MDPI

Article

A Log-Det Heuristics for Covariance Matrix Estimation: The Analytic Setup

Enrico Bernardi [†] and Matteo Farnè *,[†]

Dipartimento di Scienze Statistiche, Università di Bologna, 40126 Bologna, Italy; enrico.bernardi@unibo.it
* Correspondence: matteo.farne@unibo.it
† These authors contributed equally to this work.

Abstract: This paper studies a new nonconvex optimization problem aimed at recovering high-dimensional covariance matrices with a low rank plus sparse structure. The objective is composed of a smooth nonconvex loss and a nonsmooth composite penalty. A number of structural analytic properties of the new heuristics are presented and proven, thus providing the necessary framework for further investigating the statistical applications. In particular, the first and the second derivative of the smooth loss are obtained, its local convexity range is derived, and the Lipschitzianity of its gradient is shown. This opens the path to solve the described problem via a proximal gradient algorithm.

Keywords: nonconvex optimization; local convexity; nuclear norm; covariance matrix; high dimension

1. Introduction

The estimation of large covariance or precision matrices is a relevant challenge nowadays, due to the increasing availability of datasets composed of a large number of variables p compared to the sample size n in many fields. The urgency of this topic is testified by several recent books [1–3], and comprehensive reviews [4–6]. In this paper, we assume for the $p \times p$ covariance matrix Σ^* a low rank plus sparse decomposition, that is

$$\Sigma^* = L^* + S^* = BB' + S^*, \tag{1}$$

where $L^* = BB' = U_L \Lambda_L U_L'$, U_L is a $p \times r$ matrix such that $U_L' U_L = I_r$, Λ_L is a $r \times r$ diagonal matrix, and S^* is element-wise sparse, i.e. it contains only $s \ll \frac{p(p-1)}{2}$ off-diagonal non-zero elements. Since [7] proposed their approximate factor model, structure (1) has become the reference model for many high-dimensional covariance matrix estimators, like POET [8].

The recovery of structure (1) is a statistical problem of primary relevance. Ref. [7] proposed to consistently estimate L^* (as $p \to \infty$) by means of principal component analysis (PCA, see [9]), assuming that the eigenvalues of L^* diverge with the dimension p while the eigenvalues of S^* remain bounded. [8] proposes to estimate L^* by the top r principal components of the sample covariance matrix Σ_n (as $p \to \infty$) and to estimate S^* by thresholding their orthogonal complement. In [10], L^* and S^* are recovered by nuclear norm plus l_1 penalization, that is by computing

$$\left(\widehat{L}, \widehat{S} \right) = \arg \min_{L \succeq 0, S \succ 0} \mathfrak{L}(L, S) + \mathfrak{P}(L, S), \tag{2}$$

where $\mathfrak{L}(L, S)$ is a smooth loss function, $\mathfrak{P}(L, S)$ is a nonsmooth penalty function, $L \succeq 0$ denotes positive semidefiniteness for L and $S \succ 0$ denotes positive definiteness for S. In particular, denoting by $\lambda_i(M)$, $i = 1, \ldots, p$, the eigenvalues of a $p \times p$ matrix M sorted in descending order, $\mathfrak{L}(L, S) = \frac{1}{2}\|\Sigma_n - (L + S)\|_F^2$, $\mathfrak{P}(L, S) = \psi\|L\|_* + \rho\|S\|_1$, where

$\|\mathbf{L}\|_* = \sum_{i=1}^p \lambda_i(\mathbf{L})$ (the nuclear norm of $\mathbf{L}$), $\|\mathbf{S}\|_1 = \sum_{i=1}^p \sum_{j=i}^p |\mathbf{S}_{ij}|$ (the l_1 norm of $\mathbf{S}$), and ψ and ρ are non-negative threshold parameters.

The nuclear norm was first proposed in [11] as an alternative to PCA. Ref. [12] furnishes a proof that $\psi\|\mathbf{L}\|_* + \rho\|\mathbf{S}\|_1$ is the tightest convex relaxation of the original nonconvex penalty $\psi \mathrm{rk}(\mathbf{L}) + \rho\|\mathbf{S}\|_0$. Ref. [13] proves that the l_1 norm minimization provides the sparsest solution to most large underdetermined linear systems, while [14] proves that the nuclear norm minimization provides guaranteed rank minimization under a set of linear equality constraints. Ref. [15] shows that l_1 norm minimization selects the best linear model in a wide range of situations. The nuclear norm has instead been used to solve large matrix completion problems, like in [16–18], and [19]. Nuclear norm plus l_1 norm minimization was first exploited in [20] to provide a robust version of PCA under grossly corrupted or missing data.

The pair of estimators (2) derived in [10] is named ALCE (ALgebraic Covariance Estimator). Although ALCE has many desirable statistical properties, there is room to further improve it by replacing $\frac{1}{2}\|\Sigma_n - (\mathbf{L} + \mathbf{S})\|_F^2$ by a different loss. The Frobenius loss optimizes in fact the entry by entry performance of $\widehat{\Sigma}$, while a loss able to explicitly control the spectrum estimation quality may be desirable. In this paper, we consider the loss

$$\mathfrak{L}(\mathbf{L}, \mathbf{S}) = \frac{1}{2} \log \det(\mathbf{I}_p + \boldsymbol{\Delta}_n \boldsymbol{\Delta}_n'), \tag{3}$$

where $\boldsymbol{\Delta}_n = \Sigma - \Sigma_n$, and $\Sigma = \mathbf{L} + \mathbf{S}$. Heuristics (3) is controlled by the individual singular values of $\boldsymbol{\Delta}_n$, because

$$\log \det(\mathbf{I}_p + \boldsymbol{\Delta}_n \boldsymbol{\Delta}_n') = \log \prod_{i=1}^p (\lambda_i(\mathbf{I}_p + \boldsymbol{\Delta}_n \boldsymbol{\Delta}_n')) \le \sum_{i=1}^p (1 + \lambda_i(\boldsymbol{\Delta}_n)^2) = p + \sum_{i=1}^p \lambda_i(\boldsymbol{\Delta}_n)^2, \tag{4}$$

and, therefore, it is better suited for the estimation of the underlying spectrum.

To the best of our knowledge, the mathematical properties of (3) have not been extensively studied. Analogously to the univariate context ($p = 1$), (3) is not a convex function. According to ongoing works like [21], nonconvex problems may be approached either by searching for approximate solutions instead of global solutions, or by exploiting the geometric structure of the objective function. Furthermore, in this case, the idea of restricting the analysis to the convexity region of the objective, a region that may be indefinitely extended (see the concept of Extendable Local Strong Convexity in [22]), is the key to apply, for instance, existing proximal gradient algorithms for convex functions (see [23]). For this reason, in this paper we calculate the first and second derivatives of (3), we derive its range of local convexity, and the Lipschitzianity of (3) and of its gradient. This opens the path to using the usual proximal gradient algorithms (see [23]) to solve problem (2) with $\mathfrak{L}(\mathbf{L}, \mathbf{S})$ as in (3).

