

REVUE DE STATISTIQUE APPLIQUÉE

STÉPHANE GIRARD

BERNARD CHALMOND

JEAN-MARC DINTEN

Une ACP non linéaire basée sur l'approximation par variété

Revue de statistique appliquée, tome 46, n° 3 (1998), p. 5-19

http://www.numdam.org/item?id=RSA_1998__46_3_5_0

© Société française de statistique, 1998, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

UNE ACP NON LINÉAIRE BASÉE SUR L'APPROXIMATION PAR VARIÉTÉ

Stéphane Girard*, Bernard Chalmond**, Jean-Marc Dinten***

* INRIA Rhône-Alpes, ZIRST 655 avenue de l'Europe, 38330 Montbonnot Saint Martin

** Université de Cergy-Pontoise et ENS Cachan (DIAM-CMLA) 2, avenue A. Chauvin,
95302 Cergy Cedex

*** LETI (CEA – Technologies Avancées) DSYS-CEN/G 17 avenue des Martyrs,
38054 Grenoble Cedex 9

RÉSUMÉ

Cet article concerne la généralisation non linéaire de l'Analyse en Composantes Principales (ACP). Il est connu que l'ACP construit, sous forme d'un espace vectoriel, le modèle optimal d'un jeu de données gaussien. Lorsque les données n'ont pas une distribution gaussienne, l'ACP ne construit pas le modèle optimal du point de vue des résidus d'approximation. Nous proposons une méthode d'approximation par variété généralisant en de nombreux points les propriétés de l'ACP mais permettant de dépasser ses performances d'approximation. Notre modèle bénéficie d'une mise en œuvre très simple, basée sur une construction incrémentale. A chaque itération de sa construction, la dimension de la variété est incrémentée et les résidus d'approximation diminués. Nous montrons une application à la définition d'un modèle de l'arcade dentaire humaine.

Mots-clés : analyse en composantes principales, modèles auto-associatifs non linéaires, variétés, splines.

ABSTRACT

This article deals with the non linear generalization of the Principal Component Analysis (PCA). It is well-known that PCA builds, thanks to a linear subspace, the optimal model of Gaussian data. When the set of points is not Gaussian, PCA does not build the optimal model from the approximation residuals point of view. We propose a manifold approximation method both generalizing in many points the PCA properties and improving its approximation performances. This model benefits of a very simple implementation based on an incremental scheme. At each iteration of the model design, the manifold dimension is incremented and residual errors decreased. An application to the modelling of human dental arches is presented.

Keywords : principal component analysis, non-linear auto-associative models, manifolds, splines.

Nous nous intéressons au problème de la généralisation de l'ACP du point de vue de l'approximation. En d'autres termes, nous étudions les méthodes de construction de modèles qui, étant donné un nuage d'un grand nombre de points dans un espace n -dimensionnel, cherchent à le représenter de manière concise. Ces modèles sont appelés auto-associatifs (AA) et ils sont définis par une équation implicite. L'ACP fait partie de ces méthodes, elle construit un modèle auto-associatif linéaire. Cet article est consacré plus particulièrement aux modèles auto-associatifs définissant une variété (AAV). Du point de vue géométrique, il s'agit de construire une variété approchant au mieux le nuage de points de départ. En ce sens, ces modèles généralisent l'ACP qui réalise une approximation par un sous-espace vectoriel.

La première partie de cet article est consacrée à la présentation de l'ACP et à la mise en évidence de ses limites. Nous poursuivons en présentant une généralisation non linéaire de celle-ci permettant de dépasser ces limitations. Dans une troisième partie, nous détaillons la mise en œuvre de notre méthode pour l'appliquer dans une dernière partie à la modélisation de l'arcade dentaire humaine. Nous concluons en présentant nos directions de recherche actuelles.

1. Position du problème

Notons $\{x^j\}_{j=1\dots N}$ le nuage de N points de \mathbb{R}^n à modéliser. La $i^{\text{ème}}$ ($1 \leq i \leq n$) composante du point x^j ($1 \leq j \leq N$) est notée x_i^j . Dans un premier temps, nous rappelons le principe de la modélisation par ACP, puis nous illustrons ses limites sur un exemple simple avant de présenter quelques propositions récentes pour dépasser ces limites.

1.1. L'Analyse en Composantes Principales

Par souci de simplicité, nous supposons dorénavant que les données sont centrées. L'ACP construit alors un modèle linéaire de l'ensemble de points en l'approchant par des sous-espaces vectoriels d'équation :

$$x - \sum_{k=1}^d \langle x, a^k \rangle a^k = 0$$

d désigne la dimension du sous-espace vectoriel, et l'ensemble d'axes $\{a_k\}_{k=1\dots d}$ une base orthonormée de ce sous-espace. Les axes sont choisis de façon à minimiser la distance quadratique moyenne entre les données et le sous-espace vectoriel. Par conséquent, l'ACP construit le meilleur modèle linéaire dans le sens où il minimise la norme \mathcal{L}_2 du modèle aux observations. De plus, on peut montrer que dans le cas où les données sont distribuées suivant une loi gaussienne, cette propriété reste vraie parmi l'ensemble des modèles de type auto-associatif [8]. Ainsi, il est clair que les limitations de l'ACP apparaîtront dès lors que les données ne seront pas issues d'une loi gaussienne. Prenons par exemple le cas du nuage de points de \mathbb{R}^3 présenté figure 1. Bien que cet ensemble de point soit de dimension intrinsèque 1, l'ACP demande de construire un modèle de dimension 3 afin d'obtenir une fraction d'information représentée supérieure à 75 %.

