CAHIERS DU BURO

ERIC-OLIVIER LOCHARD
DANIÈLE FLORENS

Taxonomie automatique : la sous-dominante dans son cadre général et « la » plus proche au sens des moindres carrés

Cahiers du Bureau universitaire de recherche opérationnelle. Série Recherche, tome 25 (1976), p. 39-60
<http://www.numdam.org/item?id=BURO_1976__25__39_0>

TAXONOMIE AUTOMATIQUE : LA SOUS-DOMINANTE DANS SON CADRE GÉNÉRAL ET "LA" PLUS PROCHE AU SENS DES MOINDRES CARRÉS

Danièle FLORENS et Eric-Olivier LOCHARD

Résumé : On montre la "trivialité" de la sous-dominante en généralisant sa construction pour des dissimilarités sur un ensemble quelconque à valeur dans un ensemble totalement ordonné, stable par inf et sup. On donne une démonstration formelle de la continuité de la sous-dominante d'une dissimilarité réelle sur un ensemble fini. On montre l'unicité et la continuité presque partout de l'ultramétrique la plus proche au sens des moindres carrés d'une dissimilarité réelle sur un ensemble fini.

INTRODUCTION

Nous avons étudié la taxonomie numérique pour répondre aux besoins des travaux que nous avons entrepris en collaboration avec des géographes de Montpellier. Nous avons essayé d'abord de dégager un langage clair pour cette théorie. Nous avons essayé ensuite de dégager les parties formelles, ce qui nous a permis de montrer la trivialité de la sous-dominante en la construisant dans le cadre le plus général possible. Nous avons aussi résolu quelques problèmes posés par l'ultramétrique la plus proche au sens des moindres carrés d'une dissimilarité donnée. Nous n'avons pas encore pu construire un algorithme performant quoique nous ayons montré que cette ultramétrique était calculable.

Nous donnons dans le chapitre 5 un algorithme très performant du calcul de la projection d'une fonction sur un cône de fonctions croissantes.

Devant le manque d'idées claires des utilisateurs de cette théorie nous abandonnons provisoirement les recherches dans ce domaine.

La meilleure bibliographie sur le sujet est à notre avis celle de Mac Cormack [2]. Nous y renvoyons donc pour les références des travaux non cités dans le nôtre.

Les démonstrations qui ne sont pas écrites suivent leur cours et sont laissées au lecteur.

CHAPITRE I

ORDONNANCES ET ULTRAORDONNANCES

§0 – 0.0 – Soit I un ensemble. L'application qui à un préordre sur I fait correspondre l'ensemble de ses sections commençantes est une bijection entre l'ensemble des préordres sur I et l'ensemble des parties de I stable par intersection et réunion quelconques.

0.1 – Si R est un préordre sur I, rappelons qu'une section commençante est une partie E de I vérifiant $E = \cup\{R^{-1}(x)/\ x \in E\}$.

0.2 – La bijection réciproque est la restriction de l'application qui à un ensemble F de parties de I fait correspondre la relation R définie par : "xRy" si et seulement si "$E \in F$ et $y \in E \Rightarrow x \in E$". Cette relation R est un préordre sur I.

0.3 – Si F est stable par réunions et intersections quelconques le préordre associé est total si et seulement si F est totalement ordonné par inclusion.

§1 – 1.0 – **Définitions.** *Soit E un ensemble. On appelle préordonnance sur E un ensemble F de relations binaires symétriques non vide sur E, totalement ordonné par inclusion et stable par intersection quelconque et réunion non vide .*

Une préordonnance sera dite une ordonnance si les relations qui la composent sont réflexives.

Une ordonnance sera dite séparée si elle contient la diagonale de $E \times E$.

1.1 – D'après ce qui précède se donner une préordonnance sur l'ensemble fini E revient à se donner un préordre total R sur $E \times E$ vérifiant $(x, y) R (y, x)$ pour tout $(x, y) \in E \times E$. Se donner un tel préordre revient à se donner un préordre total sur $\mathcal{P}_1(E) \cup \mathcal{P}_2(E)$.

Une préordonnance sera une ordonnance si et seulement si, pour le préordre ainsi construit sur $\mathcal{P}_1(E) \cup \mathcal{P}_2(E)$, les éléments de $\mathcal{P}_1(E)$ sont minimaux, i.e. se donner une ordonnance revient à se donner un préordre total sur $\mathcal{P}_2(E)$. Ceci montre que notre langage est équivalent aux langages habituels.

1.2 *Soit F une ordonnance sur E. Les propriétés suivantes sont équivalentes :*

i) *F est maximale pour l'inclusion.*

ii) *F est séparée et le préordre associé sur $\mathcal{P}_2(E)$ est un ordre.*

iii) *F est séparée, si $X \in F$ et si Y est le prédécesseur de X dans F, s'il existe, alors le cardinal de $X - Y$ est inférieur ou égal à 2.*

iv) Si de plus E est fini et a n éléments les conditions précédentes sont équivalentes à : *le cardinal de* F *est* $n(n-1)/2+1$.

1.3 – **Corollaire.** *Si* E *est de cardinal* n *il y a* $\binom{n}{2}!$ *ordonnances maximales sur* E.

1.4 – **Lemme.** *Toute ordonnance séparée est l'intersection des ordonnances maximales qui la contiennent.*

§**2** – 2.1 – **Définition.** *Soit E un ensemble. Une ordonnance sur E sera dite une ultraordonnance si les relations qui la composent sont des relations d'équivalence.*

2.2 – Un des problèmes de la taxonomie aveugle est de construire des applications de l'ensemble des ordonnances sur un ensemble dans l'ensemble des ultraordonnances sur le même ensemble.

2.3 – La première idée qui vient à l'esprit est d'associer à une ordonnance F l'ensemble des relations d'équivalence engendrées par ses éléments. Malheureusement cet ensemble n'est pas stable par intersection mais stable par réunion non vide.

