Covariance matrices with prescribed null entries

D. Chafaï and D. Concordet

http://www.lsp.ups-tlse.fr/Chafai/

Technical Report. Toulouse, May 2006.

Abstract

For a Gaussian random vector X with covariance matrix K, two distinct components X_i and X_j are marginally independent if and only if $K_{i,j} = 0$. In these notes, we consider the maximum likelihood estimation of the covariance matrix K under the constraint that K has some prescribed null entries. This estimation problem arise in many applied contexts, and is often fully addressed only when the pattern of prescribed zeroes for K has a simple structure, e.g. K has a block diagonal structure for instance. In these notes, we do not use any simplified structural assumptions on the pattern of zeroes.

Contents

The maximum likelihood (ML) for structured covariance matrices has been studied for a long time by statisticians. What is said on these notes is already present in the literature, see for instance [CDR06] and references therein (for patterns on K) and [DRV05] and references therein (for patterns on K^{-1}). For patterns in K^{-1}, the link with graphical models, chordal graphs, sparse Choleski, dual Lagrangian formulations, and max-det problems is addressed in [DRV05] and [VBW96] and references therein. The recent book [BV04] by Boyd and Vandenberghe provides many interesting material to address the questions considered in these notes.

1 Introduction

Let $X_1, \ldots, X_N$ be a random sample drawn from a multivariate centred Gaussian distribution on $\mathbb{R}^n$ with unknown non-singular covariance matrix K. Assume that some prescribed entries of K are known to be null, i.e. $K_{i,j} = 0$ for any $(i,j) \in I$, for some known subset I of $\{(i,j); 1 \leqslant i < j \leqslant n\}$. Consider now the maximum likelihood estimation (MLE) of K from $X_1, \ldots, X_N$. The log-likelihood $\mathbf{L}(K)$ is given by

$$\mathbf{L}(K) = -\frac{nN}{2}\log(2\pi) + \frac{N}{2}\log\det(K^{-1}) - \frac{1}{2}\sum_{i=1}^{N} X_i^\top K^{-1} X_i, \tag{1}$$

where $X_1, \ldots, X_N$ must be seen as column vectors. The maximum-likelihood estimator is thus obtained by the maximisation of $\mathbf{L}$ over the set $\mathcal{K}_n^{+*}(I)$ of non-singular $n \times n$ covariance matrices K such that $K_{i,j} = 0$ for any $(i,j) \in I$.

When $I = \emptyset$, then $\mathcal{K}_n^{+*}(I)$ is the whole set of non-singular covariance matrices, and the maximum-likelihood estimator is given by the empirical covariance matrix

$$\mathbb{X} := \frac{1}{N}\sum_{i=1}^{N} X_i X_i^\top, \tag{2}$$

and is almost-surely symmetric definite positive. In presence of additional parameters, the Cholesky factorisation allows for instance the set up of a gradient-like approach. Similarly, a block Cholesky factorisation allows to address the case where I is such that the elements of $\mathcal{K}_n^{+*}(I)$ are block diagonal by tensorisation.

The goal of these notes is to explore some aspects of these sets $\mathcal{K}_n^{+*}(I)$. No closed form is known for the maximum likelihood when I has a generic structure, see for instance [CDR06] and references therein.

2 Some useful sets of matrices

For any n in $\mathbb{N}^*$, let $\mathcal{M}_n$ be the vector space of square $n \times n$ matrices with real entries. In the sequel, a vector v of $\mathbb{R}^n$ is by default a column vector, and $v^\top$ is a row vector. For any A and B in $\mathcal{M}_n$, we denote by $A \circ B$ their Hadamard product, defined by $(A \circ B)_{i,j} = A_{i,j}B_{i,j}$ for any $1 \leqslant i,j \leqslant n$. We denote by GL_n the set of non-singular elements of $\mathcal{M}_n$, by $\mathcal{P}_n$ the set of $n \times n$ permutation matrices, by $\mathcal{O}_n$ the set of orthogonal elements of $\mathcal{M}_n$, and by $\mathcal{S}_n$ the set of symmetric elements of $\mathcal{M}_n$. Recall that $\mathcal{P}_n \subset \mathcal{O}_n \subset \mathrm{GL}_n$. We denote by $\mathcal{D}_n$ the set of diagonal elements of $\mathcal{M}_n$. The vector space $\mathcal{M}_n$ is isomorphic to $\mathbb{R}^{n\times n}$, and $\mathcal{S}_n$ is a sub-vector-space of $\mathcal{M}_n$ of dimension $n(n+1)/2$. Finally, GL_n is an open and dense multiplicative sub-group of $\mathcal{M}_n$, whereas $\mathcal{P}_n$ is a finite sub-group of GL_n isomorphic to the symmetric group Σ_n.

We identify $\mathcal{M}_n$ with the standard Hilbert space $\mathbb{R}^{n\times n}$. The scalar product is given by the Frobenius formula $A \cdot B := \mathrm{Tr}(AB^\top) = \sum_{i,j=1}^{n} A_{i,j}B_{i,j}$. The canonical basis $\{E^{i,j}; 1 \leqslant i,j \leqslant n\}$ of $\mathbb{R}^{n\times n}$ forms an orthonormal basis of the Hilbert space $\mathcal{M}_n$. The entries of $E^{i,j}$ are all zero except the entry at line i and column j which is equal to 1. We denote by I_n the identity matrix, by $\{e_1, \ldots, e_n\}$ the canonical basis of $\mathbb{R}^n$.

As a sub-Hilbert-space of $\mathcal{M}_n$ of dimension $n(n+1)/2$, the set $\mathcal{S}_n$ is closed with zero Lebesgue measure and empty interior in $\mathcal{M}_n$. We identify $\mathcal{S}_n$ with the Hilbert space $\mathbb{R}^{n(n+1)/2}$, equipped with the Frobenius scalar product $A \cdot B := \mathrm{Tr}(AB) = \sum_{i,j=1}^n A_{i,j} B_{i,j}$. This Hilbert structure on $\mathcal{S}_n$ coincides with the trace of the Hilbert structure of $\mathcal{M}_n$ via the natural inclusion $\mathcal{S}_n \subset \mathcal{M}_n$. It corresponds to see $\mathcal{S}_n$ an a sub-vector space of $\mathbb{R}^{n(n+1)/n}$, with the trace metric and induced topology. Notice that the scalar product on $\mathcal{S}_n$ is not the standard Euclidean scalar product of $\mathbb{R}^{n(n+1)/2}$ since the off-diagonal elements have weight 2, namely $A \cdot B = \sum_{i=1}^n A_{i,i} B_{i,i} + 2 \sum_{1 \leqslant i<j \leqslant n} A_{i,j} B_{i,j}$. The family $\{\frac{1}{2}(E^{i,j} + E^{j,i}); 1 \leqslant i \leqslant j \leqslant n\}$ forms an orthonormal basis of the Hilbert space $\mathcal{S}_n$. The set $\mathcal{S}_n \cap \mathrm{GL}_n$ is an open dense subset of the Hilbert space $\mathcal{S}_n$. We denote by $\mathcal{S}_n^+$ the set of elements of $\mathcal{S}_n$ with non-negative spectrum, and by $\mathcal{S}_n^{+*}$ the set of elements of $\mathcal{S}_n$ with positive spectrum. We also define $\mathcal{D}_n^+ := \mathcal{D}_n \cap \mathcal{S}_n^+ \simeq [0, +\infty)^n$ and $\mathcal{D}_n^{+*} := \mathcal{D}_n \cap \mathcal{S}_n^{+*} \simeq (0, +\infty)^n$.

Let $\mathcal{I}_n$ be the set of subsets of $\{(i,j); 1 \leqslant i < j \leqslant n\}$. For any $I \in \mathcal{I}_n$, let us define the sub-Hilbert-space $\mathcal{K}_n(I)$ of $\mathcal{S}_n$ by

$$\mathcal{K}_n(I) := \{A \in \mathcal{S}_n; \forall (i,j) \in I, A_{i,j} = 0\}.$$

Notice that $\mathcal{D}_n \subset \mathcal{K}_n(I)$, that $\mathcal{K}_n(\emptyset) = \mathcal{S}_n$, and that $\mathcal{K}_n(I_{\max}) = \mathcal{D}_n$ where $I_{\max} := \{(i,j); 1 \leqslant i < j \leqslant n\}$. The set $\mathcal{K}_n(I)$ is a Hilbert space equipped with the Frobenius scalar product inherited from the Hilbert space $\mathcal{S}_n$. It is the vector span of the orthonormal family $\{\frac{1}{2}(E^{i,j} + E^{j,i}); (i,j) \notin I\}$. The orthogonal $(\mathcal{K}_n(I))^\perp$ of $\mathcal{K}_n(I)$ in the Hilbert space $\mathcal{S}_n$ is the vector span of the orthonormal family $\{\frac{1}{2}(E^{i,j} + E^{j,i}); (i,j) \in I\}$.

We also define $\mathcal{K}_n^+(I) = \mathcal{K}_n(I) \cap \mathcal{S}_n^+$ and $\mathcal{K}_n^{+*}(I) = \mathcal{K}_n(I) \cap \mathcal{S}_n^{+*}$, which are both convex cones of the Hilbert spaces $\mathcal{K}_n(I)$ or $\mathcal{S}_n$. Notice that $\mathcal{K}_n^+(\emptyset) = \mathcal{S}_n^+$, and $\mathcal{K}_n^{+*}(\emptyset) = \mathcal{S}_n^{+*}$. Moreover, $I \subset J$ implies $\mathcal{K}_n(J) \subset \mathcal{K}_n(I)$, and $\cap_{I \in \mathcal{I}_n} \mathcal{K}_n(I) = \mathcal{K}_n(I_{\max}) = \mathcal{D}_n$. Similar properties hold for $\mathcal{K}_n^+$ and $\mathcal{K}_n^{+*}$ by replacing $\mathcal{D}_n$ with $\mathcal{D}_n^+$ and $\mathcal{D}_n^{+*}$ respectively.

