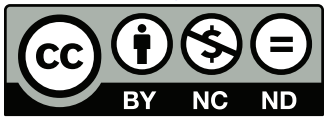
R.A. Delgado, J.C. Agüero and G.C. Goodwin. **A Rank-Constrained Optimization Approach: Application to Factor Analysis**. In *19th IFAC World Congress*, Cape Town, South Africa, 2014.

Available at http://dx.doi.org/10.3182/20140824-6-ZA-1003.02690



A Rank-Constrained Optimization approach: Application to Factor Analysis

Ramón A. Delgado * **Juan C. Agüero** * **Graham C. Goodwin** *

* *School of Electrical Engineering and Computer Science, The University of Newcastle, Australia*
{Juan.Aguero, Graham.Goodwin}@newcastle.edu.au,
Ramon.Delgado@uon.edu.au

Abstract: In this paper, we present a general method for rank-constrained optimization. We use an iterative convex optimization procedure where it is possible to include any extra convex constraints. The proposed approach has potential application in several areas. We focus on the problem of Factor Analysis. In this case, our approach provides sufficient flexibility to handle correlated errors. The benefits of the method is demonstrated via a simulation study.

1. INTRODUCTION

Rank-Constrained Optimization problems are ubiquitous in science and engineering (see e.g. Markovsky (2012)). Theses problems arise in areas such as system identification, control, statistics and signal processing (Grossmann et al., 2009; Kim and Moon, 2006; Markovsky, 2008; ten Berge and Kiers, 1991).

It is well known, that imposing rank-constraints into optimization and feasibility problems is a formidable task. Moreover, there does not currently exist a general procedure to include rank-constraints into optimization problems (Markovsky, 2008). Most of the existing methods use the specific structure of the problem of interest. For example, approaches based on Riemannian manifold optimization has been applied to linear regressions (Meyer, 2011), and to solve Lyapunov equations (Vandereycken, 2010). In addition, other approaches such as GECO (Shalev-Shwartz et al., 2011) and ADMiRA (Lee and Bresler, 2010) are based on greedy selection. Also, some Newton-like algorithms has been developed to solve Linear Matrix Inequalities subject to rank constraints (Orsi et al., 2006). Moreover, variable projection-type algorithms has been applied to structured low-rank approximation problems-Markovsky (2014). However, it is not clear how to extend these methods to other applications of interest. More recently, an heuristic approach based on the so called *nuclear norm* has been extensively used (Fazel, 2001, 2002). However, in this approach the condition on the rank is not considered as a hard constraint.

In this paper, we propose an approach to solve convex optimization problems subject to rank constraints. This approach allows one to constrain the rank of a matrix, while minimizing a cost function. In addition, it is possible to include any additional convex contraint into the proposed approach. We show the benefits of the proposed approach in a Factor Analysis problem that considers correlated errors.

The layout of remainder of the paper is as follows. In Section 2, we discuss the problem of interest. In Section 3, the proposed approach is presented. Section 4 presents an application to Factor Analysis in presence of correlated errors. A simulation study is presented in Section 5, and finally conclusions are draw in Section 6.

Notation: $\operatorname{rank} A$ denotes the rank of a matrix A. $\lambda_i(A)$ denotes the i-th largest eigenvalue of a matrix A, $A \succeq 0$ denotes that A is positive semidefinite, and $A \succeq B$ denotes that $A - B \succeq 0$. We represent the transpose of a given matrix A via $A^\top$. $\mathbb{S}^n$ denotes the set of symmetric matrices of size $n \times n$, and $\mathbb{S}^n_+$ the set of symmetric positive definite matrices, i.e. $\mathbb{S}^n_+ := \{A \in \mathbb{S}^n | A \succeq 0\}$.

2. PROBLEM DESCRIPTION

Consider the following rank-constrained optimization problem

$$\begin{aligned} \mathcal{P}_{rco}: \quad & \min_{\theta \in \mathbb{R}^p} \ f(\theta) \\ & \text{subject to } \ \theta \in \Omega \\ & \qquad\qquad \operatorname{rank} G(\theta) \leq r \\ & \qquad\qquad G(\theta) \in \mathbb{S}^n_+ \end{aligned}$$

where $\Omega \subset \mathbb{R}^p$ is a convex set, $f(\theta) : \mathbb{R}^p \to \mathbb{R}$ and $G(\theta) : \mathbb{R}^p \to \mathbb{S}^n_+$ such that θ belongs to a convex set.

The condition $G(\theta) \in \mathbb{S}^n_+$ can be relaxed in order to consider general non-square real matrices $G(\theta)$ (see e.g. (Dattorro, 2005§4)). However, for the sake of simplicity this is not considered in this paper.