Lemme. *Soit* F *une ordonnance. L'ensemble des intersections quelconques des relations d'équivalence engendrées par les éléments de* F *est une ultraordonnance (que nous noterons* ult F*).*

. Ult F est stable par intersection quelconque, par définition.

. Soit R une famille non vide d'éléments de ult F. Cette famille est totalement ordonnée. Donc sa réunion est une relation d'équivalence.

Pour la commodité notons G l'ensemble des relations d'équivalence engendrées par les éléments de F. Remarquons que G est un ensemble totalement ordonné de relations d'équivalences stables par réunion non vide.

soit $$x \in E \times E - \cup R_i$$

$$\forall_i, \exists\, G_i \in G \quad \text{tel que} \quad G_i \supset R_i \quad \text{et} \quad x \notin G_i$$

donc $$\cup G_i \in G \supset \cup R_i \quad \text{et} \quad x \notin \cup G_i \ ,$$

2.4 – **Proposition.** *Soit* F *une ultraordonnance sur l'ensemble* I. *Les propositions suivantes sont équivalentes :*

i) F *est maximale dans l'ensemble des ultraordonnances muni de l'inclusion.*

ii) *F est séparée et pour tout $S \in F$, $S \neq \Delta$, de prédecesseur R dans F, S/R est une relation d'équivalence minimale sur E/R.*

Si de plus E est de cardinal n fini, les conditions précédentes sont équivalentes à :

iii) *F est de cardinal n.*

(Sur un ensemble E une relation d'équivalence R est dite minimale si elle est minimale pour l'inclusion dans l'ensemble des relations d'équivalence différentes de la diagonale de $E \times E$; ce qui équivaut à dire qu'il y a une et une seule classe modulo R, ayant 2 éléments ; toute autre classe ayant un et un seul élément).

2.5 – **Lemme.** *Toute ultraordonnance séparée est l'intersection des ultraordonnances maximales qui la contiennent.*

§3 – 3.1 – Soit un ensemble fini de cardinal n, Q un ordre total sur I et P une relation d'équivalence sur I : la relation R sur I/P définie par : "sRt si et seulement si le plus petit élément de s pour Q est inférieur ou égal au plus petit élément de t pour Q", est un ordre total sur I que nous noterons abusivement Q/P.

Si $s : I \to [n]$ est la bijection définie par Q (i.e. le rang) nous noterons s/P la bijection définie par Q/P de $I/P \to [\text{card } I/P]$.

3.2 – Ainsi la donnée d'une partie à 2 éléments de $[k]$ $(k > 1)$ permet la construction d'une application surjective de $[k]$ dans $[k-1]$, la donnée d'un élément de $\prod_{i=2}^{n} \mathcal{P}_2([i])$ permet donc la construction d'une ultraordonnance maximale sur $[n]$. L'application ainsi construite de $\prod_{i=2}^{n} \mathcal{P}_2([i])$ dans l'ensemble des ultraordonnances maximales sur $[n]$ est une bijection.

3.3 – En conséquence il y a $\prod_{i=2}^{n} \binom{i}{2} = n\,!\,(n+1)\,!\,/2^{n-2}$ ultraordonnances maximales sur un ensemble de cardinal n.

CHAPITRE 2
DISSIMILARITES, ULTRAECART, "LA SOUS-DOMINANTE"

§0 0.1 – Soit I un ensemble totalement ordonné, stable par intersection et réunion. Nous noterons hyper I l'ensemble des sections principales de I :

$$\text{hyper I} = \{]\leftarrow, x\, [,]\leftarrow, x], \text{ où } x \in I \}.$$

Sur $F(X, I)$, ensemble des applications de X dans I, nous définissons un ordre :

$$f \text{ et } g \in F(X, I) : \quad f \leqslant g \iff f(x) \leqslant g(x) \;\; \forall\; x \in X.$$

Sur F (hyper I, $\mathcal{P}(X)$) nous définissons l'ordre :

$$\alpha \text{ et } \beta \in F(\text{hyper } I,\ \mathcal{P}(X)) : \alpha \leqslant \beta \iff \alpha(K) \supseteq \beta(K), \forall\ K \in \text{hyper } I.$$

0.2 – Nous allons définir une application Φ de $F(X, I)$ vers F (hyper I, $\mathcal{P}(X)$) :

Si $f : X \to I$ alors nous avons les propriétés suivantes :

1) $\Phi(f)$ commute aux bornes sup et inf.

2) $f(x) = \sup \ \cap\, \Phi(f)^{-1} \ ([\{x\}, X])$

3) Φ est croissante i.e. si

$$f(x) \leqslant g(x) \quad \forall\, x \in X, \qquad \Phi(f) \leqslant \Phi(g)$$

(en effet

$$\alpha \in \Phi(g)([\leftarrow, y]) \iff \alpha \in g^{-1}([\leftarrow, y]) \text{ donc } g(\alpha) \leqslant y \text{ et } f(\alpha) \leqslant y).$$

0.3 – Nous allons définir une application ψ de F (hyper I, $\mathcal{P}(X)$) vers $F(X, I)$:

Si g : hyper $I \to \mathcal{P}(X)$, $\psi(g)(x) = \sup \cap\, g^{-1}([x], X])$ nous avons les propriétés :

1) $\psi \circ \Phi = Id_{F(X,I)}$

2) ψ est croissante.

En effet : soit $\alpha \leqslant \beta$ deux éléments de F (hyper I, $\mathcal{P}(X)$) il faut comparer $\sup \cap \alpha^{-1}([\{x\}, X])$ et $\sup \cap \beta^{-1}([\{x\}, X])$.

Si $K \in \beta^{-1}([\{x\}, X])$, c'est-à-dire si $x \in \beta(K)$ nous avons :

$$\begin{aligned} \alpha(K) \supseteq \beta(K) \Rightarrow\ & x \in \alpha(K) \Rightarrow K \in \alpha^{-1}([\{x\}, X]) \\ & \Rightarrow \cap\, \alpha^{-1}([\{x\}, X]) \subseteq \cap\, \beta^{-1}([\{x\}, X]) \\ & \Rightarrow \sup \cap \ldots \qquad \leqslant \sup \cap \ldots \\ & \Rightarrow \psi\, \alpha(x) \leqslant \psi\, \beta(x)\ \forall x \in X : \psi \text{ est croissante.} \end{aligned}$$

3) On peut montrer que si un élément g de F (hyper I, $\mathcal{P}(X)$) commute aux bornes inf et sup, on a : $\Phi\, \psi\, (g) = g$.