2. Analytic Setup

We consider the objective function

$$\phi(\mathbf{L}, \mathbf{S}) = \mathfrak{L}(\mathbf{L}, \mathbf{S}) + \mathfrak{P}(\mathbf{L}, \mathbf{S}), \tag{5}$$

where $\mathfrak{L}(\mathbf{L}, \mathbf{S}) = \frac{1}{2} \log \det(\mathbf{I}_p + \boldsymbol{\Delta}_n \boldsymbol{\Delta}_n')$ is the smooth part of $\phi(\mathbf{L}, \mathbf{S})$ and $\mathfrak{P}(\mathbf{L}, \mathbf{S}) = \psi\|\mathbf{L}\|_* + \rho\|\mathbf{S}\|_1$ is the non-smooth (but convex) part of $\phi(\mathbf{L}, \mathbf{S})$. First, we calculate the derivative of the smooth component $\mathfrak{L}(\mathbf{L}, \mathbf{S})$ *wrt* $\mathbf{L}$ and $\mathbf{S}$, which is

$$\frac{\delta \frac{1}{2} \log \det(\mathbf{I}_p + \boldsymbol{\Delta}_n \boldsymbol{\Delta}_n')}{\delta \mathbf{L}} = \frac{\delta \frac{1}{2} \log \det(\mathbf{I}_p + \boldsymbol{\Delta}_n \boldsymbol{\Delta}_n')}{\delta \mathbf{S}} = (\mathbf{I}_p + \boldsymbol{\Delta}_n \boldsymbol{\Delta}_n')^{-1} \boldsymbol{\Delta}_n. \tag{6}$$

Proof. Let us consider two generic $p \times p$ matrices $\mathbf{L}$ and $\mathbf{S}$, their sum $\Sigma = \mathbf{L} + \mathbf{S}$, and the matrix $\boldsymbol{\Delta}_n = \Sigma - \Sigma_n$. Let us define the matrix function $\varphi(\Sigma) = \mathbf{I}_p + \boldsymbol{\Delta}_n \boldsymbol{\Delta}_n'$ and the function $\phi(\Sigma) = \log \det \varphi(\Sigma)$. We denote by $\mathbf{e}_i$ the i-th canonical basis vector, by $\mathbf{e}_i^l = \delta_{il}$

its l-th element, and by σ_{ij} the ij entry of $\boldsymbol{\Sigma}$. Then, following [24], for each $i, j = 1, \ldots, p$, we can write

$$\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_{ij}} = m_{ij} = \mathbf{e}_i' \mathbf{e}_j$$

$$\frac{\partial}{\partial \sigma_{ij}} \log \det \varphi(\boldsymbol{\Sigma}) = \mathrm{Tr}\left(\varphi^{-1}(\boldsymbol{\Sigma}) \frac{\partial \varphi}{\partial \sigma_{ij}} \right)$$

$$\frac{\partial \varphi}{\partial \sigma_{ij}} = \boldsymbol{\Delta}_n \frac{\partial \boldsymbol{\Delta}_n'}{\partial \sigma_{ij}} + \frac{\partial \boldsymbol{\Delta}_n}{\partial \sigma_{ij}} \boldsymbol{\Delta}_n'$$

$$\frac{\partial \boldsymbol{\Delta}_n}{\partial \sigma_{ij}} = m_{ij}.$$

Therefore,

$$\begin{aligned}
\frac{\partial}{\partial \sigma_{ij}} \log \det \varphi(\boldsymbol{\Sigma}) &= \mathrm{Tr}\left(\varphi^{-1}(\boldsymbol{\Sigma}) \left[\boldsymbol{\Delta}_n m_{ji} + m_{ij} \boldsymbol{\Delta}_n' \right] \right) \\
&= \mathrm{Tr}\left(\varphi^{-1}(\boldsymbol{\Sigma}) \boldsymbol{\Delta}_n m_{ji} \right) + \mathrm{Tr}\left(\varphi^{-1}(\boldsymbol{\Sigma}) m_{ij} \boldsymbol{\Delta}_n' \right).
\end{aligned}$$

Since for $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ conformable matrices

$$\mathrm{Tr}(\mathbf{ABC}) = \mathrm{Tr}((\mathbf{ABC})') = \mathrm{Tr}(\mathbf{C}'\mathbf{B}'\mathbf{A}') = \mathrm{Tr}(\mathbf{A}'\mathbf{B}'\mathbf{C}'),$$

we get

$$\begin{aligned}
\frac{\partial}{\partial \sigma_{ij}} \log \det \varphi(\boldsymbol{\Sigma}) &= 2\mathrm{Tr}\left(\varphi^{-1}(\boldsymbol{\Sigma}) \boldsymbol{\Delta}_n m_{ji} \right) \\
&= 2 \sum_\nu \left(\varphi^{-1}(\boldsymbol{\Sigma}) \boldsymbol{\Delta}_n m_{ji} \right)_{\nu\nu} \\
&= 2 \sum_{\nu, \rho} \left(\varphi^{-1}(\boldsymbol{\Sigma}) \right)_{\nu\rho} \left(\boldsymbol{\Delta}_n m_{ji} \right)_{\rho\nu} \\
&= 2 \sum_{\nu, \rho, \sigma} \left(\varphi^{-1}(\boldsymbol{\Sigma}) \right)_{\nu\rho} \boldsymbol{\Delta}_{n, \rho\sigma} \left(m_{ji} \right)_{\sigma\nu}.
\end{aligned} \tag{7}$$

Finally, considering that

$$\left(m_{ji} \right)_{\sigma\nu} = \left(\mathbf{e}_j \otimes \mathbf{e}_i \right)_{\sigma\nu} = \delta_{j\sigma} \delta_{i\nu},$$

we get

$$\begin{aligned}
\frac{\partial}{\partial \sigma_{ij}} \log \det \varphi(\boldsymbol{\Sigma}) &= 2 \sum_{\nu, \rho, \sigma} \left(\varphi^{-1}(\boldsymbol{\Sigma}) \right)_{\nu\rho} \boldsymbol{\Delta}_{n, \rho\sigma} \delta_{j\sigma} \delta_{i\nu} \\
&= 2 \sum_\rho \left(\varphi^{-1}(\boldsymbol{\Sigma}) \right)_{i\rho} \boldsymbol{\Delta}_{n, \rho j} \\
&= 2 \left(\varphi^{-1}(\boldsymbol{\Sigma}) \boldsymbol{\Delta}_n \right)_{ij}.
\end{aligned}$$

To sum up,

$$\frac{\partial}{\partial \boldsymbol{\Sigma}} \left[\frac{1}{2} \log \det \varphi(\boldsymbol{\Sigma}) \right] = \frac{\partial}{\partial \mathbf{L}} \left[\frac{1}{2} \log \det \varphi(\boldsymbol{\Sigma}) \right] = \frac{\partial}{\partial \mathbf{S}} \left[\frac{1}{2} \log \det \varphi(\boldsymbol{\Sigma}) \right] = \varphi^{-1}(\boldsymbol{\Sigma}) \boldsymbol{\Delta}_n. \tag{8}$$

