p_i} \in S | \mathbf{q_i} \in V_S(\mathbf{p_i})\} \quad \text{où} \quad \mathbf{q_i} = \gamma \cdot \mathbf{p_i} + (1-\gamma) \cdot \mathbf{q}$$

Les éléments du voisinage γ-Observable de **q** sont appelés les «Voisins γ-Observables» de **q** (cf. Figure 2).

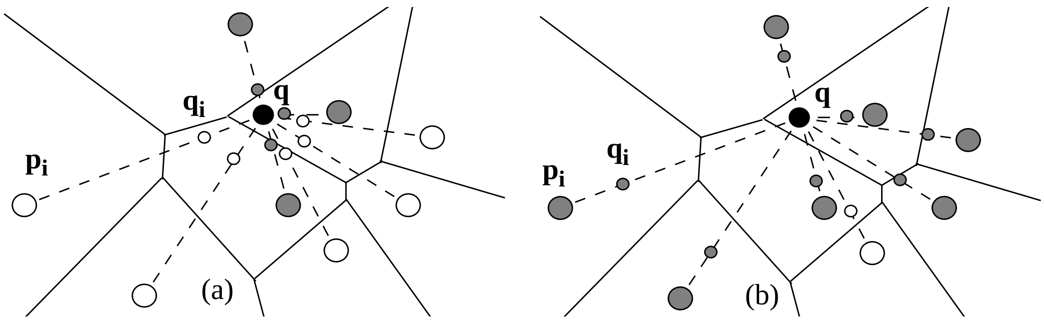

***Figure 2**: les Voisins γ-Observables (grands disques gris) d'un point **q** (noir) avec γ=0.25 (a) et 0.75 (b). Si $\mathbf{q_i}$ (petits disques gris ou blancs) est dans la région de Voronoï de $\mathbf{p_i}$ ($\mathbf{q_i}$ est en gris dans ce cas), alors $\mathbf{p_i}$ fait partie du voisinage γ-Observable de **q**.*

Considérant la définition 3, nous avons:

Conséquence 3.1

$$N0_S(\mathbf{q}) = \{\mathbf{p_{win}}\} \quad \text{où} \quad \mathbf{q} \in V_S(\mathbf{p_{win}})$$

ce qui signifie que pour γ=0, seuls les représentants $\mathbf{p_{win}}$ les plus proches de **q** sont Voisins γ-Observables de **q**. En général, $\mathbf{p_{win}}$ est unique sauf lorsque **q** se trouve sur la frontière de polytopes de Voronoï (cf. Figure 3).

Par ailleurs nous avons:

Conséquence 3.2

$$\forall \mathbf{q} \in R^D, N1_S(\mathbf{q}) = S$$

ce qui signifie que pour γ=1, tous les représentants sont des Voisins γ-Observables de **q** (cf. Figure 3). Ces conséquences sont immédiates en utilisant la définition 3.

Régler le paramètre γ entre 0 et 1 permet de régler la taille de N_γ entre 1 et n.

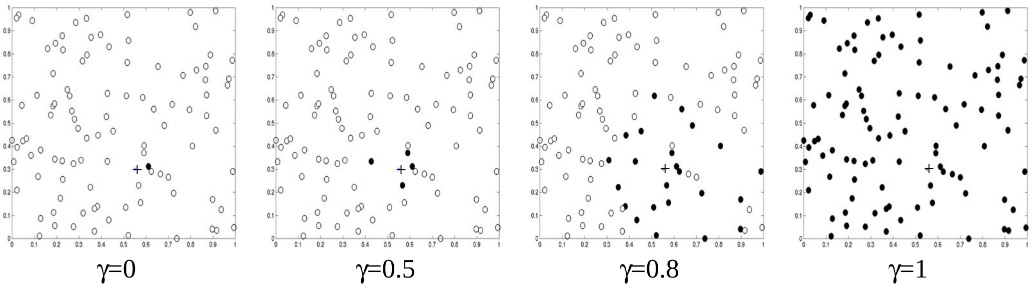

Figure 3*: différents voisinages γ-Observables (cercles noirs) d'un point* ***q*** *(croix noire) définis sur un même ensemble de représentants (cercles) pour différentes valeurs de γ*

La Figure 3 montre le voisinage γ-Observable d'un point pour différentes valeurs de γ pour un même ensemble de 100 représentants uniformément répartis dans le carré unité.

La Figure 3, pour γ=0.8, montre qu'il existe des représentants à l'intérieur de l'enveloppe convexe des Voisins γ-Observables de **q**, qui ne font pas eux-même partie de ce voisinage. Le voisinage γ-Observable peut être qualifié de «non-convexe» dans ce cas, contrairement aux voisinages définis par les k-NN et les VN qui sont toujours «convexes» (Cf. [25] pour les fondements mathématiques et l'origine de cette définition de la convexité d'ensembles discrets). Nos expériences en Quantification Vectorielle montrent l'intérêt de cette caractéristique du voisinage γ-Observable qui tend à homogénéiser la répartition des représentants voisins.

