KYBERNETIKA — VOLUME *55* (2019), NUMBER 4, PAGES 740–754

AN ALTERNATING MINIMIZATION ALGORITHM FOR FACTOR ANALYSIS

Valentina Ciccone, Augusto Ferrante and Mattia Zorzi

The problem of decomposing a given covariance matrix as the sum of a positive semi-definite matrix of given rank and a positive semi-definite diagonal matrix, is considered. We present a projection-type algorithm to address this problem. This algorithm appears to perform extremely well and is extremely fast even when the given covariance matrix has a very large dimension. The effectiveness of the algorithm is assessed through simulation studies and by applications to three real benchmark datasets that are considered. A local convergence analysis of the algorithm is also presented.

Keywords: matrix decomposition, factor analysis, covariance matrices, low rank matrices, projections

Classification: 62H25

1. INTRODUCTION

The problem of decomposing a given covariance matrix into the sum of a low rank matrix L plus a diagonal matrix D bursts more than a century of tradition in scientific literature. In fact, it may be viewed as a linear algebraic counterpart of a *Factor Analysis* problem which is a problem in multivariate statistics, see [7, 9, 13, 16, 22, 25], with applications in countless fields of science. Factor analysis aims to extract statistical commonalities among data and is therefore a tool of great importance in signal processing as pointed out in [5] and [20]: we refer to these works for an extensive discussion on the importance of the problem, on its applications, on the formidable stream of literature produced on this topic in the last century, and on the numerous variants in which the problem can be formulated.

This work takes an optimization-oriented approach: for a given covariance matrix Σ and a given rank r, we seek a positive semidefinite matrix L with rank not larger than r and a positive semidefinite diagonal matrix D such that their sum is as close as possible to Σ. A closed-form solution for this problem appears to be out of reach. We propose an easy-to-implement iterative algorithm, based on alternating minimization, to solve numerically the considered problem. This algorithm appears to perform extremely well and in simulations converges very rapidly to the solution. Despite the simplicity of the algorithm, the convergence analysis is non trivial due to the non-convexity of the set of low rank matrices.

DOI: 10.14736/kyb-2019-4-0740

Notation. Given a matrix M, $M^\top$ denotes its transpose and $\mathrm{tr}(M)$ denotes its trace (for a square M). The symbol $\mathbf{Q}_n$ denotes the space of real symmetric matrices of size n. If $X \in \mathbf{Q}_n$ is positive definite or positive semi-definite we write $X \succ 0$ or $X \succeq 0$, respectively. We denote by $\mathbf{D}_n$ the space of diagonal matrices of size n an by $\mathbf{O}_n$ the set of orthogonal matrices of size n. The Frobenius norm is denoted by $\|\cdot\|_F$ while $\|\cdot\|$ denotes the Euclidean norm.

2. PRELIMINARIES AND PROBLEM DEFINITION

Factor models are used to described high dimensional vectors of data in terms of a small number of common latent factors. In its simplest formulation, the classic (linear static) factor model is given by

$$y = Ax + z \tag{1}$$

where $A \in \mathbb{R}^{n\times r}$, with $r \ll n$, is the so-called factor loading matrix, x is the vector of (independent) latent factors and z represents the idiosyncratic component. Here, x and z are zero-mean, independent Gaussian random vectors; the covariance matrix of x is the identity matrix of dimension r and the covariance matrix of z is a diagonal matrix $D \in \mathbf{D}_n$. Note that, Ax represents the latent variable. Clearly, y is itself a Gaussian random vector with zero mean and we denote by Σ its covariance matrix. Since x and z are independent it holds that

$$\Sigma = L + D \tag{2}$$

where $L := AA^\top$ and D are the covariance matrices of Ax and z, respectively. Thus, L has rank equal to r, and D is diagonal. Hence, in its original conception the construction of a factor model is mathematically equivalent to a matrix additive decomposition problem which seeks, for a given Σ, a decomposition of the type of (2). Of course the model is maximally parsimonious if the rank of L is minimum. The problem of minimizing the rank of L in decomposition (2) is known as Frisch's problem and, to date, no exact solution for such a problem is actually available, with the only exception of the special case when this minimum rank is $r = n - 1$, in which case a closed-form parametrization of the solutions is provided in [21]. This lack of explicit formulas has motivated a rich stream of literature and different numerical approaches which have been proposed over the years. A relaxation of this problem has also been considered in which the matrix D is only required to be diagonal but not positive semi-definite. The main difficulty in these problems is related to the non-convexity of the rank function so that a viable alternative is to consider the so called *minimum trace factor analysis* problem, [10, 23]:

$$\min_{L,D\in\mathbf{Q}_n} \quad \mathrm{tr}(L) \quad \text{subject to} \quad L, D \succeq 0,\ \Sigma = L + D,\ D \in \mathbf{D}_n, \tag{3}$$

where the trace of L is used as convex surrogate of the rank function as in [11, 12].

Note that, in many cases the equality constraint in (3) may be too compelling. Therefore, an alternative approach is to allow for residuals in the decomposition. Typically, this leads to an optimization problem where the residual $\Sigma - L - D$ is minimized with respect to a chosen norm under a constraint limiting the rank of L. This approach is known as *minimum residual factor analysis*, see [5, 15, 24]. Note that the presence of the rank constraint makes such problems non convex and several heuristic have been proposed to deal with it. Other approaches to factor analysis encompass: principal component factor analysis as in [4], maximum likelihood methods as in [2], or the establishing of a certificate of optimal low rank as in [14]. Moreover, several variants of the mentioned approaches have been proposed by weakening modelling assumptions or by introducing additional constraints for example to account for errors in the covariance matrix estimation as in [8, 20] and [1].