The convex cone $\mathcal{K}_n^{+*}(I)$ is not empty nor closed in the Hilbert space $\mathcal{S}_n$, since it is the intersection of the closed set $\mathcal{K}_n(I)$ with the open set $\mathcal{S}_n^{*+}$. However, $\mathcal{K}_n^{+*}(I)$ is open in the Hilbert space $\mathcal{K}_n(I)$, as the trace of the open set $\mathcal{S}_n^{+*}$ onto $\mathcal{K}_n(I)$. The closure of $\mathcal{K}_n^{+*}(I)$ is given by the closed convex cone $\mathcal{K}_n^+(I)$ (both in the Hilbert space $\mathcal{K}_n(I)$ and in the Hilbert space $\mathcal{S}_n$).

For any $I \in \mathcal{I}_n$, we denote by C_I the symmetric $n \times n$ matrix defined by $C_I := \sum_{(i,j) \notin I} (E^{i,j} + E^{j,i})$. For any $A \in \mathcal{S}_n$, the matrix $A \circ C_I$ is the orthogonal projection $p_{\mathcal{K}_n(I)}(A)$ of A onto $\mathcal{K}_n(I)$.

We define also the set $\mathcal{R}_n^{+*}(I) := \{A; A^{-1} \in \mathcal{K}_n^{+*}(I)\}$, which is suitable for patterns of zeroes on the inverse of non-singular covariance matrices.

3 Marginal versus mutual independence for Gaussian vectors

For any vector X and any unordered sequence I of components indexes, we denote by X_I the corresponding components of X taken in the natural order. Similarly, for a matrix K, we denote by $K_{I,J}$ the sub-matrix of K corresponding to the entries of K with indexes in $I \times J$, taken in the natural order. The components of X_I are said to be *conditionally independent given* X_J when the conditional law $\mathcal{L}(X_I \,|\, X_J)$ is a product distribution. The conditional independence of the components of X_I given X_J means that the marginal components of the multivariate law $\mathcal{L}(X_I \,|\, X_J)$ are mutually independent. The conditional independence of the components of X_I given X_J corresponds also to the fact that the conditional law $\mathcal{L}(X_{I_1} \,|\, X_{J \cup I_2})$ does not depend on X_{I_2}, for any partition $I_1 \cup I_2 = I$.

Recall that the components of a Gaussian random vector X of covariance matrix K are mutually independent if and only if they are two by two independent, and if and only if K is diagonal. Let X be now a centred Gaussian random vector with non-singular covariance matrix K. Let I and J be two disjoint sequences of indexes. It is a classical result of probability Theory that the conditional law $\mathcal{L}(X_I \mid X_J)$ is a multivariate centred Gaussian distribution with covariance matrix[1]

$$K_{I,I} - K_{I,J}(K_{J,J})^{-1}K_{J,I} = ((K^{-1})_{I,I})^{-1}.$$

Consequently, the components of X_I are conditionally independent given X_J if and only if $((K^{-1})_{I,I})^{-1}$ is diagonal. This holds if and only if $(K^{-1})_{I,I}$ is diagonal. In particular, let us take $I = \{i, j\}$ with $i \neq j$ and $J = \{k; k \notin I\}$. We have that X_i and X_j are conditionally independent given the rest of the components of X if and only if $(K^{-1})_{i,j} = 0$. The reader may find more details on these classical results in [Dem72] and [Lau96] and references therein.

Let X be a Gaussian random vector with non-singular covariance matrix K, and let $i \neq j$ be two components indexes. Then, $(K_{i,j})^{-1} = 0$ if and only if X_i and X_j are mutually independent given the other components of X. Prescribing zeroes in the inverse covariance matrix K^{-1} corresponds to impose conditional independence on the components. Patterns of zeroes in K^{-1} arise in Gaussian graphical models, see for instance [Lau96] and references therein. In contrast, $K_{i,j} = 0$ if and only if X_i and X_j are independent, i.e. X_i and X_j are marginally independent. Patterns of zeroes in K arise for instance in linear models for longitudinal data. Therefore, we should not confuse patterns of zeroes for the covariance matrix K with patterns of zeroes for the inverse covariance matrix K^{-1}. Of course, K^{-1} is itself a covariance matrix, but the random vector X is drawn from K and not from K^{-1}, and thus, the estimation problem is not the same.

We emphasise that most of patterns of zeroes are not stable by matrix inversion, even for symmetric positive definite matrices. There are however notable exceptions. Namely, a non-singular matrix K is block diagonal if and only if K^{-1} is block diagonal. A covariance matrix of a Gaussian random vector is block diagonal if and only if the corresponding blocks of components of the vector are two by two marginally independent. Two by two marginal independence is equivalent to mutual independence for blocks of components of Gaussian vectors. It implies conditional independence of any couple of blocks given the other blocks. Notice that the blocks can be shuffled by permutations of the coordinates, which corresponds to consider $PKP^\top$ where P is the suitable permutation matrix. In particular, if $K_{i,j} = 0$ for any $i \neq j$, then $(K^{-1})_{i,j} = 0$ for any $j \neq i$. This corresponds to the fact that X_i and X_j are marginally independent for any $j \neq i$ if and only if X_i and X_j are mutually independent (given the others components) for any $i \neq j$.

Finally, it follows from the discussion above that for any Gaussian random vector X with non-singular covariance matrix K, and for any $i \neq j$, the marginal independence of X_i and X_j where $X \sim \mathcal{N}(0, K^{-1})$ is equivalent to conditional independence of Y_i and Y_j where $Y \sim \mathcal{N}(0, K)$. Furthermore, since $K = (K^{-1})^{-1}$, the two notions of independence can be swapped in the latter sentence.

Remark 3.1. The joint estimation of the mean can be done quite easily and separately, since the estimator of the mean μ is the empirical mean $\widehat{\mu}$. The empirical covariance matrix (2) must be replaced by $N^{-1}\sum_{i=1}^N (X_i - \widehat{\mu})(X_i - \widehat{\mu})^\top$.

[1] It is customary to attribute this matrix identity to Issai Schur (1875-1941). The term $K_{I,I} - ((K^{-1})_{I,I})^{-1} = K_{I,J}(K_{J,J})^{-1}K_{J,I}$ is known as the *Schur complement*.

Proposition 3.2. *Let $I \in \mathcal{I}_n$, and let $X_1, \ldots, X_N$ be a random sample of a centred Gaussian distribution of $\mathbb{R}^n$ with covariance matrix $K \in \mathcal{K}_n^{+*}(I)$. If $\mathbb{X}$ is the empirical covariance matrix of the sample defined by (2), then a.-s., we have $\mathbb{X} \circ C_I \in \mathcal{K}_n^{+*}(I)$ for large enough sample size N. Moreover, as an element of $\mathcal{K}_n(I)$, the matrix $N^{-1}\mathbb{X} \circ C_I$ is a strongly consistent and unbiased estimator of K.*

Proof. We have $\mathbb{E}(\mathbb{X} \circ C_I) = \mathbb{E}(\mathbb{X}) \circ C_I = K \circ C_I = K$. The law of large numbers ensures that $\lim_{N\to\infty} \mathbb{X} = K$ a.-s and thus that $\lim_{N\to\infty} \mathbb{X} \circ C_I = K \circ C_I = K$ a.-s. This gives the consistency and the zero bias. Notice that $\mathbb{X} \circ C_I \in \mathcal{K}_n(I)$. Now, Since $K \in \mathcal{K}_n^{+*}(I)$ and since $\mathcal{K}_n^{+*}(I)$ is an open subset of the Hilbert space $\mathcal{K}_n(I)$, we get $\mathbb{X} \circ C_I \in \mathcal{K}_n^{+*}(I)$ for large enough N, a.-s. □

Proposition 3.3. *Let $I \in \mathcal{I}_n$ and let $X_1, \ldots, X_N$ be a random sample of a centred Gaussian distribution of $\mathbb{R}^n$ with covariance matrix $K \in \mathcal{R}_n^{+*}(I)$. If $\mathbb{X}$ is the empirical covariance matrix of the sample defined by (2), then, a.-s., and for large enough N, we have $\mathbb{X} \in \mathcal{S}_n^{+*}$ and $\mathbb{X}^{-1} \circ C_I \in \mathcal{R}_n^{+*}(I)$. Moreover, a.-s. $\lim_{N\to\infty}(\mathbb{X}^{-1} \circ C_I)^{-1} = K$.*

Proof. Notice that $\mathbb{X} \in \mathcal{S}_n$. By the law of large numbers, a.-s. $\lim_{N\to\infty} \mathbb{X} = K$. Now, $K \in \mathcal{S}_n^{+*}$ and $\mathcal{S}_n^{+*}$ is an open subset of the Hilbert space $\mathcal{S}_n$. Therefore, a.-s., for large enough N, $\mathbb{X} \in \mathcal{S}_n^{+*}$. Also, a.-s. $\lim_{N\to\infty} \mathbb{X}^{-1} = K^{-1}$, and thus, $\lim_{N\to\infty} N\mathbb{X}^{-1} \circ C_I = K^{-1} \circ C_I = K^{-1}$. Notice that a.-s. $\mathbb{X}^{-1} \circ C_I \in \mathcal{K}_n(I)$ for large enough N. Since $K^{-1} \in \mathcal{K}_n^{+*}(I)$ where $\mathcal{K}_n^{+*}(I)$ is an open subset of the Hilbert space $\mathcal{K}_n(I)$, we get that a.-s. $\mathbb{X}^{-1} \circ C_I \in \mathcal{K}_n^{+*}(I)$ for large enough N. □

4 The log-likelihood on $\mathcal{K}_n^{+*}(I)$

Let $I \in \mathcal{I}_n$, and let $X_1, \ldots, X_N$ be a random sample drawn from a centred Gaussian law on $\mathbb{R}^n$ with covariance matrix $K \in \mathcal{K}_n^{+*}(I)$. Here $\mathcal{K}_n^{+*}(I)$ is viewed as a convex open cone of the Hilbert space $\mathcal{K}_n(I)$. Let us consider the estimation of K by the arg-maximum of the log-likelihood $\mathbf{L}$ defined by (1). For the moment, we do not know actually if the log-likelihood has a maximum on $\mathcal{K}_n^{+*}(I)$. For any $A \in \mathcal{K}_n^{+*}(I)$ and $B \in \mathcal{K}_n(I)$, we have

$$\begin{aligned}
2N^{-1}D(\mathbf{L})(A)(B) &= -A^{-1} \cdot B + N^{-1}\sum_{i=1}^{N}(A^{-1}X_i)^\top B(A^{-1}X_i) \\
&= -A^{-1} \cdot B + N^{-1}\left(\sum_{i=1}^{N}(A^{-1}X_i)(A^{-1}X_i)^\top\right) \cdot B \\
&= (-A^{-1} + (A^{-1}\mathbb{X}A^{-1})) \cdot B,
\end{aligned}$$

where $\mathbb{X}$ is the empirical covariance matrix of the sample defined by (2). Since $B \in \mathcal{K}_n(I)$, the gradient $\nabla\mathbf{L}(A)$ in the Hilbert space $\mathcal{K}_n(I)$ is given by

$$2N^{-1}\nabla\mathbf{L}(A) = C_I \circ (-A^{-1} + A^{-1}\mathbb{X}A^{-1}).$$

If the likelihood has an arg-maximum A over $\mathcal{K}_n^{+*}(I)$, then necessarily $\nabla\mathbf{L}(A) = 0$, which can be written as follows.

$$C_I \circ A^{-1} = C_I \circ (A^{-1}\mathbb{X}A^{-1}).$$

In other words, A^{-1} and $A^{-1}\mathbb{X}A^{-1}$ must have the same entries on the complement of I. Moreover, A must have null entries on I. The reader may find in [CDR06] a review on several algorithm for the computation of the MLE.