D'un point de vue numérique, la complexité de calcul des Voisins γ-Observables est $O(D.n^2)$, si l'on utilise un algorithme de tri en O(D.n) pour trouver le plus proche voisin de $\mathbf{q_i}$ (cf. définition 3). Elle est linéaire avec la dimension de l'espace d'entrée, cette dimension peut donc être «grande» comparée à la limite pratique pour les Voisins Naturels (*i.e.* au-delà de D=5 pour n=25). Dans le cas d'applications où les temps de calculs sont primordiaux, il est possible d'utiliser des algorithmes de recherche des PPV plus performant pour réduire cette complexité à O(D.n.log(n)) [3], voire compte tenu de la simplicité des opérations mises en jeu, d'utiliser des architectures parallèles [2][10] ce que ne permettent pas les VN.

Analogie du concert de plein-air

On peut montrer que le voisinage γ-Observable est une solution du problème du «concert de plein-air» qui est posé comme suit: «Soit une foule de spectateurs répartis aléatoirement autour de la scène d'un concert de plein-air, quels sont ceux qui peuvent voir la scène?»

Les spectateurs qui peuvent observer la scène pour une hauteur donnée de celle-ci, sont ceux qui appartiennent à son voisinage γ-Observable. Dans cette analogie, γ règle la hauteur de la scène et donc la taille du voisinage (cf. Figure 4).

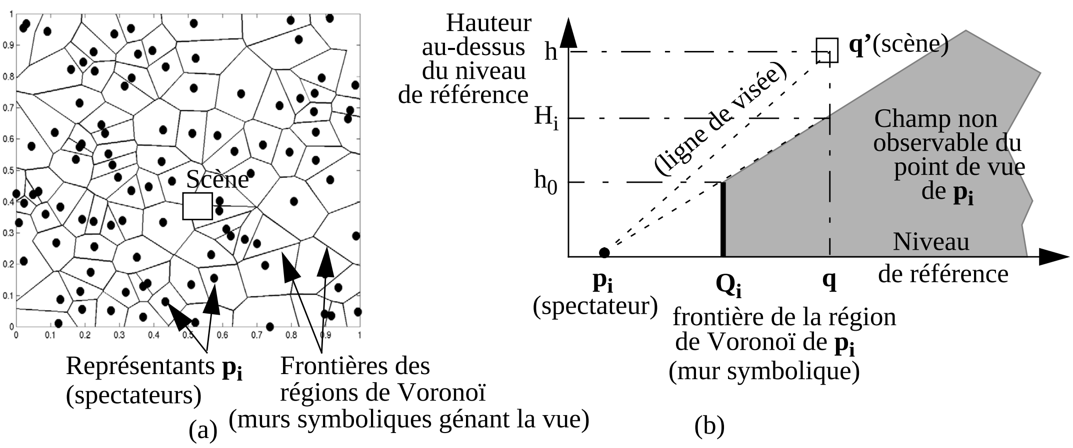

Figure 4*: (a) Vue de dessus du concert de plein-air. (b) Vue de côté du champ visuel d'un spectateur en fonction de la position de la scène.*

III. Propriétés

III.1 Propriété 1: taille du voisinage

La démonstration de la propriété suivante est donnée en Annexe:

Propriété 1

$$\boxed{\forall(\gamma_0, \gamma_1) \in [0, 1]^2, \forall \mathbf{q} \in R^D, (\gamma_1 \geq \gamma_0) \Leftrightarrow N\gamma_{0_S}(\mathbf{q}) \subseteq N\gamma_{1_S}(\mathbf{q})}$$

La propriété 1 assure que, toutes choses égales par ailleurs, la taille du voisinage γ-Observable N_γ croît avec γ (cf. Figure 3).

III.2 Propriété 2: lien avec les voisins naturels

Nous considérons le cas particulier du voisinage 0.5-Observable défini par:

$$N_{0,5_S}(\mathbf{q}) = \{\mathbf{p_i} \in S | (0,5 \cdot (\mathbf{p_i} + \mathbf{q})) \in V_S(\mathbf{p_i})\} \qquad (3, 1)$$

Nous démontrons en Annexe la propriété suivante:

Propriété 2

$$\boxed{\forall \mathbf{q} \in R^D, N_{0,5_S}(\mathbf{q}) \subseteq VN_S(\mathbf{q})}$$

Pour un point donné, les Voisins 0.5-Observables sont un sous-ensemble de ses Voisins Naturels. La propriété 1 permet d'étendre ce résultat à tout γ inférieur à 0.5.

Le nombre de Voisins 0.5-Observables est une borne inférieure du nombre de Voisins Naturels, facilement calculable même en dimensions élevées.

Les Voisins 0.5-Observables sont aussi un moyen de déterminer si une triangulation de Delaunay est aussi une triangulation de Pitteway (cf. [20] (pp.90-92)). La Figure 5a montre les Voisins 0.5-Observables comparés aux Voisins Naturels d'un point.

