Convex Optimization Methods for Dimension Reduction and Coefficient Estimation in Multivariate Linear Regression

Zhaosong Lu* Renato D. C. Monteiro† Ming Yuan‡

January 10, 2008

Abstract

In this paper, we study convex optimization methods for computing the trace norm regularized least squares estimate in multivariate linear regression. The so-called factor estimation and selection (FES) method, recently proposed by Yuan et al. [17], conducts parameter estimation and factor selection simultaneously and have been shown to enjoy nice properties in both large and finite samples. To compute the estimates, however, can be very challenging in practice because of the high dimensionality and the trace norm constraint. In this paper, we explore Nesterov's first-order methods [12, 13] and interior point methods for computing the penalized least squares estimate. The performance of these methods is then compared using a set of randomly generated instances. We show that the best of Nesterov's first-order methods substantially outperforms the interior point method implemented in SDPT3 version 4.0 (beta) [15]. Moreover, the former method is much more memory efficient.

Key words: Cone programming, smooth saddle point problem, first-order method, interior point method, multivariate linear regression, trace norm, dimension reduction.

AMS 2000 subject classification: 90C22, 90C25, 90C47, 65K05, 62H12, 62J05

1 Introduction

Multivariate linear regression is routinely used in statistics to model the predictive relationships of multiple related responses on a common set of predictors. In general multivariate linear regression, we have n observations on q responses $\mathbf{b} = (b_1, \ldots, b_q)'$ and p explanatory variables $\mathbf{a} = (a_1, \ldots, a_p)'$, and

$$B = AU + E, \tag{1}$$

*Department of Mathematics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada (Email: `zhaosong@sfu.ca`).
This author was supported in part by SFU President's Research Grant and NSERC Discovery Grant.

†School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, USA (Email: `monteiro@isye.gatech.edu`).
This author was supported in part by NSF Grant CCF-0430644 and ONR grant N00014-05-1-0183.

‡School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, USA (Email: `myuan@isye.gatech.edu`).
This author was supported in part by NSF Grants DMS-0624841 and DMS-0706724.

where $B = (\mathbf{b}^1, \ldots, \mathbf{b}^n)' \in \Re^{n\times q}$ and $A = (\mathbf{a}^1, \ldots, \mathbf{a}^n)' \in \Re^{n\times p}$ consists of the data of responses and explanatory variables, respectively, $U \in \Re^{p\times q}$ is the coefficient matrix, $E = (\mathbf{e}^1, \ldots, \mathbf{e}^n)' \in \Re^{n\times q}$ is the regression noise, and all $\mathbf{e}^i$s are independently sampled from $\mathcal{N}(0, \Sigma)$.

Classical estimators for the coefficient matrix U such as the least squares estimate are known to perform sub-optimally because they do not utilize the information that the responses are related. This problem is exacerbated when the dimensionality p or q is moderate or large. Linear factor models are widely used to overcome this problem. In the linear factor model, the response B is regressed against a small number of linearly transformed explanatory variables, which are often referred to as factors. More specifically, the linear factor model can be expressed as

$$B = F\Omega + E, \tag{2}$$

where $\Omega \in \Re^{r\times q}$, and $F = A\Gamma$ for some $\Gamma \in \Re^{p\times r}$ and $r \le \min\{p, q\}$. The columns of F, namely, F_j $(j = 1, \ldots, r)$ represent the so-called factors. Clearly (2) is an alternative representation of (1) with $U = \Gamma\Omega$, and the dimension of the estimation problem reduces as r decreases. Many popular methods including canonical correction (Hotelling [7, 8]), reduced rank (Anderson [1], Izenman [9], Reinsel and Velu [14]), principal components (Massy [11]), partial least squares (Wold [16]) and joint continuum regression (Brooks and Stone [5]) among others can all be formulated in the form of linear factor regression. They differ in the way in which the factors are determined.

Given the number of factors r, estimation in the linear factor model most often proceeds in two steps: the factors, or equivalently Γ, are first constructed, and then Ω is estimated by least squares for (2). It is obviously of great importance to be able to determine r for (2). For a smaller number of factors, a more accurate estimate is expected since there are fewer free parameters. But too few factors may not be sufficient to describe the predictive relationships. In all of the aforementioned methods, the number of factors r is chosen in a separate step from the estimation of (2) through either hypothesis testing or cross-validation. The coefficient matrix is typically estimated on the basis of the number of factors selected. Due to its discrete nature, this type of procedure can be very unstable in the sense of Breiman [4]: small changes in the data can result in very different estimates.

Recently, Yuan et al. [17] proposed a novel method that can simultaneously choose the number of factors, determine the factors and estimate the factor loading matrix Ω. It has been demonstrated that the so-called factor estimation and selection (FES) method combines and retains the advantages of the existing methods. FES is a constrained least square estimate where the trace norm (or the Ky Fan m-norm where $m = \min\{p, q\}$) of the coefficient matrix U is forced to be smaller than an upper bound:

$$\begin{array}{ll} \min\limits_{U} & \mathrm{Tr}((B - AU)W(B - AU)') \\ \text{s.t.} & \sum\limits_{i=1}^{m} \sigma_i(U) \le M. \end{array} \tag{3}$$

where W is a positive definite weight matrix. Common choices of the weight matrix W include Σ^{-1} and I. To fix ideas, we assume throughout the paper that $W = I$. Under this assumption, (3) is equivalent to

$$\begin{array}{ll} \min\limits_{U} & \|B - AU\|_F^2 \\ \text{s.t.} & \sum\limits_{i=1}^{m} \sigma_i(U) \le M. \end{array} \tag{4}$$

It is shown in Yuan et al. [17] that the constraint used by FES encourages sparsity in the factor space and at the same time gives shrinkage coefficient estimates and thus conducts dimension reduction and estimation simultaneously in the multivariate linear model.

The goal of this paper is to explore convex optimization methods, namely, Nesterov's first-order methods [12, 13], and interior point methods for solving (4). We also compare the performance of these methods on a set of randomly generated instances. We show that the best of Nesterov's first-order methods substantially outperforms the interior point method implemented in the code SDPT3 version 4.0 (beta) [15], and that the former method requires much less memory than the latter one.

This paper is organized as follows. In Subsection 1.1, we introduce the notation that is used throughout the paper. In Section 2, we present some technical results that are used in our presentation. In Section 3, we provide a simplification for problem (4), and present cone programming and smooth saddle point reformulations for it. In Section 4, we review two convex minimization methods due to Nesterov [12, 13] and discuss the details of their implementation for solving the aforementioned smooth saddle point reformulations of (4). In Section 5, we present computational results comparing a well-known second-order interior-point method applied to the aforementioned cone programming reformulation of (4) with Nesterov's first-order approaches for solving the different saddle point reformulations of (4). Finally, we present some concluding remarks in Section **??**.

1.1 Notation

The following notation is used throughout our paper. For any real number α, $[\alpha]^+$ denotes the nonnegative part of α, that is, $[\alpha]^+ = \max\{\alpha, 0\}$. The symbol $\Re^p$ denotes the p-dimensional Euclidean space. We denote by e the vector of all ones whose dimension should be clear from the context. For any $w \in \Re^p$, $\mathrm{Diag}(w)$ denotes the $p \times p$ diagonal matrix whose ith diagonal element is w_i for $i = 1, \ldots, p$. The Euclidean norm in $\Re^p$ is denoted by $\|\cdot\|$.

We let $\mathcal{S}^n$ denote the space of $n \times n$ symmetric matrices, and $Z \succeq 0$ indicate that Z is positive semidefinite. We also write $\mathcal{S}^n_+$ for $\{Z \in \mathcal{S}^n : Z \succeq 0\}$, and $\mathcal{S}^n_{++}$ for its interior, the set of positive definite matrices in $\mathcal{S}^n$. For any $Z \in \mathcal{S}^n$, we let $\lambda_i(Z)$, for $i = 1, ..., n$, denote the ith largest eigenvalue of Z, $\lambda_{\min}(Z)$ (resp., $\lambda_{\max}(Z)$) denote the minimal (resp., maximal) eigenvalue of Z, and define $\|Z\|_\infty := \max_{1 \le i \le n} |\lambda_i(Z)|$ and $\|Z\|_1 = \sum_{i=1}^n |\lambda_i(Z)|$. Either the identity matrix or operator will be denoted by I.

The space of all $p \times q$ matrices with real entries is denoted by $\Re^{p\times q}$. Given matrices X and Y in $\Re^{p\times q}$, the standard inner product is defined by $X \bullet Y = \mathrm{Tr}(X^T Y)$, where $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix. The operator norm and the Frobenius norm of a $p \times q$-matrix X are defined as $\|X\| := \max\{\|Xu\| : \|u\| \le 1\} = [\lambda_{\max}(X^T X)]^{1/2}$ and $\|X\|_F := \sqrt{X \bullet X}$, respectively. Given any $X \in \Re^{p\times q}$, we let $\mathrm{vec}(X)$ denote the vector in $\Re^{pq}$ obtained by stacking the columns of X according to the order in which they appear in X, and $\sigma_i(X)$ denote the ith largest singular value of X for $i = 1, \ldots, \min\{p, q\}$. (Recall that $\sigma_i(X) = [\lambda_i(X^T X)]^{1/2} = [\lambda_i(XX^T)]^{1/2}$ for $i = 1, \ldots, \min\{p, q\}$.) Also, let $\mathcal{G} : \Re^{p\times q} \to \Re^{(p+q)\times(p+q)}$ be defined as

$$\mathcal{G}(X) := \begin{pmatrix} 0 & X^T \\ X & 0 \end{pmatrix}, \ \forall X \in \Re^{p\times q}. \tag{5}$$

The following sets are used throughout the paper:

$$\begin{aligned} \mathcal{B}_F^{p\times q}(r) &:= \{X \in \Re^{p\times q} : \|X\|_F \le r\}, \\ \Delta^n_=(r) &:= \{Z \in \mathcal{S}^n : \|Z\|_1 = r, \, Z \succeq 0\}, \\ \Delta^n_\le(r) &:= \{Z \in \mathcal{S}^n : \|Z\|_1 \le r, \, Z \succeq 0\}, \\ \mathcal{L}^p &:= \left\{x \in \Re^p : x_1 \ge \sqrt{x_2^2 + \ldots + x_p^2}\right\}, \end{aligned}$$

where the latter is the well-known p-dimensional second-order cone.