On $\mathcal{K}_n^{+*}(\emptyset) = \mathcal{S}_n^{+*}$, the matrix $\mathbb{X}$ belongs almost-surely to $\mathcal{S}_n^{+*}$ for $N > n$. Thus, almost-surely, for $N > n$, the maximiser of $\mathbf{L}$ on $\mathcal{K}_n^{+*}(I)$ exists, is unique, and is given by $\mathbb{X}$. When $I \neq \emptyset$, this shows that on $\mathcal{K}_n^{+*}(I)$, the function $\mathbf{L}$ is bounded above by its value at point $\mathbb{X}$.

The maximisation of $\mathbf{L}$ on $\mathcal{K}_n^{+*}(I)$ is equivalent (up to matrix inversion of the maximisers) to the maximisation of $\mathbf{H} : A \mapsto \mathbf{L}(A^{-1})$ on $\mathcal{R}_n^{+*}(I)$. The function $\mathbf{H}$ is strictly concave on $\mathcal{S}_n^{+*}$. Suppose that $\mathbf{H}$ has a maximiser of on $\mathcal{R}_n^{+*}(I)$, then we can deduce that it is unique if $\mathcal{R}_n^{+*}(I)$ is connected (for example if it is convex).

Let us consider the case where I is such that $\mathcal{K}_n(I)$ is the set of block diagonal matrices, with blocks of indexes $I_1, \ldots, I_r$. We have in particular that $K_{I_i,I_i} \in \mathcal{S}_{|I_i|}^{+*}$ for any $1 \leqslant i \leqslant r$. By the law of large numbers, almost-surely, for large enough sample size N, we have $\mathbb{X}_{I_i,I_i} \in \mathcal{S}_{|I_i|}^{+*}$ for any $1 \leqslant i \leqslant r$. Let $B \in \mathcal{K}_n(I)$ be defined by $B_{I_i,I_i} := (\mathbb{X}_{I_i,I_i})^{-1}$ for any $1 \leqslant i \leqslant r$. Then $B \in \mathcal{K}_n^{+*}(I) \cap \mathcal{R}_n^{+*}(I)$ and $B = C_I \circ B = C_I \circ (B\mathbb{X}B)$ (matrix multiplication by blocks). We have thus constructed a stationary point $A = B^{-1}$ of the likelihood.

5 The log-likelihood on $\mathcal{R}_n^{+*}(I)$

Let $I \in \mathcal{I}_n$, and let $X_1, \ldots, X_N$ be a random sample drawn from a centred Gaussian law on $\mathbb{R}^n$ with covariance matrix $K \in \mathcal{R}_n^{+*}(I)$. Here $\mathcal{K}_n^{+*}(I)$ is viewed as a convex open cone of the Hilbert space $\mathcal{K}_n(I)$. Let us consider the estimation of K by the arg-maximum of the log-likelihood $\mathbf{L}$ defined by (1). For the moment, we do not know actually if the log-likelihood has a maximum on $\mathcal{R}_n^{+*}(I)$. We replace the maximisation problem of $\mathbf{L}$ on $\mathcal{R}_n^{+*}(I)$ by the maximisation problem of $\mathbf{H}$ on $\mathcal{K}_n^{+*}(I)$, where $\mathbf{H}(K) := \mathbf{L}(K^{-1})$ for any $K \in \mathcal{R}_n^{+*}(I)$. The functional $\mathbf{H}$ is defined on the open convex cone $\mathcal{K}_n^{+*}(I)$ of the Hilbert space $\mathcal{K}_n(I)$. For any $A \in \mathcal{K}_n^{+*}(I)$ and $B \in \mathcal{K}_n(I)$, we have

$$\begin{aligned} 2N^{-1}D(\mathbf{H})(A)(B) &= A^{-1} \cdot B - N^{-1}\sum_{i=1}^{N} X_i^\top B X_i \\ &= (A^{-1} - \mathbb{X}) \cdot B, \end{aligned}$$

where $\mathbb{X}$ is the empirical covariance matrix of the sample defined by (2). Since $B \in \mathcal{K}_n(I)$, the gradient $\nabla\mathbf{H}(A)$ in the Hilbert space $\mathcal{K}_n(I)$ is given by

$$2\nabla\mathbf{H}(A) = C_I \circ (A^{-1} - \mathbb{X}).$$

If the likelihood has an arg-maximum A^{-1} over $\mathcal{R}_n^{+*}(I)$, then necessarily $\nabla\mathbf{H}(A) = 0$, which can be written as follows.

$$C_I \circ A^{-1} = C_I \circ \mathbb{X}.$$

In other words, A^{-1} and $\mathbb{X}$ must have the same entries outside I. Moreover, A must have null entries on I. The reader may find a dual Lagrangian expression of the ML in [DRV05], which sees the MLE as the max-det (maximum of Shannon entropy!) under the above constraints. The article [DRV05] gives also links with graphical models, with chordal graphs, with sparse Choleski factorisation, and numerical algorithms. There is an explicit expression of the MLE in terms of sparse Choleski factorisation when I is related to a chordal graph, cf. [DRV05].

On $\mathcal{K}_n^{+*}(\emptyset) = \mathcal{S}_n^{+*}$, the matrix $\mathbb{X}$ belongs almost-surely to $\mathcal{S}_n^{+*}$ for $N > n$. Thus, almost-surely, for $N > n$, the maximiser of $\mathbf{H}$ on $\mathcal{S}_n^{+*}$ exists, is unique, and is given by

$\mathbb{X}$. When $I \neq \emptyset$, we have $\mathcal{K}_n^{+*}(I) \subset \mathcal{S}_n^{+*}$, and thus $\mathbf{H}$ is bounded above on $\mathcal{K}_n^{+*}(I)$ by its value at point $\mathbb{X}$. Since $\mathbf{H}$ is strictly concave on $\mathcal{S}_n^{+*}$, a maximiser of $\mathbf{H}$ on the convex set $\mathcal{K}_n^{+*}(I)$ (and thus the MLE) is necessarily unique if it exists (we used that convexity implies connectivity). It remains to find necessary and/or sufficient conditions on N, I, $\mathbb{X}$, and K in order to ensure that the maximiser of $\mathbf{H}$ (and thus the MLE) on $\mathcal{K}_n^{+*}(I)$ exists.

6 Chordal graphs and the example of block diagonal patterns

7 Variance-Correlation representation of $\mathcal{S}_n^+$

Definition 7.1 (Correlation matrices)**.** An element A of $\mathcal{S}_n$ is a *correlation matrix of size* n if and only if A belongs to $\mathcal{S}_n^+$ and has unit diagonal entries. We denote by $\mathcal{C}_n$ the set of correlation matrices of size n. We also denote by $\mathcal{C}_n^*$ the set $\mathcal{C}_n \cap \mathcal{S}_n^{+*}$ of non-singular correlation matrices of size n.

The set $\mathcal{C}_n$ is thus a subset of the closed convex cone $\mathcal{S}_n^+$, obtained as the intersection of $\mathcal{S}_n^+$ with the affine sub-space (of the Hilbert space $\mathcal{S}_n$) constituted by the elements of $\mathcal{S}_n$ with unit diagonal entries.

Theorem 7.2 (Random vectors)**.** *The set $\mathcal{C}_n$ is equal to the set of covariance matrices of all random vectors of $\mathbb{R}^n$ with unit variances (i.e. the variance of each of the n components is 1). Moreover, the random vectors can be assumed centred and Gaussian.*

Proof. Follows from the fact that the set of covariance matrices of size n is precisely $\mathcal{S}_n^+$. The random vectors can be taken centred Gaussian since every element of $\mathcal{S}_n^+$ is the covariance matrix of a centred Gaussian vector of $\mathbb{R}^n$. □

Theorem 7.3 (Convex & compact)**.** *The off-diagonal entries of the elements of $\mathcal{C}_n$ belong to $[-1,+1]$, whereas the diagonal entries are all equal to 1. The set $\mathcal{C}_n$ is a non-empty convex compact subset of the Hilbert space $\mathcal{S}_n$, included in $\mathcal{S}_n^+$, and stable by the Hadamard product.*

