Semaine d'Etude Mathématiques et Entreprises 1 : Géométrie des matrices de covariance pour le traitement de signaux radars

Etienne Bernard, Aurélien Bosche, Nicolas Charon, Salima El Kolei, Julie Lapebie, Etienne Le Masson, Jean-Marie Mirebeau, Thomas Richard

Semaine d'Etude Maths–Entreprises 1

4–8 avril 2011, Institut Henri Poincaré (Paris)

Géométrie des matrices de covariance pour le traitement de signaux radars

E. Bernard[a] A. Bosche[b]
N. Charon[c] S. Kolei[d]
J. Lapebie[e] E. Le Masson[f]
J-M. Mirebeau[g] T. Richard[b]

[a] *CMLS, Ecole Polytechnique, 91128 Palaiseau, France*
[b] *Institut Fourier, Université Grenoble I, 38402 St Martin d'Hères, France*
[c] *CMLA, ENS Cachan, 94235 Cachan, France*
[d] *Laboratoire J.A. Dieudonné, Université de Nice Sophia-Antipolis, 06108 Nice, France*
[e] *Master 2 Mathématiques et applications, Université de Provence, Marseilles, France*
[f] *Laboratoire de Mathématiques d'Orsay, Université Paris-Sud, 91405 Orsay, France*
[g] *Laboratoire Jacques-Louis Lions, UPMC, 75005 Paris, France*

Sujet proposé par :

THALES

Correspondant : F. Barbaresco (Thalès)

MATHS & ENTREPRISES

cnrs dépasser les frontières

Résumé

Les radars Doppler permettent de détecter des objets volants petits ou de faible signature radar. Thalès propose ici de réfléchir aux techniques permettant de faire ressortir de la masse des données radar celles qui sont "aberrantes" afin de repérer parmi le bruit de fond dû aux milieux environnants (nuages de pluie,...) la trace d'un objet volant.

Numéro de publication : SEME001-2011-04-D

Introduction

Le sujet d'étude proposé par F. Barbaresco porte sur les radars Doppler, et son objectif est le suivant : améliorer si possible la détection grâce à de tels radars d'objets (volants) petits ou de faible signature radar. En d'autres termes il s'agit, parmi la masse de données générées par un radar Doppler, de faire ressortir du bruit de fond les données qui sont "aberrantes" et donc correspondent à la détection d'un objet. Notre étude n'est pas fondée sur des données réelles rentrant dans ce cadre, car celles-ci sont classifiées. Nous avons travaillé sur des données simulées, des modèles mathématiques, et des données de radar météorologique.

La première partie de ce rapport est consacrée à la description d'un radar Doppler. Elle ne comporte pas d'apport personnel, mais permet de préciser le cadre de notre étude. La seconde partie décrit une première approche de la résolution du problème posé, par des méthodes fondées sur la nature statistique d'un signal radar. La troisième partie décrit, dans un cadre mathématique abstrait et général, deux méthodes d'identification de données aberrantes qui pourraient être appliquées au problème posé. Ces méthodes et algorithmes sont liés à d'autres questions importantes de mathématiques appliquées : les théories dites de persistance topologique [5], et l'algorithme PageRank [8] de Google respectivement. La dernière partie est consacrée à l'étude, par les méthodes de la géométrie différentielle, de certains espaces de matrices liés au problème étudié.

1 Principe d'un radar Doppler

Rappelons dans un premier temps le fonctionnement d'un radar standard (non Doppler). Nous considérons un émetteur et un récepteur d'ondes électromagnétiques, placés à la même position que l'on note par convention 0. A un temps T, l'émetteur génère une impulsion électromagnétique, dans une direction θ_0. Un objet présent dans cette direction, dont la position sera notée $x_0(t)$, reçoit cette onde et en réfléchit une certaine proportion vers le récepteur. La réception d'une impulsion électromagnétique par le radar, au temps $T + \Delta T$ permet à son opérateur d'affirmer qu'un objet est présent dans la direction θ_0, et d'en évaluer la distance grâce à la formule

$$r_0 = |x_0(t)| = c\Delta T/2,$$

où c désigne la vitesse de la lumière.

Dans le cas d'un radar Doppler, les ondes électromagnétiques émises et reçues sont oscillantes, et leur phase et leurs fréquences jouent un rôle important. Pour simplifier l'exposition nous nous focaliserons sur la *phase complexe* de ces ondes, et nous mettrons de côté leur amplitude et leur polarisation.

La phase complexe de l'onde émise au temps t est $\exp(j\omega t)$ (où nous notons j le nombre imaginaire pur unité, $j^2 = -1$, conformément à la convention physique). Compte tenu du temps de propagation de la lumière, la phase complexe de l'onde reçue par l'objet au temps t est

$$\exp(j\omega(t - |x_0(t)|/c)).$$

Compte tenu du temps de retour de la lumière vers le radar, celui-ci reçoit en première approximation une onde réfléchie de phase complexe

$$\exp(j\omega(t - 2|x_0(t)|/c)). \tag{1}$$

Cette approximation sera suffisante dans le cadre de notre étude, mais on pourrait être plus précis en prenant en compte le déplacement de l'objet entre les temps $t - 2|x_0(t)|/c$ et t, et éventuellement des effets relativistes.

