

UNE EXTENSION DE L'ACP : LES MODÈLES AUTO-ASSOCIATIFS

STÉPHANE GIRARD & SERGE IOVLEFF

RÉSUMÉ. Dans cet article, nous proposons les modèles Auto-Associatifs (AA) comme candidats à la généralisation de l'Analyse en Composantes Principales (ACP). Les modèles AA ont été introduits en analyse des données du point de vue géométrique. Ils reposent sur l'approximation du nuage des observations par une variété différentiable. Nous proposons ici une interprétation en termes de modèles de Poursuite de Projection en Régression adaptés au cas auto-associatif. Nous établissons leurs propriétés théoriques et montrons comment elles étendent celles de l'ACP. Nous proposons également un algorithme de construction itératif dont nous illustrons le fonctionnement sur données simulées d'une part et données réelles issues de l'analyse d'images d'autre part.

1. INTRODUCTION AUX MODÈLES AUTO-ASSOCIATIFS.

L'Analyse en Composantes Principales (ACP) [22] est une méthode couramment utilisée pour la réduction de dimension en analyse des données. Elle possède une interprétation géométrique simple. Etant donné un ensemble de points de \mathbb{R}^p , et pour un entier $0 \leq d \leq p$, l'ACP construit le sous-espace affine de dimension d approchant au mieux les points au sens de la distance euclidienne [28]. Partant de ce point de vue, de nombreux auteurs ont proposé des extensions non-linéaires de cette méthode. Les approches de type courbes ou surfaces principales [19, 7] font partie de cette famille de méthodes. L'ACP peut également être interprétée en termes de Poursuite de Projection [21, 23]. Elle recherche le sous-espace affine de dimension d maximisant la variance projetée [20]. Ainsi, l'introduction de critères autres que la variance permet de définir autant de méthodes d'exploration des données [12, 27]. Dans les approches de type ACPVI-Spline (Analyse en Composantes Principales de Variables Instrumentales [9]) et ACP curvilinéaire [1], l'introduction de transformations non-linéaires des coordonnées permet de conserver un critère de variance projetée sur les données transformées. Enfin, il est également possible d'associer un modèle probabiliste gaussien à l'ACP [29], le sous-espace affine étant alors obtenu par maximisation d'une vraisemblance. Cette approche permet d'obtenir d'autres méthodes de réduction

de dimension en considérant des modèles non-gaussiens, par exemple des modèles de mélange.

L'extension de l'ACP au cas non-linéaire est donc un problème difficile si l'on ne souhaite pas "perdre" ses trois interprétations (géométrique, en terme de variance projetée, probabiliste). Ainsi, la construction d'un modèle probabiliste satisfaisant est souvent impossible sans spécifier la loi des observations. La méthode ainsi construite est alors "ad-hoc" et de peu d'utilité en pratique. De plus, l'introduction d'une non-linéarité peut faire perdre l'interprétation géométrique du modèle construit. Les notions de variables principales, de directions principales, d'inertie expliquée et résiduelle se généralisent alors malaisément. Ajoutons, que la non-linéarité du modèle peut également mener à des problèmes de mise en œuvre. Peuvent se poser les problèmes de l'existence, unicité et calculabilité d'un estimateur du modèle.

Dans cet article, nous proposons les Auto-Associatifs (AA) comme candidats à la généralisation de l'ACP. Les modèles AA ont été introduits dans [15] du point de vue géométrique. Ils reposent sur l'approximation du nuage des observations par une variété différentiable. Nous montrons ici que ces modèles peuvent également être interprétés comme des modèles de Poursuite de Projection en Régression [11, 24] adaptés au cas auto-associatif. De ce fait, nous proposons un algorithme de mise en œuvre aisée, directement inspiré de celle de l'ACP itérative. Nous proposons également un cadre probabiliste permettant d'établir de nombreuses propriétés théoriques.

Pour cela, considérons tout d'abord l'ACP du point de vue de la Poursuite de Projection. Si X est un vecteur aléatoire de \mathbb{R}^p possédant un moment du second ordre, il est toujours possible de le décomposer en une somme de d variables aléatoires orthogonales (variables principales) et d'un résidu en appliquant de manière itérative les étapes [A] (recherche d'Axes), [P] (Projection), [R] (Régression) et [M] (Mise à jour) suivantes (pour la démonstration voir paragraphe 3.1) :

Algorithme 1.1.

- Pour $j = 0$, on pose $R^0 = X - \mathbb{E}[X]$.

- Pour $j = 1, \dots, d$:

$$[A] \text{ Déterminer } a^j = \arg \max_{x \in \mathbb{R}^p} \mathbb{E} \left[\langle x, R^{j-1} \rangle^2 \right]$$

$$s.c. \|x\| = 1 \text{ et } \langle x, a^k \rangle = 0, 1 \leq k < j.$$

$$[P] \text{ Calculer } Y^j = \langle a^j, R^{j-1} \rangle.$$

$$[R] \text{ Déterminer } b^j = \arg \min_{x \in \mathbb{R}^p} \mathbb{E} \left[\|R^{j-1} - Y^j x\|^2 \right] \text{ s.c. } \langle x, a^j \rangle = 1,$$

$$(\text{on trouve } b^j = a^j) \text{ puis poser } s^j(Y^j) = Y^j b^j.$$

$$[M] \text{ Calculer } R^j = R^{j-1} - s^j(Y^j).$$

Nous appelons les vecteurs a^j les directions révélatrices, les variables aléatoires Y^j les variables principales, les fonctions s^j les fonctions de régression, et les vecteurs aléatoires R^j les résidus. L'étape [A] consiste à rechercher un axe privilégié perpendiculaire aux précédents, qui maximise un certain critère : ici la variance projetée. L'étape [P] consiste à projeter les résidus sur l'axe trouvé pour déterminer les variables principales, l'étape [R] consiste à chercher la meilleure fonction linéaire des variables principales qui approche les résidus. L'étape [M] consiste à mettre à jour les résidus.

Considérons les étapes [A] et [R] en détail. La propriété cruciale de l'étape [A] consiste à déterminer une nouvelle direction révélatrice dans l'espace des résidus obtenus à l'itération précédente. Le fait que l'axe retenu soit obtenu en maximisant la variance projetée est (selon notre opinion) un choix arbitraire lorsque X n'est pas gaussien. L'étape [R] détermine la fonction linéaire de la variable principale qui approche au mieux les résidus au sens des moindres carrés. De même, le fait de se restreindre à une classe de fonctions linéaires peut être un choix trop restrictif dès que X n'est plus gaussien.

Les modèles AA étendent l'algorithme précédent en considérant des étapes [A] et [R] plus générales. Pour l'étape [A], nous considérons un certain nombre d'autres critères issus de la Poursuite de Projection correspondant à des objectifs différents. Nous envisageons l'étape [R] comme un problème de régression pouvant être abordé par des outils de type splines ou estimateurs à noyaux. Nous montrons que ce type de généralisation de l'ACP permet de conserver ses principales propriétés théoriques (construction d'un modèle exact, décroissance des résidus, ...) ou de les étendre (approximation des réalisations de X non plus par un sous-espace affine mais par une variété différentiable).

Finalement il est important de noter que si la matrice de variance-covariance du vecteur aléatoire X est de rang d alors l'algorithme 1.1 donne une décomposition exacte de X . Dans le cas des modèles AA, cette propriété se généralise ainsi : si la loi du vecteur aléatoire X est concentrée sur une variété de dimension d paramétrable par les d axes a^1, \dots, a^d alors les modèles AA fournissent une décomposition exacte de X .

Ce document est organisé de la manière suivante. Paragraphe 2 nous définissons les modèles auto-associatifs et énonçons leurs propriétés générales. Nous présentons au paragraphe 3 deux principes de construction de modèles AA et nous étudions leur particularités. Paragraphe 4, nous présentons différents critères provenant de la poursuite de projection dans le cadre des modèles AA, ainsi que différentes méthodes pour estimer les fonctions de régression. Finalement, nous montrons au paragraphe 5 quelques illustrations sur données simulées ainsi qu'une application à l'analyse d'images.

