PH. BOURGEOIS

Extension de la méthode Procuste à la comparaison de nuages de points situés dans des espaces de dimensions différentes

RAIRO. Recherche opérationnelle, tome 16, nº 1 (1982), p. 45-63
<http://www.numdam.org/item?id=RO_1982__16_1_45_0>

R.A.I.R.O. Recherche opérationnelle/Operations Research
(vol. 16, n° 1, février 1982, p. 45 à 63)

EXTENSION DE LA MÉTHODE PROCUSTE A LA COMPARAISON DE NUAGES DE POINTS SITUÉS DANS DES ESPACES DE DIMENSIONS DIFFÉRENTES (*)

par Ph. BOURGEOIS (1)

Résumé. — *On rappelle d'abord les principaux résultats connus de la méthode Procuste et les difficultés qu'elle soulève. On montre ensuite qu'en modifiant légèrement le problème initial, on peut se ramener à la recherche du maximum d'une fonction concave sur un domaine convexe, ce qui permet d'obtenir une caractérisation de l'optimum global.*

Mots-clés : Superposition de nuages, décomposition en valeurs singulières, programme concave, reconnaissance de forme.

Abstract. — *First, the main points of Procrustes problem and its difficulties are summarized. Next, a slight modification of the initial problem is suggested, which reduces it to maximization of a concave function over a convex set. Necessary and sufficient conditions for a global optimum are given.*

Key words : Superposing configurations, singular value decomposition, concave programming, pattern recognition.

1. RÉSULTATS DE LA MÉTHODE PROCUSTE CLASSIQUE

1.1. Méthode Procuste pour deux nuages de points

Dans un espace $E=\mathbb{R}^n$ on considère deux nuages de points indicés par un même ensemble fini I de cardinal m. On note X_1 (resp. X_2) le tableau à n lignes et m colonnes des coordonnées du premier (resp. deuxième) nuage.

On cherche à vérifier si les deux nuages ont une *forme* identique, i. e. s'ils sont deux représentations euclidiennes de l'ensemble I associées à une même distance définie sur $I \times I$. Pour ce faire on déplace le nuage X_2 sans le déformer, à l'aide d'une translation et d'une isométrie linéaire, de façon à superposer le mieux possible les deux nuages :

$$X_2 \varsubsetneq PX_2 + tu',$$

(*) Reçu en avril 1981.
(1) Université Paris 10, Nanterre.

où P est une matrice d'isométrie : $P'P=PP'=I_n$; t est un vecteur de translation : $t \in \mathbb{R}^n$; u' est le vecteur $(1, 1 \ldots 1)$; $u \in \mathbb{R}^m$.

On mesure la qualité de la superposition des deux nuages par leur distance résiduelle après déplacement de X_2. On utilise pour cela la distance *euclidienne* usuelle entre matrices de mêmes dimensions, définie par :

$$d^2(A, B)=\operatorname{tr}[(A-B)(A-B)']$$

(où tr désigne la trace d'une matrice carrée).

d^2 est associée de façon naturelle au produit scalaire :

$$\langle A, B \rangle = \operatorname{tr}(AB').$$

On cherche donc le vecteur de translation t et la matrice d'isométrie P qui rendent minimum la quantité :

$$d^2(X_1, PX_2+tu')$$

On montre [1] que la translation optimale de X_2 fait coïncider les centres de gravité des deux nuages et s'écrit :

$$t=-(\frac{1}{m}PX_2-X_1)u.$$

Pour alléger les notations on peut supposer, sans perte de généralité, que les nuages X_1 et X_2 sont désormais centrés sur l'origine des coordonnées. On montre alors que la matrice P est optimale si et seulement si elle maximise la quantité : $\operatorname{tr}(PX_2X_1')$.

La solution de ce problème est connue :

on effectue la *décomposition en valeurs singulières* [4] de la matrice X_2X_1' :

$$X_2X_1'=UDV', \tag{1}$$

où $U'U=UU'=I_n$, $V'V=VV'=I_n$ et D est une matrice diagonale dont les éléments diagonaux, *positifs ou nuls*, sont appelés : valeurs singulières de la matrice X_2X_1'.

On tire de (1) :

$$\operatorname{tr}(PX_2X_1')=\operatorname{tr}(PUDV').$$

En utilisant l'inégalité de Schwarz on montre [2] que :

$$\operatorname{tr}(PUDV') \leqq \operatorname{tr}(D),$$

ce qui s'écrit encore :

$$\operatorname{tr}(PX_2 X_1') \leqq \operatorname{tr}[(X_1 X_2' X_2 X_1')^{1/2}]. \tag{2}$$

Il y a égalité si et seulement si : $P = VU'$.

Lorsque la matrice $X_2 X_1'$ est régulière, P est déterminée de façon univoque et s'écrit :

$$P = X_1 X_2' \, (X_2 X_1' X_1 X_2')^{-1/2}. \tag{3}$$

COROLLAIRE : *A l'optimum la matrice $PX_2 X_1'$ est nécessairement symétrique et définie ou semi-définie positive puisqu'on a :*

$$PX_2 X_1' = VU' \, UDV' = VDV'. \tag{4}$$

Enfin, on montre que l'écart minimal entre les deux nuages X_1 et X_2 (centrés) vaut :

$$d_P^2(X_1, X_2) = \operatorname{tr}(X_1 X_1') + tr(X_2 X_2') - 2 \operatorname{tr}[(X_1 X_2' X_2 X_1')^{1/2}]. \tag{5}$$

1.2. Méthode Procuste pour plus de deux nuages

On considère maintenant un ensemble $\{X_1, X_2 \ldots X_l\}$ de l nuages de points situés dans $\mathbb{R}^n$ et indicés par le même ensemble I de cardinal m. On désire superposer simultanément ces nuages.