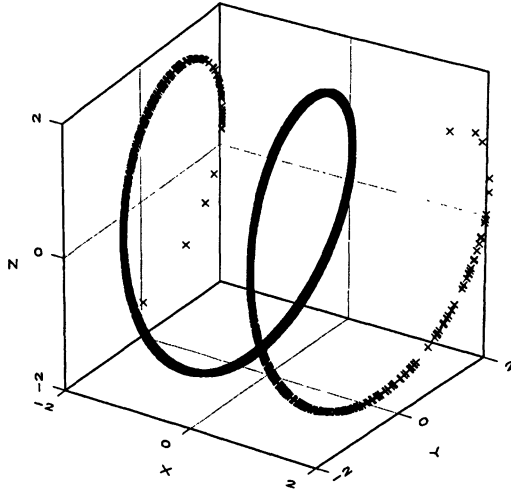


FIGURE 1
Nuage «hélice».

1.2. Quelques généralisations proposées

Les «Courbes Principales» [12] sont sans doute la généralisation la plus intuitive de l'ACP. L'idée est de substituer aux axes a^k de l'ACP des courbes. Cependant, il n'est pas possible d'obtenir une expression analytique de l'opérateur de projection sur les courbes, et le modèle obtenu est donc non paramétrique. De plus, cette approche est limitée à la construction de courbes, variétés de dimension 1. Citons également les approches de type réseaux de neurones, et plus précisément les perceptrons en mode auto-associatif [14]. L'équation du modèle est obtenue en introduisant un opérateur non linéaire σ (appelé fonction d'activation) dans le modèle de l'ACP :

$$x - \sum_{k=1}^d \sigma^k (\langle x, a^k \rangle) = 0$$

Il a été constaté [3] que les modèles ainsi construits ne permettent pas d'obtenir des résidus inférieurs à ceux de l'ACP en raison des difficultés de minimisation de la distance modèle-observation.

Une approche alternative est l'ACP curvilinéaire qui consiste à rechercher des transformations non linéaires des variables [5] de manière à maximiser la fraction d'information représentée par l'ACP sur les données ainsi prétraitées.

Nous allons quant à nous présenter un principe de construction de modèles auto-associatifs, généralisant les modèles de type ACP ou réseaux de neurones. Ces modèles sont définis de façon à, d'une part conserver les propriétés théoriques de l'ACP, et d'autre part permettre d'obtenir aisément de meilleures approximations dans un contexte non linéaire.

2. Modèle non linéaire

Nous présentons tout d'abord les principes de base qui nous ont conduit dans la construction d'un modèle non linéaire. Ensuite, nous illustrons ces principes dans des cas simples puis nous donnons les propriétés importantes que vérifie notre modèle.

2.1. Principe

Nous avons vu que l'ACP réalise une approximation du nuage de points par un sous-espace vectoriel. Afin d'obtenir une plus grande classe de modèles, nous cherchons pour notre part à construire des approximations par des variétés.

Définition 1 *Un modèle auto-associatif défini par une variété (AAV) est représenté par l'équation implicite suivante :*

$$G(\beta, x) = 0 \quad (1)$$

- β est un vecteur de paramètres de \mathbb{R}^p , p quelconque ;
- $G(\beta, \cdot)$, à β fixé est une fonction de \mathbb{R}^n dans \mathbb{R}^n , supposée continûment dérivable, et de telle sorte que (1) soit l'équation d'une variété ;
- On appelle alors dimension du modèle, la dimension de la variété.

Nous focalisons notre attention sur les modèles dits «composés». Nous verrons par la suite que ces modèles peuvent être construits itérativement. A chaque itération, la dimension du modèle est incrémentée de manière à augmenter la précision avec laquelle on représente les données.

Définition 2 *Le modèle est dit composé (AAVC) si de plus G s'écrit¹ :*

$$G(\beta, x) \stackrel{\text{def}}{=} \left(\prod_{k=d}^1 G^k(\beta^k, \cdot) \right) (x)$$

Avec $\forall k \in \{1, \dots, d\}$:

- $G^k(\beta^k, \cdot)$ $k^{\text{ème}}$ composante du modèle, fonction de \mathbb{R}^n dans \mathbb{R}^n ;
- β^k $k^{\text{ème}}$ vecteur de paramètres, de dimension p_k quelconque.