The problem we are going to consider is a minimum residual type problem: for a given r and a given matrix Σ we want to find a positive semidefinite matrix L with rank at most r and a positive semidefinite diagonal matrix D such that their sum is as close as possible to Σ. This can be formalized as follows:

$$(L^*, D^*) := \underset{L\in\mathcal{L}_{n,r}, D\in\mathcal{D}_n}{\arg\min} \quad \|\Sigma - L - D\|_F^2 \tag{4}$$

where the sets $\mathcal{L}_{n,r}$ and $\mathcal{D}_n$ are defined as:

$$\mathcal{L}_{n,r} := \{X \in \mathbf{Q}_n : X \succeq 0,\ \operatorname{rank}(X) \le r\}, \quad \mathcal{D}_n := \{X \in \mathbf{D}_n : X \succeq 0\}.$$

Note that, in practice, r can be obtained by resorting to available methods for estimating the number of factors, see [3, 6, 8] and references therein. Alternatively, the problem can be solved for increasing values of r until the residue $\|\Sigma - L^* - D^*\|_F$ is not greater than a certain tolerance. In the case when Σ is the sample covariance matrix estimated from the data, the residue $\Sigma - L^* - D^*$ accounts for the uncertainty in the estimation of Σ, and our problem is equivalent to find a good trade-off between the fit term (i. e. the residue) and the complexity of the model (i. e. r).

3. THE PROPOSED ALGORITHM

A closed-form solution for Problem (4) appears to be out of reach. However, this Problem appears to be well suited for a coordinate descent type iterative algorithm. Such algorithm alternates between solving a minimization problem with respect to L and a minimization problem with respect to D:

$$L_k = \underset{L\in\mathcal{L}_{n,r}}{\arg\min}\, \|\Sigma - L - D_{k-1}\|_F, \qquad D_k = \underset{D\in\mathcal{D}_n}{\arg\min}\, \|\Sigma - L_k - D\|_F, \tag{5}$$

where L_k and D_k denote the values of L and D at the kth iteration. Both these sub-problems admit explicit solutions which are provided by the projection operators onto the sets $\mathcal{L}_{n,r}$ and $\mathcal{D}_n$, respectively. Indeed, let $X \in \mathbf{Q}_n$ and consider its spectral decomposition $X = USU^\top$, $U \in \mathbf{O}_n$ and $S = \operatorname{diag}(s_1, \ldots, s_n)$ with $s_1 \ge s_2 \ge \ldots \ge s_n$ being the eigenvalues of X arranged in decreasing order. Then, the matrix with rank at most r that is closest to X in the Frobenius norm is obtained applying the projector $P_{\mathcal{L}_{n,r}}$:

$$P_{\mathcal{L}_{n,r}}(X) := U \operatorname{diag}(f_l(s_1), \ldots, f_l(s_n))U^\top$$

with $f_l(\cdot)$ defined as

$$f_l(s_i) := \begin{cases} s_i & \text{for } i \le r \wedge s_i > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

On the other hand, the projector $P_{\mathcal{D}_n}$ onto the set $\mathcal{D}_n$ is:

$$P_{\mathcal{D}_n}(X) := \operatorname{diag}(f_d(X_{11}), \ldots, f_d(X_{nn})) \tag{7}$$

with $f_d(\cdot)$ defined as

$$f_d(X_{ii}) := \max\{X_{ii},\ 0\} \tag{8}$$

and X_{ii} is the entry of X in row i and column i.

Then, at kth iteration the algorithm computes $L_k := P_{\mathcal{L}_{n,r}}(\Sigma - D_{k-1})$ and $D_k := P_{\mathcal{D}_n}(\Sigma - L_k)$. The complete procedure is outlined in Algorithm 1: $\varepsilon > 0$ is the maximum error allowed in the relative decomposition error, while N represents the maximum number of iterations.

Algorithm 1

```
Input: $\Sigma$, $r$, $\epsilon$, N
Output: $L^*$, $D^*$
Initialize: initialize $D$ randomly, i=0
while $\|\Sigma - L - D\|_F^2 / \|\Sigma\|_F^2 < \epsilon$ and i < N
    $L = P_{\mathcal{L}_{n,r}}(\Sigma - D)$
    $D = P_{\mathcal{D}_n}(\Sigma - L)$
    i=i+1
end while
    $L^* = L$, $D^* = D$
```

4. NUMERICAL SIMULATIONS

To provide empirical evidence of the convergence properties of the algorithm simulations studies have been performed by using the software Matlab-R2012b on a 2014 laptop MacBook Pro, Quad-i7 2.0 GHz.

Simulations with synthetic data. To begin with, we have considered the case of a covariance matrix, Σ, computed as the sum of a randomly generated positive semidefinite low-rank matrix L of dimension n and rank r, and a randomly generated positive definite diagonal matrix D. We have performed 200 Monte Carlo runs with $n = 40$ and $r = 4$ and 200 runs with $n = 40$ and $r = 10$. The original low-rank and diagonal matrices are recovered with negligible numerical errors. Indeed the following quantities:

- the relative decomposition error on $L + D$: $\|\Sigma - L^* - D^*\| / \|\Sigma\|$;
- the relative error on L: $\|L - L^*\| / \|L\|$;
- the relative error on D: $\|D - D^*\| / \|D\|$;

are all of the order of 10^{-10}. The average computational time for each experiment is less than five hundredths of a second: in less than half minute all 2 × 200 runs converged.