$\square$

In the following, we explicit the second derivative of $\mathfrak{L}(\mathbf{L}, \mathbf{S}) = \frac{1}{2} \log \det \varphi(\boldsymbol{\Sigma})$, with $\varphi(\boldsymbol{\Sigma}) = (\mathbf{I}_p + \boldsymbol{\Delta}_n \boldsymbol{\Delta}_n')$ and $\boldsymbol{\Sigma} = \mathbf{L} + \mathbf{S}$:

$$\frac{\partial^2}{\partial\sigma_{ij}\partial\sigma_{hk}}\frac{1}{2}\log\det\varphi(\boldsymbol{\Sigma}) = \left(\frac{1}{2}\text{Hess}\log\det\varphi(\boldsymbol{\Sigma})\right)_{ijhk}$$
$$= \delta_{jk}\left(\varphi^{-1}(\boldsymbol{\Sigma})\right)_{ih} - \sum_{\mu,\sigma}\left(\varphi^{-1}(\boldsymbol{\Sigma})\right)_{h\mu}\boldsymbol{\Delta}_{\mu j}\left(\varphi^{-1}(\boldsymbol{\Sigma})\right)_{i\sigma}\boldsymbol{\Delta}_{\sigma k} \quad (9)$$
$$- \left(\varphi^{-1}(\boldsymbol{\Sigma})\right)_{ih}\sum_{\mu,\lambda}\left(\varphi^{-1}(\boldsymbol{\Sigma})\right)_{\lambda\mu}\boldsymbol{\Delta}_{\mu j}\boldsymbol{\Delta}_{\lambda k}.$$

More, if $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_n$, we get

$$\left(\frac{1}{2}\text{Hess}\log\det\varphi(\boldsymbol{\Sigma})\right)_{ijhk} = \delta_{jk}\otimes\delta_{ih} = \left(\mathbf{I}_p\otimes\mathbf{I}_p\right)_{ijhk}, \quad (10)$$

that is,

$$\frac{1}{2}\text{Hess}\log\det\varphi(\boldsymbol{\Sigma}) = \mathbf{I}_p\otimes\mathbf{I}_p. \quad (11)$$

Proof. From [24], we write

$$\frac{\partial^2}{\partial\sigma_{ij}\partial\sigma_{hk}}\frac{1}{2}\log\det\varphi(\boldsymbol{\Sigma}) = \frac{1}{2}\left(\text{Tr}\left(\varphi^{-1}(\boldsymbol{\Sigma})\frac{\partial^2\varphi(\boldsymbol{\Sigma})}{\partial\sigma_{ij}\partial\sigma_{hk}}\right) - \text{Tr}\left(\varphi^{-1}(\boldsymbol{\Sigma})\frac{\partial\varphi(\boldsymbol{\Sigma})}{\partial\sigma_{ij}}\varphi^{-1}(\boldsymbol{\Sigma})\frac{\partial\varphi(\boldsymbol{\Sigma})}{\partial\sigma_{hk}}\right)\right),$$

and we recall that

$$\frac{\partial^2\varphi(\boldsymbol{\Sigma})}{\partial\sigma_{ij}\partial\sigma_{hk}} = \sigma_{hk}\sigma_{ji} + \sigma_{ij}\sigma_{kh}.$$

Then, we can calculate

$$\text{Tr}\left(\varphi^{-1}(\boldsymbol{\Sigma})\frac{\partial^2\varphi(\boldsymbol{\Sigma})}{\partial\sigma_{ij}\partial\sigma_{hk}}\right) = \text{Tr}\left(\varphi^{-1}(\boldsymbol{\Sigma})\sigma_{hk}\sigma_{ji} + \varphi^{-1}(\boldsymbol{\Sigma})\sigma_{ij}\sigma_{kh}\right)$$
$$= \text{Tr}\left(\varphi^{-1}(\boldsymbol{\Sigma})\sigma_{hk}\sigma_{ji}\right) + \text{Tr}\left(\varphi^{-1}(\boldsymbol{\Sigma})\sigma_{ij}\sigma_{kh}\right)$$
$$= 2\text{Tr}\left(\varphi^{-1}(\boldsymbol{\Sigma})\sigma_{hk}\sigma_{ji}\right)$$
$$= 2\sum_{\nu}\left(\varphi^{-1}(\boldsymbol{\Sigma})\sigma_{hk}\sigma_{ji}\right)$$
$$= 2\delta_{jk}(\varphi^{-1}(\boldsymbol{\Sigma}))_{ih}. \quad (12)$$

The second summand

$$\text{Tr}\left(\varphi^{-1}(\boldsymbol{\Sigma})\frac{\partial\varphi(\boldsymbol{\Sigma})}{\partial\sigma_{ij}}\varphi^{-1}(\boldsymbol{\Sigma})\frac{\partial\varphi(\boldsymbol{\Sigma})}{\partial\sigma_{hk}}\right)$$

can be derived from (12) as

$$2\text{Tr}\left(\varphi^{-1}(\boldsymbol{\Sigma})\boldsymbol{\Delta}\sigma_{ji}\varphi^{-1}(\boldsymbol{\Sigma})\boldsymbol{\Delta}\sigma_{kh}\right) + 2\text{Tr}\left(\varphi^{-1}(\boldsymbol{\Sigma})\boldsymbol{\Delta}\sigma_{ji}\varphi^{-1}(\boldsymbol{\Sigma})\sigma_{hk}\boldsymbol{\Delta}\right).$$

Equation (9) is consequently proved. $\square$

3. Local Convexity

The aim of this section is to determine the range of convexity for $\mathfrak{L}(\mathbf{L},\mathbf{S}) = \frac{1}{2}\log\det(\mathbf{I}_p + \boldsymbol{\Delta}_n\boldsymbol{\Delta}'_n)$, $\boldsymbol{\Delta}_n = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}_n$, *wrt* to the semidefinite positive matrix $\boldsymbol{\Delta}_n\boldsymbol{\Delta}'_n$. In the univariate context, the function $\frac{1}{2}\ln\det(1 + x^2)$ is convex if and only if $|x| < \frac{1}{\sqrt{2}}$. In the multivariate context, it is therefore reasonable to suppose that a similar condition on $\boldsymbol{\Delta}_n\boldsymbol{\Delta}'_n$ ensures local convexity. A proof can be given by showing the positive definiteness of the Hessian of $\mathfrak{L}(\mathbf{L},\mathbf{S})$ for some range of $\|\boldsymbol{\Delta}_n\boldsymbol{\Delta}'_n\|$. In other words, we need to show that there exists a positive δ such that, whenever $\|\boldsymbol{\Delta}_n\boldsymbol{\Delta}'_n\| < \delta$, the function $\frac{1}{2}\log\det(\mathbf{I}_p + \boldsymbol{\Delta}_n\boldsymbol{\Delta}'_n)$ is convex.