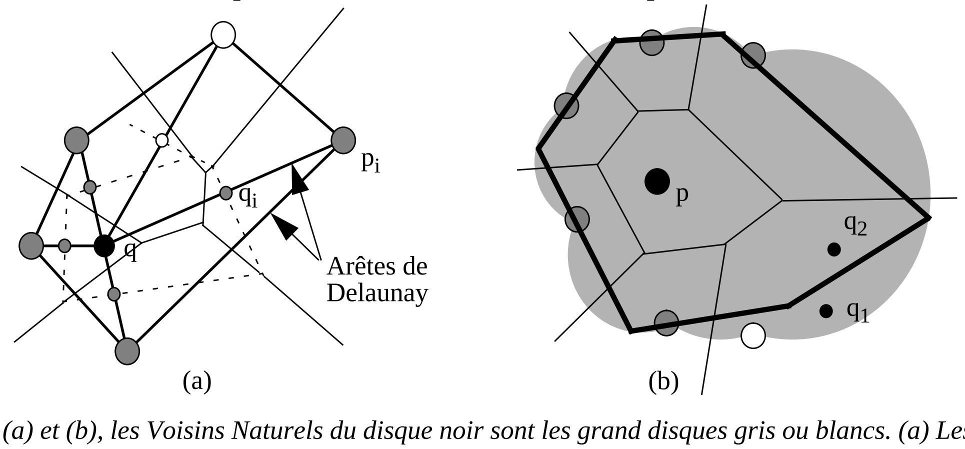

***Figure 5**: en (a) et (b), les Voisins Naturels du disque noir sont les grand disques gris ou blancs. (a) Les Voisins 0.5-Observables de **q** (grands disques gris), sont un sous-ensemble des Voisins Naturels de **q** tels que l'arête de Delaunay entre **q** et $\mathbf{p_i}$ coupe la frontière commune à $V_{S\ U\{\mathbf{q}\}}(\mathbf{q})$ et $V_{S\ U\{\mathbf{q}\}}(\mathbf{p_i})$ en $\mathbf{q_i}$. (b) montre les régions d'influence superposées du représentant (noir) pour les Voisins 0.5-Observables (région grise intérieure au trait fort) et les Voisins Naturels (région grise entière). La région d'influence de **p** au sens des Voisins 0.5-Observables est homothétique à la région de Voronoï de **p**. **p** est un Voisin Naturel de $\mathbf{q_1}$ et $\mathbf{q_2}$, mais en utilisant les Voisins 0.5-Observables, **p** n'est détecté que comme Voisin Naturel de $\mathbf{q_2}$.*

Pour quantifier la relation entre les Voisins 0.5-Observables et les Voisins Naturels, dans l'expérience suivante, nous calculons le rapport du nombre de Voisins 0.5-Observables sur le nombre total de Voisins Naturels de chaque représentant parmi tous les autres.

__Tableau 1__: rapport moyen et écart type du nombre de Voisins 0.5-Observables sur le nombre de Voisins Naturels de chaque représentant parmi eux, sur 10 expériences pour un ensemble de n représentants uniformément distribués dans un hypercube unité D-dimensionnel.

n	D+2		25		50		100		500	
D	mean	std	mean	std	mean	std	mean	std	mean	std
2	0.72	0.25	0.63	0.20	0.64	0.20	0.64	0.20	0.66	0.20
3	0.65	0.25	0.48	0.18	0.48	0.17	0.47	0.16	0.47	0.16
5	0.69	0.20	0.41	0.15	0.35	0.12	0.32	0.12	0.29	0.10
10	0.77	0.16	0.65	0.15	-	-	-	-	-	-
15	0.87	0.11	0.83	0.11	-	-	-	-	-	-
20	0.94	0.07	0.92	0.08	-	-	-	-	-	-

Le Tableau 1 donne le rapport moyen obtenu pour différents nombres de représentants et différentes dimensions sur 10 expériences dans chaque cas. Une partie du Tableau 1 n'a pu être complétée du fait de la grande complexité des calculs et du manque de mémoire lors de l'utilisation du logiciel `QHull` [7] pour le calcul des Voisins Naturels grâce à la triangulation de Delaunay. Le calcul des Voisins 0.5-Observables ne souffre pas de ce problème puisqu'il consiste à calculer des distances euclidiennes (O(D)) et rechercher des minima (O(n)) ce qui n'est pas coûteux en temps de calcul ni en place mémoire même en grandes dimensions (D>5).

Nous pouvons voir que pour un nombre de représentants n donné, plus la dimension D est élevée meilleure est l'approximation des Voisins Naturels par les Voisins 0.5-Observables. De même, pour un espace de dimension donnée D, plus le nombre de représentants n est faible, meilleure est cette approximation.

On peut envisager d'utiliser les Voisins γ-Observables, particulièrement en grande dimension, pour trouver les Voisins Naturels restant en utilisant la technique proposée dans [14]. Une voie à explorer est l'utilisation directe du Voisinage γ-Observable avec différentes valeurs de γ pour tenter d'encadrer avec plus de précision l'ensemble des Voisins Naturels voire pour les trouver tous.

III.3 Propriété 3: forme des régions d'influence

Nous définissons la Région d'Influence $RI_N(\mathbf{p_i})$ d'un représentant $\mathbf{p_i}$ correspondant à un voisinage N, par:

Définition 4

$$\boxed{\forall \mathbf{p_i} \in S, \quad RI_N(\mathbf{p_i}) = \left\{ \mathbf{q} \in R^D \middle| \mathbf{p_i} \in N(\mathbf{q}) \right\}}$$

La Figure 5b montre la région d'influence d'un représentant en fonction du type de voisinage utilisé. Il apparaît sur la Figure 5b, que la région d'influence d'un représentant correspondant au Voisins 0.5-Observables, est incluse dans celle correspondant aux Voisins Naturels, ce qui est décrit par la propriété 2.