Cette décomposition demande cependant encore de construire des fonctions à n variables. Or ceci pose problème lorsque la dimension n de l'espace est importante, car apparaît le «fléau de la dimension» ou «curse of dimensionality». Huber [13] a en effet mis en évidence sur un exemple très simple que dans ce cas, même pour des jeux de données de taille N très grande, l'espace reste relativement vide.

L'idée est alors de construire des modèles compressant les données pour s'affranchir autant que possible du fléau de la dimension. Pour cela, nous introduisons un opérateur de projection choisi linéaire de façon à éviter les limitations des «courbes principales» déjà évoquées.

¹ Le signe \prod désigne ici la composition des opérateurs.

Définition 3 Nous appelons projection sur l'axe a , normé, l'opérateur suivant :

$$P^a : x \in \mathbb{R}^n \rightarrow \langle a, x \rangle \in \mathbb{R}.$$

La seconde brique de construction des modèles auto-associatifs est l'opérateur «dual» qui permet d'associer à une valeur dans \mathbb{R} un point de l'espace \mathbb{R}^n .

Définition 4 Nous appelons restauration l'opérateur :

$$S^\alpha : t \in \mathbb{R} \rightarrow S^\alpha(t) \in \mathbb{R}^n$$

- S^α est supposé continûment dérivable ;
- α est un ensemble de paramètres.

Le choix du type d'opérateur S^α , appelé aussi fonction de lissage en raison de sa régularité, définit en grande partie le type de modèle construit. L'utilisation combinée des deux opérateurs S^α et P^a par composition $S^\alpha P^a$ permet d'associer à tout point de \mathbb{R}^n un autre point «lissé» de \mathbb{R}^n , mais en ayant dans l'intervalle compressé l'information. Cette technique est aussi utilisée dans la construction des réseaux de neurones où l'on parle de «bottleneck». Plus généralement, les modèles auto-associatifs sont basés sur cette approche. Nous montrons maintenant comment construire la fonction G à partir de $S^\alpha P^a$, le gain apporté par cette approche étant de ramener la construction de G , fonction de n variables, à celui de S^α , fonction d'une seule variable.

Définition 5 Composantes du modèle AAVC.

- Les composantes du modèle sont définies par $G^k(\beta^k, \cdot) \stackrel{\text{def}}{=} Id_{\mathbb{R}^n} - S^{\alpha^k} P^{a^k}$, $\forall k \in \{1 \dots d\}$, l'ensemble de paramètres étant $\beta^k = (\alpha^k, a^k)$.
- On note $r^j(d)$, le résidu d'approximation de l'observation x^j par le modèle, $r^j(d) = G(\beta, x^j)$.

$$\text{Le résidu total est donné par } R_d = \sum_{j=1}^N ||r^j(d)||^2.$$

Ainsi chaque composante du modèle évalue la différence entre un point et sa projection-restauration. Afin d'avoir un point de vue encore plus intuitif sur ces modèles, nous montrons dans le paragraphe suivant comment les interpréter dans les cas $d = 1$ puis $d = 2$.

2.2. Illustrations

2.2.1. Le cas $d = 1$

Le modèle s'écrit dans ce cas :

$$x - S^{\alpha^1} P^{a^1} x = 0$$

Sous réserve d'imposer une condition naturelle, nous exhibons deux propriétés essentielles du modèle.

Lemme 1 *Si on impose que la restauration-projection soit l'identité :*

$$P^{a^1} S^{\alpha^1} = Id_{\mathbb{R}} \quad (2)$$

alors on a les deux propriétés suivantes :

1. Le modèle représente une variété de dimension 1.
2. Le résidu d'approximation de chaque point est orthogonal à l'axe choisi :

$$\langle a^1, r^j(1) \rangle = 0 \quad \forall j \in \{1, \dots, N\} \quad (3)$$

Preuve : La démonstration du premier résultat est technique et peut être trouvée dans [8].

Le second point est vérifié en formant le produit scalaire

$$\langle a^1, r^j(1) \rangle = \langle a^1, x^j - S^{\alpha^1} P^{a^1} x^j \rangle = \left(Id_{\mathbb{R}} - P^{a^1} S^{\alpha^1} \right) \left(P^{a^1} x^j \right) = 0$$

en utilisant la propriété (2).

Le premier point établit la nature du modèle à une composante : il représente une courbe d'un certain type (sans point de rebroussement, et qui ne s'auto-intersecte pas). La condition (2) signifie que nous cherchons à construire une fonction de restauration qui soit l'inverse à droite de la projection : si l'on restaure un réel puis que l'on le reprojete, on désire retrouver le même point de \mathbb{R} . La question naturelle est alors de savoir si l'on peut satisfaire la condition inverse, à savoir $S^{\alpha^1} P^{a^1}(x^j) = x^j$, $\forall j \in \{1, \dots, N\}$, de façon à obtenir une restauration parfaite. La réponse est malheureusement négative, l'opérateur de projection étant, dans la majorité des cas, non injectif sur le jeu de données. Nous verrons cependant lors de la mise en œuvre de la méthode que l'on peut construire un axe a^1 favorisant une restauration satisfaisante : $S^{\alpha^1} P^{a^1}(x^j) \simeq x^j$, $\forall j \in \{1, \dots, N\}$. Le résultat $\langle a^1, r^j(1) \rangle = 0$, $\forall j \in \{1, \dots, N\}$ découlant de la condition (2) est très important car il va permettre d'itérer simplement la construction du modèle.