Let $\mathcal{U}$ be a normed vector space whose norm is denoted by $\|\cdot\|_{\mathcal{U}}$. The dual space of $\mathcal{U}$, denoted by $\mathcal{U}^*$, is the normed vector space consisting of all linear functionals of $u^* : \mathcal{U} \to \Re$, endowed with the dual norm $\|\cdot\|^*_{\mathcal{U}}$ defined as

$$\|u^*\|^*_{\mathcal{U}} = \max_u \{\langle u^*, u\rangle : \ \|u\|_{\mathcal{U}} \le 1\}, \quad \forall u^* \in \mathcal{U}^*,$$

where $\langle u^*, u\rangle := u^*(u)$ is the value of the linear functional u^* at u.

If $\mathcal{V}$ denotes another normed vector space with norm $\|\cdot\|_{\mathcal{V}}$, and $\mathcal{E} : \mathcal{U} \to \mathcal{V}^*$ is a linear operator, the operator norm of $\mathcal{E}$ is defined as

$$\|\mathcal{E}\|_{\mathcal{U},\mathcal{V}} = \max_u \{\|\mathcal{E}u\|^*_{\mathcal{V}} : \|u\|_{\mathcal{U}} \le 1\}. \tag{6}$$

A function $f : \Omega \subseteq U \to \Re$ is said to be L-Lipschitz-differentiable with respect to $\|\cdot\|_{\mathcal{U}}$ if it is differentiable and

$$\|\nabla f(u) - \nabla f(\tilde{u})\|^*_{\mathcal{U}} \le L\|u - \tilde{u}\|_{\mathcal{U}} \quad \forall u, \tilde{u} \in \Omega. \tag{7}$$

2 Technical results

In this section, we provide some technical results that are used in our presentation. In particular, we present some unconstrained reformulations of a constrained nonlinear programming problem in Subsection 2.1 and establish some results about eigenvalues and singular values in Subsection 2.2.

2.1 Unconstrained reformulation of constrained problems

In this subsection, we discuss two ways of solving a constrained nonlinear programming problem based on some unconstrained nonlinear programming reformulations.

Given a set $\emptyset \ne X \subseteq \Re^n$ and functions $f : X \to \Re$ and $h : X \to \Re^k$, consider the nonlinear programming problem:

$$f^* = \inf \{f(x) : x \in X,\, h_i(x) \le 0,\, i = 1, \ldots, k\}. \tag{8}$$

The first reformulation of (8) is based on the exact penalty approach, which consists of solving the exact penalization problem

$$f_\gamma{}^* = \inf \{f_\gamma(x) := f(x) + \gamma[g(x)]^+ : x \in X\}, \tag{9}$$

for some large penalty parameter $\gamma > 0$, where $g(x) = \max\{h_i(x) : i = 1, \ldots, k\}$. To obtain stronger consequences, we make the following assumptions about problem (8):

A.1) The set X is convex and functions f and h_i are convex for each $i = 1, \ldots, k$;

A.2) $f^* \in \Re$ and there exists a point $x^0 \in X$ such that $g(x^0) < 0$.

We will use the following notion throughout the paper.

Definition 1 Consider the problem of minimizing a real-valued function $f(x)$ over a certain nonempty feasible region $\mathcal{F}$ contained in the domain of f and let $\bar{f} := \inf\{f(x) : x \in \mathcal{F}\}$. For $\epsilon \ge 0$, we say that x_ϵ is an **ϵ-optimal solution** of this problem if $x_\epsilon \in \mathcal{F}$ and $f(x_\epsilon) \le \epsilon + \bar{f}$.

We note that the existence of an ϵ-optimal solution for some $\epsilon > 0$ implies that $\bar{f}$ is finite.

Theorem 2.1 *Suppose Assumptions A.1 and A.2 hold and define*

$$\bar{\gamma} := \frac{f(x^0) - f^*}{|g(x^0)|} \geq 0.$$

For $x \in X$, define

$$z(x) := \frac{x + \theta(x)x^0}{1 + \theta(x)}, \qquad \textit{where} \;\; \theta(x) := \frac{[g(x)]^+}{|g(x^0)|}. \tag{10}$$

Then, the following statements hold:

a) for every $x \in X$, the point $z(x)$ is a feasible solution of (8);

b) ${f_\gamma}^* = f^*$ for every $\gamma \geq \bar{\gamma}$;

c) for every $\gamma \geq \bar{\gamma}$ and $\epsilon \geq 0$, any ϵ-optimal solution of (8) is also an ϵ-optimal solution of (9);

d) if $\gamma \geq \bar{\gamma}$, $\epsilon \geq 0$ and x_ϵ^γ is an ϵ-optimal solution of (9), then the point $z(x_\epsilon^\gamma)$ is an ϵ-optimal solution of (8).

e) if $\gamma > \bar{\gamma}$, $\epsilon \geq 0$ and x_ϵ^γ is an ϵ-optimal solution of (9), then $f(x_\epsilon^\gamma) - f^* \leq \epsilon$ and $[g(x_\epsilon^\gamma)]^+ \leq \epsilon/(\gamma - \bar{\gamma})$.

Proof. Let $x \in X$ be arbitrarily given. Clearly, convexity of X, the assumption that $x^0 \in X$ and the definition of $z(x)$ imply that $z(x) \in X$. Moreover, Assumption A.1 implies that $g : X \to \Re$ is convex. This fact, the assumption that $g(x^0) < 0$, and the definitions of $z(x)$ and $\theta(x)$ then imply that

$$g(z(x)) \leq \frac{g(x) + \theta(x)g(x^0)}{1 + \theta(x)} \leq \frac{[g(x)]^+ - \theta(x)|g(x^0)|}{1 + \theta(x)} = 0.$$

Hence, statement (a) follows.

To prove statement (b), assume that $\gamma \geq \bar{\gamma}$ and let $x \in X$ be given. Convexity of f yields $(1 + \theta(x))f(z(x)) \leq f(x) + \theta(x)f(x^0)$, which, together with the definitions of $\bar{\gamma}$ and $\theta(x)$, imply that

$$\begin{aligned} f_\gamma(x) - f^* &= f(x) + \gamma[g(x)]^+ - f^* \\ &\geq (1 + \theta(x))f(z(x)) - \theta(x)f(x^0) + \gamma[g(x)]^+ - f^* \\ &= (1 + \theta(x))(f(z(x)) - f^*) - \theta(x)(f(x^0) - f^*) + \gamma[g(x)]^+ \\ &= (1 + \theta(x))(f(z(x)) - f^*) + (\gamma - \bar{\gamma})[g(x)]^+. \end{aligned} \tag{11}$$

In view of the assumption that $\gamma \geq \bar{\gamma}$ and statement (a), the above inequality implies that $f_\gamma(x) - f^* \geq 0$ for every $x \in X$, and hence that ${f_\gamma}^* \geq f^*$. Since the inequality ${f_\gamma}^* \leq f^*$ obviously holds for any $\gamma \geq 0$, we then conclude that ${f_\gamma}^* = f^*$ for any $\gamma \geq \bar{\gamma}$. Statement (c) follows as an immediate consequence of (b).

For some $\gamma \geq \bar{\gamma}$ and $\epsilon \geq 0$, assume now that x_ϵ^γ is an ϵ-optimal solution of (9). Then, statement (b) and inequality (11) imply that

$$\epsilon \geq f_\gamma(x_\epsilon^\gamma) - {f_\gamma}^* \geq (1 + \theta(x_\epsilon^\gamma))(f(z(x_\epsilon^\gamma)) - f^*) + (\gamma - \bar{\gamma})[g(x_\epsilon^\gamma)]^+. \tag{12}$$

Using the assumption that $\gamma \geq \bar{\gamma}$, the above inequality clearly implies that $f(z(x_\epsilon^\gamma)) - f^* \leq \epsilon/(1 + \theta(x_\epsilon^\gamma)) \leq \epsilon$, and hence that $z(x_\epsilon^\gamma)$ is an ϵ-optimal solution of (8) in view of statement (a). Hence, statement (d) follows. Moreover, if $\gamma > \bar{\gamma}$, we also conclude from (12) that $[g(x_\epsilon^\gamma)]^+ \leq \epsilon/(\gamma - \bar{\gamma})$. Also, the first inequality of (12) implies that $f(x_\epsilon^\gamma) - f^* \leq f(x_\epsilon^\gamma) + \gamma[g(x_\epsilon^\gamma)]^+ - f^* = f_\gamma(x_\epsilon^\gamma) - {f_\gamma}^* \leq \epsilon$, showing that statement (e) holds. ■

We observe that the threshold value $\bar{\gamma}$ depends on the optimal value f^*, and hence can be computed only for those problems in which f^* is known. If instead a lower bound $f_l \leq f^*$ is known, then choosing the penalty parameter γ in problem (9) as $\gamma := (f(x^0) - f_l)/|g(x^0)|$ guarantees that an ϵ-optimal solution x_ϵ^γ of (9) yields the ϵ-optimal solution $z(x_\epsilon^\gamma)$ of (8), in view of Theorem 2.1(c).