Nous avons alors la propriété suivante démontrée en Annexe:

Propriété 3

$$\boxed{\forall \mathbf{p_i} \in S, RI_{N\gamma_S}(\mathbf{p_i}) = H_{\mathbf{p_i}, \frac{1}{1-\gamma}}(V_S(\mathbf{p_i}))}$$

où $H_{\Omega,k}$ est l'homothétie de centre Ω et de rapport k. Cette propriété signifie que la région d'influence d'un représentant au sens du voisinage γ-Observable est homothétique à sa région de Voronoï. La Figure 5b illustre cette propriété pour les Voisins 0.5-Observables.

IV. Caractérisation et critère de γ-observabilité

Dans cette section, nous définissons un critère algébrique d'appartenance au voisinage γ-Observable ainsi qu'une valeur critique γ_c au-dessus de laquelle tous les représentants sont Voisins γ-Observables de tout point du domaine de R^D considéré.

IV.1 Caractérisation d'un représentant γ-Observable

Pour un point $\mathbf{q}$ donné de R^D, nous définissons $C_j(S,\mathbf{q},\gamma)$ par:

Définition 5

$$\boxed{C_j(S, \mathbf{q}, \gamma) = \prod_{\substack{l = 1 \\ l \neq j}}^{n} H(d_{lj}^2 - d_{jj}^2)}$$

avec

$$\forall j \in (1, \ldots, n), \mathbf{q_j} = \gamma \cdot \mathbf{p_j} + (1-\gamma) \cdot \mathbf{q} \qquad (4, 1)$$

$$\forall (l, j) \in (1, \ldots, n)^2, d_{lj} = \|\mathbf{p_l} - \mathbf{q_j}\| \qquad (4, 2)$$

et la fonction d'Heavyside

$$H(x) = \begin{cases} 1, \forall x \geq 0 \\ 0, \forall x < 0 \end{cases} \qquad (4, 3)$$

Conséquence 5.1

$$\boxed{\mathbf{q_j} \in V_S(\mathbf{p_j}) \Leftrightarrow C_j(S, \mathbf{q}, \gamma) = 1}$$

$C_j(S,\mathbf{q},\gamma)$ est une fonction caractéristique de l'appartenance du représentant $\mathbf{p_j}$ au voisinage γ-Observable de $\mathbf{q}$.

En effet, par construction, pour tout indice $l \neq j$:

$$C_j(S, \mathbf{q}, \gamma) = 1 \Leftrightarrow d_{lj}^2 - d_{jj}^2 \geq 0 \Leftrightarrow \mathbf{q_j} \in V_S(\mathbf{p_j})$$

IV.2 Critère de γ–Observabilité

Soit Δ un domaine borné de R^D de diamètre δ contenant les représentants $\mathbf{p_i}$ de S et les points $\mathbf{q}$. Nous avons la propriété suivante

Propriété 4

$$\forall j \in (1, \ldots, n), \forall \mathbf{q} \in \Delta \subset R^D, \forall S \subset \Delta,\ C_j(S, \mathbf{q}, \gamma_c) = 1$$

$$\text{avec} \qquad \gamma_c = 1 - \min_{j \in (1, \ldots, n)}\left(\min_{\substack{l \in (1, \ldots, n) \\ l \neq j}}\left(\frac{1}{2}\frac{\|\mathbf{p_l} - \mathbf{p_j}\|}{\delta + \|\mathbf{p_j}\|}\right)\right)$$

Cette propriété est démontrée en Annexe. Le choix de γ supérieur ou égal à cette valeur critique γ_c garantit que tous les représentants $\mathbf{p_j}$ appartiennent au voisinage γ-Observable de tout point $\mathbf{q}$ de Δ.

Il découle des propriétés 1 et 4 que:

$$\boxed{\forall \mathbf{q} \in \Delta \subset R^D, \forall \gamma \in [\gamma_c, 1], N_{\gamma_S}(\mathbf{q}) = S \qquad (4, 4)}$$

ce qui généralise la conséquence 3.2 de la définition 3.

Par construction, γ_c est minorée strictement par la valeur 0.5. Cette valeur particulière de γ (cf. Section III.2) apparaît de nouveau ici soulignant les liens qui unissent les Voisins γ-Observables et les Voisins Naturels.

IV.3 Etude de $N_{\gamma S}$ en fonction de γ et de D

La Figure 6 montre l'évolution de la taille du voisinage γ-Observable en fonction de la valeur de γ pour différentes dimensions.

Elle a été calculée en utilisant n={100,1000} représentants aléatoires uniformément répartis dans un hypercube unitaire D-dimensionnel $[0,1]^D$, et en calculant le nombre de Voisins γ-Observables de **q**(0.5,0.5,...,0.5), centre de cet hypercube, pour γ prenant ses valeurs dans [0,1] de 0.1 en 0.1. Les courbes résultantes sont la moyenne sur 10 expériences avec différentes configurations aléatoires des représentants.

Nous observons que la taille du voisinage γ-Observable n'est pas linéaire avec γ, et pour un nombre n de représentants, cette taille augmente d'autant plus vite avec γ que la dimension D est grande.