2.2.2. Le cas $d = 2$

Le modèle avec $d = 2$ est de la forme :

$$G(\beta, x) \stackrel{\text{def}}{=} \left(Id_{\mathbb{R}^n} - S^{\alpha^2} P^{a^2} \right) \left(Id_{\mathbb{R}^n} - S^{\alpha^1} P^{a^1} \right) (x)$$

Le principe est simple, les points x^j sont projetés et restaurés une première fois puis l'écart de restauration est calculé par $r^j(1) = (Id_{\mathbb{R}^n} - S^{\alpha^1} P^{a^1})(x^j)$ et le même travail est effectué sur ces résidus. De plus, nous savons grâce à (3) que

ceux-ci sont dans l'hyperplan de normale a^1 . On peut donc choisir a^2 et S^{α^2} dans ce même hyperplan. Ce principe sera à la base de toutes les propriétés du modèle (cf. paragraphe 2.4).

Finalement, on a $G(\beta, x^j) = (Id_{\mathbb{R}^n} - S^{\alpha^2} P^{a^2})(r^j(1))$, ce qui se note également $r^j(2)$. L'algorithme de construction itérative d'un modèle composé obéit à cette même logique. Ainsi, les données sont modélisées suivant deux directions orthogonales a^1 et a^2 . Le problème du fléau de la dimension est contourné en ne considérant que quelques directions intéressantes.

On montre alors que sous certaines hypothèses la dimension du modèle obtenu est le nombre d d'axes choisis (cf. paragraphe 2.4). Le problème principal est le choix des axes, problème pouvant être résolu par des techniques dites de Poursuite de Projection (PP). Une approche classique en analyse des données est de choisir des axes suivant lesquels la distribution est la plus éloignée d'une loi gaussienne [6] de façon à montrer à l'utilisateur la spécificité du nuage de points. Dans le cas général, on maximise une fonctionnelle $I(X, a)$ appelée index. A titre d'exemple, l'index dans le cas de l'ACP est la variance projetée : $I(X, a) = \text{var}(\langle X, a \rangle)$. Un index adapté à la spécificité de notre problème est proposé dans le paragraphe 3.

Ce type d'approche est utilisé dans les problèmes d'approximation [11] où les fonctions S^{α^k} sont fixées par le choix de la fonction de régularisation, ou dans les algorithmes de type PPR (Poursuite de Projection en Régression) [7] dans le cas de la régression, le choix de ces fonctions étant alors laissé à l'utilisateur. Dans notre cas, le type des fonctions de lissage est également un des degrés de liberté du modèle. Une famille de fonctions splines est proposé dans le paragraphe 3.

2.3. Résumé

En résumé, le modèle AAVC s'écrit :

$$\left(\prod_{k=d}^1 (Id_{\mathbb{R}^n} - S^{\alpha^k} P^{a^k}) \right) (x) = 0$$

Avec les conditions suivantes :

$$\begin{aligned} \forall k \in \{1 \dots d\} P^{a^k} S^{\alpha^k} &= Id_{\mathbb{R}} \\ \forall k, l \in \{1 \dots d\} \langle a^k, a^l \rangle &= \delta_{kl} \end{aligned} \quad (4)$$

Il est construit selon le principe présenté dans le paragraphe précédent, itéré d fois.

Les étapes (5) et (6) de l'algorithme ci-dessous seront précisées dans le paragraphe 3. On y proposera notamment un famille de fonctions de lissage $\{S^{\alpha}\}_{\alpha}$, un index de choix d'axe de projection I , ainsi que des techniques pour les optimiser. Auparavant, nous montrons dans le paragraphe suivant quelques propriétés théoriques de ce type de modèle, indépendantes de ces choix.

Algorithme de construction d'un modèle AAVC	
1. Initialisation :	
$d \leftarrow 0$	
$r^j(0) \leftarrow x^j \quad j = 1 \dots N$	
2. Calcul des nouvelles $d + 1^{\text{ème}}$ caractéristiques :	
– Calcul de l'axe a^{d+1} .	$a^{d+1} = \arg \max_a I(r(d), a) \quad (5)$
– Calcul des paramètres α^{d+1} .	$\alpha^{d+1} = \arg \min_{\alpha} \sum_{j=1}^N \left\ r^j(d) - S^{\alpha} P^{a^{d+1}} r^j(d) \right\ ^2 \quad (6)$
3. Mise à jour des résidus.	
$r^j(d+1) \leftarrow r^j(d) - S^{\alpha^{d+1}} P^{a^{d+1}} r^j(d) \quad j = 1 \dots N$	
$d \leftarrow d + 1$	
4. Retour en 2. si les résidus sont trop importants.	