2.1 Existing Results

Optimization problems subject to rank-constraints are difficult problems, since they are combinatorial in nature. However, there are some special cases, such as, unconstrained low-rank approximation problem, that admits a closed-form solution.

Theorem 1. (Eckart and Young, 1936) Given a matrix $X \in \mathbb{R}^{m \times n}$, then the solution of the following low-rank approximation problem

$$\hat{X}^* = \arg\min_Z \|X - Z\|_F \text{ s.t } \operatorname{rank} Z \leq r$$

is given by the truncated Singular Value Decomposition (SVD), i.e. if $X = USV^T$ is a SVD of X, then the

minimizer is given by $\hat{X}^* = U_{1:r}S_{1:r,1:r}V_{1:r}^T$ where $U_{1:r}$ denotes the first r columns of U, and $S_{1:r,1:r}$ denotes the submatrix composed by the first r rows and the first r columns of S, i.e. the submatrix containing the r-largest singular values of X. Moreover, the minimizer $\hat{X}^*$ is unique if and only if $\sigma_{r+1} \neq \sigma_r$.

Eckart-Young theorem has become the cornerstone of most of existing approaches. one of the main reasons for this is that the SVD can be efficiently computed in an numerically robust way. However, for problems such as structured low-rank approximation, the SVD-based methods can be seen as relaxation of the original rank-constrained problem (Markovsky, 2008).

An interesting approach to impose rank constraints has been proposed in (Dattorro, 2005; Kim and Moon, 2006; ten Berge and Kiers, 1991) which is related to the ideas presented in (d'Aspremont, 2003). The main idea is to minimize the sum of the smallest eigenvalues i.e. exploits the equivalence between imposing the constraint $\text{rank}(G) \leq r$ (with $G \in \mathbb{S}_+^n$) and imposing the constraint that the sum of the $n-r$ smallest eigenvalues of G is equal to zero.

In order to describe the idea, we require of the following definition

$$\Phi_{n,r} = \{W \in \mathbb{S}^n,\ 0 \preceq W \preceq I, \text{trace}(W) = n - r\} \quad (1)$$

This set corresponds to the convex hull of the rank-(n-r) projection matrices, i.e the smallest convex set that contain such set of matrices, see e.g. (Dattorro, 2005). The set $\Phi_{n,r}$ can be used to compute the sum of the (n-r) smallest eigenvalues of a matrix by solving an optimization problem, as follows

Lemma 2. Consider $G \in \mathbb{S}^n$ whose eigenvalues are $\lambda_1(G) \geq \cdots \geq \lambda_n(G)$, then

$$\mathcal{P}_1 : \qquad \sum_{i=r+1}^{n} \lambda_i(G) = \min_{W \in \Phi_{n,r}} \ \text{trace}(W^\top G)$$

Proof. Direct from (Overton and Womersley, 1993, Theorem 3.4) and by considering that $\sum_{i=r+1}^{n} \lambda_i(G) = \text{trace}(G) - \sum_{i=1}^{r} \lambda_i(G)$. □

Notice that Problem $\mathcal{P}_1$ has a closed-form solution. In fact, let us consider the diagonalisation $G^* = Q\Lambda Q^\top$, then the direction matrix $W = U^* U^{*\top}$ is optimal, where U^* correspond to the directions of Q corresponding to the $n-r$ smallest entries of the diagonal matrix Λ.

In some cases (for example, when the solution is not unique), it is preferable to numerically solve $\mathcal{P}_1$ rather than to use the closed-form solution.

In (Dattorro, 2005) Lemma 2 has been used to solve the following rank-constrained feasibility problem

$$\mathcal{P}_{feas} : \quad \begin{aligned} \underset{G \in \mathbb{S}_+^n}{\text{find}} \ & G \\ \text{subject to} \ & G \in \mathcal{C} \\ & \text{rank}\, G \leq r, \end{aligned}$$

where $\mathcal{C}$ is a convex set. This rank-constrained feasibility problem $\mathcal{P}_{feas}$ can be equivalently stated as the following minimization problem.

$$\mathcal{P}_2 : \quad \min_{G \in \mathbb{S}_+^n} \min_{W \in \Phi_{n,r}} \ \text{trace}(W^\top G) \quad \text{subject to} \ \ G \in \mathcal{C}$$

Problem $\mathcal{P}_2$ can be solved by iteratively alternating minimization between G and W. In more detail, given a current estimate of G, $\hat{G}^m$, at the iteration m. Then, the optimization update is as follows:

$$\hat{W}^{m+1} = \arg \left\{ \min_{W \in \Phi_{n,r}} \text{trace}(W^\top \hat{G}^m) \right\} \quad (2)$$

$$\hat{G}^{m+1} = \arg \left\{ \min_{G \in \mathbb{S}_+^n} \text{trace}((\hat{W}^{m+1})^\top G) \quad s.t\ G \in \mathcal{C} \right\} \quad (3)$$