To account for how the algorithm scales with the dimensionality of the problem two further numerical experiments have been conducted with increasing values of the dimension n: $n = 20 * 2^j$, with $j = 0 \dots 5$. In the first experiment, the rank was fixed at the value $r = 8$, while in the second experiment the ratio r/n was fixed the value 0.2. In both experiments, for each value of n, 50 factor models have been generated and the resulting covariance matrices have been used as input for our algorithm which recovered the correct decomposition except for a handful of times. The statistics of the execution time (in seconds) are summarized in Table 1 for the first experiment and in Table 2 for the second. Both experiments provide evidence that the algorithm scales extremely well with dimensionality.

n	r	mean	st. dev.
20	8	0.0610	0.0545
40	8	0.0349	0.0092
80	8	0.0642	0.0081
160	8	0.1927	0.0173
320	8	1.2286	0.0875
640	8	5.4973	0.2793
1280	8	26.3725	0.9813

Tab. 1. For each value of n the table displays the mean execution time (in seconds) and standard deviation across 50 experiments.

n	r	mean	st. dev.
20	4	0.0219	0.0452
40	8	0.0362	0.0093
80	16	0.1069	0.0171
160	32	0.3842	0.0444
320	64	2.7974	0.1565
640	128	14.7181	0.6733
1280	256	85.5031	26.8018

Tab. 2. For each couple of (n, r) the table displays the mean execution time (in seconds) and standard deviation across 50 experiments.

Finally, we have considered the case of a covariance matrix which admits only approximately a "low-rank plus diagonal" decomposition. Given a covariance matrix Σ generated as before (which therefore admits an exact "low-rank plus diagonal" decomposition), we have generated a sample of numerosity N from the distribution $\mathcal{N}(0, \Sigma)$ and we have estimated the corresponding sample covariance $\hat{\Sigma}_N$ which have been used as input for the algorithm. We have considered the same setting as before with $n = 40$, $r = 4$ and $n = 40$, $r = 10$. In both cases for each sample size $N = 200, 500, 1000$ we have performed 200 Monte Carlo runs. The 6×200 simulations took less than 5 minutes to converge and we observed the following:

1. In all the 6×200 simulations the sequence (D_k, L_k) produced by Algorithm 1

converged to a stationary point (L^*, D^*) and, as discussed in Proposition 5.1 below, this point is a (at least) local optimum.

2. In all the 6 × 200 simulations, the inequality

$$\|L^* + D^* - \hat{\Sigma}_N\|_F - \|\Sigma_{true} - \hat{\Sigma}_N\|_F \leq 0$$

is satisfied which provides a sanity check on the performance of the proposed algorithm. In fact, especially for $N = 1000$, Σ_{true} may be viewed as a good approximation of $\hat{\Sigma}_N$ and, on the other hand, we know that, by construction, Σ_{true} may be decomposed as the sum of a low rank positive semidefinite matrix and a diagonal positive matrix. Hence, $\Sigma_{true} = L_{true} + D_{true}$ may be viewed as a benchmark which is always outperformed by the decomposition provided by the proposed algorithm.

The results for the decomposition error $\|\hat{\Sigma}_N - L^* - D^*\|/\|\hat{\Sigma}\|$ are summarized in Figures 1 and 2.

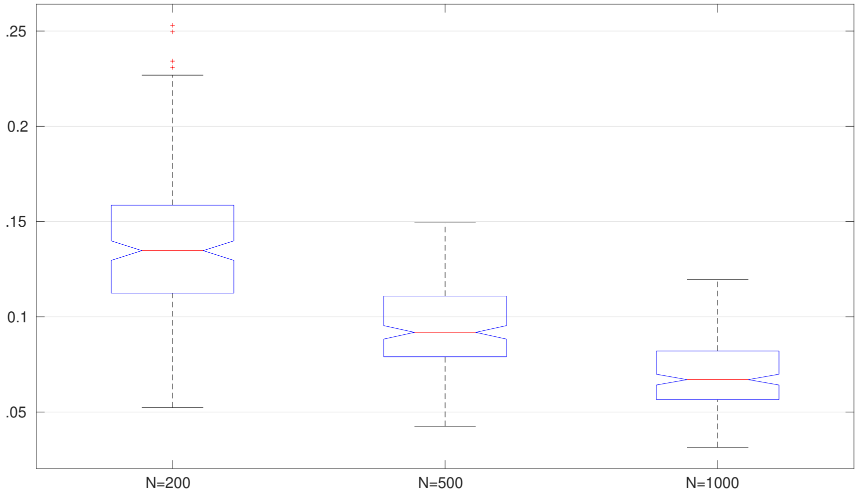

Fig. 1. Case $r = 4$. Decomposition errors $\|\hat{\Sigma}_N - L^* - D^*\|/\|\hat{\Sigma}_N\|$ with $N = 200$, $N = 500$ and $N = 1000$.

Figures 3 and 4 display the following quantities:

- the relative decomposition error on $L + D$: $\|\Sigma - L^* - D^*\|/\|\Sigma\|$;
- the relative error on L: $\|L - L^*\|/\|L\|$;
- the relative error on D: $\|D - D^*\|/\|D\|$.

The obtained results appear extremely promising.

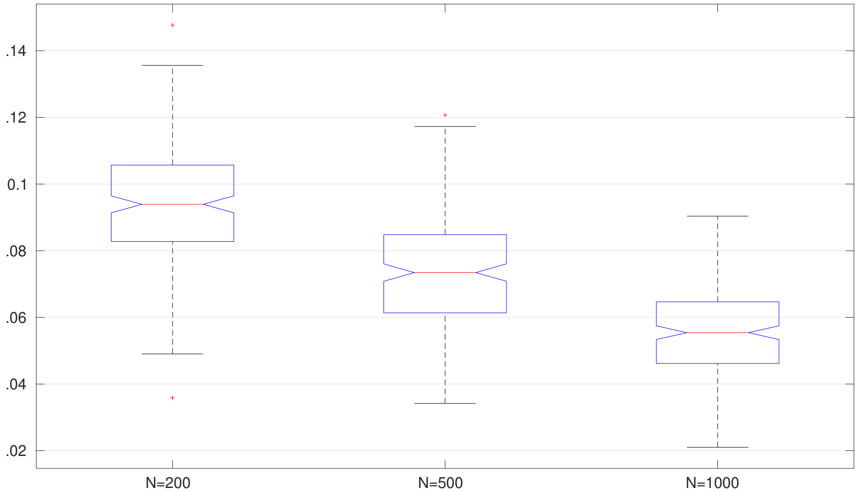

Fig. 2. Case $r = 10$. Decomposition errors $\|\hat{\Sigma}_N - L^* - D^*\|/\|\hat{\Sigma}_N\|$ with $N = 200$, $N = 500$ and $N = 1000$.