Proof. The fact that $\mathcal{C}_n$ is closed follows from its definition as the intersection of an affine sub-vector space of $\mathcal{S}_n$ (which is closed) with the closed convex cone $\mathcal{S}_n^+$. We have $I_n \in \mathcal{C}_n$ and thus $\mathcal{C}_n$ is non empty. Another simple example of element of $\mathcal{C}_n$ is given by the matrix with all entries equal to 1. Any $C \in \mathcal{C}_n$ can by viewed as the correlation matrix of a centred random vector V of $\mathbb{R}^n$. Now, by the Cauchy-Schwarz inequality, $|\mathbb{E}(V_iV_j)|^2 \leqslant \mathbb{E}(V_i^2)\mathbb{E}(V_j^2)$. Thus the absolute value of the off-diagonal entries of C are bounded by 1 since the diagonal entries are equal to 1. Thus $\mathcal{C}$ is bounded, as expected. For the convexity, let C and C' be in $\mathcal{C}_n$ and let $t \in [0,1]$. Let X and Y be two independent and centred Gaussian vectors of $\mathbb{R}^n$ with respective covariance matrices C and C'. Then $Z := \sqrt{t}X + \sqrt{1-t}Y$ is a centred Gaussian vector with covariance matrix $tC+(1-t)C'$. Moreover, the variances of the components of Z are all equal to 1. Therefore, $tC + (1-t)C' \in \mathcal{C}_n$. It remains to show that $\mathcal{C}_n$ is stable by the Hadamard product. The covariance matrix of the centred random vector $X \circ Y$ (i.e. with i^{th} component X_iY_i) is $C \circ C'$. Moreover, the variances of the components of $X \circ Y$ are all equal to 1. Notice that $X \circ Y$ is not Gaussian. Notice finally that it is possible to give a proof of this Theorem by using an inductive construction of $\mathcal{C}_n$, but it is slightly longer. □

It is tempting to ask if the upper off-diagonal entries of the elements of $\mathcal{C}_n$ cover completely the cube $[-1,+1]^{n(n-1)/2}$ (which is the centred ℓ^∞ ball of $\mathbb{R}^{n(n-1)/2}$ of unit radius). The answer in negative. Apart from the low dimensional cases $n = 1$ and $n = 2$, the upper off-diagonal entries of the elements of $\mathcal{C}_n$ form a strict non-empty convex compact subset of $[-1, +1]^{n(n-1)/2}$. Let us give counter examples for $n = 3$. The following two elements of $\mathcal{S}_n$ do not belong to $\mathcal{S}_n^+$, even if they look like elements of $\mathcal{C}_n$.

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \quad \text{and} \quad \frac{1}{3}\begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \\ 2 & -2 & 3 \end{pmatrix}.$$

Hence, the Cauchy-Schwarz inequality between diagonal and off-diagonal entries does not gather all the constraints defining $\mathcal{C}_n$ as a subset of $\mathcal{S}_n^+$.

Theorem 7.4 (Variance-Correlation representation of $\mathcal{S}_n^+$). *We have the following representation of the convex closed cone $\mathcal{S}_n^+$.*

$$\mathcal{S}_n^+ = \{DCD; \text{ where } (D, C) \in \mathcal{D}_n^+ \times \mathcal{C}_n\}.$$

Moreover, $D = \mathrm{Diag}(\sqrt{A_{1,1}}, \dots, \sqrt{A_{n,n}})$ where $A := DCD$, and C is uniquely determined by A on the set of indexes $\{(i, j) \in \{1, \dots, n\}^2; A_{i,i}A_{j,j} > 0\}$ via $C_{i,j} = A_{i,i}/D_iD_j$. Finally, DCD belongs to $\mathcal{S}_n^{+*}$ if and only if $(D, C) \in \mathcal{D}_n^{+*} \times \mathcal{C}_n^*$, and the couple (D, C) is uniquely determined by DCD in that case.

Proof. Let $D \in \mathcal{D}_n^+$ and $C \in \mathcal{C}_n$. Let X be a (column) random centred Gaussian vector of $\mathbb{R}^n$ with covariance matrix C. Then DX is a centred Gaussian random vector of $\mathbb{R}^n$ with covariance DCD. Therefore, $DCD \in \mathcal{S}_n^+$. Conversely, for any $A \in \mathcal{S}_n^+$, let X be a centred Gaussian random vector of $\mathbb{R}^n$ with covariance matrix A. Let $D \in \mathcal{D}_n^+$ be defined for any $1 \leqslant i \leqslant n$ by $D_i = \sqrt{\mathbb{E}(X_i^2)}$. Notice that by the Cauchy-Schwarz inequality, $|A_{i,j}| \leqslant D_iD_j$ for any $1 \leqslant i, j \leqslant n$. Thus, $A_{i,j} = 0$ for any $1 \leqslant i, j \leqslant n$ such that $D_iD_j = 0$. Let Y be a centred random vector of $\mathbb{R}^n$ with unit variances. Let Z be the centred random vector given for any $1 \leqslant i \leqslant n$ by $Z_i := X_i/D_i$ if $D_i > 0$ and $Z_i = Y_i$ if not. Then the n components of Z have unit variance, and thus its covariance matrix C belongs to $\mathcal{C}_n$. Moreover $\sqrt{D_iD_j}C_{i,j} = A_{i,j}$ for any $1 \leqslant i, j \leqslant n$. Therefore, we get $A = DCD$ as expected.

From $A = DCD$ we get $A_{i,j} = D_iC_{i,j}D_j$ for any $1 \leqslant i, j \leqslant n$. In particular, $A_{i,i} = D_i^2$ for any $1 \leqslant i \leqslant n$, since $C_{i,i} = 1$. Thus, $C_{i,j} = A_{i,j}/D_iD_j$ for any $1 \leqslant i, j \leqslant n$ such that $A_{i,i}A_{j,j} = D_i^2D_j^2 > 0$. Finally, let $(A, C, D) \in \mathcal{S}_n^+ \times \mathcal{D}_n^+ \times \mathcal{C}_n$ such that $A = DCD$. Since $\det(A) = \det(D)^2 \det(C)$, we get that $A \in \mathcal{S}_n^{+*}$ if and only if $D \in \mathcal{D}_n^{+*}$ and $C \in \mathcal{C}_n^*$. The uniqueness of (D, C) comes from the fact that the diagonal of A is strictly positive in that case. □

Let us define an equivalence relation on $\mathcal{S}_n^{+*}$ by $A \sim B$ if and only if they have the same correlation matrix. Then the quotient set $\mathcal{S}_n^{+*}/\sim$ is by definition the set $\mathcal{C}_n^*$ of non-singular correlation matrices of size n.

Notice that $(DMD')_{i,j} = D_iM_{i,j}D'_j$, for any $M \in \mathcal{M}_n$, any $D, D' \in \mathcal{D}_n$, and any $1 \leqslant i, j \leqslant n$. The matrix product and the Hadamard product coincide on $\mathcal{D}_n$. Moreover, if $M_1, M_2 \in \mathcal{M}_n$ and $D_1, D_2, D'_1, D'_2 \in \mathcal{D}_n$, then

$$(D_1M_1D'_1) \circ (D_2M_2D'_2) = (D_1D_2)(M_1 \circ M_2)(D'_1D'_2).$$

Combined with the variance-correlation representation of $\mathcal{S}_n^+$ and with the stability of $\mathcal{C}_n$ by the Hadamard product, the identity above contains the stability of $\mathcal{S}_n^+$ by the Hadamard product (the Schur Theorem).

Theorem 7.5 (Inductive construction of $\mathcal{C}_n$). *We have $\mathcal{C}_1 = \{1\}$, and for any $n > 1$,*

$$\mathcal{C}_{n+1} = \{C \vee c \text{ where } C \in \mathcal{C}_n \text{ and } c \in \mathcal{E}(C)\},$$

where $\mathcal{E}(C)$ is the image by C of the ellipsoid $\{\alpha \in \mathbb{R}^n;$ such that $\alpha^\top C\alpha \leqslant 1\}$; and where $C' := C \vee c$ is the element of $\mathcal{S}_{n+1}$ defined by $C'_{i,j} := C'$ for any $1 \leqslant i, j \leqslant n$ and $C'_{n+1,n+1} := 1$ and $C'_{i,n+1} := C'_{n+1,i} := c_i$ for any $1 \leqslant i \leqslant n$.

Proof. Let us assume that the construction is valid for some $n \geqslant 1$. Let $C' \in \mathcal{C}_{n+1}$, and let X be a centred Gaussian random vector of $\mathbb{R}^{n+1}$, with covariance matrix C'. The component X_{n+1} can be written $X_{n+1} = \alpha_1 X_1 + \cdots + \alpha_n X_n + Z$ where $\alpha \in \mathbb{R}^n$ and where Z is a centred univariate Gaussian random variable, independent of $X_1, \ldots, X_n$. Let $C \in \mathcal{C}_n$ be the covariance matrix of the first n components of X. We have $C' = C \vee c$ where $c := C\alpha$. Since X has unit variances, we get also $1 = \tau + \alpha^\top C\alpha$, where τ is the variance of Z. Thus, we get $\alpha^\top C\alpha = 1 - \tau \leqslant 1$, and the vector $c = C\alpha$ belongs to $\mathcal{E}(C)$.

Conversely, let $C \in \mathcal{C}_n$ and $c \in \mathcal{E}(C)$. By definition of $\mathcal{E}(C)$, there exists $\alpha \in \mathbb{R}^n$ such that $\alpha^\top C\alpha \leqslant 1$ and $c = C\alpha$. Let X be a centred Gaussian random vector of $\mathbb{R}^n$ with covariance matrix C. Let Z be a univariate centred Gaussian random variable, independent of X, and of variance $1 - \alpha^\top C\alpha \geqslant 0$. Then $(X, \alpha_1 X_1 + \cdots + \alpha_n X_n + Z)$ is a centred Gaussian random vector of $\mathbb{R}^{n+1}$, with covariance matrix $C \vee c$. Consequently, $C \vee c \in \mathcal{C}_{n+1}$. □

Similarly, the same line of reasoning gives that $\mathcal{S}_1^+ = [0, +\infty)$ and for any $n > 0$,

$$\mathcal{S}_{n+1}^+ = \{A \vee_\tau c \text{ where } A \in \mathcal{S}_n^+, \text{ and } c \in \mathrm{Im}(A), \text{ and } \tau \geqslant 0\},$$

where this time $(A \vee_\tau c)_{i,j} := A_{i,j}$ for any $1 \leqslant i, j \leqslant n$, and $(A \vee_\tau c)_{n+1,i} := (A \vee_\tau c)_{i,n+1} = c_i$ for any $1 \leqslant i \leqslant n$, and $(A \vee_\tau c)_{n+1,n+1} := \tau + \alpha^\top A\alpha$ with α such that $A\alpha = c$. Actually, this method allows more generally to inductively construct the subsets of $\mathcal{S}_n^+$ defined by prescribing a constant diagonal. The inductive definition of $\mathcal{C}_n$ given above is in a way dimensionally oriented. It can be symmetrised by considering the action of the symmetric group Σ_n (conjugacy by permutation matrices). Namely,

$$\mathcal{C}_{n+1} = \{P(C \vee c)P^\top; (C, P) \in \mathcal{C}_n \times \mathcal{P}_n, c \in \mathcal{E}(C)\}.$$

Notice that I_n belongs to $\mathcal{C}_n$ and that $\mathcal{E}(I_n)$ is the centred unit Euclidean ball of radius 1 of $\mathbb{R}^n$. The elements of $\mathcal{C}_{n+1}$ which can be written $I_n \vee c$ are characterised by $\|c\|_2 \leqslant 1$, which is much stronger than $\|c\|_\infty \leqslant 1$ in high dimension n.