The condition under which equivalence between $\mathcal{P}_{feas}$ and $\mathcal{P}_2$ holds, is given in the following result:

Lemma 3. The rank of a matrix $G \in \mathbb{S}_+^n$ is less that r, if and only if, there exists a $W \in \Phi_{n,r}$, such that

$$\text{trace}\, W^\top G = 0 \quad (4)$$

Proof. From Lemma 2 we have that

$$\sum_{i=r+1}^{n} \lambda_i(G) \leq \text{trace}(W^\top G) \quad (5)$$

and since $G \in \mathbb{S}_+^n$, we have that $\sum_{i=r+1}^{n} \lambda_i(G) \geq 0$. Then, (4) provide a sufficient condition for $\text{rank}(G) \leq r$. The necessity of (4) follows from the equality between the sum of the (n-r) smallest eigenvalues of G and the optimal value of the cost function in problem $\mathcal{P}_1$. □

Condition (4) indicates that feasibility problem $\mathcal{P}_{feas}$ has been solved. It remains an open problem to state conditions under which (4) is achieved by iteratively alternating minimization between G and W in $\mathcal{P}_2$. However, there exist some cases in which alternating minimization of G and W in problem $\mathcal{P}_2$ does not converge to the optimal solution as is shown in the following example.

Example 1. (Kim and Moon, 2006) Consider the following feasibility problem were we want to find x_1 and x_2 such that

$$G = \begin{bmatrix} x_1 & 0 \\ 0 & x_2 \end{bmatrix};\ x_2 \geq 1 \quad (6)$$

and the $\text{rank}\, G = 1$. Note that the solution of this feasibility problem is given by $x_1 = 0$ and $x_2 \geq 1$. If the initial values for x_1 and x_2 are such that $x_2 = 1$ and $x_1 > 1$ then, the smallest eigenvalue of G is 1 with corresponding eigenvector $[0\ 1]^\top$. The optimization problem becomes

$$\min x_2 \ \ s.t\ \ (6),$$

and its solution can be any $x_1 \geq 0$ and $x_2 = 1$. Thus x_1 is not guarantee to be zero and the alternating minimization between G and W does not converge to the correct result.

In order to overcome such cases (Kim and Moon, 2006) proposes the inclusion of an extra regularization, $\alpha\, \text{trace}(G)$, in the cost function on $\mathcal{P}_2$. Next, the regularization parameter α is gradually reduced to zero. This implies that the speed of convergence of the method is affected by the speed at which α tends to zero.

In our experience, only a small perturbation on the solution of the alternating minimization is required to overcome this special case. In fact, this special case is handled by solving $\mathcal{P}_1$ numerically instead of using the closed-form

Algorithm 1 Method 1 to rank-constrained optimization.

Require: $0 < \epsilon_0 \leq 0.5$
Require: $\theta^0 \in \Omega$ such that $\text{rank}(G(\theta^0)) \leq r$
  $m \leftarrow 0$
  while $\epsilon_m \geq$ Tolerance **do**
    Solve problem $\mathcal{P}_{inner}$.
    if in $\mathcal{P}_{inner}$ ($\text{trace}(W^\top G(\theta)) \leq Tolerance$) **then**
      Denote the solution as θ^{m+1}
      $\epsilon_{m+1} \leftarrow \epsilon_m$
    else
      $\theta^{m+1} \leftarrow \theta^m$
      $\epsilon_{m+1} \leftarrow 0.5\epsilon_m$
    end if
    $m \leftarrow m + 1$
  end while
  return $\theta = \theta^m$

solution. In fact, including a small perturbation such as the numerical errors helps to overcome the non-convergence of the alternating minimization between G and W. Minimizing the bilinear term $\text{trace}(W^\top G)$ is known to be difficult, since it is non-convex. However, there exist global optimization methods that can be used when a bilinear term is present in the objective function. See e.g. (Gorski et al., 2007).

It is worth mentioning that iterations of alternating minimization between G and W can be easily implemented using standard convex optimization software, such as CVX (Grant et al., 2011).