Applications to real data. We now investigate the performance of the proposed method on three real world datasets which are popular benchmark in factor analysis:

- the bfi dataset, from the R library psych, which consists of 2800 observations on 28 variables: 25 variables represent personality self-reported items and while 3 are demographic variables;
- the neo dataset, also from the R library psych, which consists of a correlation matrix of size 30 × 30 estimated from 1000 observations;
- the Harman dataset, from the R library datasets, which consists of a correlation matrix of size 24×24 estimated from 145 observations: the cross-section represents psychological tests carried out to seventh- and eighth-grade children.

These datasets have been used in [5, Section 5.3] to compare the performance of their approach, which minimizes the q-norm of the residue (with $q = 1$), against different factor analysis methods. This approach can be considered as the state of the art as it outperforms the other available methods. In this section we take it as benchmark for comparisons and we repeat the analysis in [5, Section 5.3].
The adopted measure of performance is the explained variance, defined as

$$\sum_{i=1}^{r} \lambda_i(L^*)/\sum_{i=1}^{n} |\lambda_i(\Sigma - D^*)|,$$

where $\lambda_i(\cdot)$ denotes the ith largest eigenvalue. For each dataset Problem (4) is solved for the values of r considered in [5]. The results are depicted in Figure 5. The proposed

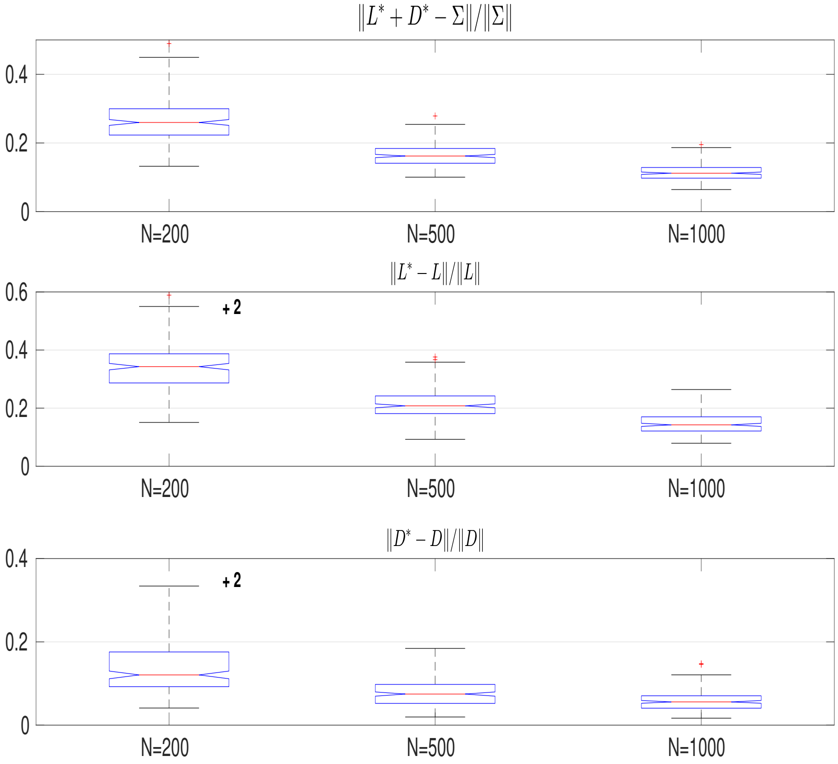

Fig. 3. Case $r = 4$. The displayed quantities are: $\|\Sigma - L^* - D^*\|/\|\Sigma\|$, $\|L - L^*\|/\|L\|$ and $\|D - D^*\|/\|D\|$, where L^* and D^* represent the estimates with $N = 200$, $N = 500$ and $N = 1000$.

method provides a higher amount of explained variance with respect to [5]. Similarly to [5], our method shows a flexibility in delivering different models with varying r.

5. CONVERGENCE ANALYSIS

We now discuss the convergence of the proposed algorithm to a local minimum. First, we observe that the iterative minimization in (5) produces a sequence of values for the objective function that is monotonically non-increasing. Since the objective function is bounded from below, we have the following obvious result.

Lemma 5.1. For $h \in \mathbb{N}$, define the sequence F_h by $F_h := \|\Sigma - L_k - D_k\|_F^2$ for $h = 2k$ (even), and $F_h := \|\Sigma - L_{k+1} - D_k\|_F^2$, for $h = 2k+1$ (odd), where L_k, D_k is the sequence produced by Algorithm 1. Then the sequence F_h is monotonically non-increasing and has limit as $h \to \infty$.

