INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Auto-Associative Models and Generalized Principal Component Analysis

Stéphane Girard — Serge Iovleff

N° 4364

Janvier 2002

THÈME 4



ISSN 0249-6399 ISRN INRIA/RR--4364--FR+ENG

INRIA
RHÔNE-ALPES

Auto-Associative Models and Generalized Principal Component Analysis

Stéphane Girard , Serge Iovleff *

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet IS2

Rapport de recherche n° 4364 — Janvier 2002 — 23 pages

Abstract: In this paper, we propose the auto-associative (AA) model to generalize the Principal component analysis (PCA). AA models have been introduced in data analysis from a geometrical point of view. They are based on the approximation of the observations scatterplot by a differentiable manifold. We propose here to interpret them as Projection Pursuit models adapted to the auto-associative case. We establish their theoretical properties and show how they extend the PCA ones. An iterative algorithm of construction is proposed and its principle is illustrated both on simulated and real data from image analysis.

Key-words: Auto-Associative models, Principal Component Analysis, Projection Pursuit, Regression.

* SAMOS, Université Paris 1, C2102, Centre Pierre Mendès France, 90 Rue de Tolbiac, 75634 Paris Cedex 13. Serge.Iovleff@univ-ubs.fr

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe, 38330 Montbonnot-St-Martin (France)
Téléphone : +33 4 76 61 52 00 — Télécopie +33 4 76 61 52 52

Modèles Auto-Associatifs et Analyse en Composantes Principales Généralisée

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.

Mots-clés : Modèles auto-associatifs, analyse en composantes principales, poursuite de projection, régression.

1 Introduction to auto-associative models.

Principal component analysis (PCA) [23] is a widely used method for dimension reduction in multivariate data analysis. It benefits from a simple geometrical interpretation. Given a set of points from $\mathbb{R}^p$ and an integer $0 \leq d \leq p$, PCA builds the d-dimensional affine subspace minimizing the Euclidean distance to the scatterplot [29]. Starting from this point of view, many authors have proposed nonlinear extensions of this technique. Principal curves or principal surfaces methods [19, 7] belong to this family of approaches. PCA can also be interpreted in terms of Projection pursuit [22, 24]. It builds the d-dimensional affine subspace maximizing the projected variance [21]. Indeed, the introduction of other criteria than the variance leads to various data exploration methods [12, 28]. In PCAIV-Spline (Principal component analysis of instrumental variables [9]) and curvilinear PCA [1] approaches, the introduction of nonlinear transformations of the coordinates is combined with a criteria of projected variance on the transformed data. Finally, it is also possible to associate a Gaussian probabilistic model to PCA [30], the affine subspace is then obtained through a maximization-likelihood estimation. This approach can lead to new dimension-reduction methods by considering some non Gaussian models, such as mixture models.

The extension of PCA to the nonlinear case without losing these interpretations is a difficult problem. Moreover, the definition of a satisfying probabilistic model is often impossible without specifying the observations distribution. As a consequence, such a method would be very specific and thus of little practical interest. Besides, the introduction of nonlinearity can lead to lose the geometrical interpretation of the model and the related concepts of principal variables, principal directions or residual inertia. Furthermore, the introduction of nonlinearity often yields existence, unicity and implementation problems.

We propose the auto-associative (AA) models as candidates to the generalization of PCA. AA models have been introduced in [15] from a geometrical point of view. They are based on the approximation of the observations scatterplot by a manifold. We show here that these models can also be interpreted as Pursuit Projection Regression models (PPR) [11, 25] adapted to the auto-associative case. Consequently, we propose a simple algorithm, similar to an iterative PCA, to implement them. We propose as well a probabilistic framework permitting to prove many theoretical properties.

First, we consider PCA from the Projection Pursuit point of view. If X is a $\mathbb{R}^p$ random vector with finite second order moment, it can be expanded as a sum of d orthogonal random variables and a residual by applying iteratively the following steps: [A] computation of the Axes, [P] Projection, [R] Regression and [U] Update (for a proof, see Section 3.1) :

Algorithm 1.1

- *For* $j = 0$, *define* $R^0 = X - \mathbb{E}[X]$.
- *For* $j = 1, \ldots, d$:

  [A] *Determine* $a^j = \arg\max_{x \in \mathbb{R}^p} \mathbb{E}\left[\left\langle x, R^{j-1}\right\rangle^2\right]$
  u.c. $\|x\| = 1$ *and* $\left\langle x, a^k\right\rangle = 0,\ 1 \leq k < j$.

  [P] *Compute* $Y^j = \left\langle a^j, R^{j-1}\right\rangle$.

  [R] *Determine* $b^j = \arg\min_{x \in \mathbb{R}^p} \mathbb{E}\left[\left\|R^{j-1} - Y^j x\right\|^2\right]$ *u.c.* $\left\langle x, a^j\right\rangle = 1$,
  (we find $b^j = a^j$*) and define* $s^j(t) = tb^j$, $t \in \mathbb{R}$.

  [U] *Compute* $R^j = R^{j-1} - s^j(Y^j)$.

The vectors a^j are called the principal directions, the random variables Y^j the principal variables, the functions s^j the regression functions and the random vectors R^j the residuals. Step [A] consists of computing an axis perpendicular to the previous ones maximizing a given criteria: Here the projected variance. In our opinion, this is an arbitrary choice when X is not Gaussian. Step [P] consists of projecting the residuals on this axis to determine the principal variables, and step [R] is devoted to the search of the linear function of the principal variables best approaching the residuals. Moreover, the limitation to a class of linear functions can be a too restrictive choice as soon as X is not Gaussian. Step [U] simply consists of updating the residuals.
AA models extend the previous algorithm by considering more general steps [A] and [R]. Step [A] is considered as a Projection Pursuit step, where many different criteria can be implemented. Step [R] is seen as a regression problem that can be addressed by general tools such as spline or kernel estimates. We show that this kind of generalization benefits from PCA main theoretical properties (construction of an exact model, decrease of the residuals, ...) or extends them (approximation of the scatterplot by a manifold instead of a linear subspace).
This article is organized as follows. In Section 2, auto-associative models are defined and their main properties are given. Two particular AA models are presented in Section 3 and their characteristics are studied. In Section 4, we review different criteria coming from Projection Pursuit algorithm and adapted to the framework of AA models. Several methods to estimate the regression functions are also presented. Finally, some illustrations are provided in Section 5 both on simulated data and for an application to image analysis.

