

Espace Riemannien symétrique et géométrie des espaces de matrices de covariance : équations de diffusion et calculs de médianes

FRÉDÉRIC BARBARESCO¹, GUILLAUME BOUYT²

¹ Thales Air Systems, Département Stratégie Technologie et Innovation, Hameau de Roussigny, 91470 Limours, France
² Mines ParisTech, 60, Boulevard Saint-Michel 75272 PARIS cedex 06

Résumé - Les matrices de covariance sont des objets fondamentaux en traitement du signal et souvent à la base des algorithmes en Science de l'information. Nous replaçons l'étude de ces matrices de covariance dans un cadre purement géométrique lié aux espaces métriques à courbure négative en soulignant les liens avec les espaces symétriques riemanniens et le demi-plan de Siegel. Sur la base du flot de gradient de Karcher permettant de façon itérative le calcul du barycentre de N matrices de covariance, nous définissons la notion d'équation de diffusion sur un graphe de matrices de covariance. Nous introduisons enfin le calcul des médianes via un modèle autorégressif complexe pour conserver la structure Toeplitz et dans le cas de lois multivariées gaussiennes de moyenne non nulle.

Abstract – Covariance matrices are fundamental tools in signal processing and used in most algorithms in Information Science. We replace the study of these covariance matrices in a purely geometrical framework given by metric spaces of negative curvature and we underline links with symmetric Riemannian spaces and Siegel half-plane geometry. On the basis of Karcher gradient flow for computation of N matrices barycenter, we define the notion of Fourier heat flow equation on a graph of covariance matrices. We introduce finally median computation by complex autoregressive model to preserve Toeplitz structure and in case of multivariate Gaussian laws of non-zero mean.

1 Géométrie des matrices de covariances

1.1 Demi-plan de Siegel & espace de matrices hermitiennes définies positives

Classiquement en traitement du signal, on manipule des matrices de covariance $R_n = E[S_n S_n^+]$ d'une série temporelle ou spatiale de données $S_n = [s_1 \ s_2 \ \dots \ s_n]^T$. La matrice de covariance ainsi définie est Toeplitz Hermitienne définie positive : $\forall W \in C^n, W^+ R_n W > 0$ et $R_n^+ = R_n$. Pour manipuler habituellement de telles matrices et définir des distances, on utilise la norme classique de Frobenius donnée par l'expression $\|A - B\|_F = \langle A - B, A - B \rangle = \text{Tr}((A - B)(A - B)^+)$. Cette hypothèse suppose un espace vectoriel normé de courbure nulle pour lequel la moyenne entre 2 matrices et la géodésique sont alors données par : $A \circ B = \frac{A+B}{2}$, $\gamma(t) = A + t(B - A)$. En fait, l'espace des matrices hermitiennes définies positives définit un cône symétrique et doit se traiter dans le cadre des espaces métriques à courbure négative. Nous proposons ici de formaliser le problème avec la modélisation proposée par Carl Ludwig Siegel [1]. Le demi-plan de Siegel correspond à l'espace des matrices complexes dont la partie imaginaire est définie positive $SH_n = \{Z = X + iY \in \text{Sym}(n, C) / \text{Im}(Z) = Y > 0\}$ (il s'agit d'une extension dans le cas matriciel du demi-plan de Poincaré). Dans cet espace, l'action du groupe dit Symplectique $Sp_{2n}R$ agit transitivement (il s'agit d'une généralisation du groupe $SL_2R = Sp_2R$ des matrices inversibles de déterminant 1, qui intervient dans l'espace de Poincaré). Le groupe symplectique est

constitué des matrices $M \in GL_{2n}F$ (F étant égale à R ou C) de dimension $2n$, qui satisfont aux relations :

$$Sp(n, F) \equiv \left\{ M \in GL(2n, F) / M^T J M = J, J = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix} \right\} \in SL(2n, R)$$

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in Sp(n, F) \Leftrightarrow \begin{cases} A^T C \text{ et } B^T D \text{ symétrique} \\ A^T D - C^T B = I_n \end{cases}$$

On considère alors le groupe quotient $PSp(n, R) \equiv Sp(n, R) / \{\pm I_{2n}\}$ qui agit comme un sous-groupe d'isométrie sur SH_n via la transformation de Möbius généralisée : $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \Rightarrow M(Z) = (AZ + B)(CZ + D)^{-1}$. Siegel a montré que l'unique métrique invariante par cette isométrie était : $ds_{\text{Siegel}}^2 = \text{Tr}(Y^{-1}(dZ)Y^{-1}(d\bar{Z}))$ avec $Z = X + iY$.