Nous constatons que pour n=100 et n=1000 représentants, plus de 95% de ces points sont des Voisins 0.5-Observables du point **q** en dimension D=20, donc plus de 95% des représentants sont des Voisins Naturels de **q**. Ce résultat expérimental tend à confirmer un résultat théorique établi par Demartines et Fort [26], qui montre que plus la dimension D est grande, plus la norme de vecteurs tirés aléatoirement dans D est proche de la moyenne de ces normes, i.e. les points tirés semblent tous être équidistants les uns des autres, donc tous situés approximativement aux sommets d'un simplexe équilatéral. Or dans un tel simplexe, les sommets sont tous Voisins Naturels les uns des autres, et on peut montrer en particulier qu'ils sont tous Voisins 0.5-Observables les uns des autres, ce qui correspond approximativement à la situation observée.

Nous observons comme à la section III.2, que plus la dimension D est élevée pour un nombre de représentants n donné, plus la proportion de Voisins 0.5-Observables parmi les représentants est grande, donc meilleure est l'approximation des Voisins Naturels par les Voisins 0.5-Observables.

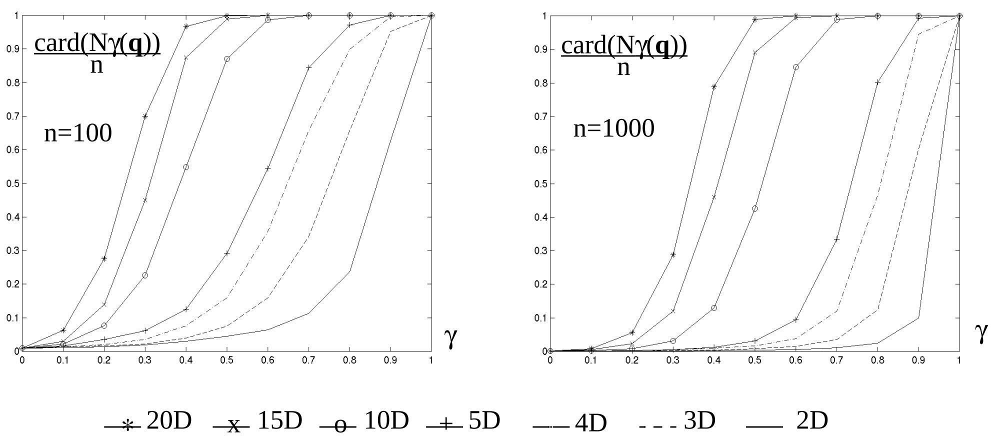

Figure 6*: taille du voisinage γ-Observable en fonction de D et γ pour n=100 et n=1000.*

V. Discussion et conclusions

Nous avons défini un nouveau voisinage appelé voisinage γ-Observable, qui rassemble les propriétés intéressantes des k-Plus Proches Voisins et des Voisins Naturels: sa taille est réglable, il implique des opérations simples et peu de mémoire, ses voisins entourent le point considéré. Nous avons donné une caractérisation algébrique de ce voisinage et étudié quelques-unes de ses propriétés. Ce voisinage peut éventuellement être implémenté sur une architecture parallèle *a priori* sans grande difficulté. Nous avons étudié plus particulièrement le cas des Voisins 0.5-Observables qui peuvent être utilisés pour approcher les Voisins Naturels plus difficiles à calculer et qui ouvre de nouvelles voies en analyse de données de grandes dimensions. Les Voisins γ-Observables font le lien entre les k-Plus Proches Voisins, voisinage simple à calculer mais relativement pauvre, et les Voisins Naturels, voisinage plus riche puisque porteur d'une information sur la topologie des données (dimension intrinsèque [18] et positions relatives) mais plus complexe voire impossible à calculer en grande dimension. Le voisinage γ-Observable est donc un outil supplémentaire pour la caractérisation et l'analyse de données en grandes dimensions.

Notre intérêt se porte actuellement sur l'utilisation du voisinage γ-Observable en Quantification Vectorielle [27], où il donne déjà des résultats prometteurs comparés à ceux du Neural-Gas [17] sur diverses distributions. La propriété de «non-convexité» semble jouer un rôle important dans ce type d'applications.

Plus généralement, l'utilisation du voisinage γ-Observable dans le domaine des systèmes coopératifs (automates cellulaires, réseaux neuronaux auto-organisés) pourrait mener à des comportements dynamiques intéressants dans ces systèmes où la définition d'un voisinage joue un grand rôle.

Nous étudions aussi les rapports entre ce voisinage et la technique d'interpolation des Noyaux de Voronoï que nous avons développée récemment [28] et qui respecte la triangulation induite (i.e. partielle) de Delaunay [18]. Cette technique peut s'appliquer dans des espaces de grande dimension sur des triangulations de Delaunay partielles dont la dimension intrinsèque varie spatialement, ce qui est difficilement envisageable avec d'autres approches de type B-spline par exemple.

Le voisinage γ-Observable ouvre donc des perspectives dans les domaines de la géométrie algorithmique, de l'analyse de données, de la Quantification Vectorielle, de la dynamique des systèmes coopératifs, et de l'interpolation.

Remerciements

Nous remercions les relecteurs pour leurs commentaires qui ont permis d'améliorer la lisibilité de cet article. Nous remercions également Volkmar Welker (Philipps-Universität Marburg, Allemagne) pour son aide concernant la définition de la convexité pour des ensembles discrets.