Les radars comportent des circuits de *démodulation*, qui permettent essentiellement de multiplier le signal reçu (1) par $\exp(-j\omega t)$, et donc d'identifier la contribution $\exp(-2j\omega|x_0(t)|/c)$ de l'objet à la phase. En présence de plusieurs objets, de positions $x_1(t)$, ..., $x_m(t)$ proches du point de coordonnées polaires (θ_0, r_0) visé par le radar, et d'une source de bruit ξ, les contributions s'additionnent et le signal reçu prend la forme

$$f(t) = \sum_{1\leq k\leq m} \exp(-2j\omega|x_k(t)|/c) + \xi(t).$$

On peut écrire en première approximation

$$|x_k(t)| \simeq u_k + tv_k,$$

où v_k désigne la vitesse radiale de la k-ième cible. Dans cette approximation, le signal reçu prend donc la forme

$$f(t) \simeq \sum_{1\leq k\leq m} \exp(-2j\omega(u_k + tv_k)/c) + \xi(t). \tag{2}$$

Ce signal est échantillonné à une série de temps généralement régulièrement espacés : $T,\ T+\delta T,\ \cdots, T+(n-1)\delta T$. Nous posons donc pour tout $0 \leq i \leq n-1$,

$$z_i^0 = f(T + i\delta T). \tag{3}$$

Rappelons que le vecteur $z^0 = (z_i^0)_{0\leq i\leq n-1} \in \mathbb{C}^n$ correspond à une observation du radar dans la direction θ_0, à la distance r_0. Une autre valeur des paramètres (θ, r) donnerait un autre vecteur $z \in \mathbb{C}^n$. Compte tenu de (2), les coefficients de z^0 ont la forme

$$z_i^0 \simeq \sum_{1\leq k\leq m} \lambda_k \exp(-2ij\omega v_k \delta T/c) + \xi_i, \tag{4}$$

où $\lambda_k = \exp(-2j\omega(u_k + v_k T)/c)$. Finalement, on procède à l'affichage du spectre Doppler, défini pour tout $s \in \mathbb{R}$ par

$$I(s) = \left| \sum_{0\leq i\leq n} z_i^0 \exp(ijs) \right|.$$

Compte tenu de (4), ce spectre présente des pics (cf Fig 1) aux valeurs $s_k = 2\omega v_k \delta T/c$, ce qui permet d'identifier les vitesses $v_1, \cdots, v_k$ des objets présents aux voisinage du point de coordonnées (r_0, θ_0).

Pour des données réelles on a typiquement $n = 8$ (donc $z^0 \in \mathbb{C}^8$), en d'autres termes la fonction f est échantillonnée à 8 temps différents, et $0 \leq m \leq 2$, en d'autres termes il n'y a au voisinage du point de coordonnées polaires (θ_0, r_0) visé par le radar, que deux objets au plus de vitesses différentes.

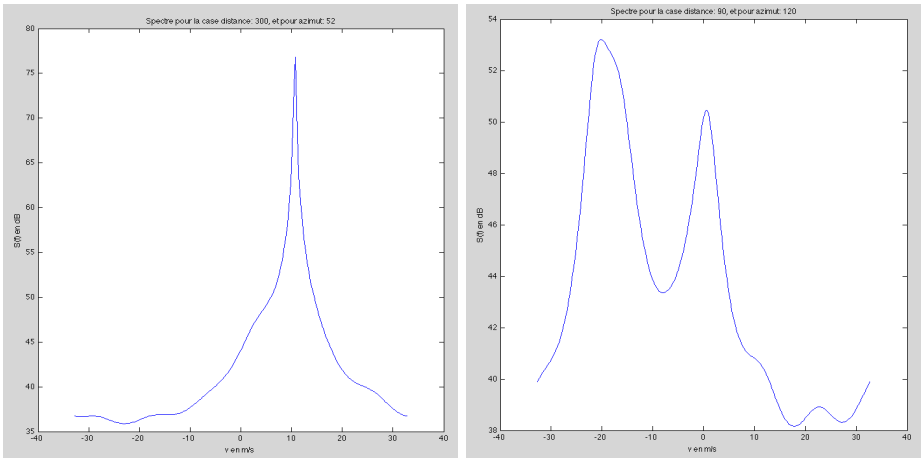
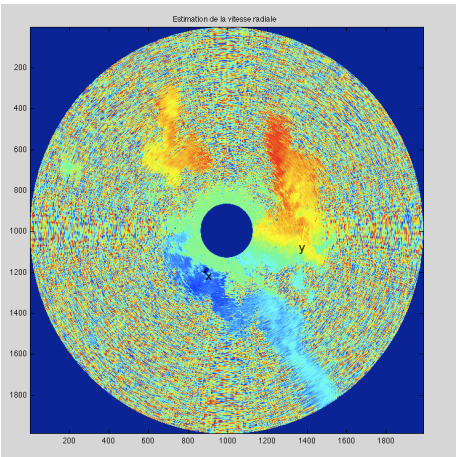


FIGURE 1 – Spectre Doppler issu de données météorologiques pour deux jeux de coordonnées polaires (θ, r). Le ou les pics donnent les vitesses radiales des objets, ici des nuages, présents aux voisinages de ces coordonnées. Image inférieure : vitesse radiale associée au pic principal, en code couleur, pour toute la zone couverte par le radar : $r_0 \leq r \leq R_0$ et $0 \leq \theta \leq 2\pi$. (Données et programmes fournis par F. Barbaresco.)