2. LES MODÈLES AUTO-ASSOCIATIFS.

Dans un premier temps, nous donnons la définition des modèles auto-associatifs ainsi que de quelques notions connexes. Dans un second temps, nous proposons un algorithme de construction dont nous prouvons les propriétés théoriques.

2.1. Définitions.

Définition 2.1. Une application F de $\mathbb{R}^p \rightarrow \mathbb{R}^p$ est dite auto-associative de dimension d , s'il existe d vecteurs orthonormés a^j et d fonctions s^j de $\mathbb{R} \rightarrow \mathbb{R}^p$ vérifiant $P_{a^j} \circ s^j = \text{Id}_{\mathbb{R}^p}$ et $P_{a^k} \circ s^j = 0$, $1 \leq k < j \leq d$, où $P_{a^j}(x) = \langle a^j, x \rangle$, tels que

$$F = \left(\text{Id}_{\mathbb{R}^p} - s^d \circ P_{a^d} \right) \circ \dots \circ \left(\text{Id}_{\mathbb{R}^p} - s^1 \circ P_{a^1} \right) = \prod_{k=d}^1 \left(\text{Id}_{\mathbb{R}^p} - s^k P_{a^k} \right).$$

Les vecteurs a^j sont appelés les directions révélatrices, les fonctions s^j sont appelées les fonctions de régression et on écrira $F \in \mathcal{A}_{a,s}^d$.

Par la suite, par souci de concision, le produit représentera la composition. Il apparaît qu'un modèle auto-associatif de dimension d est déterminé par la donnée d'un ensemble de d fonctions de régression et de d directions révélatrices. La démonstration du lemme suivant peut être trouvée dans [13].

Lemme 2.1. Soit $F \in \mathcal{A}_{a,s}^d$, et supposons que les s^j , $j = 1, \dots, d$ soient $C^1(\mathbb{R}, \mathbb{R}^p)$. L'équation $F(x) = 0$ définit alors une sous-variété différentiable de dimension d .

Considérons un vecteur aléatoire X de \mathbb{R}^p de carré intégrable défini sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. Nous notons \mathbb{P}_X sa loi de probabilité sur \mathbb{R}^p , et $L^2_X(\mathbb{R}, \mathbb{R}^p)$ l'espace des fonctions de \mathbb{R} dans \mathbb{R}^p telles que sP_a soit de carré \mathbb{P}_X -intégrable pour tout $a \in \mathbb{R}^p$.

Définition 2.2. On dit que X vérifie un modèle auto-associatif de dimension d de directions révélatrices (a^1, \dots, a^d) , de fonctions de régression (s^1, \dots, s^d) et de résidu ε , si X vérifie $F(X - \mu) = \varepsilon$ où $F \in \mathcal{A}_{a,s}^d$, $\mu \in \mathbb{R}$ et où ε est un vecteur aléatoire centré.

En outre, on dit que X vérifie un modèle AA linéaire si les fonctions de régressions sont linéaires.

Donnons deux exemples de modèles auto-associatifs triviaux :

- (1) Tout X satisfait un modèle AA de dimension 0. Il suffit de choisir $F = \text{Id}$, $\mu = \mathbb{E}[X]$ et $\varepsilon = X - \mathbb{E}[X]$. On a alors $\text{Var} \left[\|\varepsilon\|^2 \right] = \text{Var} \left[\|X\|^2 \right]$.
- (2) X satisfait toujours également un modèle AA de dimension p . Dans ce cas $F = 0$, $\mu = 0$ et $\varepsilon = 0$ conduisent à $\text{Var} \left[\|\varepsilon\|^2 \right] = 0$.

Dans la pratique, il s'agit de réaliser un compromis entre ces deux extrêmes en construisant un modèle de dimension $d \ll p$ et tel que $\text{Var} \left[\|\varepsilon\|^2 \right] \ll \text{Var} \left[\|X\|^2 \right]$. Par exemple, dans le cas où X est centré et possède une matrice de variance-covariance Σ de rang d , il satisfait un modèle AA linéaire de dimension d de résidu nul. En effet, notons a^j , $j = 1, \dots, d$ les vecteurs propres de Σ associés aux valeurs propres non nulles. Nous montrons au Corollaire 3.1 que

$$F(x) = \prod_{k=d}^1 \left(\text{Id}_{\mathbb{R}^p} - P_{a^k} a^k \right) (x) = x - \sum_{k=1}^d \langle a^k, x \rangle a^k$$

et que $\varepsilon = 0$ \mathbb{P} .p.s. définissent un modèle auto-associatif linéaire pour X . Il s'agit en fait de la décomposition de X obtenue par ACP.

Notre but dans cet article est de fournir un algorithme efficace pour construire des modèles auto-associatifs non nécessairement linéaires de dimension réduite avec une variance résiduelle faible. Dans cette optique, nous introduisons deux définitions :

Définition 2.3. *On dit qu'un ensemble $\mathcal{S}(\mathbb{R}, \mathbb{R}^p)$ de fonctions mesurables de \mathbb{R} dans \mathbb{R}^p est admissible s'il est un sous-ensemble fermé de $L^2_X(\mathbb{R}, \mathbb{R}^p)$ et s'il vérifie la condition suivante :*

$$(\mathcal{R}) : \begin{cases} \forall b \in \mathbb{R}^p & s \in \mathcal{S}(\mathbb{R}, \mathbb{R}^p) \Rightarrow s + b \in \mathcal{S}(\mathbb{R}, \mathbb{R}^p) \\ & \text{Id}_{\mathbb{R}} b \in \mathcal{S}(\mathbb{R}, \mathbb{R}^p). \end{cases}$$

La condition (\mathcal{R}) s'interprète comme une invariance par translation. Un choix possible de $\mathcal{S}(\mathbb{R}, \mathbb{R}^p)$ est l'ensemble des fonctions affines de \mathbb{R} dans \mathbb{R}^p . Cet exemple est traité en détails dans le paragraphe 3.1.

Définition 2.4. *Soit a un vecteur unitaire de \mathbb{R}^p . Un index I est une fonctionnelle qui associe à la projection du vecteur aléatoire X sur a (i.e. $\langle a, X \rangle$) un réel positif.*

Un choix de I possible est $I(\langle a, X \rangle) = \text{Var}[\langle a, X \rangle]$, la variance projetée. D'autres exemples sont présentés au paragraphe 4.2.

2.2. Construction des modèles auto-associatifs. Soient $\mathcal{S}(\mathbb{R}, \mathbb{R}^p)$ un ensemble de fonctions admissibles et $d \in \{0, \dots, p\}$. On introduit l'algorithme suivant :

Algorithme 2.1.

- Pour $j = 0$, on pose $\mu = \mathbb{E}[X]$ et $R^0 = X - \mu$.
- Pour $j = 1, \dots, d$:
 - [A] Déterminer $a^j = \arg \max_{x \in \mathbb{R}^p} I(\langle x, R^{j-1} \rangle)$
s.c. $\|x\| = 1, \langle a^k, x \rangle = 0, 1 \leq k < j$.
 - [P] Calculer $Y^j = \langle a^j, R^{j-1} \rangle$.
 - [R] Choisir $s^j \in \arg \min_{s \in \mathcal{S}(\mathbb{R}, \mathbb{R}^p)} \mathbb{E} \left[\|R^{j-1} - s(Y^j)\|^2 \right]$ *s.c.* $P_{a^j} s^j = \text{Id}$.
 - [M] Calculer $R^j = R^{j-1} - s^j(Y^j)$.