2.4. Propriétés

Nous énonçons deux types de résultats, les propriétés d'approximation du modèle, puis nous donnons des éléments de comparaison du modèle AAVC avec l'ACP.

Proposition 1 *Les modèles AAVC ont les propriétés suivantes :*

1. *Le modèle avec d composantes représente une variété de dimension d .*
2. *La suite des résidus moyens est décroissante.*
3. *Pour $d = n$, le modèle est exact.*

Ainsi, l'algorithme de construction de modèle réalise des approximations de plus en plus précises par des variétés de dimensions croissantes. Nous avons déjà vu avec le lemme 1 que pour une dimension $d = 1$, le modèle représente un certain type de courbe, pour $d = 2$, il s'agit de surfaces. Dans le cas où d est quelconque, on parle de variété, partie de \mathbb{R}^n localement semblable à une partie ouverte de \mathbb{R}^d [1]. La preuve de cette propriété repose essentiellement sur la condition (2) et est établie dans [8]. On met à profit ces propriétés pour définir la fraction d'information.

Définition 6 *On définit Q_d la fraction d'information représentée par le modèle à d axes :*

$$Q_d = 1 - \frac{\sum_{j=1}^N \|r^j(d)\|^2}{\sum_{j=1}^N \|x^j\|^2}$$

De façon évidente, on a $0 \leq Q_d \leq 1$ et on déduit des résultats précédents que la suite $(Q_d)_d$ est croissante et que $Q_n = 1$ (figure 2). Cette grandeur permet de choisir la dimension du modèle en fonction de la fraction de l'information que l'on désire représenter. Remarquons que ces propriétés sont les généralisations directes de celles de l'ACP. La comparaison de ces deux types de modèle est l'objet de la proposition suivante.

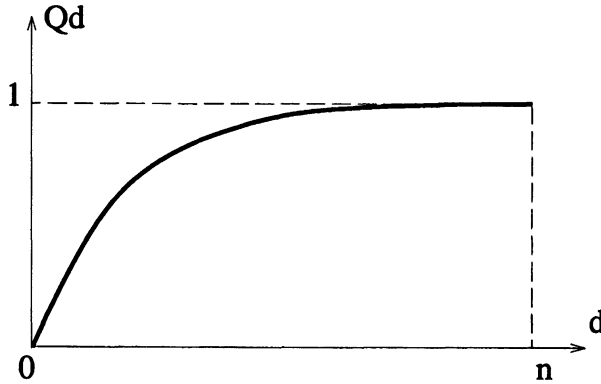


FIGURE 2
Comportement de la fraction d'information.

Proposition 2 *Le choix de fonctions de restauration $S^{a^k}(t) = ta^k$, conduit au modèle de l'ACP.*

$$G(a, x) = x - \sum_{k=1}^d \langle a^k, x \rangle a^k$$

Il apparaît donc que l'ACP est un cas particulier des modèles AAVC où l'on choisit comme fonction de restauration la rétroprojection. On peut aller plus loin dans la comparaison entre les deux méthodes en montrant que le seul modèle AAVC additif est le modèle de l'ACP [8].

Nous abordons les problèmes pratiques de mise en œuvre de la méthode dans le paragraphe suivant.

3. Mise en œuvre

Afin d'implémenter l'algorithme de construction de modèle, nous devons en préciser sa seconde étape et plus particulièrement les optimisations (5) et (6). Pour cela nous allons décrire le choix des fonctions de restauration ainsi que celui de l'index de projection.

Le type de fonction de restauration détermine en grande partie la nature du modèle. Par exemple, nous avons vu avec la proposition 2 que choisir des fonctions de restauration linéaires conduit à une ACP. Nous proposons de choisir des fonctions de lissage splines qui sont performantes dans les problèmes de régression [4].

Pour cela, on se donne T_M une subdivision d'un intervalle $[a, b]$. $T_M = \{a = t_0, t_1, \dots, t_M, t_{M+1} = b\}$. Un point $t_i, 0 \leq i \leq M + 1$ est appelé un nœud, et un point $t_i, 1 \leq i \leq M$ est appelé un nœud intérieur (M représente donc le nombre de nœuds intérieurs). On choisit alors comme fonctions de lissage :

$$S_i^{\alpha_i^k} \in s_3(T_M) \quad 1 \leq i \leq n, \quad 1 \leq k \leq d$$

Avec $s_3(T_M)$ l'ensemble des splines cubiques construites sur la subdivision T_M à M nœuds intérieurs

$$s_3(T_M) = \{s : [a, b] \rightarrow \mathbb{R}, s \text{ polynôme de degré 3 sur } [t_i, t_{i+1}] \forall i \in \{0, M\}, s \in C^2([a, b])\}$$

