Modèles autorégressifs : du coefficient de réflexion à la géométrie Riemannienne de l'information

Autoregressive models: from reflection coefficient to Riemannian information geometry

par Frédéric BARBARESCO

THOMSON-CSF AIRSYS RD/RDTA
Unité de Développement Radar, Direction Technique Opérationnelle,
Service Algorithmes & Etudes Fonctionnelles, 7/9 rue des Mathurins, 92221 Bagneux Cedex,
E-mail : Frederic.F.B.Barbaresco@airsys.thomson-csf.com

résumé et mots clés

Nous mettons en évidence les liens étroits qui unissent l'analyse AR et l'analyse en sous-espaces. Nous mettons en relation, deux types d'algorithmes qui utilisent ces liens et possèdent les avantages communs d'être récursifs en ordre, et à chaque ordre d'être parallèlisables suivant le principe d'entrelacement des valeurs propres. Finalement, nous appliquons les concepts de géométrie Riemannienne de l'information, développés par Chentsov, dans le cas AR pour établir le calcul récursif de différentes distances paramétriques, possedant la propriété d'être invariantes par changement de paramétrisation : la métrique de Siegel, la géodésique minimale de Jensen ainsi qu'une expression récursive de la divergence de Kullback.

Modèles AR, coefficient de réflexion, analyse en sous-espaces propres, géométrie de Chentsov, métrique de Siegel, géodésique minimale, divergence de Kullback, géométrie de Finsler.

abstract and key words

We describe close connections between AR & eigen-subspaces analysis. We underline relations between two kinds of algorithms which exploit these connections and which have same both advantages to be order-based recursive and, at each order, to compute results in parallel, thanks to eigenvalues interlacing principle. Finally, we apply Riemannian Information geometry concepts, developed by Chentsov, for AR applications, to obtain recursive computation of new parametric distances, invariant by parameters changes : Siegel metric, Jensen's minimal geodesic, and a recursive formula for Kullback divergence.

AR models, reflection coefficient, eigen-subspaces analysis, Chentsov geometry, Siegel metric, minimal geodesic, Kullback divergence, Finsler geometry.

1. préambule

Elève du Professeur Picinbono à Supelec en 1991 en section automatique, je suis entré l'année suivante à THOMSON où l'on m'a sollicité pour approfondir plus particulièrement l'analyse Doppler radar. C'est à la relecture du cours du professeur Picinbono, qui a donné lieu plus tard à l'ouvrage [Pic95], que j'ai pu établir les résultats que j'expose ici. Les équations, que nous établirons dans la suite, ont été définies dans le cas complexe en nous plaçant dans l'hypothèse de signaux circulaires. Une généralisation au cas non circulaire serait à développer, et je renvoie le lecteur aux travaux du professeur Picinbono [Pic95, p. 18][Pic94].

La plupart des résultats établis sont issus d'une utilisation systématique de la structure blocs de la matrice de corrélation des données et de son inverse, paramétrées par les coefficients autorégressifs. La structure de l'inverse de la matrice de corrélation est démontrée dans [Pic95, p. 137] :

$$R_n^{-1} = \begin{bmatrix} \alpha_{n-1} & \alpha_{n-1} \cdot A_{n-1}^{+} \\ \alpha_{n-1} \cdot A_{n-1} & R_{n-1}^{-1} + \alpha_{n-1} \cdot A_{n-1} \cdot A_{n-1}^{+} \end{bmatrix} \quad (1)$$

et similairement

$$R_n = \begin{bmatrix} \alpha_{n-1}^{-1} + A_{n-1}^{+} \cdot R_{n-1} \cdot A_{n-1} & -A_{n-1}^{+} \cdot R_{n-1} \\ -R_{n-1} \cdot A_{n-1} & R_{n-1} \end{bmatrix} \quad (2)$$

avec $\alpha_n^{-1} = \left[1 - |\mu_n|^2\right] \cdot \alpha_{n-1}^{-1}$ et $A_n = \begin{bmatrix} A_{n-1} \\ 0 \end{bmatrix} + \mu_n \cdot \begin{bmatrix} A_{n-1}^{(-)} \\ 1 \end{bmatrix}$.

Nous utiliserons la notation de B. Picinbono [Pic95, p.130] : $A_n^{(-)} = J_n \cdot A_n^*$ avec J_n matrice anti-diagonale.

Nous établirons [Barb97] les liens étroits qui unissent l'analyse en sous-espaces propres et l'analyse autorégressive. En particulier, nous ferons le lien entre deux algorithmes qui possèdent une structure de mise en œuvre comparable basée sur le double principe : récursivité sur l'ordre de l'espace et à chaque ordre une parallélisation du calcul des valeurs propres. Les relations algébriques établies, dans ce contexte, permettront de proposer un algorithme rapide et robuste de calcul des éléments propres d'une matrice d'autocorrélation à partir des paramètres autorégressifs, d'interpréter géométriquement les équations et de donner une nouvelle définition du coefficient de réflexion. Enfin, nous montrerons la géométrie particulière induite par les coefficients de réflexion à l'intérieur du polydisque complexe $\{\mu = (\mu_1, \ldots, \mu_n)/|\mu_k| < 1, \ \ \forall k = 1, \ldots, n\}$, à partir des outils de la géométrie différentielle appliquée à la statistique. Un modèle statistique est spécifié par une famille de distributions de probabilité, usuellement décrites par un ensemble de paramètres continus appelé espace des paramètres. Cet espace possède des propriétés géométriques qui sont induites par le contenu local d'information et les structures des distributions. En 1945, Rao introduisit une métrique Riemannienne à partir de la matrice d'information de Fisher dans l'espace des paramètres d'une famille paramétrique de distributions de probabilité, et proposa une distance géodésique induite par cette métrique comme étant une mesure de dissimilarité entre deux distributions de probabilité. Ces travaux ont été étendus par Reed, Dawid et Amari. Atkinson et Mitchell [Atk81], et indépendamment Chentsov [Chen72] dans ses travaux théoriques, ont calculé les distances de Rao pour un certain nombre de distributions de probabilités. Enfin, récemment, une approche unifiée pour la construction de distance et de mesures de dissimilarité dans les espaces de probabilités a été développée par, Burbea [Burb82], Rao, Oller [Cal90] et Cuadras. Nous avons appliqué ces approches au cas de l'analyse autorégressive et trouvé de nouvelles distances paramétriques robustes.