Lemma 1. *Given* $0 < \mu \leq \frac{1}{3p}$, *we have that the function*

$$\log \det(\mathbf{I}_p + \mathbf{A}\mathbf{A}^*) \tag{13}$$

is convex on the set $\mathcal{C}_\mu = \{\mathbf{A} | \mathbf{A} \text{ is a real } p \times p \text{ matrix}, \|\mathbf{A}\|_2 \leq \mu\}$ *where* $\|\mathbf{A}\|_2$ *denotes the spectral norm of* $\mathbf{A}$.

Proof. We proceed using the criterion of convexity estimating the second derivative with respect to t of

$$\phi(t) = \log \det(\mathbf{I}_p + (t\mathbf{A} + (1-t)\mathbf{B})(t\mathbf{A} + (1-t)\mathbf{B})^*). \tag{14}$$

Let us recall that

$$\frac{d}{dt} \log \det \mathbf{G}(t) = \text{Tr}(\mathbf{G}(t)^{-1}\mathbf{G}'(t)), \tag{15}$$

where $\mathbf{G}(t)$ is a differentiable square matrix-valued function and (15) holds for those values of t for which $\mathbf{G}(t)$ is invertible.

Furthermore, we have as well

$$\frac{d}{dt}\text{Tr}(\mathbf{A}(t)) = \text{Tr}(\frac{d}{dt}\mathbf{A}(t)), \quad \frac{d}{dt}\mathbf{A}^{-1} = -\mathbf{A}^{-1}\left(\frac{d}{dt}\mathbf{A}\right)\mathbf{A}^{-1}, \tag{16}$$

for any differentiable square matrix-valued function $\mathbf{A}(t)$. (See [25] or [26] e.g., how to prove these identities).

Calling $\mathbf{G}(t) = \mathbf{I}_p + (t\mathbf{A} + (1-t)\mathbf{B})(t\mathbf{A} + (1-t)\mathbf{B})^*$, we see that $\mathbf{G}'(t) = 2t\mathbf{\Lambda}\mathbf{\Lambda}^* + \mathbf{R}$ and $\mathbf{G}''(t) = 2\mathbf{\Lambda}\mathbf{\Lambda}^*$ where

$$\mathbf{\Lambda} = \mathbf{A} - \mathbf{B}, \mathbf{R} = \mathbf{B}\mathbf{\Lambda}^* + \mathbf{\Lambda}\mathbf{B}^*$$

and $\mathbf{G}(t) = \frac{1}{3p\mu}\mathbf{I}_p + t^2\mathbf{\Lambda}\mathbf{\Lambda}^* + t\mathbf{R} + \mathbf{B}\mathbf{B}^*$.

Thus, applying (15) and (16) we get

$$\phi'(t) = \text{Tr}(\mathbf{G}(t)^{-1}\mathbf{G}'(t))$$

and

$$\phi''(t) = \text{Tr}\left((\mathbf{G}(t)^{-1}\mathbf{G}'(t))'\right) = \text{Tr}\left(\mathbf{G}^{-1}\mathbf{G}'' - \mathbf{G}^{-1}\mathbf{G}'\mathbf{G}^{-1}\mathbf{G}'\right) \tag{17}$$

Convexity will follow once we have proven that (17) is non-negative for every $t \in [0,1]$ and every $\mathbf{A}$ and $\mathbf{B}$ in $\mathcal{C}_\mu$.

Due to the circularity of the trace function we also have

$$\phi''(t) = \text{Tr}\left(\mathbf{G}^{-1/2}\mathbf{G}''\mathbf{G}^{-1/2} - \mathbf{G}^{-1/2}\mathbf{G}'\mathbf{G}^{-1}\mathbf{G}'\mathbf{G}^{-1/2}\right), \tag{18}$$

that is

$$\phi''(t) = \text{Tr}\left(\mathbf{G}^{-1/2}\mathbf{G}''\mathbf{G}^{-1/2}\right) - \text{Tr}\left(\mathbf{G}^{-1/2}\mathbf{G}'\mathbf{G}^{-1/2}\,\mathbf{G}^{-1/2}\mathbf{G}'\mathbf{G}^{-1/2}\right). \tag{19}$$

This can be written as

$$\phi''(t) = 2\text{Tr}\left(\mathbf{G}^{-1/2}\mathbf{\Lambda}\mathbf{\Lambda}^*\mathbf{G}^{-1/2}\right) - \text{Tr}\left(\mathbf{G}^{-1/2}(2t\mathbf{\Lambda}\mathbf{\Lambda}^* + \mathbf{R})\mathbf{G}^{-1/2}\,\mathbf{G}^{-1/2}(2t\mathbf{\Lambda}\mathbf{\Lambda}^* + \mathbf{R})\mathbf{G}^{-1/2}\right). \tag{20}$$

We recall that $\mathbf{G}(t)$ is self-adjoint so that denoting by $\mathbf{H}$ the matrix $\mathbf{G}^{-1/2}\mathbf{\Lambda}$ and $\mathbf{K}$ the matrix $\mathbf{G}^{-1/2}\mathbf{G}'\mathbf{G}^{-1/2}$ we get that (19) can be written as

$$\phi''(t) = 2\text{Tr}(\mathbf{H}\mathbf{H}^*) - \text{Tr}(\mathbf{K}\mathbf{K}^*). \tag{21}$$

We also recall that $\text{Tr}(\mathbf{AB}^*)$ induces a scalar product to which the trace norm is attached:

$$||\mathbf{A}||_{\text{tr}} = \sqrt{\text{Tr}(\mathbf{AA}^*)} = \sum_i \sigma_i(\mathbf{A}),$$

where $\sigma_i(\mathbf{A})$ are the singular values of $\mathbf{A}$. In particular we have $||\mathbf{A}||_2 \leq ||\mathbf{A}||_{\text{tr}} \leq p||\mathbf{A}||_2$ for every $\mathbf{A}$. Now from (21) convexity can be checked as

$$\phi''(t) = 2||\mathbf{H}||_{\text{tr}} - ||\mathbf{K}||_{\text{tr}} \geq 0. \tag{22}$$

Let us consider

$$||\mathbf{K}||_2 = ||\mathbf{G}^{-1/2}\mathbf{G}'\mathbf{G}^{-1/2}||_2 = ||\mathbf{G}^{-1/2}(2t\mathbf{\Lambda\Lambda}^* + \mathbf{R})\mathbf{G}^{-1/2}||_2.$$

We have

$$||\mathbf{K}||_2 \leq 2t||\mathbf{G}^{-1/2}\mathbf{\Lambda\Lambda}^*\mathbf{G}^{-1/2}||_2 + ||\mathbf{G}^{-1/2}\mathbf{R}\mathbf{G}^{-1/2}||_2.$$