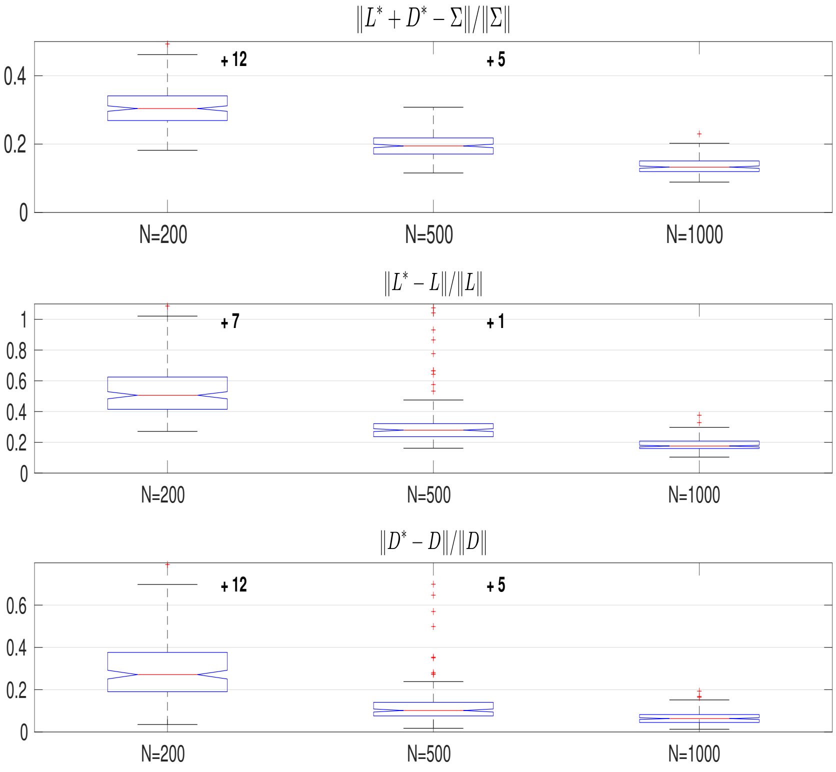

Fig. 4. Case $r = 10$. The displayed quantities are: $\|\Sigma - L^* - D^*\|/\|\Sigma\|$, $\|L - L^*\|/\|L\|$ and $\|D - D^*\|/\|D\|$, where L^* and D^* represent the estimates with $N = 200$, $N = 500$ and $N = 1000$.

Establishing the convergence for L_k and D_k is less trivial. We start with D_k. To this aim we observe that as a consequence of Lemma 5.1, we have that $\varepsilon_k := F_{2k-1} - F_{2k}$ not only converges to zero but it converges sufficiently fast.

Lemma 5.2. Assume that $\varepsilon_k := F_{2k-1} - F_{2k}$ tends to zero faster than $1/k^{2q}$ with $q > 1$ and let D_k be the sequence of diagonal matrices produced by Algorithm 1. Then the sequence D_k converges to a certain diagonal matrix $D^* \in \mathcal{D}_n$.

Proof. We have

$$F_{2k} = \|\Sigma - L_k - D_k\|_F^2 = F_{2k-1} - \varepsilon_k = \|\Sigma - L_k - D_{k-1}\|_F^2 - \varepsilon_k.$$

Let $s_k(i) := [\Sigma - L_k]_{ii}$ be the ith element in the diagonal of $\Sigma - L_k$ and $d_k(i) := [D_k]_{ii}$ be the ith element in the diagonal of D_k. Since in (7) for each i, $d_k(i)$ is chosen

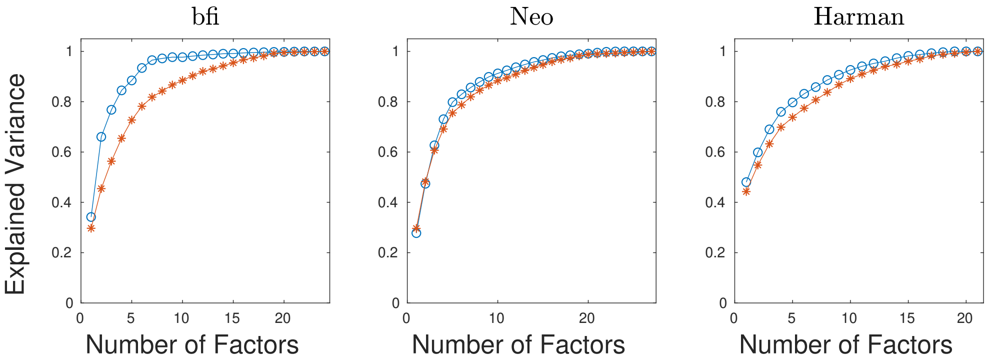

Fig. 5. Proportion of variance explained by our method (blue circles) and by the benchmark method (red stars).

independently of the others in order to minimize $\|\Sigma - L_k - D_k\|_F^2$, we have that

$$\begin{aligned}-\varepsilon_k &= \|\Sigma - L_k - D_k\|_F^2 - \|\Sigma - L_k - D_{k-1}\|_F^2\\ &= \sum_{i=1}^{n}\{[s_k(i) - d_k(i)]^2 - [s_k(i) - d_{k-1}(i)]^2\} \leq [s_k(i) - d_k(i)]^2 - [s_k(i) - d_{k-1}(i)]^2\end{aligned}$$

which yields

$$\varepsilon_k \geq [d_{k-1}(i) - d_k(i)][d_k(i) + d_{k-1}(i) - 2s_k(i)].$$

Now, we can consider two cases: if $s_k(i) \geq 0$, then the minimizer $d_k(i)$ is equal to $s_k(i)$, so that we have

$$\varepsilon_k \geq [d_{k-1}(i) - d_k(i)]^2.$$

If $s_k(i) < 0$, then $d_k(i) = 0$ so that we have again

$$\varepsilon_k \geq d_{k-1}(i)[d_{k-1}(i) - 2s_k(i)] \geq [d_{k-1}(i) - d_k(i)]^2.$$