Theorem 7.6. *For any $1 \leqslant i < j \leqslant n$, the map $\varphi_{i,j} : C \in \mathcal{C}_n \mapsto C_{i,j} \in [-1, +1]$ has range $[-1, +1]$. Moreover, for any $C \in \mathcal{C}_n$, the eigenvalues of C sum to n and belong to $[0, n]$.*

Proof. For the first statement, consider $P(I_n \vee c)P^\top$ with a suitable $P \in \mathcal{P}_n$ and a suitable $c \in \mathbb{R}^{n-1}$ with $\|c\|_2 \leqslant 1$. For the second statement, use Gershgorin-Hadamard Theorem and notice that $C \in \mathcal{S}_n^+$. The eigenvalue n is obtained with $C = (1, \ldots, 1)^\top (1, \ldots, 1)$. □

More generally, let C be in $\mathcal{C}_n$, and let $w_1, \ldots, w_n$ be the orthonormal eigenvectors of C associated to the eigenvalues $\lambda_1 \geqslant \ldots \geqslant \lambda_n$ given by the spectral Theorem. Then $\mathcal{E}(C)$ is the image of the ellipsoid $\{\beta \in \mathbb{R}^n; \sum_{i=1}^n \lambda_i \beta_i^2 \leqslant 1\}$ by the endomorphism $\beta \mapsto \sum_{i=1}^n \lambda_i \beta_i w_i$ of $\mathbb{R}^n$. This gives $\mathcal{E}(C) = \{c \in \mathbb{R}^n;$ where $c_i = 0$ if $\lambda_i = 0$ and $\sum_{\lambda_i > 0} \lambda_i^{-1} (w_i^\top c)^2 \leqslant 1\}$. It is an ellipsoid of $\mathbb{R}^n$ when C is not singular, and an imbedded lower dimensional ellipsoid when C is singular. Notice that $\mathcal{E}(C)$ contains the collection of n vectorial segments

$\{tw_i; t \in \sqrt{\lambda}[-1,+1]\}$ of $\mathbb{R}^n$. Let us consider for instance the example where $C \in \mathcal{C}_n$ is the matrix with all entries equal to 1. In that case, $C = (1,\dots,1)^\top(1,\dots,1)$ is of rank 1 with eigenvalues 0 (multiplicity $n-1$) and n (multiplicity 1). Then $\mathcal{E}(C)$ is in that case the vectorial segment $\{(t,\dots,t)^\top; t \in [-1,+1]\}$ of $\mathbb{R}^n$, which is a diagonal of the cube $[-1,+1]^n$ of $\mathbb{R}^n$.

8 Special sub-space $\mathcal{K}_n(I)$ of $\mathcal{S}_n$ and sub-cone $\mathcal{K}_n^+(I)$ of $\mathcal{S}_n^+$

Theorem 8.1 (Invariances of $\mathcal{K}_n^+(I)$)**.** *For any $I \in \mathcal{I}_n$, the set $\mathcal{K}_n^+(I)$ is a closed convex cone of $\mathcal{K}_n(I)$ or $\mathcal{S}_n$, stable by the Hadamard product. Moreover, for any $P \in \mathcal{P}_n$, there exists $I_P \in \mathcal{I}_n$ with* $\mathrm{card}(I_P) = \mathrm{card}(I)$ *such that $PAP^\top \in \mathcal{K}_n^+(I_P)$ for any $A \in \mathcal{K}_n^+(I)$.*

The last statement of the Theorem above corresponds to the action by conjugacy of the symmetric group Σ_n on the finite set $\{\mathcal{K}_n^+(I); I \in \mathcal{I}_n\}$ of cardinal $2^{n(n-1)/2}$. It basically expresses that the number of null entries is preserved by the permutations of lines and columns.

The orthogonal projection $p_{\mathcal{K}_n(I)}$ in $\mathcal{S}_n$ onto $\mathcal{K}_n(I)$ can be expressed as a Hadamard product with a special matrix constructed from I. Namely, for any $I \in \mathcal{I}_n$ and any $A \in \mathcal{S}_n$, we have $p_{\mathcal{K}_n(I)}(A) = A \circ C_I$, where

$$C_I := \sum_{(i,j)\notin I} (E^{i,j} + E^{j,i}).$$

The projection onto $\mathcal{K}_n(I)$ corresponds to put zeroes in the entries of A prescribed by I.

Definition 8.2 (Admissible elements of $\mathcal{I}_n$)**.** We say that an element $\mathcal{I}_n$ of $\mathcal{I}_n$ is *admissible* if and only if $C_I \in \mathcal{S}_n^+$. We denote by $\mathcal{J}_n$ the subset of admissible element of $\mathcal{I}_n$.

Notice that $\emptyset$ is admissible, and that $C_\emptyset = (1,\dots,1)^\top(1,\dots,1)$. Moreover, for any $I \in \mathcal{I}_n$, we have $C_I = p_{\mathcal{K}_n(I)}(C_\emptyset)$, and that $C_\emptyset$ is the neutral element of the Hadamard product. Notice also that the set $I_{\max}$ is admissible, and that $C_{I_{\max}} = I_n$. Recall that $\mathcal{K}_n(I_{\max}) = \mathcal{D}_n$. More generally, are admissible the elements I of $\mathcal{I}_n$ which corresponds to the case where $\mathcal{K}_n(I)$ is the sub-vector-space of $\mathcal{S}_n$ constituted with block diagonal symmetric matrices with prescribed blocks positions. However, a singleton I is not admissible in general in dimension $n > 2$. Namely, here is a counter example in dimension $n = 3$, which corresponds to the singleton $I = \{(1,3)\}$.

$$A := \begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix} \in \mathcal{S}_n^+ \quad \text{but} \quad p_{\mathcal{K}_n(I)}(A) = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix} \notin \mathcal{S}_n^+.$$

The matrix C_I associated to this counter example is given by

$$C_I = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \notin \mathcal{S}_n^+.$$

Theorem 8.3 (Admissibility and projection onto $\mathcal{K}_n(I)$)**.** *The set $\{C_I; I \in \mathcal{J}_n\}$ is a finite subset of $\mathcal{C}_n$, containing $C_\emptyset$ and I_n. It is exactly the subset of $\mathcal{C}_n$ constituted with the correlation matrices of size n with all entries in $\{0,1\}$. In particular, for any $I \in \mathcal{I}_n$, we have that $I \in \mathcal{J}_n$ if and only if $C_I \in \mathcal{C}_n$. Last but not least, for any $I \in \mathcal{I}_n$, we have $I \in \mathcal{J}_n$ if and only if $p_{\mathcal{K}_n(I)}(A) \in \mathcal{S}_n^+$ for any $A \in \mathcal{S}_n^+$.*

Proof. The first parts are immediate. Let us establish the last equivalence. If $I \in \mathcal{J}_n$, then $C_I := \sum_{(i,j)\notin I}(E^{i,j} + E^{j,i}) \in \mathcal{S}_n^+$. By the stability of $\mathcal{S}_n^+$ by the Hadamard product, $p_{\mathcal{K}_n(I)}(A) = A \circ C_I \in \mathcal{S}_n^+$ provided that $A \in \mathcal{S}_n^+$. Conversely, assume that $p_{\mathcal{K}_n(I)}(A) \in \mathcal{S}_n^+$ for any $A \in \mathcal{S}_n^+$. Then, $p_{\mathcal{K}_n(I)}(C_\emptyset) = C_\emptyset \circ C_I = C_I$ since $C_\emptyset$ is the neutral element of the Hadamard product. Therefore, $C_I \in \mathcal{S}_n^+$ and thus $I \in \mathcal{J}_n$. □

Theorem 8.4 (Inductive construction of admissible elements of $\mathcal{I}_n$). *We use here the notations used for the inductive construction of $\mathcal{C}_n$. We have $\mathcal{J}_1 = \{\emptyset\}$ and for any $n > 0$,*

$$\mathcal{J}_{n+1} = \{I' \in \mathcal{I}_{n+1} \text{ such that } C_{I'} = C_I \vee c \text{ with } I \in \mathcal{J}_n \text{ and } c \in \mathcal{E}(C_I) \cap \{0,1\}^n\}.$$

For any $I \in \mathcal{J}_n$, the element $I' \in \mathcal{I}_{n+1}$ defined by $C_{I'} := C_I \vee 0$ belongs to $\mathcal{J}_{n+1}$.

Proof. The inductive construction of $\mathcal{J}_n$ follows from the inductive construction of $\mathcal{C}_n$, and the fact that $\{C_I; I \in \mathcal{J}_n\}$ is exactly the subset of $\mathcal{C}_n$ constituted with the correlation matrices of size n with all entries in $\{0,1\}$. For the last statement, we notice that the null vector of $\mathbb{R}^n$ has entries in $\{0,1\}$ and belongs to the centred ellipsoid $\mathcal{E}(C_I)$. Thus, the element I' of $\mathcal{I}_{n+1}$ defined by $C_{I'} = C_I \vee 0$ belongs to $\mathcal{J}_{n+1}$. Notice that we can then move the zeroes (of the last column and line of $C_{I'}$) elsewhere by a permutation matrix. □

In particular, if $I \in \mathcal{J}_n$ is such that C_I is not singular, then C_I has full range and thus, for any $J \subset \{(i, n+1); 1 \leqslant i \leqslant n\}$, we have $I \cup J \in \mathcal{J}_{n+1}$. Notice that the matrix C_I in the counter example above is singular!

