

# Algorithmes stochastiques pour calculer les $p$ -moyennes de mesures de probabilité et géométrie des matrices de covariance Toeplitz

Marc ARNAUDON<sup>1</sup>, Le YANG<sup>1</sup>, Frédéric BARBARESCO<sup>2</sup>

<sup>1</sup>Laboratoire de Mathématiques et Applications, CNRS: UMR 6086, Université de Poitiers, Téléport 2 – BP30179, Boulevard Marie et Pierre Curie, F – 86962 Futuroscope Chasseneuil Cedex, France

<sup>2</sup>Thales Air Systems, Surface Radar, Technical Directorate, Advanced Developments Dept. Voie Pierre-Gilles de Gennes, F-91470 Limours, France

Marc.Arnaudon@math.univ-poitiers.fr

Le.Yang@math.univ-poitiers.fr

frederic.barbaresco@thalesgroup.com

**Résumé** – On développe un algorithme stochastique pour calculer les  $p$ -moyennes d’une mesure de probabilité  $\mu$  dont le support est contenu dans une boule géodésique régulière d’une variété riemannienne. On montre la convergence presque sûre de cet algorithme. De plus, sous une condition de régularité, la vitesse de convergence est donnée par un théorème central limite. Dans la deuxième partie de cet article, on munit les matrices de covariance Toeplitz d’une métrique riemannienne via une fonction potentielle de Kähler et les coefficients de réflexion. On montre qu’avec cette métrique l’espace des matrices de covariance Toeplitz est une variété de Cartan-Hadamard. La distance riemannienne et les géodésiques dans cet espace sont aussi calculées.

**Abstract** – We develop a stochastic algorithm to compute the  $p$ -means of a probability measure whose support is included in a regular geodesic ball of a Riemannian manifold. The almost sure convergence of this algorithm is proved. Moreover, under a regularity condition the convergence rate of the algorithm is given by a central limit theorem. In the second part of this article, we give the Toeplitz covariance matrices a Riemannian metric by using a Kähler potential function and the reflection coefficients. We show that with this metric the space of Toeplitz covariance matrices is a Cartan-Hadamard manifold. The Riemannian distance and the geodesics are also computed.

## 1 Introduction

L’analyse statistique des données est une méthode importante dans la théorie et la pratique du traitement de l’image et du signal. Un problème souvent rencontré lorsqu’on fait l’analyse statistique d’une population de données est de calculer leur centre, dont la définition est choisie à l’avance en fonction du problème à résoudre. La notion de centralité la plus usuelle est celle du barycentre. C’est le point qui minimise la somme des carrés des distances aux points de données. Le barycentre d’une mesure de probabilité dans une variété riemannienne a été très bien étudié depuis ces trente dernières années, voir par exemple [6], [7], [5]. Un algorithme par descente de gradient pour calculer le barycentre est proposé par Le dans son article [9].

Cependant, le barycentre est non robuste et sensible au bruit, ce qui le rend inadéquat dans les situations avec beaucoup de valeurs aberrantes. C’est pourquoi on introduit la notion de  $p$ -moyenne, qui fournit un estimateur statistique plus flexible et plus robuste que le barycentre lorsque  $p < 2$ . L’existence et l’unicité des  $p$ -moyennes sont montrées par Afsari dans [1]. Dans le cas où  $p = 1$  on retrouve la notion de la médiane. Un algorithme de sous-gradient pour calculer la médiane peut être trouvé dans [10].

Afin d’effectuer l’analyse statistique des données dans les

cas pratiques, il est nécessaire de trouver une façon efficace de calculer les  $p$ -moyennes. On considère cette question pour des données à valeurs dans une variété riemannienne, qui est un cadre suffisamment général pour la plupart des applications, et on donne une réponse en y introduisant un algorithme stochastique qui converge presque sûrement vers les  $p$ -moyennes. L’avantage de cet algorithme est que l’on n’a pas besoin de calculer le gradient de la fonction à minimiser. L’idée est de se diriger vers la direction d’un point dans le support de  $\mu$  à chaque étape. Ce point est choisi d’une façon aléatoire selon  $\mu$ . De plus, le pas de déplacement est une fonction bien choisie de la distance à ce point,  $p$  et le nombre de l’étape. Des simulations de cet algorithme seront aussi données.