2 Auto-associative models.

In a first time, we define auto-associative models as well as some related objects. In a second time, we propose an algorithm to compute them and establish its theoretical properties.

2.1 Definitions

Definition 2.1 *An application F: $\mathbb{R}^p \to \mathbb{R}^p$ is auto-associative of dimension d, if there exist d unit orthogonal vectors a^j and d functions s^j: $\mathbb{R} \to \mathbb{R}^p$ such that*

$$F = \left(\mathrm{Id}_{\mathbb{R}^p} - s^d \circ P_{a^d}\right) \circ \ldots \circ \left(\mathrm{Id}_{\mathbb{R}^p} - s^1 \circ P_{a^1}\right) = \coprod_{k=d}^{1} \left(\mathrm{Id}_{\mathbb{R}^p} - s^k \circ P_{a^k}\right),$$

$P_{a^j} \circ s^j = \mathrm{Id}_{\mathbb{R}^p}$ and $P_{a^k} \circ s^j = 0$, $1 \leq k < j \leq d$, with $P_{a^j}(x) = \langle a^j, x \rangle$. The vectors a^j are called the principal directions, the functions s^j are called the regression functions and we note $F \in \mathcal{A}^d_{a,s}$.

In the following, for sake of conciseness, the product will represent the composition. The proof of the following lemma can be found in [13].

Lemma 2.1 *Consider $F \in \mathcal{A}^d_{a,s}$, and suppose that the s^j, $j = 1, \ldots, d$ are $C^1(\mathbb{R}, \mathbb{R}^p)$. Then, the equation $F(x) = 0$ defines a differentiable d-dimensional manifold.*

Consider a square integrable random vector $X \in \mathbb{R}^p$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We note $\mathbb{P}_X$ the law of X on $\mathbb{R}^p$, and $\mathrm{L}^2_X(\mathbb{R}, \mathbb{R}^p)$ the set of the functions s from $\mathbb{R}$ to $\mathbb{R}^p$ such that $s \circ P_a$ is $\mathbb{P}_X$ square integrable for every normed vector $a \in \mathbb{R}^p$.

Definition 2.2 *X verifies a d-dimensional auto-associative model with principal directions $(a^1, \ldots, a^d)$, regression functions $(s^1, \ldots, s^d)$ and residual ε, if X verifies $F(X - \mu) = \varepsilon$ where $F \in \mathcal{A}^d_{a,s}$, $\mu \in \mathbb{R}$ and where ε is a centered random vector.*

Besides, we say that X verifies a linear AA model when the regression functions are linear. Let us give two simple examples of auto-associative models:

- Every X satisfies a 0-dimensional AA model (choose $F = \mathrm{Id}$, $\mu = \mathbb{E}[X]$ and $\varepsilon = X - \mathbb{E}[X]$). We then have $\mathrm{Var}\left[\|\varepsilon\|^2\right] = \mathrm{Var}\left[\|X\|^2\right]$.
- Similarly, X always satisfies a p-dimensional AA model. In this case $F = 0$, $\mu = 0$ and $\varepsilon = 0$ yield $\mathrm{Var}\left[\|\varepsilon\|^2\right] = 0$.

In practice, it is important to find a balance between these two extreme cases by constructing a d-dimensional model with $d \ll p$ and $\mathrm{Var}\left[\|\varepsilon\|^2\right] \ll \mathrm{Var}\left[\|X\|^2\right]$. For example, in the case where X is centered with a covariance matrix Σ of rank d, it satisfies a d-dimensional linear AA model with a null residual. Indeed, note a^j, $j = 1, \ldots, d$ the eigenvectors of Σ associated to the positive eigenvalues. We show in Corollary 3.1 that

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

with $\varepsilon = 0$ $\mathbb{P}$-.a.s. define a linear auto-associative model for X. This is the expansion of X obtained by PCA.
We now propose an efficient algorithm to build some auto-associative models which are not necessarily linear, with small dimension and small residual variance. In this aim, we introduce two definitions:

Definition 2.3 *A set $\mathcal{S}(\mathbb{R}, \mathbb{R}^p)$ of measurable functions from $\mathbb{R}$ to $\mathbb{R}^p$ is admissible whenever it is a closed subset of $\mathrm{L}^2_X(\mathbb{R}, \mathbb{R}^p)$ and it verifies the following condition:*

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

$(\mathcal{R})$ can be interpreted as an invariance condition with respect to translation. A possible choice of $\mathcal{S}(\mathbb{R}, \mathbb{R}^p)$ is the set of affine functions from $\mathbb{R}$ to $\mathbb{R}^p$. This example is treated in detail in Section 3.1.

Definition 2.4 *Let $a \in \mathbb{R}^p$ be an unit vector. An index I: $\mathbb{R} \to \mathbb{R}$ is a functional measuring the interest of the projection of the random vector X on a (i.e. $\langle a, X \rangle$) with a non negative real number.*

A possible choice of I is $I(\langle a, X \rangle) = \mathrm{Var}\left[\langle a, X \rangle\right]$, the projected variance. Some other examples are presented in Section 4.2.

