

Géométrie des matrices positives semi-définies de rang fixé : un peu de théorie et beaucoup d'applications

Rodolphe SEPULCHRE¹, Pierre-Antoine ABSIL², Silvère BONNABEL³

¹Institut Montefiore, B28
Université de Liège
4000 Liège Sart-Tilman, Belgique.
r.sepulchre@ulg.ac.be

²Département d'Ingénierie Mathématique
Université catholique de Louvain
Av. Georges Lemaître 4, B1348 Louvain La Neuve, Belgique.
absil@inma.ucl.ac.be

³Centre de Robotique
Mines Paris-Tech
60 Bd Saint-Michel, 75006 Paris, France.
silvere.bonnabel@mines-paristech.fr

Résumé— Cet article est une introduction au calcul et à l'optimisation sur les matrices symétriques positives semi-définies de rang (faible) fixé. L'approche proposée est basée sur deux géométries riemanniennes quotient, qui permettent de calculer efficacement tout en préservant le rang et le caractère positif des matrices considérées. Le champ d'applications est vaste, et l'article survole quelques développements récents qui illustrent l'intérêt de l'approche considérée dans les problèmes de très grande taille rencontrés en contrôle, statistiques, et apprentissage.

Mots-clés— Optimisation, géométrie riemannienne, matrices de rang faible, cône SDP, moyennes, apprentissage automatique.

I. INTRODUCTION

Cet article est une invitation au calcul et à l'optimisation sur l'espace des matrices symétriques $n \times n$ positives semi-définies de rang fixé p , en particulier lorsque $p \ll n$. Cet ensemble est noté $S^+(p, n)$ et, dans la suite de l'article, on parle de matrices *positives* ou de matrices *positives de rang faible* pour désigner les éléments de $S^+(p, n)$. L'article suit la structure de la conférence plénière donnée par le premier auteur au congrès CIFA 2010. Il constitue une introduction au sujet et pointe vers quelques contributions récentes d'un sujet en plein développement.

Les travaux présentés ont un fondement *géométrique* et une motivation *algorithmique*. On cherche à munir l'espace non-linéaire $S^+(p, n)$ d'une structure géométrique riemannienne efficace pour le calcul matriciel sur les matrices positives de rang faible. Dans ce sens, les travaux présentés s'inscrivent dans le cadre plus général de l'*optimisation sur les variétés* et s'appuient en particulier sur le point de vue développé dans l'ouvrage [4]. L'exemple principal de cet ouvrage concerne le calcul et l'optimisation sur la variété de Grassmann $Gr(p, n)$, qui désigne l'ensemble des sous-espaces linéaires de dimension p dans \mathbb{R}^n . Là aussi, les motivations proviennent essentiellement du rang faible

$p \ll n$. Une géométrie mise en avant dans cet ouvrage est celle où un sous-espace \mathcal{Y} est représenté par une matrice $Y \in \mathbb{R}^{n \times p}$ de rang plein qui sous-tend \mathcal{Y} . La représentation n'est pas unique, puisque la matrice YM sous-tend le même sous-espace pour n'importe quelle matrice M carrée de taille $p \times p$. Un sous-espace est donc représenté par une classe d'équivalence de matrices dans un espace de matrices linéaire. La géométrie *quotient* qui en résulte est particulièrement adaptée au calcul sur les sous-espaces parce qu'elle peut tirer profit de la simplicité du calcul dans un espace linéaire ambiant de dimension comparable à celle de l'espace d'intérêt. L'ouvrage [4], qui s'appuie lui-même sur des travaux antérieurs importants tels que [12], [30], [16], [3], [1], indique que la géométrie riemannienne associée à cette structure quotient est efficace pour le calcul et l'optimisation sur les sous-espaces.

Les travaux présentés ici s'inscrivent dans la même logique. Ils s'appuient sur deux géométries quotient pour les matrices positives de rang faible. Ces géométries trouvent leur source dans des factorisations classiques de matrices positives X sous la forme $X = YY^T$, où la matrice $Y \in \mathbb{R}^{n \times p}$ est de rang plein, et sous la forme $X = UR^2U^T$, où la matrice $U \in \mathbb{R}^{n \times p}$ satisfait la contrainte d'orthogonalité $U^TU = I_p$ et où la matrice R^2 est une matrice définie positive de dimension p . Comme dans le cas de la variété des sous-espaces, ces géométries permettent une paramétrisation peu coûteuse de l'ensemble $S^+(p, n)$ tout en tirant profit de la simplicité du calcul dans l'espace ambiant.

Le calcul sur les sous-espaces est motivé depuis longtemps par un grand nombre d'applications qui gravitent autour de la réduction de données et en particulier du calcul de sous-espaces propres dominants dans des grandes matrices. De la même manière, l'article qui suit suggère l'étendue du champ d'applications qui motivent le cal-

cul sur les matrices positives de rang faible. L'intérêt est plus récent que pour le calcul de sous-espaces, mais il a connu une accélération dans les dernières années avec l'utilisation croissante de relaxations convexes pour résoudre des problèmes d'optimisation durs et leur application dans des problèmes de très grande dimension. L'approche considérée ici est non-convexe mais elle permet de traiter des problèmes de très grande taille qui ne peuvent pas (encore) être traités dans le cadre de l'optimisation convexe.