Annexes

Dans toutes les démonstrations qui suivent, nous considérons un ensemble S=($\mathbf{p}_1,\ldots,\mathbf{p}_n$) de n points représentants distincts dans un domaine Δ borné de l'espace euclidien R^D, $\mathbf{q}$ un point de Δ, δ le diamètre de Δ, et γ ou γ_j un nombre réel compris entre 0 et 1.

Démonstration de la Propriété 1

Soit $\mathbf{q}_{\mathbf{i},\mathbf{j}}$ défini par:

$$\mathbf{q}_{\mathbf{i},\,\mathbf{j}} = \gamma_j \cdot \mathbf{p}_\mathbf{i} + (1-\gamma_j) \cdot \mathbf{q} \qquad (1)$$

alors on a:

$$\begin{aligned} \forall(\gamma_0, \gamma_1) \in [0,1]^2, (\gamma_1 \geq \gamma_0) &\Leftrightarrow (1-\gamma_1) \cdot \|\mathbf{q}-\mathbf{p}_\mathbf{i}\| \leq (1-\gamma_0) \cdot \|\mathbf{q}-\mathbf{p}_\mathbf{i}\| \\ &\Leftrightarrow \|\mathbf{q}_{\mathbf{i},\,\mathbf{1}} - \mathbf{p}_\mathbf{i}\| \leq \|\mathbf{q}_{\mathbf{i},\,\mathbf{0}} - \mathbf{p}_\mathbf{i}\| \\ &\Leftrightarrow (\mathbf{q}_{\mathbf{i},\,\mathbf{0}} \in V_S(\mathbf{p}_\mathbf{i}) \Rightarrow \mathbf{q}_{\mathbf{i},\,\mathbf{1}} \in V_S(\mathbf{p}_\mathbf{i})) \\ &\Leftrightarrow (\mathbf{p}_\mathbf{i} \in N\gamma_{0_S}(\mathbf{q}) \Rightarrow \mathbf{p}_\mathbf{i} \in N\gamma_{1_S}(\mathbf{q})) \\ &\Leftrightarrow (N\gamma_{0_S}(\mathbf{q}) \subseteq N\gamma_{1_S}(\mathbf{q})) \end{aligned}$$

CQFD.

Démonstration de la Propriété 2

Soit $\mathbf{q_i}$ défini par:

$$\mathbf{q_i} = 0{,}5 \cdot (\mathbf{p_i} + \mathbf{q}) \qquad (2)$$

alors:

$$\|\mathbf{q_i} - \mathbf{q}\| = \|\mathbf{q_i} - \mathbf{p_i}\| \qquad (3)$$

et par la définition 3, on obtient:

$$\mathbf{p_i} \in N_{0,5_S}(\mathbf{q}) \Leftrightarrow \mathbf{q_i} \in V_S(\mathbf{p_i}) \qquad (4)$$

par suite de (3) et (4), on a:

$$(4) \Leftrightarrow \mathbf{q_i} \in V_S(\mathbf{p_i}) \Leftrightarrow \forall \mathbf{p_j} \in S, \begin{cases} \|\mathbf{q_i} - \mathbf{p_i}\| \le \|\mathbf{q_i} - \mathbf{p_j}\| \\ \|\mathbf{q_i} - \mathbf{q}\| \le \|\mathbf{q_i} - \mathbf{p_j}\| \end{cases}$$

$$\Leftrightarrow \begin{cases} \mathbf{q_i} \in V_{S \cup \{\mathbf{q}\}}(\mathbf{p_i}) \\ \mathbf{q_i} \in V_{S \cup \{\mathbf{q}\}}(\mathbf{q}) \end{cases}$$

$$\Leftrightarrow \quad \mathbf{q_i} \in (V_{S \cup \{\mathbf{q}\}}(\mathbf{p_i}) \cap V_{S \cup \{\mathbf{q}\}}(\mathbf{q})) \qquad (5)$$

Donc,

$$(5) \Rightarrow (V_{S \cup \{\mathbf{q}\}}(\mathbf{p_i}) \cap V_{S \cup \{\mathbf{q}\}}(\mathbf{q})) \neq \emptyset \qquad (6)$$

$$(6) \Leftrightarrow \mathbf{p_i} \in VN_S(\mathbf{q})$$

ce qui mène à:

$$\mathbf{p_i} \in N_{0,5_S}(\mathbf{q}) \Rightarrow \mathbf{p_i} \in VN_S(\mathbf{q})$$

$$\Leftrightarrow \forall \mathbf{q} \in R^D, N_{0,5_S}(\mathbf{q}) \subseteq VN_S(\mathbf{q})$$

CQFD.