Le choix du nombre de nœuds M est alors crucial, il détermine la dimension de l'espace vectoriel $s_3(T_M)$ et donc la complexité du modèle. Nous le déterminons par validation croisée [17] de façon à obtenir un bon compromis entre approximation et généralisation, l'ensemble des paramètres α est alors obtenu par (6) qui se résume dans ce cas à l'inversion d'un système linéaire. Cette estimation peut aussi être interprétée comme la construction d'une fonction de restauration S^α vérifiant $S^\alpha P^a(x^j) \simeq x^j, \forall j \in \{1, \dots, N\}$. En comparant cette condition avec (2) $P^a S^\alpha = Id_{\mathbb{R}}$, il apparaît que les performances d'approximation du modèle seront très liées aux choix de l'axe de projection. Nous avons déjà mentionné que le meilleur cas $S^\alpha P^a(x^j) = x^j, \forall j \in \{1, \dots, N\}$, n'apparaît que rarement, puisqu'il correspond au cas où l'opérateur de projection est injectif sur le jeu de données (figure 3b). Le rôle de l'index de projection est alors de favoriser des fonctions de projections pour lesquelles cette condition est «presque» vérifiée. Il est possible de quantifier ceci en comptant le nombre de points du jeu d'exemples qui sont plus proches voisins dans \mathbb{R}^n et qui ne le sont plus après projection. La figure 3 illustre ce principe.

Dans le premier cas (figure 3a), deux points voisins dans \mathbb{R} , ne le sont pas dans \mathbb{R}^n , et la projection conduit à des superpositions. Intuitivement, l'index de projection peut se décrire ainsi :

$$I(x, a) = \sum_{i=1}^N \sum_{j \neq i} \Phi [(x^j \text{ plus proche voisin de } x^i)] \Rightarrow (P^a(x^j) \text{ plus proche voisin de } P^a(x^i))].$$

Φ représente la fonction indicatrice : $\Phi(P) = 1$ si P est vraie, 0 sinon. L'expression analytique exacte de l'index ne présente pas d'intérêt particulier. Notons cependant que cette fonction n'est pas continue par rapport à a , ce qui interdit toute maximisation basée sur des techniques de gradient. Nous avons mis en œuvre un algorithme de recuit simulé [2] pour obtenir son maximum global et ainsi réaliser l'opération (5) de l'algorithme. L'ensemble de ce schéma est testé dans le paragraphe suivant sur des données réelles, pour une validation sur des données simulées, on pourra se référer à [10].

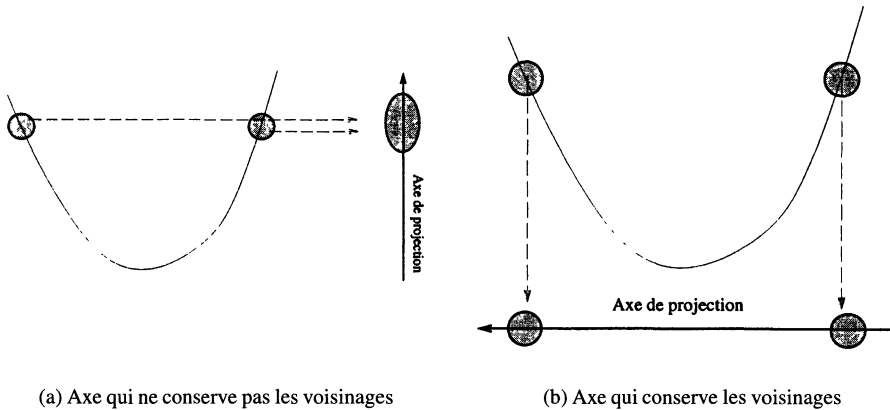


FIGURE 3
Choix de l'axe de projection.

4. Exemple d'application

Afin d'illustrer cette étude, nous reprenons le problème de la modélisation de l'arcade dentaire humaine déjà abordé dans [16]. Pour les raisons évoquées par les auteurs, nous focalisons notre attention sur la modélisation de l'arcade mandibulaire ou inférieure. Chaque dent étant repérée par une ou plusieurs points, l'arcade mandibulaire est alors représentée par un ensemble de 22 points du plan, après suppression des points correspondant aux dents de sagesse. Nous disposons d'un jeu de 71 données (figure 4). A titre d'exemple, la figure 6 montre l'ensemble des points caractérisant l'arcade de l'individu 13. Notons alors $\{(\eta_i^j, \xi_i^j) \mid 1 \leq i \leq 22\}$ l'ensemble des couples de coordonnées des points formant l'arcade mandibulaire du $j^{\text{ème}}$ individu. On obtient alors un ensemble de $N = 71$ observations de dimension $n = 44$ en posant :

$$\begin{cases} x_{2i-1}^j = \eta_i^j \\ x_{2i}^j = \xi_i^j \end{cases} \text{ pour } i = 1 \dots 22 \text{ et } j = 1 \dots 71.$$

L'ACP permet alors de construire un modèle linéaire du nuage obtenu et, de ce fait, un modèle linéaire paramétré de l'arcade mandibulaire humaine. Notons que l'utilisation de l'ACP pour représenter des déformations de courbes est due à Rice et Silverman [15]. La figure 5 montre la décroissance du résidu d'approximation en fonction de la dimension du modèle choisi. Pour obtenir une fraction d'information de 95 %, il apparaît qu'un modèle de dimension $d = 9$ est nécessaire.