En général il est impossible de réaliser simultanément le minimum de tous les écarts entre nuages pris deux à deux. Pour pallier cette difficulté, Gower a suggéré [3] de minimiser plutôt la somme de tous ces écarts.

On montre sans difficulté que le minimum n'est atteint que lorsque tous les centres de gravité des nuages sont confondus en un même point, ce qui s'obtient en centrant tous les nuages sur l'origine du repère de coordonnées.

Le « Problème Procuste généralisé » s'écrit alors :

$$\operatorname*{Min}_{(P_1, P_2 \ldots P_l)} \sum_h \sum_{<k} d^2(P_h X_h, P_k X_k),$$

$$P_h' P_h = I; \qquad h = 1, 2 \ldots l.$$

On montre que ce problème équivaut à la recherche du maximum, sous les mêmes contraintes, de la quantité :

$$\sum_h \sum_{<k} \operatorname{tr}(P_h X_h X_k' P_k')$$

Condition nécessaire d'optimum [5]

Pour tout indice h_0 fixé, le nuage $P_{h_0} X_{h_0}$ doit être placé à distance minimale de la somme de tous les autres nuages, i. e. du nuage :

$$C_{h_0} = \sum_{h \neq h_0} P_h X_h.$$

Pour démontrer cette condition, il suffit d'isoler les termes dépendant de l'indice h_0 dans l'expression à maximiser :

$$\sum_h \sum_{<k} \operatorname{tr}(P_h X_h X'_k P'_k) = \operatorname{tr}(P_{h_0} X_{h_0} C'_{h_0}) + \left\{ \begin{array}{l} \text{des termes indé-} \\ \text{pendants de } h_0 \end{array} \right\} \tag{6}$$

Or on a vu au paragraphe 1.1 que la quantité $\operatorname{tr}(P_{h_0} X_{h_0} C'_{h_0})$ est maximale quand P_{h_0} est l'isométrie plaçant le nuage X_{h_0} *à distance minimale du nuage* C_{h_0}.

COROLLAIRE : *Il découle de la relation* (4) *qu'à l'optimum les matrices* $W_h = P_h X_h C'_h$, $h \in \{1, 2 \ldots 1\}$, *sont nécessairement symétriques et définies* (*ou semi-définies*) *positives.*

*
* *

L'inconvénient majeur du problème Procuste généralisé est qu'il conduit à maximiser sous contraintes une fonction non concave. Aussi les algorithmes de recherche de l'optimum publiés ([3], [5]) jusqu'ici peuvent converger vers une solution qui n'est pas forcément optimale, même localement, ainsi qu'on peut le constater en construisant un exemple adéquat.

C'est pourquoi nous étudions dans la suite comment modifier le problème initialement posé afin de nous ramener à la résolution d'un *programme concave.*

2. QUELQUES FORMULATIONS ÉQUIVALENTES DU PROBLÈME PROCUSTE

Gower propose (*cf.* supra) pour rapprocher au maximum les nuages X_1, $X_2 \ldots X_l$ les uns des autres de minimiser le critère :

$$\sum_h \sum_{<k} d^2(P_h X_h, P_k X_k).$$

En revenant à la définition de la distance d, et en faisant jouer les propriétés de la trace et les contraintes d'orthogonalité : $P'_h P_h = P'_k P_k = I$, on montre que :

$$d^2(P_h X_h, P_k X_k) = \operatorname{tr}(X_h X'_h) + \operatorname{tr}(X_k X'_k) - 2 \operatorname{tr}(P_h X_h X'_k P'_k) \tag{7}$$

Posons :

$$\overline{X}=\frac{1}{l}\sum_{h=1}^{l} P_h X_h, \tag{8}$$

$\overline{X}$ est le barycentre des l nuages, après déplacement; on peut encore qualifier $\overline{X}$ de nuage-moyen.

En sommant les relations (7) sur h et k et en utilisant (8), on obtient après regroupement des termes :

$$\sum_{h}\sum_{<k} d^2(P_h X_h, P_k X_k)=l\sum_{h=1}^{l} \operatorname{tr}(X_h X'_h)-l^2 \operatorname{tr}(\overline{XX'}). \tag{9}$$

Par un procédé analogue on montre également qu'on a :

$$\sum_{h=1}^{l} d^2(P_h X_h, \overline{X})=\sum_{h=1}^{l} \operatorname{tr}(X_h X'_h)-l \operatorname{tr}(\overline{XX'}). \tag{10}$$

Enfin, en rapprochant les relations (9) et (10) on établit la double identité :

$$\sum_{h}\sum_{<k} d^2(P_h X_h, P_k X_k)=l\sum_{h=1}^{l} \operatorname{tr}(X_h X'_h) -l^2 \operatorname{tr}(\overline{XX'})=l\sum_{h=1}^{l} d^2(P_h X_h, \overline{X}).$$

On voit ainsi que le problème de Gower peut être présenté sous deux autres formes équivalentes :

– trouver le nuage-moyen après déplacement, $\overline{X}=1/l\sum_{h=1}^{l} P_h X_h$, dont l'inertie, i.e. tr $(\overline{XX'})$ soit maximale;

– trouver le nuage-moyen après déplacement qui minimise la dispersion de la classe de nuages $(X_1, X_2 \ldots X_l)$, i.e. la quantité $\sum_{h=1}^{l} d^2(P_h X_h, \overline{X})$.

A partir de cette formulation, il est facile de voir que l'optimum ne peut être atteint que si chaque nuage $P_h X_h$ se trouve à distance minimale du nuage-moyen $\overline{X}$. Chaque distance $d^2(P_h X_h, \overline{X})$ coïncide alors avec l'écart de Procuste $d_P^2(X_h, \overline{X})$ qui mesure, rappelons-le, la qualité de la superposition des deux nuages.