2. analyse autorégressive et analyse en sous-espaces

L'analyse en sous-espace est un outil puissant utilisé en traitement du signal qui a souvent relégué l'analyse autorégressive au second plan. Nous allons montrer qu'en fait ces deux méthodes sont intimement liées l'une à l'autre et que l'analyse autorégressive fournit une méthode rapide et robuste de calcul des élèments propres d'une matrice Toeplitz d'autocorrélation. Nous soulignons le lien avec l'approche de Bunse-Gerstner [Bun95], les deux schémas numériques utilisant la notion d'entrelacement des valeurs propres. Les deux approches donnent lieu à des algorithmes récursifs sur l'ordre du modèle et qui à chaque étape, sont parallélisables. Dans le premier cas, on applique [Barb97a] en parallèle, à chaque ordre, une méthode dichotomique sur des intervalles distincts contigus, et dans le second, on emploie en parallèle, une méthode de bissection [Bun95] sur des arcs jointifs sur le cercle complexe unité.

La structure matricielle donnée par (1), nous permet de façon simple de montrer que vecteurs propres et valeurs propres de l'inverse de la matrice de corrélation R_n^{-1} se déduisent récursivement sur l'ordre à partir des paramètres autorégressifs. Grâce au théorème d'entrelacement des valeurs propres de matrices imbriquées (théorème de Wilkinson) que notre algorithme vérifie, les valeurs propres $\{\eta_k^{(n)}\}_{k=1,\ldots,n}$ de la matrice R_n^{-1} sont calculées récursivement à chaque ordre n de façon parallèle comme racine d'une fonction $F^n(\eta)$ définie par parties, et dont les bornes de définition sont simplement données par les valeurs propres à l'ordre précédent $\{\eta_k^{(n-1)}\}_{k=1,\ldots,n-1}$ (on prendra simplement comme majorant la trace calculée également récursivement et comme minorant zéro). Sur chaque intervalle la fonction $F^n(\eta)$ est strictement croissante et ne possède qu'une seule racine, ce qui permet le calcul parallèle des valeurs propres $\{\eta_k^{(n)}\}_{k=1,\ldots,n}$ par dichotomie sur chaque intervalle de définition $]\eta_{k-1}^{(n-1)}, \eta_k^{(n-1)}[$.

Les vecteurs propres $\{X_k^{(n)}\}_{k=1,\ldots,n}$ de R_n^{-1} se déduisent d'une seconde formule et sont renormalisés à chaque ordre n. Il suffit pour cela de poser les formules d'analyse en sous-espaces propres de R_n^{-1} :

$$U_{n-1}^{+} \cdot R_{n-1}^{-1} \cdot U_{n-1} = \Lambda_{n-1} = \mathrm{diag}\{\ldots, \eta_k^{(n-1)}, \ldots\} \quad (3)$$

avec

$$U_{n-1} = \left[X_1^{(n-1)} \ldots X_{n-1}^{(n-1)}\right]$$

et

$$U_{n-1}^{+} \cdot U_{n-1} = U_{n-1} \cdot U_{n-1}^{+} = I_{n-1}$$

L'application simple de la structure blocs définie en (1) permet de trouver le schéma numérique associé à partir des paramètres de l'analyse autorégressive α_{n-1}, A_{n-1}, qui fournit valeurs propres $\eta_k^{(n)}$ et vecteurs propres $X_k^{(n)}$ à chaque ordre n :

$$F^{(n)}\left(\eta_k^{(n)}\right) = \eta_k^{(n)} - \alpha_{n-1} + \alpha_{n-1} \cdot \eta_k^{(n)} \sum_{i=1}^{n-1} \frac{\left|A_{n-1}^{+} \cdot X_i^{(n-1)}\right|^2}{\left(\eta_i^{(n-1)} - \eta_k^{(n)}\right)} = 0$$

et

$$X_k^{(n)} = \begin{bmatrix} X_{k,1}^{(n)} \\ -\eta_k^{(n)} \cdot X_{k,1}^{(n)} \cdot U_{n-1} \cdot \left(\Lambda_{n-1} - \eta_k^{(n)} \cdot I_{n-1}\right)^{-1} \cdot U_{n-1}^{+} \cdot A_{n-1} \end{bmatrix} \tag{4}$$

Il apparaît de façon claire que les équations (4) permettent à partir de la connaissance de

$$\alpha_{n-1}, A_{n-1}, \quad \text{et} \quad \{\eta_k^{(n-1)}, X_k^{(n-1)}\}_{k=1,\ldots,n-1}$$

à l'ordre $n-1$, de calculer les nouveaux élèments propres à l'ordre n :

$$\{\eta_k^{(n)}, X_k^{(n)}\}_{k=1,\ldots,n}.$$

Ainsi, nous sommes ramenés à la recherche parallèle, par dichotomie, des n racines de la fonction $F^{(n)}(\eta)$, qui est définie par morceaux contigus $\bigcup_{i=1}^{n-2} \left]\eta_{n-i}^{(n-1)}, \eta_{n-i-1}^{(n-1)}\right[$ et sur lesquels il n'y a qu'une seule racine sur chaque intervalle de définition puisque la dérivée de $F^{(n)}(\eta)$ est strictement supérieure à l'unité :

$$\frac{\partial F^{(n)(\eta)}}{\partial \eta} = \frac{\alpha_{n-1}}{\eta_k^{(n)} \cdot \left|X_{k,1}^{(n)}\right|^2} = 1 + \alpha_{n-1} \cdot \sum_{k=1}^{n-1} \frac{\eta_k^{(n-1)} \cdot \left|A_{n-1}^{+} \cdot X_k^{(n-1)}\right|^2}{\left(\eta_k^{(n-1)} - \eta\right)^2} > 1 \tag{5}$$

Nous verrons dans la suite que l'inverse de cette dérivée, aux endroits où la fonction s'annule, pourra être interprétée géométriquement.