§ 1 – 1.1 – **Définition** : *on appelle dissimilarité sur un ensemble X une application symétrique de $X \times X$ sur un ensemble I totalement ordonné et stable par intersection et réunion. Une dissimilarité est dite réelle si $I = \overline{\mathbb{R}}$.*

Un ultraécart est une dissimilarité vérifiant l'inégalité ultramétrique :

$$\forall\, x, y, z \in X,\ f(x, y) \leqslant \operatorname{Max}\left(f(x, z),\ f(y, z)\right).$$

1.2 – Soit $f : X \times X \to I$ où I est un ensemble totalement ordonné et stable par intersection et réunion.

f est un ultraécart $\Longleftrightarrow \forall\, K \in$ hyper I, $\Phi\,(f)\,(K)$ est une relation d'équivalence.

1.3 – Un des problèmes de la taxonomie aveugle est de construire des applications de l'ensemble des dissimilarités sur un ensemble fini dans l'ensemble des ultraécarts sur le même ensemble. L'idée qui vient le plus immédiatement à l'esprit est la suivante : à une dissimilarité g sur un ensemble X on fait correspondre ψ (equ $\circ\ \Phi\,(g)$) où equ est l'application de $\mathcal{P}(X \times X)$ dans $\mathcal{P}(X \times X)$ qui à une partie de $X \times X$ fait correspondre la relation d'équivalence engendrée. L'application equ est croissante.

§ 2 – 2.1 – **Théorème** – *Soit u une application de $\mathcal{P}(X \times X)$ dans $\mathcal{P}(X \times X)$ dont l'image est composée de relations d'équivalence. Si u est croissante (i.e. $A \subset B \Rightarrow u(A) \subset u(B)$) alors $\psi\,(u \circ \Phi(g))$ est un ultraécart.*

Démonstration :

1) Symétrie : il faut montrer que :

$$\psi\,(u \circ \Phi(g))\ (x, y) = \psi\,(u \circ \Phi\,(g))\ (y, x),$$

avec

$$\psi\,(u \circ \Phi\,(g))\,(x,\,y) \;=\; \mathrm{Sup}\;\cap\;(u \circ \Phi\,(g))^{-1}\,([\{(x,\,y)\},X\times X]).$$

Soit K un élément de

$$(u \circ \Phi\,(g))^{-1}\;\;([\{(x,\,y)\},X\times X]),\ \text{i.e. } (x,\,y) \in (u \circ \Phi\,(g))\,(K).$$

Mais $(u \circ \Phi\,(g))\,(K)$ est symétrique donc

$$K \in (u \circ \Phi\,(g))^{-1}\;\;([\{y,\,x)\},X\times X])$$

ce qui suffit à démontrer la symétrie.

2) Inégalité ultramétrique :

$$u \circ \Phi\,(g) : \mathrm{hyper}\,I \to \mathcal{P}\,(X\times X)$$

est croissante. Donc l'image réciproque d'une section finissante de $\mathcal{P}\,(X\times X)$ est une section finissante de hyper I.

On veut montrer que :

$$\sup \cap\,(u \circ \Phi(g))^{-1}\;([\{x,\,y)\}\,X\times X]) \leqslant \max\,(\sup\;\cap\,(u \circ \Phi\,(g))^{-1}\;([\{(x,\,z)\},X\times X]),$$
$$\sup\;\cap\,(u \circ \Phi\,(g))^{-1}\,([\{y\,,z)\}\,,X\times X])).$$

Considérons les deux ensembles :

$$(u \circ \Phi\,(g))^{-1}\;([\{(y,\,z)\},X\times X]) \;\text{ et }\; (u \circ \Phi\,(g))^{-1}\;([\{(x,\,z)\},\,X\times X]).$$

Ces deux ensembles sont des sections finissantes dans un ensemble totalement ordonné, donc comparables. Supposons par exemple qu'on aie l'inclusion $\subset$. Alors

$$\cap\,(u \circ \Phi\,(g))^{-1}\;\;([\{(y,\,z),\,X\times X]) \subset \cap\,(u \circ \Phi\,(g))^{-1}\;\;([\{(x\,,z)\,,X\times X])$$

et sup " " $\geqslant$ sup " "

De même :

$$\sup \cap\,(u \circ \Phi(g))^{-1}\;\;([\{(y,\,z)\},X\times X]) \geqslant\; \sup \cap\,(u \circ \Phi(g))^{-1}([\{(x,z)\},X\times X)].$$

$\psi(u \circ \Phi(g))$ est bien un ultraécart.

Corrollaire : ψ (equ $\circ$ Φ $(g))$ *est un ultraécart qui est la borne supérieure des ultraécarts plus petits que g. On l'appelle la sous-dominante : SD(g).*

Démonstration :

1) $\psi(\mathrm{equ} \circ \Phi(g)) \leqslant g$.

En effet equ $\circ \Phi(g) \leqslant \Phi(g)$ (la relation engendrée contient la partie.) Et ψ étant croissante $\psi(\mathrm{equ} \circ \Phi(g)) \leqslant \psi \Phi(g) = g$.

2) C'est le plus grand des ultraécarts inférieurs à g.

Soit h un ultraécart plus petit que g.

$$\Phi(h) \leqslant \Phi(g)$$

$$\Phi(h) = \mathrm{equ} \circ \Phi(h) \leqslant \mathrm{equ} \circ \Phi(g) \qquad \text{(h est un ultraécart)}$$

$$h = \psi \Phi(h) \leqslant \psi(\mathrm{equ} \circ \Phi(g)).$$

2.2 – **Définition** *: Soit f une dissimilarité $X \times X \to I$. L'image réciproque par f de l'ensemble des sections commençantes non vides de I est une ordonnance sur X qu'on appelle l'ordonnance associée à f.*

2.3 – On n'a pas cherché à démontrer et on ne sait donc pas si le fait suivant est vrai :

Soit $f : X \times X \to I$ une dissimilarité ; $\boldsymbol{F}$ l'ordonnance associée.