2.2 Construction of auto-associative models

Let $\mathcal{S}(\mathbb{R}, \mathbb{R}^p)$ be a set of admissible functions and $d \in \{0, \ldots, p\}$. Consider the following algorithm:

Algorithm 2.1

- *For $j = 0$, define $\mu = \mathbb{E}[X]$ and $R^0 = X - \mu$.*
- *For $j = 1, \ldots, d$:*
  - [A] *Determine* $a^j = \arg\max_{x \in \mathbb{R}^p} I(\langle x, R^{j-1} \rangle)$
    u.c. $\|x\| = 1$, $\langle x, a^k \rangle = 0$, $1 \leq k < j$.
  - [P] *Compute* $Y^j = \langle a^j, R^{j-1} \rangle$.
  - [R] *Choose* $s^j \in \arg\min_{s \in \mathcal{S}(\mathbb{R}, \mathbb{R}^p)} \mathbb{E}\left[\left\| R^{j-1} - s(Y^j) \right\|^2 \right]$ *u.c.* $P_{a^j} \circ s^j = \mathrm{Id}$.
  - [U] *Compute* $R^j = R^{j-1} - s^j(Y^j)$.

We prove in Theorem 2.1 that this algorithm builds a d-dimensional auto-associative model. We also show that it builds an exact representation of X in p iterations.
It is clear that step [R] strongly depends on the choice of $\mathcal{S}(\mathbb{R}, \mathbb{R}^p)$. The existence of a solution to the minimization problem is established thanks to the conditions imposed on $\mathcal{S}(\mathbb{R}, \mathbb{R}^p)$.

In particular, condition $(\mathcal{R})$ ensures that there exist some functions in $\mathcal{S}(\mathbb{R}, \mathbb{R}^p)$ verifying the constraint of the minimization problem. The unicity of the solution is not established without an additional convexity condition. In this paper we focus on two extreme cases. The choice $\mathcal{S}(\mathbb{R}, \mathbb{R}^p) = \mathcal{A}(\mathbb{R}, \mathbb{R}^p)$, the set of the affine functions from $\mathbb{R}$ to $\mathbb{R}^p$ is examined in Section 3.1, and the choice $\mathcal{S}(\mathbb{R}, \mathbb{R}^p) = \mathrm{L}^2_X(\mathbb{R}, \mathbb{R}^p)$ is considered in Section 3.2. The choice of the index I is discussed in Section 4.2.

Theorem 2.1 *Algorithm 2.1 builds a d-dimensional AA model with principal directions* $(a^1, \ldots, a^d)$*, regression functions* $(s^1, \ldots, s^d)$ *and residuals* $\varepsilon = R^d$*. Moreover, when* $d = p$ *then* $\varepsilon = R^p = 0$ *and the exact expansion holds:*

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

Note that these properties are quite general, since they do not depend neither on the index I, nor on the subset of admissible functions $\mathcal{S}(\mathbb{R}, \mathbb{R}^p)$. We provide in Section 3 a few additional properties corresponding to some particular choices of I and $\mathcal{S}(\mathbb{R}, \mathbb{R}^p)$.
We first prove the following proposition:

Proposition 2.1 *The residuals and the regression functions obtained with Algorithm 2.1 verify the following properties :*

(i) For all $1 \leq j \leq d$, $\mathbb{E}\left[R^j\right] = \mathbb{E}\left[Y^j\right] = \mathbb{E}\left[s^j(Y^j)\right] = 0$.

(ii) For all $1 \leq k \leq j \leq d$, $\left\langle a^k, R^j \right\rangle = 0$, $\mathbb{P}$*-a.s.*

(iii) For all $1 \leq k < j \leq d$, $\left\langle a^k, s^j(Y^j) \right\rangle = 0$, $\mathbb{P}$*-a.s.*

(iv) The sequence of the residual norms is $\mathbb{P}$*-a.s. non increasing.*

Proof :

(i) The proof is done recursively on j. Let us note H_j the hypothesis $\mathbb{E}\left[R^j\right] = 0$. It is clear that H_0 is true. Suppose H_{j-1} is verified. We thus have,

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

Now, s^j is a solution of step [R] and then $\mathbb{E}\left[s^j(Y^j)\right] = 0$. This last equality can be proved by contradiction. If $\mathbb{E}\left[s^j(Y^j)\right] \neq 0$, then introduce $\mu^j = \mathbb{E}\left[s^j(Y^j)\right]$ and $s'^j = s^j - \mu^j$. Since $\left\langle a^j, s^j \right\rangle = \mathrm{Id}$ and $\mathbb{E}\left[Y^j\right] = \mathbb{E}\left[\left\langle a^j, R^{j-1} \right\rangle\right] = 0$ by H_{j-1}, we have $\left\langle a^j, \mu \right\rangle = 0$ and thus we also have $\left\langle a^j, s'^j \right\rangle = \mathrm{Id}$. Moreover, from condition $(\mathcal{R})$, we have $s'^j \in \mathcal{S}(\mathbb{R}, \mathbb{R}^p)$, and therefore

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

since R^{j-1} is centered. This contradicts the hypothesis of minimality of s^j. As a conclusion, $\mathbb{E}\left[R^j\right] = -\mathbb{E}\left[s^j(Y^j)\right] = 0$.

(ii) and (iii) The proof is also done by induction on j. Note H_j the hypothesis $\forall k \leq j$, $\langle a^k, R^j \rangle = 0$. H_1 is true since

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

Suppose H_{j-1} is verified and let us prove H_j. For $k = j$, we have

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

For $k < j$, we have with H_{j-1}:

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

Now, s^j is a solution of step [R] and thus minimizes

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

From H_{j-1} and condition $(\mathcal{R})$, the minimum is reached for a function s^j such that $\langle a^k, s^j \rangle = 0$ (the proof is done by contradiction as in (i)). To conclude, $\langle a^k, R^j \rangle = 0$ and $\langle a^k, s^j \rangle = 0$, which both prove H_j and (iii).

(iv) Consider $j \geq 1$ and $s'^j \in \mathcal{S}(\mathbb{R}, \mathbb{R}^p)$ given by $s'^j = \langle s^j, a^j \rangle a^j$. We have

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

The first term is null in view of (ii). Condition $\langle a^j, s^j \rangle = \mathrm{Id}$ entails that the second term is null too. Finally, in view of the definition of s'^j:

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

■

The proof of Theorem 2.1 is now straightforward. It only remains to show that $R^p = 0$ $\mathbb{P}$-a.s. In view of (ii) and Proposition 2.1, $\mathbb{R}^p$ is orthogonal to a $\mathbb{R}^p$ basis, and therefore it is $\mathbb{P}$-a.s null.
The following corollary will reveal useful to choose the dimension of a model.