Pour ce faire, nous allons identifier la matrice R_n^{-1} donnée par sa structure blocs (1), avec son expression à partir de ses éléments propres $\{\eta_k^{(n)}, X_k^{(n)}\}$:

$$R_n^{-1} = \sum_{k=1}^{n} \eta_k^{(n)} X_k^{(n)} \cdot X_k^{(n)+} \tag{6}$$

Par une simple identification terme à terme de R_n^{-1} (1) et (6) et en utilisant la dérivée de $F^{(n)}(\eta)$ donnée par (5), on se ramène immédiatement à :

$$\sum_{k=1}^{n} \pi_k^{(n)} = 1 \quad \text{et} \quad T_{n-1} = \begin{bmatrix} 1 \\ A_{n-1} \end{bmatrix} = \sum_{k=1}^{n} \pi_k^{(n)} \cdot \frac{X_k^{(n)}}{X_{k,1}^{(n)}} \tag{7}$$

avec $\pi_k^{(n)} = \left(\dfrac{\partial F^{(n)}(\eta_k^{(n)})}{\partial \eta}\right)^{-1}$ où $\pi_k^{(n)}$ l'inverse de la dérivée de $F^{(n)}(\eta)$, à la valeur au $F^{(n)}(\eta)$ s'annule (la valeur propre $\eta_k^{(n)}$), apparaît comme le projeté du vecteur de prédiction linéaire T_{n-1} sur le vecteur propre associé $X_k^{(n)}$, normalisé par rapport à sa première composante $X_{k,1}^{(n)}$. Comme par définition T_{n-1}, le vecteur de prédiction linéaire est orthogonal à l'espace signal, il appartient à l'espace bruit. Ainsi, on peut vérifier que pour les valeurs propres de l'espace signal, la valeur $\pi_k^{(n)}$ est plus faible que pour les valeurs propres correspondant à l'espace bruit. Comme la première relation de (7) nous indique que la somme des valeurs $\pi_k^{(n)}$ est normalisée à 1, ce paramètre est un moyen de détermination de la dimension de l'espace signal, en observant le comportement de ces valeurs.

Les relations précédentes permettent également de donner une nouvelle expression du coefficient de réflexion, qui fournit une nouvelle interprétation de ce dernier, soit :

$$\mu_{n-1} = \frac{2 \cdot \sum_{k=1}^{n} \eta_k^{(n)} \cdot X_{k,n}^{(n)} \cdot X_{k,1}^{(n)*}}{\sum_{k=1}^{n} \eta_k^{(n)} \cdot \left[|X_{k,n}^{(n)}| + |X_{k,1}^{(n)}|^2\right]} \underset{\substack{\eta_k^{(n)} = \alpha_n \\ n>M}}{\approx} \underset{k>M}{\operatorname{cov}} [X_{k,n}^{(n)}, X_{k,1}^{(n)}] \tag{8}$$

Ainsi, le coefficient de réflexion à l'ordre $(n-1)$, μ_{n-1} apparaît explicitement comme étant la covariance normalisée entre la première $X_{k,1}^{(n)}$ et la dernière composante $X_{k,n}^{(n)}$ des vecteurs propres $\{X_k^{(n)}\}_{k=1,\ldots,n}$ de la matrice R_n^{-1}, pondérée par les valeurs propres $\{\eta_k^{(n)}\}_{k=1,\ldots,n}$ associées.

Cette approche est comparable à la méthode de Bunse-Gerstner [Bun95] où la parallélisation s'obtient par une méthode de bissection sur des arcs jointifs du cercle complexe unité, en analysant les concordances de signes d'une pseudo séquence de Sturm construite à partir des polynômes récurrents à trois termes [Del91] définis à partir des coefficients de réflexion.

3. analyse autorégressive et géométrie de Chentsov

Il est possible de formaliser certains problèmes d'analyse spectrale en terme de distance entre modèles paramétriques de distributions de probabilité. Cette distance est formulée dans l'espace des paramètres et s'exprime comme la géodésique minimale sur la variété définie par l'entropie. Lorsque cette entropie correspond à l'entropie de Shannon, la métrique différentielle associée est alors fournie par la matrice d'information de Fisher. La géométrie obtenue s'appelle la géométrie de l'information des familles paramétriques. L'intérêt de cette géométrie réside dans le fait qu'elle permet de définir une distance invariante par toute transformation non singulière dans l'espace des paramètres. On rappelle que la divergence de Kullback coincide avec cette métrique différentielle pour des faibles variations dans l'espace des paramètres. En particulier, pour les modèles autorégressifs, il est possible de se ramener à la métrique différentielle de Siegel [Sieg64] que ce dernier a étudiée pour le groupe des matrices Hermitiennes définies positives $GL_n(C)$.

La notion de métrique d'information pour les familles paramétriques de distributions de probabilités utilise les résultats de la géométrie différentielle appliquée aux statistiques et introduit la notion de distance géodésique, proposée initialement par Rao [Burb82] via la matrice d'information de Fisher en généralisant les travaux de Mahalanobis. Etant donné une famille paramétrique de fonctions de densités $G_\Theta = \{p(./\theta) : \theta \in \Theta\}$ avec Θ l'espace des paramètres, on peut définir, sous certaines conditions de régularité, une structure de variété Riemannienne des variétés G_θ ou Θ, par l'intermédiaire de la matrice d'information de Fisher $(g_{ij}(\theta))_{ij}$:

$$I(\theta) = [g_{ij}(\theta)] \tag{9}$$

avec $g_{ij}(\theta) = E\left[\frac{\partial \ln p(x/\theta)}{\partial \theta_i} \cdot \frac{\partial \ln p(x/\theta)}{\partial \theta_j^*}\right]$

La géométrie obtenue est appelée « géométrie de l'information » des familles paramétriques. Cette géométrie possèdent des propriétés intéressantes d'un point de vue statistique, en particulier son invariance par transformations non-singulière des paramètres. Dans ce chapitre, nous allons appliquer ces concepts de géométrie différentielle à l'analyse autorégressive [Barb98,99] (intégration de la métrique de Siegel [Sieg64]). Nous proposerons des schémas numériques de calcul des distances géodésiques de Jensen et montrerons les liens algébriques avec le calcul récursif de la divergence de Kullback dans le cas AR.

