SEPTIEME COLLOQUE SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS

NICE du 28 MAI au 2 JUIN 1979

APPLICATIONS DE L'ANALYSE FACTORIELLE

M. H. DEBART

CIT-ALCATEL 1 AVENUE ARISTIDE BRIAND ARCUEIL

RESUME

L'analyse factorielle est une méthode de calcul générale qui permet de retrouver la structure aléatoire d'un ensemble de variables gaussiennes lorsque on a mesuré, imparfaitement, la matrice de covariance. Le principe consiste à modeler cette matrice de façon à retrouver une matrice possédant la propriété cherchée : c'est-à-dire la présence, dans les variables aléatoires étudiées, d'un certain nombre de facteurs communs.

On en trouve tout naturellement l'application dans la reconstitution des ondes frappant un certain nombre de capteurs en acoustique sous-marine et ainsi à la détection des signaux dans le bruit.

SUMMARY

Factor analysis is a general method of computation whose purpose consists of reconstructing the random pattern of a set of gaussian variables, whose covariance matrix has been experimentally evaluated. The process is a modification of this matrix so as to build a new matrix (as "close" as possible to the former one) having the desired features : i.e. showing the presence, in the set of random variables on study, of a number of common factors.

Applications of this method are found, of course, in reconstructing the shape of narrow-band waves hitting a set of sensors, in underwater acoustics, and, of the other hand, in detecting narrow-band signals affected by stationary white noise.

1. ESTIMATION D'UNE MATRICE DE COVARIANCE

a) Matrice réelle

On dispose de p variables aléatoires gaussiennes $X_1 \ldots X_p$ et on veut estimer leur matrice de covariance. Pour cela, on dispose de N observations indépendantes sur l'ensemble des variables.

La variable X_i donne N réalisations $x_{i1} \ldots x_{iN}$ et la variable x_j donne N réalisations $x_{j1} \ldots x_{jN}$.

Le terme (i,j) de la matrice de covariance, soit

$$\overline{(X_i - \overline{X}_i)(X_j - \overline{X}_j)}$$

est estimé par la somme

$$S_{ij} = \frac{1}{N-1} \Sigma (x_{im} - \overline{X}_i)(x_{jm} - \overline{X}_j) \qquad (N-1 = n)$$

L'ensemble des estimations forme une matrice $S = \{S_{ij}\}$ symétrique et définie positive.

Si on appelle Σ la matrice de covariance, une distribution de probabilité peut être affectée à la matrice S. Elle est connue sous le nom de distribution de WISHART (voir réf. 1 p 154).

La densité de probabilité de la matrice S est de la forme :

$$m(S) = \frac{|S|^{\frac{n-p-1}{2}} e^{-\frac{1}{2} Tr\{\Sigma^{-1}S\}}}{2^{\frac{np}{2}} \pi^{\frac{p(p-1)}{4}} \prod_1^p \Gamma\{\frac{n+1-i}{2}\} |\Sigma|^{\frac{n}{2}}}$$

On peut déduire de cette densité une fonction de vraisemblance attachée à une matrice S mesurée, à partir de laquelle on infère une matrice de covariance Σ. Cette fonction n'est définie qu'à l'addition près d'une fonction des observations.

On choisit :

$$F = \log|\Sigma| + Tr|\Sigma^{-1} S| - Log|S| - p$$

qui a en particulier l'avantage d'être nulle pour une mesure parfaite ($S \equiv \Sigma$).

Le problème général de transformation de la matrice observée S peut alors s'exprimer ainsi :

- trouver une matrice Σ^* obéissant à un certain nombre de contraintes et telles que la fonction de vraisemblance correspondant à cette matrice soit minimale.

2. ANALYSE FACTORIELLE : RECHERCHE DE LA SOLUTION OPTIMALE PAR ITERATION

On veut reconnaître si les mesures faites sur l'ensemble des variables aléatoires permettent de les décomposer en sommes de facteurs communs ; chaque mesure faite sur une variable aléatoire de l'ensemble est supposée affectée d'un bruit indépendant.

S'il existe K facteurs communs, une variable aléatoire X_i du système peut se représenter par :

$$X_i = \lambda_{i1}f_1 + \lambda_{i2}f_2 + \lambda_{ik}f_k + e_i$$

Les $f_i \ldots f_k$ sont les facteurs communs qu'on ramène à des variables aléatoires normales. Les $\lambda_{i1} \ldots \lambda_{ik}$ sont les coefficients de pondération. Les variables aléatoires $f_i \ldots f_k$ et e_i (bruit de mesure) sont indépendantes entre elles.