Notice that the spectral norm $|| \bullet ||_2$ is self-adjoint, that is $||\mathbf{M}^*||_2 = ||\mathbf{M}||_2$ for every $\mathbf{M}$ (see e.g., [27]). Then

$$||\mathbf{K}||_2 \leq 2t||\mathbf{HH}^*||_2 + ||\mathbf{G}^{-1/2}(\mathbf{B\Lambda}^* + \mathbf{\Lambda B}^*)\mathbf{G}^{-1/2}||_2,$$

that is

$$||\mathbf{K}||_2 \leq 2||\mathbf{H}||_2||\mathbf{G}^{-1/2}\mathbf{\Lambda}||_2 + 2||\mathbf{G}^{-1/2}\mathbf{\Lambda}||_2||\mathbf{G}^{-1/2}\mathbf{B}||_2.$$

Thus,

$$||\mathbf{K}||_2 \leq 2||\mathbf{H}||_2\Big(||\mathbf{G}^{-1/2}\mathbf{\Lambda}||_2 + ||\mathbf{G}^{-1/2}\mathbf{B}||_2\Big). \tag{23}$$

Assume now that $\mathbf{A}, \mathbf{B} \in \mathcal{C}_\mu$: $||\mathbf{A}||_2 \leq \mu$, $||\mathbf{B}||_2 \leq \mu$. We deduce that $||\mathbf{\Lambda}||_2 \leq 2\mu$ and due to the structure of $\mathbf{G}(t) = \mathbf{I}_p + \mathbf{Q}(t)\mathbf{Q}(t)^*$ we also have

$$||\mathbf{G}^{-1/2}\mathbf{\Lambda}||_2 \leq 2\mu \, , ||\mathbf{G}^{-1/2}\mathbf{B}||_2 \leq \mu. \tag{24}$$

Finally, we have

$$||\mathbf{K}||_2 \leq 6\mu||\mathbf{H}||_2.$$

Going back to (22) we have

$$\phi''(t) = 2||\mathbf{H}||_{\text{tr}} - ||\mathbf{K}||_{\text{tr}} \geq 2||\mathbf{H}||_2 - p||\mathbf{K}||_2 \geq 2(1 - 3p\mu)||\mathbf{H}||_2 \geq 0, \tag{25}$$

since $0 < \mu \leq \frac{1}{3p}$. □

By means of a simple change of variable, the following result can be proven.

Lemma 2. *For any $\delta > 0$ the function*

$$\log \det\Big(\delta^{-2}\mathbf{I}_p + \mathbf{AA}^*\Big) \tag{26}$$

is convex on the closed ball $\mathcal{C}_\delta = \{\mathbf{A}|\mathbf{A} \text{ is a real } p \times p \text{ matrix} , ||\mathbf{A}||_2 \leq \frac{1}{3\delta p}\}$.

In conclusion, even though the function $\log \det(\mathbf{I}_p + \mathbf{A})$ is always concave, Lemma 2 shows that the function $\log \det(\delta^{-2}\mathbf{I}_p + \mathbf{AA}^*)$ can be made locally convex into any ball centered in 0, just choosing a suitable δ.

4. Lipschitz-Continuity

In this section, we prove the Lipschitzianity of the smooth function $\mathfrak{L}(\mathbf{L}, \mathbf{S}) = \frac{1}{2}\ln \det(\mathbf{I}_p + \mathbf{\Delta}_n\mathbf{\Delta}'_n)$, and of its gradient function, $\frac{\delta\mathfrak{L}(\mathbf{L},\mathbf{S})}{\delta\mathbf{L}} = \frac{\delta\mathfrak{L}(\mathbf{L},\mathbf{S})}{\delta\mathbf{S}} = (\mathbf{I}_p + \mathbf{\Delta}_n\mathbf{\Delta}'_n)^{-1}\mathbf{\Delta}_n$ (see (6)).

Lemma 3. *The function $\mathfrak{L}(\mathbf{L}, \mathbf{S}) = \frac{1}{2} \ln \det(\mathbf{I}_p + \mathbf{\Delta}_n \mathbf{\Delta}_n')$ is Lipschitz continuous in Euclidean norm with Lipschitz constant equal to 1:*

$$|\log \det \varphi(\mathbf{\Sigma}_1) - \log \det \varphi(\mathbf{\Sigma}_2)| \le \|\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2\|_2. \tag{27}$$

Proof. Let us recall that $\mathbf{L}$ and $\mathbf{S}$ are two generic $p \times p$ matrices, $\mathbf{\Sigma} = \mathbf{L} + \mathbf{S}$ is their sum, and $\mathbf{\Delta}_n = \mathbf{\Sigma} - \mathbf{\Sigma}_n$. We reconsider the matrix function $\varphi(\mathbf{\Sigma}) = \mathbf{I}_p + \mathbf{\Delta}_n \mathbf{\Delta}_n'$ and the function $\phi(\mathbf{\Sigma}) = \log \det \varphi(\mathbf{\Sigma})$. We recall from (6) that

$$\frac{\partial}{\partial \mathbf{\Sigma}} \left[\frac{1}{2} \log \det \varphi(\mathbf{\Sigma}) \right] = \varphi^{-1}(\mathbf{\Sigma}) \mathbf{\Delta}_n.$$

Given two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^p$, let us define the Euclidean inner product $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}' \mathbf{v}$. We consider

$$\langle \varphi(\mathbf{\Sigma}) \mathbf{v}, \mathbf{v} \rangle = \langle (\mathbf{I}_p + \mathbf{\Delta}_n \mathbf{\Delta}_n') \mathbf{v}, \mathbf{v} \rangle = |\mathbf{v}|^2 + |\mathbf{\Delta}_n' \mathbf{v}|^2,$$

where $|\mathbf{v}|$ is the Euclidean norm of $\mathbf{v} \in \mathbb{R}^p$. Then we have

$$|\mathbf{v}|^2 + |\mathbf{\Delta}_n' \mathbf{v}|^2 \le \frac{1}{2} \left(\frac{1}{\delta^2} |\varphi(\mathbf{\Sigma}) \mathbf{v}|^2 + \delta^2 |\mathbf{v}|^2 \right),$$

via Cauchy-Schwarz, for any $\delta \in \mathbb{R}$. Now choose $\delta = \sqrt{2}$, then we have

$$4 |\mathbf{\Delta}_n' \mathbf{v}|^2 \le |\varphi(\mathbf{\Sigma}) \mathbf{v}|^2$$

for every $\mathbf{v} \in \mathbb{R}^p$. Noticing that $\varphi(\mathbf{\Sigma})$ is invertible and plugging in the previous inequality $\mathbf{v} = \varphi(\mathbf{\Sigma})^{-1} \mathbf{w}$, $\mathbf{w} \in \mathbb{R}^p$, we obtain

$$\max_{\mathbf{w} \ne \mathbf{0}} \frac{|\mathbf{\Delta}_n' \varphi(\mathbf{\Sigma})^{-1} \mathbf{w}|}{|\mathbf{w}|^2} \le \frac{1}{2}.$$