Corollary 2.1 *Let Q_d be the information ratio represented by the d-dimensional AA model:*

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

Then, $Q_0 = 0$, $Q_p = 1$ and the sequence (Q_d) is non decreasing.

3 Two particular auto-associative models.

We consider two important cases in practice where step [R] has an explicit solution: the linear auto-associative models (LAA) and the auto-associative regression models (AAR). Clearly, these models inherit from the properties established in the previous section. In both cases, we precise these general properties by giving some further characteristics.

3.1 Linear auto-associative models

We focus on the case where $\mathcal{S}(\mathbb{R}, \mathbb{R}^p) = \mathcal{A}(\mathbb{R}, \mathbb{R}^p)$. From Proposition 2.1(i), it is straightforward that we can restrict ourselves to linear regression fonctions s i.e. such that $s(t) = tb$, $t \in \mathbb{R}$, $b \in \mathbb{R}^p$. Thus, step [R] can be rewritten as:

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

and we have a result similar to Theorem 2.1:

Theorem 3.1 *Algorithm 2.1 builds a d-dimensional LAA model with regression functions $s^j(t) = tb^j$. Moreover, for $d = p$, the following expansion holds:*

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

and the principal variables Y^k, $k = 1, \ldots, p$ are orthogonal.

We first prove the following properties.

Proposition 3.1 *Let Σ^j be the covariance matrix of R^j. The regression functions and the principal variables obtained with Algorithm 2.1 share the following properties :*

(i) For all $1 \leq j \leq d$, $b^j = \Sigma^{j-1} a^j / ({}^t a^j \Sigma^{j-1} a^j)$.

(ii) For all $1 \leq i < j \leq p$, $\mathbb{E}\left[Y^i Y^j\right] = 0$.

Proof :

(i) Let $\mathcal{L}(x,\lambda)$ be the Lagrangian associated to the minimization problem of step [R]:

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

Annulating its gradient with respect to x, we obtain the equation

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

and projecting on the axis a^j, it yields $\lambda = 0$ leading to

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

(ii) The result can be proved by induction by noting H_k : $\mathbb{E}\left[Y^i Y^j\right] = 0$, $1 \leq i < j \leq k$. H_1 is straightforwardly true. Let us suppose that H_k is true and prove H_{k+1}. The random vector X can be expanded as :

$$X = \mathbb{E}[X] + \sum_{i=1}^{k} Y^i b^i + R^k. \tag{1}$$

Hence, by projection,

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

and for $1 \leq j < k+1$ we thus obtain:

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

by H_k. Taking into account (i), we have $b^j = \mathbb{E}\left[R^{j-1}Y^j\right]/\mathbb{E}\left[{Y^j}^2\right]$, and consequently,

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

An expansion similar to (1) yields

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

and then

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

by H_k since $j - 1 < k$. ■

Theorem 3.1 is then an immediate consequence of Theorem 2.1 and Proposition 3.1. Let us note that, from the part (i) of the proof, the constraint of step [R] is always satisfied and thus inactive.

Corollary 3.1 *If, moreover, the index I of step [A] is the projected variance, i.e.* $I(\langle x, R^{j-1}\rangle) = \mathrm{Var}\left[\langle x, R^{j-1}\rangle\right]$, *then Algorithm 2.1 computes the PCA model of* X.

Proof : It is well-known that the solution a^j of step [A] is the eigenvector associated to the larger eigenvalue λ_j of Σ^{j-1}. From Proposition 3.1(i), we then obtain $b^j = a^j$. Introducing $A^j = a^j\,{}^t a^j$ we consider the induction hypothesis

$$H_k : \Sigma^k = \Sigma^0 - \sum_{j=1}^{k} \lambda_j A^j,\ R^k = R^0 - \sum_{j=1}^{k} A^j R^0.$$

H_0 is straightfowardly true. Supposing H_k is true we now prove that H_{k+1} is also true. We have on one hand :

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

and on the other hand :

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

and thus H_{k+1} is true. It yields

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

which proves that a^{k+1} is also the eigenvector of Σ^0 associated to the eigenvalue λ_{k+1}. Introducing the Jordan's expansion

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

we deduce from H_d that $\Sigma^d = 0$ and thus that R^d is almost surely constant. Since the residuals are centered, we have $R^d = 0$, $\mathbb{P}$-a.s. and

$$X = \mathbb{E}[X] + \sum_{k=1}^{d} \langle a^k, X - \mathbb{E}[X]\rangle\, a^k,\ \ \mathbb{P} - a.s. \tag{2}$$

which is the expansion produced by a PCA. ■

Remark that the auto-associative function F associated to a PCA by (2) is linear. It is possible to show that, conversely, PCA is the only AA model associated to a linear function F [14].

3.2 Auto-associative regression models

Herein, we consider the case where $\mathcal{S}(\mathbb{R}, \mathbb{R}^p) = \mathrm{L}^2_X(\mathbb{R}, \mathbb{R}^p)$. In this case, step [R] has an explicit solution:

[R] $s^j(Y^j) = \mathbb{E}\left[R^{j-1}|Y^j\right]$,

since the conditional expectation is an orthogonal projector in L^2_X and it satisfies the constraint. We thus have the following result:

Theorem 3.2 *Algorithm 2.1 builds a d-dimensional auto-associative model. Moreover, when $d = p$, we have the exact expansion:*

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

where the principal variables Y^j et Y^{j+1} are orthogonal, $j = 1, \ldots, p-1$.

We first prove the following proposition:

Proposition 3.2 *The residuals and the principal variables obtained with Algorithm 2.1 verify the following properties :*

(i) For all $1 \leq j \leq d$, $\mathbb{E}\left[\, R^j \,\middle|\, Y^j\right] = 0$, $\mathbb{P}$-a.s.

(ii) For all $1 \leq j < d$, $\mathbb{E}\left[Y^j Y^{j+1}\right] = 0$.