2 Approche Statistique

Dans ce paragraphe, nous supposons donnés un ensemble fini de vecteurs $z^u \in \mathbb{C}^n$, $u \in U$. Chaque vecteur z^u correspond aux données issues du radar pour un point de coordonnées (θ_u, r_u).

Notre objectif est le suivant : déterminer si ces données radar correspondent à l'observation des mêmes objets (par exemple des nuages étendus et de vitesse fixe), où si l'une d'entre elles trahit la présence d'un objet supplémentaire, petit, de vitesse différente et de signature radar potentiellement faible (par exemple un missile). S'il existe, nous notons u_0 l'indice correspondant.

Les vecteurs z^u, $u \in U \setminus \{u_0\}$, correspondent à la détection des mêmes objets. Leurs coordonnées prennent donc, d'après (4), la forme

$$z_i^u \simeq \sum_{1 \leq k \leq m} \lambda_k^u \exp(-ijs_k) + \xi_i^u,$$

où $s_k := 2\omega v_k \delta T/c \in \mathbb{R}$. Notez que les coefficients λ_k^u sont a-priori différents pour chaque observation. Définissons un polynôme $P \in \mathbb{C}[X]$ et des coefficients $a_1, \cdots, a_m \in \mathbb{C}$ par l'égalité

$$P(X) = \prod_{1 \leq k \leq m} (X - \exp(-js_k)) = X^m - a_1 X^{m-1} - a_2 X^{m-2} - \cdots - a_m.$$

Si $\lambda_1, \cdots, \lambda_k \in \mathbb{C}$ sont des nombres complexes donnés, et si $(y_i)_{i \in \mathbb{Z}}$ est la suite définie par

$$y_i = \sum_{1 \leq k \leq m} \lambda_k \exp(-ijs_k),$$

alors nous avons par construction pour tout $i \in \mathbb{Z}$

$$y_i = a_1 y_{i-1} + \cdots + a_m y_{i-m}.$$

Ceci suggère de modéliser les vecteurs z^u, $u \in U \setminus \{u_0\}$, comme les réalisations d'un processus auto-régressif stationnaire, de la forme

$$Z_i = a_1 Z_{i-1} + \cdots + a_m Z_{i-m} + \xi_i, \tag{5}$$

où ξ_i désigne un bruit gaussien. Cette modélisation est pertinente si la longueur n des vecteurs z^u est "grande" devant l'ordre m du modèle. On peut supposer que c'est le cas dans la situation étudiée, pour laquelle $n = 8$ et $0 \leq m \leq 2$ typiquement.

Pour tester le modèle (5), et déterminer s'il existe un indice u_0 aberrant, on pourra utiliser la méthode suivante, dont les idées directrices ont été suggérées par Josselin Garnier.

Si l'hypothèse (5) est satisfaite alors les vecteurs z^u, qui ne correspondent pas à des observations aberrantes, sont des réalisations indépendantes d'un vecteur gaussien à valeurs dans $\mathbb{C}^n$ centré mais de matrice de covariance M inconnue. En effet cette matrice de covariance dépend des coefficients $(a_k)_{1 \leq k \leq m}$, qui dépendent eux mêmes des vitesses $(v_k)_{1 \leq k \leq m}$, qui sont a-priori inconnues. Nous introduisons donc la matrice de covariance empirique N, dont les coefficients sont

$$N_{ij} := \frac{1}{\#(U)} \sum_{u \in U} z_i^u z_j^u.$$

Cette matrice est hermitienne positive, et génériquement définie positive si $\#(U) \geq n$ ce que nous supposerons. Nous définissons par ailleurs les points

$$y^u := N^{-\frac{1}{2}} z^u.$$

Si le modèle (5) était vérifié, et si l'on avait $N = M$, alors les points $y^u \in \mathbb{C}^n$, qui ne correspondent pas à des observations aberrantes, seraient des réalisations indépendantes d'un vecteur gaussien centré de matrice de covariance identité.

Nous avons construit la matrice N et les vecteurs $(y^u)_{u\in S}$ dans le cas de certaines données de radar météorologique fournies par F. Barbaresco. Nous avons utilisé les tests de normalité disponibles, comme le test de Kolgomorov-Smirnov, et obtenu que les vecteurs $(y^u)_{u\in U}$ étaient compatibles avec des réalisations indépendantes d'un vecteur gaussien centré de matrice de covariance identité.