In conclusion, in both cases, we have

$$|d_k(i) - d_{k-1}(i)| \leq \alpha_k := \sqrt{\varepsilon_k}.$$

As a consequence, we have

$$\begin{aligned}|d_{k+m}(i) - d_k(i)| &\leq |d_{k+m}(i) - d_{k+m-1}(i)| + |d_{k+m-1}(i) - d_{k+m-2}(i)| + \dots\\ &\quad + |d_{k+1}(i) - d_k(i)|\\ &\leq \alpha_{k+m} + \dots + \alpha_{k+1} \leq \sum_{l=1}^{m}\frac{M}{(k+l)^q} \leq \sum_{l=1}^{\infty}\frac{M}{(k+l)^q}\\ &= \sum_{h=k+1}^{\infty}\frac{M}{h^q} = \left[\sum_{h=1}^{\infty}\frac{M}{h^q} - \sum_{h=1}^{k}\frac{M}{h^q}\right]\end{aligned}$$

where M is a constant and $q > 1$ so that all the infinite sums converge to a finite value. Since we have

$$\lim_{k\to\infty}\left[\sum_{h=1}^{\infty}\frac{M}{h^q}-\sum_{h=1}^{k}\frac{M}{h^q}\right]=\sum_{h=1}^{\infty}\frac{M}{h^q}-\lim_{k\to\infty}\sum_{h=1}^{k}\frac{M}{h^q}=0,$$

we can conclude that $\lim_{l,k\to\infty}|d_l(i)-d_k(i)|=0$, so that $d_k(i)$ is a Cauchy sequence and hence it converges. Since this holds for each $i=1,\dots,n$, we have that the sequence D_k converges to a certain diagonal matrix D^*. Finally, since $\mathcal{D}_n$ is closed, clearly $D^*\in\mathcal{D}_n$. □

For the convergence of the sequence L_k we need to rule out a pathological situation.

Lemma 5.3. Under the assumptions of Lemma 5.2, let $S := \Sigma - D^*$ with D^* being the limit of the sequence of diagonal matrices D_k produced by Algorithm 1. If S has n distinct eigenvalues then the sequence of rank r matrices L_k produced by Algorithm 1 converges to a rank r matrix L^*.

Proof. Let $s_1 > s_2 > \ldots > s_n$ be the eigenvalues of S arranged in decreasing order. By continuity of the eigenvalues, for a sufficiently large k, $S_k := \Sigma - D_k$ has n distinct eigenvalues $s_{k,1} > s_{k,2} > \cdots > s_{k,n}$ and $\lim_{k\to\infty} s_{k,i} = s_i$. According to [18, Chapter 9, Theorem 8], for each $i = 1,\dots,n$, we can select an eigenvector (and hence a normalized eigenvector) $v_{k,i}$ of S_k associated with the eigenvalue $s_{k,i}$ in such a way that $v_{k,i}$ converges to a normalized eigenvector of S associated with the eigenvalue s_i. Now recall that

$$L_{k+1} = P_{\mathcal{L}_{n,r}}(S_k) = U_k \operatorname{diag}(f_l(s_{k,1}), \ldots, f_l(s_{k,n}))U_k^\top$$

where the ith column of U_k is a normalized eigenvector of S_k associated with the eigenvalue $s_{k,i}$. As a normalized eigenvector is unique up to its sign, we have $U_k = V_k\Delta_k$ with $V_k := [v_{k,1} \mid v_{k,2} \mid \cdots \mid v_{k,n}]$ and Δ_k is a diagonal matrix whose diagonal entries can only be ±1. We easily see that the contribution of the Δ_k cancels and we have

$$L_{k+1} = V_k \operatorname{diag}(f_l(s_{k,1}), \ldots, f_l(s_{k,n}))V_k^\top$$

so that L_{k+1} is the product of three matrices each one of which converges as k tends to infinity. □

Proposition 5.1. Assume that the hypothesis of Lemma 5.3 holds and let L^* be the matrix defined in the same lemma and r be its rank. Assume also that the tangent space of $\mathcal{L}_{n,r}$ at L^* does not contain diagonal matrices.[1] Then the sequence (D_k, L_k) produced by Algorithm 1 converges to a point corresponding to a local minimum of the cost function.

[1] This assumption is reasonable as the dimension of $\mathcal{L}_{n,r}$, and hence of its tangent space at L^*, is $rn - r(r-1)/2$, [17]. Therefore, whenever r is not too close to n (which is the interesting case) the number $n(n-1)/2$ of the constraints needed to impose that an element of the tangent space at L^* is diagonal is larger than the number $rn - r(r-1)/2$ of available parameters. Hence, except for very special cases, the assumption indeed holds.

Proof. By the previous results, we know that D_k converges to D^* and L_k converges to L^*. Assume by contradiction that (D^*, L^*) is not a minimum. Then, for any $\varepsilon > 0$, there exists δD and δL such that $\|\delta D\|_F < \varepsilon$, $\|\delta L\|_F < \varepsilon$, $(L + \delta L) \in \mathcal{L}_{n,r}$, $(D + \delta D) \in \mathcal{D}_n$ and

$$\|\Sigma - L^* - D^*\|_F^2 > \|\Sigma - L^* - D^* - \delta L - \delta D\|_F^2.$$

Now let δT be the projection of δL on the tangent space of $\mathcal{L}_{n,r}$ at L^*. For a sufficiently small ε we have

$$\|\Sigma - L^* - D^*\|_F^2 \geq \|\Sigma - L^* - D^* - \delta T - \delta D\|_F^2.$$

By setting $R := \Sigma - L^* - D^*$ and computing the Frobenius norms in the previous formula, we get

$$2(\mathrm{tr}[R\delta T] + \mathrm{tr}[R\delta D]) - \|\delta T + \delta D\|_F^2 \geq 0.$$

By assumption $\delta T + \delta D \neq 0$ so that at least one of the two quantities $2\,\mathrm{tr}[R\delta L]$ and $2\,\mathrm{tr}[R\delta D]$ is positive. In the case of $\mathrm{tr}[R\delta D] > 0$ we have that for all κ sufficiently small,

$$\min_{D\in\mathcal{D}_n} \|\Sigma - L^* - D\|_F^2 \leq \|\Sigma - L^* - D^* - \kappa\delta D\|_F^2 = \|R\|_F^2 + \kappa^2\|\delta D\|_F^2 - 2\kappa\,\mathrm{tr}[R\delta D] < \|R\|_F^2$$

which is contradiction because the algorithm converges so that $\min_{D\in\mathcal{D}_n} \|\Sigma - L^* - D\|_F^2 = \|R\|_F^2$.