L'espace des matrices de covariance correspond au cas $X = 0$ et $Y = R$, soit $Z = iR$. Dans ce cas, en développant l'expression de la métrique de Siegel, on retrouve la métrique classique introduite en géométrie de l'Information par Rao [6] : $ds^2 = \text{Tr}((R^{-1}(dR))^2)$. Si on intègre cette métrique pour calculer la distance entre 2 matrices R_1 et R_2 , on obtient l'expression suivante, qui peut se calculer à partir des valeurs propres étendues $\{\lambda_k\}_{k=1}^n$ ($\det(R_2 - \lambda R_1) = 0$) entre les deux matrices R_1 et R_2 :

$$d^2(R_1, R_2) = \left\| \log(R_1^{-1/2} R_2 R_1^{-1/2}) \right\|^2 = \sum_{k=1}^n \log^2(\lambda_k)$$

Dans le cas général, ou $X \neq 0$, Siegel a donné l'expression de la distance qui est la suivante :

$$d_{\text{Siegel}}^2(Z_1, Z_2) = \left(\sum_{k=1}^n \log^2 \left(\frac{1 + \sqrt{r_k}}{1 - \sqrt{r_k}} \right) \right) \text{ avec } Z_1, Z_2 \in SH_n$$

où les r_k ($\det(R - r.I) = 0$) sont les valeurs propres de :

$$R(Z_1, Z_2) = (Z_1 - Z_2)(Z_1 - \bar{Z}_2)^{-1}(\bar{Z}_1 - \bar{Z}_2)(Z_1 - Z_2)^{-1}$$

Il suffit de remarquer que la différentielle 2^{nd} de $Z \rightarrow R(Z_1, Z)$ en $Z_1 = Z$ est donnée par l'expression :

proposé une moyenne log-euclidienne $A = \exp\left(\frac{1}{N} \sum_{k=1}^N B_k\right)$,

reprise dans [5] mais la moyenne obtenue n'est pas invariante par n'importe quelle transformation $\Sigma \rightarrow W\Sigma W^+$, $W \in GL_n(C)$. En fait c'est vrai seulement si W est une matrice de similitude. Or, cette invariance générale est fondamentale (transformations qui résultent d'un changement de base sur le vecteur de données).

3 Equation de diffusion & graphe de matrices

Dans un espace vectoriel normé dans le cas unidimensionnel, l'équation de diffusion s'écrit :

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} \Rightarrow \frac{\partial u}{\partial t} = \frac{1}{\nabla x} \left(\frac{u_{n+1} - u_n}{\nabla x} - \frac{u_n - u_{n-1}}{\nabla x} \right) = \frac{2}{\nabla x^2} (\hat{u}_n - u_n) \text{ avec}$$

$\hat{u}_n = \frac{u_{n+1} + u_{n-1}}{2}$ la moyenne arithmétique. L'équation de

Fourier discrétisée peut également s'écrire :

$$u_{n,t+1} = u_{n,t} + \frac{2\nabla t}{\nabla x^2} [\hat{u}_{n,t} - u_{n,t}] = (1-\rho)u_{n,t} + \rho\hat{u}_{n,t}. \text{ Par analogie sur}$$

la base du flot de Karcher, il est possible d'écrire une équation de diffusion sur un graphe 1D de matrice, via la moyenne géométrique $\hat{A}_{n,t}$ des voisins $(A_{n-1,t}, A_{n+1,t})$:

$$\boxed{\begin{aligned} A_{n,t+1} &= A_{n,t}^{1/2} e^{\frac{2\nabla t}{\nabla x^2} \log(A_{n,t}^{1/2} \hat{A}_{n,t} A_{n,t}^{1/2})} A_{n,t}^{1/2} = A_{n,t}^{1/2} (A_{n,t}^{-1/2} \hat{A}_{n,t} A_{n,t}^{-1/2})^\rho A_{n,t}^{1/2} \\ \text{avec } \rho &= \frac{2\nabla t}{\nabla x^2} \text{ et } \hat{A}_{n,t} = A_{n+1,t}^{1/2} (A_{n+1,t}^{-1/2} A_{n-1,t} A_{n+1,t}^{-1/2})^{1/2} A_{n+1,t}^{1/2} \end{aligned}}$$