Nous avons également utilisé notre procédure de construction de modèle non linéaire. Par souci de rapidité, nous avons utilisé les mêmes axes de projection que l'ACP, les axes obtenus par maximisation de I n'apportant pas, sur cet exemple, un gain significatif. Les fonctions de restauration sont des splines cubiques construites sur une subdivision de $M = 40$ nœuds. De même que pour l'ACP, nous avons représenté

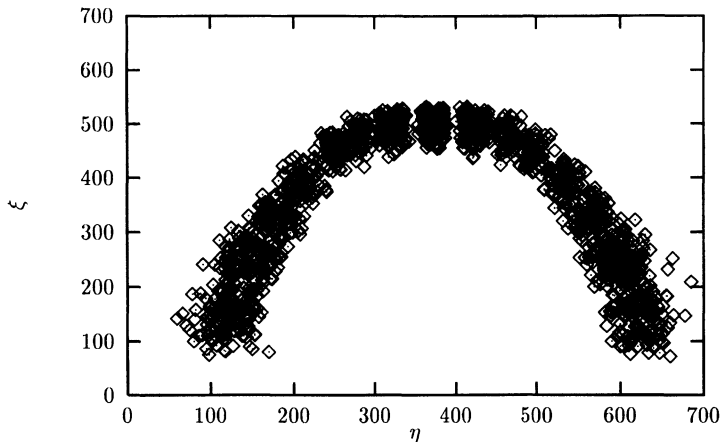


FIGURE 4
Superposition des 71 arcades de 22 points.

les résidus d'approximation à diverses étapes de la construction du modèle. Il apparaît que ceux-ci sont toujours très inférieurs, pour des modèles de même dimension à ceux de l'ACP. Par exemple, pour avoir une fraction d'information de 95 %, un modèle de dimension 2 est suffisant alors que l'ACP demandait un modèle de dimension 9. En fait, un modèle AAVC de dimension 4 permet une représentation quasiment exacte de l'arcade inférieure. Les 4 paramètres de ce modèle non linéaire fournissent alors autant d'indices de morphologie dento-faciale. Remarquons que le modèle associé permet de représenter aussi bien des arcades «saines» (figure 6) que pathologiques (figure 7), alors que les limites du modèle linéaire de l'ACP sont surtout marquées sur ces dernières. Cette nette distinction entre les deux méthodes est due à l'utilisation de fonctions de restauration non linéaires par le modèle AAVC. Les fonctions splines de $s_3(T_{40})$ peuvent beaucoup mieux s'adapter aux données que les droites utilisées par l'ACP.

Le lecteur intéressé pourra également consulter [9] pour une application des modèles AAVC à la construction de modèles de coupes de fémur.

5. Conclusion

Nous avons proposé une généralisation non linéaire de l'ACP. Le point de départ de cette étude a été de remarquer que l'ACP construit un modèle linéaire en approchant les données par un sous-espace vectoriel. Afin d'obtenir de meilleures approximations, il est alors naturel de considérer des modèles de variétés. Nous avons proposé des modèles de type composé offrant le double avantage d'une construction simple, basée sur un algorithme itératif d'approximations successives, et de propriétés théoriques intéressantes. Ces modèles ont montré une grande efficacité sur des problèmes concrets. Nos recherches actuelles concernent la définition et la

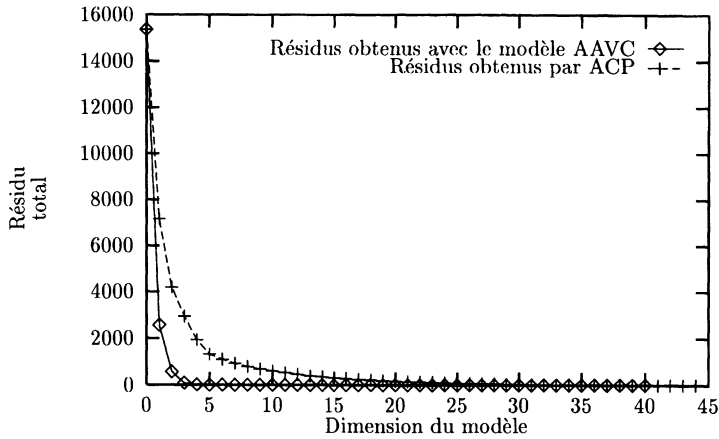


FIGURE 5

Evolution du résidu total en fonction de la dimension du modèle.