L'article est organisé comme suit : la Section 2 développe quelque peu le cadre d'applications qui motive le calcul sur les matrices de rang faible. La Section 3 fournit une introduction aux géométries quotient de $S^+(p, n)$ et indique en quel sens ces géométries quotient ouvrent l'accès aux algorithmes d'optimisation sur les variétés quotient développés dans les dernières années. Les Sections 4, 5, et 6, donnent des pointeurs vers des contributions récentes qui ont appliqué ce cadre de calcul à des problèmes particuliers d'optimisation. La Section 7 conclut avec quelques perspectives sur les recherches en cours dans ce cadre.

A. Notations

- P_n est l'ensemble des matrices définies positives $n \times n$.
- $S^+(p, n)$ est l'ensemble des matrices symétriques semi-définies positives $n \times n$ de rang $p \leq n$.
- $GL(n)$ est le groupe général linéaire, c'est-à-dire l'ensemble des matrices carrées inversibles de taille n .
- $\mathbb{R}_*^{n \times p}$ est l'ensemble des matrices de rang plein $n \times p$.
- $V_{n,p} = O(n)/O(n-p)$ est la variété de Stiefel : l'ensemble des matrices $n \times p$ avec la contrainte d'orthogonalité $U^T U = I_p$.
- $Gr(p, n)$ est la variété de Grassman : l'ensemble des sous-espaces de dimension p dans \mathbb{R}^n .
- $Sym(n)$ est l'espace linéaire des matrices symétriques de taille n . $Skew(n)$ est l'espace linéaire des matrices anti-symétriques de taille n .
- $T_X \mathcal{M}$ est l'espace tangent à la variété \mathcal{M} au point X .

II. CALCULER AVEC DES MATRICES POSITIVES DE RANG FAIBLE

Les matrices symétriques définies positives occupent une place importante dans le calcul matriciel. A titre d'illustration, on peut citer les domaines suivants où elles jouent un rôle fondamental :

- Statistiques : les matrices semi-définies positives apparaissent comme matrices de covariance. La géométrie des matrices de covariance est étudiée dans [31].
- Optimisation : les matrices définies positives sont les variables d'un grand nombre de problèmes d'optimisation convexe et non-convexe [10].
- Théorie des systèmes et contrôle : les formes quadratiques caractérisent la solution de plusieurs problèmes centraux de la théorie des systèmes linéaires : filtrage optimal (équation de Kalman), étude de stabilité et construction de modèles d'états équilibrés (équation de Lyapunov), la commande optimale LQ (équation de Riccati). Plus généralement, de nombreux problèmes de contrôle peuvent être reformulés sous la forme d'inégalités matricielles linéaires (LMIs) dont les inconnues sont des matrices définies positives. [9]

- Apprentissage automatique : les matrices définies positives caractérisent un produit scalaire, et donc une distance euclidienne. Plus généralement, elles sont les paramètres des méthodes d'apprentissage à base de noyaux [21].
- Théorie des graphes : les matrices positives interviennent comme Laplacien d'un graphe ou comme matrices de similarité entre graphes [24].
- Théorie de l'information quantique : les matrices positives interviennent comme matrices de probabilité dans la généralisation de la théorie des probabilités sur les espaces non-commutatifs [27].

Bon nombre d'algorithmes dans ces différentes disciplines font usage de la géométrie riemannienne du cône des matrices positives. Deux raisons principales motivent la généralisation de ce cadre algorithmique aux matrices positives de rang faible : la taille des problèmes et les contraintes de rang. On pense notamment aux problèmes d'inférence statistique en grande dimension, qui constituent l'un des domaines de recherche les plus actifs des mathématiques appliquées.

Dans un algorithme itératif, les opérations matricielles courantes associées à une matrice carrée de taille $n \times n$ telles que le calcul de valeurs propres ou l'inversion ont un coût numérique d'ordre $O(n^3)$. Mais les problèmes de très grande taille demandent des algorithmes de complexité linéaire $O(n)$. Une solution consiste à approximer la matrice $n \times n$ par une matrice de rang $p \ll n$ qui peut être manipulée par des opérations de complexité $O(np^2)$ par exemple. Une telle approche a déjà été préconisée en océanographie (filtrage de Kalman de rang faible [29]), en apprentissage, ou en réduction de modèles [23]. Citons deux articles récents illustrant dans quel sens le cadre présenté ici est adapté au contexte.

Exemple 1 : L'article récent [35] étudie l'équation de Lyapunov généralisée

$$AXM + MXA = C$$

où A , M , et C sont des matrices symétriques dans $\mathbb{R}^{n \times n}$. La complexité d'un algorithme de résolution directe de cette équation est $O(n^3)$. Pour se ramener à une complexité linéaire, les auteurs utilisent une approximation de rang faible de la solution. En particulier, ils proposent de résoudre le problème d'optimisation :

$$\min f : S^+(p, n) \rightarrow \mathbb{R} : X \rightarrow tr(XAXM) - tr(XC)$$

sur l'ensemble des matrices positives de rang p . Pour résoudre ce problème d'optimisation, les auteurs utilisent un algorithme de région de confiance sur une variété riemannienne. Ils développent à cette fin une géométrie riemannienne (plongée) pour $S^+(p, n)$, proche des géométries (quotient) considérées dans le présent article. \square

Une deuxième motivation pour le développement d'algorithmes sur $S^+(p, n)$ provient de contraintes de rang rencontrées dans une multitude de problèmes. Les contraintes de rang apparaissent notamment dans l'apprentissage ou la complétion d'une matrice de distance [32], et dans la relaxation de problèmes d'optimisation non-linéaire ou/et combinatoire. L'exemple qui suit fournit une illustration simple en automatique.