The following result, which is a slight variation of a result due to H. Everett (see for example pages 147 and 163 of [6]), shows that approximate optimal solutions of Lagrangian subproblems associated with (8) yield approximate optimal solutions of a perturbed version of (8).

Theorem 2.2 *(Approximate Everett's theorem) Suppose that for some $\lambda \in \Re^k_+$ and $\epsilon \geq 0$, x_ϵ^λ is an ϵ-optimal solution of the problem*

$$f_\lambda^* = \inf \left\{ f(x) + \sum_{i=1}^k \lambda_i h_i(x) : x \in X \right\}. \tag{13}$$

Then, x_ϵ^λ is an ϵ-optimal solution of the problem

$$f_{\epsilon\lambda}^* = \inf \left\{ f(x) : x \in X,\, h_i(x) \leq h_i(x_\epsilon^\lambda),\, i = 1, \ldots, k \right\}. \tag{14}$$

Proof. Let $\tilde{x}$ be a feasible solution of (14). Since x_ϵ^λ is an ϵ-optimal solution of (13), we have $f(x_\epsilon^\lambda) + \sum_{i=1}^k \lambda_i h_i(x_\epsilon^\lambda) \leq f_\lambda^* + \epsilon$. This inequality together with the definition of f_λ^* in (13) implies that

$$f(x_\epsilon^\lambda) \;\leq\; f_\lambda^* - \sum_{i=1}^k \lambda_i h_i(x_\epsilon^\lambda) + \epsilon \;\leq\; f(\tilde{x}) + \sum_{i=1}^k \lambda_i [h_i(\tilde{x}) - h_i(x_\epsilon^\lambda)] + \epsilon \;\leq\; f(\tilde{x}) + \epsilon,$$

where the last inequality is due to the fact that $\lambda_i \geq 0$ for all $i = 1, \ldots, k$ and $\tilde{x}$ is feasible solution of (14). Since the latter inequality holds for every feasible solution $\tilde{x}$ of (14), we conclude that $f(x_\epsilon^\lambda) \leq f_{\epsilon\lambda}^* + \epsilon$, and hence that x_ϵ^λ is an ϵ-optimal solution of (14). ■

If our goal is to solve problem $\inf\{f(x) : x \in X,\, h_i(x) \leq b_i,\, i = 1, \ldots, k\}$ for many different right hand sides $b \in \Re^k$, then, in view of the above result, this goal can be accomplished by minimizing the Lagrangian subproblem (13) for many different Lagrange multipliers $\lambda \in \Re^k_+$. We note that this idea is specially popular in statistics for the case when $k = 1$.

2.2 Some results on eigenvalues and singular values

In this subsection, we establish some technical results on eigenvalues and singular values that will be subsequently used.

The first result gives some well-known identities involving the maximum eigenvalue of a real symmetric matrix.

Lemma 2.3 *For any $Z \in \mathcal{S}^n$ and scalars $\alpha > 0$ and $\beta \in \Re$, the following statements hold:*

$$\lambda_{\max}(Z) = \max_{W \in \Delta^n_=(1)} Z \bullet W, \tag{15}$$

$$[\alpha \lambda_{\max}(Z) + \beta]^+ = \max_{W \in \Delta^n_\le(1)} \alpha Z \bullet W + \beta \mathrm{Tr}(W). \tag{16}$$

Proof. Identity (15) is well-known. We have

$$\begin{aligned}
[\alpha \lambda_{\max}(Z) + \beta]^+ &= [\lambda_{\max}(\alpha Z + \beta I)]^+ = \max_{t \in [0,1]} t \lambda_{\max}(\alpha Z + \beta I) \\
&= \max_{t \in [0,1], W \in \Delta^n_=(1)} t(\alpha Z + \beta I) \bullet W = \max_{W \in \Delta^n_\le(1)} (\alpha Z + \beta I) \bullet W,
\end{aligned}$$

where the third equality is due to (15) and the fourth equality is due to the fact that tW takes all possible values in $\Delta^n_\le(1)$ under the condition that $t \in [0,1]$ and $W \in \Delta^n_=(1)$. ∎

The second result gives some characterizations of the sum of the k largest eigenvalues of a real symmetric matrix.

Lemma 2.4 *Let $Z \in \mathcal{S}^n$ and integer $1 \le k \le n$ be given. Then, the following statements hold:*

a) For $t \in \Re$, we have

$$\sum_{i=1}^k \lambda_i(Z) \le t \;\Leftrightarrow\; \left\{ \begin{array}{lcl} t - ks - \mathrm{Tr}(Y) & \ge & 0, \\ Y - Z + sI & \succeq & 0, \\ Y & \succeq & 0, \end{array} \right.$$

for some $Y \in \mathcal{S}^n$ and $s \in \Re$;

b) The following identities hold:

$$\begin{aligned}
\sum_{i=1}^k \lambda_i(Z) &= \min_{Y \in \mathcal{S}^n_+} \max_{W \in \Delta^n_=(1)} k(Z - Y) \bullet W + \mathrm{Tr}(Y) && (17) \\
&= \max_{W \in \mathcal{S}^n} \{ Z \bullet W : \mathrm{Tr}(W) = k, \, 0 \preceq W \preceq I \}. && (18)
\end{aligned}$$

c) For every scalar $\alpha > 0$ and $\beta \in \Re$, the following identities hold:

$$\begin{aligned}
\left[\alpha \sum_{i=1}^k \lambda_i(Z) + \beta \right]^+ &= \min_{Y \in \mathcal{S}^n_+} \max_{W \in \Delta^n_\le(1)} k(\alpha Z - Y) \bullet W + [\beta + \mathrm{Tr}(Y)] \, \mathrm{Tr}(W) && (19) \\
&= \max_{W \in \mathcal{S}^n, \, t \in \Re} \{ \alpha Z \bullet W + \beta t : \mathrm{Tr}(W) = tk, \, 0 \preceq W \preceq tI, \, 0 \le t \le 1 \} && (20)
\end{aligned}$$

Proof. a) This statement is proved on pages 147-148 of Ben-Tal and Nemirovski [2].
b) Statement (a) clearly implies that

$$\sum_{i=1}^k \lambda_i(Z) = \min_{s \in \Re, Y \in \mathcal{S}^n} \{ ks + \mathrm{Tr}(Y) : \; Y + sI \succeq Z, \, Y \succeq 0 \}. \tag{21}$$

Noting that the condition $Y+sI \succeq Z$ is equivalent to $s \geq \lambda_{\max}(Z-Y)$, we can eliminate the variable s from the above min problem to conclude that

$$\sum_{i=1}^{k} \lambda_i(Z) = \min\{k\lambda_{\max}(Z-Y) + \mathrm{Tr}(Y) : Y \in \mathcal{S}^n_+\}. \tag{22}$$

This relation together with (15) clearly implies identity (17). Moreover, noting that the max problem (18) is the dual of min problem (21) and that they both have strictly feasible solutions, we conclude that identity (18) holds in view of a well-known strong duality result.

c) Using (22), the fact that $\inf_{x\in X}[x]^+ = [\inf\, X]^+$ for any $X \subseteq \Re$ and (16), we obtain

$$\begin{aligned}
\left[\alpha\sum_{i=1}^{k}\lambda_i(Z)+\beta\right]^+ &= \left[\sum_{i=1}^{k}\lambda_i\left(\alpha Z+\frac{\beta}{k}I\right)\right]^+ \\
&= \left[\min_{Y\in\mathcal{S}^n_+} k\lambda_{\max}\left(\alpha Z+\frac{\beta}{k}I-Y\right)+\mathrm{Tr}(Y)\right]^+ \\
&= \min_{Y\in\mathcal{S}^n_+}\left[k\lambda_{\max}\left(\alpha Z+\frac{\beta}{k}I-Y\right)+\mathrm{Tr}(Y)\right]^+ \\
&= \min_{Y\in\mathcal{S}^n_+}\max_{W\in\Delta^n_{\leq}(1)} k\left(\alpha Z+\frac{\beta}{k}I-Y\right)\bullet W+\mathrm{Tr}(Y)\mathrm{Tr}(W),
\end{aligned}$$

from which (19) immediately follows. Moreover, using (18), the fact that $[\gamma]^+ = \max_{t\in[0,1]} t\gamma$ for every $\gamma \in \Re$ and performing the change of variable $Y = t\tilde{Y}$ in the last equality below, we obtain

$$\begin{aligned}
\left[\alpha\sum_{i=1}^{k}\lambda_i(Z)+\beta\right]^+ &= \left[\sum_{i=1}^{k}\lambda_i\left(\alpha Z+\frac{\beta}{k}I\right)\right]^+ \\
&= \left[\max_{\tilde{Y}\in\mathcal{S}^n}\left\{\left(\alpha Z+\frac{\beta}{k}I\right)\bullet\tilde{Y} : \mathrm{Tr}(\tilde{Y})=k,\, 0\preceq\tilde{Y}\preceq I\right\}\right]^+ \\
&= \max_{\tilde{Y}\in\mathcal{S}^n,\, t\in\Re}\left\{t\left(\alpha Z+\frac{\beta}{k}I\right)\bullet\tilde{Y} : \mathrm{Tr}(\tilde{Y})=k,\, 0\preceq\tilde{Y}\preceq I,\, 0\leq t\leq 1\right\} \\
&= \max_{Y\in\mathcal{S}^n,\, t\in\Re}\left\{\left(\alpha Z+\frac{\beta}{k}I\right)\bullet Y : \mathrm{Tr}(Y)=tk,\, 0\preceq Y\preceq tI,\, 0\leq t\leq 1\right\},
\end{aligned}$$