Nous proposons un second critère qui permet d'identifier une donnée aberrante potentielle, et de déterminer si elle est statistiquement significative. Comme la densité de probabilité d'un vecteur Gaussien centré et de matrice de covariance identité est radiale et radialement décroissante, une observation aberrante parmi les $(y^u)_{u\in U}$ se trahit typiquement par une norme excessivement grande. Nous posons donc

$$m := \max_{u\in U} |y^u|,$$

et nous désignons comme potentiellement aberrante la donnée y^{u_0} telle que $m = |y^{u_0}|$. Pour savoir si elle est statistiquement significative nous remarquons que si $Y_1, \cdots, Y_N \in \mathbb{C}^n$, $N = \#(U)$, désignent des réalisations indépendantes d'un vecteur gaussien centré, complexe et de matrice de covariance identité, alors

$$1 - P\left(\max_{1\leq i\leq N} |Y_i| \geq m\right) = P\left(\max_{1\leq i\leq N} |Y_i| < m\right) = P(|Y_1| < m)^N = P\left(X_{2n} < m^2\right)^N \quad (6)$$

où X_{2n} désigne une variable aléatoire suivant la loi $\chi^2(2n)$, dite du "chi-deux à $2n$ degrés de liberté" (car $\mathbb{C}^n$ est de dimension $2n$ sur $\mathbb{R}$). Cette loi étant connue explicitement on peut calculer la probabilité (6) et conclure du caractère aberrant ou non, d'un point de vue statistique, de la donnée y^{u_0}. En l'absence de données, nous n'avons cependant pas pu tester cette méthode.

3 Identification de points aberrants

Nous avons élaboré deux méthodes de détection de points aberrants, qui sont posées dans un cadre très général et indépendant du modèle statistique discuté dans la partie précédente. Ces méthodes traitent le problème suivant : étant donné un espace métrique fini (X, d), identifier ceux de ses points qui sont “isolés” (dans un sens heuristique, et non mathématique).

Cette généralité peut paraître trop grande et dommageable vis à vis du problème posé. Cependant, certains algorithmes de traitement du signal utilisés décrits par F. Barbaresco transforment les données radar $z \in \mathbb{C}^n$ en objets complexes, par exemple des matrices hermitiennes positives et de structure Toeplitz (des matrices de covariance empiriques des processus auto-régressifs stationnaires discutés dans la partie précédente). Leurs distances

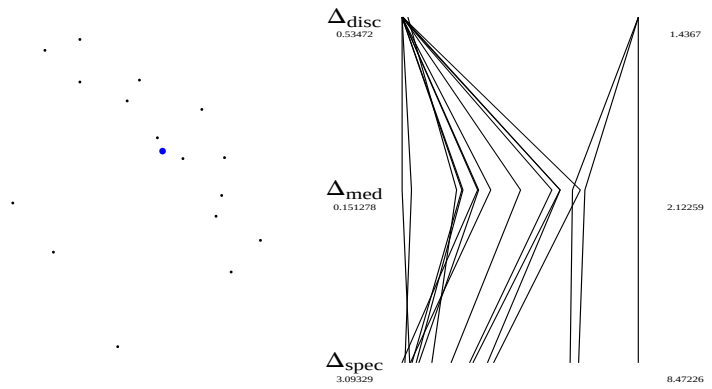


FIGURE 2 – Un ensemble de points du plan, et sa médiane (en bleu). Valeurs des fonctions Δ_{med} (définie ci-dessus), Δ_{disc} et Δ_{spec} (définies ci-dessous) pour chacun des ces points. La fonction Δ_{disc} permet de clairement distinguer les trois points isolés du groupe.

réciproques sont mesurées via différentes distances issues des théories de l'information ou du transport. Une approche assez générale était nécessaire pour englober ce cadre.

Précisément, notre objectif est le suivant : étant donné un espace métrique fini (X, d), construire une fonction $\Delta : X \to \mathbb{R}_+$ qui prenne des valeurs plus grandes pour les points "isolés" de X que pour les autres. Nous cherchons en particulier à construire une fonction plus discriminante que les choix naturels suivants : les fonctions distance $\Delta_{\text{bar}}(x) = |x - b|$ au barycentre b, ou distance $\Delta_{\text{med}}(x) = |x - m|$ à la médiane m. Rappelons que le barycentre et la médiane sont définis par les problèmes de minimisation (qui peuvent n'avoir aucune solution, ou en avoir plusieurs)

$$b = \operatorname*{argmin}_z \sum_{x \in X} d(z, x)^2, \quad \text{et} \quad m = \operatorname*{argmin}_z \sum_{x \in X} d(z, x),$$

où z parcourt X ou un espace métrique contenant X.

3.1 Méthodes issues de la géométrie discrète

La méthode décrite dans ce paragraphe est inspirée des travaux sur la persistance topologique [5]. L'heuristique de cette théorie est la suivante : définir des notions topologiques pour un nuage fini de points X inclus dans un espace métrique. Ces notions doivent être suffisamment "persistantes", en d'autres termes stables, pour que l'on puisse retrouver la topologie d'une variété $\mathcal{X}$ si celle-ci est échantillonnée suffisamment finement par l'ensemble X. La topologie de X étant discrète strictement parlant, car c'est un ensemble fini, il s'agit d'utiliser la structure métrique et de regarder l'ensemble X à différentes échelles $\delta \geq 0$.