Réciproquement

Supposons que l'on cherche d'emblée le nuage X qui minimise la dispersion des nuages calculée avec l'écart de Procuste, i.e. la fonction-objectif :

$$f(X)=\sum_{h=1}^{l} d_P^2(X_h, X).$$

Considérons un nuage X fixé. Par définition même de l'écart de Procuste, il existe un système de déplacements $P_1, P_2 \ldots P_l$ dépendants de X, tel qu'on puisse écrire :

$$f(X) = \sum_{h=1}^{l} d^2(P_h X_h, X). \tag{11}$$

Mais puisque d est une distance euclidienne, par application du théorème bien connu de Huyghens, on a l'inégalité :

$$\sum_{h=1}^{l} d^2(P_h X_h, X) \geqq \sum_{h=1}^{l} d^2(P_h X_h, \overline{X}). \tag{12}$$

On a également, par définition de l'écart de Procuste :

$$\sum_{h=1}^{l} d^2(P_h X_h, \overline{X}) \geqq \sum_{h=1}^{l} d_P^2(X_h, \overline{X}) = f(\overline{X}). \tag{13}$$

Des relations (11), (12), (13) on tire :

$$f(X) \geqq f(\overline{X}).$$

De plus, l'inégalité (12) étant stricte lorsque $X \neq \overline{X}$, on a dans ce cas : $f(X) > f(\overline{X})$.

Ainsi la fonction f ne peut atteindre son minimum que pour un nuage s'écrivant comme le barycentre des nuages $X_1, X_2 \ldots X_l$, après déplacement. Compte tenu de ce qui précède, la minimisation de $f(X)$ est donc équivalente au problème de Gower.

Remarque : A l'optimum, le nuage-moyen $\overline{X}$ est le meilleur représentant de la classe de nuages $(X_1, X_2 \ldots X_l)$ au sens de Procuste, c'est-à-dire du point de vue de leur forme. Ceci suggère une application intéressante en *reconnaissance de forme :* il s'agit de tester l'hypothèse que les nuages en question sont les observations éventuellement perturbées par un bruit de mesure, d'un même objet indéformable susceptible d'occuper une position différente dans l'espace lors de chaque observation effectuée. On a donc un problème de filtrage consistant à éliminer ces différences de position afin d'obtenir la meilleure image-moyenne de l'objet observé.

3. NOUVELLE APPROCHE DU PROBLÈME PROCUSTE GÉNÉRALISÉ A PLUS DE DEUX NUAGES

3.1. Résultats préliminaires

Nous venons d'établir que le problème de Gower revenait à chercher un nuage X minimisant la fonction :

$$f(X) = \sum_{h=1}^{l} d_P^2(X_h, X).$$

En utilisant la formule (5) du paragraphe 1 définissant l'écart de Procuste, il vient :

$$f(X)=\sum_{h=1}^{l}\left\{\operatorname{tr}(X_h X_h')+\operatorname{tr}(XX')-2\operatorname{tr}\left[(X_h X' X X_h')^{1/2}\right]\right\}. \tag{14}$$

Comme $\operatorname{tr}(XX')=\operatorname{tr}(X'X)$, la fonction à minimiser ne dépend en fait que de la matrice $F=X'X$ (symétrique et positive) des produits scalaires entre individus centrés du nuage X. F est en quelque sorte la « forme » du nuage X.

On constate sans difficulté que minimiser $f(X)$ équivaut alors à maximiser :

$$q(F)=\frac{1}{l}\sum_{h=1}^{l}\operatorname{tr}(X_h F X_h')^{1/2}-\frac{1}{2}\operatorname{tr}(F). \tag{15}$$

Dans la méthode Procuste habituelle, le nuage X doit être situé dans le même espace E que chacun des nuages $X_1, X_2 \ldots X_l$. Aussi la matrice $F=X'X$ ne peut avoir un rang supérieur à la dimension n de l'espace E en question. La légère modification que nous proposons consiste à supprimer la contrainte $\operatorname{rg}(F)\leqq n$, afin de rechercher l'optimum sur le cône-convexe $\mathscr{C}(m)$ des matrices semi-définies positives de dimension m.

Cette modification nous semble particulièrement justifiée quand les nuages examinés sont situés dans des espaces de dimensions différentes car on ne voit pas très bien dans ce cas selon quel critère choisir *a priori* la dimension du nuage-moyen X (*cf.* infra, § 4.3).

Nous donnons ci-après quatre propositions, démontrées en annexe, nécessaires à la poursuite de notre étude.

PROPOSITION 1 : *Les fonctions matricielles :*

$$\tau_h(F)=\operatorname{tr}\left[(X_h F X_h')^{1/2}\right]; \qquad h=1,\,2,\,\ldots,\,l,$$

sont concaves sur $\mathscr{C}(m)$.

COROLLAIRE : *La fonction-objectif* $q(F)$ *est concave sur* $\mathscr{C}(m)$.
En effet, d'après (15) on peut écrire :

$$q(F)=\frac{1}{l}\sum_{h=1}^{l}\tau_h(F)-\frac{1}{2}\operatorname{tr}(F). \tag{16}$$

La fonction $q(F)$ *est donc somme de* l *fonctions concaves et d'une fonction linéaire.*

PROPOSITION 2 : *On peut toujours limiter l'étude du maximum de* $q(F)$ *au cas où à l'optimum les matrices* $X_h F X_h'$ *sont nécessairement toutes régulières.*

PROPOSITION 3 : *Lorsque les matrices* $X_h F X'_h$ *sont régulières, les fonctions* τ_h *définies ci-dessus sont différentiables au point F et leur différentielle s'écrit :*

$$d\tau_h = \frac{1}{2} \operatorname{tr}(dF X'_h (X_h F X'_h)^{-1/2} X_h); \qquad h=1, 2, \ldots, l. \tag{17}$$

COROLLAIRE : *Il découle des résultats* (16) *et* (17) *que la fonction* $q(F)$ *est différentiable au point F et que sa différentielle s'écrit :*

$$dq = \frac{1}{l} \sum_{h=1}^{l} \operatorname{tr}(dF X'_h (X_h F X'_h)^{-1/2} X_h) - \frac{1}{2} tr(dF). \tag{18}$$

En posant :

$$M = \frac{1}{l} \sum_{h=1}^{l} X'_h (X_h F X'_h)^{-1/2} X_h, \tag{19}$$

la relation (18) *se note sous forme plus condensée :*

$$dq = \frac{1}{2} \operatorname{tr}[dF(M-I)]. \tag{20}$$

Dans la suite on prendra garde au fait que la matrice M dépend de F bien que pour simplifier l'écriture cette dépendance n'apparaisse pas dans la notation.