3.1. théorème de Pythagore en théorie de l'information

Chentsov [Chen72], élève de Smirnov, a essayé de fonder une « géométrie statistique », par analogie avec la physique où la géométrie différentielle a changé en profondeur notre compréhension du monde. En physique, une loi générale doit rester valide quelque soit le système référentiel considéré. En géométrie classique, de tels changements de situations forment un groupe. De façon équivalente, en statistique [Egu85], le système de toutes les règles de décision de tous les problèmes statistiques concevables, associé aux opérations naturelles de composition, forme une catégorie algébrique. Cette catégorie génère une géométrie uniforme des familles de lois de probabilités, dans laquelle les « éléments » sont les familles statistiques et les « mouvements » sont les règles de décisions. Les distributions de probabilités des familles statistiques (comme les familles paramétrées $\{p(./\theta),\ \theta \in \Theta\}$) peuvent être considérées en terme de variété différentielle. En considérant cette espace comme une variété, et les groupes de transformations associés, l'objectif de cette géométrie est de rechercher les propriétés des éléments appartenant à cette variété qui demeurent inchangées par rapport aux transformations dans ce groupe. Ceci a amené Chentsov [Chen72] à développer le concept de géométrie statistique.

En premier lieu, Chentsov [Chen72] a mis en évidence, au-delà de sa simple propriété de convexité, les propriétés géométriques de la divergence de Kullback :

$$K(p,q) = \int p(x) \cdot \ln \frac{p(x)}{q(x)} \cdot dx = E_p\left[\ln\left[\frac{p}{q}\right]\right] \tag{10}$$

Il a en particulier établit l'équivalent du théorème de Pythagore dans le cas de la géométrie de l'information :

Soit $B_k(p,R)$, $0 < R < \frac{1}{2}r^2 < +\infty$, une K-boule et les états p, q, m tels que :

$$\int m \cdot \ln\left(\frac{q}{p}\right) \cdot d\xi = R \quad ou \quad \int (q-m) \cdot \ln\left(\frac{q}{p}\right) \cdot d\xi = 0 \tag{11}$$

Alors, $D_K^2(m,p) = D_K^2(m,q) + D_K^2(q,p)$

avec $D(m,p) = \sqrt{2 \cdot K(m,p)}$

Malgré son caractère non topologique et non métrique, Chentsov [Chen72] l'a appelée géométrie Pythagorienne asymmétrique et a montré que localement cette géométrie est assimilable à une géométrie Riemannienne symétrique.

3.2. géométrie Riemannienne de l'information

En effet, en appliquant un développement de Taylor d'ordre 2 à la divergence de Kullback pour une petite variation $d\theta$ dans l'espace des paramètres entre les distributions $p(./\theta)$ et $p(.\theta + d\theta)$:

$$K[p(.\theta), p(./\theta + d\theta)] = \frac{1}{2!} \sum_{i,j} \cdot d\theta_i \cdot d\theta_j^* + O\left(|d\theta|^3\right) \tag{12}$$

nous faisons apparaître une métrique Riemannienne invariante, pour toute transformation non-singulière des paramètres, à partir de la matrice d'information de Fisher $I(\theta)$ donnée par (9) :

$$ds^2 = \sum_{i,j} g_{ij}(\theta) d\theta_i \cdot d\theta_j^* = d\theta^+ \cdot I(\theta) \cdot d\theta = \mathrm{Var}\left[d \ln p(x/\theta)\right] \tag{13}$$

C'est la matrice de Fisher qui assure l'invariance par transformation $W(\theta)$ non-singulière des paramètres :

$$w = W(\theta) \Rightarrow I(w) = \left[\frac{\partial\theta_i}{\partial w_j}\right] \cdot I(\theta) \cdot \left[\frac{\partial\theta_i}{\partial w_j}\right] \Rightarrow ds^2(w) = ds^2(\theta) \tag{14}$$

Cette métrique a été généralisée par Burbea et Rao [Burb82] en montrant que le tenseur métrique $g_{ij}(\theta)$ pouvait être interprété comme le Hessien de l'entropie de Shannon, et donc être étendu pour d'autres entropies. Il est alors possible de définir une métrique associée à une fonctionnelle d'entropie quelconque :

$$g_{ij}(\theta) = -E\left[\frac{\partial^2 \ln p(x/\theta)}{\partial\theta_i \cdot \partial\theta_j^*}\right] = \frac{\partial^2 H(p)}{\partial\theta_i \partial\theta_j^*}$$

avec $H(p) = -\int p(x/\theta) \cdot \ln p(x/\theta) \cdot dx$

Si la matrice de Fisher est proportionnelle à l'identité $I(\theta) = 1/\sigma^2 \cdot I_d$, nous sommes ramenés au cas d'une métrique euclidienne $ds^2 = 1/\sigma^2 \cdot \sum_i |d\theta_i|^2$. Ceci suppose l'indépendance entre paramètres θ_i et une covariance identique σ^2. De toutes les manières, nous resterons contraints par la borne de Cramer-Rao $E\left[(\hat{\theta} - \theta) \cdot (\hat{\theta} - \theta)^+\right] \geq I(\theta)^{-1}$, avec égalité pour le Maximum de vraisemblance. Aussi, dans certains cas applicatifs, nous pourrions remplacer la matrice de Fisher $I(\theta)$, par la matrice de covariance sur les paramètres $E\left[(\hat{\theta} - \theta) \cdot (\hat{\theta} - \theta)^+\right]$ obtenue par une méthode « Bootstrap », étudiée par Bühlmann dans le cas autorégressif [Bühl95].

3.3. géométrie différentielle et métrique de Siegel

Dans le cas de l'analyse autorégressive, nous allons essayer de calculer formellement la métrique différentielle associée à la matrice d'information de Fisher, et son intégration pour définir une distance en terme de géodésique minimale. Considérons donc un processus multivarié gaussien :

$$p(X_n/R_n) = (2\pi)^{-n/2} \cdot |R_n|^{-1/2} \cdot e^{-\frac{1}{2}\mathrm{Tr}\,[\hat{R}_n \cdot R_n^{-1}]} \tag{15}$$

avec $\hat{R}_n = (X_n - m_n) \cdot (X_n - m_n)^+$ et $E[\hat{R}_n] = R_n$

Les éléments de la matrice de Fisher ont déja été calculés par différents auteurs, et le résultat suivant est désormais classique en traitement de signal [Kopp86] :

$$g_{ij}(\theta) = -\frac{1}{2}\mathrm{Tr}\,[\partial_i R_n \cdot \partial_j R_n^{-1}] + \partial_i m_n^+ \cdot R_n^{-1} \cdot \partial_j m_n \tag{16}$$

avec $\partial_i = \dfrac{\partial}{\partial\theta_i}$.

Pour nous ramener à une expression simple de la métrique différentielle ds^2, nous ferons en premier lieu l'hypothèse que le processus est de moyenne nulle $m_n = 0$, puis nous appliquerons un développement de Taylor d'ordre 1 à la variation élémentaire de l'inverse de la matrice de corrélation $dR_n^{-1} = \sum_i \partial_i R_n^{-1} \cdot d\theta_i$.