Exemple 2 : Les auteurs de l'article [17] montrent que l'existence d'un contrôleur qui stabilise simultanément N fonctions de transfert (dans une configuration de retour unitaire) peut-être reformulée comme l'existence d'une solution à l'inégalité matricielle bilinéaire (BMI)

$$\text{Trouver } x \in \mathbb{R}^n : \sum_{i=1}^n \sum_{j=1}^n x_i x_j H_{ij} > 0 \quad (BMI)$$

Une telle inégalité peut être convertie en une inégalité matricielle linéaire en définissant les éléments de la matrice X comme $x_{ij} = x_i x_j$. L'inégalité linéaire conduit à une relaxation convexe du problème mais le problème n'est équivalent au problème initial que sous une contrainte de rang 1. Une approche consiste à moduler le degré de relaxation du problème en cherchant des solutions pour différentes valeurs (fixées) du rang. \square

La formulation de différents problèmes de contrôle sous la forme d'inégalités matricielles bilinéaires et leur relaxation convexe sont considérées dans [15]. Plus généralement, tout problème d'optimisation avec des contraintes et un objectif polynomial peut être reformulé comme une inégalité matricielle linéaire avec contrainte de rang 1. L'article [28] propose une méthode de Newton sur $S^+(p, n)$ pour la résolution de tels problèmes. La géométrie riemannienne (plongée) considérée est proche des géométries (quotients) considérées ici.

III. GÉOMÉTRIE DES MATRICES DE RANG FAIBLE

A. Distances invariantes et paramétrisations

Les deux géométries considérées dans ce papier s'appuient sur deux factorisations matricielles importantes. La factorisation

$$X = YY^T, \quad Y \in \mathbb{R}^{n \times p} \quad (1)$$

généralise la représentation d'un nombre positif comme le carré d'un nombre réel : $x = y^2$. La représentation n'est pas unique puisque $YY^T = (YO)(O^T Y)$ pour n'importe quelle matrice orthogonale $O \in O(p)$. Mais la factorisation est efficace car en déplaçant les opérations de calcul de X sur Y , on peut calculer "comme dans un espace linéaire". Puisque le représentant Y n'est pas unique, il faut seulement veiller à ce que les calculs ne dépendent pas d'un choix particulier. On peut illustrer ce raisonnement sur le choix d'une distance entre deux matrices $X_1 = Y_1 Y_1^T$ et $X_2 = Y_2 Y_2^T$ de $S^+(p, n)$. Le choix

$$d(X_1, X_2) = \min_{O \in O(p)} \|Y_1 - Y_2 O\|_F \quad (2)$$

s'impose naturellement parce qu'il ramène le calcul de distance dans un espace euclidien, et le minimum permet de trouver le point de la fibre sur laquelle se trouve Y_2 le plus proche de Y_1 (voir Fig.1), ce qui rend la distance indépendante d'un choix particulier de représentants.

Dans la factorisation (1), la matrice Y peut elle même être factorisée sous la forme $Y = UR$ où U est une matrice $n \times p$ qui satisfait $U^T U = I_p$ et R est une matrice $p \times p$ définie positive. Cette factorisation matricielle généralise la forme polaire d'un nombre complexe $y = e^{i\theta} r$. On obtient de la sorte une deuxième factorisation

$$X = UR^2 U^T, \quad U \in V_{n,p}, \quad R^2 \in P(p), \quad (3)$$

et ainsi, une deuxième paramétrisation des matrices de $S^+(p, n)$ par un couple de matrices (U, R^2) . Tout comme dans le cas précédent, la factorisation n'est pas unique : le couple (U, R^2) et $(UO, O^T R^2 O)$ représentent la même matrice X pour n'importe quelle matrice orthogonale $O \in O(p)$. La paramétrisation (3) met en évidence la représentation géométrique d'une matrice positive de rang faible comme un ellipsoïde "plat", c'est-à-dire un ellipsoïde de \mathbb{R}^n contenu dans un sous-espace de dimension p . Dans la représentation (3), la matrice U identifie le sous-espace $\text{sp}(U) = \text{sp}(UO)$ tandis que la matrice $R^2 \succ 0_p$ identifie un ellipsoïde plein dans ce sous-espace. Cette géométrie suggère le choix d'une distance sur $S^+(p, n)$ qui pondère séparément la distance entre les sous-espaces sous-tendus par les deux ellipsoïdes, et la distance entre les deux ellipsoïdes dans un sous-espace commun. La distance entre deux matrices $X_1 = U_1 R_1^2 U_1^T$ et $X_2 = U_2 R_2^2 U_2^T$ pourrait alors prendre la forme

$$d(X_1, X_2) = d_{Gr(p,n)}(\text{sp}(U_1), \text{sp}(U_2)) + d_{P(p)}(R_1^2, R_2^2) \quad (4)$$

sachant que les deux variétés $Gr(p, n)$ et $P(p)$ ont des géométries riemanniennes bien étudiées. La distance naturelle sur Grassmann fait intervenir les angles principaux entre sous-espaces. Elle est invariante par rotation des matrices U_1 et/ou U_2 . En revanche, la distance naturelle sur le cône est invariante par l'action simultanée de conjugaison sur R_1^2 et R_2^2 mais, en général, $d_{P(p)}(R_1^2, R_2^2) \neq d_{P(p)}(R_1^2, O^T R_2^2 O)$. Les représentants R_1^2 et R_2^2 ne peuvent donc pas être choisis arbitrairement dans l'expression (4). Dans [6], le représentant R_2 est fixé en définissant la rotation d'énergie minimale nécessaire à ramener l'ellipsoïde X_2 dans le plan sous-tendu par l'ellipsoïde X_1 . Cette opération peut être vue comme le choix d'un représentant U_2 le plus proche de U_1 dans $V_{n,p}$. On montre dans [6] que (4) n'est pas une distance riemannienne (par exemple elle ne satisfait pas l'inégalité triangulaire). Elle en est cependant une bonne approximation, dans un sens qui, pour être précisé, demande une caractérisation plus précise de la structure riemannienne correspondante. Elle fournit donc une mesure de proximité entre matrices de $S^+(p, n)$.