Ensuite, la vitesse de convergence sera étudiée. On montrera que la chaîne de Markov inhomogène renormalisée associée à l’algorithme stochastique converge en loi vers un processus de diffusion inhomogène. C’est un résultat de type du principe d’invariance.

Les matrices de covariance Toeplitz ont une importance fondamentale dans le traitement de signaux stationnaires. Il arrive souvent qu’une observation soit la réalisation d’un processus autorégressif dont on estime les covariances. Avec la métrique de type Bergman proposée dans [2], dont l’indépendance quant

au choix du système de coordonnées est montrée dans [8], on déduit que les matrices de covariance Toeplitz forment une variété de Cartan-Hadamard. Si on veut comparer, moyenner ou filtrer les observations, il est nécessaire de calculer des  $p$ -moyennes. Afin d'obtenir ceci en utilisant l'algorithme stochastique, on est obligé d'étudier la géométrie riemannienne de l'espace des matrices de covariance Toeplitz pour connaître la distance riemannienne et les géodésiques entre deux telles matrices.

## 2 Algorithmes stochastiques et théorème central limite

Les détails mathématiques de cette section peuvent être trouvés dans l'article [3].

### 2.1 $p$ -moyennes dans une boule géodésique régulière

Soit  $M$  une variété riemannienne avec la distance riemannienne  $\rho$  et le rayon d'injectivité  $\text{inj}(M)$ . Supposons que les courbures sectionnelles  $K(\sigma)$  vérifient  $-\beta^2 \leq K(\sigma) \leq \alpha^2$ . Soit  $\mu$  une mesure de probabilité dont le support est contenu dans un sous-ensemble compact  $K$  d'une boule géodésique régulière  $B(a, r) \subset M$ . Fixons  $p \in [1, \infty)$ . On va toujours supposer que l'hypothèse suivante est vérifiée:

**Hypothèse 1.** *Le support de  $\mu$  n'est pas réduit à un point. Si  $p = 1$ , le support de  $\mu$  n'est pas contenu dans une seule géodésique. Le rayon  $r$  satisfait*

$$r < r_{\alpha,p} = \begin{cases} \frac{1}{2} \min \left\{ \text{inj}(M), \frac{\pi}{2\alpha} \right\}, & \text{si } p \in [1, 2); \\ \frac{1}{2} \min \left\{ \text{inj}(M), \frac{\pi}{\alpha} \right\}, & \text{si } p \in [2, \infty). \end{cases}$$

Sous l'Hypothèse 1, il est montré dans [1] que la fonction

$$H_p : M \longrightarrow \mathbb{R}_+, x \longmapsto \int_M \rho^p(x, y) \mu(dy)$$

a un seul minimiseur  $e_p$ , la  $p$ -moyenne de  $\mu$ , et de plus  $e_p \in B(a, r)$ . Si  $p = 1$ ,  $e_1$  est la médiane de  $\mu$ .

Il est facile de voir que si  $p \in [1, 2)$ , alors  $H_p$  est strictement convexe sur  $B(a, r)$ . De plus, si  $p \geq 2$ , alors  $H_p$  est de classe  $C^2$  sur  $B(a, r)$ .

### 2.2 Algorithmes stochastiques pour calculer les $p$ -moyenne

Maintenant on peut énoncer le résultat principal dans le théorème suivant: on définit un algorithme de gradient stochastique pour approximer les  $p$ -moyennes.

**Théorème 2.1.** *Soient  $(P_k)_{k \geq 1}$  une suite de variables aléatoires indépendantes de loi  $\mu$  à valeurs dans  $B(a, r)$  et  $(t_k)_{k \geq 1}$  une suite de nombres strictement positifs telle que*

$$t_k \leq C_{p,r,\mu,K}, \quad \sum_{k=1}^{\infty} t_k = +\infty \quad \text{et} \quad \sum_{k=1}^{\infty} t_k^2 < \infty,$$

où  $C_{p,r,\mu,K}$  est une certaine constante. Fixons  $x_0 \in B(a, r)$ , on définit inductivement une marche aléatoire  $(X_k)_{k \geq 0}$  par

$$X_k = \begin{cases} x_0, & \text{si } k = 0; \\ \exp_{X_{k-1}}(-t_k \text{grad}_{X_{k-1}} F_p(\cdot, P_k)), & \text{si } k \geq 1, \end{cases}$$

où  $F_p(x, y) = \rho^p(x, y)$ , avec la convention  $\text{grad}_x F_p(\cdot, x) = 0$ . Alors la marche aléatoire  $(X_k)_{k \geq 1}$  converge dans  $L^2$  et presque sûrement vers la  $p$ -moyenne  $e_p$ .

