

A priori spatiaux et analyse factorielle de séquences scintigraphiques

Marc FILIPPI¹, Eric MOISAN¹ Michel DESVIGNES¹, Catherine GHEZZI², Pascale PERRET², Daniel FAGRET²

¹GIPSA Lab

11 rue des Mathématiques, 38400 Saint Martin d'Hères, France

²INSERM, U1039, Radiopharmaceutiques Biocliniques, faculté de Médecine de Grenoble, France

Marc.Filippi@gipsa-lab.fr, Eric.Moisan@gipsa-lab.fr, Michel.Desvignes@gipsa-lab.fr,
Catherine.Ghezzi@ujf-grenoble.fr, Pascale.Perret@ujf-grenoble.fr,
dfagret@chu-grenoble.fr

Résumé – L'objectif de ces travaux est de séparer des mélanges de signaux provenant de différents organes dans des séquences scintigraphiques. L'approche classique utilise un modèle linéaire ne prenant pas en compte les a priori spatiaux. Nous proposons d'introduire des informations spatiales sous forme de contraintes d'absence au sein de cette méthode. Nous introduisons aussi une nouvelle procédure d'initialisation de cette analyse. Ces modifications améliorent qualitativement et quantitativement les résultats obtenus sur des séquences synthétiques et réelles.

Abstract – The aim of this work is to separate signals from different organs in scintigraphic sequences. The classical approach use a linear model which does not take into account spatial information. We propose to add spatial information to this method in the form of absence constraints. We also add a new initialization procedure. These modifications improve significantly both qualitative and quantitative results on synthetic and real sequences.

1 Introduction

La résistance à l'insuline peut être la cause de maladies cardiovasculaires et d'un diabète de type 2. Pour évaluer l'indice d'insulino-résistance d'un individu, on injecte au patient un traceur du glucose par intra-veineuse. Le parcours et la concentration du traceur sont suivis et enregistrés par une gamma-caméra qui comptabilise les rayons gamma et produit en sortie une séquence d'images 2D. Chaque image est donc la projection de l'activité du traceur dans les organes (figure 1 et 2). Sur un même pixel, plusieurs compartiments physiologiques (ventricule gauche (VG), ventricule droit (VD), myocarde, ...) peuvent donc se superposer. Pour évaluer l'indice d'insulino-résistance du patient, les courbes de concentration du traceur dans le VG et le myocarde sont comparées. Le problème est donc de retrouver les courbes du traceur et de segmenter chaque organe (VD, VG, myocarde, poumons, ...). La difficulté essentielle tient à un RSB très faible des signaux (figure 2).

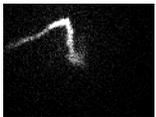


FIGURE 1 – Arrivée du traceur dans le ventricule droit, $t=4s$

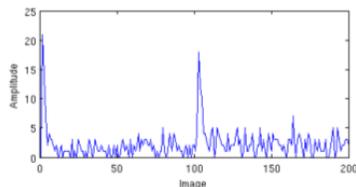


FIGURE 2 – Signal d'un pixel situé au niveau du ventricule droit

En imagerie médicale, l'analyse factorielle est une méthode

utilisée pour séparer les sources physiologiques. La méthode classique [1], a connu plusieurs variations, comme l'utilisation d'une nouvelle mesure [4], ou bien une résolution passant par une optimisation avec des moindres carrés pénalisés [5]. Ce type d'analyse, tout comme les autres méthodes classiques de séparation de sources, ne prend pas en compte les informations spatiales a priori. Par exemple, le signal du foie ne peut pas se trouver dans un pixel du VG. Nous proposons d'ajouter des a priori spatiaux à l'analyse factorielle afin de faciliter la résolution du problème. L'initialisation des facteurs est également modifiée.

2 L'Analyse Factorielle classique

2.1 Modèle

L'analyse factorielle ([1]) est une méthode de réduction de dimension et de séparation de sources. Elle suppose que les données à analyser sont des combinaisons linéaires d'un petit nombre de signaux appelés facteurs : $X = AF + \varepsilon$. La matrice $X \in R^{N \times P}$ est la séquence d'images à analyser où N est le nombre de pixels et P le nombre d'images. Les K images facteur sont représentées dans $A \in R^{N \times K}$ où K est le nombre de facteurs, et $F \in R^{K \times P}$ est la matrice des facteurs. Le bruit est représenté par $\varepsilon \in R^{N \times P}$. Chaque pixel i est ainsi une combinaison linéaire des facteurs F_k dont les poids sont les $A_{i,k}$ qui forment les images facteur. Cette factorisation doit avoir une signification physique ou physiologique (un facteur est le signal

provenant d'un organe).