3. A SOLUTION TO RANK-CONSTRAINED OPTIMIZATION

In this section, we describe the proposed approach to solve the Rank-Constrained Optimization problem $\mathcal{P}_{rco}$. This problem can be solved by using an iterative procedure that in each iteration solves a feasibility problem. In more detail, let $\theta^m \in \Omega$ be the current value of $\theta \in \Omega$, at iteration m. Then, problem $\mathcal{P}_{rco}$ can be solved, by iteratively solving the following feasibility problem

$$\begin{aligned} \mathcal{P}_{inner}: \quad & \underset{\theta \in \mathbb{R}^p}{\text{find}} \quad \theta \\ & \text{subject to} \quad f(\theta) \leq f(\theta^m) \cdot (1 - \epsilon_m) \\ & \qquad \theta \in \Omega \\ & \qquad \text{rank}(G(\theta)) \leq r \\ & \qquad G(\theta) \in \mathbb{S}_+^n \end{aligned}$$

with $0 < \epsilon_m < 1$ is a parameter chosen by the user. If the feasibility problem $\mathcal{P}_{inner}$ cannot be solved, we compute $\epsilon_{m+1} = 0.5\epsilon_m$ and proceed with the algorithm, until problem $\mathcal{P}_{inner}$ cannot be solved for a sufficiently small ϵ_m.

In the method described above (Method 1), the speed of convergence is controlled by the value of $\epsilon_m > 0$. Note that, choosing an $\epsilon_m > 0$ in $\mathcal{P}_{inner}$ is a mechanism to reduce $f(\theta)$ by a non-negligible amount at each step. The procedure to problem $\mathcal{P}_{rco}$ is described in Algorithm 1.

Notice that the Method 1 can be extended to consider non-convex functions $f(\theta)$ in problem $\mathcal{P}_{rco}$ by using the general optimization framework of MM algorithms (Lange et al., 2000).

The proposed scheme could be computationally expensive in the sense that require iterations of $\mathcal{P}_{inner}$, that is also solved in an iterative procedure. In order to overcome this issue, a relaxed version of this algorithm is proposed in the next section.

3.1 Method 2: Relaxed Rank-Constrained Optimization

In this method (Method 2) a regularization parameter $\alpha > 0$ is added to manage the tradeoff between minimizing the sum of the smallest eigenvalues of $G(\theta)$ and the cost function $f(\theta)$. Thus, the relaxed problem is stated as follows

$$\begin{aligned} \mathcal{P}_{rrco}: \quad & \min_{\theta \in \mathbb{R}^p} \min_{W \in \Phi_{n,k}} \quad \text{trace}(W^\top G(\theta)) + \alpha f(\theta) \\ & \text{subject to} \quad \theta \in \Omega \\ & \qquad\qquad\quad G(\theta) \in \mathbb{S}_+^n \end{aligned}$$

This problem can be solved by using an iterative alternating minimization procedure, similar to (2)-(3).

4. APPLICATION EXAMPLE: FACTOR ANALYSIS

The methods described in the previous section are powerful tools that can be applied to a large range of optimization problems. In order to show the potential of them, we apply these methods to the problem of Factor Analysis in presence of correlated errors.

In Factor Analysis (FA) a collection of n random variables are measured. It is assumed that these variables can be decomposed in two parts: a common part describing the co-movement between the random variables, and an idiosyncratic part describing the individual movement of each random variable. The common part is modelled as a linear combination of r random variables, called *factors*.

Classical FA is based on the strict assumption that idiosyncratic movements must be uncorrelated, i.e. the idiosyncratic covariance must be diagonal (Chamberlain and Rotchschild, 1983). There has been a renewed interest in stating a more general problem that, among others things, relaxes the assumption of diagonal idiosyncratic covariance. One approach is the "approximate factor model" (Bai, 2003; Stock and Watson, 2002), where (by introducing the assumption that $n \to \infty$) "weak" correlation within the idiosyncratic covariance is allowed. An advantage of this approach is that it provides a theoretical framework for the use of standard FA tools when "weak" idiosyncratic correlation is present.

Our proposed method ensures that: (i) the number of factors is less than a pre-specified bound; (ii) the idiosyncratic covariance matrix is semidefinite positive; (iii) it allow us to relax relax the assumption of diagonal idiosyncratic covariance, by assuming, instead, that this matrix is sparse.

4.1 Problem formulation

Consider a measured output $y_k \in \mathbb{R}^n$, latent factors $f_k \in \mathbb{R}^r$, idiosyncratic noise $v_k \in \mathbb{R}^n$, and a model:

$$y_k = Af_k + v_k \tag{7}$$

where $A \in \mathbb{R}^{n \times r}$ is the matrix of factor loadings. We assume that f_k and v_k are mutually uncorrelated i.i.d.

zero-mean Gaussian processes, with covariances Γ and Ψ, respectively. Thus, the measured output y_k is an i.i.d. zero mean Gaussian process with covariance

$$\Sigma = A\Gamma A^\top + \Psi \tag{8}$$

There are two important issues in (classical) FA: (i) any rotation of the factors will produce the same output characteristics, (ii) the components of the idiosyncratic noise v_k must be mutually uncorrelated, i.e. Ψ must be a diagonal matrix.