Afin d'illustrer le Théorème 2.1, considérons le cas où  $M = \mathbb{R}^d$  et  $\mu$  une mesure de probabilité de support compact dans  $\mathbb{R}^d$ . Supposons que  $p = 2$ , alors  $e_2 = \mathbf{E}[P_1]$  et  $\text{grad}_x F_p(\cdot, y) = 2(x - y)$ . Maintenant, en prenant  $t_k = \frac{1}{2k}$  on obtient

$$X_k = \frac{1}{k} \sum_{j=1}^k P_j.$$

Par conséquent, dans ce cas simple, le Théorème 2.1 n'est rien d'autre que la loi forte des grands nombres.

Dans la plupart des cas pratiques, la mesure de probabilité  $\mu$  est la loi uniforme sur un nombre fini de points de données différents  $q_1, \dots, q_N$ . Afin de faire marcher l'algorithme dans ce cas il suffit à chaque fois de tirer un point  $q_i$  de façon aléatoire et uniforme dans l'ensemble  $\{q_1, \dots, q_N\}$ , puis de se déplacer le long de la géodésique joignant l'état actuel  $x_k$  et le point de données  $q_i$  d'une longueur  $t_k p \rho^{p-1}(x_k, q_i)$  afin d'arriver à l'état suivant  $x_{k+1}$ . De plus, on peut arrêter l'algorithme à l'étape  $n$  si la critère d'arrêt suivante est vérifiée:

$$\left| \frac{p}{N} \sum_{\substack{i=1 \\ q_i \neq x_n}}^N \rho^{p-2}(x_n, q_i) \exp_{x_n}^{-1} q_i \right| < \varepsilon,$$

où  $\varepsilon > 0$  est une précision fixée a priori.

## 2.3 Simulations des algorithmes

### 2.3.1 La médiane de la mesure uniforme sur un triangle équilatéral dans le plan

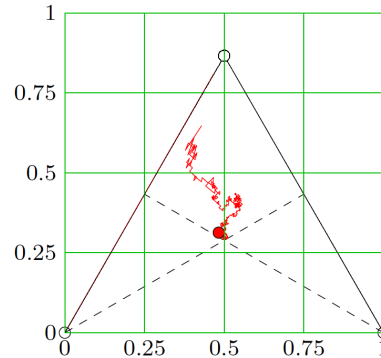


FIG. 1: Médiane d'un triangle régulier

Ici  $M$  est le plan  $\mathbb{R}^2$  et  $\mu$  est la mesure uniforme sur un triangle équilatéral. Le chemin rouge représente la trajectoire de la chaîne de Markov inhomogène  $(X_k)_{k \geq 0}$  avec  $p = 1$ . Le point rouge est la médiane  $e_1$ .

### 2.3.2 La médiane d'une mesure non uniforme sur le carré unité dans le plan

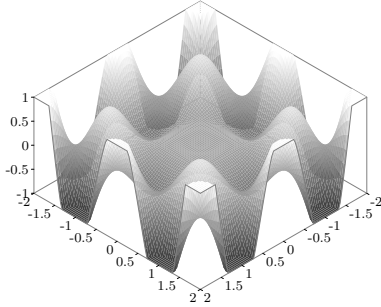


FIG. 2: Fonction de densité sur  $[0, 1] \times [0, 1]$

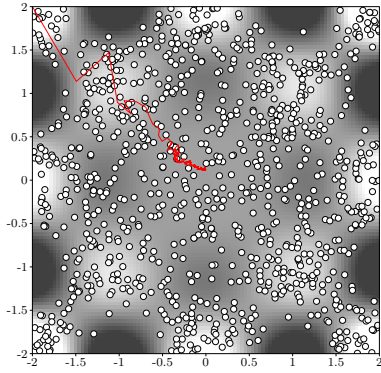


FIG. 3: Médiane de la mesure à densité

Ici  $M$  est le plan  $\mathbf{R}^2$  et  $\mu$  est une mesure à densité donnée par FIG. 2. Le chemin rouge dans FIG. 3 représente la trajectoire de la chaîne de Markov inhomogène  $(X_k)_{k \geq 0}$  avec  $p = 1$ . Le point rouge est la médiane  $e_1$ . Les cercles noirs représentent les valeurs  $(P_k)_{k \geq 1}$ .