Les deux géométries esquissées ci-dessus conduisent à deux structures riemanniennes distinctes pour l'espace $S^+(p, n)$. Dans le premier cas, un point de $S^+(p, n)$ est identifié à une classe d'équivalence de matrices dans l'espace $\mathbb{R}_*^{n \times p}$. Dans le deuxième cas, un point de $S^+(p, n)$ est identifié à une classe d'équivalences de matrices dans l'espace produit $V_{n,p} \times P(p)$. Dans les deux cas, la classe d'équivalence est générée par le groupe des rotations $O(p)$. Formellement, on note cette identification

$$S^+(p, n) \approx \mathbb{R}_*^{n \times p} / O(p) \approx (V_{n,p} \times P(p)) / O(p)$$

L'espace des classes d'équivalence est appelé un espace *quotient* tandis que l'espace ambiant est appelé espace *total*. Une classe d'équivalence dans l'espace total est appelée *fibre*.

B. Géométrie riemannienne des variétés quotient

La section précédente suggère de décrire l'espace non-linéaire $S^+(p, n)$ comme un quotient (ou ensemble de classes d'équivalence) dans des espaces tels que $\mathbb{R}_*^{n \times p}$, $V_{n,p}$,

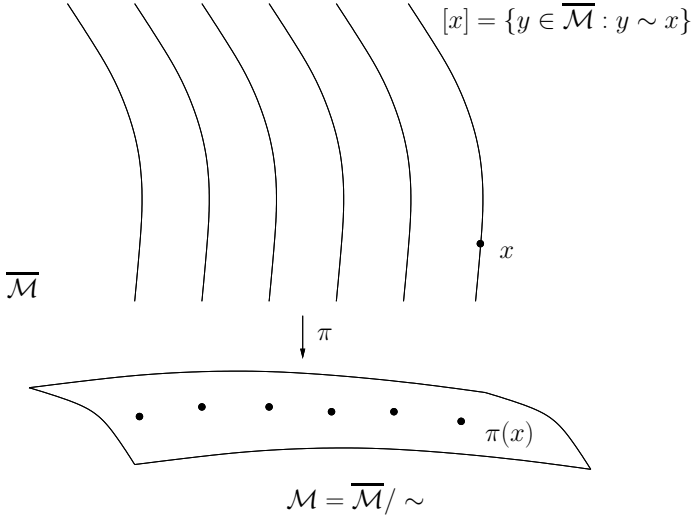


Fig. 1. L'ensemble $\overline{\mathcal{M}}/\sim := \{[x] : x \in \overline{\mathcal{M}}\}$ est une *variété différentielle quotient*. Le projecteur π projette chaque $x \in \overline{\mathcal{M}}$ de l'espace total sur la variété.

et $P(p)$. Ces espaces ont l'avantage de posséder une structure de variété différentielle riemannienne bien étudiée. La théorie des variétés quotient tire profit de cet avantage en inférant une géométrie riemannienne sur l'espace quotient à partir de la géométrie riemannienne de l'espace total. De nombreuses variétés rencontrées dans les applications matricielles peuvent avantageusement être décrites comme des quotients d'espaces plus simples. Pour cette raison, la géométrie riemannienne des variétés quotient occupe une place centrale dans l'ouvrage [4]. Par exemple, la variété de Grassmann $Gr(p, n)$ y est décrite comme le quotient $Y \in \mathbb{R}_*^{n \times p} / GL(p)$ parce que le sous-espace $\text{sp}(Y)$ est caractérisé par l'ensemble des matrices $[Y] = \{YM \mid M \in GL(p)\}$. La relation $Y_1 = Y_2 M$ est une relation d'équivalence parce que l'espace $GL(p)$ est un groupe (pour la multiplication matricielle).

La structure différentielle d'une variété quotient est schématisée sur la Figure 1. Sous des conditions de régularité de la relation d'équivalence (Proposition 3.4.2. dans [4]), elle est entièrement caractérisée par la structure différentielle de l'espace total.

Au-delà de la structure différentielle d'une variété quotient, la caractérisation de sa géométrie riemannienne suit également le schéma de la Figure 1. La caractérisation du plan tangent, de la métrique, et des géodésiques s'appuie sur leur caractérisation dans l'espace total, en veillant à ce que la caractérisation induite dans l'espace quotient ne dépende pas du représentant choisi dans la fibre.

Cette caractérisation peut être illustrée sur la construction du plan tangent. Soit un point x de l'espace total $\overline{\mathcal{M}}$. Puisque la fibre est une sous-variété de l'espace total, l'espace tangent à la fibre au point x est un sous-espace de l'espace tangent $T_x \overline{\mathcal{M}}$. On appelle ce sous-espace l'espace *vertical* \mathcal{V}_x et un espace complémentaire (transverse aux fibres) l'espace horizontal \mathcal{H}_x . Tout vecteur tangent au point x dans l'espace total admet donc la décomposition

$$\xi_x = \xi_x^v + \xi_x^h, \xi_x^v \in \mathcal{V}_x, \xi_x^h \in \mathcal{H}_x$$

Le vecteur ξ_x^h est appelé le relèvement horizontal (*hori-*

zontal lift) du vecteur tangent $\xi_{\pi(x)}$. Il constitue une représentation du vecteur tangent à la variété quotient par un vecteur tangent à l'espace total.