Si cette hypothèse est vérifiée, la matrice de corrélation Σ prend une forme particulière. Si on appelle Λ la matrice des coefficients de pondération et si

$\overline{e_i^2} = \psi_i$

$$\Lambda = \begin{pmatrix} \lambda_{i1} & \cdots & \lambda_{ik} \\ \lambda_{p1} & \cdots & \lambda_{pk} \end{pmatrix}$$ est une matrice à p lignes et k colonnes

$$\psi = \begin{pmatrix} \psi_1 & \cdots & 0 \\ \cdots & \cdots & \psi_p \end{pmatrix}$$ est une matrice diagonale

On peut écrire alors :

$$\Sigma = \Lambda\Lambda^T + \psi \qquad (1)$$

Pour chercher s'il existe K facteurs communs, on doit donc rechercher la matrice la plus proche de S (au sens défini précédemment et possédant la propriété (1).

Ce calcul se fait par itération ; on se donne une matrice diagonale d'essai $\begin{pmatrix} \psi_{1_0} & \\ & \psi_{p_0} \end{pmatrix}$ et on recherche la matrice Λ qui minimise la fonction de vraisemblance.

A l'aide des dérivées $\frac{\partial F}{\partial \lambda}$, $\frac{\partial F}{\partial \psi}$, $\frac{\partial^2 F}{\partial \lambda^2}$, $\frac{\partial}{\partial \psi^2}$, on définit une nouvelle matrice diagonale $\begin{pmatrix} \psi_{1_1} & \\ & \psi_{p_1} \end{pmatrix}$ améliorée, et on recommence le calcul jusqu'à converger suffisamment.

1) Recherche de la matrice Λ quand ψ est donné.

Il faut remarquer que la matrice Λ n'est pas définie de façon unique. En effet, si on a trouvé une solution Λ_0 et si m est une matrice de rotation (telle que $mm^T = I$) la matrice Λm^T est aussi solution car $\Lambda m^T \Lambda^T m = \Lambda\Lambda^T$.
Il faut donc pour la rechercher, lui imposer une condition supplémentaire.
Il est commode de poser cette condition sous la forme :

$$\Delta = \Lambda^T \psi^{-1} \Lambda \text{ est diagonale}$$

Le calcul matriciel montre que les dérivées $\frac{\partial F}{\partial \Lambda}$, $\frac{\partial F}{\partial \psi}$ sont respectivement égales à

$$\frac{\partial F}{\partial \Lambda} = 2\Sigma^{-1}\{\Sigma - S\}\Sigma^{-1}\Lambda$$

$$\frac{\partial F}{\partial \psi} = Diag\{\Sigma^{-1}(\Sigma - S)\Sigma^{-1}\}$$

On recherche donc une matrice p x k : Λ satisfaisant aux conditions :

1) $$\frac{\partial F}{\partial \lambda} = 0 \quad \text{soit } (\Sigma - S)\Sigma^{-1}\Lambda = 0$$

2) $$\Delta = \Lambda^T \psi^{-1} \Lambda \text{ est diagonale.}$$

Ces deux conditions peuvent être combinées sous la forme :

$$(\Sigma - S)\psi^{-1}\Lambda(I + \Delta)^{-1} = 0$$

Si I désigne la matrice unité. Ou encore :

$$S\psi^{-1}\Lambda = \Lambda(I + \Delta)$$

Cette relation se met sous une forme directement utilisable quand on la multiplie à gauche par $\psi^{-\frac{1}{2}}$:

$$\{\psi^{-\frac{1}{2}} S\psi^{-\frac{1}{2}}\}\{\psi^{-\frac{1}{2}}\Lambda\} = \{\psi^{-\frac{1}{2}}\Lambda\}\{I + \Delta\}$$

Comme la matrice $I + \Delta$ est diagonale, cette relation montre que la matrice $\{\psi^{-\frac{1}{2}}\Lambda\}$ a pour colonnes les vecteurs propres de la matrice :

$$\{\psi^{-\frac{1}{2}} S\psi^{-\frac{1}{2}}\} = S^*$$

Les éléments diagonaux de $I + \Delta$ sont les valeurs propres associées. On peut montrer que ce sont les k plus grandes ; on les appelle :

$$\theta_1 > \theta_2 \dots > \theta_k$$

et on les suppose plus grandes que 1.

Ces vecteurs propres associés sont $\omega_1 \dots \omega_k$

On les norme à l'unité : $|\omega_1|^2 = \dots = |\omega_k|^2 = 1$.