En se reportant alors à l'expression (13) de la métrique ds^2, nous obtenons :

$$ds^2(\theta) = \frac{1}{2}\mathrm{Tr}\,\left[(R_n \cdot dR_n^{-1})^2\right] \tag{17}$$

Cette métrique a été étudiée par Siegel [Sieg64] pour l'ensemble des matrices Hermitiennes dans le contexte de la géométrie symplectique et récemment par Calvo & Oller [Cal90]. Plus généralement, elle définit la notion de produit scalaire pour le groupe $GL_n(C)$:

$$\langle X, Y\rangle_{R_n} = \frac{1}{2}\mathrm{Tr}\,[R_n \cdot X \cdot R_n \cdot Y] \tag{18}$$

où X et $Y \in GL_n(C)$

3.4. distance de Möbius et métrique différentielle

Nous allons essayer de développer l'expression de cette métrique (17) de Siegel dans le cas autorégressif, ainsi que son intégration pour finalement proposer une distance qui s'exprime de façon simple à partir des coefficients de réflexion. Encore une fois, nous allons faire appelle à la structure blocs définie en (1). Il suffit alors d'appliquer la formule (17), pour faire apparaître que cette dernière s'exprime de façon récursive comme suit :

$$ds_n^2 = ds_{n-1}^2 + \frac{1}{2} \cdot \left(\frac{d\alpha_{n-1}}{\alpha_{n-1}}\right)^2 + \alpha_{n-1} \cdot dA_{n-1}^+ \cdot R_{n-1} \cdot dA_{n-1} \tag{19}$$

En différençiant, l'équation de Levinson donnée en (2), il est alors possible de trouver une borne inférieure à cette métrique à partir des paramètres $\{\alpha_k, \mu_k\}$:

$$ds_n^2 > \frac{1}{2}\sum_{k=0}^{n-1}\left(\frac{d\alpha_k}{\alpha_k}\right)^2 + \sum_{k=1}^{n-1}\frac{|d\mu_k|^2}{1 - |\mu_k|^2} \tag{20}$$

Il reste alors à intégrer cette métrique (20) en remarquant que $|dz|/\sqrt{1 - |z|^2} = d[\arcsin|z|]$. On utilise alors la transformation de Möbius définie par la relation $v_k = \phi_{\tau_k}(\omega_k) = \dfrac{\omega_k - \tau_k}{\tau_k^* \omega_k - 1} \cdot e^{i\varphi}$ pour définir un automorphisme holomorphique ϕ du polydisque complexe unité D^{n-1} dans D^{n-1}, tel que $S(\omega, \tau) = S(\phi_\tau(\omega), \phi_\tau(\tau)) = S(v, 0)$. Il en découle alors la définition d'une nouvelle distance définie à partir de la transformation de Möbius sur les coefficients de réflexion :

$$S^2\left(R_n^{(1)}, R_n^{(2)}\right) = \frac{1}{2}\sum_{k=0}^{n-1}\ln^2(\beta_k) + \sum_{k=1}^{n-1}[\arcsin(\delta_k)]^2 \tag{21}$$

avec $\beta_k = \dfrac{1-|\mu_k^{(1)}|^2}{1-|\mu_k^{(2)}|^2}\beta_{k-1}$ et $\delta_k = \left|\dfrac{\mu_k^{(1)}-\mu_k^{(2)}}{1-\mu_k^{(1)}\cdot\mu_k^{(2)*}}\right|$

On remarquera que la distance obtenue est différente de la distance classique dans le polydisque complexe D^{n-1} [Pflu93].

3.5. géodésique minimale et théorème de Jensen dans le cas autorégressif

Nous proposons également une autre manière de calculer la distance minimale entre densités de probabilité dans le cas autorégressif, en utilisant le théorème de Jensen qui fournit une expression formelle de la géodésique minimale, pour des variables multivariées gaussiennes $\{R_n^{(1)} \neq R_n^{(2)}$ et $m_n^{(1)} = m_n^{(2)}\}$, dans le cas de la métrique différentielle de Fisher. Jensen a montré que cette géodésique minimale s'obtenait par calcul de valeurs propres sur la base du théorème suivant :

$$\begin{aligned} S^2\left(R_n^{(1)},R_n^{(2)}\right) &= \frac{1}{2}\cdot\left|\left|\ln\left(R_n^{(1)1/2+}\cdot R_n^{(2)-1}\cdot R_n^{(1)1/2}\right)\right|\right|^2 \\ &= \frac{1}{2}\sum_{k=1}^{n}\ln{}^2(\lambda_k^{(n)}) \end{aligned} \tag{22}$$

avec $\left|R_n^{(1)1/2+}\cdot R_n^{(2)-1}\cdot R_n^{(1)1/2}-\lambda_k^{(n)}\cdot I_n\right| = 0.$

Ce résultat a été complété par James [James73] qui a établit la distribution asymptotique de la statistique et a montré que si $R_n^{(1)} = R_n^{(2)}$ alors $\dfrac{n}{2}\sum_{i=1}^{n}\ln^2(\lambda_i)$ suit une loi de Khi-2 à $n\cdot(n+1)/2$ degrés de libertés. Swain [Swai75] a généralisé le théorème de Jensen en introduisant une classe de fonctions de contraste $S_v^2\left(R_n^{(1)},R_n^{(2)}\right) = \sum_{k=1}^{n} v\left(\lambda_k^{(n)}\right)$ avec les $\lambda_k^{(n)}$ valeurs propres de $R_n^{(1)1/2+}\cdot R_n^{(2)-1}\cdot R_n^{(1)1/2}$ et $v(\lambda)$ une fonction C^3 sur $[0,+\infty[$ satisfaisant $v(1) = v'(1) = 0$, $v(\lambda) = \dfrac{1}{2}$ et $v(\lambda) > 0$, si $\lambda > 0, \lambda \neq 1$ en remarquant que $S_v^2\left(R_n^{(1)},R_n^{(2)}\right) = S_w^2\left(R_n^{(2)},R_n^{(1)}\right)$ avec $w(\lambda) = v(\lambda^{-1})$. Dans notre logique, nous avons utilisé de nouveau les structures blocs données en (1) pour mettre en évidence que la matrice apparaissant dans le théorème (22) de Jensen $R_n^{(1)1/2+}\cdot R^{(2)-1}\cdot R_n^{(1)1/2}$ possédait une structure blocs identique à R_n^{-1} :