PROPOSITION 4 : *Lorsque les matrices* $X_h F X'_h$ *sont régulières, alors pour toute matrice* $G \in \mathscr{C}(m)$ *on a les inégalités :*

$$\left.\begin{aligned} &\operatorname{tr}[(X_h G X'_h)^{1/2}] \leqq \operatorname{tr}[(X_h F X'_h)^{1/2}] + \frac{1}{2} \operatorname{tr}[(\mathrm{G}-\mathrm{F}) X'_h (X_h F X'_h)^{-1/2} X_h], \\ &\qquad h=1, 2, \ldots, 1. \end{aligned}\right\} \tag{21}$$

COROLLAIRE : *En sommant les relations* (21) *pour tous les indices h et à l'aide des définitions* (15) *et* (19), *on établit alors la propriété :*

$$\forall G \in \mathscr{C}(m): \quad q(G) \leqq q(F) + \frac{1}{2} \operatorname{tr}[(\mathrm{G}-\mathrm{F})(\mathrm{M}-\mathrm{I})] \tag{22}$$

Nota : En se souvenant que q est concave et à l'aide de (20), on pourra vérifier qu'on retrouve dans (22) l'inégalité de concavité bien connue.

3.2. Caractérisation de l'optimum global

THÉORÈME : *Dans le cadre des hypothèses faites au paragraphe* 3.1, *pour que la matrice F soit un maximum global de la fonction* $q(F)$ *sur* $\mathscr{C}(m)$ *il faut et il suffit que soient vérifiées* simultanément *les deux conditions suivantes :*

1° *les valeurs propres de la matrice M* [cf. *définition* (19)] *sont au plus égales à* 1;

2° *On a l'égalité :* $F = FM$.

Démonstration : (*a*) Les conditions sont suffisantes.
Supposons que les deux conditions ci-dessus soient vérifiées.
Puisque :

$$F=FM \quad \Leftrightarrow \quad F(M-I)=0,$$

on a donc :

$$\operatorname{tr}[(G-F)(M-I)]=\operatorname{tr}[G(M-I)]. \tag{23}$$

Comme les valeurs propres de M sont inférieures ou égales à 1, la matrice $M-I$ est semi-définie négative.

La matrice G étant positive, puisque $G\in\mathscr{C}(m)$, on sait que les valeurs propres de la matrice $G(M-I)$ seront toutes réelles et négatives ou nulles. Il s'ensuit que :

$$\operatorname{tr}[G(M-I)]\leqq 0,$$

En appliquant ce résultat aux relations (23) puis (22), on en déduit :

$$\forall\, G\in\mathscr{C}(m), \quad q(G)\leqq q(F),$$

donc F est maximum global de q sur $\mathscr{C}(m)$.

(*b*) Les conditions sont nécessaires.

Supposons que F soit un maximum de la fonction q. D'après la proposition 2 du paragraphe 3.1, la différentielle de q existe et vérifie par définition [*cf.* (20)] :

$$\left| q(F+dF)-q(F)-\frac{1}{2}\operatorname{tr}[dF(M-I)]\right| = 0\ \| dF \|.$$

Autrement dit, pour toute variation infinitésimale *dF admissible*, i. e. telle que $F+dF\in\mathscr{C}(m)$, on a l'équivalent du 1[er] ordre :

$$q(F+dF)-q(F)\sim\frac{1}{2}\operatorname{tr}[dF(M-I)]. \tag{24}$$

Puisque q est maximum en F, on a nécessairement :

$$q(F+dF)-q(F)\leqq 0,$$

ce qui entraîne d'après (24) :

$$\operatorname{tr}[dF(M-I)]\leqq 0. \tag{25}$$

Nous allons maintenant examiner deux cas particuliers.

1. $dF = v\,\varepsilon\,v'$, où v est un vecteur normé quelconque de $\mathbb{R}^m$ et ε un réel positif arbitrairement petit.

On vérifie aisément que la matrice $F + dF$ est par construction symétrique et non négative, donc que dF est un accroissement admissible.

En faisant $dF = v\,\varepsilon\,v'$ dans la relation (25) on obtient :

$$\varepsilon\,v'(M-I)\,v \leqq 0,$$

inégalité qui ne peut être satisfaite que si :

$$v'(M-I)\,v \leqq 0,$$

soit encore : $v'\,M\,v \leqq v'\,v = 1$.

Puisque le vecteur v est arbitraire, les valeurs propres de M doivent donc être inférieures ou égales à 1.

2. $dF = -v_i\,\varepsilon_i\,v_i'$, où v_i est un vecteur propre normé de F associé à une valeur propre λ_i *strictement positive*, et ε_i un réel arbitrairement petit choisi dans $]0, \lambda_i[$. On sait que la matrice $F + dF$ est alors symétrique, qu'elle a les mêmes vecteurs propres que F et les mêmes valeurs propres sauf λ_i qui est déflatée de la quantité ε_i. Mais puisque $\varepsilon_i \in]0, \lambda_i[\Rightarrow \lambda_i - \varepsilon_i > 0$, la matrice $F + dF$ est non négative donc dF est un accroissement admissible.