Remark 4. The assumption that Ψ is diagonal is considered very restrictive (Stock and Watson, 2002). Approximate factor models relax this assumption by considering that $n \to \infty$, $T \to \infty$. Under this assumption, estimators (such as principal components and quasi maximum likelihood (Doz et al., 2012; Stock and Watson, 2002)) can be shown to be consistent in the presence of "weak" idiosyncratic cross-correlation. However, it has been pointed out that when n is small, the presence of cross-correlation could severally deteriorate the performance of such estimators (Boivin and Ng, 2006).

4.2 Existing Methods

In this section we briefly review the more relevant existing methods for Factor Analysis.

Principal Components Analysis (PCA) A *principal components* estimator minimizes the residual sum of squares $\sum_{k=1}^{T}(y_k - Af_k)^\top(y_k - Af_k)$ subject to that $A^\top A = I$. The Principal Component (PC) estimate $\hat{A}$ can be computed as the eigenvectors corresponding to the r-largest eigenvectors of $S = \sum_{t=1}^{T} y_k y_k^\top$. Then $\hat{f}_k = (\hat{A}^\top\hat{A})^{-1}A^\top y_k$. This allows one to obtain the PC estimates in a computationally efficient way. However, there is a drawback, for fixed n and when $T \to \infty$, the PC estimator is inconsistent unless $\Psi = \sigma^2 I$.

PC estimator can be applied to FA provided the variables are a-priori normalized, i.e. such that the variables have unitary covariance. In fact, for the special case that y_k is normally distributed, and $\Psi = \sigma^2 I$ the PC estimator is the maximum likelihood estimator. Also, it has been shown that when $n \to \infty$ and $T \to \infty$, the PC estimator is consistent even when there is "weak" cross-correlation in Ψ. For technical details see (Bai, 2003).

Robust PCA (RPCA) In RPCA, the focus is to discover the structure of Ψ. In order to achieve this, the cost function to be minimized is $\|\Psi\|_F + \alpha\|\Psi\|_1$, where the regularization term $\alpha\|\Psi\|_1$ allows Ψ to have a non-diagonal structure. Notice that, existing RPCA algorithms do not ensure $\Psi \succeq 0$. Thus, they are unsuitable for FA. Moreover, existing RPCA algorithms, such as (Lin et al., 2009), are based on the trace heuristic to induce the rank constraint on $A\Phi A^\top$. Then, the rank of the solution cannot be specified a priori.

Expectation-Maximization Expectation - Maximization (EM) algorithms (Dempster et al., 1977) have been successfully applied in several areas, such as, system Identification (see e.g. (Agüero et al., 2012; Delgado et al., 2012)), Channel Estimation (see e.g. (Carvajal et al., 2011)) and the computation of principal components estimates (see e.g. (Roweis, 1998)).

EM algorithm is a two-step iterative procedure designed to compute the maximum likelihood estimate. EM algorithms introduce the concept of complete data. The complete data is assumed to be composed of a set of measured variables, $\mathcal{Y}$, and also of a set of unmeasured variables, known as the *hidden variables*, $\mathcal{H}$.

In FA, the *factors* are a natural choice for hidden variables. An EM algorithm for FA can be described as follows: Given a current estimate of the parameters $(\hat{A}, \hat{\Psi})$ and setting $\Phi = I$, then the associated EM iteration is given by:

E-step Compute

$$\mu_{f_k|\mathcal{Y}} = \hat{A}^\top(\hat{A}\hat{A}^\top + \Psi)^{-1} y_k \tag{9}$$

$$\Sigma_{f_k|\mathcal{Y}} = I - \hat{A}^\top(\hat{A}\hat{A}^\top + \Psi)^{-1}\hat{A} \tag{10}$$

where $\mu_{f_k|\mathcal{Y}}$ and $\Sigma_{f_k|\mathcal{Y}}$ are the mean and covariance of the factors, conditioned on $\mathcal{Y}$ assuming the estimates $(\hat{A}, \hat{\Psi})$ are available.

M-step

$$\hat{A} = \left(\sum_{k=1}^{T} y_k \mu_{f_k|\mathcal{Y}}^\top\right)\left(\sum_{k=1}^{T} \Sigma_{f_k|\mathcal{Y}} + \mu_{f_k|\mathcal{Y}}\mu_{f_k|\mathcal{Y}}^\top\right)^{-1} \tag{11}$$

$$\hat{\Psi} = \text{diag}^*\{1/T \sum_{k=1}^{T}(y_k y_k^\top - y_k \mu_{f_k|\mathcal{Y}}^\top \hat{A}^\top - \hat{A}\mu_{f_k|\mathcal{Y}} y_k^\top + \hat{A}(\Sigma_{f_k|\mathcal{Y}} + \mu_{f_k|\mathcal{Y}}\mu_{f_k|\mathcal{Y}}^\top)\hat{A}^\top)\} \tag{12}$$

where the operator $\text{diag}^*\{M\}$ returns a diagonal matrix with values equal to the values on the diagonal of M (i.e. if $L = \text{diag}^*\{M\}$, thus $L_{ii} = M_{ii}$ and $L_{ij} = 0$ for $i \neq j$).