A partir des géodésiques $\gamma_{A_{n,t+1}}^{A_{n,t}}(\lambda)$ et $\gamma_{\hat{A}_{n,t}}^{A_{n,t}}(\lambda)$ entre $A_{n,t}$,

$$A_{n,t+1} \text{ et } \hat{A}_{n,t} : \log(A_{n,t}^{-1/2} A_{n,t+1} A_{n,t}^{-1/2}) = \frac{2\nabla t}{\nabla x^2} \log(A_{n,t}^{-1/2} \hat{A}_{n,t} A_{n,t}^{-1/2})$$

$$\Rightarrow \left. \frac{d\gamma_{A_{n,t+1}}^{A_{n,t}}(\lambda)}{d\lambda} \right|_{\lambda=0} = \frac{2\nabla t}{\nabla x^2} \left. \frac{d\gamma_{\hat{A}_{n,t}}^{A_{n,t}}(\lambda)}{d\lambda} \right|_{\lambda=0}$$

Regardons maintenant le cas anisotrope classique :

$$\frac{\partial u}{\partial t} = \frac{1}{\sqrt{\det(g)}} \sum_{i,j} \frac{\partial}{\partial x_i} \left(\sqrt{\det(g)} g^{ij} \frac{\partial u}{\partial x_j} \right). \text{ Dans le cas 1 D, on a :}$$

$$ds^2 = dx^2 + du^2 = (1 + (\nabla_x u)^2) dx^2 \text{ avec } \nabla_x u = \frac{\partial u}{\partial x} \text{ et le flot :}$$

$$\frac{\partial u}{\partial t} = (1 + (\nabla_x u)^2)^{-1/2} \frac{\partial}{\partial x} \left((1 + (\nabla_x u)^2)^{1/2} \frac{\partial u}{\partial x} \right) \text{ qui est discrétisé :}$$

$$u_{n,t+1} = u_{n,t} + \frac{\nabla t \alpha_{n,t}}{\nabla x} \left[\alpha_{n,t}^+ \left(\frac{u_{n+1,t} - u_{n,t}}{\nabla x} \right) - \alpha_{n,t}^- \left(\frac{u_{n,t} - u_{n-1,t}}{\nabla x} \right) \right]$$

$$u_{n,t+1} = u_{n,t} + \gamma [\hat{u}_{n,t}^\beta - u_{n,t}] \text{ avec } \hat{u}_{n,t}^\beta = \beta u_{n+1,t} + (1-\beta)u_{n-1,t}$$

Par analogie, nous proposons un flot anisotrope sur un graphe 1D de matrices comme suit :

$$A_{n,t+1} = A_{n,t}^{1/2} e^{\gamma \log(A_{n,t}^{-1/2} \hat{A}_{n,t} A_{n,t}^{-1/2})} A_{n,t}^{1/2} = A_{n,t}^{1/2} (A_{n,t}^{-1/2} \hat{A}_{n,t} A_{n,t}^{-1/2})^\gamma A_{n,t}^{1/2}$$

$$\text{avec } \hat{A}_{n,t} = A_{n+1,t}^{1/2} e^{\beta \log(A_{n+1,t}^{-1/2} A_{n-1,t} A_{n+1,t}^{-1/2})} A_{n+1,t}^{1/2} = A_{n+1,t}^{1/2} (A_{n+1,t}^{-1/2} A_{n-1,t} A_{n+1,t}^{-1/2})^\beta A_{n+1,t}^{1/2}$$

$$\beta = \frac{\alpha_{n,t}^+}{(\alpha_{n,t}^+ + \alpha_{n,t}^-)}, \gamma = \frac{\nabla t \alpha_{n,t} (\alpha_{n,t}^+ + \alpha_{n,t}^-)}{\nabla x^2}, \alpha_{n,t} = \left(1 + \frac{d^2(A_{n+1,t}, A_{n-1,t})}{\nabla x^2} \right)^{-1/2}$$

$$\alpha_{n,t}^+ = \left(1 + \frac{d^2(A_{n+1,t}, A_{n,t})}{\nabla x^2} \right)^{-1/2} \text{ et } \alpha_{n,t}^- = \left(1 + \frac{d^2(A_{n,t}, A_{n-1,t})}{\nabla x^2} \right)^{-1/2}$$