$$\begin{aligned} \Sigma_n &= R_n^{(1)1/2+}\cdot R_n^{(2)-1}\cdot R_n^{(1)1/2} \\ &= \begin{bmatrix} \beta_{n-1} & \beta_{n-1}\cdot V_{n-1}^+ \\ \beta_{n-1}\cdot V_{n-1} & \Sigma_{n-1}+\beta_{n-1}\cdot V_{n-1}\cdot V_{n-1}^+ \end{bmatrix} \end{aligned} \tag{23}$$

où $V_{n-1} = F_{n-1}^{(1)+}\cdot\left[A_{n-1}^{(2)}-A_{n-1}^{(1)}\right]$ et $\beta_{n-1} = \dfrac{\alpha_{n-1}^{(2)}}{\alpha_{n-1}^{(1)}}$

avec $\alpha_n^{(1)}\cdot R_n^{(1)} = F_n^{(1)}\cdot F_n^{(1)+}$.

Nous avons fait apparaitre la différence entre vecteurs autorégressifs $\left[A_{n-1}^{(2)}-A_{n-1}^{(1)}\right]$, à nouveau le terme β_{n-1}, ainsi que la racine carré $F_n^{(1)}$ de $\alpha_n^{(1)}\cdot R_n^{(1)}$. Nous avons montré que cette dernière expression $F_n^{(1)}$ [Barb99] s'exprime de façon récursive en faisant apparaître une matrice de type compagnon qui s'exprime à partir des paramètres autorégressifs. On voit de façon évidente que Σ_n (23) et R_n^{-1} (1) possèdent la même structure blocs. Aussi, nous pouvons appliquer le même schéma numérique qu'au chapitre 2 pour calculer les valeurs propres $\lambda_k^{(n)}$ de $\sum_n$ et d'après (22) la géodésique minimale fournie par le théorème de Jensen :

avec $\lambda_k^{(n)}$ racine de

$$G^{(n)}\left(\lambda_k^{(n)}\right) = \lambda_k^{(n)}-\beta_{n-1}+\beta_{n-1}\cdot\lambda_k^{(n)}\cdot\sum_{i=1}^{n-1}\frac{|V_{n-1}^+\cdot X_i^{(n-1)}|^2}{\left(\lambda_i^{(n-1)}-\lambda_k^{(n)}\right)} = 0 \tag{24}$$

3.6. équations d'Euler-Lagrange pour les modèles autorégressifs

Par définition, la plus courte distance entre deux points θ_1, θ_2 le long de la variété différentielle, de métrique associée $ds^2(\theta)$, est donnée par la géodésique minimale solution du problème variationnel suivant : $\underset{C_{\theta_1,\theta_2}}{\text{Inf}}\left|\int\sqrt{ds^2(\theta)}\cdot d\theta\right|$ fournie par les équations d'Euler- Lagrange :

$$\sum_{i=1}^{n} g_{ik}(\theta)\cdot\frac{d^2\theta}{dt^2}+\frac{1}{2}\sum_{i,j=1}^{n}\left[\frac{\partial g_{ik}(\theta)}{\partial\theta_j}+\frac{\partial g_{ik}(\theta)}{\partial\theta_i}-\frac{\partial g_{ij}(\theta)}{\partial\theta_k}\right]\cdot\frac{d\theta_i}{dt}\cdot\frac{d\theta_j}{dt} = 0 \quad i,j,k = 1,2,\ldots,n \tag{25}$$

La géodésique (la borne inf. de la distance de Rao) est une distance et elle est invariante par action du groupe affine dans l'espace des paramètres. Skovgaard [Skov84] a montré que la géodésique minimale définissant la distance entre deux distributions gaussiennes multivariées (m_n, R_n) était solution de l'équation d'Euler-Lagrange associée suivante :

$$\frac{d^2m_n}{dt^2}-\frac{dR_n}{dt}\cdot R_n^{-1}\cdot\frac{dm_n}{dt} = 0$$

et

$$\frac{d^2R_n}{dt^2}+\frac{dm_n}{dt}\cdot\frac{dm_n^+}{dt}-\frac{dR_n}{dt}\cdot R_n^{-1}\cdot\frac{dR_n}{dt} = 0 \tag{26}$$

En utilisant, la structure blocs (1) de R_n, il serait alors possible de déterminer les équations d'Euler-Lagrange pour les paramètres de l'analyse autorégressive $\{\alpha_{n-1}, A_{n-1}\}$ récursivement sur l'ordre n. Ce travail est encore à développer.

Ce travail est encore à développer à partir des équations différentielles associées déduites des structures blocs (1) et (2) :

$$\begin{cases} d^2 \ln \alpha_{n-1} - \alpha_{n-1} \cdot dA^+_{n-1} \cdot R_{n-1} \cdot dA_{n-1} = 0 \\ \alpha_{n-1} \cdot R_{n-1} \cdot dA_{n-1} \cdot dA^+_{n-1} = 0 \\ d^2 A_{n-1} + [d \ln \alpha_{n-1} \cdot I_{n-1} - R_{n-1} \cdot dR^{-1}_{n-1}] \cdot dA_{n-1} = 0 \end{cases}$$

3.7. divergence de Kullback pour les modèles autorégressifs

Au chapitre 3.5, nous avons mis en évidence la structure blocs (23) de la matrice $\Sigma_n = R_n^{(1)1/2+} \cdot R_n^{(2)-1} \cdot R_n^{(1)1/2}$. Or, il est possible de faire apparaître cette matrice Σ_n dans le calcul de la divergence de Kullback entre variable multivariée gaussiennes $\{R_n^{(1)} \neq R_n^{(2)}$ et $m_n^{(1)} = m_n^{(2)}\}$. Ceci va nous permettre de proposer un calcul récursif de la divergence de Kullback $K(R_n^{(1)}, R_n^{(2)})$ à partir des paramètres de l'analyse autorégressive $V_{n-1} = F_{n-1}^{(1)+} \cdot \left[A_{n-1}^{(2)} - A_{n-1}^{(1)}\right]$ et β_{n-1}. Ainsi, la divergence de Kullback peut se réécrire :

$$K\left(R_n^{(1)}, R_n^{(2)}\right) = \frac{1}{2}\left[\mathrm{Tr}[\Sigma_n] - \ln|\Sigma_n| - n\right].$$