Nous vérifions théorème 2.1 que cet algorithme construit un modèle auto-associatif de dimension d . Nous montrons de plus qu'à l'issue de p itérations il construit une représentation exacte de X .

Il est clair que l'étape [R] dépend fortement du choix de $\mathcal{S}(\mathbb{R}, \mathbb{R}^p)$. L'existence d'une solution au problème de minimisation est acquise grâce aux conditions imposées à $\mathcal{S}(\mathbb{R}, \mathbb{R}^p)$. En particulier, la condition (\mathcal{R}) nous assure qu'il existe des fonctions dans $\mathcal{S}(\mathbb{R}, \mathbb{R}^p)$ qui vérifient la contrainte du problème de minimisation. L'unicité de la solution n'est pas assurée sans une condition supplémentaire de convexité. Dans cet article nous étudions deux cas extrêmes : le cas où $\mathcal{S}(\mathbb{R}, \mathbb{R}^p) = \mathcal{A}(\mathbb{R}, \mathbb{R}^p)$ l'ensemble des fonctions affines de \mathbb{R} dans \mathbb{R}^p , dans le paragraphe 3.1 et le cas où $\mathcal{S}(\mathbb{R}, \mathbb{R}^p) = L^2_X(\mathbb{R}, \mathbb{R}^p)$, dans le paragraphe 3.2. Le choix de l'index I est discuté paragraphe 4.2.

Théorème 2.1. *L'algorithme 2.1 construit un modèle auto-associatif de dimension d de directions révélatrices (a^1, \dots, a^d) , de fonctions de régression (s^1, \dots, s^d) et de résidu $\varepsilon = R^d$. De plus, si $d = p$ alors $\varepsilon = R^p = 0$ et on a la décomposition exacte :*

$$X = \mathbb{E}[X] + \sum_{k=1}^p s^k(Y^k).$$

Notons que ces propriétés sont générales dans le sens où elles ne dépendent ni de l'index I choisi, ni de l'ensemble de fonctions admissible $\mathcal{S}(\mathbb{R}, \mathbb{R}^p)$. Nous prouverons dans le paragraphe 3 quelques propriétés complémentaires correspondant à des choix de I et $\mathcal{S}(\mathbb{R}, \mathbb{R}^p)$ particuliers.

Nous allons tout d'abord prouver la proposition suivante :

Proposition 2.1. *Les résidus et les fonctions de régression obtenus avec l'algorithme 2.1 vérifient les propriétés suivantes :*

- (i) Pour tout $1 \leq j \leq d$, $\mathbb{E}[R^j] = \mathbb{E}[Y^j] = \mathbb{E}[s^j(Y^j)] = 0$.

- (ii) Pour tous $1 \leq k \leq j \leq d$, $\langle a^k, R^j \rangle = 0$ \mathbb{P} -p.s.
- (iii) Pour tous $1 \leq k < j \leq d$, $\langle a^k, s^j(Y^j) \rangle = 0$, \mathbb{P} -p.s.
- (iv) La suite des normes des résidus est \mathbb{P} -p.s. décroissante.

Preuve :

- (i) La preuve est faite par récurrence sur j . Notons HR_j l'hypothèse $\mathbb{E}[R^j] = 0$. Il est clair que HR_0 est vraie. Supposons HR_{j-1} vérifiée. On a alors,

$$\mathbb{E}[R^j] = \mathbb{E}[R^{j-1}] - \mathbb{E}[s^j(Y^j)] = -\mathbb{E}[s^j(Y^j)].$$

Or, s^j est solution de l'étape [R] et donc $\mathbb{E}[s^j(Y^j)] = 0$. Cette dernière égalité se prouve par l'absurde. En effet, si $\mathbb{E}[s^j(Y^j)] \neq 0$, on introduit $\mu = \mathbb{E}[s^j(Y^j)]$, et $s'^j = s^j - \mu$. Comme $\langle a^j, s^j \rangle = \text{Id}$ et $\mathbb{E}[Y^j] = \mathbb{E}[\langle a^j, R^{j-1} \rangle] = 0$ par HR_{j-1} , on a $\langle a, \mu \rangle = 0$ et on a encore $\langle a^j, s'^j \rangle = \text{Id}$. De plus, d'après la condition (\mathcal{R}) , on a $s'^j \in \mathcal{S}(\mathbb{R}, \mathbb{R}^p)$, et donc

$$\mathbb{E}[\|R^{j-1} - s'^j(Y^j)\|^2] < \mathbb{E}[\|R^{j-1} - s^j(Y^j)\|^2],$$

puisque R^{j-1} est centré, ce qui contredit l'hypothèse de minimalité de s^j . En conclusion, $\mathbb{E}[R^j] = -\mathbb{E}[s^j(Y^j)] = 0$.

- (ii) et (iii) La preuve se fait également par récurrence sur j . Notons HR_j l'hypothèse $\forall k \leq j$, $\langle a^k, R^j \rangle = 0$. HR_1 est vraie car

$$\langle a^1, R^1 \rangle = \langle a^1, R^0 \rangle - \langle a^1, s^1(Y^1) \rangle = Y^1 - Y^1 = 0.$$

Supposons HR_{j-1} vérifiée et prouvons HR_j . Pour $k = j$, on a

$$\langle a^j, R^j \rangle = \langle a^j, R^{j-1} \rangle - \langle a^j, s^j(Y^j) \rangle = Y^j - Y^j = 0.$$

Pour $k < j$, on a par HR_{j-1} :

$$\langle a^k, R^j \rangle = \langle a^k, R^{j-1} \rangle - \langle a^k, s^j(Y^j) \rangle = \langle a^k, s^j(Y^j) \rangle.$$

Or, s^j est solution de l'étape [R] et donc $\langle a^k, s^j(Y^j) \rangle = 0$ ce qui prouve (iii). En effet,

$$\|R^{j-1} - s^j(Y^j)\|^2 = \langle a^k, R^{j-1} - s^j(Y^j) \rangle^2 + \sum_{i \neq k} \langle a^i, R^{j-1} - s^j(Y^j) \rangle^2.$$

D'après HR_{j-1} , $\langle a^k, R^{j-1} \rangle = 0$ et la condition (\mathcal{R}) assurant la stabilité de $\mathcal{S}(\mathbb{R}, \mathbb{R}^p)$ par projection, le minimum est atteint pour une fonction s^j telle que $\langle a^k, s^j \rangle = 0$. La preuve se fait par l'absurde de la même manière qu'en (i). En conclusion, $\langle a^k, R^j \rangle = 0$ et $\langle a^k, s^j \rangle = 0$, ce qui montre à la fois HR_j et (iii).

(iv) Soit $j \geq 1$ et s'^j la fonction de $\mathcal{S}(\mathbb{R}, \mathbb{R}^p)$ obtenue par projection orthogonale de s^j sur a^j . On a

$$\begin{aligned} \|R^j\|^2 &= \|R^{j-1} - s^j(Y^j)\|^2 \\ &\leq \|R^{j-1} - s'^j(Y^j)\|^2 \\ &= \sum_{k=1}^{j-1} \langle a^k, R^{j-1} - s'^j(Y^j) \rangle^2 + \langle a^j, R^{j-1} - s'^j(Y^j) \rangle^2 \\ &\quad + \sum_{k=j+1}^p \langle a^k, R^{j-1} - s'^j(Y^j) \rangle^2. \end{aligned}$$

Le premier terme est nul d'après (ii) et (iii). Le second terme est nul lui aussi d'après la condition $\langle a^j, s^j \rangle = \text{Id}$. Finalement, d'après la définition de s'^j :

$$\|R^j\|^2 \leq \sum_{k=j+1}^p \langle a^k, R^{j-1} - s'^j(Y^j) \rangle^2 = \sum_{k=j+1}^p \langle a^k, R^{j-1} \rangle^2 \leq \|R^{j-1}\|^2.$$

■

La preuve du théorème est alors immédiate. Il suffit de montrer que $R^p = 0$ \mathbb{P} -p.s. Or, d'après (ii) de la proposition 2.1, il est orthogonal à une base de \mathbb{R}^p , donc \mathbb{P} -p.s nul.