C. Géométries riemanniennes quotient des matrices de rang fixé

Les deux paramétrisations de $S^+(p, n)$ discutées dans la Section 4 sous-tendent des géométries riemanniennes quotients étudiées respectivement dans les références [20], [2], [34] et [6].

C.1 Première géométrie

Le quotient

$$S^+(p, n) \approx \mathbb{R}_*^{n \times p} / O(p)$$

donne lieu à une géométrie riemannienne quotient particulièrement simple et analogue à la structure quotient $Gr(p, n) \approx \mathbb{R}_*^{n \times p} / GL(p)$ puisque l'espace total $\mathbb{R}_*^{n \times p}$ possède une géométrie euclidienne. Son plan tangent en tout point est l'espace linéaire $\mathbb{R}^{n \times p}$, qui peut être muni du produit scalaire usuel $\langle \xi_1, \xi_2 \rangle = \text{tr}(\xi_1^T \xi_2)$. La fibre correspondant à une classe d'équivalence est le groupe orthogonal $O(p)$, dont la géométrie est bien caractérisée. Son plan tangent en un point $Y \in \mathbb{R}_*^{n \times p}$ est l'ensemble des matrices de la forme $Y\Omega$ où $\Omega \in \text{skew}(p)$ est anti-symétrique. On en déduit une expression de l'espace horizontal

$$\mathcal{H}_Y = \{\xi_Y \in \mathbb{R}^{n \times p} \mid \xi_Y^T Y = Y^T \xi_Y\}$$

qui est le complément orthogonal de l'espace vertical pour la métrique euclidienne. La restriction de la métrique sur l'espace horizontal induit une structure riemannienne sur l'espace quotient parce que la métrique est invariante le long de la fibre.

Il résulte de cette structure riemannienne quotient que les objets géométriques peuvent être caractérisés sur l'espace quotient à partir de leur caractérisation dans l'espace total. Par exemple, si \bar{f} est une fonction définie dans l'espace total qui est invariante le long des fibres (i.e. $\bar{f}(YO) = \bar{f}(Y)$ pour toute matrice de rotation O), alors elle induit une fonction unique f sur l'espace quotient et on a la propriété

$$\overline{\text{grad} f_x} = \text{grad} \bar{f}(\bar{x})$$

qui exprime que le lift horizontal du gradient sur la variété quotient est simplement le gradient dans l'espace total. En d'autres mots, le gradient dans l'espace total est un représentant adéquat du gradient dans l'espace quotient. Cette propriété est implicitement utilisée dans nombre d'algorithmes écrits dans l'espace total.

On montre dans [4] que la géométrie de second ordre (caractérisation de la connection riemannienne et des géodésiques) peut également tirer profit des relations entre espace total et espace quotient. Ainsi, lorsque l'espace horizontal est le complément *orthogonal* de l'espace vertical, les géodésiques de l'espace quotient $S^+(p, n) \approx \mathbb{R}_*^{n \times p} / O(p)$ sont des projections des géodésiques de l'espace total, c'est-à-dire des droites dans l'espace $\mathbb{R}^{n \times p}$. Il en résulte que la distance (2) correspond bien à la distance riemannienne (la longueur de la géodésique) de l'espace $S^+(p, n) \approx \mathbb{R}_*^{n \times p} / O(p)$.

C.2 Deuxième géométrie

Le quotient $S^+(p, n) \approx (V_{n,p} \times P(p))/O(p)$ donne également lieu à une géométrie riemannienne étroitement liée à celle de l'espace total produit $V_{n,p} \times P(p)$.

Variété de Stiefel $V_{n,p}$: La géométrie riemannienne de la variété de Stiefel $V_{n,p}$ est par exemple étudiée en détail du point de vue algorithmique dans l'article [12]. La métrique usuelle $\langle \xi_1, \xi_2 \rangle = \text{tr}(\xi_1^T \xi_2)$ induit une métrique dans l'espace quotient $V_{n,p}/O(p) \approx Gr(p, n)$ car elle est invariante le long des fibres. Au point $U \in V_{n,p}$, un vecteur horizontal est de la forme $\xi_U = U_\perp B$ pour une matrice quelconque $B \in \mathbb{R}^{(n-p) \times p}$. La métrique $g_U(\xi_U^1, \xi_U^2) = \text{tr}(B_1^T B_2)$ est une métrique invariante le long de la fibre $UO(p)$.

Cône des matrices $P(p)$: La géométrie riemannienne du cône $P(p)$ est étudiée dans [13]. Le cône possède une géométrie réductive obtenue par action du groupe de conjugaison sur l'identité : $\Phi_R(X) = RXR^T$. Au point $R^2 \in P(p)$, un vecteur tangent est de la forme $\xi_{R^2} = RDR$ où $D \in \text{Sym}(p)$. La métrique $g_{R^2}(\xi_{R^2}^1, \xi_{R^2}^2) = \text{tr}(R^{-1}D_1R^{-2}D_2R^{-1})$ est une métrique invariante le long de la fibre $O(p)R^2O(p)^T$.