Proof :

(i) Since $R^j = R^{j-1} - s^j(Y^j)$, we have $\mathbb{E}\left[R^j \left| Y^j\right.\right] = \mathbb{E}\left[R^{j-1} \left| Y^j\right.\right] - \mathbb{E}\left[s^j(Y^j) \left| Y^j\right.\right]$ and consequently $\mathbb{E}\left[R^j|Y^j\right] = 0$, $\mathbb{P}$-a.s.

(ii) We have $\mathbb{E}\left[Y^j Y^{j+1}\right] = \mathbb{E}\left[Y^j \left\langle a^{j+1}, R^j\right\rangle\right] = \mathbb{E}\left[Y^j \left\langle a^{j+1}, \mathbb{E}\left[R^j|Y^j\right]\right\rangle\right] = 0$ from (i). ■

Theorem 3.2 is a direct consequence of Theorem 2.1 and Proposition 3.2(ii).

4 Implementation.

Consider a sample $(X_1, \ldots, X_n)$ iid from an unknown distribution $\mathbb{P}_X$. The parameter μ is estimated by the empirical mean $\bar{X} = 1/n \sum X_i$. The two crucial steps in Algorithm 2.1 are [A] and [R]: the determination of the principal directions and the estimation of the regression functions. The choices of the index I and of the class of functions $\mathcal{S}(\mathbb{R}, \mathbb{R}^p)$ both determine the nature of the obtained model and the complexity of the computation associated to the optimization problems [A] and [R].

4.1 Estimation of the regression function

Remark that, when $\mathcal{S}(\mathbb{R}, \mathbb{R}^p)$ is the set of linear functions from $\mathbb{R}$ to $\mathbb{R}^p$ then, from Proposition 3.1, a unique solution exists and is given by $b^j = \Sigma^{j-1}a^j / ({}^t a^j \Sigma^{j-1} a^j)$ where Σ^{j-1} is the covariance matrix of the residual R^{j-1}. Then, b^j is estimated by replacing in this formula Σ^{j-1} by its empirical estimate and a^j by the estimation obtained at step [A].
In the case of AAR models, the problem reduces to estimating the conditional expectation of R^{j-1} given Y^j. This standard problem [18] can be solved (for example) by kernel regression [2] or spline regression [16].
Here, we have chosen a kernel estimate to deal with the simulated and real data. For an example of the use of spline regression in a similar context, we refer to [5]. Compared to a classical regression problem, we have a additional constraint on the function to estimate. At iteration j, it has to verify $P_{a^j} \circ s^j = \mathrm{Id}$. Fortunately, in the orthogonal basis B^j of $\mathbb{R}^p$ obtained by completing $\{a^1, \ldots, a^j\}$, step [R] reduces to $(p-j)$ independent regressions. Hence, each coordinate $k \in \{j+1, \ldots, p\}$ of the estimate can be written in the basis B^j as:

$$\tilde{s}_k^j(u) = \sum_{i=1}^{n} \tilde{R}_{i,k}^{j-1} K_h(u - Y_i^j) \Big/ \sum_{i=1}^{n} K_h(u - Y_i^j) \,, \tag{3}$$

where $\tilde{R}_{i,k}^{j-1}$ represents the k-th coordinate of the residual of the observation i at the $(j-1)$-th iteration in the basis B^j, Y_i^j represents the value of the j-th principal variable for the observation i and the kernel K_h is for example the density of a Gaussian variable with zero mean and standard deviation h, called window in this context. More generally, any Parzen-Rosenblatt kernel is convenient. For an automatic choice of h, we refer to [20], chapter 6.

4.2 Computation of principal directions

The choice of the index I is the key point of any Projection Pursuit problem where it is needed to find "interesting" directions. We refer to [22] and [24] for a review on this topic. The meaning of the word "interesting" depends on the considered data analysis problem. For instance, Friedman *et al* [10, 12], and more recently Hall [17], have proposed an index to find clusters or use deviation from the normality measures to reveal more complex structures of the scatterplot. An alternative approach can be found in [4] where a particular metric is introduced in PCA in order to detect clusters. We can also mention the indices dedicated to outliers detection [28]. We plan to combine those indices with LAA models in a next article.
In the framework of AAR models, we are interested in finding parametrization directions for the manifold to be estimated. In this aim, Demartines [8] proposes an index that favours the directions in which the projection approximatively preserves distances. From a similar principle, Girard [5] proposes an index revealing the directions in which the neighbourhood structure is invariant with respect to projection. Both criteria require complex optimization algorithms.
We have chosen an approach similar to Lebart one's [26]. It consists of defining a contiguity coefficient whose minimization allows to unfold nonlinear structures. In terms of index, at

each iteration j, we have to maximize with respect to x the ratio of quadratic functions:

$$I(\langle x, R^{j-1}\rangle) = \sum_{i=1}^{n} \left\langle x, R_i^{j-1}\right\rangle^2 \Bigg/ \sum_{k=1}^{n}\sum_{\ell=1}^{n} m_{k\ell} \left\langle x, R_k^{j-1} - R_\ell^{j-1}\right\rangle^2 . \tag{4}$$

The matrix $M = (m_{k\ell})$ is a contiguity matrix of order 1, whose value is 1 when R_ℓ^{j-1} is the nearest neighbor of R_k^{j-1}, 0 otherwise. The resulting principal direction a^j is then given by the eigenvector associated to the largest eigenvalue of the matrix $V_j^{\star -1} V_j$ where

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

is proportional to the local covariance matrix. The matrix

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

is proportional to the empirical covariance matrix of R^{j-1}. $V_j^{\star -1}$ should be read as the generalized inverse of $V_j^\star$ since it is not invertible, R^j being orthogonal to $\{a^1, \ldots, a^j\}$ from Proposition 2.1(ii). Note that this approach is equivalent to the one of Lebart when the contiguity matrix M is symmetric.