Le corollaire suivant est utile en pratique pour choisir la dimension d'un modèle.

Corollaire 2.1. *Soit Q_d la fraction d'information représentée par un modèle AA de dimension d :*

$$Q_d = 1 - \mathbb{E} \left[\left\| R^d \right\|^2 \right] / \text{Var} \left[\|X\|^2 \right].$$

On a alors $Q_0 = 0$, $Q_p = 1$ et la suite (Q_d) est croissante.

3. ETUDE DE DEUX MODÈLES AUTO-ASSOCIATIFS PARTICULIERS.

Nous allons considérer deux cas importants en pratique où il est possible de fournir une solution explicite à l'étape [R] : les modèles Auto-Associatifs Linéaires (AAL) et les modèles Auto-Associatifs de Régression (AAR). Ces modèles héritent des propriétés établies dans le paragraphe précédent. Dans chaque cas, nous complétons ces propriétés générales par quelques caractéristiques propres à ces deux cas particuliers.

3.1. Les modèles auto-associatifs linéaires. Nous nous intéressons au cas où $\mathcal{S}(\mathbb{R}, \mathbb{R}^p) = \mathcal{A}(\mathbb{R}, \mathbb{R}^p)$. De la proposition 2.1(i), il est immédiat qu'il suffit en fait de se restreindre aux fonctions de régression s linéaires, c'est à dire de la forme $s(t) = tb$, $t \in \mathbb{R}$, $b \in \mathbb{R}^p$. L'étape [R], s'écrit donc

$$[\mathbf{R}] \text{ Trouver } b^j = \arg \min_{x \in \mathbb{R}^p} \mathbb{E} [\|R^{j-1} - Y^j x\|^2], \text{ s.c. } \langle a^j, x \rangle = 1,$$

et l'on a l'analogie du théorème 2.1 suivant :

Théorème 3.1. *L'algorithme 2.1 construit un modèle auto-associatif linéaire de dimension d pour X de fonctions de régression $s^j(t) = tb^j$. De plus pour $d = p$, on a la décomposition :*

$$(1) \quad X = \mathbb{E}[X] + \sum_{k=1}^p Y^k b^k, \quad \mathbb{P} - p.s.$$

où les variables aléatoires Y^k , $k = 1, \dots, p$ sont orthogonales.

Avant de démontrer ce théorème, nous prouvons les propriétés suivantes.

Proposition 3.1. *Soit Σ^j la matrice de variance-covariance de R^j . Les fonctions de régression et les variables principales obtenues avec l'algorithme 2.1 vérifient les propriétés suivantes :*

- (i) Pour tout $1 \leq j \leq d$, $b^j = \Sigma^{j-1} a^j / ({}^t a^j \Sigma^{j-1} a^j)$.
- (ii) Pour tous $1 \leq i < j \leq p$, $\mathbb{E}[Y^i Y^j] = 0$.

Preuve :

- (i) Soit $\mathcal{L}(x, \lambda)$ le lagrangien associé au problème de minimisation de l'étape [R] :

$$\mathcal{L}(x, \lambda) = \mathbb{E} [\|R^{j-1} - Y^j x\|^2] + \lambda (\langle a^j, x \rangle - 1).$$

En annulant son gradient par rapport à x , on obtient l'équation

$$2\mathbb{E}[R^{j-1} Y^j] - 2x \mathbb{E}[Y^{j2}] + \lambda a^j = 0,$$

et par projection sur l'axe a^j , on obtient $\lambda = 0$ d'où

$$(2) \quad b^j = \mathbb{E}[R^{j-1} Y^j] / \mathbb{E}[Y^{j2}] = \Sigma^{j-1} a^j / ({}^t a^j \Sigma^{j-1} a^j).$$

- (ii) Le résultat se prouve par récurrence en notant $\text{HR}_k : \mathbb{E}[Y^i Y^j] = 0$, $1 \leq i < j \leq k$. HR_1 est trivialement vraie, supposons alors HR_k vraie et démontrons HR_{k+1} . Le vecteur aléatoire X se décompose ainsi :

$$(3) \quad X = \mathbb{E}[X] + \sum_{i=1}^k Y^i b^i + R^k,$$

soit, par projection,

$$\langle X - \mathbb{E}[X], a^{k+1} \rangle = \sum_{i=1}^k Y^i \langle b^i, a^{k+1} \rangle + Y^{k+1},$$

et pour $1 \leq j < k + 1$ on a alors :

$$\begin{aligned} \mathbb{E}[Y^j Y^{k+1}] &= \mathbb{E}\left[Y^j \langle X - \mathbb{E}[X], a^{k+1} \rangle\right] - \sum_{i=1}^k \mathbb{E}[Y^i Y^j] \langle b^i, a^{k+1} \rangle \\ &= \mathbb{E}\left[Y^j \langle X - \mathbb{E}[X], a^{k+1} \rangle\right] - \mathbb{E}[Y^{j^2}] \langle b^j, a^{k+1} \rangle, \end{aligned}$$

par HR_k . Or d'après (2) on a $b^j = \mathbb{E}[R^{j-1} Y^j] / \mathbb{E}[Y^{j^2}]$, par conséquent,

$$\mathbb{E}[Y^j Y^{k+1}] = \mathbb{E}\left[Y^j \langle a^{k+1}, X - \mathbb{E}[X] - R^{j-1} \rangle\right].$$

Par une décomposition analogue à (3), on a

$$X - R^{j-1} = \sum_{i=1}^{j-1} Y^i b^i,$$

et donc

$$\mathbb{E}[Y^j Y^{k+1}] = \sum_{i=1}^{j-1} \mathbb{E}[Y^i Y^j] \langle a^{k+1}, b^i \rangle = 0.$$

par HR_k puisque $j - 1 < k$. ■

La preuve du théorème 3.1 est alors une conséquence immédiate du théorème 2.1 et de la proposition 3.1. Remarquons de plus que, d'après la partie (i) de ma preuve, la contrainte de l'étape [R] est toujours satisfaite et est donc redondante.

Corollaire 3.1. *Si de plus l'index I de l'étape [A] est la variance projetée, i.e. $I(\langle x, R^{j-1} \rangle) = \text{Var}[\langle x, R^{j-1} \rangle]$, alors l'algorithme 2.1 effectue une ACP de X .*

Preuve : Il est bien connu que la solution a^j de l'étape de recherche de direction révélatrice [A] est le vecteur propre associé à la plus grande valeur propre λ_j de Σ^{j-1} . D'après la proposition 3.1(i), on a alors $b^j = a^j$.