It is well known that EM algorithms converge to a stationary point of the likelihood function (Dempster et al., 1977).

Minimum Rank Factor Analysis Minimum Rank Factor Analysis (MRFA) was originally proposed in (ten Berge and Kiers, 1991). The approach minimizes the sum of the smallest $n - r$ eigenvalues of $\Sigma - \Psi$. This can be interpreted as minimizing the unexplained co-movement. An advantage of MRFA is that includes the constraints that $\Sigma - \Psi \succeq 0$ and $\Psi \succeq 0$. MRFA can be formulated as follows:

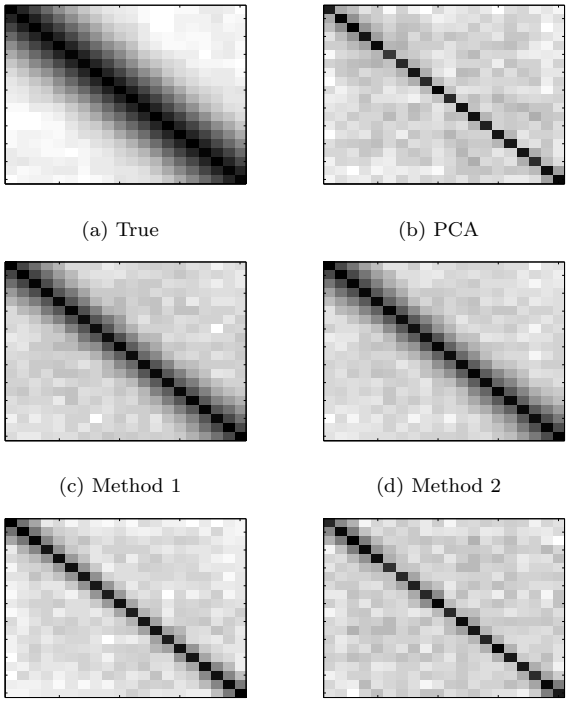
$$\mathcal{P}_{MRFA}: \quad \underset{\Psi}{\text{minimize}} \;\; h(\Psi) = \sum_{i=r+1}^{n} \lambda_i(\Sigma - \Psi)$$
$$\text{subject to} \;\; \Sigma - \Psi \succeq 0$$
$$\Psi \succeq 0$$
$$\Psi \text{ diagonal}$$

The minimum of $h(\Psi)$ coincides with the minimum of the function

$$g(\Psi, W) = \text{trace}(W^\top(\Sigma - \Psi)) \tag{13}$$

where $W = XX^\top$ and X is an $N \times (n-r)$ columnwise orthonormal matrix (ten Berge and Kiers, 1991). Then (13) can be monotonically minimized by alternating between optimizing Ψ and W (Shapiro and ten Berge, 2002; ten Berge and Kiers, 1991).

Fig. 1. Mean value over $N_{mc} = 100$ Monte Carlo simulations on the magnitud of each entry of $\bar{\Psi}$ for each method



(a) True (b) PCA

(c) Method 1 (d) Method 2

(e) EM (f) RPCA

4.3 Proposed approach to Factor Analysis with Correlated Errors

In Classical FA, the idiosyncratic covariance matrix Ψ must be diagonal. This is considered very restrictive (Doz et al., 2012). An alternative weaker assumption is to impose a sparsity assumption, such that, most of the entries of Ψ are zeros.

In this paper we propose to drop the assumption of diagonal Ψ, by assuming instead that Ψ is "sparse". This problem can be solved by using the methods developed in Section 3 by using $f(\Psi) = \|\Psi\|_1$, i.e. by solving the following optimization problem

$$\begin{aligned} \mathcal{P}_3 : \qquad & \min_{\Psi} \ \|\Psi\|_1 \\ & \text{subject to } \ \Sigma - \Psi \succeq 0 \\ & \qquad\qquad\ \ \Psi \succeq 0 \\ & \qquad\qquad\ \ \text{rank}(\Sigma - \Psi) \leq r \end{aligned}$$

Notice that, the proposed approach ensures both, that $\Psi \succeq 0$, and $\Sigma - \Psi \succeq 0$.

Remark 5. A variant of problem $\mathcal{P}_3$, could be achieved by changing the cost function to $f(\Psi) = \|\Psi\|_F + \beta\|\Psi\|_1$. Notice it is possible to use different cost functions with minimal changes in the code.