2.2 Non unicité et contraintes

La décomposition de X en une matrice F des facteurs et une matrice A des images facteur n'est pas unique. Il faut ajouter des contraintes aux matrices A et F . Des normalisations sont donc imposées aux matrices A et F afin de restreindre l'espace des solutions : $\sum_{k=1}^K A_{i,k} = 1$ ($\forall i$) et $\sum_{t=1}^P F_{k,t} = 1$ ($\forall k$). Les données subissent la même normalisation que les facteurs au début de l'analyse. Pour que l'analyse factorielle donne des résultats qui aient ici une signification physique, on utilise des contraintes de non négativité sur les coefficients des matrices A et F ($\forall(i, k) A_{i,k} \geq 0$ et $\forall(k, t) F_{k,t} \geq 0$).

2.3 Interprétation géométrique

Le modèle linéaire additif doté de la normalisation et des contraintes de non négativité a une interprétation géométrique très forte [2]. En l'absence de bruit, les données se retrouvent à l'intérieur d'un K -simplexe dont les sommets sont les facteurs (figure 3). Le signal d'un pixel est en fait un barycentre des facteurs, dont les poids sont décrits dans l'image facteur. L'analyse factorielle recherche un ensemble de K facteurs qui englobe les données afin de satisfaire la contrainte $A \geq 0$. La contrainte $F \geq 0$ impose au facteur de ne pas être trop éloigné du centre des données, pour éviter d'avoir un terme $F_{k,t}$ négatif. Les contraintes ne garantissent pas l'unicité de la solution.

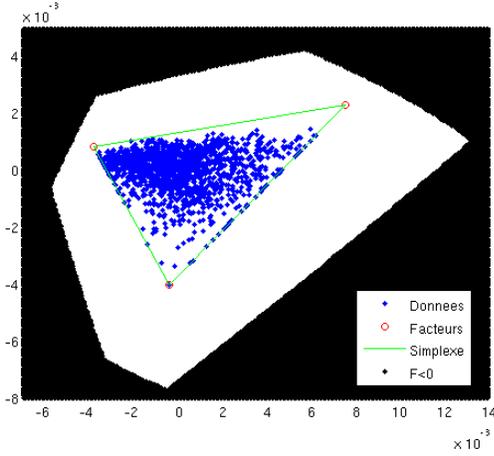


FIGURE 3 – Le simplexe englobant les données ($K=3$). La zone noire est la zone où les signaux comportent au moins un coefficient négatif.

2.4 Influence du bruit

En pratique, les séquences d'images scintigraphiques présentent un bruit poissonien. Les données ne sont donc plus nécessairement contenues dans le K -simplexe lié aux facteurs, et

ne vivent plus dans un sous espace affine de dimension $K - 1$. En particulier les données qui étaient proches des bords du K -simplexe, c'est à dire celles qui avaient au moins un poids très proche de 0, risquent de sortir du K -simplexe.

2.5 Algorithme

L'analyse factorielle se décompose en deux étapes [2]. La première étape est une analyse orthogonale (analyse en composantes principales) qui réduit le nombre de dimensions des données. Idéalement, ces données vivent en dimension $K - 1$. En pratique le bruit ne nous permet que d'approcher cet espace de dimension $K - 1$. Les données sont tout d'abord centrées et réduites : $Y_i = \frac{X_i - \bar{X}}{\|X_i - \bar{X}\|_1}$ où $\bar{X} = \frac{1}{N} \sum_{n=1}^N X_n$ est le centre des données, et X_i le signal du i^{eme} pixel. Les données centrées réduites sont ensuite projetées dans le sous-espace affine \mathcal{S} de centre \bar{X} et parallèle à l'espace engendré par les $K - 1$ plus grandes valeurs propres de la matrice d'auto-corrélation de Y . On néglige la projection dans \mathcal{S} du bruit ϵ compte tenu de la faible dimension de cet espace. Les projections des données Y_i dans \mathcal{S} sont notées $V_i \in R^{1 \times K-1}$. Une alternative à l'ACP, est l'analyse factorielle des correspondances ([4]). On note :

$$U = \begin{pmatrix} \bar{X} \\ U_1 \\ \vdots \\ U_{K-1} \end{pmatrix} \in R^{K \times P} \text{ et } V = \begin{pmatrix} 1 & V_1 \\ \vdots & \vdots \\ 1 & V_N \end{pmatrix} \in R^{N \times K}.$$