En faisant comme ci-dessus $dF = -v_i\,\varepsilon_i\,v_i'$ dans la relation (25) on obtient :

$$-\varepsilon_i\,v_i'(M-I)\,v_i \leqq 0.$$

Cette inégalité entraîne aussitôt :

$$v_i'(M-I)\,v_i \geqq 0$$

et finalement :

$$v_i'\,M\,v_i \geqq v_i'\,v_i = 1.$$

Comme on a montré ci-dessus que M n'a aucune valeur propre supérieure à 1, on ne peut avoir l'inégalité stricte; il s'ensuit que : $v_i'\,M\,v_i = 1$.

De plus 1 étant valeur extrême du produit scalaire $\langle v, M\,v \rangle$ pour v appartenant à la sphère-unité de $\mathbb{R}^m$, cette valeur ne peut être atteinte que pour un vecteur propre de M associé à la valeur propre 1.

On a donc nécessairement les relations :

$$M\,v_i = v_i \qquad \text{et} \qquad v_i'\,M = v_i', \tag{26}$$

qui doivent être vérifiées par tout vecteur propre v_i de F associé à une valeur propre λ_i non nulle.

Écrivons maintenant la décomposition spectrale de F :

$$F = \sum_{i=1}^{m} v_i \lambda_i v_i',$$

On en déduit aussitôt :

$$FM = \sum_{i=1}^{m} v_i \lambda_i v_i' M. \tag{27}$$

Comme on vient de montrer que pour tout $i=1, 2, \ldots, m$:

$$\begin{cases} \text{ou bien } \lambda_i = 0, \\ \text{ou bien } v_i' M = v_i', \end{cases}$$

dans tous les cas la relation $\lambda_i v_i' M = \lambda_i v_i'$ est vérifiée. En reportant ce résultat dans (27) on montre ainsi qu'on a nécessairement l'égalité $F = FM$.

Nota : Dans certains cas il peut arriver qu'on ait à l'optimum :

$$rg(F) = rg(M).$$

Les valeurs propres de M sont alors toutes égales à 1 ou 0. M est donc une matrice de projecteur orthogonal. On montre qu'il s'agit plus précisément de la projection orthogonale (dans $\mathbb{R}^m$) sur le sous-espace vectoriel $\mathrm{Vec}(X_1, X_2 \ldots X_l)$ engendré par les l nuages examinés.

4. UN ALGORITHME DE RECHERCHE AUTOMATIQUE DE L'OPTIMUM

Nous proposons ici une version améliorée de l'algorithme décrit par Ten Berge [5]; elle permet notamment de choisir librement la dimension du nuage-moyen $\overline{X}$ (cf. paragraphe 3.1) et de contrôler qu'on a bien atteint l'optimum global.

4.1 Fonctionnement de l'algorithme

On a vu au paragraphe 1.2 que le problème Procuste généralisé consiste à chercher un système de déplacements $(P_1, P_2 \ldots P_l)$ qui rende maximale la quantité :

$$\theta(P_1, P_2 \ldots P_l) = \sum_h \sum_{<k} \mathrm{tr}\,(P_h X_h X_k' P_k'). \tag{28}$$

L'algorithme de Ten Berge effectue cette recherche par itérations successives. A chaque itération on déplace *successivement* tous les nuages de façon à obtenir l'augmentation maximale de la fonction-objectif.

Cette méthode repose sur la condition nécessaire d'optimum du paragraphe 1.2 : celle-ci montre, rappelons-le, qu'on peut améliorer la fonction-objectif tant que chaque nuage $P_h X_h$ n'est pas placé à distance minimale de la somme de tous les autres, i. e. du nuage : $C_h = \sum_{k \neq h} P_k X_k$.

On examine donc chaque distance $d^2(P_h X_h, C_h)$ pour $h=1$ à l. Si une distance n'est pas minimale, on recalcule le déplacement P_h de façon à ce qu'elle le devienne. Les résultats du paragraphe 1.1 assurent qu'on obtient ainsi la plus grande augmentation de la fonction-objectif par déplacement d'un seul nuage.

On répète ce processus tant qu'il est possible d'améliorer encore la fonction θ.

4.2. Convergence de l'algorithme

Considérons l'un des termes $\operatorname{tr}(P_h X_h X'_k P'_k)$ figurant dans la définition (28) de la fonction-objectif. On montre, par un même argument qu'au paragraphe 1.1, que ce terme est majoré par la quantité $\operatorname{tr}[(X_h X'_k X_k X'_h)^{1/2}]$.

On déduit alors de (28) l'inégalité :

$$\theta(P_1, P_2 \ldots P_l) \leqq \sum_h \sum_{<k} \operatorname{tr}[(X_h X'_k X_k X'_h)^{1/2}] = c.$$

Ainsi l'algorithme réalise à chaque itération un accroissement strict de la fonction θ qui est bornée supérieurement par la constante c. Cela suffit à assurer la convergence en pratique. Compte tenu des limites de la représentation des nombres en machine, on arrête l'algorithme dès que le taux d'accroissement de la fonction-objectif est devenu inférieur à un seuil de précision donné.

Théoriquement, l'algorithme peut converger vers une solution non optimale : en effet, le fait de ne plus pouvoir améliorer la fonction-objectif en déplaçant un seul nuage n'assure pas pour autant qu'il serait impossible de le faire en déplaçant deux ou plusieurs nuages à la fois.

En pratique, on évite cet inconvénient dans la quasi-totalité des cas en recourant à un procédé efficace d'amorçage de l'algorithme. Il est cependant prudent de contrôler qu'on a bien atteint l'optimum en vérifiant que les deux conditions du paragraphe 3.2 sont satisfaites.