i.e., (20) holds. ■

Corollary 2.5 *Let $Z \in \mathcal{S}^n$ and integer $1 \leq k < n$ be given. Then, the following statements hold:*

a) For any $\delta \geq 0$, we have

$$\sum_{i=1}^{k}\lambda_i(Z) \leq \min_{Y\in\Delta^n_{\leq}(\delta)}\max_{W\in\Delta^n_{=}(1)} k(Z-Y)\bullet W+\mathrm{Tr}(Y), \tag{23}$$

and equality holds whenever

$$\delta \geq \delta_1(Z) := \left(n\sum_{i=1}^{k}\lambda_i(Z) - k\mathrm{Tr}(Z)\right)/(n-k). \tag{24}$$

b) For any $\delta \geq 0$, $\alpha > 0$ and $\beta \in \Re$, we have

$$\left[\alpha \sum_{i=1}^{k} \lambda_i(Z) + \beta\right]^+ \leq \min_{Y \in \Delta^n_{\leq}(\delta)} \max_{W \in \Delta^n_{\leq}(1)} k(\alpha Z - Y) \bullet W + [\beta + \mathrm{Tr}(Y)]\, \mathrm{Tr}(W) \tag{25}$$

and equality holds whenever

$$\delta \geq \delta_2(Z) := \alpha \left(n \sum_{i=1}^{k} \lambda_i(Z) - k\mathrm{Tr}(Z) \right) /(n-k). \tag{26}$$

Proof. a) Inequality (23) is an immediate consequence of (17). Now, let $Y^* \in \mathcal{S}^n_+$ be an optimal solution of (17), or equivalently, problem (22). Then, we have

$$\sum_{i=1}^{k} \lambda_i(Z) = k\lambda_{\max}(Z - Y^*) + \mathrm{Tr}(Y^*) \geq k\mathrm{Tr}(Z - Y^*)/n + \mathrm{Tr}(Y^*),$$

or equivalently, $\mathrm{Tr}(Y^*) \leq \delta_1(Z)$, where $\delta_1(Z)$ is given by (24). This conclusion together with (22) implies that

$$\sum_{i=1}^{k} \lambda_i(Z) = \min_{Y \in \Delta^n_{\leq}(\delta)} k\lambda_{\max}(Z - Y) + \mathrm{Tr}(Y) \tag{27}$$

or equivalently, equality holds in (23), for any $\delta \geq \delta_1(Z)$.

b) Inequality (25) is an immediate consequence of (19). In view of (24) and (26), we can easily verify that $\delta_1(\alpha Z + \beta I/k) = \delta_2(Z)$. Using this identity, and replacing Z in (27) by $\alpha Z + \beta I/k$, we obtain that

$$\sum_{i=1}^{k} \lambda_i \left(\alpha Z + \frac{\beta}{k} I \right) = \min_{Y \in \Delta^n_{\leq}(\delta)} k\lambda_{\max} \left(\alpha Z + \frac{\beta}{k} I - Y \right) + \mathrm{Tr}(Y)$$

for any $\delta \geq \delta_2(Z)$. The rest of proof is similar to the one used to show (19) in Lemma 2.4. ∎

Lemma 2.6 *Let $X \in \Re^{p \times q}$ be given. Then, the following statements hold:*

a) the $p+q$ eigenvalues of the symmetric matrix $\mathcal{G}(X)$ defined in (5), arranged in nonascending order, are

$$\sigma_1(X), \cdots, \sigma_m(X), 0, \cdots, 0, -\sigma_m(X), \cdots, -\sigma_1(X),$$

where $m := \min(p, q)$;

b) For any positive integer $k \leq m$, we have

$$\sum_{i=1}^{k} \sigma_i(X) = \sum_{i=1}^{k} \lambda_i(\mathcal{G}(X)).$$

Proof. Statement (a) is proved on page 153 of [2] and statement (b) is an immediate consequence of (a). ∎

The following two results about the sum of the k largest singular values of a matrix follow immediately from Lemmas 2.4 and 2.6 and Corollary 2.5.

Proposition 2.7 *Let $X \in \Re^{p\times q}$ and integer $1 \le k \le \min\{p,q\}$ be given and set $n := p+q$. Then:*

a) For $t \in \Re$, we have

$$\sum_{i=1}^k \sigma_i(X) \le t \;\Leftrightarrow\; \left\{ \begin{array}{lcl} t - ks - \mathrm{Tr}(Y) & \ge & 0, \\ Y - \mathcal{G}(X) + sI & \succeq & 0, \\ Y & \succeq & 0, \end{array} \right.$$

for some $Y \in \mathcal{S}^n$ and $s \in \Re$;

b) The following identities hold:

$$\sum_{i=1}^k \sigma_i(X) = \min_{Y\in\mathcal{S}^n_+} \max_{W\in\Delta^n_=(1)} k(\mathcal{G}(X)-Y)\bullet W + \mathrm{Tr}(Y) \tag{28}$$

$$= \max_{W\in\mathcal{S}^n} \{\mathcal{G}(X)\bullet W : \mathrm{Tr}(W) = k,\ 0 \preceq W \preceq I\}. \tag{29}$$

c) For every scalar $\alpha > 0$ and $\beta \in \Re$, the following identities hold:

$$\left[\alpha \sum_{i=1}^k \sigma_i(X) + \beta\right]^+ = \min_{Y\in\mathcal{S}^n_+} \max_{W\in\Delta^n_\le(1)} k(\alpha\mathcal{G}(X)-Y)\bullet W + [\beta + \mathrm{Tr}(Y)]\,\mathrm{Tr}(W) \tag{30}$$

$$= \max_{W\in\mathcal{S}^n,\, t\in\Re} \{\alpha\mathcal{G}(X)\bullet W + \beta t : \mathrm{Tr}(W) = tk,\ 0 \preceq W \preceq tI,\ 0 \le t \le 1\}. \tag{31}$$

Corollary 2.8 *Let $X \in \Re^{p\times q}$ and integer $1 \le k \le \min\{p,q\}$ be given and define $n := p+q$ and $\delta(X) := (n \sum_{i=1}^k \sigma_i(X))/(n-k)$. Then, the following statements hold:*

a) For any $\delta \ge 0$, we have

$$\sum_{i=1}^k \sigma_i(X) \le \min_{Y\in\Delta^n_\le(\delta)} \max_{W\in\Delta^n_=(1)} k(\mathcal{G}(X)-Y)\bullet W + \mathrm{Tr}(Y),$$

and equality holds whenever $\delta \ge \delta(X)$.

b) For every $\delta \ge 0$, $\alpha > 0$ and $\beta \in \Re$, we have

$$\left[\alpha \sum_{i=1}^k \sigma_i(X) + \beta\right]^+ \le \min_{Y\in\Delta^n_\le(\delta)} \max_{W\in\Delta^n_\le(1)} k(\alpha\mathcal{G}(X)-Y)\bullet W + [\beta + \mathrm{Tr}(Y)]\,\mathrm{Tr}(W)$$

and equality holds whenever $\delta \ge \alpha\delta(X)$.

Proof. We first observe that (5) implies that $\mathrm{Tr}(\mathcal{G}(X)) = 0$ for every $X \in \Re^{p\times q}$. Applying Corollary 2.5 to the matrix $Z = \mathcal{G}(X)$ and noting Lemma 2.6, the conclusions of Corollary 2.8 follow. ∎

3 Problem reformulations

This section consists of three subsections. The first subsection shows that the restricted least squares problem (4) can be reduced to one which does not depend on the (usually large) number of rows of the matrices A and/or B. In the second and third subsections, we provide cone programming and smooth saddle point reformulations for (4), respectively.

3.1 Problem simplification

Observe that the number of rows of the data matrices A and B which appear in (4) is equal to the number of observations n, which is usually quite large in many applications. However, the size of the decision variable U in (4) does not depend on n. In this subsection we show how problem (4) can be reduced to similar types of problems in which the new matrix A is a $p \times p$ diagonal matrix and hence to problems which do not depend on n. Clearly, from a computational point of view, the resulting formulations need less storage space and can be more efficiently solved.

Since in most applications, the matrix A has full column rank, we assume that this property holds throughout the paper. Thus, there exists an orthonormal matrix $Q \in \Re^{p\times p}$ and a positive diagonal matrix $\Lambda \in \Re^{p\times p}$ such that $A^T A = Q\Lambda^2 Q^T$. Letting

$$X := Q^T U, \quad H := \Lambda^{-1} Q^T A^T B, \tag{32}$$

we have

$$\begin{aligned} \|B - AU\|_{\mathrm{F}}^2 - \|B\|_F^2 &= \|AU\|_F^2 - 2\,(AU) \bullet B \;=\; \mathrm{Tr}(U^T A^T A U) - 2\,\mathrm{Tr}(U^T A^T B) \\ &= \mathrm{Tr}(U^T Q \Lambda^2 Q^T U) - 2\,\mathrm{Tr}(U^T Q \Lambda H) \\ &= \|\Lambda X\|_F^2 - 2(\Lambda X) \bullet H \;=\; \|\Lambda X - H\|_F^2 - \|H\|_F^2. \end{aligned}$$

Noting that the singular values of $X = Q^T U$ and U are identical, we immediately see from the above identity that (4) is equivalent to

$$\begin{array}{ll} \min\limits_X & \frac{1}{2}\|\Lambda X - H\|_{\mathrm{F}}^2 \\ \text{s.t.} & \sum\limits_{i=1}^m \sigma_i(X) \le M, \end{array} \tag{33}$$

where Λ and H are defined in (32).