La matrice diagonale formée à partir de $\theta_1 \dots \theta_k$ sera appelée Θ. La matrice formée par la réunion $(\omega_1 \dots \omega_k)$ sera appelée Ω

$\Omega^T\Omega = I_k$ car les vecteurs propres sont orthonormaux. L'équation satisfaite par Λ a pour solution :

$$\Lambda = \psi^{\frac{1}{2}}\Omega(\Theta - I)^{\frac{1}{2}}$$

Si on est parti d'un vecteur d'essai ψ_o, les opérations définies ci-dessous, c'est-à-dire :

- calcul de $S^* = \psi_o^{-\frac{1}{2}} S\psi_o^{-\frac{1}{2}}$
- calcul de Θ, Ω à partir des vecteurs propres et des valeurs propres en nombre k) de S^*
- calcul de $\Lambda_o = \psi_o^{\frac{1}{2}}\Omega\,(\Theta - I)^{\frac{1}{2}}$

permettent de définir le vecteur Λ_o associé.

Il faut ensuite améliorer le vecteur Λ_o en changeant de matrice ψ de façon à diminuer la fonction de vraisemblance.

On utilise pour cela une méthode du type NEWTON-RAPHSON en évaluant :

(a) le vecteur-colonne des dérivées premières de la fonction de vraisemblance : $G = \{\frac{\partial F}{\partial \psi i}\}$

(b) la matrice carrée des dérivées secondes :

$$\Phi = \{\frac{\partial^2 F}{\partial\psi_i \partial\psi_j}\}$$

Ce qui exige les calculs suivants :

(a) $$\frac{\partial F}{\partial\psi_i} = \frac{1}{\psi_i}\{\sum_1^k(\theta_r - 1)\,\omega_{ir}^2 + 1 - \frac{S_{ii}}{\psi_i}\}$$

(b) On forme la matrice :

$$\Gamma = \psi^{-\frac{1}{2}}\{I - \Omega\Omega^T\}\psi^{-\frac{1}{2}} \qquad (\text{éléments } \Gamma_{ij})$$

et la matrice $\Phi = \{\Gamma_{ij}^2\}$

On peut alors passer de la matrice originale $\{\psi_o\}$ à une nouvelle matrice $\{\psi_1\}$ par la transformation de NEWTON-RAPHSON :

$$\{\psi_1\} = \{I\}\{\Phi\}^{-1}\{G\}.k$$

Le facteur de correction k doit être estimé de façon arbitraire. Toute l'itération est ainsi définie. On fait la suite des calculs :

$$\psi_o \to \Lambda_o \to \psi_1 \to \Lambda_1 \to \psi_2 \to \Lambda_2 \text{ - etc...}$$

Le calcul est arrêté quand les éléments de G sont tous inférieurs en module à une quantité prédéterminée. On peut alors calculer la matrice Σ reconstituée :

$$\Sigma = \Lambda_n\Lambda_n^T + \psi_n$$

et la fonction de vraisemblance associée.
Ce calcul est évidemment beaucoup plus long si on n'a aucune idée du nombre des facteurs indépendants à rechercher ; il varie entre 1 et p-1.

2) Matrice hermitique

Le calcul précédent s'étend sans difficulté à une matrice de covariance hermitique définie positive dont les valeurs propres sont réelles et positives. Il suffit de remplacer la grandeur :

$$\{I - \Omega\Omega^T\}$$

par la grandeur :

$$\{I - \Omega\Omega^{T*}\}$$

et Γ_{ij}^2 par $\Gamma_{ij}\Gamma_{ij}^*$

3. APPLICATIONS PRATIQUES

1) Mesure angulaire à partir d'un ensemble de capteurs.

Un ensemble de capteurs, en écoute passive à bande étroite, recueille des signaux analytiques déphasés et des bruits indépendants. A partir de n capteurs, on peut mesurer une matrice hermitique de covariance : S. Si on lui applique la méthode d'analyse factorielle, on peut retrouver les signaux analytiques (jusqu'à n-1) incidents et les bruits indépendants associés.

Un exemple numérique sera donné plus loin.

2) Application à la détection

L'analyse factorielle peut être utile pour détecter un ou plusieurs signaux sinusoïdaux dans un bruit, stationnaire, associée à l'algorithme adaptatif de WIDROW.

a) Détection adaptative de WIDROW

Quand on reçoit un signal affecté de bruit blanc additif stationnaire, les échantillons disponibles à l'instant j forment un vecteur-colonne.