Démonstration de la Propriété 3

Considérant les définitions 3 et 4, on obtient:

$$\forall \mathbf{p_i} \in S, \quad RI_{N\gamma_S}(\mathbf{p_i}) = \left\{ \mathbf{q} \in R^D \middle| \mathbf{p_i} \in N\gamma_S(\mathbf{q}) \right\} \qquad (7)$$

$$(7) \Leftrightarrow \forall \mathbf{p_i} \in S, \quad RI_{N\gamma_S}(\mathbf{p_i}) = \left\{ \mathbf{q} \in R^D \middle| \mathbf{q_i} \in V_S(\mathbf{p_i}) \right\} \qquad (8)$$

et suivant la définition 3, l'équation (8) se traduit par:

$$\forall \mathbf{p_i} \in S, \forall \mathbf{q} \in RI_{N\gamma_S}(\mathbf{p_i}), \mathbf{q_i} \in V_S(\mathbf{p_i}) \text{ et } (\mathbf{q} - \mathbf{p_i}) = \frac{1}{1-\gamma} \cdot (\mathbf{q_i} - \mathbf{p_i}) \qquad (9)$$

$$(9) \Leftrightarrow \forall \mathbf{p_i} \in S, \forall \mathbf{q} \in RI_{N\gamma_S}(\mathbf{p_i}), \mathbf{q_i} \in V_S(\mathbf{p_i}) \text{ et } \mathbf{q} = H_{\mathbf{p_i}, \frac{1}{1-\gamma}}(\mathbf{q_i})$$

$$(9) \Leftrightarrow \forall \mathbf{p_i} \in S, RI_{N\gamma_S}(\mathbf{p_i}) = H_{\mathbf{p_i}, \frac{1}{1-\gamma}}(V_S(\mathbf{p_i}))$$

où $H_{\Omega,k}$ est l'homothétie de centre Ω et de rapport k.

CQFD.

Démonstration de la Propriété 4

D’après la caractérisation des Voisins γ-Observables (Conséquence 5.1):

$$C_j(S, \mathbf{q}, \gamma) = 1 \Leftrightarrow d_{lj}^2 - d_{jj}^2 \geq 0 \qquad \forall l \neq j$$

$$\Leftrightarrow \|\mathbf{p_l} - \mathbf{q_j}\|^2 - \|\mathbf{p_j} - \mathbf{q_j}\|^2 \geq 0 \qquad \forall l \neq j$$

$$\Leftrightarrow \|\mathbf{p_l} - \mathbf{p_j} + (1-\gamma)(\mathbf{p_j} - \mathbf{q})\|^2 - \|(1-\gamma)(\mathbf{p_j} - \mathbf{q})\|^2 \geq 0 \qquad \forall l \neq j$$

$$\Leftrightarrow \|\mathbf{p_l} - \mathbf{p_j}\|^2 - 2(1-\gamma)\underbrace{\langle \mathbf{p_l} - \mathbf{p_j} | \mathbf{q} - \mathbf{p_j} \rangle}_{P} \geq 0 \qquad \forall l \neq j$$

Si le produit scalaire P est négatif ou nul, cette inégalité est vérifiée automatiquement et $C_j(S,\mathbf{q},\gamma)=1$ est réalisé.

Si P est positif, alors pour réaliser $C_j(S,\mathbf{q},\gamma)=1$, on doit avoir:

$$1 - \gamma \leq \frac{1}{2} \frac{\|\mathbf{p_l} - \mathbf{p_j}\|^2}{\langle \mathbf{p_l} - \mathbf{p_j} | \mathbf{q} - \mathbf{p_j} \rangle} \qquad \forall l \neq j$$

or

$$\langle \mathbf{p_l} - \mathbf{p_j} | \mathbf{q} - \mathbf{p_j} \rangle \leq \|\mathbf{p_l} - \mathbf{p_j}\| \|\mathbf{q} - \mathbf{p_j}\| \leq \|\mathbf{p_l} - \mathbf{p_j}\| (\|\mathbf{q}\| + \|\mathbf{p_j}\|) \leq \|\mathbf{p_l} - \mathbf{p_j}\| (\delta + \|\mathbf{p_j}\|)$$

Ainsi, pour j fixé, $C_j(S,\mathbf{q},\gamma)=1$ est réalisé pour tout **q** de Δ lorsque

$$1 - \gamma = \min_{\substack{l \in (1, \ldots, n) \\ l \neq j}} \left(\frac{1}{2} \frac{\|\mathbf{p_l} - \mathbf{p_j}\|}{\delta + \|\mathbf{p_j}\|} \right)$$

Lorsque j décrit {1,...,n}, $C_j(S,\mathbf{q},\gamma)=1$ est réalisé si l'on choisit γ tel que

$$1 - \gamma = \min_{j \in (1, \ldots, n)} \left(\min_{\substack{l \in (1, \ldots, n) \\ l \neq j}} \left(\frac{1}{2} \frac{\|\mathbf{p_l} - \mathbf{p_j}\|}{\delta + \|\mathbf{p_j}\|} \right) \right)$$

CQFD.

Références

[1] N.Amenta, M.Bern and M.Kamvysselis - *A New Voronoï-Based Surface Reconstruction Algorithm* - Proc. of Siggraph’98, 1998

[2] F. Ancona, S. Rovetta, and R. Zunino - *Implementation of Neural Gas Training in Analog VLSI* - Proc. International Symposium on Intelligent Systems, ISIS'97 - Reggio Calabria, Italy - September 1997

[3] S.Arya, D.M.Mount, N.S.Netanyahu, R.Silverman and A.Y.Wu - *An Optimal Algorithm for Approximate Nearest Neighbor Searching in Fixed Dimensions* - Proc. of the 5th ACM-SIAM Symposium on Discrete Algorithms, pp. 573-582, 1994

[4] M.Aupetit, P.Massotte and P.Couturier - *Function Approximation with Continuous Self-Organizing Maps using Neighboring Influence Interpolation* - To appear in Proc. of Neural Computation, Berlin, Germany, May 2000