In view of Theorem 2.2, we observe that for any $\lambda \ge 0$ and $\epsilon \ge 0$, any ϵ-optimal solution X_ϵ of the following Lagrangian relaxation problem

$$\min_X \frac{1}{2}\|\Lambda X - H\|_{\mathrm{F}}^2 + \lambda \sum_{i=1}^m \sigma_i(X). \tag{34}$$

is an ϵ-optimal solution of problem (33) with $M = \sum_{i=1}^m \sigma_i(X_\epsilon)$. In practice, we often need to solve problem (33) for a sequence of Ms. Hence, one way to solve such problems is to solve problem (34) for a sequence of λs.

We will later present convex optimization methods for approximately solving the formulations (33) and (34), and hence, as a by-product, formulation (4).

Before ending this subsection, we provide bounds on the optimal solutions of problems (33) and (34).

Lemma 3.1 *For every $M > 0$, problem (33) has a unique optimal solution X_M^*. Moreover,*

$$\|X_M^*\|_F \le \tilde{r}_x := \min\left\{\frac{2\|\Lambda H\|_F}{\lambda_{\min}^2(\Lambda)}, M\right\}. \tag{35}$$

Proof. Using the fact that Λ is a $p \times p$ positive diagonal matrix, it is easy to see that the objective function of (33) is a (quadratic) strongly convex function, from which we conclude that (33) has a unique optimal solution X_M^*. Since $\|H\|_F^2/2$ is the value of the objective function of (33) at $X = 0$, we have $\|\Lambda X_M^* - H\|_F^2/2 \le \|H\|_F^2/2$, or equivalently $\|\Lambda X_M^*\|_F^2 \le 2(\Lambda H) \bullet X_M^*$. Hence, we have

$$(\lambda_{\min}(\Lambda))^2 \; \|X_M^*\|_F^2 \le \|\Lambda X_M^*\|_F^2 \le 2(\Lambda H) \bullet X_M^* \le 2\|X_M^*\|_F \, \|\Lambda H\|_F,$$

which implies that $\|X_M^*\|_F \le 2\|\Lambda H\|_F / \lambda_{\min}^2(\Lambda)$. Moreover, using the fact that $\|X\|_F^2 = \sum_{i=1}^m \sigma_i^2(X)$ for any $X \in \Re^{p \times q}$, we easily see that

$$\|X\|_F \le \sum_{i=1}^m \sigma_i(X). \tag{36}$$

Since X_M^* is feasible for (33), it then follows from (36) that $\|X_M^*\|_F \le M$. We have thus shown that inequality (35) holds. ■

Lemma 3.2 *For every $\lambda > 0$, problem (34) has a unique optimal solution X_λ^*. Moreover,*

$$\|X_\lambda^*\|_F \le \sum_{i=1}^m \sigma_i(X_\lambda^*) \le r_x := \min\left\{ \frac{\|H\|_F^2}{2\lambda}, \sum_{i=1}^m \sigma_i(\Lambda^{-1}H) \right\}. \tag{37}$$

Proof. As shown in Lemma 3.1, the function $X \in \Re^{p \times q} \to \|\Lambda X - H\|_F^2$ is a (quadratic) strongly convex function. Since the term $\lambda \sum_{i=1}^m \sigma_i(X)$ is convex in X, it follows that the objective function of (34) is strongly convex, from which we conclude that (34) has a unique optimal solution X_λ^*. Since $\|H\|_F^2/2$ is the value of the objective function of (34) at $X = 0$, we have

$$\lambda \sum_{i=1}^m \sigma_i(X_\lambda^*) \le \frac{1}{2}\|\Lambda X_\lambda^* - H\|_F^2 + \lambda \sum_{i=1}^m \sigma_i(X_\lambda^*) \le \frac{1}{2}\|H\|_F^2. \tag{38}$$

Also, considering the objective function of (34) at $X = \Lambda^{-1}H$, we conclude that

$$\lambda \sum_{i=1}^m \sigma_i(X_\lambda^*) \le \frac{1}{2}\|\Lambda X_\lambda^* - H\|_F^2 + \lambda \sum_{i=1}^m \sigma_i(X_\lambda^*) \le \lambda \sum_{i=1}^m \sigma_i(\Lambda^{-1}H). \tag{39}$$

Now, (37) follows immediately from (36), (38) and (39). ■

3.2 Cone programming reformulations

In this subsection, we provide cone programming reformulations for problems (33) and (34), respectively.

Proposition 3.3 *Problem (34) can be reformulated as the following cone programming:*

$$\begin{array}{cl} \min\limits_{r,s,t,X,Y} & 2r + \lambda t \\ s.t. & \begin{pmatrix} r+1 \\ r-1 \\ \mathrm{vec}(\Lambda X - H) \end{pmatrix} \in \mathcal{L}^{pq+2}, \\ & Y - \mathcal{G}(X) + sI \succeq 0, \\ & ms + \mathrm{Tr}(Y) - t \le 0, \;\; Y \succeq 0, \end{array} \tag{40}$$

where $(r, s, t, X, Y) \in \Re \times \Re \times \Re \times \Re^{p\times q} \times \mathcal{S}^n$ with $n := p + q$ and $\mathcal{G}(X)$ is defined in (5).

Proof. We first observe that (34) is equivalent to

$$\begin{array}{ll} \min\limits_{r,X} & 2r + \lambda t \\ \text{s.t.} & \|\Lambda X - H\|_{\mathrm{F}}^2 \leq 4r \\ & \sum\limits_{i=1}^m \sigma_i(X) - t \leq 0. \end{array} \tag{41}$$

Using Lemma 2.6 and the following relation

$$4r \geq \|v\|^2 \Leftrightarrow \begin{pmatrix} r+1 \\ r-1 \\ v \end{pmatrix} \in \mathcal{L}^{k+2},$$

for any $v \in \Re^k$ and $r \in \Re$, we easily see that (41) is equivalent to (40) ■

The following proposition can be similarly established.

Proposition 3.4 *Problem (33) can be reformulated as the following cone programming:*

$$\begin{array}{ll} \min\limits_{r,s,X,Y} & 2r \\ s.t. & \begin{pmatrix} r+1 \\ r-1 \\ \mathrm{vec}(\Lambda X - H) \end{pmatrix} \in \mathcal{L}^{pq+2}, \\ & Y - \mathcal{G}(X) + sI \succeq 0, \\ & ms + \mathrm{Tr}(Y) \leq M, \quad Y \succeq 0, \end{array} \tag{42}$$

where $(r, s, X, Y) \in \Re \times \Re \times \Re^{p\times q} \times \mathcal{S}^n$ with $n := p + q$ and $\mathcal{G}(X)$ is defined in (5).

3.3 Smooth saddle point reformulations

In this section, we provide smooth saddle point reformulations for problems (33) and (34).

3.3.1 Smooth saddle point reformulations for (34)

In this subsection, we reformulate (34) into smooth saddle point problems that can be suitably solved by Nesterov's algorithms as described in Subsection 4.1.

Theorem 3.5 *Let $\epsilon \geq 0$ be given and define $n := p + q$, $m := \min(p, q)$ and $r_y := nr_x/(n - m)$, where r_x is defined in (37). Assume that (X_ϵ, Y_ϵ) is an ϵ-optimal solution of the smooth saddle point problem*

$$\min_{\{X \in \mathcal{B}_F^{p\times q}(r_x),\ Y \in \Delta_{\leq}^n(r_y)\}} \max_{W \in \Delta_{=}^n(1)} \left\{ \frac{1}{2}\|\Lambda X - H\|_F^2 + \lambda m(\mathcal{G}(X) - Y) \bullet W + \lambda \mathrm{Tr}(Y) \right\}, \tag{43}$$

where $\mathcal{G}(X)$ is defined in (5). Then, X_ϵ is an ϵ-optimal solution of problem (34).

Proof. Let $\phi_\lambda(X,Y,W)$ denote the quantity within brackets in relation (43). Let $f_\lambda : \Re^{p\times q} \to \Re$ be defined as

$$f_\lambda(X) = \min_{Y\in\Delta^n_{\leq}(r_y)} \max_{W\in\Delta^n_{=}(1)} \phi_\lambda(X,Y,W), \tag{44}$$

and observe that (43) is equivalent to the problem $\min\{f_\lambda(X) : X \in \mathcal{B}_F^{p\times q}(r_x)\}$. In view of Corollary 2.8(a), f_λ majorizes the objective function of (34) over the space $\Re^{p\times q}$, and agrees with it whenever $\delta(X) \leq r_y$, where $\delta(\cdot)$ is defined in Corollary 2.8. Also, by Lemma 3.2, the optimal solution X^*_λ of (34) satisfies $\|X^*_\lambda\|_F \leq \sum_{i=1}^m \sigma_i(X^*_\lambda) \leq r_x$, and hence the relations $X^*_\lambda \in \mathcal{B}_F^{p\times q}(r_x)$ and $\delta(X^*_\lambda) \leq nr_x/(n-m) = r_y$. The above two observations clearly imply that (34) and (43) have the same optimal value f^*_λ. Assume now that, for some $\epsilon \geq 0$, (X_ϵ, Y_ϵ) is an ϵ-optimal solution of (43). Then, by (44), we have

$$f_\lambda(X_\epsilon) \leq \max_{W\in\Delta^n_{=}(1)} \phi_\lambda(X_\epsilon, Y_\epsilon, W) \leq f^*_\lambda + \epsilon.$$

Since f_λ majorizes the objective function of problem (34), which has f^*_λ as its optimal value, we conclude from the above relation that X_ϵ is an ϵ-optimal solution of (34). ∎

We next provide another smooth saddle point formulation for problem (34), which outperforms the formulation (43) in our computational experiments.