Theorem 8.5 (Just put zeroes where they must be!). *Let $I \in \mathcal{J}_n$ be some fixed admissible element of $\mathcal{I}_n$. Then, we have $p_{\mathcal{K}_n^+(I)} = p_{\mathcal{K}_n(I)}$ on $\mathcal{S}_n^+$. In other words, provided that $A \in \mathcal{S}_n^+$, the orthogonal projection of A onto the closed convex cone $\mathcal{K}_n^+(I)$ is equal to its orthogonal projection onto the sub-vector space $\mathcal{K}_n(I)$ of $\mathcal{S}_n$. In other words, the orthogonal projection of an element A of $\mathcal{S}_n^+$ onto $\mathcal{K}_n^+(I)$ is obtained by putting zeroes at the entries given by I.*

Proof. Since $I \in \mathcal{J}_n$, we have $p_{\mathcal{K}_n(I)}(A) \in \mathcal{S}_n^+$ for any $A \in \mathcal{S}_n^+$. Then the Lemma below with $(\mathcal{H}, \mathcal{A}, \mathcal{B}) = (\mathcal{S}_n, \mathcal{K}_n(I), \mathcal{S}_n^+)$ gives that $p_{\mathcal{K}_n^+(I)} = p_{\mathcal{K}_n(I)}$ on $\mathcal{S}_n^+$. □

Lemma 8.6 (Nested projections). *Let $\mathcal{A}$ and $\mathcal{B}$ be two non-empty closed convex subsets of a Hilbert space $\mathcal{H}$. Let $x \in \mathcal{H}$ such that $p_{\mathcal{A}}(x) \in \mathcal{B}$, then $p_{\mathcal{A}\cap\mathcal{B}}(x) = p_{\mathcal{A}}(x)$. In other words, if $p_{\mathcal{A}}(x) \in \mathcal{B}$ for any x in some subset $\mathcal{G}$ of $\mathcal{H}$, then $p_{\mathcal{A}} = p_{\mathcal{A}\cap\mathcal{B}}$ on $\mathcal{G}$.*

Proof. We have $p_{\mathcal{A}}(x) = \arg\inf_{y\in\mathcal{A}} \|x - y\|$ and $p_{\mathcal{A}\cap\mathcal{B}}(x) = \arg\inf_{y\in\mathcal{A}\cap\mathcal{B}} \|x - y\|$ for any $x \in \mathcal{H}$. First, we have $\arg\inf_{y\in\mathcal{A}} \|x - y\| \leqslant \arg\inf_{y\in\mathcal{A}\cap\mathcal{B}} \|x - y\|$. If $p_{\mathcal{A}}(x) \in \mathcal{B}$, then $p_{\mathcal{A}}(x) \in \mathcal{A} \cap \mathcal{B}$ and thus $\inf_{y\in\mathcal{A}\cap\mathcal{B}} \|x - y\| = \inf_{y\in\mathcal{A}} \|x - y\| = \|x - p_{\mathcal{A}}(x)\|$. Therefore, $p_{\mathcal{A}\cap\mathcal{B}}(x) = p_{\mathcal{A}}(x)$. □

A Some basic facts

Theorem A.1 (Spectral Theorem). *An element A of $\mathcal{M}_n$ belongs to $\mathcal{S}_n$ if and only if there exists O in $\mathcal{O}_n$ such that $OAO^\top \in \mathcal{D}_n$. In particular, every element A of $\mathcal{S}_n$ is diagonalisable and has a real spectrum. The column of O are the normalised eigenvectors of A, and the diagonal entries of $OAO^\top$ are the eigenvalues of A.*

For any $A \in \mathcal{S}_n^+$ with spectrum $\{\lambda_1, \ldots, \lambda_n\} \in [0, +\infty)^n$, the spectral Theorem provides in particular a sort of square root matrix $B := O^\top \mathrm{Diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})O$, in the sense that $BB = A$. The following Theorem provides another type of matrix square root, in relation with the quadratic form associated to the element of $\mathcal{S}_n^+$. It is related to the Gram-Schmidt orthonormalisation algorithm.

Theorem A.2 (Cholesky[2] factorisation). *An element A of $\mathcal{M}_n$ belongs to $\mathcal{S}_n^+$ if and only if there exists a lower triangular element L of $\mathcal{M}_n$ such that $A = LL^\top$. Moreover, $\det(A) = \prod_{i=1}^n L_{i,i}^2$, and A belongs to $\mathcal{S}_n^{+*}$ if and only if L has non-null diagonal entries. Additionally, for any $A \in \mathcal{S}_n^+$, there is a unique such L with non-negative diagonal entries, which can be computed recursively from the entries of A via the formulas*

$$L_{i,j} = L_{j,j}\left(A_{i,j} - \sum_{k=1}^{j-1} L_{i,k}L_{j,k}\right) \quad \text{and} \quad L_{i,i} = \sqrt{A_{i,i} - \sum_{k=1}^{i-1} L_{i,k}^2}.$$

Theorem A.3 (Probability folklore). *There is a one to one correspondence between $\mathcal{S}_n^+$ and the set of covariance matrices of random vectors of dimension n. Moreover, the random vectors can be assumed Gaussian, and $\mathcal{S}_n^{+*}$ corresponds exactly to the covariance matrices of Gaussian vectors with absolute continuous law with respect to the Lebesgue measure.*

Proof. Follows from the usage of the Fourier transform of probability measures on $\mathbb{R}^n$. The reduction to Gaussian random vectors makes use of the spectral Theorem or of the Choleski factorisation. □

Recall that if X is a centred Gaussian random vector of $\mathbb{R}^n$ with covariance matrix K, then for any M in $\mathcal{M}_n$, the random vector MX is a centred Gaussian random vector with covariance matrix $MKM^\top$. The particular case $K = I_n$ is very useful. This mechanism is reversible, as expressed by the Choleski factorisation. The Cholesky factorisation allows to compute efficiently the determinant and the inverse of A, and to simulate multivariate Gaussian random variables with prescribed covariance matrix A from simulations of the univariate standard Gauss distribution. The Cholesky factorisation in $\mathcal{S}_n$ is a particular case of the LU factorisation[3] in $\mathcal{M}_n$.

Theorem A.4 (Schur). *The convex cone $\mathcal{S}_n^+$ is stable by the Hadamard product.*

Proof. We extract the proof from [HJ94, Chap. 5] (see also [HJ90, Section 7.5]). Let A and B be two elements of $\mathcal{S}_n^+$. Let X and Y be two centred random vectors of $\mathbb{R}^n$ with respective covariance matrices A and B. We can construct them explicitly from A and B as Gaussian random vectors by using the spectral Theorem of the Choleski factorisation. Now, the vector $X \circ Y$ (i.e its i^{th} entry is X_iY_i) is a centred random vector with covariance matrix $A \circ B$. Thus, $A \circ B \in \mathcal{S}_n^+$. Beware that $X \circ Y$ is not a Gaussian vector, even if X and Y are. □

B The det and log det functions on GL_n

For any $A \in \mathcal{M}_n$, we denote by cofA the cofactors matrix[4] of A. Recall that $A(\mathrm{cof}A)^\top = \det(A)I_n$, in such a way that $\det(A)A^{-1} = (\mathrm{cof}A)^\top$ when $A \in \mathrm{GL}_n$.

[2] André-Louis Cholesky, 1875-1918.

[3] Do not confuse the LU factorisation with the polar decomposition or the QR decomposition in $\mathcal{M}_n$.

[4] The cofactors matrix $\mathrm{cof}A$ of A is defined for any i and j in $\{1, \ldots, n\}$ by $(\mathrm{cof}A)_{i,j} := (-1)^{i+j}\det(\tilde{A}_{i,j})$, where $\tilde{A}_{i,j}$ is the $(n-1) \times (n-1)$ sub-matrix of A obtained by removing the i^{th} line and the j^{th} column. By definition, $\det(\tilde{A}_{i,j})$ is a *minor* of A.

Theorem B.1 (Determinant). *The function $A \in \mathrm{GL}_n \mapsto \det(A)$ is $\mathcal{C}^\infty$. Moreover, for any $A \in \mathrm{GL}_n$ and $B \in \mathcal{M}_n$, we have the so called Jacobi formulas*

$$(D\det)(A)(B) = B \cdot \mathrm{cof} A \quad \text{and} \quad \nabla \det(A) = \mathrm{cof} A = \det(A) A^{-T},$$

where $A^{-\top} := (A^\top)^{-1} = (A^{-1})^\top$.

Proof. The determinant of an $n \times n$ matrix is an alternate n-multi-linear form. This gives the smoothness. The first formula follows from $\det(A) = \sum_{i,j=1}^n A_{i,i}(\mathrm{cofA})_{i,j}$. □

Theorem B.2 (Inverse). *The map $\mathrm{inv} : A \in \mathrm{GL}_n \mapsto A^{-1} \in \mathrm{GL}_n$ is differentiable and its differential is given by $D(\mathrm{inv})(A)(B) = -A^{-1}BA^{-1}$ for any $A \in \mathrm{GL}_n$ and $B \in \mathcal{M}_n$.*

Proof. We have[5] $(A+B)^{-1} = (I_n + A^{-1}B)^{-1}A^{-1} = \sum_{i=0}^\infty (-1)^i (A^{-1}B)^i A^{-1}$ for any $A \in \mathrm{GL}_n$ and small enough $B \in \mathcal{M}_n$ (in norm). Thus, for any $A \in \mathrm{GL}_n$ and any $B \in \mathcal{M}_n$, $(A+B)^{-1} = A^{-1} - A^{-1}BA^{-1} + o(\|B\|)$. □

Theorem B.3 ($\log\det$). *The function $A \in \mathrm{GL}_n \mapsto \log\det(A)$ is C^∞ on GL_n. Moreover, for any $A \in \mathrm{GL}_n$ and any $U, V \in \mathcal{M}_n$,*

$$D(\log\det)(A)(U) = U \cdot A^{-\top} \quad \text{and} \quad \nabla(\log\det)(A) = A^{-\top},$$

and

$$D^2(\log\det)(A)(U,V) = -U \cdot (A^{-T}V^\top A^{-T}) = -\mathrm{Tr}(A^{-1}UA^{-1}V).$$

Proof. The smoothness and the first derivative come by composition (chain rule). The second derivative comes from the expression of the derivative of $A \mapsto A^{-1}$. □

The open set GL_n is not convex, but is locally convex. We can ask if the function $\log\det$ is concave. Unfortunately, for any $A \in \mathrm{GL}_n$ and any $B \in \mathcal{M}_n$, we have $D^2(\log\det)(A)(U,U) = -\mathrm{Tr}((UA^{-1})^2)$ which is not necessarily in $(-\infty, 0]$. Of course, if UA^{-1} is diagonalisable, then $\mathrm{Tr}((UA^{-1})^2) \geqslant 0$ as the sum of squares of the eigenvalues of UA^{-1}.