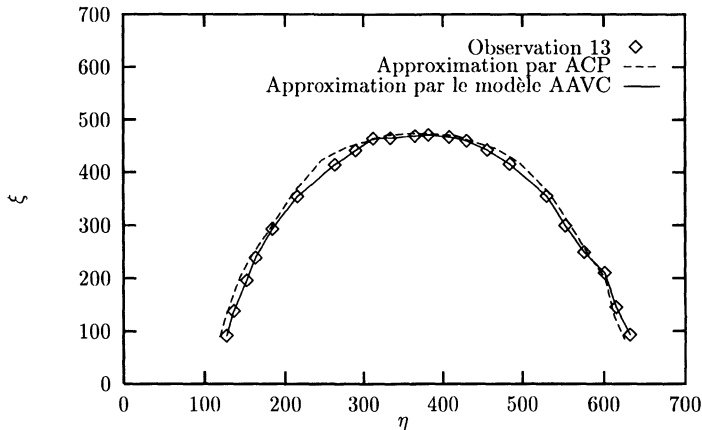


FIGURE 6

Représentation d'une arcade «saine» et de ses différentes approximations.

comparaison de nouveaux index de projection. En effet, l'index utilisé pour l'ACP (la variance projetée) est aisé à maximiser mais n'est pas le plus adapté à notre modèle. A l'opposé, l'index *ad hoc* que nous proposons est pénalisé par une maximisation délicate à mettre en œuvre. Il serait alors intéressant d'avoir un index offrant un meilleur compromis efficacité-simplicité.

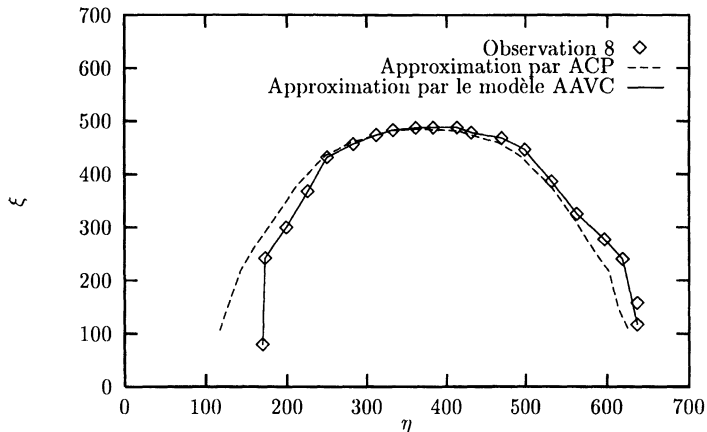


FIGURE 7

Représentation d'une arcade pathologique et de ses différentes approximations.

Références

- [1] AVEZ A. (1983). *Calcul différentiel*, Masson, Paris.
- [2] AZENCOTT R. (1988). Simulated annealing. *Seminaire BOURBAKI*, 40^{ème} année, 697, 1-15, juin.
- [3] COTTRELL M. (1994). Analyse de données et réseaux de neurones. Préprint SAMOS-Université Paris I, février.
- [4] EUBANK R.L. (1990). *Spline smoothing and non-parametric regression*. Marcel Decker.
- [5] FERRATY F. (1997). Estimations de transformations optimales en ACP curvilinéaire. *Revue de Statistique Appliquée*, XLV (1), 5-39.
- [6] FRIEDMAN J.H. (1987). Exploratory projection pursuit. *Journal of the American Statistical Association*, 82 (397), 249-266.
- [7] FRIEDMAN J.H., STUETZLE W. (1981). Projection pursuit regression. *Journal of the American Statistical Association*, 76 (376), 817-823.
- [8] GIRARD S. (1996). *Construction et apprentissage statistique de modèles auto-associatifs non-linéaires*. PhD thesis, Université de Cergy-Pontoise, octobre.
- [9] GIRARD S., CHALMOND B., DINTEN J.M. (1997). Apprentissage de modèles flexibles pour la reconnaissance des formes. In *Seizième colloque GRETSI*, pages 1295-1298, Grenoble, septembre.
- [10] GIRARD S., CHAMOND B., DINTEN J.M. (1997). Designing non linear models for flexible curves. *Curves and Surfaces in Geometric Design*, A. Le Méhauté, C. Rabut, and L.L. Schumaker (eds.), 135-142.

- [11] GIROSI F., JONES M., POGGIO T. (1995). Regularization theory and neural networks architectures. *Neural Computation*, pages 219-269, july.
- [12] HASTIE T., STUETZLE W. (1989). Principal curves. *Journal of the American Statistical Association*, 84 (406), 502-516.
- [13] HUBER P.J. (1985). Projection pursuit. *The Annals of Statistics*, 13 (2), 435-525.
- [14] KARHUNEN J., JOUTSENSALO J. (1995). Generalizations of principal component analysis, optimization problems, and neural networks. *Neural Networks*, 8 (4), 549-562.
- [15] RICE J.A., SILVERMAN B.W. (1991). Silverman. Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal Royal Statistical Society B*, 53 (1), 233-243.
- [16] ROTENBERG M., DEJEAN P., DUCROS D., FALGUEROLLES A. (1995). Quelques modèles pour l'arcade dentaire humaine. *La Revue de Modulad*, 16, 19-40, décembre.
- [17] STONE M. (1974). Cross-validated choice and assessment of statistical prediction. *Journal Royal Statistical Society B*, 36, 111-147.