5. EXAMPLES

Consider model (7)-(8), where $r = 3$, $n = 20$, $T = 100$, $\Phi = I$ and the matrix of factor loadings, A, is constructed as a random matrix, in which each matrix entry is independent and normal distributed, i.e $A_{ij} \sim N(0, 1)$. The covariance matrix Ψ is constructed such that the matrix entries $\Psi_{ij} = \tau^{|i-j|}$ with $\tau = 0.7$. In order to adjust the signal to noise ratio we fix the ratio $\|AA^\top\|_F / \|\Psi\|_F = 2$.

Table 1. Mean value over $N_{mc} = 100$ Monte Carlo simulations of the performance index $d_c(\cdot)$, of $\|\hat{\breve{\Psi}}\|_F$ and the execution time (Self-time and time to get initial estimate plus self-time).

Method	$d_c(\cdot)$	$\|\breve{\Psi}\|_F$	Self-Time [s]	Total Time [s]
PCA	0.1992	23.7653	0.02055	0.02055
Method 1	0.0996	17.6631	170.1203	194.6211
Method 2	0.1001	17.5749	24.4803	24.5008
EM	0.1330	16.6911	0.0093	0.0298
RPCA	0.2260	25.4499	0.3964	0.3964

We run $N_{mc} = 100$ Monte Carlo simulations, with different realizations of f_k and v_k, and different factor loadings, A.

We solve problem $\mathcal{P}_3$ using both proposed methods. We compare the proposed methods against the PCA, the EM algorithm described in section 4.2.3, and a RPCA method (Lin et al., 2009). In RPCA we choose a regularization parameter such that the rank constraint is satisfied. In Method 2 we chose the regularization parameters as $\alpha = 1/\sqrt{r}$.

We use the performance index

$$d_c(P_m) = 1 - tr\{AP_mA^\top\}/tr\{AA^\top\},$$

where P_m corresponding to the orthogonal projection of the $\hat{A}$ for the method m. We also consider the covariance residual

$$\breve{\Psi} = \hat{\Sigma} - \hat{A}\hat{A}^\top,$$

where $\hat{\Sigma} = \frac{1}{T}\sum_{k=1}^{T} y_k y_k^\top$. Notice, that that $\breve{\Psi}$ may differ from the estimate of Ψ provided by the method, since $\breve{\Psi}$ is not constrained to have an specific structure.

Table 1 shows the mean values over the Monte Carlo simulations of the execution time, $d_c(P_m)$, and $\|\breve{\Psi}\|_F$. The execution time is decomposed in self-time, and total time, where the total time includes the execution time required to get the initial estimate plus the self-time. PCA and RPCA do not require an initial estimate. The EM algorithm and Method 2, where initialized with the PCA estimate, and Method 1 it is initialized with Method 2 estimate.

In Table 1 we notice that Method 1 provides the best estimate of the subspace of A. Method 2 also provides a good estimate of the subspace of A. Notice that EM achieves in average the lowest value of $\|\breve{\Psi}\|_F$. However, EM algorithm doesn't provides the best estimate of the subspace of A. This is possible due to that Ψ is constrained to have a diagonal structure.

As expected, Table 1 shows that Method 1 and Method 2 are computationally expensive compare to others methods.

Figure 1 shows the mean magnitude of $\breve{\Psi}$ for all methods over the Monte Carlo simulations, denoted by $\bar{\Psi} = 1/N_{mc}\sum_{i=1}^{N_{mc}} |\breve{\Psi}_i|$, where in here $|\cdot|$ denotes the entry-wise absolute value. The magnitudes are shown on a logarithmic scale. Figure 1b shows $\bar{\Psi}$ for PCA, which produces a $\bar{\Psi}$ matrix which is consistent with the assumption of diagonal Ψ. Figures 1c-1d show the resultant $\bar{\Psi}$ for the proposed

methods. We see that non-diagonal entries appear more frequently than in other methods. Figure 1e shows $\bar{\Psi}$ of EM, that seems to be consistent with the assumption that matrix Ψ is diagonal. Figure 1f shows $\bar{\Psi}$ of RPCA, that on average "discovers" a $\bar{\Psi}$ diagonal, but in average does not properly recover the non-diagonal entries.

6. CONCLUSION

We have presented a general approach to rank constrained optimization. The approach can be applied to several areas, such as, system identification, control, and signal processing. The method is applied to Factor Analysis. In this framework, the approach provides sufficient flexibility to handle non-diagonal idiosyncratic covariance. The numerical example has shown the efficacy of the approach. In many cases it outperforms existing methods.

REFERENCES

J.C. Agüero, W. Tang, J.I. Yuz, R.A. Delgado, and G.C. Goodwin. Dual time–frequency domain system identification. *Automatica*, 48(12):3031 – 3041, 2012.