C Spectral functions on $\mathcal{S}_n$

By the spectral Theorem, an element A of $\mathcal{S}_n$ is in $\mathcal{S}_n^+$ (resp. $\mathcal{S}_n^{+*}$) if and only if $x^\top Ax \geqslant 0$ (resp. > 0) for any non null vector x of $\mathbb{R}^n$. Namely, we just need to write $x^\top Ax = x^\top ODOx = (Ox)^\top D(Ox)$ where D is diagonal and $O \in \mathcal{O}_n$. As a consequence, for any A and B in $\mathcal{S}_+$ (resp. $\mathcal{S}_n^{+*}$), the sum $A+B$ belongs to $\mathcal{S}_n^+$ (resp. $\mathcal{S}_n^{+*}$). Namely, $x^\top Ax + x^\top Bx = (Ox)^\top D(Ox) + (O'x)^\top D'(O'x)$. The sets $\mathcal{S}_n^+$ and $\mathcal{S}_n^{+*}$ are thus stable by vector additions and positive dilations, in other words, they are convex cones. The cone $\mathcal{S}_n^+$ is anti-polar, cf. [Mal01].

In the sequel, we denote by $\lambda : \mathcal{S}_n \to \mathbb{R}^n$ the non-increasing spectrum functional. This function is given by $\lambda(A) = (\lambda_1(A), \dots, \lambda_n(A))$ where $\mathrm{spec(A)} = \{\lambda_1(A), \dots, \lambda_n(A)\}$ with $\lambda_1(A) \geqslant \cdots \geqslant \lambda_n(A)$. The following Theorem gives the orthogonal projection on $\mathcal{S}_n^+$ in $\mathcal{S}_n$.

[5] Sometimes referred as the "von Neumann formula". If $M \in \mathcal{M}_n$ is such that $\|M\|_\# < 1$ for some matrix norm $\|\cdot\|_\#$, then $I_n - M \in \mathrm{GL}_n$ and $(I_n - M)^{-1} = \sum_{i=0}^\infty M^k$. See [HJ90, Corollary 5.6.16].

Theorem C.1 (Eckart-Young-Higham). *The orthogonal projection $p_{\mathcal{S}_n^+}(A)$ of an element A of $\mathcal{S}_n$ onto $\mathcal{S}_n^+$ is obtained by replacing the negative eigenvalues of A with zero. In other words, $p_{\mathcal{S}_n^+}(A) = O\mathrm{Diag}(\max(0,\lambda_1),\ldots,\max(0,\lambda_n))O^\top$ where $A = O\mathrm{Diag}(\lambda_1,\ldots,\lambda_n)O^\top$ with $O \in \mathcal{O}_n$.*

Let $f : \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}$ be a smooth function defined on an open convex subset $\mathcal{D}$ of $\mathbb{R}^n$. Let $S(\mathcal{D})$ be an open subset of $\mathcal{S}_n$ such that $\lambda(S(\mathcal{D})) \subset \mathcal{D}$. The composed function $f \circ \lambda : S(\mathcal{D}) \subset \mathcal{S}_n \to \mathbb{R}$ induces a function $S(\mathcal{D}) \to \mathcal{S}_n$ which maps $A = O\mathrm{Diag}(\lambda_1,\ldots,\lambda_n)O^\top$ onto $O\mathrm{Diag}(f(\lambda_1),\ldots,f(\lambda_n))O^\top$. Beware that the cone $\mathcal{S}_n$ is not an open subset of $\mathcal{M}_n$. The derivative and the gradient are thus defined in the Hilbert space $\mathcal{S}_n$.

Theorem C.2 (Lewis, 1996). *If $f : \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}$ is symmetric convex, then the matrix function $f \circ \lambda : S(\mathcal{D}) \subset \mathcal{S}_n \to \mathbb{R}$ is also convex. Let $A \in \mathcal{S}_n$ such that $\lambda(A) \in S(\mathcal{D})$. Then f is differentiable at the point $\lambda(A)$ if and only if $f \circ \lambda$ is differentiable at the point A. In that case, we have the following formula*

$$\nabla(f \circ \lambda)(A) = O\mathrm{Diag}(\nabla f(\lambda(A)))O^\top$$

where $O \in \mathcal{O}_n$ is such that $A = O\mathrm{Diag}(\lambda(A))O^\top$.

A striking example is given by the function $F : A \in \mathcal{S}_n^{+*} \to \log\det(A) \in \mathbb{R}$. Since $F(A) = \sum_{i=1}^n \lambda_i(A)$, one can write F as $F = f \circ \lambda$ where $f : \mathbb{R}^n \to \mathbb{R}$ is the symmetric concave function defined by $f(x) = \log(x_1) + \cdots + \log(x_n)$. Here $\mathcal{D} = (0,+\infty)^n$ and $S(\mathcal{D}) = \mathcal{S}_n^{+*}$. We get in particular from Lewis Theorem that $\nabla F(A) = A^{-1} = A^{-\top}$. We thus recovered on $\mathcal{S}_n^{+*}$ the result already established on GL_n (Jacobi formula).

Theorem C.3. *The function $A \in \mathcal{S}_n^{+*} \mapsto \log\det(A)$ is strictly concave.*

The strictness comes from the fact that the gradient never vanishes. In other words, $\det(tA+(1-t)B) \geqslant \det(A)^t \det(B)^{1-t}$, for any $t \in [0,1]$ and any $A, B \in \mathcal{S}_n^+$, with equality if and only if $\lambda \in \{0,1\}$ or if $A = B$.

D Surface representation of $\mathcal{S}_n^+$

The spectral Theorem has the following interesting immediate Corollary.

Corollary D.1. *Every element A of $\mathcal{S}_n^+$ of rank k can be written as a sum of k elements of $\mathcal{S}_n^+$ of rank 1. In other words, there exists k column vectors $v_1,\ldots,v_k$ of $\mathbb{R}^n$ such that $A = v_1 v_1^\top + \cdots + v_k {v_k}^\top$.*

The vectors v_i appearing above are just the column of O, scaled by the square root of the associated eigenvalue of A. More generally, the spectral Theorem yields the following structural result.

Corollary D.2. *If $\mathbb{B}_n$ is the set of orthonormal bases of $\mathbb{R}^n$, then*

$$\mathcal{S}_n = \left\{ \sum_{i=1}^n \lambda_i v_i v_i^\top ; \{v_1,\ldots,v_n\} \in \mathbb{B}_n, \lambda \in \mathbb{R}^n \right\}.$$

Moreover, the components of λ and the vectors $\{v_1,\ldots,v_n\}$ are respectively the spectrum and the eigenvectors of the symmetric matrix $\sum_{i=1}^n \lambda_i v_i v_i^\top$. The elements of rank k of $\mathcal{S}_n$ correspond to $\sum_{i=1}^n \mathrm{I}_{\{\lambda_i \neq 0\}} = k$. Additionally, the elements of $\mathcal{S}_n^+$ correspond to $\lambda \in [0,+\infty)^n$ whereas the elements of $\mathcal{S}_n^{+*}$ correspond to $\lambda \in (0,+\infty)^n$. For any element A of $\mathcal{S}_n$ (resp. $\mathcal{S}_n^+$) of rank r and any $n_1 + \cdots + n_l = r$, there exists elements $A_1,\ldots,A_l$ of $\mathcal{S}_n$ (resp. $\mathcal{S}_n^+$) of respective ranks $n_1,\ldots,n_l$ such that $A = A_1 + \cdots + A_l$. Finally, the interior of the closed convex cone $\mathcal{S}_n^+$ in the Hilbert space $\mathcal{S}_n$ is given by $\mathcal{S}_n^{+*} = \mathcal{S}_n^+ \cap \mathrm{GL}_n$. In other *words, the surface of $\mathcal{S}_n^+$ is the set of symmetric matrices with non-negative spectrum and rank strictly less that n.*

Proof. For any column vector $v \in \mathbb{R}^n$, the symmetric matrix $vv^\top$ has columns $v_1 v, \ldots, v_n v$. Consequently, when $v \neq 0$, the matrix $vv^\top$ is of rank 1 with spectrum $\{0, v^\top v\}$, and $vv^\top v = (v^\top v)v$. Every element $A \in \mathcal{S}_n$ of rank 1 can be written in the form $vv^\top$ where v is an eigenvector associated to the unique non null eigenvalue of A.