Tout combinaison de ces deux métriques avec $\lambda > 0$

$$g_{(U, R^2)} = g_U + \lambda g_{R^2}$$

est donc invariante le long des fibres, et induit une métrique sur l'espace horizontal défini comme le produit des espaces tangents à $V_{n,p}$ et $P(p)$:

$$\mathcal{H}_{(U, R^2)} = \{(U_\perp B, RDR) : B \in \mathbb{R}^{(n-p) \times p}, D \in \text{Sym}(p)\}$$

Puisque l'espace vertical est automatiquement défini comme l'espace tangent aux fibres

$$\mathcal{V}_{(U, R^2)} = \{(U\Omega, R^2\Omega - \Omega R^2) : \Omega \in \text{skew}(p)\}$$

on remarquera que l'espace horizontal est complémentaire mais non orthogonal à l'espace vertical pour la métrique de l'espace total $g_{(U, R^2)}$. Il s'agit d'une différence notable par rapport à la géométrie quotient $S^+(p, n) \approx \mathbb{R}_*^{n \times p}/O(p)$, avec comme conséquence, en particulier, que les géodésiques ne coïncident pas avec les géodésiques de l'espace total. C'est la raison pour laquelle la distance riemannienne associée à cette géométrie ne se réduit pas à l'expression (4) même si cette expression semble une bonne mesure de proximité entre éléments de $S^+(p, n)$. Le lecteur est renvoyé à la référence [6] pour plus de détails.

IV. MOYENNES DE RANG FAIBLE

Munir l'espace $S^+(p, n)$ d'une géométrie riemannienne ouvre la voie à la définition et au calcul de quantités statistiques telles que la moyenne ou la médiane d'un ensemble de points. La capacité à établir des moyennes sur des sous-ensembles de matrices, permet notamment de faire du filtrage sur ces sous-ensembles. La capacité à définir des médianes permet en particulier d'étendre à des données matricielles les filtres médians, connus en traitement d'image pour leur grande robustesse aux points aberrants. Sur une variété riemannienne, la moyenne (de Karcher) est définie comme le point (ou l'ensemble des points)

$$\arg \min \sum_{i=1}^N d^2(x, x_i)$$

tandis que la médiane est définie comme le point (ou l'ensemble des points)

$$\arg \min \sum_{i=1}^N d(x, x_i)$$

Les moyennes de matrices positive définies ont été abondamment étudiées (voir par exemple le chapitre 4 dans [5]). La moyenne de Karcher de deux matrices sur le cône $P(n)$ muni de la métrique invariante est une généralisation de la moyenne géométrique de deux nombres positifs. En effet elle admet l'expression

$$M(X_1, X_2) = X_1^{1/2}(X_1^{1/2}X_2X_1^{1/2})^{-1/2}X_1^{1/2}$$

qui est une version "symétrisée" de $(X_1X_2)^{1/2}$. De manière plus générale, le cône $P(n)$ muni de la métrique invariante est une variété géodésiquement complète, qui admet une moyenne et une médiane unique pour un ensemble quelconque de N points.

La généralisation du concept de moyenne à des matrices de rang fixé est considérée dans [7]. La géométrie $S^+(p, n) \approx (V_{n,p} \times P(p))/O(p)$ suggère une définition basée sur le sous-espace moyen calculé sur la variété $Gr(p, n)$ et sur un ellipsoïde moyen dans ce sous espace calculé sur la variété $P(p)$. Chaque ellipsoïde est ramené dans le sous-espace moyen par une rotation minimale. Lorsque l'on particularise cette définition à des projecteurs, on obtient le projecteur sur l'espace moyen calculé sur $Gr(p, n)$.

Cette approche diffère de l'approche plus classique qui consiste à étendre la définition de moyenne à la fermeture de $P(n)$ par un argument de continuité. Dans ce cas, on peut montrer que la moyenne ne préserve pas le rang. Avec cette définition la moyenne de deux ellipsoïdes de $S^+(p, n)$ pour $p < n$ est presque-toujours nulle.

V. OPTIMISATION SUR LES MATRICES DE RANG FAIBLE

Munir l'espace $S^+(p, n)$ de géométries riemanniennes ouvre la voie à l'optimisation numérique sur les matrices de rang faible en adpotant le cadre général des algorithmes d'optimisation sur les variétés considéré dans [4]. C'est l'objet principal de l'article [20] (voir aussi la thèse [19] pour plus de détails), qui considère plus particulièrement des problèmes d'optimisation sur le spectahedron (l'ensemble des matrices positives de trace unitaire) et sur l'elliptope (l'ensemble des matrices positives dont les éléments diagonaux sont tous unitaires). Un problème classique sur l'elliptope est le problème de partition optimale de graphes (max-cut) ; un problème classique sur le spectahedron est le problème de recherche de composantes principales creuses (sparse PCA).

Pour aborder de tels problèmes d'optimisation dans des problèmes de grande dimension, on suppose que le problème admet une solution optimale de rang faible $p^* \ll n$. Le rang p^* n'est pas nécessairement connu a priori. L'algorithme proposé effectue une optimisation sur $S^+(p, n)$, pour des valeurs croissantes du rang p . Pour chaque valeur de p , l'algorithme trouve un minimum (local). Les conditions d'optimalité permettent de vérifier si la solution de rang p est une solution optimale du problème original. Si ce n'est pas le cas, le rang est augmenté et la solution de

rang p est utilisée comme condition initiale. Un algorithme de région de confiance permet d'identifier rapidement une direction de descente à partir de l'ancien minimum local.

L'algorithme proposé permet en principe de résoudre des problèmes de très grande taille pour autant qu'ils possèdent une solution optimale de rang faible. Dans l'article [18], l'algorithme est appliqué avec succès dans un problème de segmentation d'images de taille $0(10^3)$.