Alors l'ordonnance associée à la sous-dominante de f est ult $\boldsymbol{F}$. (voir définition chapitre 1).

§3 – 3.1 – Soit I un ensemble totalement ordonné ayant des bornes supérieures quelconques.

Soit $\Phi : X \times X \to I$ une dissimilarité. Il est évident que l'ensemble des ultraécarts plus petits que Φ a un plus grand élément : SD (Φ) :

Soit V cet ensemble d'ultraécarts et soit ψ sa borne supérieure. Soient x, y, z trois éléments de X. Nous devons montrer que

$$\psi(x, y) \leqslant \max(\psi(x, z), \psi(y, z)).$$

Supposons

$$\psi(y, z) = \max(\psi(x, z), \psi(y, z)).$$

Pour tout $\varphi \in V$ nous avons

$$\max (\varphi (x, z), \varphi (y, z)) \leqslant \psi (y, z)$$

donc $$\psi (x, y) \leqslant \psi (y, z).$$

Le seul avantage de la construction 2.1 est de faire apparaître l'application "relation d'équivalence engendrée".

3.2 – Maintenant, si I est un ensemble totalement ordonné ayant des bornes supérieures quelconques et des bornes inférieures pour les parties finies non vides, on montre que :

$SD(\Phi)(x, y) = \text{Sup} \{\inf \Phi (t, y), t \in F, z \in F / F$ parcourant les parties finies de X contenant x et $y\}$.

Pour montrer cela, on démontre successivement que : $SD(\Phi) \leqslant \Phi$, que $SD(\Phi)$ est croissante en Φ, que si Φ est un ultraécart $\Phi = SD(\Phi)$ et enfin que $SD(\Phi)$ est un ultraécart.

CHAPITRE 3
CONTINUITE DE LA SOUS-DOMINANTE D'UNE DISSIMILARITE REELLE SUR UN ENSEMBLE FINI

§0 – Dans [4], Jardine et Sibson démontrent la continuité de l'application qui à une dissimilarité réelle sur un ensemble fini fait correspondre sa sous-dominante.

Nous en présentons ici une démonstration formelle.

Dans [1], (pages 196 et 197, note 2) J.P. Benzécri et ses collaborateurs nient un tel résultat sans démonstration en s'appuyant sur des arguments mathématiques qui oublient simplement que sur un ensemble fini il n'y a pas que la topologie discrète.

§1 – 1.1 – **Définition** : *Soient X et Y deux espaces topologiques, une relation* $\Gamma : X \to Y$ *est dite d'image réciproque ouverte si et seulement si* Γ^{-1} *(O) est ouvert pour tout ouvert 0 de Y (en abréviation I.R.O.).*

1.2 – **Lemme** : *Soient X et Y deux espaces topologiques et* $\Gamma : X \to Y$ *une relation. Les propriétés suivantes sont équivalentes :*

i) *pour tout ouvert F de Y,* Γ^{-1} *(F) est ouvert,*

ii) *il existe une représentation* $\begin{array}{ccc} G & \xrightarrow{g} & X \\ f \downarrow & & \\ Y & & \end{array}$ *de* Γ *où g est une application ouverte (i.e. g (0) est ouvert pour tout ouvert 0 de G) et f est une application continue,*

iii) *l'application :* $G_\Gamma \to X$ *est ouverte lorsqu'on munit* G_Γ *de la topologie image réciproque par :* $G_\Gamma \to Y$ *de la topologie de Y,*

iv) *l'application* $\Gamma : X \to \mathcal{P}(Y)$ *est continue lorsqu'on munit* $\mathcal{P}(Y)$ *de la topologie la moins fine parmi celles qui vérifient : pour toute partie fermée F de Y,* $\mathcal{P}(F)$ *est fermée.*

Corollaire :

i) *La composée de deux I.R.O est une I.R.O.*

ii) *I.R.O. est une propriété universelle. (i.e. si* $G : X \to Y$ *est I.R.O. alors pour tout espace topologique Z :* $G \times 1_z : X \times Z \to Y \times Z$ *est I.R.O.).*

§2 – 2.1 – **Lemme** : *Soit* $f : X \to \bar{R}$ *une application.*

i) *si elle est s.c.i. (i.e. pour tout k,* $f^{-1}(]k, +\infty])$ *est ouvert) alors* $f_\leqslant$ *définie par* $f_\leqslant(x) = \{y \mid y \in \bar{R} \text{ et } y \leqslant f(x)\}$ *est I.R.O.*

ii) *si elle est s.c.s. alors* $f_\geqslant$ *est I.R.O.*

Démonstration

i) $f_\leqslant^{-1}(]\alpha,\beta[) = \{x \mid x \in X \text{ et } f_\leqslant(x) \cap]\alpha,\beta[\neq \phi\}$

$$= \{x \mid f(x) > \alpha\} = f^{-1}(]\alpha, +\infty[)$$

2.2 – **Lemme** : *Si* $\Gamma : X \to \bar{R}$ *est I.R.O alors* Γ_m *définie par* $\Gamma_m(x) =$ sup. $\Gamma(x)$ *est s.c.i.*

Démonstration

$$\begin{aligned} \Gamma_m^{-1}(]\alpha, +\infty[) &= \{x / \Gamma_m(x) > \alpha\} \\ &= \{x / \sup \Gamma(x) > \alpha\} \\ &= \{x / \Gamma(x) \cap]\alpha, +\infty[\neq \phi\} \\ \text{(Rappelons que } \sup \Phi = +\infty) \quad &= \Gamma^{-1}(]\alpha, +\infty]) \end{aligned}$$

§3 – 3.1 – **Lemme** : *Soit $\varphi : X \times Y \to \bar{R}$ une application s.c.i. et $\Gamma : X \to Y$ une relation I.R.O. L'application $\Pi : X \to \bar{R}$ définie par $\Pi(x) = \sup\{\varphi(x, y) / y \in \Gamma(x)\}$ est s.c.i.*

Posons $\gamma = \varphi_{\leqslant}(y)$, γ est I.R.O. Nous avons $\Pi = \operatorname{Sup} \gamma$ d'où la conclusion.