$X_i = \begin{pmatrix} X_{oj} \\ X_{nj} \end{pmatrix}$ qu'on veut filtrer linéairement, en formant le produit scalaire de ce vecteur pour un vecteur-poids:

$$g_j = X_j^T W = W^T X_j$$

$W = \begin{pmatrix} W_o \\ W_n \end{pmatrix}$ est ce produit scalaire.

La différence entre la réponse désirée du filtre et la sortie du même filtre est :

$$\varepsilon_j = d_j - X_j^T W = d_j - W^T X_j$$

On peut définir le filtre de façon à minimiser l'erreur quadratique moyenne :

$$E(\varepsilon_j^2) = E(d_j^2) - 2E(d_i X_j^T)W + W^T E\{X_j X_j^T\}W$$

Dans cette expression :

$$E(X_j X_j^T) = E\begin{pmatrix} X_{oj}X_{oj} & X_{oj}X_{1j}\dots \\ \dots\dots\dots\dots & X_{nj}X_{nj} \end{pmatrix} = R$$

est la matrice de covariance du signal affecté de bruit :

$$P = E\{d_j X_j\} = E \begin{pmatrix} d_j X_{oj} \\ d_j X_{ij} \end{pmatrix}$$

est le vecteur-colonne caractérisant la réponse désirée.
Ce filtre optimal qui minimise $E(\varepsilon_j^2)$ est défini par :

$$W^* = R^{-1}\,P.$$

(équation de WIENER-HOPF).

Quand on ignore tout de la matrice de covariance, on peut partir d'un vecteur-poids arbitraire et le corriger à l'aide de l'algorithme de WIDROW :

$$W_{j+1} = W_j + 2\mu\varepsilon_j X_j$$

formule dans laquelle μ représente un facteur arbitraire de correction.
En pratique, l'opération est très difficile quand on ignore tout de la matrice R ; si μ est trop élevé, il y a divergence (la théorie prévoit la valeur maximale autorisée), s'il est trop faible, la convergence est trop lente.

On rend la méthode efficace en partant d'une estimation assez bonne de la matrice R, ce qui peut se faire d'une façon simple en utilisant une méthode inspirée de l'analyse factorielle.

b) Estimation de la matrice de covariance

En opérant sur un certain nombre d'échantillons, on mesure une matrice de covariance S (n x n). On veut l'ajuster du mieux possible à une matrice de covariance Σ d'un signal sinusoïdal et d'un bruit blanc stationnaire :

$$\Sigma = \begin{vmatrix} a & b\cos\phi & \dots & b\cos n\phi \\ & a & b\cos\phi & \\ & & & a \end{vmatrix}$$

On peut construire au départ une matrice Σ_0 formé à l'aide d'estimations grossières des paramètres a, b, ϕ et évaluer les dérivées :

$$\frac{\partial F}{\partial a}, \frac{\partial F}{\partial b}, \frac{\partial T}{\partial \phi} \text{ (vecteur-colonne)}$$

$$\frac{\partial^2 F}{\partial a^2}, \frac{\partial^2 F}{\partial a \partial b} \dots \frac{\partial^2 F}{\partial \phi^2} \quad \text{(matrice)}$$

Pour appliquer ensuite la méthode de NEWTON-RAPHSON permettant d'améliorer la fonction de vraisemblance Γ ... et utiliser l'itération.
On rappelle que :

$$F = \mathrm{Log}|\Sigma| + \mathrm{Tr}\{S\Sigma^{-1}\} - \mathrm{Log}|S| - p$$

Les dérivées se calculent à partir des formules générales :

$$\frac{\partial \mathrm{Log}|\Sigma|}{\partial \sigma_{ii}} = \sigma^{ii}$$

$$\frac{\partial \mathrm{Log}|\Sigma|}{\partial \sigma_{ij}} = 2\sigma^{ij}$$

(σ^{ii} et σ^{ij} sont les mineurs de σ_{ii}, σ_{ij})

$$\frac{\partial \mathrm{Tr}(S\Sigma^{-1})}{\partial \sigma_{ii}} = -\{\Sigma^{-1}S\Sigma^{-1}\}_{ii}$$

$$\frac{\partial \mathrm{Tr}(S\Sigma^{-1})}{\partial \sigma_{ij}} = -2\{\Sigma^{-1}S\Sigma^{-1}\}_{ij} \qquad (i \neq j)$$

On calcule donc les différentes dérivées en (a,b,ϕ) en calculant d'abord l'inverse de Σ_0 : Σ_0^{-1} et on en déduit des valeurs améliorées de (a,b,ϕ), etc...