VI. RÉGRESSION LINÉAIRE DANS $S^+(p, n)$

Munir l'espace $S^+(p, n)$ de géométries riemanniennes ouvre la voie aux algorithmes de régression sur cet espace. C'est l'objet principal des articles [26], [25] qui utilisent le cadre général des méthodes de gradient stochastique [8] pour l'apprentissage de noyaux ou de distances de rang faible. Le problème de régression (linéaire) consiste à apprendre un modèle $\hat{y} = \text{tr}(XW)$ à partir d'une séquence de données $W_t \succeq 0 \in \mathbb{R}^{n \times n}$ et d'observations $y_t \in \mathbb{R}^+$, $t = 1, 2, \dots$. Dans une majorité d'applications, la matrice W est de rang 1 et on écrira $W = ww^T$ avec $w \in \mathbb{R}^n$. Les algorithmes d'apprentissage sont des algorithmes itératifs qui consistent en une mise à jour du modèle estimé à l'instant t

$$X_{t+1} = \arg \min_{X \in S^+(p, n)} D(X, X_t) + \eta_t L_t(X)$$

où le terme $D(X, X_t)$ est une mesure de proximité sur $S^+(p, n)$ à X_t , qui représente l'ensemble de la connaissance acquise sur le modèle à l'instant t , et où $L(X)$ est une erreur de prédiction basée sur un certain nombre d'observations passées, par exemple $L_t(X) = \sum_{k=t-T}^t (y_k - \hat{y}_k)^2$.

Des algorithmes d'apprentissage sont développés dans [26], [25] sur base des deux géométries quotients discutées plus haut. La géométrie $S^+(p, n) \approx \mathbb{R}_*^{n \times p} / O(p)$ et la distance 2 donnent lieu à une itération particulièrement simple pour Y_t dans la factorisation $X_t = Y_t Y_t^T$:

$$Y_{t+1} = Y_t - \eta_t (\hat{y}_t - y_t) W_t Y_t \quad (5)$$

Cette itération peut être rapprochée d'heuristiques proposées dans la littérature, par exemple [11]. La géométrie $S^+(p, n) \approx (\mathbb{V}_{n,p} \times P(p)) / O(p)$ et la mesure de proximité (4) donnent lieu à des itérations séparées sur le sous-espace $\text{sp}(U_t)$ et sur la métrique dans le sous-espace R_t^2 . Cette géométrie est particulièrement intéressante pour comprendre la régression sur $S^+(p, n)$ comme une généralisation de travaux antérieurs traitant soit de la régression sur $Gr(p, n)$ (apprentissage de sous-espaces, voir par exemple [22]), soit de la régression sur le cône des matrices définies positives, voir par exemple [33].

La régression sur $S^+(p, n)$ offre un cadre pour l'apprentissage simultané d'un sous-espace et d'une métrique dans le sous-espace. Les résultats présentés dans [25] suggèrent qu'il s'agit d'un cadre novateur et intéressant pour les problèmes de grande taille. En particulier, et pour des raisons évidentes, l'apprentissage simultané du sous-espace et de la métrique peut conduire à des performances largement supérieures à l'alternative qui consiste à réduire la dimension des données dans un premier temps par l'apprentissage d'un sous-espace et à apprendre une métrique dans le sous-espace des données réduites.

VII. CONCLUSION

L'ensemble des matrices symétriques positives semi-définies de rang fixé peut être muni de plusieurs géométries riemanniennes qui offrent un cadre rigoureux et général pour le développement d'algorithmes de calcul efficaces qui préservent le rang et la structure de ces matrices. L'utilisation de ces fondements géométriques à des fins algorithmiques est un domaine de recherche assez récent. Deux géométries riemanniennes de variété *quotient* sont discutées dans cet article. Elles trouvent leurs racines dans des factorisations matricielles classiques qui, parce qu'elles ne sont pas uniques, donnent lieu à des représentations d'une même matrice positive de rang fixé par une classe d'équivalence de matrices dans un espace ambiant. Les géométries quotient donnent lieu à des calculs efficaces parce qu'elles tirent profit de la géométrie simplifiée de l'espace ambiant sans augmenter significativement la dimension de la paramétrisation de l'ensemble visé. L'article pointe vers diverses contributions récentes qui illustrent le champ d'applications très vaste de l'optimisation sur les matrices positives de rang faible en statistiques, en théorie du contrôle, et en théorie de l'apprentissage.

Deux perspectives intéressantes peuvent être identifiées pour la poursuite du cadre algorithmique considéré. D'une part, le présent article considère des géométries *quotients* pour décrire l'espace $S^+(p, n)$. Il reste à quantifier l'avantage relatif de ces géométries par rapport à d'autres géométries riemanniennes pour le même ensemble de matrices, en particulier les géométries *plongées* considérées par exemple dans [28], [34], [35]. D'autre part, le présent article considère un cadre d'optimisation non convexe dans lequel le rang des matrices positives considérées est fixé. Une approche alternative qui a reçu une attention croissante dans les dernières années consiste à privilégier un cadre d'optimisation convexe sur le cône $P(n)$ et à induire un rang faible – non-fixé a priori – par des termes de régularisation dans la fonction objectif (voir par exemple [14]). Les liens entre les algorithmes qui résultent des deux approches restent à identifier.

VIII. REMERCIEMENTS

L'article présente des résultats de recherches du réseau belge DYSCO (Dynamical Systems, Control, and Optimization), financé par le Programme interuniversitaire des pôles d'attraction, initié par le service de la politique scientifique. La responsabilité scientifique incombe aux auteurs.