Theorem 3.6 *For some $\epsilon \geq 0$, assume that X_ϵ is an ϵ-optimal solution of the smooth saddle point problem*

$$\min_{X\in\mathcal{B}_F^{p\times q}(r_x)} \max_{W\in\Omega} \left\{ \frac{1}{2}\|\Lambda X - H\|_F^2 + \lambda m \mathcal{G}(X) \bullet W \right\}, \tag{45}$$

where $\mathcal{G}(X)$ and r_x are defined in (5) and (37), respectively, and the set Ω is defined as

$$\Omega := \{W \in \mathcal{S}^{p+q} : 0 \preceq W \preceq I/m, \mathrm{Tr}(W) = 1\}. \tag{46}$$

Then, X_ϵ is an ϵ-optimal solution of problem (34).

Proof. This result follows as immediately from Lemma 3.2 and relations (29) and (34). ∎

Generally, both minimization problems (43) and (45) have nonsmooth objective functions. In Subsection 4.1.2, we will describe a suitable method, namely Nesterov's smooth approximation scheme, for solving these min-max type problems. Also, in contrast with the dual of (43), which also has a nonsmooth objective function, the dual of (45), namely the problem

$$\max_{W\in\Omega} \min_{X\in\mathcal{B}_F^{p\times q}(r_x)} \left\{ \frac{1}{2}\|\Lambda X - H\|_F^2 + \lambda m \mathcal{G}(X) \bullet W \right\} \tag{47}$$

has the nice property that its objective function is a smooth function with Lipschitz continuous gradient (see Subsection 4.2.3 for details). In Subsections 4.1.1 and 4.2.3, we describe an algorithm, namely Nesterov's smooth method, for solving the dual of (45) which, as a by-product, yields a pair of primal and dual nearly-optimal solutions, and hence a nearly-optimal solution of (45). Finally, we should mention that our computational results presented in Sections 4 and 5 evaluate the efficiency of the aforementioned methods for solving formulations (43), (45) and the dual of (45). It also compares the performance of second-order methods for solving the cone programming formulation (40) with that of aforementioned first-order approaches.

3.3.2 Smooth saddle point reformulations for (33)

In this subsection, we will provide smooth saddle point reformulations for (33) that can be suitably solved by Nesterov's algorithms as described in Subsection 4.1.

By directly applying Theorem 2.1 to problem (33), we obtain the following result.

Lemma 3.7 *Let $m := \min(p, q)$. Suppose that $\bar{X} \in \Re^{p\times q}$ satisfies $\sum_{i=1}^{m} \sigma_i(\bar{X}) < M$ and let γ be a scalar such that $\gamma \geq \bar{\gamma}$, where $\bar{\gamma}$ is given by*

$$\bar{\gamma} = \frac{\|\Lambda\bar{X} - H\|_{\mathrm{F}}^2/2}{M - \sum_{i=1}^{m} \sigma_i(\bar{X})}. \tag{48}$$

Then, the following statements hold:

a) The optimal values of (33) and the penalized problem

$$\min_{X\in\Re^{p\times q}} \left\{ \frac{1}{2}\|\Lambda X - H\|_{\mathrm{F}}^2 + \gamma \left[\sum_{i=1}^{m} \sigma_i(X) - M\right]^+ \right\} \tag{49}$$

coincide, and the optimal solution solution X_M^* of (33) is an optimal solution of (49);

b) if $\epsilon \geq 0$ and X_ϵ is an ϵ-optimal solution of problem (49), then the point X^ϵ defined as

$$X^\epsilon := \frac{X_\epsilon + \theta\bar{X}}{1+\theta}, \qquad \textit{where } \theta := \frac{\left[\sum_{i=1}^{m} \sigma_i(X_\epsilon) - M\right]^+}{M - \sum_{i=1}^{m} \sigma_i(\bar{X})}, \tag{50}$$

is an ϵ-optimal solution of (33).

Theorem 3.8 *Let $n := p + q$, $m := \min(p, q)$, $\tilde{r}_y := nM/(n - m)$, and $\tilde{r}_x$ be defined as in (35). Suppose that $\bar{X} \in \Re^{p\times q}$ satisfies $\sum_{i=1}^{m} \sigma_i(\bar{X}) < M$ and let γ be a scalar such that $\gamma \geq \bar{\gamma}$, where $\bar{\gamma}$ is defined in (48). For $\epsilon \geq 0$, assume that (X_ϵ, Y_ϵ) is an ϵ-optimal solution of the problem*

$$\min_{\{X\in\mathcal{B}_F^{p\times q}(\tilde{r}_x),\ Y\in\Delta_{\leq}^n(\tilde{r}_y)\}} \max_{W\in\Delta_{\leq}^n(1)} \left\{ \frac{1}{2}\|\Lambda X - H\|_{\mathrm{F}}^2 + \gamma m(\mathcal{G}(X) - Y) \bullet W + \gamma(\mathrm{Tr}(Y) - M)\mathrm{Tr}(W) \right\}. \tag{51}$$

and let X^ϵ be given by (50). Then, the matrix X^ϵ is an ϵ-optimal solution of (33).

Proof. We first show that X_ϵ is an ϵ-optimal solution of (49). Let X_M^* and f_M^* be the optimal solution and optimal value of (33), respectively. Then, it follows from Lemma 3.7 that they are also the optimal solution and optimal value of (49), respectively. Let $\phi_M(X, Y, W)$ denote the quantity within brackets in relation (51) and let $f_M : \Re^{p\times q} \to \Re$ be defined as

$$f_M(X) = \min_{Y\in\Delta_{\leq}^n(\tilde{r}_y)} \max_{W\in\Delta_{\leq}^n(1)} \phi_M(X, Y, W). \tag{52}$$

Note that (51) is equivalent to the problem $\min\{f_M(X) : X \in \mathcal{B}_F^{p\times q}(\tilde{r}_x)\}$. In view of Corollary 2.8(b), f_M majorizes the objective function of (49) over $\Re^{p\times q}$ and agrees with it whenever $\delta(X) \le \tilde{r}_y$, where $\delta(\cdot)$ is defined in Corollary 2.8. Also, by Lemma 3.1, the optimal solution X_M^* of (33) satisfies $\|X_M^*\|_F \le \tilde{r}_x$, or equivalently $X_M^* \in \mathcal{B}_F^{p\times q}(\tilde{r}_x)$. Moreover, since X_M^* satisfies the constraints of (33), we have $\sum_{i=1}^m \sigma_i(X_M^*) \le M$, and hence that $\delta(X_M^*) \le nM/(n-m) = \tilde{r}_y$. All the above conclusions imply that problems (51) and (49) have the same optimal value f_M^*. Assume now that, for some $\epsilon \ge 0$, (X_ϵ, Y_ϵ) is an ϵ-optimal solution of (51). Then, by (52), we have

$$f_M(X_\epsilon) \le \max_{W\in\Delta^n_{\le}(1)} \phi_M(X_\epsilon, Y_\epsilon, W) \le f_M^* + \epsilon.$$

Since f_M majorizes the objective function of problem (49) which has f_M^* as its optimal value, we conclude from the above relation that X_ϵ is an ϵ-optimal solution of (49). The later conclusion together with Lemma 3.7(b) yields the conclusion of the theorem. ∎

The above theorem gives one possible smooth saddle point reformulation, namely (51), of problem (33). We next provide an alternative smooth saddle point formulation for problem (33).

Theorem 3.9 *Suppose that $\bar{X} \in \Re^{p\times q}$ satisfies $\sum_{i=1}^m \sigma_i(\bar{X}) < M$ and let γ be a scalar such that $\gamma \ge \bar{\gamma}$, where $\bar{\gamma}$ is defined in (48). For some $\epsilon \ge 0$, assume that X_ϵ is an ϵ-optimal solution of the problem*

$$\min_{X\in\mathcal{B}_F^{p\times q}(\tilde{r}_x)} \max\left\{\frac{1}{2}\|\Lambda X - H\|_{\rm F}^2 + \gamma(q\mathcal{G}(X)\bullet W - Mt) : \mathrm{Tr}(W) = tk,\ 0 \preceq W \preceq tI,\ 0 \le t \le 1\right\}, \tag{53}$$

where $\tilde{r}_x$ is defined in (35), and let X^ϵ be defined in (50). Then, X^ϵ is an ϵ-optimal solution of (33).

Proof. By Lemma 3.7, the optimal solution X_M^* of (33) is also that of (49). Moreover, X_M^* satisfies $X_M^* \in \mathcal{B}_F^{p\times q}(\tilde{r}_x)$ due to Lemma 3.1. Further, in view of (31), the objective functions of problems (49) and (53) are equal to each other over the space $\Re^{p\times q}$. We then easily see that X_M^* is also an optimal solution of (53), and problems (53) and (49) have the same optimal value. Thus, it follows that X_ϵ is also an ϵ-optimal solution of problem (49). This observation together with Lemma 3.7 immediately yields the conclusion of this theorem.