On calcule donc le nuage-moyen $\overline{X} = 1/l \sum_{h=1}^{l} P_h X_h$; sa forme $F = \overline{X}' \overline{X}$; la matrice $M = 1/l \sum_{h=1}^{l} X'_h (X_h F X'_h)^{-1/2} X_h$.

On diagonalise M pour vérifier que ses valeurs propres sont inférieures ou égales à 1.

PROPOSITION : *Après arrêt de l'algorithme, la condition $F = FM$ est satisfaite* ipso facto.

Preuve : L'algorithme s'arrête quand chaque nuage est à distance minimale de la somme de tous les autres. On montre ([2], [5]) que chaque nuage $P_h X_h$ est *a fortiori* placé à distance minimale du nuage-moyen $\overline{X}$.

Le résultat (3) du paragraphe 1.1 permet alors d'écrire :

$$P_h = \overline{X} X_h' (X_h F X_h')^{-1/2}; \qquad h = 1, 2, \ldots, l. \tag{30}$$

Nota : On admettra que la matrice $X_h F X_h'$ est régulière (*cf.* la démonstration de la proposition 2 en annexe).

On tire alors des relations (30) :

$$\overline{X} = \frac{1}{l} \sum_{h=1}^{l} P_h X_h = \frac{1}{l} \sum_{h=1}^{l} \overline{X} X_h' (X_h F X_h')^{-1/2} X_h = \overline{X} M. \tag{31}$$

On en déduit aussitôt :

$$\overline{X}' \overline{X} = \overline{X}' \overline{X} M,$$

soit : $F = FM$.

4.3. Amorçage de l'algorithme

On doit d'abord fixer la dimension n de l'espace E dans lequel on cherchera le nuage-moyen $\overline{X}$. Pour que la méthode Procuste s'applique, il faut qu'on puisse considérer que chacun des nuages $X_1, X_2 \ldots X_l$ est situé dans un sous-espace vectoriel de ce même espace E, ce qui implique que sa dimension soit au moins égale à la plus grande des dimensions $n_1, n_2, \ldots, n_l$ des sous-espaces en question.

Posons : $n^* = \operatorname{Sup}_h n_h$. On doit donc imposer la condition : $n^* \leqq n$.

Par ailleurs, puisque le nuage-moyen s'écrit :

$$\overline{X} = \frac{1}{l} \sum_{h=1}^{l} P_h X_h,$$

il est certainement contenu dans un sous-espace vectoriel de l'espace $\mathrm{Vec}(X_1, X_2 \dots X_l)$ engendré par les l nuages examinés. On a donc la relation : $E \subset \mathrm{Vec}(X_1, X_2 \dots X_l)$. En notant p la dimension de $\mathrm{Vec}(X_1, X_2 \dots X_l)$, on déduit de ce qui précède la double inégalité :

$$n^* \leqq n \leqq p.$$

En fait, pour être sûr de pouvoir trouver le nuage-moyen $\overline{X}$, il suffirait de le chercher dans l'espace $\mathrm{Vec}\,(X_1, X_2 \dots X_l)$ tout entier, i. e. choisir $n=p$; au vu de nos expériences personnelles et pour des raisons d'économie de place-mémoire et de temps de calcul, il nous semble cependant plus efficace d'adopter la tactique inverse!

On choisit au départ une valeur de n pas trop élevée, voire la plus petite possible (i. e. $n=n^*$), quitte à revenir sur ce choix si l'algorithme n'a pas convergé vers l'optimum. Dans un tel cas, nous avons obtenu de bons résultats expérimentaux en incrémentant n du nombre s de valeurs propres de M structement supérieures à 1.

Une fois choisie la dimension n de l'espace ambiant E, il est important de bien initialiser les déplacements $P_1, P_2, \dots, P_l$, afin de minimiser le risque de non-convergence vers l'optimum.

Pour cela, on cherche dans E un premier nuage X suffisamment représentatif de la forme des nuages $X_1, X_2, \dots, X_l$; on détermine ensuite, pour $h=1$ à l, le déplacement P_h amenant le nuage $P_h X_h$ à distance minimale de X.

Construction de X

Soit $F_h = X'_h X_h$ la « forme » du nuage X_h; $h=1, 2, \dots, l$. On peut penser que $\overline{F} = 1/l \sum_{h=1}^{l} F_h$ offre déjà une bonne image de la forme des l nuages examinés.

Nota : On montre aisément que $\mathrm{rg}\,(\overline{F})=p$. On retient alors le nuage X dont la forme $X' X$ est la meilleure approximation d'Eckart et Young [4] de $\overline{F}$.

* * *

Après avoir calculé les déplacements $P_1, P_2 \dots P_l$, on exécute l'algorithme décrit au paragraphe 4.1. Il est fréquent qu'on obtienne alors 99 % du maximum de la fonction $\theta(P_1, P_2 \dots P_l)$ au cours des trois ou quatre premières itérations.

ANNEXE

Démonstration des propositions 1, 2, 3, 4 du paragraphe 3.1

PROPOSITION 1 : *Montrons que la fonction matricielle : $A \hookrightarrow \operatorname{tr}(A^{1/2})$ est concave.*

Soit A et B, deux matrices symétriques semi-définies positives et de même dimension n; soit r un réel choisi arbitrairement dans $[0, 1]$.

Posons : $C=(1-r)A+rB$.

Puisque C est symétrique, elle est diagonalisable et admet une base de vecteurs propres orthonormés $u_1, u_2, \ldots, u_n$.