5 Examples.

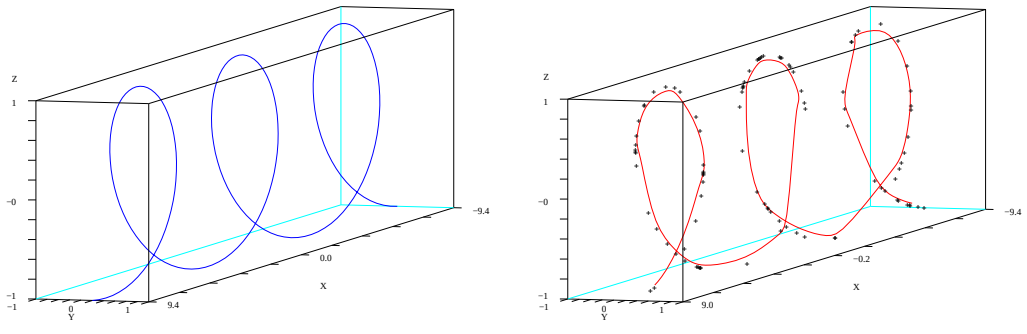
We first present two illustrations of the construction principle of AAR models on low dimensional data (Section 5.1 and 5.2). Second, AAR models are applied to an image analysis problem in Section 5.3. In all cases, the principal directions are computed thanks to the contiguity index (4). Similarly, we always use a Gaussian kernel method (3) for the regression step [R].

5.1 First example on simulated data

The data are simulated from a distribution whose support is a one-dimensional manifold in $\mathbb{R}^3$. The equation of the manifold is given by

$$x \to (x, \sin x, \cos x). \tag{5}$$

The first coordinate of the random vector is uniformly distributed on the interval $[-3\pi, 3\pi]$ and $n = 100$ points are simulated. We use only one iteration of Algorithm 2.1. The square cosine between the natural axis of parameterization (the x-axis) and the axis estimated in step [A] is as high as 0.998. The window of the kernel estimate is chosen equal to $h = 0.3$. After the first iteration, the residual variance equals 0.03%. The theoretical manifold, the simulated scatterplot and the estimated manifold are represented on Figure 1 for comparison.



(a) (b)

Figure 1: Representation of the manifold (a), the simulated scatterplot and the estimated manifold (b).

5.2 Second example on simulated data

The data are simulated from a distribution whose support is a two-dimensional manifold in $\mathbb{R}^3$. The equation of the manifold is given by

$$(x, y) \to \left(x, y, \cos(\pi\sqrt{x^2 + y^2})(1 - \exp\{-64(x^2 + y^2)\})\right). \tag{6}$$

The first two coordinates of the random vector are uniformly distributed on $[-1/2, 1/2] \times [-1, 1]$ and $n = 1000$ points are simulated.
We limit ourselves to two iterations. The square cosine between the first natural axis of parameterization (the y-axis) and the first estimated axis a^1 is as high as 0.998 and the square cosine between the second natural axis of parameterization (the x-axis) and the second estimated axis a^2 is 0.999. The window of the kernel estimate is chosen equal to $h = 0.12$. After the first and second iterations, the residual variance is respectively equal to 15.9% and 2.38%.
The manifold (6) and the simulated scatterplot are represented on Figure 2(a)–(b). The first regression function s^1 is plotted in blue on Figure 2(c). It approximatively represents the shape of the scatterplot in the y-direction. It can be noted that it does not take into account the hole induced by the exponential function. The corresponding residuals (after the first iteration) are represented on Figure 2(e). Remark that, accordingly to Proposition 2.1(ii), they are orthogonal to the first principal direction a^1. The second regression function is drawn in red on Figure 2(c). Figure 2(d) shows the estimated manifold after 2 iterations. The associated residuals are represented on Figure 2(f). They are orthogonal to the two principal directions a^1 and a^2. In fact, they are a consequence of the poor reconstruction of the hole due to the non additive nature of the manifold equation (6).