Or d'après la structure blocs (23) de Σ_n, $|\Sigma_n| = \beta_{n-1} \cdot |\Sigma_{n-1}|$ et $\mathrm{Tr}[\Sigma_n] = \mathrm{Tr}[\Sigma_{n-1}] + \beta_{n-1} \cdot [1 + V^+_{n-1} \cdot V_{n-1}]$, ce qui permet finalement de trouver l'expression finale récursive de la divergence de Kullback :

$$K\left(R_n^{(1)}, R_n^{(2)}\right) = K\left(R_{n-1}^{(1)}, R_{n-1}^{(2)}\right) + \frac{1}{2}\left[(\beta_{n-1} - 1) + \beta_{n-1} \cdot V^+_{n-1} \cdot V_{n-1} - \ln(\beta_{n-1})\right] \quad (27)$$

3.8. géométrie Finslérienne de l'information

Jusqu'à présent, seul un développement de Taylor à l'ordre deux de la divergence de Kullback a été employé pour mettre en évidence la structure de géométrie Riemannienne à partir de la matrice de Fisher. Cependant, pour des variations plus importantes dans l'espace des paramètres $d\theta$, on ne peut plus négliger les termes d'ordre supérieur à deux dans le développement de Taylor de la divergence de Kullback. L'espace apparaît alors comme explicitement non-Riemannien et asymétrique :

$$K[p(./\theta), p(./\theta + d\theta)] = \frac{1}{2!}\sum_{i,j} g_{ij}(\theta) \cdot d\theta_i \cdot d\theta_j + \frac{1}{3!}\sum_{i,j,k} g_{i,j,k}(\theta) \cdot d\theta_i \cdot d\theta_j \cdot d\theta_k + (|d\theta|^4) \quad (28)$$

avec

$$g_{ijk}(\theta) = \int p(x/\theta) \cdot \left[3 \cdot \frac{\partial \ln p(x/\theta)}{\partial \theta_i} \cdot \frac{\partial \ln p(x/\theta)}{\partial \theta_j} \cdot \frac{\partial \ln p(x/\theta)}{\partial \theta_k} + \sum_{\text{cycl}} \frac{\partial \ln p(x/\theta)}{\partial \theta_i} \cdot \frac{\partial^2 \ln p(x/\theta)}{\partial \theta_j \cdot \partial \theta_k}\right] \cdot dx$$

Ingarden [Ing81] a appelé cette structure une géométrie Finslérienne non-homogène, étant une extension inhomogène de la géométrie de Finsler classique [Cart34], pour laquelle la formulation variationnelle de la géodésique $K(p,q) = \int_p^q ds$ ne peut plus être définie et où la notion de connexion n'a plus de définition au sens habituel. Une équivalence est possible avec le principe de Glansdorff-Prigogine de production minimum d'entropie. Ainsi des contraintes doivent être rajoutées à travers le semi-groupe dynamique Λ_t et la production d'entropie minimum $\sigma(p)$ et donner un sens à la notion de chemin minimal :

$$\sigma(p) = \left.\frac{dK(\Lambda_t p, p_0)}{dt}\right|_{t=0} \quad \text{avec } \Lambda_t p_0 = p_0 \qquad (29)$$

La géometrie de Finsler fait partie d'un domaine de recherche mathématique assez actif. Aussi, son utilisation en géométrie de l'information en est à son début.

4. conclusion

J'ai tenté dans cet article de montrer les richesses algébriques de l'analyse autorégressive. Mais, je tenais avant tout à sensibiliser la communauté traitement de signal, sur l'ouverture immense que nous offre la géométrie Riemannienne dans le cadre de la statistique, pour définir de nouveaux algorithmes, fournir de nouvelles interprétations, et placer notre discipline au rang de celles, comme la physique théorique, qui doivent leur renommée aux outils mathématiques qu'elles utilisent. Ainsi des passerelles peuvent se créer entre traitement de signal statistique et mécanique quantique [Petz96] sur la base de la géométrie de l'information. Nous voyons également se généraliser, l'utilisation de la géométrie différentielle en traitement du signal et de l'image : Brockett (Gradient Flow on Manifolds), Lions (mean curvature flow), Karmarkar (Interior Point Flow), Kulhavy (Projection Filter), ...

Ainsi de nouvelles approches peuvent être envisagées pour l'estimation spectrale à partir d'une approche EDP de type flot géodésique. Plus particulièrement, l'analyse AR régularisée peut être formalisée en terme variationnel en faisant l'analogie suivante :

$$\operatorname*{Min}_{A_n}\left[\int E^{(n)}(f) \cdot df + \lambda \cdot \int \left|\frac{\partial A^{(n)}(f)}{\partial f}\right| \cdot df\right] \Leftrightarrow \operatorname*{Min}_{A_n}\left[\int \sqrt{E^{(n)}(f)} \cdot \left|\frac{\partial A^{(n)}(f)}{\partial f}\right| \cdot df\right] \quad (30)$$

avec les notations suivantes :

$$E^{(n)}(f) = \sum_{k=-\infty}^{+\infty} E[e_n(i) \cdot e_n^*(i-k)] \cdot e^{-j2\pi f \cdot k}$$

et

$$A^{(n)}(f) = \sum_{k=0}^{n} a_k^{(n)} \cdot e^{-j2\pi f \cdot k}$$

avec $e_n(i) = \sum_{k=0}^{n} a_k^{(n)} \cdot x_{i-k}$

Le problème se ramène alors à la détermination de la géodésique minimale suivante :

$$\underset{A_n}{\mathrm{Min}} \left[\int \sqrt{E^{(n)}(f)} \cdot ds \right] = \underset{A_n}{\mathrm{Min}} \left[\int ds_E \right]$$

avec

$$ds_E^2 = E^{(n)}(f) \cdot ds^2 \quad \text{et} \quad ds = \left| \frac{\partial A^{(n)}(f)}{\partial f} \right| \cdot df \qquad (31)$$

Le problème se résoud alors en utilisant un généralisation de l'équation de Fourier de la chaleur à partir de l'opérateur de Beltrami :

$$\begin{aligned} \frac{\partial A^{(n)}}{\partial t} &= \Delta_E A^{(n)}(f) \\ &= \frac{1}{\sqrt{|E^{(n)}(f)|}} \cdot \frac{\partial}{\partial f} \cdot \left[\frac{\sqrt{|E^{(n)}(f)|}}{E^{(n)}(f)} \cdot \frac{\partial A^{(n)}(f)}{\partial f} \right] \end{aligned} \qquad (32)$$

A partir de la transformée en z de cette EDP et par identification terme à terme, il est possible de déduire un schéma numérique fournissant la solution de l'analyse spectrale régularisée par le flot géodésique.