In the case of $\mathrm{tr}[R\delta T] > 0$ we have that

$$\min_{L\in\mathcal{L}_{n,r}} \|\Sigma - L - D^*\|_F^2 \leq \|\Sigma - D^* - P_{\mathcal{L}_{n,r}}(L^* + \kappa\delta T)\|_F^2 \tag{9}$$

where $P_{\mathcal{L}_{n,r}}(\cdot)$ is the projection onto $\mathcal{L}_{n,r}$. Thus we have $P_{\mathcal{L}_{n,r}}(L^* + \kappa\delta T) = L^* + \kappa\delta T + E$ where $\lim_{\kappa\to 0} \|E\|_F/\kappa = 0$.

Thus, for $\kappa > 0$ sufficiently small, we have

$$\begin{aligned} q &:= \|\Sigma - D^* - P_{\mathcal{L}_{n,r}}(L^* + \kappa\delta T)\|_F^2 = \|\Sigma - D^* - L^* - \kappa\delta T - E\|_F^2 \\ &= \|R\|_F^2 + \kappa^2\|\delta T\|_F^2 + \|E\|_F^2 - 2\kappa\,\mathrm{tr}(R\delta T) - 2\,\mathrm{tr}(RE) + 2\kappa\,\mathrm{tr}(\delta T E) < \|R\|_F^2. \end{aligned}$$

In conclusion, we have

$$\min_{L\in\mathcal{L}_{n,r}} \|\Sigma - L - D^*\|_F^2 < \|R\|_F^2, \tag{10}$$

that, as in the previous case leads to a contradiction. □

Remark 1. It is quite intuitive that the conditions of Proposition 5.1 are not very stringent: in fact in all the practical situations that we have studied in simulations those conditions are satisfied.

Remark 2. Since Problem (4) is non-convex and may have, in general, many local minima, we cannot make any claim on the global optimality. However, we have performed massive simulations for the case when Σ is synthetically produced as the sum of a positive definite diagonal matrix D and a positive semi-definite matrix L. In this case, if we select $r = \mathrm{rank}(L)$, the (globally) optimal solution clearly corresponds to a zero residual error and we can therefore evaluate the performance of our algorithm. The empirical evidence shows that the number of cases in which our algorithm does not produce a zero residual error is negligible. Even in these cases the residual error norm was never larger than 1% of the norm of Σ and a numerically zero residual error was always obtained by repeating a few times the minimization with perturbed initial conditions.

6. AN ALTERNATING PROJECTION TYPE ALGORITHM

In this section we present our algorithm under a different perspective that may be useful in addressing questions on the properties of the proposed method. In fact, by suitably translating $\mathcal{D}_n$, we easily see that this method can be viewed as an alternating projection type algorithm for which a very rich literature has been developed. To this aim, define

$$\tilde{\mathcal{D}}_n := \Sigma - \mathcal{D}_n = \{X \in \mathbf{Q}_n : X_{ij} = \Sigma_{ij}, \forall i \neq j,\, X_{ii} \leq \Sigma_{ii}, i = 1, \dots, n\} \tag{11}$$

and notice that the projection in this affine set is easily obtained as:

$$P_{\tilde{\mathcal{D}}_n}(X) := \begin{cases} X_{ij} = X_{ij} & \text{for } i = j \wedge X_{ii} < \Sigma_{ii} \\ X_{ij} = \Sigma_{ij} & \text{for } (i = j \wedge X_{ii} \geq \Sigma_{ii}) \vee i \neq j. \end{cases} \tag{12}$$

We consider now the sequences L_k and D_k produced by our algorithm. We recall that our D_k is given by $D_k = P_{\mathcal{D}_n}(\Sigma - L_k)$. By taking this formula into account, a direct computation shows that the matrix $\tilde{D}_k := P_{\tilde{\mathcal{D}}_n}(L_k)$ may be written as $\Sigma - D_k$ so that, in view of the formula $L_k = P_{\mathcal{L}_{n,r}}(\Sigma - D_{k-1})$, we immediately get that

$$L_{k+1} = P_{\mathcal{L}_{n,r}}(P_{\tilde{\mathcal{D}}_n}(L_k))$$

which shows that the iteration for L_k is the result of an alternating projection algorithm. These kind of algorithms burst a long tradition which dates back to Von Neumann in the '30s. While for alternating projection onto convex sets the convergence results are well established, for the non-convex case much less is known. In our case $\tilde{\mathcal{D}}_n$ is a convex set of dimension n, but the set $\mathcal{L}_{n,r}$ is a non-convex embedded manifold of $\mathbb{R}^{n \times n}$ with dimension $nr - r(r-1)/2$ and it is smooth at those points for which the rank is exactly r. In [19] a proof of local convergence (at a linear rate) for alternating projection onto smooth manifolds is provided under the assumption of transversal intersection. In our case, transversal intersection cannot hold when r is small with respect to n but it may be possible to generalise that approach to provide a further analysis of the algorithm properties and, in particular, of its convergence rate.

Finally, the set $\tilde{D}_n$ is particularly interesting because of the following interpretation that is particularly evident when r is such that Σ can be decomposed exactly as $L^* + D^*$ so that (L^*, D^*) is clearly an optimal solution of (4). In this case, $D^* = \Sigma - L^*$ and thus $\Sigma - L^* \in \mathcal{D}_n$. The latter condition is equivalent to the condition $L^* \in \tilde{\mathcal{D}}_n$ Therefore the problem (4) can reformulated only in terms of L as follows:

$$L^* := \underset{L \in \mathcal{L}_{n,r} \cap \tilde{\mathcal{D}}_n,}{\arg\min} \quad \|\Sigma - L\|_F^2. \tag{13}$$

7. CONCLUSIONS

We have proposed an alternating minimization algorithm for decomposing a covariance matrix as sum of a low rank matrix L, whose maximal rank is a priori fixed, plus a diagonal matrix D. This algorithm minimizes the norm of the residual difference between the covariance matrix and the sum $L + D$. Simulation results showed that the algorithm performs extremely well and converges very rapidly to the solution. Finally, we have proved that, under reasonable assumptions, this algorithm converges to a local minimum.