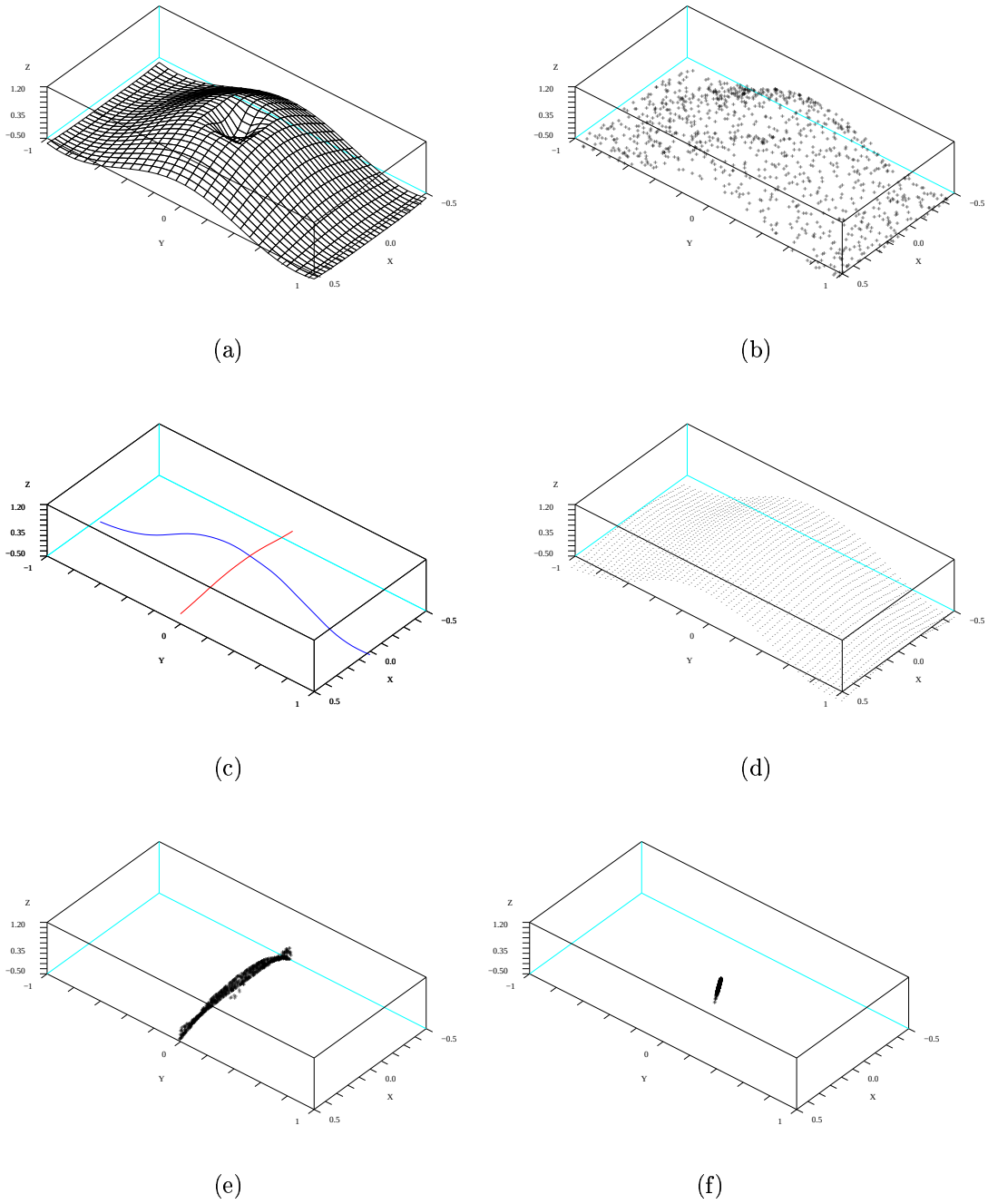


Figure 2: Representation of the manifold (a), of the simulated scatterplot (b), of the two regression functions (c), of the two-dimensional manifold estimated and sampled (d), and of the residuals after the first iteration (e), and the second one (f).

5.3 Example in image analysis

Image analysis is a privileged application field for multivariate analysis [6], since an image with $M \times M$ pixels can be represented by a vector of $\mathbb{R}^p$ with $p = M^2$. Even with images of moderate size, this yields data in spaces of extremely large dimension. PCA is a tool usually very efficient to reduce the dimension of such data [27, 31]. However, even some very simple deformation in the image space can lead to consequent nonlinearities in the space $\mathbb{R}^p$. In such situation, the PCA efficiency is significatively decreased. This remark is the starting point of the work of Capelli *et al* [3] who propose a "piecewise" PCA. The idea is to split the nonlinear structure of $\mathbb{R}^p$ into approximatively linear sub-structures. We study here a database of 45 images of size 256×256 taken from the archive of Centre For Intelligent Systems, Faculty of Human Sciences and Faculty of Technology, University of Plymouth. It is composed of images of a synthesis object viewed under different elevation and azimuth angles. A sample from the database is presented on Figure 3.

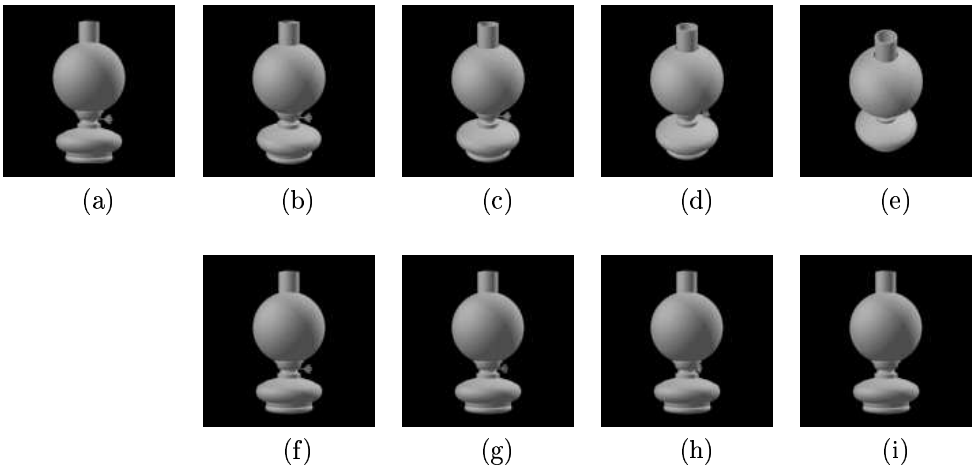


Figure 3: A sample from the image database. (a) reference image, (b-e) rotation using the elevation angle, (f-i) rotation using the azimuth angle.

Each image is represented as a vector of dimension $M^2 = 256^2$. We then obtain a scatterplot of $n = 45$ points in dimension 65536. However, a simple rotation of axes allows to represent this set of points in dimension $p = 44$. In the following, our aim is to compare the modelling results obtained by a classical PCA and by AAR models. For those last ones, we select $h = 200$ for the smoothing parameter. Figure 4 shows the compared information percentage $100Q_d$ represented by AAR and PCA models of increasing dimension $d = 0, \ldots, 10$ (see Corollary 2.1).
The one-dimensional AAR model allows to represent more than 96% of the information. As a comparison, a linear model built by PCA should be of dimension 4 to reach this percentage. Moreover, the elbow in the curve associated to AAR models seems to indicate