3.2 – **Corollaire** : *Soit (X, d) un espace métrique et $\Gamma : X \to X$ une relation I.R.O. et U.I.R.F. telle que $\Gamma(x) \neq \phi$ pour tout x. Alors l'application qui à x fait correspondre $d(x, \Gamma(x))$ est continue et la relation qui à x fait correspondre $\{y / y \in \Gamma(x)$ et $d(x, y) = d(x, \Gamma(x))\}$ est* U.I.R.F.

Proposition :

§4 – 4.1 – *Soit I un ensemble fini, $\mathcal{D}$ l'ensemble des dissimilarités sur I, la relation Γ qui à d fait correspondre l'ensemble des ultraécarts sur I plus petits que d est I.R.O. et U.I.R.F.*

Le fait qu'elle soit U.I.R.F est relativement trivial. On applique le théorème classique : soit $f : X \to Y$ une application continue, où X est séparé et Y localement compact, alors f est universellement fermée si et seulement si $f^{-1}(K)$ est compact, pour toute partie compacte K de Y.

Soit $p : \Gamma \to \mathcal{D}$ la première projection de Γ. Elle est continue, Γ est séparée et $\mathcal{D}$ localement compact.

Soit K un compact de $\mathcal{D}$, il est contenu dans le cube

$$K_\alpha = \{f / 0 \leqslant f(x, y) \leqslant \alpha, \ \forall (x, y) \in I^2\}$$

$f^{-1}(K)$ est fermé et contenu dans le cube $K_\alpha \times K_\alpha$ donc est compact.

Le fait que Γ soit I.R.O est un peu plus subtil. Du fait de la structure de l'ensemble des ultraécarts sur I et du fait que la réunion de relations I.R.O est une relation I.R.O. on se ramène à démontrer le fait suivant :

Soient I un ensemble fini, Q un préordre total sur I et C le cône des fonctions positives croissantes, la relation $\Gamma : R^I_+ \to R^I_+$ qui à d fonction sur I fait correspondre l'ensemble des éléments de C plus petits que d est I.R.O.

Soit $\Pi \,]a_i, b_i[$ un pavé ouvert de R^I, $a_i < b_i$ et $b_i > 0$. Nous avons : $(\Pi \,]a_i, b_i[) \cap C \neq \phi$ si, et seulement si, pour tout $i, \delta(i) = \operatorname{Sup}_{j Q i}\{a_j < b_i\}$: en effet, soit d croissante et appartenant à $\Pi \,]a_i, b_i[$, nous avons : $d(i) \geqslant \delta(i)$ pour tout i ; inversement supposons que pour tout $i, \delta(i) < b_i$, notons ϵ la quantité $\inf (b_i - \delta_i)/2$, $\delta + \epsilon \in C \cap (\Pi \,]a_i, b_i[)$.

Supposons donc que $C \cap (\Pi \,]a_i, b_i[) \neq \phi$, alors $\Gamma^{-1}(\Pi(]a_i, b_i[)) = \{ d / d(i) \geqslant \epsilon(i) \text{ si } \delta(i) = 0$ et $d(i) > \delta(i) \text{ si } \delta(i) \neq 0 \}$

est un ouvert de $\boldsymbol{R}^I_+$.

4.2 – **Corollaire** : *L'application qui à une dissimilarité d fait correspondre sa sous-dominante est continue.*

CHAPITRE 4

L'ULTRAMETRIQUE LA PLUS PROCHE AU SENS DES MOINDRES CARRES D'UNE DISSIMILARITE REELLE SUR UN ENSEMBLE FINI

§0 – 0.0 – Soit X un ensemble fini. On notera $\theta(X)$ l'ensemble des ordonnances sur X et $\mathcal{F}_S(X \times X, \boldsymbol{R})$ l'ensemble des dissimilarités réelles finies sur X.

0.1 – A $g \in \mathcal{F}_S(X \times X, \boldsymbol{R})$ on sait faire correspondre une ordonnance notée $\theta(g)$ qui est l'ensemble des images réciproques des sections commençantes de $\boldsymbol{R}$.

0.2 – Si g est une dissimilarité, g est un ultraécart si et seulement si $\theta(g)$ est une ultraordonnance sur X.

0.3 – Soit F une ordonnance sur X. $\theta^{-1}([\leftarrow, F])$ est le cône fermé des fonctions de $X \times X$ dans $\boldsymbol{R}$ qui sont croissantes pour le préordre Q défini sur $X \times X$ par F (voir chap. 1). On notera $\mathcal{C}(F)$ ce cône. C'est aussi le cône convexe fermé engendré par les vecteurs qui sont les fonctions carractéristiques des sections finissantes non vides de $X \times X$ pour le préordre Q et l'opposé des fonctions caractéristiques des sections commençantes non vides de $X \times X$ pour Q. On a :

$$F \subset G \Rightarrow \mathcal{C}(F) \subset \mathcal{C}(G).$$

0.4 – Si F est une ultraordonnance, $\mathcal{C}(F)$ est un ensemble d'ultraécarts. Donc l'ensemble des ultraécarts est la réunion des $n!\ (n+1)!/2^{n-2}$ cônes associés aux ultraordonnances maximales.

§1 – 1.0 – Sur $\mathfrak{F}_S(X \times X, R)$, on peut mettre la distance hilbertienne définie par la base "canonique" Dans ces conditions on considère la relation MC de $\mathfrak{F}_S$ sur $\mathfrak{F}_S$ qui à une dissimilarité fait correspondre l'ensemble des ultraécarts les plus proches au sens des moindres carrés de cette dissimilarité.