(Received March 6, 2019)

REFERENCES

[1] A. Agarwal, S. Negahban, and M. J. Wainwright: Noisy matrix decomposition via convex relaxation: Optimal rates in high dimensions. Ann. Statist. *40* (2012), 2, 1171–1197. DOI:10.1214/12-aos1000

[2] J. Bai, K. Li, et al.: Statistical analysis of factor models of high dimension. Ann. Statist. *40* (2012), 1, 436–465.

[3] J. Bai and S. Ng: Determining the number of factors in approximate factor models. Econometrica *70* (2002), 1, 191–221. DOI:10.1111/1468-0262.00273

[4] J. Bai and S. Ng: Large dimensional factor analysis. Found. Trends Econometr. *3* (2008), 2, 89–163.

[5] D. Bertsimas, M. S. Copenhaver, and R. Mazumder: Certifiably optimal low rank factor analysis. J. Machine Learning Res. *18* (2017), 29, 1–53.

[6] V. Ciccone, A. Ferrante, and M. Zorzi: Factor analysis with finite data. In: Proc. 56th Annual Conference on Decision and Control (CDC), VIC, Melbourne 2017, pp. 4046–4051. DOI:10.1109/cdc.2017.8264253

[7] V. Ciccone, A. Ferrante, and M. Zorzi: Robust identification of "Sparse Plus Low-rank" graphical models: An optimization approach. In: Proc. 2018 IEEE Conference on Decision and Control (CDC), Miami Beach 2018, pp. 2241–2246. DOI:10.1109/cdc.2018.8619796

[8] V. Ciccone, A. Ferrante, and M. Zorzi: Factor models with real data: A robust estimation of the number of factors. IEEE Trans. Automat. Control *64* (2019), 6, 2412–2425. DOI:10.1109/tac.2018.2867372

[9] M. Deistler and C. Zinner: Modelling high-dimensional time series by generalized linear dynamic factor models: An introductory survey. Comm. Inform. Systems *7* (2007), 2, 153–166. DOI:10.4310/cis.2007.v7.n2.a3

[10] G. Della Riccia and A. Shapiro: Minimum rank and minimum trace of covariance matrices. Psychometrika *47* (1982), 443–448. DOI:10.1007/bf02293708

[11] M. Fazel: Matrix rank minimization with applications. Elec. Eng. Dept. Stanford University *54* (2002), 1–130.

[12] M. Fazel, H. Hindi, and S. Boyd: Rank minimization and applications in system theory. In: Proc. American Control Conference *4* (2004), pp. 3273–3278. DOI:10.23919/acc.2004.1384521

[13] J. Geweke: The dynamic factor analysis of economic time series models. In: Latent Variables in Socio-Economic Models, SSRI workshop series, North-Holland 1977, pp. 365–383.

[14] L. Guttman: Some necessary conditions for common-factor analysis. Psychometrika *19* (1954), 2, 149–161. DOI:10.1007/bf02289162

[15] H. H. Harman and W. H. Jones: Factor analysis by minimizing residuals (minres). Psychometrika *31* (1066), 3, 351–368. DOI:10.1007/bf02289468

[16] C. Heij, W. Scherrer, and M. Deistler: System identification by dynamic factor models. SIAM J. Control Optim. *35* (1997), 6, 1924–1951. DOI:10.1137/s0363012995282127

[17] U. Helmke and M. A. Shayman: Critical points of matrix least squares distance functions. Linear Algebra Appl. *215* (1995), 1–19. DOI:10.1016/0024-3795(93)00070-g

[18] P. D. Lax: Linear Algebra and Its Applications. Second edition. Wiley-Interscience, 2007.

[19] A. S. Lewis and J. Malick: Alternating projections on manifolds. Math. Oper. Res. *33* (2008), 1, 216–234. DOI:10.1287/moor.1070.0291

[20] L. Ning, T. T. Georgiou, A. Tannenbaum, and S. P. Boyd: Linear models based on noisy data and the Frisch scheme. SIAM Rev. *57* (2015), 2, 167–197. DOI:10.1137/130921179

[21] O. Reiersøl: Identifiability of a linear relation between variables which are subject to error. Econometrica: J. Economet. Soc. *18* (1950), 4, 375–389. DOI:10.2307/1907835

[22] W. Scherrer and M. Deistler: A structure theory for linear dynamic errors-in-variables models. SIAM J. Control Optim. *36* (1998), 6, 2148–2175. DOI:10.1137/s0363012994262464

[23] A. Shapiro: Rank-reducibility of a symmetric matrix and sampling theory of minimum trace factor analysis. Psychometrika *47* (1982), 2, 187–199. DOI:10.1007/bf02296274

[24] A. Shapiro and J. M. F. Ten Berge: Statistical inference of minimum rank factor analysis. Psychometrika *67* (2002), 1, 79–94. DOI:10.1007/bf02294710

[25] M. Zorzi and R. Sepulchre: AR identification of latent-variable graphical models. IEEE Trans. Automat. Control *61* (2016), 9, 2327–2340. DOI:10.1109/tac.2015.2491678

Valentina Ciccone, Department of Information Engineering, University of Padova, Via Gradenigo 6/b, 35131 Padova. Italy.
e-mail: valentina.ciccone@dei.unipd.it

Augusto Ferrante, Department of Information Engineering, University of Padova, Via Gradenigo 6/b, 35131 Padova. Italy.
e-mail: augusto@dei.unipd.it

Mattia Zorzi, Department of Information Engineering, University of Padova, Via Gradenigo 6/b, 35131 Padova. Italy.
e-mail: zorzimat@dei.unipd.it