Comme dans le cas euclidien $(\beta x + (1-\beta)y)$, on remplace la moyenne pondérée par un barycentre pondéré géodésique : $A \bullet_\beta B = A^{1/2} (A^{-1/2} B A^{-1/2})^\beta A^{1/2}$. On généralise en 2D avec le barycentre de Karcher :

$\hat{A}_{(i,j),t} = \text{Barycentre}(A_{(i-1,j),t}, A_{(i+1,j),t}, A_{(i,j-1),t}, A_{(i,j+1),t})$. On peut également approximer le noyau de la chaleur :

$$p_t((x, A_{n,t}), (x+dx, A_{n+1,t})) \propto e^{-\frac{dx^2 + d^2(A_{n,t}, A_{n+1,t})}{2t}} = e^{-\frac{dx^2}{2t} - \frac{1}{2t} \sum_{k=1}^n \log^2(\lambda_{k,t})}$$

4 Médiane de matrices de covariance

Nous avons vu que le flot de Karcher précédent ne conservait pas la structure Toeplitz. Aussi, nous allons utiliser une paramétrisation autorégressive complexe, pour calculer une médiane qui conserve cette structure. Par le biais de la géométrie affine duale introduite en géométrie de l'information, on définit [6] une géométrie kählerienne en prenant comme fonction potentielle $\tilde{\Phi}(R_n)$ l'entropie du processus, $\mathcal{G}_{ij} \equiv \frac{\partial^2 \tilde{\Phi}}{\partial H_i \partial H_j}$ avec

$$\tilde{\Phi}(R_n) = -\log(\det R_n) - n \log(\pi.e) = \sum_{k=1}^{n-1} (n-k) \cdot \ln[1 - |\mu_k|^2] + n \cdot \ln[\pi.e.P_0]$$

où $\{\mu_k\}_{k=1}^{n-1}$ sont les coefficients de réflexions et P_0 la puissance du signal. En considérant le système de coordonnées suivant $\theta^{(n)} = [P_0 \ \mu_1 \ \dots \ \mu_{n-1}]^T$, la métrique

$$\text{Kählerienne est alors : } d\mathcal{S}_n^2 = n \left(\frac{dP_0}{P_0} \right)^2 + \sum_{i=1}^{n-1} (n-i) \frac{|d\mu_i|^2}{(1-|\mu_i|^2)^2}$$

A la différence de la moyenne de Karcher, la médiane s'obtient comme $\theta_{med} = \arg \min_{\theta} \sum_{k=1}^N d(\theta, \theta_k)$ où d est la

distance riemannienne découlant de la métrique précédente. On suppose que les points de données ne sont pas inclus dans une unique géodésique : dès lors, l'objectif est strictement convexe, et le problème de minimisation admet donc une unique solution sur une boule géodésique. Le gradient s'écrit : $\nabla f = \sum_{k=1}^N \frac{\exp^{-1}(\theta_k)}{d(\theta, \theta_k)}$.

Appliquer un algorithme de descente de gradient pose dans ce cas des questions théoriques de convergence, dues à la non différentiabilité de l'objectif aux points de données ; il est possible de lever ces réserves par régularisation. Pour la simplicité de l'exposé, on suppose que la médiane n'est pas un point de donnée, mais l'approche reste valable aussi dans ce cas. Pour $\varepsilon > 0$, on introduit : $h_\varepsilon(\theta) = \sum_{k=1}^N d(\theta, \theta_k) - \varepsilon \ln(d(\theta, \theta_k) + \varepsilon)$

on a $\nabla h_\varepsilon(\theta) = \sum_{k=1}^N \frac{\exp^{-1}(\theta_k)}{d(\theta, \theta_k) + \varepsilon}$. On peut alors utiliser un

argument technique de Huiling Le [9] pour s'assurer que, pour chaque $\varepsilon > 0$, la suite $\theta^{(\varepsilon)}$ définie par l'algorithme de descente de gradient [9] sur h_ε converge, et que $\lim_{k \rightarrow \infty} \nabla h_\varepsilon(\theta_k^{(\varepsilon)}) = 0$. Prenant maintenant