Dans la seconde étape, on se place dans le sous-espace affine \mathcal{S} pour identifier les K facteurs de manière itérative. Dans un premier temps, on initialise les facteurs de façon grossière. Les images facteur A sont ensuite calculées, puis dans un second temps les matrices A et F sont corrigées grâce à une boucle alternée, afin de satisfaire les contraintes de non-négativité. Cette étape est appelée l'analyse oblique, car les facteurs obtenus en sortie ne sont généralement pas orthogonaux.

Les facteurs recherchés sont les sommets du K -simplexe englobant les données dans \mathcal{S} . On note $T_i \in R^{1 \times K-1}$ la projec-

tion du i^{eme} facteur dans \mathcal{S} et $T = \begin{pmatrix} 1 & T_1 \\ \vdots & \vdots \\ 1 & T_K \end{pmatrix} \in R^{K \times K}$

la matrice des facteurs projetés. L'initialisation proposée par Bazin et al ([3]), ne permet pas de retrouver les sommets du K -simplexe. Nous proposons ici une nouvelle initialisation qui comble cette lacune (algorithme 1).

$T_1 = V_{i_1}$ avec $i_1 = \operatorname{argmax}_j \|V_j\|$ (Le pixel le plus éloigné du centre des données projetées);

$T_2 = V_{i_2}$ avec $i_2 = \operatorname{argmax}_j \|V_j - V_{i_1}\|$;

pour $k = 3$ à K **faire**

$T_k = V_{i_k}$ avec V_{i_k} la projection du pixel le plus éloigné du sous-espace affine formé par $\{T_i\}_{i=1..k-1}$.

fin

Algorithme 1 : Nouvelle initialisation des K facteurs

Une fois les projections des facteurs initialisées, les matrices A et F s'obtiennent par ces relations : $F = T \times U$ et $A = V \times T^{-1}$.

La boucle de corrections alternées intervient ensuite afin de rendre A et F conformes aux contraintes.

```

tant que  $Arret \neq VRAI$  faire
   $\forall (i, k)$  tel que  $A_{i,k} < 0$ , faire  $A_{i,k} = \alpha \times A_{i,k}$  ;
  Renormaliser  $A$  ;
  Recalculer  $T$  et  $F$  ;
   $\forall (k, t)$  tel que  $F_{k,t} < 0$ , faire
   $F_{k,t} = \beta \times F_{k,t} + (1 - \beta)\bar{X}$  ;
  Renormaliser  $F$  ;
  Recalculer  $A$  avec la relation  $A = V \times T^{-1}$ 

```

fin

Algorithme 2 : Boucle de corrections alternées

Les paramètres α et β sont choisis dans l'intervalle $[0, 1[$. Le critère d'arrêt est satisfait lorsque le nombre de coefficients négatifs des matrices A et F devient suffisamment petit.

3 Ajout d'a priori spatial

3.1 Contraintes d'absence

Même après la normalisation et l'ajout des contraintes de non négativité, une grande liberté est laissée aux facteurs et aux images facteur (voir figure 3). La solution fournie par l'analyse factorielle peut rester lointaine de la solution ayant le sens physiologique recherché. Pour s'approcher de cette solution idéale, des a priori peuvent être ajoutés à l'analyse factorielle. Dans l'application médicale présentée, on peut déterminer grossièrement les positions des compartiments physiologiques avec les premières images de la séquence. Cette information peut être ajoutée comme contrainte à l'analyse factorielle, en imposant la nullité de certains coefficients des images facteur, ce qui traduit la certitude de l'absence d'un compartiment physiologique sur certains pixels. Par exemple, au niveau du ventricule gauche, la contribution du foie doit être nulle, et réciproquement.