[5] M.Aupetit, P.Massotte and P.Couturier - *A Continuous Self-Organizing Map using Spline Technique for Function Approximation* - Proc. of Artificial Intelligence Control and System, Cork, Ireland, September 1999

[6] F.Aurenhammer - *A Survey of a Fundamental Geometric Data Structure* - ACM Computing Surveys, Vol. 23, No. 3, September 1991

[7] C.B.Barber, D.P.Dobkin and H.Huhdanpaa - *The Quickhull Algorithm for Convex Hulls* - ACM Transactions on Mathematical Software, Vol. 22, No. 4, pp. 469-483, December 1996. http://www.geom.umn.edu/locate/qhull

[8] V.V.Belikov, V.D.Ivanov, V.K.Kontorovich, S.A.Korytnik and A.Y.Semenov - *The Non-Sibsonian Interpolation: A New Method of Interpolation of the Values of a Function on an Arbitrary Set of Points* - Computational Mathematics and Mathematical Physics, Vol. 37, No. 1, pp. 9-15, 1997

[9] L.De Floriani, P. Magillo, E. Puppo - *Applications of Computational Geometry to Geographic Information Systems* - http://www.disi.unige.it/person/PuppoE/PS/handbook.ps.gz

[10] G.De Trémiolles, P. Tannhof, B.Plougonven, C.Demarigny, K.Madani - *Visual Probe Mark Inspection, using Hardware Implementation of Artificial Neural Networks, in VLSI Production* - Proc. of IWANN'97, Lanzarote - Canary Islands, Spain - June 1997

[11] J.C.Dunn - A Fuzzy Relative of the Isodata Process and its Use in Detecting Compact Well-Separated Clusters - Journal of Cybernetics, Vol. 3, No. 3, pp. 32-57, 1974

[12] S.Eidenbenz, C.Stamm and P. Widmayer - *Positioning Guards at Fixed Height above a Terrain, an Optimum Inapproximability Result* - Proc. of European Symposium on Algorithms, 1998

[13] S.Fortune - *Voronoï Diagrams and Delaunay Triangulations* - Computing in Euclidean Geometry, D.Z.Du, F.Hwang, eds, pp. 193-233, World Scientific, 1992

[14] M.Gerstein and F.M.Richards - *Protein Geometry: Volumes, Areas, and Distances* - Manuscript for inclusion in The International Tables for Crystallography, Vol. F, Chap. 22, 1999 - http://bioinfo.mbb.yale.edu/e-print/geom-inttab/geom-inttab.pdf

[15] N.L.Jones, S.J.Owen and E.C.Perry - *Plume Characterization with Natural Neighbor Interpolation* - Proc. of Geoenvironment, Geotechnical Engineering and Environmental Engineering Divisions/ASCE, New Orleans, Louisiana, February 2000

[16] S.P.Lloyd - Least Squares quantization in PCM - IEEE Transaction on Information Theory, Vol. IT-28, No. 2, pp. 129-137, March 1982

[17] T.M.Martinetz, S.G.Berkovich and K.J.Schulten - *"Neural-Gas" Network for Vector Quantization and its Application to Time-Series Prediction* - IEEE Transaction on Neural Networks, Vol. 4, No. 4, July 1993

[18] T.Martinetz and K.Schulten - *Topology Representing Networks* - Neural Networks, Vol. 7, No. 3, pp. 507-522, 1994

[19] P.J.Moran and M.Wagner - *Introducing Alpha Shapes for the Analysis of Path Integral Monte Carlo Results* - Proc. of Visualization'94, pp. 52-59, 1994

[20] A.Okabe, B.Boots and K.Sugihara - *Spatial Tessellations: Concepts and Applications of Voronoï Diagrams* - John Wiley, Chichester 1992

[21] M.Sambridge, J.Braun and H.McQueen - *Geophysical Parameterization and Interpolation of Irregular Data using Natural Neighbours* - Geophysical Journal International, No. 122, pp. 837-857, June 1995

[22] R.Sibson - *A Brief Description of Natural Neighbour Interpolation* - Interpreting Multivariate Data, John Wiley, Chichester, pp. 21-36, 1981

[23] N.Sukumar - *The Natural Element Method in Solid Mechanics* - Ph.D Thesis in Theoretical and Applied Mechanics, Northwestern University, Evanston, Illinois, June 1998

[24] H-M Will - *Practical and Efficient Computation of Additively Weighted Voronoï Cells for Applications in Molecular Biology* - Technical Report #300, Departement Informatik, ETH Zürick, 1998 - ftp://ftp.inf.ethz.ch/pub/publications/tech-reports/

[25] P.H. Edelman, R. Jamison - *The theory of convex geometries* - Geom. Dedicata 19, pp. 247-270, 1985

[26] P. Demartines- *Analyse de données par réseaux de neurones auto-organisés* - Thèse de Doctorat de l'INPG - laboratoire TIRF, Novembre 1994

[27] M. Aupetit, P. Couturier and P. Massotte - *Vector Quantization with γ-Observable Neighbors* - Workshop on Self-Organizing Maps (WSOM'2001), Juin 2001

[28] M. Aupetit, P. Couturier and P. Massotte - *Induced Voronoï Kernels for Principal Manifolds Approximation* - Workshop on Self-Organizing Maps (WSOM'2001), Juin 2001