Nous considérons donc un espace métrique fini (X, d). Pour chaque $\delta \geq 0$ nous définissons l'ensemble d'arêtes

$$E_\delta := \left\{(x, x') \in X^2;\ d(x, x') \leq \delta\right\},$$

en d'autres termes nous joignons deux points par une arête si et seulement si leur distance est inférieure à δ. Nous notons $C_\delta(x)$ la composante connexe de x dans le graphe (X, E_δ), et nous introduisons la fonction $\Delta_{\text{disc}} : X \to \mathbb{R}_+^*$ définie comme suit

$$\Delta_{\text{disc}}(x) := \inf\left\{\delta \geq 0;\ \#(C_\delta(x)) \geq \alpha \#(X)\right\},$$

où $\alpha \in (0,1)$ est fixé (nous avons choisi $\alpha = 1/2$ pour tous nos résultats numériques).

Si par exemple [1] $X \subset \mathbb{R}^n$, l'interprétation géométrique est la suivante : posons pour tout $\delta \geq 0$

$$X_\delta := \bigcup_{x \in X} \overline{B}(x, \delta/2),$$

c'est à dire que l'on regarde l'ensemble X à l'échelle δ. Le réel $\Delta_{\text{disc}}(x)$ est la valeur minimale de δ telle que la composante connexe (au sens topologique) du point x dans X_δ contienne une proportion r de l'ensemble X. A titre d'illustration supposons que, comme sur la Figure 2, les points de X sont regroupés en un nuage dense, excepté quelques points aberrants. La valeur $\Delta_{\text{disc}}(x)$ est faible pour les points du nuage, elle est proche de la plus petite valeur de δ qui rend ce nuage connexe dans E_δ. Pour un point aberrant elle est de l'ordre de la distance séparant ce point du nuage, qui est largement supérieure.

Le calcul de la fonction $\Delta_{\text{disc}} : X \to \mathbb{R}_+^*$ se fait en $\mathcal{O}(\#(X)^2 \ln \#(X))$ opérations grâce à un algorithme dont le principe est le suivant. On ordonne dans un premier temps l'ensemble des distances mutuelles entre points

$$0 = \delta_0 < \delta_1 < \cdots < \delta_K,$$

où $K \leq \#(X)(\#(X)-1)/2$. Nous calculons ensuite récursivement l'ensemble des composantes connexes du graphe (X, E_{δ_k}), $k = 0, \cdots, K$, ce qui donne accès à la fonction Δ_{disc}. Précisément, nous remarquons que les composantes connexes du graphe (X, E_{δ_0}) sont les singletons $\{x\}$, $x \in X$. Connaissant les composantes connexes du graphe $(X, E_{\delta_{k-1}})$, où $1 \leq k \leq K$, nous regardons si les points $x_k, y_k \in X$ tels que $d(x_k, y_k) = \delta_k$ appartiennent à la même composante connexe dans $(X, E_{\delta_{k-1}})$. Si ce n'est pas le cas, et si x_k, y_k sont les seuls points de X de distance mutuelle δ_k, alors les composantes connexes du graphe (X, E_{δ_k}) sont d'une part la réunion des composantes connexes de x_k et y_k dans $(X, E_{\delta_{k-1}})$, et d'autre part les autres composantes connexes de $(X, E_{\delta_{k-1}})$.

Cet algorithme joint récursivement par paires des sous-ensembles de X, et crée donc un arbre dont les points de X sont les feuilles. Cette propriété est anecdotique dans notre cas, mais elle est centrale dans [9], où une variante de cette algorithme est utilisée pour créer un arbre philogénique des espèces.

3.2 Méthodes spectrales

La seconde méthode d'identification est inspirée des algorithmes PageRank de Google [8], et des outils d'analyse de données développés dans [4].

Son principe est le suivant. On définit une marche aléatoire sur l'ensemble fini X comme suit : à partir d'un point $x \in X$, la probabilité de passer au point $x' \in X$ au temps suivant est proportionnelle à $g(d(x,x'))$, où $g : \mathbb{R}_+ \to \mathbb{R}_+$ est une fonction donnée. Sous des conditions assez générales [2] on associer une à cette marche une mesure invariante μ unique sur X. Nous posons pour chaque $x \in X$

$$\Delta_{\text{spec}}(x) := \mu(\{x\}), \tag{7}$$

1. Il suffit que X soit inclus dans un espace métrique $\mathcal{X}$ dont les boules sont connexes et tel que pour tous $x, x' \in X$ il existe $y \in \mathcal{X}$ (un "point milieu") tel que $d(x,y) = d(y,x') = d(x,x')/2$. C'est le cas si $\mathcal{X}$ est une variété riemannienne complète.