1.1 – Ce qui précède et les théorème de projections montrent que pour tout d appartenant à $\mathfrak{F}_S$, MC (d) est fini non vide.

Nous allons montrer que :

1/ $\{ d \in \mathfrak{F}_S / \mathrm{MC}(d)$ a plus d'un élément$\}$ est de mesure de Lebesgue nulle.

2/ La relation MC est U.I.R.F. [3]

3/ La relation MC est calculable.

§2 – 2.0 – Soit C une partie convexe fermée d'espace de Hilbert H. Soit P la projection de H sur C. Pour tout x appartenant à C on sait que

$$y \in P^{-1}(\{x\}) \iff < y - x, x - z > \geqslant 0, \forall z \in C.$$

Ceci définit clairement $P^{-1}(\{x\})$ comme une intersection de demi-espace fermé $P^{-1}(\{x\})$ est donc clairement un cône convexe de sommet x. Alors :

2.1 – **Définition** : *Soit C un cône convexe de sommet 0 dans l'espace de Hilbert H. Soit P la projection sur C. Le cône convexe fermé $P^{-1}(\{0\})$ s'appelle le cône dual de C et se note $C^{\perp}$.*

2.2 – **Lemme de Moreau** : *Avec les notations de 2.1 nous avons :*

1/ $< P, Id - P > = 0$,

2/ *l'opérateur projecteur sur $C^{\perp}$ est Id-P,*

3/ *si* $y = x_1 + x_2$ avec $x_1 \in C$ et $x_2 \in C^{\perp}$ et $x_1 \perp x_2$, *alors* $x_1 = P(y)$.

Démonstration :

1/ Soit $y \in H$ nous devons démontrer que $<P(y), y - P(y)> = 0$.
Si $P(y) = 0$, c'est vrai.
Si $P(y) \neq 0$, les points 0 et $P(y)$ définissent une demi-droite du cône et $P(y)$ est la projection de y sur la droite en question. D'où la conclusion.

2/ Il faut démontrer d'abord que pour tout y, $y - P(y) \in C^{\perp}$ cequi est évident puisque $< y - P(y), P(y) - z > \geqslant 0 \; \forall z \in C$ et le premier membre de l'inégalité égale $< y - P(y), - z >$ i.e. 0 est la projection de $y - P(y)$ sur C.

Soit $z \in C^{\perp}$. Considérons $< y - (y - P(y)), y - P(y) - z >$

$$= < P(y), y - P(y) - z >$$
$$= < P(y), -z >$$
$$= - < P(y), z > \text{ qui est positif ou nul.}$$

3/ On peut écrire

$$< y - x_1 , x_1 - z > = < x_2 , x_1 - z > = - < x_2 , z > \geqslant 0$$

puisque $\qquad x_2 \in C^{\perp} \quad$ et $\quad z \in C.$

Ce qui montre que $x_1 = P(y)$.

2.3 – Nous allons étudier le cône dual d'un cône simplicial et l'opérateur projection sur un tel cône. Soit $(u_i)_{i \in I}$ une base de l'espace de Hilbert de dimension finie H. Pour tout i définissons le vecteur $\widetilde{u}_i$ par les équations :

$$< \widetilde{u}_i , \widetilde{u}_j > = 0 \quad \text{si} \quad i \neq j \quad ; \quad < \widetilde{u}_i , \widetilde{u}_i > = -1.$$

Il est facile de voir que les $\widetilde{u}_i$ forment une base de H et que le cône convexe fermé engendré par les $\widetilde{u}_i$ est $C^{\perp}$.

2.4 – Ce qui précède, joint au lemme de Moreau *montre que tout $x \in H$ s'écrit de manière unique sous la forme :*

$$x = \sum_{i \in K} \alpha_i u_i + \sum_{j \in \overline{K}} \beta_j \widetilde{u}_j \quad \textit{avec} \quad \alpha_i \geqslant 0 , \beta_j > 0$$

et K une partie de I.

L'ensemble des x qui s'écrivent de la manière précédente pour une partie K de I fixée s'appellera la région $\mathcal{K}$ de l'espace.

Soit H_k le sous-espace engendré par les u_i où $i \in K$; nous avons $P(x) = P_{H_k}(x)$ si $x \in$ région $\mathcal{K}$.

2.5 – **Corollaire** : *Dans un espace de Hilbert de dimension finie la projection sur un cône simplicial est calculable.*

2.6 – **Corollaire** : *Soit C_1 et C_2 deux cônes simpliciaux de sommet 0, l'ensemble des points x de H défini par*

$$\| x - P_{C_1}(x) \| = \| x - P_{C_2}(x) \| \quad \text{et} \quad P_{C_i}(x) \neq P_{C_2}(x)$$

est de mesure de Lebesgue nulle.

Démonstration :

1er cas : C_1 et C_2 sont deux sous-espaces vectoriels. Si $C_1 = C_2$, l'ensemble en question est vide. Sinon

$$\| x - P_{C_1}(x) \|^2 - \| x - P_{C_2}(x) \|^2$$

est un polynome homogène de degré 2 non identiquement nul. L'ensemble des points qui l'annule est de mesure de Lebesgue nulle.

2ème cas : C_1 et C_2 sont deux cônes simpliciaux. Auquel cas l'ensemble en question est une réunion finie d'ensembles de cette forme.

2.7 – **Corollaire** : *L'ensemble des points de* $\mathcal{F}_S(X \times X, R)$ *où MC a plus d'un élément est de mesure nulle.*

§3 – **Lemme** : *Soit* $(C_i)_{i \in I}$ *une famille finie de convexes fermés d'un espace de Hilbert. La relation* $\Gamma : x \to$ *l'ensemble des points de* $\bigcup_{i \in I} C_i$ *les plus proches est U.I.R.F.*

Démonstration :

$$\Gamma(x) = \bigcup_{i \in I} \left(P_{C_i}(x)\right)$$

est U.I.R.F. et I.R.O. parce que les P_{C_i} sont continues. $\Gamma(x) \neq \phi$ et on applique le corollaire 3.2 du chapitre 3.