Les nouvelles contraintes d'absence sont injectées au niveau de l'analyse oblique. Les coefficients de A qui ne sont pas suffisamment proches de 0 alors que la contrainte l'impose sont réduits (voir algorithme 3).

```

tant que  $Arret \neq VRAI$  faire
   $\forall (i, k)$  tel que  $A_{i,k} < 0$ , faire  $A_{i,k} = \alpha \times A_{i,k}$  ;
   $\forall (i, k)$  tel que  $A_{i,k} \neq 0$  ET  $C_{i,k} = 0$ , faire
   $A_{i,k} = \gamma \times A_{i,k}$  ;
  Renormaliser  $A$  ;
  :

```

fin

Algorithme 3 : Nouvelle boucle de corrections alternées de l'analyse oblique

Le paramètre γ est choisi dans l'intervalle $[0, 1[$. Le nouveau

critère d'arrêt de la boucle de corrections alternées doit vérifier que les coefficients a priori nuls de l'image facteur le sont bien. La matrice $C \in R^{N \times K}$ est la matrice d'a priori. Elle contient des 0 là où l'on souhaite que la matrice des images facteur A en contienne. L'utilisation de la matrice C soulève une difficulté : il faut être capable d'associer chaque colonne de A à un des compartiments physiologiques recherchés afin que les corrections portent sur les bonnes images facteur. Il est par exemple possible d'ajouter une information a priori sur l'allure des facteurs (voir section 4.2).

4 Résultats

4.1 Présentation des jeux de données

Les performances de l'analyse factorielle, avec et sans a priori sont évaluées sur trois types de jeux de données. Dans le premier, les facteurs et images facteur sont générées aléatoirement. Ce jeu de données permet de tester les méthodes sur une grande diversité de configurations, avec une vérité-terrain à disposition. Les figures 4 et 5 montrent un exemple de facteur et d'image facteur générés par l'algorithme. La séquence d'images finale est le résultat de réalisations de lois de Poisson dont le paramètre est proportionnel au produit des images facteur et des facteurs générés aléatoirement.

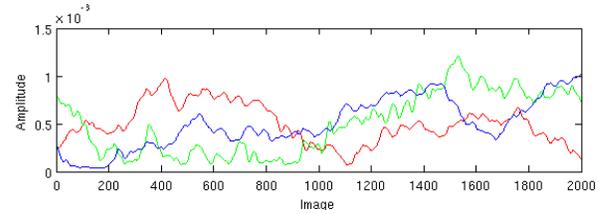


FIGURE 4 – Facteurs générés aléatoirement

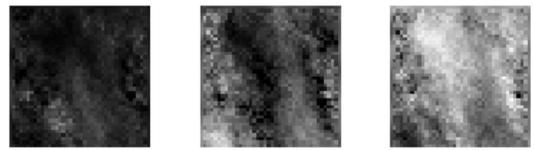


FIGURE 5 – Images facteur générées aléatoirement

Le deuxième jeu de données est un jeu créé grâce à une modélisation de la dilution du traceur dans l'organisme via le sang. Cela simule grossièrement la circulation du traceur dans l'organisme, ce qui permet d'obtenir des données similaires aux données réelles, qui constituent le troisième jeu. Dans ce jeu, les compartiments physiologiques (ventricule gauche, foie, ...) sont uniformes, et n'ont pas de variations internes. Cette modélisation permet d'avoir à disposition une vérité-terrain sur des séquences d'images proches des données réelles.

Le troisième jeu de données contient les données réelles acquises par gamma-caméra pour l'évaluation de l'indice d'insulino-résistance. Les séquences d'images contiennent l'arrivée du sang riche en traceur et sa dilution dans l'organisme (figures 1 et 2). Les images représentent l'activité du traceur à 0.5Hz. La particularité de ces images est que le signal y est extrêmement faible, sauf pendant des injections de bolus de traceur

4.2 Performances

Pour calculer les performances des algorithmes, il a été choisi $\alpha = 0.2$, $\beta = 0.7$ et $\gamma = 0.2$.

Le premier jeu de données est utilisé en générant des séquences de 1000 images de 5400 pixels à partir de 5 facteurs. La connaissance des facteurs simulés permet de réordonner les colonnes de la matrice C conformément à celle de A . Pour le second jeu de données, le même protocole d'évaluation et les mêmes paramètres sont utilisés. Les compartiments injectés dans le modèle pour l'évaluation sont le foie, le VD, le VG, les poumons et la veine cave supérieure droite. L'étude porte sur une séquence de 2000 images de 2075 pixels. Les EQM des évaluations des facteurs et images facteur trouvés pour les jeux de données 1 et 2 avec l'analyse factorielle avec et sans injection d'a priori sont répertoriés dans la table 1. L'a priori apporte une nette amélioration de l'estimation des images facteur, ainsi que des facteurs.