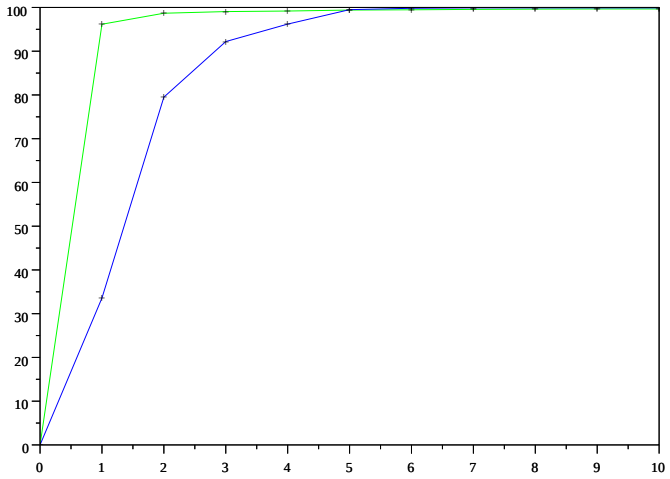
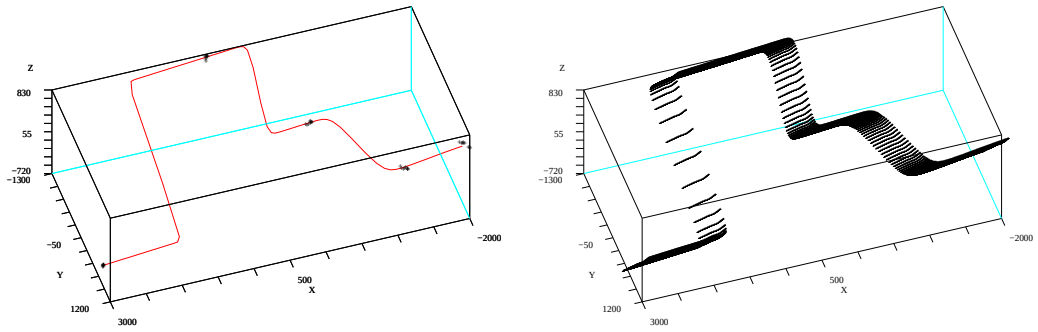


Figure 4: Evolution of the percentage of information represented with respect to the model dimension (blue : PCA model, green: AAR model).

that $d = 1$ is a convenient choice. The projection of the corresponding manifold in the linear subspace spanned by the first three PCA axes is represented Figure 5(a) where it is superposed to the scatterplot projection. Modelling this scatterplot by a two-dimensional manifold could also be justified since the image database is generated by rotating the object in two orthogonal directions. The projection of the two-dimensional manifold estimated and sampled is presented on Figure 5(b).
It is worth remarking that the principal variable Y^1 associated to the one-dimensional AAR model has a simple interpration. It corresponds to the rotation with respect to the elevation angle. As an illustration, we simulate uniform realizations of this variable and represent the corresponding images obtained with the one-dimensional AAR model (Figure 6). The variable Y^2 is not so easily interpretable. For this reason, the one-dimensional AAR model should be prefered.

6 Conclusion and further work.

As a conclusion, AA models offer a nice theoretical framework to the generalization of PCA. Moreover, they benefit from a simple implementation thanks to an interative algorithm. Its behaviour has been illustrated by building AAR models on simulated and real data from



(a) (b)

Figure 5: Projections in the subspace spanned by the first three PCA axes : (a) the one-dimensional manifold estimated and superposed to the scatterplot, (b) the two-dimensional manifold estimated and sampled.

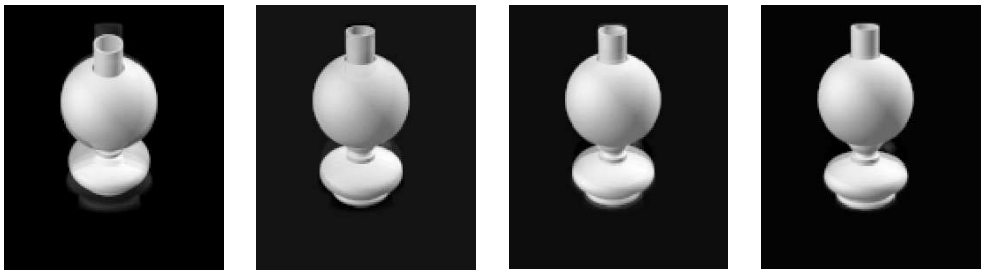


Figure 6: Simulation of 4 images with the one-dimensional AAR model. The variable Y^1 is simulated uniformly on the interval $[\min_i Y_i^1, \max_i Y_i^1]$.

image analysis. We also plan to compare PCA and LAA models in real situations. From a theoretical point of view, it would be also interesting to establish the asymptotic properties of the estimates (3) and (4) in order to propose some tests.

References

[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, Institut National Polytechnique de Grenoble (in french).

[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.

[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] T. Hastie, R. Tibshinari & J. Friedman (2001). *The elements of statistical learning*, Springer Series in Statistics, Springer.

[21] H. Hotelling (1933). "Analysis of a complex of statistical variables into principal components", *Journal of Educational Psychology*, 24, p. 417–441.

[22] P.J. Huber (1985). "Projection Pursuit". *The Annals of Statistics*, 13(2), p. 435–475.

[23] I. Jolliffe (1986). *Principal Component Analysis*, Springer-Verlag, New York.

[24] M.C. Jones & R. Sibson (1987). "What is projection pursuit?", *Journal of the Royal Statistical Society, Ser. A*, 150, p. 1–36.

[25] S. Klinke & J. Grassmann (2000). "Projection pursuit regression", *Wiley Series in Probability and Statistics*, p. 471–496.

[26] L. Lebart (2000). "*Contiguity analysis and classification*", Data Analysis, Gaul W., Opitz O., Schader M., (eds), Springer, Berlin, p. 233–244.

[27] B. Moghaddam & A. Pentland (1997). "Probabilistic visual learning for object representation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19 (7), p. 696–710.

[28] 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.

[29] 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.

[30] M.E. Tipping & C.M. Bishop (1999). "Probabilistic principal component analysis", *Journal of the Royal Statistical Society, Ser. B*, 61(3), p. 611–622.

[31] 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.

Contents

1 Introduction to auto-associative models. 3

2 Auto-associative models. 4
2.1 Definitions 5
2.2 Construction of auto-associative models 6

3 Two particular auto-associative models. 9
3.1 Linear auto-associative models 9
3.2 Auto-associative regression models 12

4 Implementation. 12
4.1 Estimation of the regression function 13
4.2 Computation of principal directions 13

5 Examples. 14
5.1 First example on simulated data 14
5.2 Second example on simulated data 15
5.3 Example in image analysis 17

6 Conclusion and further work. 18



Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399