### 2.3.3 Une mesure non uniforme sur la sphère $S^2$

Dans FIG. 4,  $M$  est la sphère  $S^2$  et  $\mu$  est une mesure non uniforme. Le chemin rouge représente la trajectoire de la chaîne de Markov inhomogène  $(X_k)_{k \geq 0}$  avec  $p = 1$ . Le point rouge est la médiane  $e_1$ . Les cercles noirs représentent les valeurs  $(P_k)_{k \geq 1}$ . On constate que même si la convexité n'existe plus, il y a toujours convergence.

## 2.4 Théorème central limite

La vitesse de convergence de l'algorithmes stochastique est donnée par le théorème suivant.

**Théorème 2.2.** Soit  $(X_k)_{k \geq 0}$  la chaîne de Markov inhomogène en temps à valeurs dans  $M$  définie dans le Théorème 2.1 avec  $t_k = \min\left(\frac{\delta}{k}, C_{p,r,\mu,K}\right)$  pour un certain  $\delta > 0$ . On définit pour

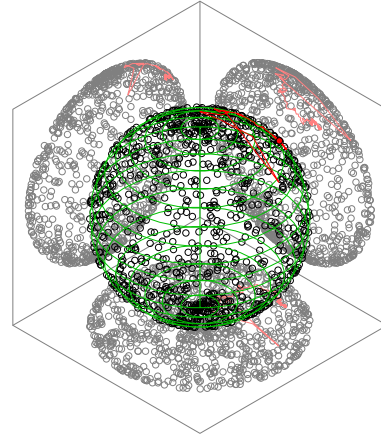


FIG. 4: Médiane sur la sphère

$n \geq 1$  la chaîne de Markov  $(Y_k^n)_{k \geq 0}$  à valeurs dans  $T_{e_p}M$  par

$$Y_k^n = \frac{k}{\sqrt{n}} \exp_{e_p}^{-1} X_k.$$

Supposons que  $H_p$  est  $C^2$  dans un voisinage de  $e_p$  et que  $\delta > C'_{p,\mu,K}$  qui est une certaine constante. Alors la suite de processus  $(Y_{[nt]}^n)_{t \geq 0}$  converge faiblement dans  $\mathbb{D}((0, \infty), T_{e_p}M)$  vers un processus de diffusion  $y_\delta$  donné par

$$y_\delta(t) = \sum_{i=1}^d t^{1-\delta\lambda_i} \int_0^t s^{\delta\lambda_i-1} \langle \delta\sigma dB_s, e_i \rangle e_i, \quad t \geq 0,$$

où  $B_t$  est le mouvement brownien standard dans  $T_{e_p}M$  et  $\sigma \in \text{End}(T_{e_p}M)$  satisfait

$$\sigma\sigma^* = \mathbb{E} \left[ \text{grad}_{e_p} F_p(\cdot, P_1) \otimes \text{grad}_{e_p} F_p(\cdot, P_1) \right],$$

$(e_i)_{1 \leq i \leq d}$  est une base orthonormée diagonalisant la forme bilinéaire symétrique  $\nabla dH_p(e_p)$  et  $(\lambda_i)_{1 \leq i \leq d}$  sont les valeurs propres associées.

Dans le cas où  $M = \mathbf{R}^d$ ,  $Y_k^n = \frac{k}{\sqrt{n}}(X_k - e_p)$  et l'espace tangent  $T_{e_p}M$  est identifié à  $\mathbf{R}^d$ . En particulier, si  $t = 1$  alors Théorème 2.2 dit que  $\sqrt{n}(X_n - e)$  converge vers une loi gaussienne de dimension  $d$ . C'est un théorème central limite: les fluctuations de l'algorithme stochastique sont asymptotiquement gaussiennes.