Choisissons arbitrairement un indice $j \in \{1, 2, \ldots, n\}$. De l'inégalité de Schwarz :

$$\langle u_j, A^{1/2} u_j \rangle \leqq (\langle u_j, u_j \rangle . \langle A^{1/2} u_j, A^{1/2} u_j \rangle)^{1/2},$$

on tire immédiatement :

$$u_j' A^{1/2} u_j \leqq (u_j' A u_j)^{1/2} \qquad (a\,1)$$

et par un même argument :

$$u_j' B^{1/2} u_j \leqq (u_j' B u_j)^{1/2}. \qquad (a\,2)$$

En effectuant la combinaison convexe, membre à membre, des inégalités $(a\,1)$ et $(a\,2)$ on obtient :

$$(1-r)\, u_j' A^{1/2} u_j + r u_j' B^{1/2} u_j \leqq (1-r)(u_j' A u_j)^{1/2} + r (u_j' B u_j)^{1/2}. \qquad (a\,3)$$

Utilisant maintenant l'inégalité entre moments d'ordre 1 et 2 :

$$\forall x \geqq 0, \quad \forall y \geqq 0, \qquad (1-r)\,x + ry \leqq ((1-r)\,x^2 + ry^2)^{1/2},$$

avec $x=(u_j' A u_j)^{1/2}$ et $y=(u_j' B u_j)^{1/2}$, il vient :

$$\left.\begin{aligned}
(1-r)(u_j' A u_j)^{1/2} + r(u_j' B u_j)^{1/2} &\leqq ((1-r)\,u_j' A u_j + r u_j' B u_j)^{1/2},\\
(1-r)(u_j' A u_j)^{1/2} + r(u_j' B u_j)^{1/2} &\leqq (u_j'((1-r)\,A + r B)\,u_j)^{1/2},\\
(1-r)(u_j' A u_j)^{1/2} + r(u_j' B u_j)^{1/2} &\leqq (u_j' C u_j)^{1/2}.
\end{aligned}\right\} \qquad (a\,4)$$

Mais u_j est par hypothèse vecteur propre de C donc également de $C^{1/2}$; il s'ensuit qu'on a l'égalité :

$$(u_j' C u_j)^{1/2} = u_j' C^{1/2} u_j. \tag{a 5}$$

On déduit alors des relations $(a\,3)$, $(a\,4)$, $(a\,5)$:

$$(1-r) u_j' A^{1/2} u_j + r u_j' B^{1/2} u_j \leqq u_j' C^{1/2} u_j. \tag{a 6}$$

Utilisant un résultat connu du calcul matriciel, on obtient en sommant sur j les inégalités $(a\,6)$:

$$(1-r)\operatorname{tr}(A^{1/2}) + r\operatorname{tr}(B^{1/2}) \leqq \operatorname{tr}(C^{1/2}) \qquad \text{avec} \quad C=(1-r)A+rB \tag{a 7}$$

La fonction $A \mapsto \operatorname{tr}(A^{1/2})$ est donc concave.

Appliquons ce résultat en faisant maintenant :

$$A = X_h F X_h', \qquad B = X_h G X_h', \qquad C = X_h((1-r)F + rG)X_h'$$

avec :

$$F \in \mathscr{C}(m) \qquad \text{et} \qquad G \in \mathscr{C}(m).$$

Il suffit de revenir à la définition de la fonction τ_h pour constater que $(a\,7)$ entraîne alors :

$$(1-r)\tau_h(F) + r\tau_h(G) \leqq \tau_h((1-r)F + rG); \qquad \forall r \in [0, 1] \tag{a 8}$$

La fonction τ_h est donc concave sur $\mathscr{C}(m)$.

PROPOSITION 2 : *On a montré que la recherche du maximum de $q(F)$ équivalait au problème de Gower :*

$$\underset{(P_1, \ldots, P_l)}{\operatorname{Max}} \sum_h \sum_{<k} \operatorname{tr}(P_h X_h X_k' P_k'),$$

le lien entre les deux formulations étant que la matrice F optimale s'écrit nécessairement $F = \overline{X}' \overline{X}$, avec $\overline{X} = 1/l \sum_{h=1}^{l} P_h X_h$.

1. Nous allons utiliser ce résultat pour montrer qu'à l'optimum on a nécessairement :

$$\operatorname{rg}(X_h F X_h') \geqq \operatorname{rg}(X_h X_h'); \qquad h = 1, 2, \ldots, l.$$

La C.N. du paragraphe 1.2 et le résultat (4) du paragraphe 1.1 assurent qu'à l'optimum la matrice $W_h = P_h X_h (\sum_{k \neq h} P_k X_k)'$ est symétrique et semi-définie positive.

Revenant à la définition de $\overline{X}$, on vérifie aussitôt que :

$$P_h X_h \overline{X}' = \frac{1}{l}(P_h X_h X_h' P_h' + W_h),$$

$P_h X_h \overline{X}'$ est donc somme de deux matrices semi-définies positives; on en déduit immédiatement :

$$\operatorname{rg}(P_h X_h \overline{X}') \geqq \operatorname{rg}(P_h X_h X_h' P_h'). \qquad (a\,9)$$

Mais à cause de la contrainte d'orthogonalité $P_h' P_h = I$, on a par ailleurs :

$$\left.\begin{array}{r}\operatorname{rg}(P_h X_h X_h' P_h') = \operatorname{rg}(X_h X_h'),\\ \operatorname{rg}(P_h X_h \overline{X}') = \operatorname{rg}(P_h X_h (\overline{X}'\,\overline{X})\,X_h'\,P_h') = \operatorname{rg}(X_h\,F X_h').\end{array}\right\} \qquad (a\,10)$$

De $(a\,9)$ et $(a\,10)$ on tire alors :

$$\operatorname{rg}(X_h F X_h') \geqq \operatorname{rg}(X_h X_h').$$

Ainsi, lorsque la matrice $X_h X_h'$ est régulière, on est certain qu'à l'optimum la matrice $X_h F X_h'$ le sera également.