La méthode peut être utilisée à partir d'une matrice de taille peu élevée, d'où on tire un premier filtre de WIDROW et on augmente ensuite progressivement le nombre des échantillons pris en compte.

On peut l'appliquer également en supposant qu'il existe 2, 3 ... sinusoïdes simultanées. La conclusion sur le nombre de sinusoïdes présentes vient de l'examen de la fonction de vraisemblance.

4. APPLICATIONS NUMERIQUES

1) Ensemble de 4 capteurs

On a considéré le cas suivant : les capteurs alignés sont distants de $\frac{\lambda}{2}$. Ils reçoivent deux ondes planes, l'une à 30° ce qui correspond à un déphasage de 90° d'un capteur à l'autre et l'autre à 20° (déphasage de 62° d'un capteur à l'autre) ; les deux ondes sont de même énergie.

Les 4 capteurs sont affectés de bruit gaussiens indépendants. Les rapports signal/bruit sont respectivement de :

- 8 dB - 11 dB - 13 dB - 4 dB

pour les 4 capteurs.

On mesure une matrice de covariance sur 250 échantillons successifs. Cette matrice est :

$$\begin{vmatrix} 0{,}5256 & -0{,}0338-0{,}0369i & 0{,}0247+0{,}0179i & 0{,}0675+0{,}0603i \\ x & 1{,}0287 & 0{,}0597-0{,}1085i & -0{,}0002+0{,}0372i \\ x & x & 1{,}5372 & 0{,}1988-0{,}0695i \\ x & x & x & 1{,}0649 \end{vmatrix} = S$$

Le premier calcul partant d'une matrice d'essai :

$$\Psi_0 = \begin{pmatrix} 0{,}5 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1{,}5 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

donne des valeurs propres :

$\theta_1 = 1{,}2847$

$\theta_2 = 1{,}1211$

$\theta_3 = 0{,}9601$

$\theta_4 = 0{,}2038$

et des vecteurs propres associés ω_1, ω_2 (d'où λ_1, λ_2)

Les déphasages des composantes d'un capteur au suivant sont :

(λ_1)		(λ_2)	
	183°		97°
	115°		35°
	21°		271°

On poursuit l'itération à partir de ces résultats.
La seconde donne :

$\theta_1 = 1{,}4006$

$\theta_2 = 1{,}3236$

(les valeurs propres se sont rapprochées).

(λ_1)		(λ_2)	
	125°		79°
	110°		53°
	75°		117°

La troisième donne :

$$\theta_1 = 1{,}4201$$
$$\theta_2 = 1{,}4003$$

$$(\lambda_1)\quad \begin{matrix} 93° \\ 90° \\ 88° \end{matrix} \qquad\qquad (\lambda_2)\quad \begin{matrix} 67° \\ 60° \\ 64° \end{matrix}$$

On tend donc rapidement vers une évaluation très serrée de la réalité.

2) Détection

On est parti d'un signal sinusoïdal dont le déphasage, d'un échantillon à l'autre est de 79° affecté d'un bruit, tel que le rapport/bruit est 0 dB.
La mesure de la matrice de covariance stationnaire est faite sur 25 échantillons. Elle fournit :

$$S = \begin{vmatrix} 1 & 0{,}0906 & -0{,}3830 & -0{,}2990 \\ & 1 & 0{,}0906 & -0{,}3830 \\ & & 1 & 0{,}0906 \\ & & & 1 \end{vmatrix}$$

On part d'une matrice Σ correspondant à $\phi = 70°$.

Ces itérations successives fournissent :

$$\left\{ \begin{array}{l} \phi = 70° \\ \phi = 75° \\ \quad = 77° \\ \quad = 78°7 \end{array} \right.$$

5. CONCLUSION

On peut retenir de l'analyse factorielle une idée générale très féconde : la modification d'une matrice expérimentale de façon à l'amener à posséder des propriétés données, cette notion est très générale et déborde largement le cadre de l'analyse factorielle proprement dite ; on peut donc en envisager des applications.

La lourdeur relative des calculs numériques associés explique la lenteur de l'introduction de ces méthodes, mais cet argument disparaît avec la progression énorme des moyens de traitement qu'on constate actuellement.

REFERENCES

1) DN LAWLEY, AZ MAXWELL
Factor Analysis as a statistical Method
Butterworths Londres 1971

2) T.W ANDERSON
An introduction to multivariate statistical Analysis
WILEY NY 1958

3) B. WIDROW et al.
Adaptative Noise Cancelling
PIEEE (63) n° 12 de 1975
(p. 1692).