$\varepsilon = \frac{1}{k}, k \in N^*$, on applique un procédé d'extraction diagonal : pour chaque k , il existe m_k tel que $\|\nabla h_{1/k}(\theta_{m_k}^{(1/k)})\| \leq \frac{1}{k}$. Comme la suite $(\theta_{m_k}^{(1/k)})$ est bornée, on

peut en extraire une suite convergente, de limite θ tel que, par passage à la limite, $\nabla f(\theta) = \sum_{k=1}^N \frac{\exp^{-1}(\theta_k)}{d(\theta, \theta_k)} = 0$.

D'après l'unicité d'un tel θ , on conclut que $(\theta_{m_k}^{(1/k)})$ converge vers θ . En pratique, cependant, il suffit de prendre un pas suffisamment faible dans la descente de

gradient non régularisée pour obtenir la précision souhaitée. Par ailleurs, la métrique concernée découplant chaque coordonnée, on travaille séparément sur chacune d'entre elles. Les termes de puissance sont médianés par la formule close $P_{med} = \exp(\text{med}(\log(P_i)))$ où med désigne la médiane usuelle sur les nombres réels.

Afin de simplifier l'algorithme de calcul pour les coefficients de réflexion, on utilise l'isométrie inverse de Cayley $C^{-1} : D \rightarrow H/\mu \mapsto i \frac{1+\mu}{1-\mu}$. Les géodésiques sont

alors des cercles centrés sur l'axe réel, ce qui permet d'exprimer simplement la carte exponentielle. On peut ainsi appliquer un schéma de descente de gradient similaire à celui du paragraphe 2. On ramène le résultat dans le disque de Poincaré en fin de calcul par C (convergence en quelques dizaines d'itérations).

5 Approximation médiane de modèles multivariés Gaussien de moyenne non nulle

Dans le cas multivarié gaussien de moyenne non-nulle, Calvo et Oller [8] ont donné une forme explicite de la géodésique ; toutefois, celle-ci est peu maniable, et le calcul de la distance géodésique reste implicite. Il est par conséquent difficile d'appliquer ce résultat à la construction de médianes. Calvo & Oller [7] ont introduit le plongement isométrique suivant :

$$R^* \times R^n \times PD_n(R) \rightarrow PD_{n+1}(R), \quad (\beta, \mu, \Sigma) \mapsto \begin{pmatrix} \Sigma + \beta\mu\mu^T & \beta\mu \\ \beta\mu^T & \beta \end{pmatrix}$$

La métrique de l'information sur $PD_{n+1}(R)$ s'écrit alors

$$ds^2 = \left(\frac{d\beta}{\beta} \right)^2 + \beta d\mu^T \Sigma^{-1} d\mu + \frac{1}{2} \text{Tr}((\Sigma^{-1} d\Sigma)^2), \text{ et on obtient bien la}$$

métrique de l'information sur les lois gaussiennes multivariées sur la sous-variété V d'équation $\beta = 1$. Cependant, cette sous-variété n'est pas géodésique : lorsque les moyennes μ diffèrent, la géodésique rejoignant les deux lois gaussiennes dans $PD_{n+1}(R)$ n'est pas incluse dans V . On peut toutefois tirer parti de cette configuration pour calculer une approximation de la médiane entre modèles multivariés gaussiens. En effet, les géodésiques normales à V sont obtenues en faisant varier β paramétré par $\beta = e^{\sqrt{2}s}$ à μ et Σ constants. Ainsi, on peut calculer la médiane de plusieurs lois dans $PD_{n+1}(R)$, puis projeter orthogonalement le résultat obtenu sur V . Il reste à préciser la qualité de cette estimation en pratique.