2. Il suffit par exemple que la marche soit *transitive*. En d'autres termes que l'on puisse passer de n'importe quel point de X à n'importe quel autre point en un nombre fini de pas, chacun de probabilité strictement positive.

ce qui représente la proportion de temps passée sur x par la marche aléatoire. Si la fonction g est bien choisie, alors $\Delta_{\text{spec}}(x)$ est très faible pour les points "isolés" de X.

Les possibilités suivantes pour la fonction g sont proposés dans [4]

$$g(r) = \exp(-r^2/\delta^2); \quad g(r) = \begin{cases} 1 & \text{si } r \leq \delta \\ 0 & \text{sinon} \end{cases}, \tag{8}$$

où $\delta > 0$ est un paramètre fixé. Le problème de ces choix est que le paramètre d'échelle δ joue un rôle important, et que déterminer une valeur convenable de ce paramètre n'est pas immédiat comme en témoigne la discussion [2]. Pour contourner cette difficulté nous avons choisi d'utiliser une fonction g *homogène*, à savoir

$$g(r) := \begin{cases} 1/r & \text{si } r \neq 0 \\ 0 & \text{si } r = 0 \end{cases}. \tag{9}$$

Notons $x_1, \cdots, x_N$ les points de X. La matrice stochastique associée à cette marche aléatoire a pour entrées

$$M_{ij} = \frac{g(d(x_i, x_j))}{\sum_{1 \leq k \leq N} g(d(x_k, x_j))}. \tag{10}$$

Pour calculer la mesure invariante μ, on remarquera que $(\mu(\{x_i\}))_{1 \leq i \leq N}$ est vecteur propre de la matrice M, associé à la valeur propre maximale 1.

Dans nos tests numériques la fonction Δ_{spec}, calculée avec le choix (9) de g, discriminait de façon convaincante les points "isolés" de X, à l'exception d'un cas pathologique que nous décrivons maintenant. Si quelques points isolés de X forment un groupe extrêmement resserré, alors la marche aléatoire que nous avons définie peut rester bloquée un temps assez long dans ce groupe lorsqu'elle y tombe. En conséquence la mesure μ donne un poids important à ce groupe de points pourtant isolés.

L'algorithme PageRank [8] qui ordonne les résultats de Google est fondé sur des idées proches de celles développées ci dessus. Lors de la présentation de nos résultats M Josselin Garnier nous a fait remarquer que cette situation pathologique était liée à une vulnérabilité de l'algorithme PageRank. Pour "tromper Google" et mettre en avant un certain site internet, il suffit a-priori de créer un groupe de sites (éventuellement sans contenu, appelés "ferme de liens" ou "link farm" en anglais) fortement interconnectés entre eux et avec le site d'intérêt (cette pratique est évidement punie par Google lorsqu'elle est détectée). Pour corriger cette vulnérabilité on pourra, dans notre contexte, remplaçer la matrice M introduite en (10) par la matrice (toujours stochastique)

$$(1-\gamma)M + \frac{\gamma}{N}[1],$$

où $\gamma \in (0,1)$ est un petit paramètre et où $[1]$ désigne la matrice dont tous les coefficients valent 1.

4 Géométrie(s) Riemannienne(s) sur les matrices de covariance

Comme il été mentionné dans la partie 3, les données peuvent se présenter sous la forme d'un ensemble fini de matrices de covariance empiriques de processus auto-régressifs.

Nous noterons $Herm_+(n,\mathbb{C})$ le cône des matrices $n \times n$ hermitiennes définies positives et $Toep_n$ les matrices dans $Herm_+(n,\mathbb{C})$ qui ont en plus une structure Toeplitz (ces matrices correspondent à des matrices de covariance de processus autorégressifs).

Pour l'intuition, il peut être utile de considérer l'équivalent réel de cet espace qui a une belle représentation géométrique. L'espace $Sym_+(n,\mathbb{R})$ s'identifie à l'espace des ellipsoïdes centrés en 0 de $\mathbb{R}^n$ via l'application :

$$M \mapsto \{x \in \mathbb{R}^n | x^T M x \leq 1.\}$$

Pour pouvoir appliquer les méthodes de la *CITE partie 3*. Il nous faut munir ces espaces de distances naturelles. De plus, pour appliquer d'autres méthodes de discrimination comme la comparaison au barycentre ou au point médian, on a besoin d'une structure plus riche, il pouvoir calculer des milieux entre deux points, si l'on ajoute que l'espace métrique considéré doit être complet, on obtient la définition d'un espace géodésique.

Nous décrivons deux approches possibles, l'une fondée sur la structure de variété riemannienne symétrique (qui correspond statistiquement à l'information de Fisher) et l'autre basée sur la distance de Wasserstein. Ces deux approches ont déjà été largement étudiées. On consultera les références données dans les deux parties suivantes pour un aperçu des travaux dans ces deux domaines.

La première approche fait apparaître une structure géométrique à la fois riche et bien adaptée à des calculs explicites, elle présente cependant l'inconvénient de supposer une connaissance a priori des lois de probablilité que l'on veut comparer. La seconde a l'avantage de permettre de définir la distance entre deux mesures de probabilité quelquonques, mais donne un espace moins symétrique.