## 3 Géométrie riemannienne des matrices de covariance Toeplitz

Soit  $\mathcal{T}_n$  l'ensemble des matrices de covariance Toeplitz d'ordre  $n$ . Chaque élément  $R_n \in \mathcal{T}_n$  s'écrit comme

$$R_n = \begin{bmatrix} r_0 & \bar{r}_1 & \dots & \bar{r}_{n-1} \\ r_1 & r_0 & \dots & \bar{r}_{n-2} \\ \vdots & \ddots & \ddots & \vdots \\ r_{n-1} & \dots & r_1 & r_0 \end{bmatrix}.$$

### 3.1 Changement de Coordonnées

Pour chaque  $1 \leq k \leq n-1$ , le bloc de taille  $(k+1) \times (k+1)$  en haut à gauche de  $R_n$  est noté par  $R_k$ , qui est associé à un modèle autorégressif d'ordre  $k$  dont l'équation de Yule-Walker est

$$\begin{bmatrix} r_0 & \bar{r}_1 & \dots & \bar{r}_k \\ r_1 & r_0 & \dots & \bar{r}_{k-1} \\ \vdots & \ddots & \ddots & \vdots \\ r_k & \dots & r_1 & r_0 \end{bmatrix} \begin{bmatrix} 1 \\ a_1^{(k)} \\ \vdots \\ a_k^{(k)} \end{bmatrix} = \begin{bmatrix} P_k \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

où  $a_1^{(k)}, \dots, a_k^{(k)}$  sont les coefficients autorégressifs,  $P_k = \det R_{k+1} / \det R_k$  est l'erreur de prédiction.

Le dernier coefficient autorégressif  $a_k^{(k)}$  est appelé le  $k$ -ième coefficient de réflexion et noté par  $\mu_k$ . Il est facile de voir que  $\mu_1, \dots, \mu_{n-1}$  sont uniquement déterminés par la matrice  $R_n$ . De plus, la récursion classique de Levinson implique que  $|\mu_k| < 1$ . Par conséquent, en prenant  $P_0 = r_0$ , on obtient une application entre deux sous variétés de  $\mathbf{R}^{2n-1}$ :

$$\varphi: \mathcal{T}_n \longrightarrow \mathbf{R}_+^* \times \mathbf{D}^{n-1}, \quad R_n \longmapsto (P_0, \mu_1, \dots, \mu_{n-1}),$$

où  $\mathbf{D} = \{z \in \mathbf{C} : |z| < 1\}$  est le disque du plan complexe. Cette application est un difféomorphisme, donc elle est une reparamétrisation des matrices de covariance Toeplitz par les coefficients de réflexion. L'avantage d'utiliser cette paramétrisation est que la métrique riemannienne que l'on utilise admet une forme très simple.

### 3.2 Métrique riemannienne de $\mathcal{T}_n$

Maintenant, on peut voir  $\mathcal{T}_n$  comme une variété riemannienne dont la métrique riemannienne, qui est introduite dans [4] via la hessienne de la fonction potentielle suivante proposée par [2],

$$\Phi(R_n) = -\ln(\det R_n),$$

est donnée par

$$ds^2 = n \frac{dP_0^2}{P_0^2} + \sum_{k=1}^{n-1} (n-k) \frac{|d\mu_k|^2}{(1-|\mu_k|^2)^2},$$

où  $(P_0, \mu_1, \dots, \mu_{n-1}) = \varphi(R_n)$ . D'après [8], cette métrique est indépendante du changement de coordonnées.

Avec cette métrique, l'espace  $\mathbf{R}_+^* \times \mathbf{D}^{n-1}$  est le produit des variétés riemanniennes  $(\mathbf{R}_+^*, ds_0^2)$  et  $(\mathbf{D}, ds_k^2)_{1 \leq k \leq n-1}$ , où  $ds_0^2 = nP_0^{-2}/dP_0^2$  et  $ds_k^2 = (n-k)|d\mu_k|^2/(1-|\mu_k|^2)^2$  qui est simplement  $n-k$  fois la métrique de Poincaré classique sur  $\mathbf{D}$ . Par conséquent,  $(\mathbf{R}_+^* \times \mathbf{D}^{n-1}, ds^2)$  est une variété de Cartan-Hadamard. De plus, la distance riemannienne entre deux points différents  $x$  et  $y$  dans  $\mathbf{R}_+^* \times \mathbf{D}^{n-1}$  est donnée par

$$d(x, y) = \left( n\sigma(P, Q)^2 + \sum_{k=1}^{n-1} (n-k)\tau(\mu_k, \nu_k)^2 \right)^{1/2},$$

où  $x = (P, \mu_1, \dots, \mu_{n-1})$ ,  $y = (Q, \nu_1, \dots, \nu_{n-1})$ ,

$$\sigma(P, Q) = \left| \ln \left( \frac{Q}{P} \right) \right| \quad \text{et} \quad \tau(\mu_k, \nu_k) = \frac{1}{2} \ln \frac{1 + \frac{|\nu_k - \mu_k|}{1 - \bar{\mu}_k \nu_k}}{1 - \frac{|\nu_k - \mu_k|}{1 - \bar{\mu}_k \nu_k}}.$$