∎

4 Numerical methods

In this section, we discuss numerical methods for solving problems (34) and (33). More specifically, Subsection 4.1 reviews general purpose algorithms, due to Nesterov (see [12, 13]), for solving a convex minimization problem over a relatively simple set with either: i) a smooth objective function that has Lipschitz continuous gradient, or; ii) a nonsmooth objective function which can be arbitrarily closely approximated by a smooth function that has Lipschitz continuous gradient. In Subsection 4.2, we present the implementation details of Nesterov's algorithms for solving the three reformulations (43), (45) and (47) of problem (34) (see Subsections 4.2.1, 4.2.2 and 4.2.3, respectively).

The implementation details of the other formulations discussed in the paper, more specifically, the reformulations (51) and (53) of problem (33), as well as their duals, will not be discussed here.

We just observe that their implementation details are, for the most part, close to the ones for problem (34). The only complication might lie in the reformulation (53), whose constraint set might not be so simple in that the subproblems arising in both Nesterov's methods (see Subsection 4.1) may not have closed form solutions. We however will not investigate this issue in this paper. We should also mention that preliminary computational tests seem to indicate that the reformulations (43), (45) and (47) of problem (34) can be more easily solved than the corresponding ones for problem (33). The reason behind this might be in the large size of the penalty parameter that appears in (33), and hence their corresponding reformulations. This is the main reason for us to focus our computational investigation only on problem (34) and its corresponding reformulations.

4.1 Review of Nesterov's algorithms

In this subsection, we review Nesterov's smooth first-order method [12, 13] for solving a class of smooth convex programming (CP) problems (Subsection 4.1.1), and Nesterov's smooth approximation scheme [13] for solving a class of non-smooth CP problems (Subsection 4.1.2).

4.1.1 Nesterov's smooth method

Let $\mathcal{U}$ and $\mathcal{V}$ be normed vector spaces with the respective norms denoted by $\|\cdot\|_{\mathcal{U}}$ and $\|\cdot\|_{\mathcal{V}}$. We will discuss Nesterov's smooth first-order method for solving the class of CP problems

$$\min_{u \in U} f(u) \tag{54}$$

where the objective function $f : U \to \Re$ has the form

$$f(u) := \max_{v \in V} \phi(u, v), \quad \forall u \in U, \tag{55}$$

for some continuous function $\phi : U \times V \to \Re$ and nonempty compact convex subsets $U \subseteq \mathcal{U}$ and $V \subseteq \mathcal{V}$. We make the following assumptions regarding the function ϕ:

B.1 for every $u \in U$, the function $\phi(u, \cdot) : V \to \Re$ is *strictly* concave;
B.2 for every $v \in V$, the function $\phi(\cdot, v) : U \to \Re$ is convex differentiable;
B.3 the function f is L-Lipschitz-differentiable on U with respect to $\|\cdot\|_{\mathcal{U}}$ (see (7)).

It is well-known that Assumptions B.1 and B.2 imply that the function f is convex differentiable, and that its gradient is given by

$$\nabla f(u) = \nabla_u \phi(u, v(u)), \quad \forall u \in U, \tag{56}$$

where $v(u)$ denotes the unique solution of (55) (see for example Proposition B.25 of [3]). Moreover, problem (54) and its dual, namely:

$$\max_{v \in V} \{g(v) := \min_{u \in U} \phi(u, v)\}, \tag{57}$$

both have optimal solutions u^* and v^* such that $f(u^*) = g(v^*)$. Finally, Assumption B.3 ensures that problem (54) can be suitably solved by Nesterov's smooth minimization approach [12, 13], which we will now describe.

Let $p_U : U \to \Re$ be a differentiable strongly convex function with modulus $\sigma_U > 0$ with respect to $\|\cdot\|_{\mathcal{U}}$, i.e.,

$$p_U(u) \geq p_U(\tilde{u}) + \langle \nabla p_U(\tilde{u}), u - \tilde{u}\rangle + \frac{\sigma_U}{2}\|u - \tilde{u}\|^2_{\mathcal{U}}, \quad \forall u, \tilde{u} \in U. \tag{58}$$

Let u_0 be defined as

$$u_0 = \arg\min\{p_U(u) : \; u \in U\}. \tag{59}$$

By subtracting the constant $p_U(u_0)$ from the function $p_U(\cdot)$, we may assume without any loss of generality that $p_U(u_0) = 0$. The Bregman distance $d_{p_U} : U \times U \to \Re$ associated with p_U is defined as

$$d_{p_U}(u; \tilde{u}) = p_U(u) - l_{p_U}(u; \tilde{u}), \quad \forall u, \tilde{u} \in U, \tag{60}$$

where $l_{p_U} : \mathcal{U} \times U \to \Re$ is the "linear approximation" of p_U defined as

$$l_{p_U}(u; \tilde{u}) = p_U(\tilde{u}) + \langle \nabla p_U(\tilde{u}), u - \tilde{u}\rangle, \quad \forall (u, \tilde{u}) \in \mathcal{U} \times U.$$

Similarly, we can define the function $l_f(\cdot;\cdot)$ that will be used subsequently.

We now describe Nesterov's smooth minimization approach [12, 13] for solving problem (54)-(55), and how it simultaneously solves its dual problem (57). Nesterov's algorithm uses a sequence $\{\alpha_k\}_{k\geq 0}$ of scalars satisfying the following condition:

$$0 < \alpha_k \leq \left(\sum_{i=0}^{k} \alpha_i\right)^{1/2}, \; \forall k \geq 0. \tag{61}$$

Clearly, (61) implies that $\alpha_0 \in (0, 1]$.

Nesterov's smooth algorithm:

Let $u_0 \in U$ and $\{\alpha_k\}_{k\geq 0}$ satisfy (59) and (61), respectively.
Set $u_0^{ag} = u_0$, $v_0 = 0 \in \mathcal{V}$, $\tau_0 = 1$ and $k = 1$;

1) Compute $v(u_{k-1})$ and $\nabla f(u_{k-1})$.

2) Compute $(u_k^{sd}, u_k^{ag}) \in U \times U$ and $v_k \in V$ as

$$v_k \equiv (1 - \tau_{k-1})v_{k-1} + \tau_{k-1}v(u_{k-1}) \tag{62}$$

$$u_k^{ag} \equiv \operatorname{argmin}\left\{\frac{L}{\sigma_U}\, p_U(u) + \sum_{i=0}^{k-1} \alpha_i\, l_f(u; u_i) : u \in U\right\} \tag{63}$$

$$u_k^{sd} \equiv (1 - \tau_{k-1})u_{k-1}^{sd} + \tau_{k-1} \operatorname{argmin}\left\{\frac{L}{\sigma_U}\, d_{p_U}(u; u_{k-1}^{ag}) + \alpha_{k-1}\, l_f(u; u_{k-1}) : u \in U\right\}. \tag{64}$$

3) Set $\tau_k = \alpha_k / (\sum_{i=0}^{k} \alpha_i)$ and $u_k = (1 - \tau_k)u_k^{sd} + \tau_k u_k^{ag}$.

4) Set $k \leftarrow k + 1$ and go to step 1).

end

We now state the main convergence result regarding Nesterov's smooth algorithm for solving problem (54) and its dual (57). Its proof is given in a more special context in [13] and also in Theorem 2.2 of Lu [10] in the general framework of our current discussion.

Theorem 4.1 *The sequence $\{(u_k^{sd}, v_k)\} \subseteq U \times V$ generated by Nesterov's smooth algorithm satisfies*

$$0 \le f(u_k^{sd}) - g(v_k) \le \frac{LD_U}{\sigma_U(\sum_{i=0}^{k-1} \alpha_i)}, \quad \forall k \ge 1, \tag{65}$$

where

$$D_U = \max\{p_U(u) : \ u \in U\}. \tag{66}$$

Proof. It has been proved in Theorem 2 of [13] that the sequence $\{(u_k, u_k^{sd})\}_{k=1}^{\infty} \subseteq U \times U$ generated by Nesterov's smooth algorithm satisfies

$$\begin{aligned} \left(\sum_{i=0}^{k-1} \alpha_i\right) f(u_k^{sd}) &\le \min\left\{ \frac{L}{\sigma_U} p_U(u) + \sum_{i=0}^{k-1} \alpha_i \, l_f(u; u_i) : u \in U \right\} \\ &\le \frac{LD_U}{\sigma_U} + \min_{u \in U} \left\{ \sum_{i=0}^{k-1} \alpha_i \, l_f(u; u_i) \right\}, \quad \forall k \ge 1, \end{aligned}$$

where the last inequality follows from (66). This relation implies

$$f(u_k^{sd}) - \min_{u \in U} \left\{ \sum_{i=0}^{k-1} \tilde{\alpha}_i \, l_f(u; u_i) \right\} \le \frac{LD_U}{\sigma_U(\sum_{i=0}^{k-1} \alpha_i)}, \quad \forall k \ge 1, \tag{67}$$

where $\tilde{\alpha}_i := \alpha_i / (\sum_{i=0}^{k-1} \alpha_i)$ for all $i = 0, \ldots, k-1$. Now, the definitions of l_f and the vector $v(u)$ together with Assumption B.2 imply that

$$l_f(u; u_i) = f(u_i) + \langle \nabla f(u_i), u - u_i \rangle = \phi(u_i, v(u_i)) + \langle \nabla_u \phi(u_i, v(u_i)), u - u_i \rangle \le \phi(u, v(u_i)), \quad \forall i \ge 0, \tag{68}$$

which, together with Assumption B.1 and the definition of g, imply that

$$\min_{u \in U} \sum_{i=0}^{k-1} \tilde{\alpha}_i \, l_f(u; u_i) \le \min_{u \in U} \sum_{i=0}^{k-1} \tilde{\alpha}_i \, \phi(u, v(u_i)) \le \min_{u \in U} \phi\left(u, \sum_{i=0}^{k-1} \tilde{\alpha}_i v(u_i)\right) = g\left(\sum_{i=0}^{k-1} \tilde{\alpha}_i v(u_i)\right). \tag{69}$$

Now, using (62) and the definition of τ_k in step 3 of the algorithm, it is easy to verify that $v_k = \sum_{i=0}^{k-1} \tilde{\alpha}_i v(u_i)$. This fact together with relations (67) and (69) yield (65) for every $k \ge 1$. ■

A typical sequence $\{\alpha_k\}$ satisfying (61) is the one in which $\alpha_k = (k+1)/2$ for all $k \ge 0$. With this choice for $\{\alpha_k\}$, we have the following specialization of Theorem 4.1.