Notons $A^j = a^{j^t} a^j$ et prenons comme hypothèse de récurrence

$$\text{HR}_k : \begin{cases} \Sigma^k &= \Sigma^0 - \sum_{j=1}^k \lambda_j A^j \\ R^k &= R^0 - \sum_{j=1}^k A^j R^0. \end{cases}$$

HR_0 est évidente. Supposons HR_k vraie et prouvons que HR_{k+1} l'est également. On a alors d'une part :

$$R^{k+1} = R^k - \langle a^{k+1}, R^k \rangle a^{k+1} = R^k - \langle a^{k+1}, X \rangle a^{k+1},$$

et d'autre part :

$$\Sigma^{k+1} = \Sigma^k + {}^t a^{k+1} \Sigma^k a^{k+1} A^{k+1} - A^{k+1} \Sigma^k - \Sigma^k A^{k+1} = \Sigma^k - \lambda_{k+1} A^{k+1},$$

et donc HR_{k+1} est vraie. On en déduit alors

$$\lambda_{k+1} a^{k+1} = \Sigma^k a^{k+1} = \Sigma^0 a^{k+1} - \sum_{j=1}^k \lambda_j \langle a^j, a^{k+1} \rangle a^j = \Sigma^0 a^{k+1},$$

ce qui prouve que a^{k+1} est aussi vecteur propre de Σ^0 associé à la valeur propre λ_{k+1} . En écrivant la décomposition de Jordan

$$\Sigma^0 = \sum_{k=1}^d \lambda_k A^k,$$

on déduit de HR_d que $\Sigma^d = 0$ et donc que R^d est presque sûrement constant. Les résidus étant centrés, on a $R^d = 0$, \mathbb{P} -p.s. et

$$(4) \quad X = \mathbb{E}[X] + \sum_{k=1}^d \langle a^k, X - \mathbb{E}[X] \rangle a^k, \quad \mathbb{P} - p.s.$$

qui est la décomposition produite par une ACP. ■

Remarquons que la fonction auto-associative F associée à l'ACP par (4) est linéaire. Il est possible de montrer qu'inversement, l'ACP est le seul modèle AA dont la fonction F est linéaire [14].

3.2. Les modèles auto-associatifs de régression. Nous considérons désormais le cas où $\mathcal{S}(\mathbb{R}, \mathbb{R}^p) = L^2_X(\mathbb{R}, \mathbb{R}^p)$. Dans ce cas, l'étape [R] possède une solution explicite :

$$[\text{R}] \quad s^j(Y^j) = \mathbb{E}[R^{j-1} | Y^j].$$

En effet, l'espérance conditionnelle est un projecteur orthogonal dans L^2_X . Elle vérifie de plus la contrainte puisque $\langle a^j, R^{j-1} \rangle = Y^j$. On a alors le résultat suivant :

Théorème 3.2. *L'algorithme 2.1 construit un modèle auto-associatif de dimension d . De plus si $d = p$ alors on a la décomposition exacte :*

$$(5) \quad X = \mathbb{E}[X] + \sum_{j=1}^p s^j(Y^j), \quad \mathbb{P} - p.s.$$

où Y^j et Y^{j+1} sont orthogonales, $j = 1, \dots, p-1$.

Nous allons tout d'abord prouver la proposition suivante :

Proposition 3.2. *Les résidus et les variables principales obtenus avec l'algorithme 2.1 vérifient les propriétés suivantes :*

- (i) Pour tout $1 \leq j \leq d$, $\mathbb{E}[R^j | Y^j] = 0$, \mathbb{P} -p.s.
- (ii) Pour tout $1 \leq j < d$, $\mathbb{E}[Y^j Y^{j+1}] = 0$.

Preuve :

- (i) Comme $R^j = R^{j-1} - s^j(Y^j)$, on a $\mathbb{E}[R^j | Y^j] = \mathbb{E}[R^{j-1} | Y^j] - \mathbb{E}[s^j(Y^j) | Y^j]$ et donc $\mathbb{E}[R^j | Y^j] = 0$, \mathbb{P} -p.s.
- (ii) On a $\mathbb{E}[Y^j Y^{j+1}] = \mathbb{E}[Y^j \langle a^{j+1}, R^j \rangle] = \mathbb{E}[Y^j \langle a^{j+1}, \mathbb{E}[R^j | Y^j] \rangle] = 0$ d'après (i). ■

La démonstration du théorème 3.2 est une conséquence directe du théorème 2.1 et de la proposition 3.2(ii).

4. MISE EN ŒUVRE DES MODÈLES AUTO-ASSOCIATIFS.

Considérons un échantillon (X_1, \dots, X_n) iid de loi \mathbb{P}_X inconnue. Le paramètre μ est alors estimé par la moyenne empirique $\bar{X} = 1/n \sum X_i$. Les deux étapes cruciales dans l'algorithme 2.1 sont [A] et [R] : la détermination de la direction révélatrice et l'estimation de la fonction de régression. Le choix de l'index I et de la classe de fonctions $\mathcal{S}(\mathbb{R}, \mathbb{R}^p)$ déterminent en effet à la fois la nature du modèle obtenu et la complexité de calcul associée aux problèmes d'optimisation [A] et [R].

4.1. Estimation de la fonction de régression. Pour ce problème on peut remarquer que si $\mathcal{S}(\mathbb{R}, \mathbb{R}^p)$ est l'ensemble des fonctions linéaires de \mathbb{R} dans \mathbb{R}^p alors d'après la proposition 3.1, il existe une unique solution donnée par $b^j = \Sigma^{j-1} a^j / ({}^t a^j \Sigma^{j-1} a^j)$ où Σ^{j-1} est la matrice de variance-covariance du résidu R^{j-1} . Il suffit donc de remplacer dans cette formule Σ^{j-1} par son estimateur empirique et a^j par son estimation obtenue à l'étape [A].

Dans le cas des modèles AAR, il s'agit d'estimer l'espérance conditionnelle de R^{j-1} sachant Y^j . Ce problème standard [18] peut être résolu (par exemple) par une régression par noyau [2] ou une régression spline [16].

Dans cet article, pour traiter les simulations et les données réelles, nous avons choisi de mettre en œuvre l'estimateur à noyau. Pour un exemple d'utilisation des splines dans un contexte similaire, nous renvoyons à [5]. Par rapport au problème classique de régression, nous avons une contrainte supplémentaire sur la fonction à estimer. A l'itération j , elle doit vérifier $P_{a^j} s^j = \text{Id}$. En se plaçant dans la base orthonormée B^j obtenue en complétant $\{a^1, \dots, a^j\}$, il suffit de réaliser $p - j$ régressions, composantes par composantes. Pour la composante numéro $k \in \{j + 1, \dots, p\}$ l'estimateur s'écrit

dans la base B^j :

$$(6) \quad \tilde{s}_k^j(u) = \sum_{i=1}^n \tilde{R}_{i,k}^{j-1} K_h(u - Y_i^j) \Big/ \sum_{i=1}^n K_h(u - Y_i^j),$$

où $\tilde{R}_{i,k}^{j-1}$ représente la k -ième coordonnée du résidu de l'individu i à la $(j - 1)$ -ième itération dans la base B^j . Y_i^j représente la valeur de la j -ième variable principale pour l'individu i . K_h est par exemple une densité normale centrée d'écart-type h appelé fenêtre dans ce contexte. Plus généralement, tout noyau de type Parzen-Rosenblatt convient.

4.2. Détermination des directions révélatrices. Le choix de l'index I est à la base de tout problème de Poursuite de Projection (PP) où il s'agit de trouver des directions "intéressantes". Nous renvoyons à [21] et [23] pour des articles de synthèse sur ce type de problème. Le sens du mot "intéressantes" dépend du problème d'analyse des données considéré. Par exemple, Friedman *et al* [10, 12], puis Hall [17], ont proposé un index pour faire apparaître des groupes ou, utilisent une distance à la normalité pour mettre en évidence d'éventuelles structures plus complexes du nuage de points. Une approche alternative peut être trouvée dans [4] où une métrique particulière est introduite dans l'ACP pour rechercher des groupes dans le nuage de points. Citons également les index dédiés à la recherche de points aberrants [27]. Ces index pourraient être utilisés conjointement avec les modèles AAL et seront étudiés dans un prochain article.

Dans le cadre des modèles AAR, il s'agit de rechercher des directions qui paramétrisent au mieux la variété que l'on cherche à estimer. Dans ce cadre, Demartines [8] propose un index qui favorise les directions dans lesquelles la projection conserve approximativement les distances. Girard [5] propose, sur un principe analogue, un index qui favorise les directions dans lesquelles la structure de voisinage est respectée par projection. Ces deux critères ont le désavantage de ne pas pouvoir être maximisés de manière simple, ou par des algorithmes d'optimisation standard.