J. Bai. Inferential theory of factor models of large dimensions. *Econometrica*, 71:135–172, 2003.

J. Boivin and S. Ng. Are more data always better for factor analysis? *Journal of Econometrics*, 132(1):169–194, 2006.

R. Carvajal, J. C. Agüero, B. I. Godoy, and G. C. Goodwin. EM-based channel estimation in OFDM systems with phase noise. In *Global Telecommunications Conference (GLOBECOM 2011)*, pages 1–5, 2011.

G. Chamberlain and M. Rotchschild. Arbitrage factor structure, and mean-variance analysis of large asset markets. *Econometrika*, 51(1281–1304), 1983.

Alexandre d'Aspremont. A semidefinite representation for some minimum cardinality problems. In *42nd IEEE Conference on Decision and Control, 2003*, volume 5, pages 4985–4990, 2003.

J Dattorro. *Convex optimization and Euclidean distance geometry*. Meboo Publishing, USA, 2005.

R. A. Delgado, G. C. Goodwin, R. Carvajal, and J. C. Agüero. A novel approach to model error modelling using the expectation-maximization algorithm. In *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pages 7327–7332. IEEE, 2012.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from imcomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1): 1–38, 1977.

C. Doz, D. Giannone, and L. Reichlin. A quasi–maximum likelihood approach for large approximate dynamic factor models. *Review of economics and statistics*, 94(4): 1014–1024, 2012.

G. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1:211–218, 1936.

M. Fazel. A rank minimization heuristic with application to minimum order system approximation. In *Proceedings of the American Control Conference*, pages 4734–4739. IEEE, 2001.

M. Fazel. *Matrix Rank Minimization with Applications*. PhD thesis, Standford University, Standford, CA, 2002.

Jochen Gorski, Frank Pfeuffer, and Kathrin Klamroth. Biconvex sets and optimization with biconvex functions: a survey and extensions. *Mathematical Methods of Operations Research*, 66(3):373–407, 2007.

M. C. Grant, S. P. Boyd, and Y. Ye. CVX: Matlab software for disciplined convex programming (web page and software) v2.0. *Available at http://cvxr.com/cvx*, April 2011.

C. Grossmann, C. Jones, and M. Morari. System identification with missing data via nuclear norm regularization. In *Presented at: European Control Conference*, volume 23, page 26, 2009.

S. J. Kim and Y. H. Moon. Structurally constrained H_2 and H_∞ control: A rank-constrained LMI approach. *Automatica*, 42(9):1583 – 1588, 2006.

K. Lange, D. R. Hunter, and I. Yang. Optimization transfer using surrogate objective functions. *Journal of Computational and Graphical Statistics*, 9(1):1–20, 2000.

K. Lee and Y. Bresler. ADMiRA: Atomic decomposition for minimum rank approximation. *Information Theory, IEEE Transactions on*, 56(9):4402–4416, 2010.

Z. Lin, A. Ganesh, J. Wright, L. Wu, M. Chen, and Y. Ma. Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix. *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 61, 2009.

I. Markovsky. Structured low-rank approximation and its applications. *Automatica*, 44(4):891–909, 2008.

I. Markovsky. *Low Rank Approximation: Algorithms, Implementation, Applications*. Communications and Control Engineering. Springer, 2012. doi: 10.1007/978-1-4471-2227-2.

I. Markovsky. Recent progress on variable projection methods for structured low-rank approximation. *Signal Processing*, 96PB:406–419, 2014.

G. Meyer. *Geometric optimization algorithms for linear regression on fixed-rank matrices*. PhD thesis, University of Liège, Belgium, 2011.

R. Orsi, U. Helmke, and J. B. Moore. A newton-like method for solving rank constrained linear matrix inequalities. *Automatica*, 42(11):1875 – 1882, 2006.

M. L. Overton and R S. Womersley. Optimality conditions and duality theory for minimizing sums of the largest eigenvalues of symmetric matrices. *Mathematical Programming*, 62(1–3):321–357, 1993.

S. Roweis. EM algorithms for PCA and SPCA. *Advances in neural information processing systems*, pages 626–632, 1998.

S. Shalev-Shwartz, A. Gonen, and O. Shamir. Large-scale convex minimization with a low-rank constraint. *arXiv preprint arXiv:1106.1622*, 2011.

A. Shapiro and J. M. F. ten Berge. Statistical inference of minimum rank factor analysis. *Psychometrika*, 67(1): 79–84, 2002.

J. H. Stock and M. W. Watson. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97:1167–1179, 2002.

J. M. ten Berge and H. A. L. Kiers. A numerical approach to approximate and exact minimum rank of a covariance matrix. *Psychometrika*, 56(2):309–315, 1991.

B. Vandereycken. *Riemannian and multilevel optimization for rank-constrained matrix problems*. PhD thesis, Department of Computer Science, KU Leuven, 2010.