4.1 Métrique de Fisher

L'exposé fait ici des propriétés de la métrique de Fisher est très succint. Pour les propriétés géométriques des espaces symétriques on renvoie au chapitre 5 du livre de Jürgen Jost [6]. Pour l'origine statistique de la métrique de Fisher et ces applications au traitement de signaux radars, on pourra se reporter à l'article de F. Barbaresco [3] et aux références qu'il contient.

Le groupe de Lie $GL_n(\mathbb{C})$ agit transitivement sur $Herm_+(n,\mathbb{C})$ par :

$$g.H = gHg^*$$

où g^* désigne l'adjoint de g. Le stabilisateur de la matrice identité I de $Herm_+(n,\mathbb{C})$ sous cette action est le groupe de $U_n(\mathbb{C})$ des matrice unitaires. On obtient un difféomorphisme entre $Herm_+(n,\mathbb{C})$ et $GL_n(\mathbb{C})/U_n(\mathbb{C})$. Si l'on munit l'espace tangent à $Herm_+(n,\mathbb{C})$ en l'identité (qui n'est autre que $Herm(n,\mathbb{C})$) du produit scalaire :

$$\langle H_1, H_2 \rangle_I = tr(H_1 H_2),$$

on peut transporter ce dernier via l'action de $GL_n(\mathbb{C})$ par :

$$\langle g.H_1, g.H_2 \rangle_{g.I} = \langle H_1, H_2 \rangle_I.$$

Plus explicitement cette métrique riemannienne est donnée par :

$$\langle H_1, H_2 \rangle_S = tr(S^{-1} H_1 S^{-1} H_2).$$

Cette métrique est par définition invariante sous l'action de $GL_n(\mathbb{C})$, elle munit $Herm_+(n,\mathbb{C})$ d'une structure de variété riemannienne homogène. De plus, on peut voir assez facilement que l'application : $S \mapsto S^{-1}$ est une involution isométrique de $Herm_+(n,\mathbb{C})$ fixant la matrice identité. Cette isométrie provient du morphisme involutif de $GL_n(\mathbb{C})$ donné par $g \mapsto (g^*)^{-1}$, dont le noyau est exactement $U_n(\mathbb{C})$. Ceci munit $Herm_+(n,\mathbb{C})$ d'une structure d'espace symétrique.

Cette structure permet de mener beaucoup de calculs explicitement, on peut en particulier exprimer la courbure, les géodésiques et la distance riemannienne (voir [6] pour les calculs dans les espaces symétriques abstraits, [3] pour les formules explicites dans le cas de $Herm_+(n,\mathbb{C})$). Le calcul de la courbure sectionnelle montre que celle-ci est toujours négative ou nulle. Ceci permet d'assurer l'existence de barycentre et de points médians.

Cependant, nous n'avons pas tiré parti ici de la structure de Toeplitz des matrices de covariances considéré. Ceci peut poser problème dans les applications, rien ne garantit par exemple que le barycentre ou le point médian d'un ensemble de matrices de covariance Toeplitz soit encore Toeplitz.

Pour contourner cette difficulté, il faut étudier la géométrie des matrices de Toeplitz comme sous-variété des matrices de covariances. Ceci est exposé dans l'article [12]. On observe alors un fait surprenant : la géométrie induite par la métrique de Fisher sur les matrices de Toeplitz est très simple : $Toep_n$ est alors isométrique au produit d'une droite euclidienne et de disques de Poincaré, ce qui permet de traiter les problèmes de recherches de point médian ou de barycentre facteur par facteur, ce qui simplifie beaucoup la recherche. Des algorithmes ont été proposés dans [12].

4.2 Distance de Wasserstein L^2

La distance de Wasserstein intervient naturellement dans l'étude de la théorie du transport optimal, on pourra consulter l'ouvrage de référence [11] pour de plus amples détails.

On note $\mathcal{P}_2(\mathbb{R}^n)$ l'ensemble des mesures de probabilité boréliennes μ sur $\mathbb{R}^n$ vérifiant $\int_{\mathbb{R}^n} |x|^2 d\mu(x) < +\infty$. Étant données deux mesures de probabilités boréliennes μ et ν dans $\mathcal{P}_2(\mathbb{R}^n)$, on peut définir la distance de Wasserstein L^2 entre μ et ν par :

$$W_2(\mu,\nu) = \left(\inf \left\{ \int_{\mathbb{R}^n \times \mathbb{R}^n} |x-y|^2 d\pi(x,y) \mid \pi \in \Pi_{\mu,\nu} \right\} \right)^{1/2},$$

où $\Pi_{\mu,\nu}$ est l'ensemble des plans de transports de μ vers ν, c'est à dire les mesures π sur $\mathbb{R}^n \times \mathbb{R}^n$ dont les marginales sont données par μ et ν.

On peut montrer que cette distance fait de $\mathcal{P}_2(\mathbb{R}^n)$ un espace géodésique, où la distance entre deux mesures μ et ν est l'infimum des longueurs des courbes absolument continues dans $\mathcal{P}_2(\mathbb{R}^n)$ reliant μ à ν, de plus cet infimum est réalisé par au moins une courbe dont on dit qu'elle est une géodésique de μ à ν.