REMARQUES : CONTRE-EXEMPLES

Remarque 1

Les algorithmes ascendants connus de la littérature ne permettent pas d'atteindre l'ultramétrique la plus proche au sens des moindres carrés, (et plus particulièrement l'algorithme dit "de la moyenne").

Considérons en effet le contre-exemple suivant.

Soit

$$A = \{1, 2, \ldots, p\}, \; I = A \cup \{e\}.$$

Soit d un indice de distance sur $I \times I$ défini par

$$\begin{aligned}
&d(1, e) = \epsilon \\
&d(x, y) = \alpha \quad \forall \underset{x \neq y}{x, y} \in A \\
&d(x, y) = \beta \text{ ailleurs où } \epsilon < \alpha < \beta.
\end{aligned}$$

a) Soit δ_M l'ultramétrique dite "de la moyenne".

δ_M est égale à :

$$\begin{aligned}
&\delta_M(1, e) = \epsilon \\
&\delta_M(x, y) = \alpha \quad \forall \underset{x \neq y}{x, y \in A - \{1\}} \\
&\delta_M(x, y) = \frac{\alpha + \beta}{2} \text{ ailleurs ;}
\end{aligned}$$

alors

$$\| \delta_M - d \|^2 = (p - 1) \left(\frac{\beta - \alpha}{2}\right)^2.$$

b) Soit δ l'ultramétrique définie par :

$$\begin{aligned}
&\delta(1, e) = \beta \\
&\delta \quad = d \text{ ailleurs}
\end{aligned}$$

alors

$$\| \delta - d \|^2 = (\beta - \epsilon)^2.$$

On voit que δ n'approxime mieux d au sens des moindres carrés que si et seulement si

$$i) \qquad (\beta - \epsilon)^2 < \frac{p - 1}{2} (\beta - \alpha)^2 \quad \text{et} \quad \epsilon < \alpha < \beta.$$

Or l'ensemble des $(\alpha, \beta, \epsilon, p)$ tels que $i)$ soit vérifié est non vide. (Ex. $(2, 7, 1, 4)$ vérifie $i)$).

Remarque 2

Les algorithmes divisifs (ou descendants) connus n'ont pas permis d'atteindre l'ultramétrique au sens des moindres carrés.

Soit $I = \{1, 2, 3, 4\}$. Soit d un indice de distance sur $I \times I$ défini par le tableau suivant :

d	1	2	3	4
1	0			
2	2	0		
3	10	1	0	
4	6	9	4	0

a) L'algorithme descendant (utilisant la distance moyenne comme distance entre deux parties) donne l'ultramétrique δ définie par le tableau :

δ	1	2	3	4
1	0			
2	2	0		
3	26/4	26/4	0	
4	26/4	26/4	4	0

$$\|\delta\|^2 = (2)^2 + (4)^2 + \frac{(26)^2}{4}$$

b) Soit δ_0 l'ultramétrique suivante :

δ_0	1	2	3	4
1	0			
2	6	0		
3	6	1	0	
4	19/3	19/3	19/3	0

$$\|\delta_0\|^2 = 1 + 2 \times 6^2 + \frac{19^2}{3} > \|\delta\|^2$$

Nous avons donc

$$\|\delta_0 - d\|^2 < \|\delta - d\|^2$$

et pourtant

$$\delta\,(\{1\,,2\}\,,\{3\,,4\}) = \frac{26}{4} > \delta_0\,(\{4\}\,,\{1\,,2\,,3\}) = \frac{19}{3}.$$

CHAPITRE 5
ALGORITHME DE PROJECTION D'UNE FONCTION SUR UN CONE DE FONCTIONS CROISSANTES

§0 – 0.0 – Soit R un préordre total sur l'ensemble fini I. Soit $C\,(R)$ le cône des fonctions croissantes pour R. Nous allons maintenant donner un algorithme très performant du calcul de la projection sur $C\,(R)$ d'une fonction g à valeurs sur I. Soulignons en passant que cet algorithme est né d'une idée lumineuse de Yves CESARI.

§1 – 0.0 – **Proposition.** *Soit R un préordre total sur l'ensemble fini I. Soit $C(R)$ le cône des fonctions croissantes de I dans $\mathbf{R}$. Munissons $\mathfrak{F}(I\,,\mathbf{R})$ du produit scolaire $\sum_{x\in I} f\,(x)\,g\,(x)$. Alors $f \in C\,(R)$ est la projection de $g \in \mathfrak{F}(I\,,\mathbf{R})$ si et seulement si les deux conditions suivantes sont vérifiées*

i) pour tout $x \in I$ $\sum_{R\,(x)} (g - f)\,(y) \leqslant 0$,

ii) pour tout $x \in I$ $\sum_{f^{-1}f(x)} (g - f)\,(y) = 0$.

Démonstration :

$\Rightarrow$) i) Soit $z = f + X_{R\,(x)}$;

nous avons :

$$< g - f\,,\,f - z > = (-1)\sum_{R\,(x)} (g - f)\,(y) \geqslant 0$$

d'après la caractérisation rappelée dans 2.0 du chapitre 4.

ii) la fonction $f + (k - f(x))\, X_{f^{-1}f(x)}$ est encore un élément de $C(R)$ et on a :

$$< g - f\,,\, f - (f + (k - f(x))\, X_{f^{-1}f(x)} > \;\geqslant\; 0$$

i.e.

$$(f(x) - k) \sum_{y \in f^{-1}f(x)} (g - f)(y) \geqslant 0\,.$$

D'où la conclusion puisque $f(x) - k$ est positif ou négatif.

$\Leftarrow$ $C(R)$ est le cône convexe engendrée par les fonctionsde la forme $X_{R(x)}$ et $X_{R^{-1}(x)}$.