Now recall that (see [25] p. 312) that the spectral norm of a matrix $\mathbf{A}$, $\|\mathbf{A}\|_2$, can be computed also via the equality

$$\|\mathbf{A}\|_2 = \max_{\mathbf{x} \ne \mathbf{0}} \frac{|\mathbf{A}\mathbf{x}|}{|\mathbf{x}|},$$

and that the spectral norm is self-adjoint (again see [25] p. 309), that is $\|\mathbf{A}'\|_2 = \|\mathbf{A}\|_2$. Summing up, we have proved that

$$\|\varphi(\mathbf{\Sigma})^{-1} \mathbf{\Delta}_n\|_2 \le \frac{1}{2}. \tag{28}$$

This means that the gradient of $\log \det \varphi(\mathbf{\Sigma})$ is uniformly bounded and since $\mathbf{\Sigma} \mapsto \log \det \varphi(\mathbf{\Sigma})$ is a smooth function we have that the Lipschitz condition is satisfied with Lipschitz constant equal to 1:

$$|\log \det \varphi(\mathbf{\Sigma}_1) - \log \det \varphi(\mathbf{\Sigma}_2)| \le \|\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2\|_2. \tag{29}$$

$\square$

We have proven that the function

$$\frac{\partial^2}{\partial \sigma_{ij} \partial \sigma_{hk}} \frac{1}{2} \log \det \varphi(\mathbf{\Sigma}) = \delta_{jk} \left(\varphi^{-1}(\mathbf{\Sigma})\right)_{ih} - \sum_{\mu,\sigma} \left(\varphi^{-1}(\mathbf{\Sigma})\right)_{h\mu} \mathbf{\Delta}_{n,\mu j} \left(\varphi^{-1}(\mathbf{\Sigma})\right)_{i\sigma} \mathbf{\Delta}_{n,\sigma k}$$

$$- \left(\varphi^{-1}(\mathbf{\Sigma})\right)_{ih} \sum_{\mu,\lambda} \left(\varphi^{-1}(\mathbf{\Sigma})\right)_{\lambda\mu} \mathbf{\Delta}_{n,\mu j} \mathbf{\Delta}_{n,\lambda k}$$

$$= \delta_{jk} \left(\varphi^{-1}(\mathbf{\Sigma})\right)_{ih} - (\varphi^{-1}(\mathbf{\Sigma})\mathbf{\Delta}_n)_{hj} (\varphi^{-1}(\mathbf{\Sigma})\mathbf{\Delta}_n)_{ik}$$

$$- \left(\varphi^{-1}(\mathbf{\Sigma})\right)_{ih} \langle \varphi^{-1}(\mathbf{\Sigma})\mathbf{\Delta}_{n,j}, \mathbf{\Delta}_{n,k} \rangle \tag{30}$$

is Lipschitz continuous.

Now, we prove that the function $\frac{\delta \mathfrak{L}(\mathbf{L},\mathbf{S})}{\delta \mathbf{L}} = \frac{\delta \mathfrak{L}(\mathbf{L},\mathbf{S})}{\delta \mathbf{S}} = (\mathbf{I}_p + \mathbf{\Delta}_n \mathbf{\Delta}_n')^{-1} \mathbf{\Delta}_n$ is Lipschitz continuous.

Lemma 4. *The function* $\frac{\delta \mathfrak{L}(\mathbf{L},\mathbf{S})}{\delta \mathbf{L}} = \frac{\delta \mathfrak{L}(\mathbf{L},\mathbf{S})}{\delta \mathbf{S}} = (\mathbf{I}_p + \mathbf{\Delta}_n \mathbf{\Delta}_n')^{-1} \mathbf{\Delta}_n$ *is Lipschitz continuous with Lipschitz constant equal to* $\frac{5}{4}$:

$$\|F(\mathbf{\Delta}_n + \epsilon \mathbf{H}) - F(\mathbf{\Delta}_n)\|_2 \le \frac{5}{4} \epsilon \|\mathbf{H}\|_2 + O(\epsilon^2), \tag{31}$$

with $F(\mathbf{\Delta}_n) = \varphi^{-1}(\mathbf{\Sigma})\mathbf{\Delta}_n = (\mathbf{I}_p + \mathbf{\Delta}_n \mathbf{\Delta}_n')^{-1} \mathbf{\Delta}_n$ *and fix* $\epsilon > 0$, *for any* $\epsilon > 0$.

Proof. Let us call $F(\mathbf{\Delta}_n) = \varphi^{-1}(\mathbf{\Sigma})\mathbf{\Delta}_n = (\mathbf{I}_p + \mathbf{\Delta}_n \mathbf{\Delta}_n')^{-1} \mathbf{\Delta}_n$ and fix $\epsilon > 0$.

Let us compute

$$F(\mathbf{\Delta}_n + \epsilon \mathbf{H}) - F(\mathbf{\Delta}_n) = (\mathbf{I}_p + (\mathbf{\Delta}_n + \epsilon \mathbf{H})(\mathbf{\Delta}_n + \epsilon \mathbf{H})')^{-1}(\mathbf{\Delta}_n + \epsilon \mathbf{H}) - (\mathbf{I}_p + \mathbf{\Delta}_n \mathbf{\Delta}_n')^{-1} \mathbf{\Delta}_n.$$

We have

$$(\mathbf{I}_p + (\mathbf{\Delta}_n + \epsilon \mathbf{H})(\mathbf{\Delta}_n + \epsilon \mathbf{H})')^{-1}(\mathbf{\Delta}_n + \epsilon \mathbf{H}) = (\mathbf{I}_p + \mathbf{\Delta}_n \mathbf{\Delta}_n' + \mathbf{\Lambda})^{-1}(\mathbf{\Delta}_n + \epsilon \mathbf{H}),$$

with $\mathbf{\Lambda} = \epsilon \mathbf{\Lambda}_0 + \epsilon^2 \mathbf{\Lambda}_1$ and

$$\mathbf{\Lambda}_0 = \mathbf{H}\mathbf{\Delta}_n' + \mathbf{\Delta}_n \mathbf{H}', \quad \mathbf{\Lambda}_1 = \mathbf{H}\mathbf{H}'.$$

Calling $\mathbf{\Psi} = \mathbf{I}_p + \mathbf{\Delta}_n \mathbf{\Delta}_n'$ we have

$$(\mathbf{\Psi} + \mathbf{\Lambda})^{-1} = (\mathbf{\Psi}(\mathbf{I}_p + \mathbf{\Psi}^{-1}\mathbf{\Lambda}))^{-1} = (\mathbf{I}_p + \mathbf{\Psi}^{-1}\mathbf{\Lambda})^{-1}\mathbf{\Psi}^{-1},$$

so that we have

$$F(\mathbf{\Delta}_n + \epsilon \mathbf{H}) - F(\mathbf{\Delta}_n) = (\mathbf{I}_p + \mathbf{\Psi}^{-1}\mathbf{\Lambda})^{-1}\mathbf{\Psi}^{-1}(\mathbf{\Delta}_n + \epsilon \mathbf{H}) - \mathbf{\Psi}^{-1}\mathbf{\Delta}_n.$$