Nous avons retenu ici une approche similaire à celle de Lebart [25] qui consiste à définir un coefficient de contiguïté dont la minimisation permet de déplier les structures non linéaires. En terme d'index, il s'agit de maximiser en x à l'itération j le rapport de formes quadratiques :

$$(7) \quad I(\langle x, R^{j-1} \rangle) = \sum_{i=1}^n \langle x, R_i^{j-1} \rangle^2 \Big/ \sum_{k=1}^n \sum_{\ell=1}^n m_{k\ell} \langle x, R_k^{j-1} - R_\ell^{j-1} \rangle^2.$$

La matrice $M = (m_{k\ell})$ est une matrice de contiguïté d'ordre 1, qui vaut 1 si R_ℓ^{j-1} est le plus proche voisin de R_k^{j-1} , 0 sinon. La direction révélatrice a^j obtenue est alors donnée par le vecteur propre associé à la plus grande

valeur propre de la matrice $V_j V_j^{\star-1}$ où

$$V_j^{\star} = \sum_{k=1}^n \sum_{\ell=1}^n m_{k\ell} {}^t(R_k^{j-1} - R_\ell^{j-1})(R_k^{j-1} - R_\ell^{j-1})$$

est proportionnelle à la matrice de covariance locale. La matrice

$$V_j = \sum_{k=1}^n {}^t R_k^{j-1} R_k^{j-1}$$

est proportionnelle à la matrice de covariance empirique de R^{j-1} . $V_j^{\star-1}$ désigne son inverse généralisée, V_j^{\star} n'étant pas inversible puisque R^j est orthogonal à $\{a_1, \dots, a_j\}$ d'après la proposition 2.1(ii). Notons que cette approche est équivalente à celle de Lebart lorsque la matrice de contiguïté M est symétrique.

5. EXEMPLES.

Nous présentons tout d'abord deux illustrations du principe de construction des modèles AAR sur des données simulées de faible dimension (paragraphes 5.1 et 5.2). Puis, nous appliquons les modèles AAR à un problème d'analyse d'images dans le paragraphe 5.3. Dans tous les cas, nous avons retenu pour le calcul des directions révélatrices l'index de contiguïté (7). De même, nous utilisons toujours l'estimateur à noyau (6) pour la régression de l'étape [R] associé à un noyau gaussien.

5.1. Premier exemple sur données simulées. Les données sont simulées à partir d'une loi dont le support est une variété de dimension 1 dans \mathbb{R}^3 . L'équation de la variété est donnée par

$$(8) \quad x \rightarrow (x, \sin x, \cos x).$$

La première coordonnée du vecteur aléatoire est répartie de manière uniforme sur l'intervalle $[-3\pi, 3\pi]$. Nous effectuons $n = 100$ tirages. Nous utilisons une seule itération de l'algorithme 2.1. Le \cos^2 entre l'axe de paramétrisation naturel (l'axe des abscisses) et l'axe estimé à l'étape [A] est de 0.998. La fenêtre de l'estimateur à noyau est choisie égale à $h = 0.3$. Après la première itération, il reste 0.03% de variance résiduelle. La figure 1 montre la variété (8), le nuage de points simulé et la variété estimée.

5.2. Second exemple sur données simulées. Les données sont simulées à partir d'une loi dont le support est une variété de dimension 2 dans \mathbb{R}^3 . L'équation de la variété est donnée par

$$(9) \quad (x, y) \rightarrow \left(x, y, \cos(\pi\sqrt{x^2 + y^2})(1 - \exp\{-64(x^2 + y^2)\}) \right).$$

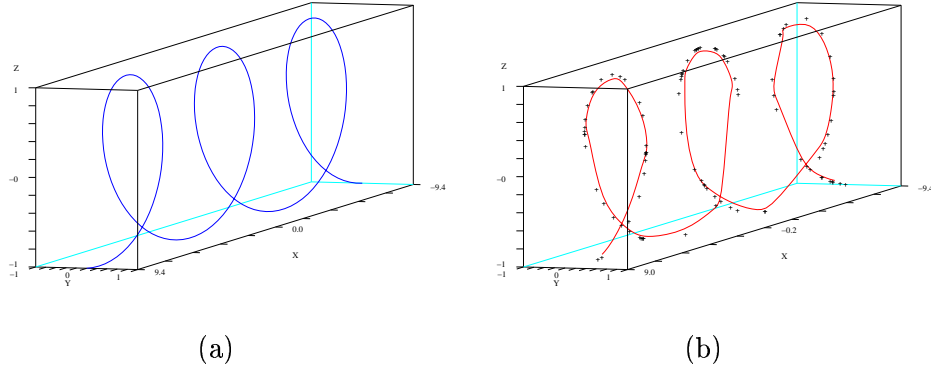


FIG. 1. Représentation de la variété (a), du nuage de points simulé et de la variété estimée (b).

Les deux premières coordonnées du vecteur aléatoire sont réparties suivant une loi uniforme sur le rectangle $[-1/2, 1/2] \times [-1, 1]$. Nous effectuons $n = 1000$ tirages.

Nous choisissons d'utiliser deux itérations. Le \cos^2 entre le premier axe de paramétrisation naturel (l'axe des ordonnées) et le premier axe a^1 estimé est de 0.998, et le \cos^2 entre le second axe de paramétrisation naturel (l'axe des abscisses) et le second axe a^2 estimé est de 0.999. La fenêtre de l'estimateur à noyau est choisie égale à $h = 0.12$. Après la première itération, il reste 15.9% de variance résiduelle, après la seconde, 2.38%.

La figure 2(a)–(b) montre la variété (9), ainsi que le nuage de points simulé. La première fonction de régression s^1 est présentée en bleu figure 2(c), elle représente de façon approximative la forme du nuage de points dans la direction des ordonnées. Elle ne prend pas en compte le creux induit par la fonction $(1 - \exp)$. Les résidus correspondants (c'est à dire après la première itération) sont représentés figure 2(e). On remarque qu'ils sont comme prévu orthogonaux à la première direction révélatrice a^1 . La seconde fonction de régression est dessinée en rouge figure 2(c). La figure 2(d) montre l'estimation de la variété obtenue après 2 itérations, les résidus associés sont présentés en figure 2(f). Ils sont orthogonaux aux deux directions révélatrices a^1 et a^2 . En fait, ils sont une conséquence de la mauvaise reconstruction du creux due au caractère non-additif de l'équation de la variété (9).

5.3. Exemple en analyse d'images. L'analyse d'images est depuis peu un domaine d'application privilégié de l'analyse des données [6]. Une image de $M \times M$ pixels peut en effet être représentée par un vecteur de \mathbb{R}^p avec $p = M^2$. Même pour des tailles d'images raisonnables, on obtient des données dans des espaces de très grande dimension. L'ACP est alors un