McCann a montré dans [7] que le sous espace $\mathcal{G}_n$ de $\mathcal{P}_2(\mathbb{R}^n)$ formé par les mesures gaussiennes est un sous espace géodésique de $\mathcal{P}_2(\mathbb{R}^n)$: toute géodésique reliant deux mesures gaussiennes est tout entière contenue dans $\mathcal{G}_n$. La distance induite sur $\mathcal{G}_n$ par W_2 peut se calculer explicitement et provient d'une métrique riemannienne sur $\mathcal{G}_n$ que l'on peut écrire :

$$\langle (X_1,U_1),(X_2,U_2) \rangle_{(X,V)} = X_1^T X_2 + tr(U_1 V U_2)$$

où l'on a identifié $\mathcal{G}_n$ avec $\mathbb{R}^n \times Sym_+(n,\mathbb{R})$ en associant à $(X,V) \in \mathbb{R}^n \times Sym_+(n,\mathbb{R})$ la mesure :

$$\gamma_{X,V} = \frac{1}{\sqrt{\det(2\pi V)}} \exp((x-X)^T V^{-1}(x-X)/2)dx$$

où dx est la mesure de Lebesgue sur $\mathbb{R}^n$. On renvoie pour ces faits ainsi que pour les formules explicites pour la distances et les géodésiques à l'article [10] d'A. Takatsu.

Les courbures sectionelles de $\mathcal{G}_n$ sont explicitement calculées dans [10]. Contrairement à la métrique de Fisher, cette métrique est à courbure positive ou nulle, les arguments standards ne permettent donc pas d'établir existence et unicité du barycentre ou de la médiane d'un ensemble fini de gaussiennes pour cette métrique. M. Agueh et G. Carlier ont néanmoins prouvé l'existence d'un unique barycentre et en ont donné une caractérisation simple permettant de la calculer par une méthode de point fixe dans [1]. Le calcul d'un point médian dans l'espace de Wasserstein reste un problème ouvert.

Par analogie avec le cas de la métrique de Fisher, on peut se demander si l'espace des matrices Toeplitz muni de la métrique de Wasserstein a des propriétés géométriques intéressantes. Des calculs en petite dimension montre que la coubure de ce sous-espace est très variable, change de signe fréquemment et n'est pas bornée. Ce qui semble exclure une description géométrique simple de structure riemannienne sur les matrices de Toeplitz.

Références

[1] M. Agueh and G. Carlier. *Barycenters in the wasserstein space*, SIAM Journal on Mathematical Analysis, 43(2) :904–924, 2011.

[2] M. Balasubramanian et E. L. Schwartz, *The Isomap Algorithm and Topological Stability*, Science, Vol 295, 2002. Suivi d'une réponse de J. B. Tenenbaum.

[3] F. Barbaresco. *Interactions between symmetric cone and information geometries : Bruhat-tits and siegel spaces models for high resolution autoregressive doppler imagery*, in *Emerging Trends in Visual Computing*, ed. F. Nielsen volume 5416 of Lecture Notes in Computer Science, pages 124–163. Springer Berlin / Heidelberg, 2009.

[4] R. R. Coifman, M. Maggioni, S. W. Zucker et I. G. Kevrekidis, *Geometric diffusions for the analysis of data from sensor networks*, Current Opinion in Neurobiology, 2005

[5] H. Edelsbrunner et J. Harer. *Persistent homology ? a survey. Surveys on Discrete and Computational Geometry. Twenty Years Later*, eds. J. E. Goodman, J. Pach and R. Pollack, Contemporary Mathematics 453, 257-282, Amer. Math. Soc., Providence, Rhode Island, 2008.

[6] J. Jost. *Riemannian geometry and geometric analysis*, Universitext. Springer-Verlag, Berlin, fifth edition, 2008.

[7] R. J. McCann. *A convexity principle for interacting gases*, Adv. Math., 128(1) :153–179, 1997.

[8] L. Page, S. Brin et R. Motwani, *The PageRank Citation Ranking : Bringing Order to the Web*, Technical Report. Stanford InfoLab, 1999

[9] N. Saitou et M. Nei, *The Neighbor-Joining Method : A New Method for Reconstructing Phylogenic Trees*, Mol. Biol. Evol., 1987

[10] A. Takatsu. *On Wasserstein geometry of Gaussian measures*, in *Probabilistic approach to geometry*, volume 57 of Adv. Stud. Pure Math., pages 463–472. Math. Soc. Japan, Tokyo, 2010.

[11] C. Villani, *Optimal transport, Old and New*, volume 338 of Grundlehren der Mathematischen Wissenschaften , Springer-Verlag, Berlin, 2009.

[12] L. Yang, M. Arnaudon, and F. Barbaresco. *Geometry of Covariance Matrices and Computation of Median*, in *American Institute of Physics Conference Series*,eds. A. Mohammad-Djafari, J.-F. Bercher, & P. Bessiére, volume 1305 of American Institute of Physics Conference Series, pages 479–486, March 2011.