6 Médiane dans le demi-espace de Siegel

On peut aussi considérer le problème de construire la médiane dans l'espace de Siegel (voir [4]). On identifie en effet le demi-espace de Siegel SH_n constitué des matrices $Z=X+iY$, (X et Y symétriques d'ordre n , Y étant de plus définie positive), avec l'espace $PSp_n = PD_{2n} \cap Sp_n$, muni de la métrique de l'information sur les matrices de PD_{2n} . On utilise la décomposition d'Iwasawa partielle:

$$S = \begin{pmatrix} I & X^T \\ 0 & I \end{pmatrix} \begin{pmatrix} Y & 0 \\ 0 & Y^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ X & I \end{pmatrix} \text{ valable pour tout } S \in PSp_n.$$

Par ailleurs, on dispose d'une expression des géodésiques reliant l'identité à $S \in PSp_n$:

$$T(t) = \exp(tU^Y BU), \text{ avec } S = \exp(U^Y BU), \text{ où } U \in O(2n)$$

Cette décomposition polaire étant toujours valable dans PSp_n , avec la matrice $B = \begin{pmatrix} H & 0 \\ 0 & -H \end{pmatrix}$ où $H = \text{diag}(h_i) \in M_n$.

La longueur géodésique est donnée par $\left(2 \sum_{i=1}^n h_i^2 \right)^{1/2}$.

Dès lors, les points Z et Z_i étant fixés dans SH_n , on utilise d'abord l'action transitive de Sp_n sur SH_n pour amener Z sur iI , les Z_i se trouvant transformés en Z_i' .

On passe alors dans PSp_n par décomposition d'Iwasawa :

iI est transformée en $I \in PSp_n$ et les Z_i' en $S_i \in PSp_n$.

L'application exponentielle inverse est ici le logarithme matriciel et d'après ce qui précède, on connaît l'expression des géodésiques partant de I , donc, des distances géodésiques entre I et les S_i ; on peut donc calculer le gradient donné au paragraphe 4. On obtient l'itéré suivant en prenant l'exponentielle du gradient, ce qui revient encore à se déplacer sur une géodésique partant de I , dont on connaît l'expression. Reste ensuite à ramener ce résultat dans l'espace SH_n de départ. Cela est possible puisque la transformation d'Iwasawa partielle et l'action de Sp_n sur SH_n sont des isométries.

Ce processus peut être itéré, et, appliqué avec un pas suffisamment faible, donner une estimation de la médiane dans le demi-plan de Siegel avec une précision arbitraire. Il pourrait néanmoins fournir une solution exacte au problème des lois multivariées gaussiennes de moyenne non nulle évoqué précédemment si un plongement pouvait être trouvé dans l'espace de Siegel.

7 References

- [1] C.L. Siegel, «Symplectic Geometry», Acad. Press, 1964
- [2] M. Émery, G. Mokobodzki, «Sur le barycentre d'une probabilité dans une variété». Séminaire de probabilités de Strasbourg, n° 25, p. 220-233, 1991
- [3] H. Karcher, «Riemannian center of mass and mollifier smoothing», Comm. Pure Applied Math., °30, pp.509-541, 1977
- [4] A. Terras, «Harmonic Analysis on Symmetric Spaces and Applications II», Springer-Verlag, 1988
- [5] V. Arsigny «Log-Euclidean metrics for fast and simple calculus on diffusion tensors», in Magnetic Resonance in Medicine, Volume 56 Issue 2, Pages 411 - 421, Jun 2006.
- [6] F. Barbaresco, «Interactions between Symmetric Cone and Information Geometries», ETCV'08, Springer Lecture Notes in Computer Science, vol.5416, pp. 124-163, 2009
- [7] M. Calvo, J. Oller, «A distance between multivariate normal distributions based in an embedding into the Siegel Group», Journal of Multivariate Analysis, vol.35, pp. 223-242, 1990
- [8] M. Calvo, J. Oller, «An explicit solution of information geodesic equations for the multivariate normal model», Statistics and Decisions., vol.9, pp. 119-138, 1991
- [9] Huiling Le, «Estimation of Riemannian Barycenters», Proc. London Math. Soc., pp. 193-200, 2004
- [10] Harris, W. F., «The average eye», Ophthalmic Physiol. Opt., n°24, pp. 580-585, 2004
- [11] «Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces» Contemporary mathematics, n°338, 2003