### 3.3 Géodésiques dans $\mathcal{T}_n$

Avec les notations précédentes, la géodésique de  $x$  à  $y$  paramétrisée par la longueur d'arc est donnée par  $\gamma(s, x, y) =$

$$\left( \gamma_0 \left( \frac{\sigma(P, Q)}{d(x, y)} s \right), \gamma_1 \left( \frac{\tau(\mu_1, \nu_1)}{d(x, y)} s \right), \dots, \gamma_1 \left( \frac{\tau(\mu_{n-1}, \nu_{n-1})}{d(x, y)} s \right) \right).$$

où  $\gamma_0$  est la géodésique dans  $(\mathbf{R}_+^*, ds_0^2)$  de  $P$  à  $Q$  paramétrisée par la longueur d'arc et pour  $1 \leq k \leq n-1$ ,  $\gamma_k$  est la géodésique dans  $(\mathbf{D}, ds_k^2)$  de  $\mu_k$  à  $\nu_k$  paramétrisée par la longueur d'arc.

Plus précisément,

$$\gamma_0 \left( \frac{\sigma(P, Q)}{d(x, y)} s \right) = P e^{\frac{(\text{sign}(Q-P))\sigma(P, Q)}{d(x, y)} s},$$

et pour  $1 \leq k \leq n-1$ ,

$$\gamma_k \left( \frac{\tau(\mu_k, \nu_k)}{d(x, y)} s \right) = \frac{(\mu_k + e^{i\theta_k}) e^{\frac{2\tau(\mu_k, \nu_k)}{d(x, y)} s} + (\mu_k - e^{i\theta_k})}{(1 + \bar{\mu}_k e^{i\theta_k}) e^{\frac{2\tau(\mu_k, \nu_k)}{d(x, y)} s} + (1 - \bar{\mu}_k e^{i\theta_k})},$$

avec

$$\theta_k = \arg \frac{\nu_k - \mu_k}{1 - \bar{\mu}_k \nu_k}.$$

## Références

- [1] B. Afsari, *Riemannian  $L^p$  center of mass : existence, uniqueness, and convexity*, Proceedings of the American Mathematical Society, S 0002-9939(2010)10541-5, Article electronically published on August 27, 2010.
- [2] S. Amari and A. Cichocki, *Information geometry of divergence functions*, Bulletin of the Polish Academy of Science, Technical Science, Vol 58 (2010), no. 1, 183-195
- [3] M. Arnaudon, C. Dombry, A. Phan and L. Yang, *Stochastic algorithms for computing means of probability measures*, preprint hal-00540623, version 1, (2010). À paraître dans Stochastic Processes and their Applications.
- [4] Barbaresco F. *Interactions between Symmetric Cone and Information Geometries*, ETVC'08, Springer Lecture Notes in Computer Science, vol. 5416, pp. 124-163, 2009
- [5] M. Emery and G. Mokobodzki, *Sur le barycentre d'une probabilité dans une variété*, Séminaire de Probabilités XXV, Lecture Notes in Mathematics 1485 (Springer, Berlin, 1991), pp. 220-233
- [6] H. Karcher, *Riemannian center of mass and mollifier smoothing*, Communications on Pure and Applied Mathematics, vol XXX (1977), 509-541
- [7] W.S. Kendall, *Probability, convexity and harmonic maps with small image I: uniqueness and fine existence*, Proc. London Math. Soc. (3) 61 no. 2 (1990) pp. 371-406
- [8] J. L. Koszul, *Sur la forme hermitienne canonique des espaces homogènes complexes*, Canadian Journal of Math., vol 7, no 4 (1955), pp. 562-576
- [9] H. Le, *Estimation of Riemannian barycentres*, LMS J. Comput. Math. 7 (2004), pp. 193-200
- [10] L. Yang, *Riemannian median and its estimation*, LMS J. Comput. Math. Vol. 13, 461-479, (2010).