Recalling that

$$(\mathbf{I}_p + \mathbf{G})^{-1} = \mathbf{I}_p - \mathbf{G} + (\mathbf{I}_p + \mathbf{G})^{-1}\mathbf{G}^2,$$

whenever $\mathbf{I}_p + \mathbf{G}$ is invertible, we have

$$F(\mathbf{\Delta}_n + \epsilon \mathbf{H}) - F(\mathbf{\Delta}_n) = (\mathbf{I}_p - \mathbf{\Psi}^{-1}\mathbf{\Lambda} + (\mathbf{I}_p + \mathbf{\Psi}^{-1}\mathbf{\Lambda})^{-1}(\mathbf{\Psi}^{-1}\mathbf{\Lambda})^2)\mathbf{\Psi}^{-1}(\mathbf{\Delta}_n + \epsilon \mathbf{H}) - \mathbf{\Psi}^{-1}\mathbf{\Delta}_n.$$

We develop in the powers of ϵ:

$$\begin{aligned} F(\mathbf{\Delta}_n + \epsilon \mathbf{H}) - F(\mathbf{\Delta}_n) &= (\mathbf{I}_p - \mathbf{\Psi}^{-1}(\epsilon \mathbf{\Lambda}_0 + \epsilon^2 \mathbf{\Lambda}_1) + (\mathbf{I}_p + \mathbf{\Psi}^{-1}(\epsilon \mathbf{\Lambda}_0 + \epsilon^2 \mathbf{\Lambda}_1)^{-1}) \\ &\quad (\mathbf{\Psi}^{-1}(\epsilon \mathbf{\Lambda}_0 + \epsilon^2 \mathbf{\Lambda}_1))^2)\mathbf{\Psi}^{-1}(\mathbf{\Delta}_n + \epsilon \mathbf{H}) - \mathbf{\Psi}^{-1}\mathbf{\Delta}_n \end{aligned}$$

A tedious but simple computation yields to

$$F(\mathbf{\Delta}_n + \epsilon\mathbf{H}) - F(\mathbf{\Delta}_n) = \epsilon\mathbf{T}_1 + O(\epsilon^2),$$

with

$$\mathbf{T}_1 = \mathbf{\Psi}^{-1}\mathbf{H} - \mathbf{\Psi}^{-1}\mathbf{\Lambda}_0\mathbf{\Psi}^{-1}\mathbf{\Delta}_n,$$

that is

$$\mathbf{T}_1 = \mathbf{\Psi}^{-1}\mathbf{H}(\mathbf{I}_p - \mathbf{\Delta}_n'\mathbf{\Psi}^{-1}\mathbf{\Delta}_n) - \mathbf{\Psi}^{-1}\mathbf{\Delta}_n\mathbf{H}'\mathbf{\Psi}^{-1}\mathbf{\Delta}_n.$$

The previous computations for the Lipschizianity of $\log\det(\mathbf{I}_p + \mathbf{\Delta}_n\mathbf{\Delta}_n')$ gave us (see (28)) that

$$\|\mathbf{\Psi}^{-1}\mathbf{\Delta}_n\|_2 \le \frac{1}{2}.$$

It is also easy to check that

$$\|\mathbf{\Psi}^{-1}\|_2 \le 1,$$

and that

$$\|\mathbf{I}_p - \mathbf{\Delta}_n'\mathbf{\Psi}^{-1}\mathbf{\Delta}_n\|_2 \le 1.$$

Putting all together we get

$$\|F(\mathbf{\Delta}_n + \epsilon\mathbf{H}) - F(\mathbf{\Delta}_n)\|_2 \le \frac{5}{4}\epsilon\|\mathbf{H}\|_2 + O(\epsilon^2), \tag{32}$$

such that we have proven that the directional derivative of the gradient is bounded in every direction by $5/4$, i.e., the gradient is Lipschitz as a function from $\mathcal{M}(p)$ to $\mathcal{M}(p)$, the vector space of $p \times p$ real matrices. $\square$

5. Discussion

In this paper, we have proved that the loss $\frac{1}{2}\log\det(\mathbf{I}_p + \mathbf{\Delta}_n\mathbf{\Delta}_n')$ has good analytic properties for the purpose of optimization, provided that the matrix $\mathbf{\Delta}_n$ fulfills certain conditions. As a consequence, by [23,28] and the supplement of [10], it follows that our analytic setup can provide a numerical solution to the problem

$$\min_{\mathbf{L} \succeq 0, \mathbf{S} \succ 0} \frac{1}{2}\log\det(\mathbf{I}_p + \mathbf{\Delta}_n\mathbf{\Delta}_n') + \psi\|\mathbf{L}\|_* + \rho\|\mathbf{S}\|_1, \tag{33}$$

by using proximal gradient algorithms. The local convexity of $\log\det(\mathbf{I}_p + \mathbf{\Delta}_n\mathbf{\Delta}_n')$ is the key to apply first-order methods to solve (33). Following [23,28] and the supplement of [10], we derive the following solution Algorithm 1.

Such algorithm may be applied in many fields, like economics, finance, biology, genetics, health, climatology, social science, among others. In future research, we plan to properly develop the selection of threshold parameters, to study how local convexity may cope with the random nature of the sample error matrix $\mathbf{\Delta}_n$, and to establish the consistency of the solution pair of (33).

Algorithm 1 Pseudocode to solve problem (33) given any input covariance matrix $\mathbf{\Sigma}_n$.

1. Set $(\mathbf{L}_0, \mathbf{S}_0) = \frac{1}{2}(\text{diag}(\mathbf{\Sigma}_n), \text{diag}(\mathbf{\Sigma}_n))$, $\eta_0 = 1$.
2. Initialize $\mathbf{Y}_0 = \mathbf{L}_0$ and $\mathbf{Z}_0 = \mathbf{S}_0$. Set $t = 1$.
3. For $t \geq 1$, repeat:

   (i) calculate calculate $\mathbf{\Delta}_{t,n} = \mathbf{Y}_{t-1} + \mathbf{Z}_{t-1} - \mathbf{\Sigma}_n$;

   (ii) compute $\frac{\partial \frac{1}{2} \log \det(\mathbf{I}_p + \mathbf{\Delta}_{t,n}\mathbf{\Delta}'_{t,n})}{\partial \mathbf{Y}_{t-1}} = \frac{\partial \frac{1}{2} \log \det(\mathbf{I}_p + \mathbf{\Delta}_{t,n}\mathbf{\Delta}'_{t,n})}{\partial \mathbf{Z}_{t-1}} = (\mathbf{I}_p + \mathbf{\Delta}_{t,n}\mathbf{\Delta}'_{t,n})^{-1}\mathbf{\Delta}_{t,n}$.

   (iii) apply the singular value thresholding (SVT, [29]) operator T_ψ to
   $\mathbf{E}_{Y,t} = \mathbf{Y}_{t-1} - \frac{1}{\ell}(\mathbf{I}_p + \mathbf{\Delta}_{t,n}\mathbf{\Delta}'_{t,n})^{-1}\mathbf{\Delta}_{t,n}$, with $\ell = \frac{10}{4}$,
   and set $\mathbf{L}_t = T_\psi(\mathbf{E}_{Y,t}) = \widehat{\mathbf{U}}\widehat{\mathbf{D}}_\psi\widehat{\mathbf{U}}^\top$.