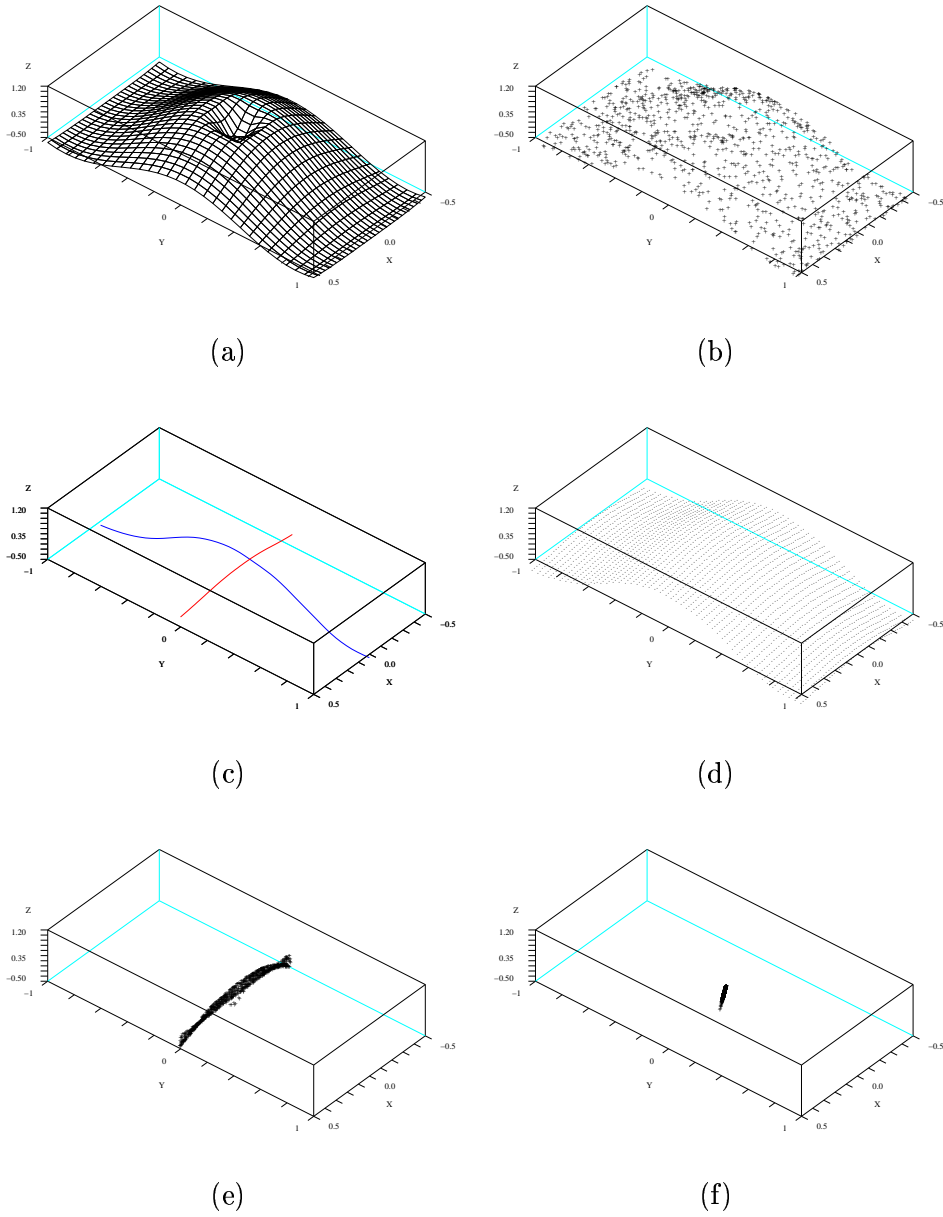


FIG. 2. Représentation de la variété (a), du nuage de points simulé (b), des deux fonctions de régression (c), de la variété de dimension 2 estimée et échantillonnée (d), des résidus à l'issue de la première itération (e), et de la seconde (f).

outil généralement très efficace pour réduire la dimension de ces données [26, 30]. Toutefois, même des déformations très simples dans l'espace de l'image peuvent se traduire par des fortes non linéarités dans l'espace \mathbb{R}^p , diminuant ainsi de beaucoup l'efficacité de l'ACP. Cette constatation est par exemple le point de départ des travaux de Capelli *et al* [3] qui proposent une ACP par groupes de façon à découper la structure non-linéaire de \mathbb{R}^p en sous-structures qui soient "plus linéaires". Nous étudions ici une base de 45 images de taille 256×256 issue de l'archive Centre For Intelligent Systems, Faculty of Human Sciences and Faculty of Technology, University of Plymouth. Il s'agit d'images d'un objet de synthèse vu sous différents angles d'élévation et d'azimuth. Un extrait de la base est représenté figure 3.

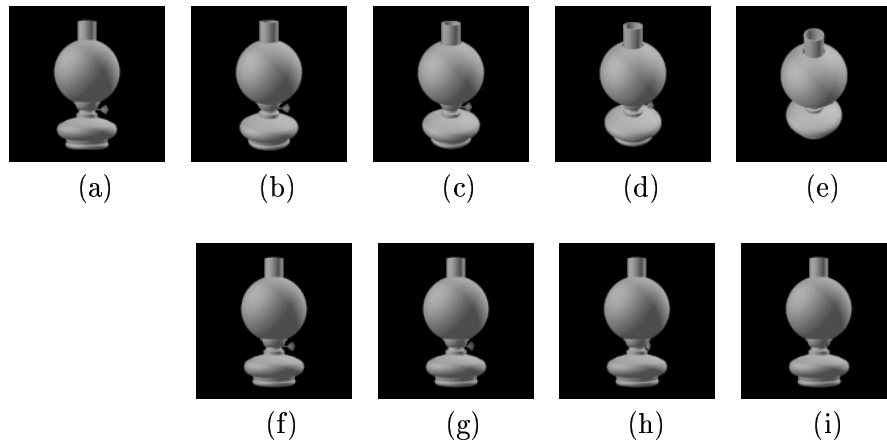


FIG. 3. Extrait de la base d'images. (a) image référence, (b-e) rotation utilisant l'angle d'élévation, (f-i) rotation utilisant l'angle d'azimuth.

Chaque image est représentée par un vecteur de dimension $M^2 = 256^2$. On obtient donc un nuage de $n = 45$ points en dimension 65536. Cependant, par un simple changement de repère, on se ramène à un ensemble de points en dimension $p = 44$. Notre but est de comparer les résultats de modélisation obtenus par une ACP classique et par les modèles AAR. Pour ces derniers, nous choisissons comme paramètre de lissage $h = 200$. La figure 4 montre l'évolution comparée du pourcentage d'information $100Q_d$ représentée par les modèles AAR et ACP de dimension d pour $d = 0, \dots, 10$ (cf corollaire 2.1).

Un modèle AAR de dimension 1 permet de représenter plus de 96% de l'information. A titre de comparaison, un modèle linéaire construit par ACP doit être de dimension 4 pour atteindre ce pourcentage. De plus, le coude dans la courbe associée aux modèles AAR semble indiquer que $d = 1$ est un choix correct de dimension de modèle. La projection de la variété correspondante dans l'espace formé par les trois premiers axes obtenus par ACP

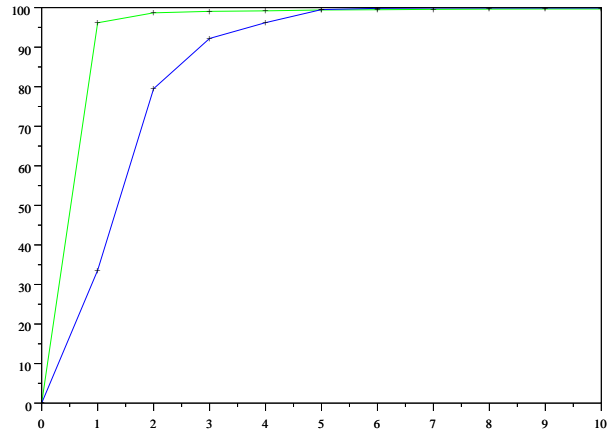


FIG. 4. Evolution du pourcentage d'information représenté en fonction de la dimension du modèle (bleu : modèle obtenu par ACP, vert : modèle AAR).

est représentée figure 5(a) où elle est superposée à la projection du nuage de points. Modéliser ce nuage de points par une variété de dimension 2 a aussi un sens puisque les images sont générées par rotation de l'objet dans deux directions orthogonales. La projection de la variété de dimension 2 estimée et échantillonnée est présentée figure 5(b).