Soit $\lambda \geqslant 0$. Nous voulons d'abord montrer que :

$$< g - f\,,\, f - \lambda\, X_{R(x)} > \;\geqslant\; 0.$$

Nous avons :

$$\begin{aligned}
< g - f, f - \lambda\;\; X_{R(x)} > \; = & \sum_{R^{-1}(x) - f^{-1}f(x)} (g - f)(y)\, f(y) \\
& + \quad f(x) \sum_{R^{-1}(x)\, \cap f^{-1} f(x) - \text{classe de } x \text{ modulo } R} (g - f)(y) \\
& + \sum_{R(x) \cap f^{-1} f(x)} (g - f)(y)\, (f(x) - \lambda) \\
& + \sum_{R(x) - f^{-1} f(x)} (g - f)(x)\, .\, (f(x) - \lambda) \\
& = - \lambda \sum_{R(x)\, \cap f^{-1} f(x)} (g - f)(y)
\end{aligned}$$

à cause de ii) et ceci est positif à cause de i et ii). La démonstration est identique pour les générateurs de la forme $- X_{R^{-1}(x)}$.

D'où la conclusion.

§2 – 2.1 – On note $(A_i)_{1 \leqslant i \leqslant n}$ la suite des classes d'équivalence. Remarquons que la projection f de g sur $C(R)$ est constante sur chaque A_i.

L'algorithme construit, par une récurrence descendante en n pas, une suite de fonctions f_k à valeurs sur $\bigcup_{i \geqslant k} A_i$ vérifiant les propriétés *i)* et *ii)* de la proposition sur $\bigcup_{i \geqslant k} A_i$. La projection cherchée est alors f_1.

2.2 – **Algorithme**

– On construit f_n : $f_n(x) = (1/\operatorname{card} A_n) \sum_{A_n} (g(y)), \forall x \in A_n, f_n$ vérifie évidemment *i)* et *ii)* sur A_n.

– Supposons qu'on aie construit f_k, à valeurs sur $\bigcup_{i \geqslant k} A_i$, qui vérifie donc :

$$\text{i')} \quad \forall\, x \in \bigcup_{i \geqslant k} A_i \quad \sum_{f_k^{-1} f_k(x)} (f_k - g)(y) = 0$$

$$\text{ii')} \quad \sum_{\substack{\cup A_i \\ i \geqslant h}} (f_k - g)(x) \geqslant 0 \quad \text{pour tout } h \geqslant k$$

et notons $m(l, h)$ la moyenne de g sur $\bigcup_{l \leqslant i \leqslant h} A_i$:

$$m(l,h) = \Big(1/\operatorname{card} \bigcup_{l \leqslant i \leqslant h} A_i\Big) \sum_{\substack{\cup A_i \\ l \leqslant i \leqslant h}} g(x).$$

Si h_1 est la borne inférieure dans $[1, n] \cap N$ des h qui vérifient :

$$m_{(k-1,h)} \leqslant f_k(A_{h+1})$$

on pose

$$\begin{cases} f_{k-1}(A_i) = m(k-1, h_1) & \text{pour} \quad k-1 \leqslant i \leqslant h_1 \\ f_{k-1}(A_i) = f_k(A_i) & \text{pour} \quad i > h_1 . \end{cases}$$

f_{k-1} est à valeurs sur $\bigcup\limits_{i \geqslant k-1} A_i$. Montrons que sur cet ensemble f_{k-1} vérifie *i*) et *ii*).

i) Si $x \in \bigcup\limits_{k-1 \leqslant i \leqslant h_1} A_i$,

$$f_{k-1}^{-1} f_{k-1}(x) = \bigcup_{k-1 \leqslant i \leqslant h_1} A_i$$

donc

$$\sum_{f_{k-1}^{-1} f_{k-1}(x)} (f_{k-1} - g)(y) = \sum_{f_{k-1}^{-1} f_{k-1}(x)} (m(k-1, h_1) - g(y)) = 0.$$

Si $x \in A_i$ avec $i > h_1$,

$$\sum_{f_{k-1}^{-1} f_{k-1}(x)} (f_{k-1} - g) = \sum_{f_k^{-1} f_k(x)} (f_k - g) = 0.$$

ii) Soit $h \geqslant k$. Sur $\bigcup\limits_{i \geqslant k} A_i$

$$f_{k-1} \geqslant f_k$$

d'où

$$\sum_{\substack{\cup A_i \\ i \geqslant h}} (f_{k-1} - g)(x) \geqslant \sum_{\substack{\cup A_i \\ i \geqslant h}} (f_k - g)(x) \geqslant 0 .$$

Il reste à montrer que :

$$\sum_{\substack{\cup A_i \\ i \geqslant k-1}} (f_{k-1} \quad g)(x) \geqslant 0$$

mais

$$\sum_{\substack{\cup A_i \\ i \geq k-1}} (f_{k-1} - g)(x) = \sum_{\substack{\cup A_i \\ k-1 \leq i \leq h_1}} (m(k-1, h_1) - g(x)) + \sum_{\substack{\cup A_i \\ i > h_1}} (f_k - g)(x)$$

Le premier terme est nul, le second positif par récurrence. D'où la conclusion.

BIBLIOGRAPHIE

[1] BENZECRI – La taxinomie. Dunod. Paris 1973.

[2] CORMACK R.M., 1971 – A review of classification. – *Journal of the Royal Statiscal Society.* p. 321-367.

[3] HAKIM, LOCHARD, OLIVIER et TEROUANNE, 1973. – Sur les traces de Spearmann (*cahier du B.U.R.O. n° 25*).

[4] JARDINE et SIBSON – Mathematical taxonomy. Wiley New-York 1971.

[5] LANCE-WILLIAM, 1967. – A general theory of classificatory sorling strategies I & II – *Comp J,* 9, 373-380
Comp J, 10, 271-277.

[6] KRASNER M., 1963. – Cours de D.E.A. de théorie des nombres.Paris.

[7] ROUX M. 1969. – An algorithm to construct a particular kind of taxonomy. Numerical taxonomy (AJ Cole). New-York : Academic Press.

[8] SIBSON, 1970. – A model for taxonomy I & II. *Math. Biosciences,* 6, 405-430.