RSB		2	3	5	8	20
Facteur Jeu 1	Sans	33.9	22.1	12.8	11.1	10.0
Facteur Jeu 1	Avec	28.3	18.1	9.3	8.2	8.9
Image Jeu 1	Sans	57.5	44.8	33.0	28.5	23.6
Image Jeu 1	Avec	52.4	35.5	22.0	19.7	20.6
Facteur Jeu 2	Sans	25.9	15.0	3.5	2.4	0.8
Facteur Jeu 2	Avec	23.7	10.8	1.4	1.0	0.6
Image Jeu 2	Sans	61.30	52.3	26.2	21.1	12.3
Image Jeu 2	Avec	51.9	33.6	17.3	14.7	11.3

TABLE 1 – EQM (en %) normalisée et moyennée sur 100 essais des facteurs et images facteur sur le premier et le second jeu de données, avec et sans a priori ($K=5$)

Enfin, pour les données réelles, une étape de pré-traitement a été effectuée afin de se focaliser sur la zone cardiaque du patient. Quatre compartiments physiologiques ont été recherchés : le VG, le VD, les poumons et le myocarde. L'a priori spatial sont les zones grossières où chaque compartiment ne peut pas se trouver. Pour associer chaque facteur à son compartiment physiologique, on cherche l'abscisse du maximum de chaque facteur trouvée à l'initialisation. On sait que le sang arrive tout d'abord par le ventricule droit, puis se dirige vers les poumons, puis vers le ventricule gauche et enfin vers le myocarde. Ainsi le facteur qui aura son maximum le plus tôt sera apparenté au ventricule droit et ainsi de suite. Les résultats obtenus sont répertoriés figure 6. Avec l'ajout d'a priori, les

images facteur des compartiments sont beaucoup plus nettes. Les facteurs sont qualitativement mieux séparés ici avec l'ajout d'a priori spatial.

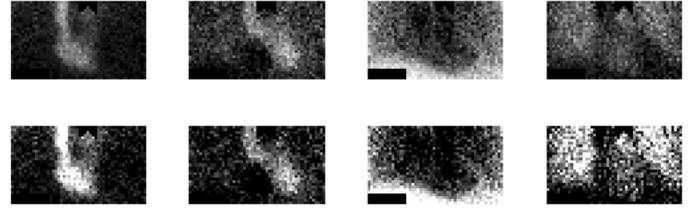


FIGURE 6 – Images facteur obtenues avec les données réelles. En haut : Sans a priori. En bas : Avec a priori. De gauche à droite : VD, VG, myocarde, et poumons

5 Conclusion

Les données synthétiques montrent que l'ajout d'a priori apporte une amélioration dans l'évaluation des facteurs et des images facteur. En cas de bruit très élevé (RSB=2 et RSB=3), les erreurs d'estimation restent toutefois importantes. La nouvelle procédure d'initialisation et l'introduction d'a priori spatial améliorent les résultats sur les données réelles en séparant mieux les organes. Cependant le signal du myocarde n'est pas encore exploitable, car son signal est très faible et sa contribution est difficile à distinguer dans le bruit. Nos perspectives sont d'améliorer la gestion conjointe des contraintes qui conditionne fortement la convergence de la méthode.

Références

- [1] Barber D C. *The use of principal components in the quantitative analysis of gamma camera dynamic studies*. Phys. Med. Biol. 25 283-92, 1980.
- [2] Di Paola R, Bazin J P, Aubry F, Aurengo A, Cavailloles F, Herry J Y and Kahn E. *Handling of dynamic sequences in nuclear medicine*. IEEE Trans. Nucl. Sci. NS-29 1310-21, 1982.
- [3] Bazin J P, Di Paola R. *Analysis applications in dynamic function studies*. 3rd World Congress of the Federation of Nuclear Medicine and Biology, Paris, 1982.
- [4] Benali H, Buvat I, Frouin F, Bazin J P et Di Paola R. *A statistical model for the determination of the optimal metric in factor analysis of medical image sequences*. Phys. Med. Biol. 38 1065-80, 1993.
- [5] Sitek A, Gullberg GT, Huesman RH. *Correction for ambiguous solutions in factor analysis using a penalized least square objective*. IEEE Trans. Med. Ima. 21 216-25, 2002.