RÉFÉRENCES

- [1] P.-A. Absil, C. G. Baker, and K. A. Gallivan. Trust-region methods on Riemannian manifolds. *Found. Comput. Math.*, 7(3) :303–330, July 2007.
- [2] P.-A. Absil, M. Ishteva, L. De Lathauwer, and S. Van Huffel. A geometric Newton method for Oja's vector field. *Neural Comput.*, 21(5) :1415–1433, May 2009.
- [3] P.-A. Absil, R. Mahony, and R. Sepulchre. Riemannian geometry of grassmann manifolds with a view on algorithmic computation. *Acta Applicandae Mathematicae*, 80 :199–220, 2004.
- [4] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008.
- [5] R. Bathia. *Positive definite matrices*. Applied Mathematics. Princeton University Press, 2007.
- [6] S. Bonnabel and R. Sepulchre. Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank. *SIAM. J. Matrix Anal. & Appl.*, 31(3) :1055–1070, August 2009.

- [7] S. Bonnabel and R. Sepulchre. Rank-preserving geometric means of positive semidefinite matrices. In *Proceedings of the 14th Symposium on Mathematical Theory of Networks and Systems*, July 8-10, Bucharest. 2010.
- [8] L. Bottou. Online algorithms and stochastic approximations. In David Saad, editor, *Online Learning and Neural Networks*. Cambridge University Press, Cambridge, UK., 1998.
- [9] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in Systems and Control Theory*, volume 15 of *Studies in Applied Mathematics*. Society for Industrial and Applied Mathematics, 1994.
- [10] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.
- [11] F. Dabek, R. Cox, F. Kaashoek, and R. Morris. Vivaldi : a decentralized network coordinate system. In *Proceedings of the 2004 conference on applications, technologies, and protocols for computer communications*, pages 15–26, 2004.
- [12] A. Edelman, T. A. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, 20(2) :303–353, 1998.
- [13] J. Faraut and A. Koranyi. *Analysis on Symmetric Cones*. Oxford Univ. Press, London, UK., 1994.
- [14] M. Fazel. *Matrix Rank Minimization with Applications*. PhD thesis, Stanford University, March 2002.
- [15] K.-C. Goh, M. G. Safonov, and J. H. Ly. Robust synthesis via bilinear matrix inequalities. *International Journal of Robust and Nonlinear Control*, 6(9-10) :1079–1095, 1996.
- [16] U. Helmke and J. B. Moore. *Optimization and Dynamical Systems*. Springer, 1994.
- [17] D. Henrion, S. Tarbouriech, and M. Sebek. Rank-one approach to simultaneous stabilization of linear systems. *Systems and Control Letters*, 38(2) :79–89, 1999.
- [18] A. Joulin, F. Bach, and J. Ponce. Discriminative clustering for image co-segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.
- [19] M. Journée. *Geometric algorithms for component analysis with a view to gene expression data analysis*. PhD thesis, Université de Liège, 2009.
- [20] M. Journée, F. Bach, P.-A. Absil, and R. Sepulchre. Low-rank optimization on the cone of positive semidefinite matrices. *SIAM J. Optimization*, in press.(Available from : <http://arxiv.org/abs/0807.4423>), 2010.
- [21] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.
- [22] J. Karhunen and J. Joutsensalo. Generalizations of principal component analysis, optimization problems, and neural networks. *Neural Networks*, 8(4) :549–562, 1995.
- [23] J.-R. Li and J. White. Low rank solutions of lyapunov equations. *SIAM J. Matrix Anal. Appl.*, 24(1) :260–280, 2002.
- [24] R. Merris. Laplacian matrices of graphs : a survey. *Linear Algebra and its Applications*, 197-198 :143 – 176, 1994.
- [25] G. Meyer, S. Bonnabel, and R. Sepulchre. Regression on fixed-rank positive semidefinite matrices : a geometric approach. *submitted to Journal of Machine Learning Research*, 2010.
- [26] G. Meyer, M. Journée, S. Bonnabel, and R. Sepulchre. From subspace learning to distance learning : A geometrical optimization approach. In *Proceedings of the IEEE Workshop on Statistical Signal Processing (SSP 2009)*, 2009.
- [27] M.A. Nielsen and I.L. Chuang. *Quantum Computation and Quantum Information*. Cambridge Series, 2000.
- [28] R. Orsi, U. Helmke, and J. Moore. A newton-like method for solving rank constraint matrix inequalities. *Automatica*, 42(11) :1875–1882, 2006.
- [29] D.T. Pham, J. Verron, and M.C. Roubaud. A singular evolutive extended kalman filter for data assimilation in oceanography. *Journal of Marine Systems*, 16(3-4) :323 – 340, 1998.
- [30] S. T. Smith. *Geometric optimization methods for adaptive filtering*. PhD thesis, Division of Applied Sciences, Harvard University, Cambridge, Massachusetts, 1993.
- [31] S. T. Smith. Covariance, subspace, and intrinsic cramér-rao bounds. *IEEE Trans. Signal Process.*, 53 :1610–1630, 2005.
- [32] M. W. Trosset. Distance matrix completion by numerical optimization. *Computational Optimization and Applications*, 17(1) :11–22, 10 2000.
- [33] K. Tsuda, G. Ratsch, and M. Warmuth. Matrix exponentiated gradient updates for on-line learning and bregman projection. *Journal of Machine Learning Research*, 6 :995–1018, 2005.
- [34] B. Vandereycken, P.-A. Absil, and S. Vandewalle. Embedded geometry of the set of symmetric positive semidefinite matrices of fixed rank. In *Proceedings of the IEEE 15th Workshop on Statistical Signal Processing*, pages 389–392, 2009.
- [35] B. Vandereycken and S. Vandewalle. A riemannian optimization approach for computing low-rank solutions of lyapunov equations. *preprint*, 2009.