The statement $A = \sum_{i=1}^n \lambda_i v_i v_i^\top$ with $\{v_1, \ldots, v_n\} \in \mathbb{B}_n$ and $\lambda \in \mathbb{R}^n$ is equivalent to $O := v_1 \otimes \cdots \otimes v_n \in \mathcal{O}_n$ and $A = O\mathrm{Diag}(\lambda_1, \ldots, \lambda_n)O^\top$. In particular, $\{\lambda_1, \ldots, \lambda_n\}$ is the spectrum of A. Additionally, if $i_1, \ldots, i_r$ are r different indexes in $\{1, \ldots, n\}$, then $\lambda_{i_1} v_{i_1} v_{i_1}^\top + \cdots + \lambda_{i_r} v_{i_r} v_{i_r}^\top$ is of rank $\mathrm{I}_{\lambda_{i_1} \neq 0} + \cdots + \mathrm{I}_{\lambda_{i_r} \neq 0} \leqslant r$, since $v_{i_1}, \ldots, v_{i_r}$ are orthonormal. Namely, let $O \in \mathcal{O}_n$ such that $Ov_i = e_i$ for any $i \in \{1, \ldots, n\}$. Then $O(\lambda_{i_1} v_{i_1} v_{i_1}^\top + \cdots + \lambda_{i_r} v_{i_r} v_{i_r}^\top)O^\top = \lambda_{i_1} e_{i_1} e_{i_1}^\top + \cdots + \lambda_{i_r} e_{i_r} e_{i_r}^\top = \mathrm{Diag}(\lambda_{i_1}, \ldots, \lambda_{i_r}, 0, \ldots, 0)$. □

These decompositions are unique up to some permutations of the v_i and λ_i. Notice that $\mathbb{B}_n$ is roughly $\mathcal{O}_n/\Sigma_n$ where the symmetric group Σ_n acts by permutation of the columns. The matrices $v_i v_i^\top$ are of rank 1 and thus belong to $\mathcal{S}_n^+ \backslash \mathcal{S}_n^{+*}$, the surface of $\mathcal{S}_n^+$. Actually, the surface of $\mathcal{S}_n^+$ corresponds exactly to the constraint $\mathrm{I}_{\lambda_1 > 0} + \cdots + \mathrm{I}_{\lambda_n > 0} < n$, which are the elements of $\mathcal{S}_n$ with rank strictly less that n. We thus have a parametrisation of $\mathcal{S}_n^+$ and its interior $\mathcal{S}_n^{+*}$ in $\mathcal{S}_n$ by the surface of $\mathcal{S}_n^+$. Notice that one can randomly generates elements of $\mathbb{B}_n$ by recursively using the uniform measures on the unit spheres of $\mathbb{R}^n, \mathbb{R}^{n-1}, \ldots, \mathbb{R}^2$, which produces recursively the vectors $v_1, \ldots, v_{n-1}$. One can move in $\mathcal{S}_n^+$ by moving slightly the scalars λ_i and by rotating the vectors v_i by an element of $\mathcal{O}_n$ close to I_n.

Notice that when $A \in \mathcal{S}_n^+$, we have $A = D(\sum_{i=1}^n v_i v_i^\top)D$ where $D \in \mathcal{D}_n^+$ is given by $D := \mathrm{Diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$. We will see in the sequel a similar decomposition where the diagonal matrix is constituted by the diagonal of A rather that the spectrum of A.

It follows any extremal element of the convex closed cone $\mathcal{S}_n^+$ is of rank 1.

E Projections onto convex closed cones

Assume that $I \notin \mathcal{J}_n$. We already gave counter examples of elements M of $\mathcal{S}_n^+$ such that $M \circ C_I \notin \mathcal{S}_n^+$. A numerical simulation shows that under $K \in \mathcal{K}_n^{+*}(I)$, it is false in general that $\mathbb{X} \circ C_I \in \mathcal{S}_n^+$ for any sample size N. A possible drawback to the fact that $\mathbb{X} \circ C_I$ does not belong to $\mathcal{K}_n^+(I)$ for small N is to consider the biased estimator $p_{\mathcal{K}_n^+(I)}(\mathbb{X})$. This corresponds to a matrix least square problem. The reader may find various algorithms for such matrix least squares problems in [MS05, Mal05, Mal01] and in [BX05] and references therein. Notice that $p_{\mathcal{K}_n^+(I)}(\mathbb{X}) + p_{(\mathcal{K}_n^+(I))^\circ}(\mathbb{X}) = \mathbb{X}$. In average, $\mathbb{E}(p_{\mathcal{K}_n^+(I)}(\mathbb{X})) + \mathbb{E}(p_{\mathcal{K}_n^+(I)^\circ}(\mathbb{X})) = K$. Therefore, the bias of the estimator $p_{\mathcal{K}_n^+(I)}(\mathbb{X})$ of K is $-\mathbb{E}(p_{\mathcal{K}_n^+(I)^\circ}(\mathbb{X}))$. Recall the following classical Theorem due to Moreau, see [HUL01, Chap. III]. Let x, x_1 and x_2 be three points in the Hilbert space $\mathbb{R}^n$, and let $\mathcal{K}$ be a convex closed cone and $\mathcal{K}^\circ$ its polar cone. Then the following two statements are equivalent.

- $x = x_1 + x_2$ with $x_1 \perp x_2$ and $(x_1, x_1) \in \mathcal{K} \times \mathcal{K}^\circ$;
- $x_1 = p_{\mathcal{K}}(x)$ and $x_2 = p_{\mathcal{K}^\circ}(x)$.

A consequence of the Moreau-Yoshida regularisation – see [Mal01, Section 2.2] and references therein – permits to show that the function $x \in \mathbb{R}^n \mapsto d_{\mathcal{K}}(x) := \frac{1}{2}\|x - p_{\mathcal{K}}(x)\|^2 = \frac{1}{2}\|p_{\mathcal{K}^\circ}(x)\|^2$ is convex with gradient given by $\nabla d_{\mathcal{K}}(x) = p_{\mathcal{K}^\circ}(x)$. Furthermore, this gradient function is Lipschitz continuous with a Lipschitz constant equal to 1. A regularity

Theorem by Rademacher shows then that $p_{\mathcal{K}}$ is twice differentiable almost everywhere in $\mathbb{R}^n$. In [Mal01, MS05, Mal05], the reader may find nice results on a dual algorithm to solve orthogonal projections in the Hilbert space $\mathcal{S}_n$ onto closed convex sub-cones of $\mathcal{S}_n^+$ (e.g. intersections of $\mathcal{S}_n^+$ with an affine sub-vector spaces of $\mathcal{S}_n$). It corresponds to least-squares in the Hilbert space $\mathcal{S}_n$ with semi-definite positiveness constraint and linear constraint. It includes the case of the projection on $\mathcal{K}_n^+(I)$. The addition of linear inequalities constraints is addressed for instance in [BX05]. We hope that for the special case of $\mathcal{K}_n^+(I)$, things are simpler and explicit.

References

[BV04] Stephen Boyd and Lieven Vandenberghe, *Convex optimization*, Cambridge University Press, Cambridge, 2004. MR MR2061575 (2005d:90002)

[BX05] Stephen Boyd and Lin Xiao, *Least squares covariance matrix adjustment*, SIAM J. Matrix. Anal. Appl. **27** (2005), no. 2, 532–548.

[CDR06] Saujay Chaudhuri, Mathias Drton, and Thomas S. Richardson, *Estimation of a Covariance Matrix with Zeros*, ArXiv.org:math.ST/0508268, 2006.

[Dem72] A. P. Dempster, *Covariance selection*, Biometrics **28** (1972), 157–175.

[DRV05] J. Dahl, V. Roychowdhury, and L. Vandenberghe, *Maximum likelihood estimation of gaussian graphical models: Numerical implementation and topology selection*, preprint, available on `http://www.ee.ucla.edu/~vandenbe/covsel.html`, september 2005.

[HJ90] Roger A. Horn and Charles R. Johnson, *Matrix analysis*, Cambridge University Press, Cambridge, 1990, Corrected reprint of the 1985 original. MR MR1084815 (91i:15001)

[HJ94] ———, *Topics in matrix analysis*, Cambridge University Press, Cambridge, 1994, Corrected reprint of the 1991 original. MR MR1288752 (95c:15001)

[HUL01] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal, *Fundamentals of convex analysis*, Grundlehren Text Editions, Springer-Verlag, Berlin, 2001, Abridged version of *Convex analysis and minimization algorithms. I* [Springer, Berlin, 1993; MR1261420 (95m:90001)] and *II* [ibid.; MR1295240 (95m:90002)]. MR MR1865628 (2002i:90002)

[Lau96] Steffen L. Lauritzen, *Graphical models*, Oxford Statistical Science Series, vol. 17, The Clarendon Press Oxford University Press, New York, 1996, Oxford Science Publications. MR MR1419991 (98g:62001)

[Lew96a] A. S. Lewis, *Convex analysis on the Hermitian matrices*, SIAM J. Optim. **6** (1996), no. 1, 164–177. MR MR1377729 (97g:49045)

[Lew96b] ———, *Derivatives of spectral functions*, Math. Oper. Res. **21** (1996), no. 3, 576–588. MR MR1403306 (97g:90141)

[LS01] Adrian S. Lewis and Hristo S. Sendov, *Twice differentiable spectral functions*, SIAM J. Matrix Anal. Appl. **23** (2001), no. 2, 368–386 (electronic). MR MR1871318 (2002j:15008)

[Mal01] Jérome Malick, *An efficient dual algorithm to solve conic least-square problems*, INRIA technical report num. 4212 available at `http://www.inrialpes.fr/bipop/people/malick/`, june 2001.

[Mal05] ———, *A dual approach to semidefinite least-squares problems*, SIAM J. Matrix. Anal. Appl. **26** (2005), 272–284.

[MS05] Jérome Malick and Hristo S. Sendov, *Clarke Generalized Jacobian of the Projection onto the Cone of Positive Semidefinite Matrices*, INRIA technical report num. 5490 available at `http://www.inrialpes.fr/bipop/people/malick/`, february 2005.

[Ser02] Denis Serre, *Matrices*, Graduate Texts in Mathematics, vol. 216, Springer-Verlag, New York, 2002, Theory and applications, Translated from the 2001 French original. MR MR1923507 (2003h:15001)

[VBW96] Lieven Vandenbergue, Stephen Boyd, and Shao-Po Wu, *Determinant maximization with linear matrix inequality constraints*, electronic documents available on Internet, april 1996.

Djalil Chafaï and Didier Concordet
Address: UMR 181 INRA/ENVT Physiopathologie et Toxicologie Expérimentales,
École Nationale Vétérinaire de Toulouse,
23 Chemin des Capelles, F-31076, Toulouse Cedex 3, France.
E-mail: mailto:d.chafai@envt.fr.nospam
Address: UMR 5583 CNRS/UPS Laboratoire de Statistique et Probabilités,
Institut de Mathématiques de Toulouse, Université Paul Sabatier,
118 route de Narbonne, F-31062, Toulouse, Cedex 4, France.
E-mail: mailto:chafai@math.ups-tlse.fr.nospam
Web: http://www.lsp.ups-tlse.fr/Chafai/