   (v) apply the soft-thresholding operator [30] T_ρ to
   $\mathbf{E}_{Z,t} = \mathbf{Z}_{t-1} - \frac{1}{\ell}(\mathbf{I}_p + \mathbf{\Delta}_{t,n}\mathbf{\Delta}'_{t,n})^{-1}\mathbf{\Delta}_{t,n}$, with $\ell = \frac{10}{4}$, and set $\mathbf{S}_t = T_\rho(\mathbf{E}_{Z,t})$.

   (vi) set $(\mathbf{Y}_t, \mathbf{Z}_t) = (\mathbf{L}_t, \mathbf{S}_t) + \left\{ \frac{\eta_{t-1}-1}{\eta_t} \right\}\{(\mathbf{L}_t, \mathbf{S}_t) - (\mathbf{L}_{t-1}, \mathbf{S}_{t-1})\}$ where
   $\eta_t = \frac{1}{2} + \frac{1}{2}\sqrt{1 + 4\eta_{t-1}^2}$.

   (vii) stop if the convergence criterion $\frac{\|\mathbf{L}_t - \mathbf{L}_{t-1}\|_F}{1 + \|\mathbf{L}_{t-1}\|_F} + \frac{\|\mathbf{S}_t - \mathbf{S}_{t-1}\|_F}{1 + \|\mathbf{S}_{t-1}\|_F} \leq \varepsilon$.

4. Set $\widehat{\mathbf{L}}_{\text{ALCE}} = \mathbf{Y}_t$ and $\widehat{\mathbf{S}}_{\text{ALCE}} = \mathbf{Z}_t$.

References

1. Pourahmadi, M. *High-Dimensional Covariance Estimation: With High-Dimensional Data*; John Wiley & Sons: Hoboken, NJ, USA, 2013; Volume 882.
2. Wainwright, M.J. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*; Cambridge University Press: Cambridge, UK, 2019; Volume 48.
3. Zagidullina, A. *High-Dimensional Covariance Matrix Estimation: An Introduction to Random Matrix Theory*; Springer Nature: Berlin/Heidelberg, Germany, 2021.
4. Fan, J.; Liao, Y.; Liu, H. An overview of the estimation of large covariance and precision matrices. *Econom. J.* **2016**, *19*, C1–C32. [CrossRef]
5. Lam, C. High-dimensional covariance matrix estimation. *Wiley Interdiscip. Rev. Comput. Stat.* **2020**, *12*, e1485. [CrossRef]
6. Ledoit, O.; Wolf, M. Shrinkage estimation of large covariance matrices: Keep it simple, statistician? *J. Multivar. Anal.* **2021**, *186*, 104796. [CrossRef]
7. Chamberlain, G.; Rothschild, M. Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica* **1983**, *51*, 1281. [CrossRef]
8. Fan, J.; Liao, Y.; Mincheva, M. Large covariance estimation by thresholding principal orthogonal complements. *J. R. Stat. Soc. Ser. (Stat. Methodol.)* **2013**, *75*, 603–680. [CrossRef] [PubMed]
9. Abdi, H; Williams, L.J. Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 433–459. [CrossRef]
10. Farnè, M.; Montanari, A. A large covariance matrix estimator under intermediate spikiness regimes. *J. Multivar. Anal.* **2020**, *176*, 104577. [CrossRef]
11. Fazel, M.; Hindi, H.; Boyd, S.P. A rank minimization heuristic with application to minimum order system approximation. In Proceedings of the American Control Conference, Arlington, VA, USA, 25–27 June 2001; Volume 6, pp. 4734–4739.
12. Fazel, M. *Matrix rank minimization with applications*. Ph.D. Thesis, Stanford University, Stanford, CA, USA, 2002.
13. Donoho, D.L. For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution. *Commun. Pure Appl. Math. J. Issued Courant Inst. Math. Sci.* **2006**, *59*, 797–829. [CrossRef]

14. Recht, B.; Fazel, M.; Parrilo, P.A. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.* **2010**, *52*, 471–501. [CrossRef]

15. Candès, J.E.; Plan, Y. Near-ideal model selection by l_1 minimization. *Ann. Stat.* **2009**, *37*, 2145–2177. [CrossRef]

16. Candès, J.E.; Tao, T. The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. Inf. Theory* **2010**, *56*, 2053–2080. [CrossRef]

17. Mazumder, R.; Hastie, T.; Tibshirani, R. Spectral regularization algorithms for learning large incomplete matrices. *J. Mach. Learn. Res.* **2010**, *11*, 2287–2322. [PubMed]

18. Srebro, N.; Rennie, J.; Jaakkola, T.S. Maximum-margin matrix factorization. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2005; pp. 1329–1336.

19. Hastie, T.; Mazumder, R.; Lee, J.D.; Zadeh, R. Matrix completion and low-rank svd via fast alternating least squares. *J. Mach. Learn. Res.* **2015**, *16*, 3367–3402. [PubMed]

20. Candès, E.J.; Li, X.; Ma, Y.; Wright, J. Robust principal component analysis? *J. ACM (JACM)* **2011**, *58*, 11. [CrossRef]

21. Danilova, M.; Dvurechensky, P.; Gasnikov, A.; Gorbunov, E.; Guminov, S.; Kamzolov, D.; Shibaev, I. Recent theoretical advances in non-convex optimization. *arXiv* **2020**, arXiv:2012.06188.

22. Dey, D.; Mukhoty, B.; Kar, P. Agglio: Global optimization for locally convex functions. In Proceedings of the 5th Joint International Conference on Data Science & Management of Data (9th ACM IKDD CODS and 27th COMAD), Bangalore, India, 8–10 January 2022; pp. 37–45.

23. Nesterov, Y. Gradient methods for minimizing composite functions. *Math. Program.* **2013**, *140*, 125–161. [CrossRef]

24. Harville, D.A. *Matrix Algebra from A Statistician's Perspective*; Springer: New York, NY, USA, 1997.

25. Horn, A.R.; Johnson, C.R. *Matrix Analysis*. Cambridge University Press: Cambridge, UK, 2012.

26. Graham, A. *Kronecker Products and Matrix Calculus: With Applications*; Ellis Horwood Limited: London, UK, 1981.

27. Lax, P.D. *Linear Algebra and Its Applications*, 2nd ed.; Pure and Applied Mathematics (Hoboken); Wiley-Interscience (John Wiley & Sons): Hoboken, NJ, USA, 2007.

28. Luo, X. High dimensional low rank and sparse covariance matrix estimation via convex minimization. *arXiv* **2011**, arXiv:1111.1133.

29. Cai, J.-F.; Candès, E.J.; Shen, Z. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* **2010**, *20*, 1956–1982. [CrossRef]

30. Daubechies, I.; Defrise, I.M.; Mol, C.D. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pure Appl. Math.* **2004**, *57*, 1413–1457. [CrossRef]