5. remerciements

Je tenais, par cet article, à exprimer toute ma reconnaissance au professeur Bernard Picinbono, qui par son enseignement rigoureux de la discipline du traitement du signal et ses ouvrages, m'a permis d'ajouter une pièce à l'édifice dont il a forgé les bases, et qu'il continue de construire et de structurer.

BIBLIOGRAPHIE

[Atk81] C. Atkinson & A.F. Mitchell, «Rao's Distance Measure», Sankya, A43, pp.345-365, 1981.

[Barb97a] F. Barbaresco, «Recursive Eigendecomposition via AR Analysis & Ago-antagonistic Regularization», *Proc. ICASSP-97*, vol. V, pp. 3969-3972, April 1997.

[Barb97b] F. Barbaresco, «Unitary Hessenberg Approach In Spectral Analysis & Comparison with Root-Music Regularization», Rapport interne THOMSON-CSF, Décembre 1997.

[Barb98] F. Barbaresco, «Half-quadratic Regularization of Time-Frequency AR Analysis for Recovery of Abrupt Spectral Discontinuities & their Detection by a Recursive Siegel Metric», *Proc. EUSIPCO-98*, pp. 621, Sept. 1998.

[Barb99] F. barbaresco, «Extrinsic & Intrinsic Entropic Priors for Time-Frequency AR Analysis : Half-Quadratic Temporal Regularization & Recursive Siegel Metric Based on Information Riemannian Geometry», *Coll. PSIP'99, ENST, Paris*, Janvier 1999.

[Bass97] M. Basseville, «Information : Entropies, Divergences et Moyennes», *IRISA Report n° 1020*, 1997.

[Bühl95] P. Bühlmann, «Sieve Bootstrap for Time Series», Tech. Report, n°431, *Dept. of Stat., Univ. of California*, Berkeley, 1995.

[Bun95] Bunse-Gerstner & He, «On a Sturm Sequence of Polynomials for Unitary Hessenberg Matrices», *SIAM Matrix Analysis*, vol. 16, n°4, pp. 1043-1055, October 1995.

[Burb82] J. Burbea & C.R. Rao, «Entropy Differential Metric, Distance and Divergence Measures in Probability Spaces : A unified Approach», *Journal of Multivariate Analysis*, vol.12, pp.575-596, 1982.

[Cal90] M. Calvo & J. Oller, «A Distance Between Multivariate Normal Distributions in an Embedding into the Siegel Group», *Journal of Multivariate Analysis*, vol. 35, n°2, pp.223-242, Nov. 1990.

[Cart34] E. Cartan, «Les espaces de Finsler», *Actualités 79*, Paris, 1934.

[Chen72] N.N. Chentsov, «Statistical Decision Rules and Optimal Inference», *Translation of Mathematical Monographs*, Vol. 53, AMS, Rhodes Island, NY, 1972.

[Del91] P. Delsarte & Y. Genin, «Tridiagonal Approach to the Algebraic Environment of Toeplitz Matrices, Part I & II, *SIAM J. Matrix Anal. Appl.*, vol.12, n°2 & n°3, April & July 1991.

[Egu85] S. Eguchi, «A Differential Geometric Approach to Statistical Inference on the Basis of Contrast Functionals», *Hirosh. Math. J.*, Vol. 15,pp. 341-391, 1985.

[Ing81] R. S. Ingarden, «Information Geometry in Functional Spaces of Classical and Quantum Finite Statistical Systems», *Int. J. Engineering Sci*, vol. 19, n°12, pp. 1609-1633, 1981.

[James73] A.T. James, «The Variance Information Manifold and Functions on it», Multivariate Analysis III, Ed. P.K. Krishnaiah, *Academic Press*, New York, pp.157-169, 1973.

[Kopp86] L. Kopp & D. Thubert, «Bornes de Cramer-Rao en traitement d'antenne. Première Partie : Formalisme «, *revue Traitement du Signal*, vol.3, n°3, pp.111-125, 1986.

[Petz96] D. Petz & H. Hasegawa, «On the Riemannian Metric of a-entropies of Density Matrices», Letter in Mathematical Physics, vol.38, pp.221-225, 1996.

[Pflu93] P. Pflug & Mareck Jarnicki, «Invariant Distances and Metrics in Complex Analysis», *Edition De Gruyter*, 1993.

[Pic94] B. Picinbono, «On Circularity», *IEEE Trans. on Signal Processing*, vol.42, pp. 3473-3482, 1994.

[Pic95] B. Picinbono, «Signaux aléatoires : Bases du traitement statistique du signal», Tome 3, Dunod, 1995.

[Skov84] L. Skovgaard, «A Riemannian Geometry of the Multivariate Normal Model», *Scand. J. Stat.*, Vol.11, pp. 211-223, 1984.

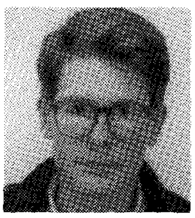
[Sieg64] Carl Ludwig Siegel, «Symplectic Geometry», *Academic Press*, New-York, 1964.

[Swai75] A.J. Swain, «A Class of Factor Analysis Estimation Procedures with Common Asymptotic Sampling Properties», *Psychometrika*, vol. 40, pp.315-335, 1975.

Manuscrit reçu le 2 avril 1998.

L' AUTEUR

Frédéric BARBARESCO



Né en 1966 à Nantes, Frédéric BARBARESCO est ingénieur SUPELEC, promotion 91 (section Automatique) et ingénieur I.N.Télécoms, promotion 89 (option communication et traitement du signal). Membre du Collège Scientifique de THOMSON-CSF, il est actuellement rattaché à la direction technique de l'unité de développement radar de THOMSON-CSF AIRSYS. Ses thèmes de recherche concernent le traitement de signal (détection et segmentation Doppler), le traitement d'image (prédiction du mouvement fluide des fouillis radar), le pistage (Track-Before-Detect) et l'optimisation (séquencement des formes d'onde radar).