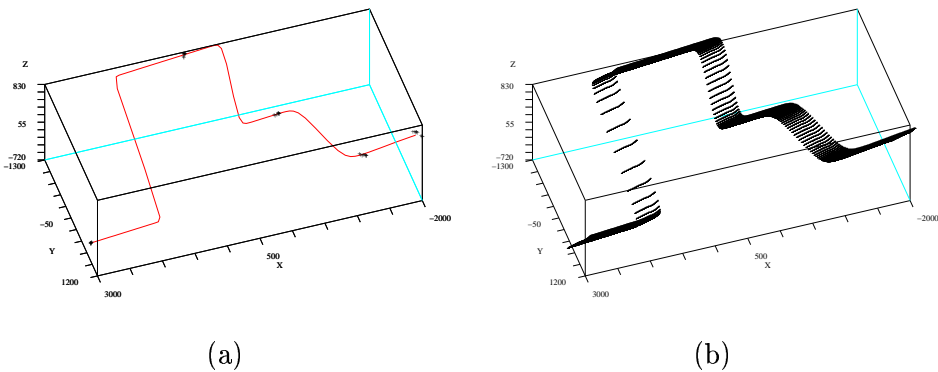


FIG. 5. Représentation des projections dans le repère formé par les trois premiers axes obtenus par ACP de : (a) la variété de dimension 1 estimée superposée au nuage de points, (b) la variété de dimension 2 estimée et échantillonnée.

Il est également intéressant de noter que la variable principale Y^1 associée au modèle AAR de dimension 1 est interprétable. Elle correspond à l'angle de rotation d'élévation. Pour s'en convaincre, il suffit de simuler des réalisations uniformes de cette variable et de représenter les images ainsi simulées avec le modèle AAR de dimension 1 (figure 6). La variable Y^2 n'est pas aussi aisément interprétable. Pour cette raison, le modèle AAR de dimension 1 semble préférable.

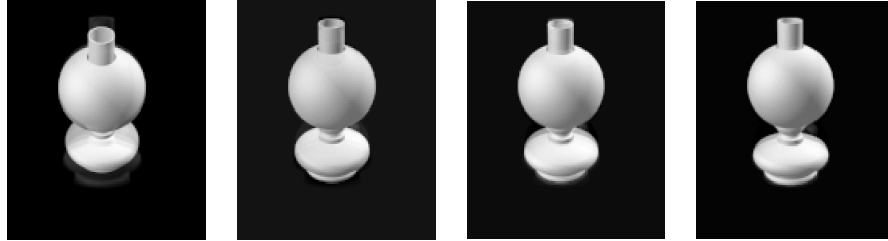


FIG. 6. Simulation de 4 images par le modèle AAR de dimension 1. La variable Y^1 est simulée uniformément sur l'intervalle $[\min_i Y_i^1, \max_i Y_i^1]$.

6. CONCLUSION ET PERSPECTIVES.

En résumé, les modèles AA offrent un cadre théorique agréable à la généralisation de l'ACP. De plus, ils bénéficient d'un algorithme de construction simple à mettre en œuvre. Nous avons illustré ici le fonctionnement des modèles AAR sur simulations et données réelles issues de l'analyse d'images. Nous projetons également de comparer ACP et modèles AAL dans des situations réelles. D'un point de vue théorique, il serait également intéressant d'établir les propriétés asymptotiques des estimateurs (6) et (7) de manière à proposer des tests.

RÉFÉRENCES

- [1] P. Besse & F. Ferraty (1995). "A fixed effect curvilinear model", *Computational Statistics*, 10(4), p. 339–351.
- [2] D. Bosq & J.P. Lecoutre (1987). *Théorie de l'estimation fonctionnelle*, Economica, Paris.
- [3] R. Capelli, D. Maio & D. Maltoni, (2001). "Multispace KL for pattern representation and classification", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(9), p. 977–996.
- [4] H. Caussinus & A. Ruiz-Gazen, (1995). "Metrics for finding typical structures by means of Principal Component Analysis", *Data science and its Applications, Harcourt Brace Japan*, p. 177–192.

- [5] B. Chalmond & S. Girard, (1999). "Nonlinear modeling of scattered multivariate data and its application to shape change", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(5), p. 422–432.
- [6] B. Chalmond, (2000). *Eléments de modélisation pour l'analyse d'images*, Springer-Verlag, Mathématiques et Applications 33.
- [7] P. Delicado (2001). "Another look at Principal curves and surfaces", *Journal of Multivariate Analysis*, 77, p. 84–116.
- [8] P. Demartines, (1994). "Analyse de données par réseaux de neurones auto-organisés", PhD thesis, (INPG).
- [9] J.F. Durand, (1993). "Generalized principal component analysis with respect to instrumental variables via univariate spline transformations", *Computational Statistics and Data Analysis*, 16, p. 423–440.
- [10] J.H. Friedman & J.W. Tukey (1974). "A Projection Pursuit algorithm for exploratory data analysis", *IEEE Trans. on computers*, C23 (9), p. 881–890.
- [11] J.H. Friedman & W. Stuetzle (1981). "Projection Pursuit Regression", *Journal of the American Statistical Association*, 76 (376), p. 817–823.
- [12] J.H. Friedman (1987). "Exploratory Projection Pursuit", *Journal of the American Statistical Association*, 82 (397), p. 249–266.
- [13] S. Girard, (2000). "A nonlinear PCA based on manifold approximation", *Computational Statistics*, 15(2), p. 145–167.
- [14] S. Girard, B. Chalmond & J.-M. Dinten (1998). "Position of Principal component analysis among auto-associative composite models", *Comptes Rendus de l'Académie des Sciences de Paris*, t.326, Série 1, p. 763–768.
- [15] S. Girard, (1996). "Design and statistical learning for nonlinear auto-associative models", PhD thesis, University of Cergy-Pontoise, (in French).
- [16] P.J. Green & B.W. Silverman (1994). *Non-parametric regression and generalized linear models*, Chapman and Hall, London.
- [17] P. Hall (1990). "On polynomial-based projection indices for exploratory projection pursuit", *The Annals of Statistics* 17(2) p. 589–605 (1989).
- [18] W. Härdle (1990). *Applied nonparametric regression*, Cambridge University Press, Cambridge.
- [19] T. Hastie & W. Stuetzle (1989). "Principal curves", *Journal of the American Statistical Association*, 84 (406), p. 502–516.
- [20] H. Hotelling (1933). "Analysis of a complex of statistical variables into principal components", *Journal of Educational Psychology*, 24, p. 417–441.
- [21] P.J. Huber (1985). "Projection Pursuit", *The Annals of Statistics*, 13(2), p. 435–475.
- [22] I. Jolliffe (1986). *Principal Component Analysis*, Springer-Verlag, New York.
- [23] M.C. Jones & R. Sibson (1987) ? "What is projection pursuit ?", *Journal of the Royal Statistical Society, Ser. A* 150, p. 1–36.
- [24] S. Klinke & J. Grassmann (2000) ? "Projection pursuit regression", *Wiley Series in Probability and Statistics*, p. 471–496.
- [25] L. Lebart (2000) ? "Contiguity analysis and classification", *Data Analysis*, Gaul W., Opitz O., Schader M., (eds), Springer, Berlin, p. 233–244.
- [26] B. Moghaddam & A. Pentland (1997). "Probabilistic visual learning for object representation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19 (7), p. 696–710.

- [27] J-X. Pan, W-K. Fung & K-T. Fang (2000), "Multiple outlier detection in multivariate data using projection pursuit techniques." *Journal of Statistical Planning and Inference*, 83(1), p. 153–167.
- [28] K. Pearson (1901). "On lines and planes of closest fit to systems of points in space", *The London, Edinburgh and Dublin philosophical magazine and journal of science*, Sixth Series 2, p. 559–572.
- [29] M.E. Tipping & C.M. Bishop (1999). "Probabilistic principal component analysis", *Journal of the Royal Statistical Society, Ser. B* 61(3), p. 611–622.
- [30] M. Uenohara & T. Kanade (1997). "Use of Fourier and Karhunen-Loeve decomposition for fast pattern matching with a large set of templates", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19 (8), p. 891–898.