2. On peut toujours se ramener, sans perte de généralité, au cas où la matrice $X_h X_h'$ est régulière.

En examinant la relation (16) on voit déjà que le nuage X_h n'intervient dans la fonction $q(F)$ que par l'intermédiaire du terme :

$$\tau_h(F) = \operatorname{tr}[(X_h F X_h')^{1/2}].$$

Mais puisque les matrices $X_h F X_h'$ et $F X_h' X_h$ ont mêmes valeurs propres, on a :

$$\tau_h(F) = \operatorname{tr}[(X_h F X_h')^{1/2}] = \operatorname{tr}[(F X_h' X_h)^{1/2}]. \qquad (a\,11)$$

Ainsi, la fonction $q(F)$ ne dépend finalement que de la forme $F_h = X_h' X_h$ du nuage X_h.

Soit r_h le rang de la matrice F_h.

En effectuant la factorisation canonique de F_h :

$$F_h = \xi_h' \xi_h,$$

on obtient une matrice ξ_h (à r_h lignes et m colonnes) qui par construction est de rang r_h; ainsi la matrice $\xi_h \xi_h'$ est *régulière*.

Compte tenu de ce qui précède, on peut affirmer que les nuages X_h et ξ_h sont rigoureusement équivalents, du point de vue de la méthode Procuste; dans la suite on pourra donc substituer le nuage ξ_h au nuage X_h chaque fois que cela est nécessaire.

Cette précaution étant prise et afin de ne pas alourdir l'exposé, nous ne distinguerons plus désormais entre les notations X_h et ξ_h, ce qui nous permettra d'écrire, d'après le (1) de cette démonstration, que les matrices $X_h F X'_h$ sont toutes régulières à l'optimum.

PROPOSITION 3 : *De l'identité* $A^{1/2}.A^{1/2}=A$, *on tire par différentiation :*

$$d(A^{1/2})\,A^{1/2}+A^{1/2}\,d(A^{1/2})=dA. \qquad (a\,12)$$

Lorsque A *est régulière on peut encore écrire :*

$$d(A^{1/2})+A^{1/2}\,d(A^{1/2})\,A^{-1/2}=dA\,A^{-1/2} \qquad (a\,13)$$

et en prenant la trace des expressions figurant dans $(a\,13)$ *:*

$$2\,\mathrm{tr}\,[d\,(A^{1/2})]=\mathrm{tr}\,(dA\,A^{-1/2}). \qquad (a\,13)$$

L'opérateur Trace étant linéaire, $(a\,14)$ *est équivalent à :*

$$d[\mathrm{tr}\,(A^{1/2})]=\frac{1}{2}\,\mathrm{tr}\,(dA\,A^{-1/2}). \qquad (a\,15)$$

Appliquons ce résultat en faisant :

$$A=X_h F X'_h, \qquad dA=X_h dF X'_h, \qquad A^{-1/2}=(X_h F X'_h)^{-1/2}$$

On obtient :

$$d\tau_h=d(\mathrm{tr}\,[(X_h F X'_h)^{1/2}])=\frac{1}{2}\,\mathrm{tr}\,(X_h\,dF X'_h\,(X_h F X'_h)^{-1/2})$$

$$=\frac{1}{2}\,\mathrm{tr}\,(dF X'_h\,(X_h F X'_h)^{-1/2}\,X_h)$$

ce qui établit la proposition 3.

PROPOSITION 4. — *Soit* A *une matrice symétrique et définie positive. Pour toute matrice* B *symétrique, non négative, et de même dimension que* A *on a alors :*

$$\mathrm{tr}\,[(B^{1/2}-A^{1/2})\,A^{-1/2}\,(B^{1/2}-A^{1/2})] \geqq 0. \qquad (a\,16)$$

Il suffit en effet de poser $T=T'=B^{1/2}-A^{1/2}$ et de remarquer que toute matrice de la forme $T'A^{-1/2}T$ est non négative quand A est positive.

En développant (a 16) on obtient après réarrangement des termes :

$$\operatorname{tr}(BA^{-1/2})+\operatorname{tr}(A^{1/2}) \geqq 2\operatorname{tr}(B^{1/2}). \qquad (a\,17)$$

En utilisant l'identité $BA^{-1/2}=(B-A)A^{-1/2}+A^{1/2}$ dans la relation (a 17) et après division par 2, il vient :

$$\frac{1}{2}\operatorname{tr}((B-A)A^{-1/2})+\operatorname{tr}(A^{1/2}) \geqq \operatorname{tr}(B^{1/2}). \qquad (a\,18)$$

Faisons alors :

$$A=X_h F X_h'; \qquad B=X_h G X_h'; \qquad B-A=X_h(G-F)X_h',$$

(a 18) se récrit :

$$\operatorname{tr}[(X_h G X_h')^{1/2}] \leqq \operatorname{tr}[(X_h F X_h')^{1/2}] + \frac{1}{2}\operatorname{tr}[(G-F)X_h'(X_h F X_h')^{-1/2}X_h]$$

ce qui démontre la proposition 4.

BIBLIOGRAPHIE

1. BOURGEOIS, *Recherche du déplacement minimisant la distance entre deux ensembles de points homologues*, Cahiers de l'Analyse des Données, vol. 3, n° 4, 1978, p. 440-448.
2. BOURGEOIS, *Thèse 3e cycle*, Statistique Mathématique, Université de Paris-6, 1980.
3. GOWER, *Generalized Procrustes Analysis*, Psychometrika, vol. 40, n° 1, 1975, p. 33-51.
4. JOHNSON, *On a Theorem by Eckart and Young, Psychometrika, vol. 28, n°* 2, 1963, p. 259-263.
5. TEN BERGE, *Orthogonal Procrustes Rotation for Two or More Matrices*, Psychometrika, vol. 42, n° 2, 1977, p. 267-276.