Theorem 4.2 *If $\alpha_k = (k+1)/2$ for every $k \ge 0$, then the sequence $\{(u_k^{sd}, v_k)\} \subseteq U \times V$ generated by Nesterov's smooth algorithm satisfies*

$$0 \le f(u_k^{sd}) - g(v_k) \le \frac{4LD_U}{\sigma_U k(k+1)}, \quad \forall k \ge 1,$$

where D_U is defined in (66). Thus, the iteration complexity of finding an ϵ-optimal solution to (54) and its dual (57) by Nesterov's smooth algorithm does not exceed $2[(LD_U)/(\sigma_U \epsilon)]^{1/2}$.

4.1.2 Nesterov's smooth approximation scheme

In this subsection, we briefly review Nesterov's smooth approximation scheme [13] for solving a class of non-smooth CP problems which admit special smooth convex-concave saddle point (i.e., min-max) reformulations.

Assume that $\mathcal{U}$ and $\mathcal{V}$ are normed vector spaces, and $U \subseteq \mathcal{U}$ and $V \subseteq \mathcal{V}$ are compact convex sets. Let $f : U \to \Re$ be a convex function given in the following form:

$$f(u) := \hat{f}(u) + \max_{v \in V}\{\langle \mathcal{E}u, v\rangle - \phi(v)\}, \quad \forall u \in U, \tag{70}$$

where $\hat{f} : U \to \Re$ is an $L_{\hat{f}}$-Lipschitz-differentiable convex function with respect to a given norm $\|\cdot\|_{\mathcal{U}}$ in $\mathcal{U}$ for some constant $L_{\hat{f}} \geq 0$, $\phi : V \to \Re$ is a continuous convex function, and $\mathcal{E}$ is a linear operator from $\mathcal{U}$ to $\mathcal{V}^*$. For the remaining subsection, our problem of interest is the following CP problem:

$$\min\{f(u) : u \in U\}, \tag{71}$$

where $f(u)$ is given by (70).

The function f defined in (70) is generally non-differentiable but can be closely approximated by a Lipschitz-differentiable function using the following construction due to Nesterov [13]. Let $p_V : V \to \Re$ be a continuous strongly convex function with modulus $\sigma_V > 0$ with respect to a given norm $\|\cdot\|_{\mathcal{V}}$ on V satisfying $\min\{p_V(v) : v \in V\} = 0$. For some *smoothness* parameter $\eta > 0$, consider the following function

$$f_\eta(u) := \hat{f}(u) + \max_{v \in V}\left\{\langle \mathcal{E}u, v\rangle - \phi(v) - \eta p_V(v)\right\}. \tag{72}$$

The following result, due to Nesterov [13], shows that f_η is a Lipschitz-differentiable function with respect to $\|\cdot\|_{\mathcal{U}}$ whose "closeness" to f depends linearly on the parameter η.

Theorem 4.3 *The following statements hold:*

a) For every $u \in U$, we have $f_\eta(u) \leq f(u) \leq f_\eta(u) + \eta D_V$, where

$$D_V := \max\{p_V(v) : v \in V\}; \tag{73}$$

b) The function $f_\eta(u)$ is L_η-Lipschitz-differentiable with respect to $\|\cdot\|_{\mathcal{U}}$, where

$$L_\eta := L_{\hat{f}} + \|\mathcal{E}\|^2_{\mathcal{U},\mathcal{V}}/(\eta\sigma_V). \tag{74}$$

Clearly, the CP problem

$$\min\{f_\eta(u) : u \in U\} \tag{75}$$

is a special case of the formulation (54)-(55) in which $\phi(u,v) = \hat{f}(u) + \langle \mathcal{E}u, v\rangle - \phi(v) - \eta p_V(v)$ for every $(u,v) \in U \times V$. Moreover, in view of the Theorem 4.3(b) and assumptions on the functions $\hat{f}$, ϕ and p_V, we conclude that problem (71)-(72) satisfies Assumptions (B.1)-(B.3). Thus, the CP problem (71)-(72) and its dual, namely:

$$\max_{v \in V}\left\{g_\eta(v) := -\phi(v) - \eta p_V(v) + \min_{u \in U}\left(\hat{f}(u) + \langle \mathcal{E}u, v\rangle\right)\right\}, \tag{76}$$

can be solved by Nesterov's smooth algorithm. In view of Theorem 4.3(a), near optimal solutions of problem (71) and its dual

$$\max_{v \in V} \left\{ g(v) := -\phi(v) + \min_{u \in U} \left(\hat{f}(u) + \langle \mathcal{E}u, v \rangle \right) \right\}, \tag{77}$$

are also obtained as a by-product.

We are now ready to describe Nesterov's smooth approximation scheme for solving problem (71) and its dual (77).

Nesterov's smooth approximation scheme:

1) Let $\epsilon > 0$ be given, and let $\eta := \epsilon/(2D_V)$.

2) Apply Nesterov's smooth algorithm to problem (75) and terminate whenever a pair of iterates $(u_k^{sd}, v_k) \in U \times V$ satisfying $f(u_k^{sd}) - g(v_k) \leq \epsilon$ is found.

end

The main convergence result regarding the above algorithm is given in Section 4 of Nesterov [13].

Theorem 4.4 *The pair of iterates $(u_k^{sd}, v_k) \in U \times V$ generated by Nesterov's smooth method applied to problem (75)-(72) with the sequence $\{\alpha_k\}$ given by $\alpha_k = (k+1)/2$ for every $k \geq 0$ satisfies*

$$0 \leq f(u_k^{sd}) - g(v_k) \leq \frac{\epsilon}{2} + \frac{8\|\mathcal{E}\|^2_{\mathcal{U},\mathcal{V}} D_U D_V}{\epsilon \sigma_U \sigma_V (k+1)^2} + \frac{4L_{\hat{f}} D_U}{\sigma_U (k+1)^2}, \quad \forall k \geq 1.$$

As a consequence, for given a scalar $\epsilon > 0$, Nesterov's smooth approximation scheme finds an ϵ-optimal solution of problem (71)-(70) and its dual (77) in a number of iterations which does not exceed

$$\sqrt{\frac{16\|\mathcal{E}\|^2_{U,V} D_U D_V}{\epsilon^2 \sigma_U \sigma_V} + \frac{8L_{\hat{f}} D_U}{\epsilon \sigma_U}}. \tag{78}$$

Proof. Using Theorems 4.2 and 4.3, and the fact that $g_\eta(v) \leq g(v)$ for all $v \in V$, we immediately have

$$\frac{4L_\eta D_U}{\sigma_U (k+1)^2} \geq f_\eta(u_k^{sd}) - g_\eta(v_k) \geq f(u_k^{sd}) - \eta D_V - g(v_k),$$

which together with (74) and the fact that $\eta := \epsilon/(2D_V)$, implies that the first statement holds. The second statement directly follows from the first one. ∎

4.2 Implementation details of Nesterov's algorithms

In this subsection, we present the implementation details of Nesterov's algorithms for solving the three reformulations (43), (45) and (47) of problem (34).

4.2.1 Implementation details of Nesterov's smooth approximation scheme for (43)

The implementation details of Nesterov's smooth approximation scheme (see Subsection 4.1.2) for solving formulation (43) are addressed in this subsection. In particular, we specify in the context of this formulation the norms, prox-functions and subproblems that appear in Nesterov's smooth approximation scheme.

For the purpose of our implementation, we reformulate problem (43) into the problem

$$\min_{\{X \in \mathcal{B}_F^{p\times q}(1),\ Y \in \Delta_{\leq}^n(1)\}} \max_{W \in \Delta_{=}^n(1)} \left\{ \frac{1}{2}\|r_x \Lambda X - H\|_F^2 + \lambda m (r_x \mathcal{G}(X) - r_y Y) \bullet W + \lambda r_y \mathrm{Tr}(Y) \right\} \tag{79}$$

obtained by scaling the variables X and Y of (43) as $X \leftarrow X/r_x$ and $Y \leftarrow Y/r_y$. From now on, our discussion in this subsection will focus on formulation (79) rather than (43).

Let $n := p + q$, $u := (X, Y)$, $v := W$ and define

$$\begin{aligned} U &:= \mathcal{B}_F^{p\times q}(1) \times \Delta_{\leq}^n(1) \subseteq \Re^{p\times q} \times \mathcal{S}^n =: \mathcal{U}, \\ V &:= \Delta_{=}^n(1) \subseteq \mathcal{S}^n =: \mathcal{V} \end{aligned}$$

and

$$\hat{f}(u) := \frac{1}{2}\|r_x \Lambda X - H\|_F^2 + \lambda r_y \mathrm{Tr}(Y), \quad \mathcal{E}u := \lambda m (r_x \mathcal{G}(X) - r_y Y), \quad \phi(v) = 0,$$

for every $u = (X, Y) \in U$ and $v \in V$. Then, in view of these definitions, it follows that problem (79) can be expressed in the form (70), and thus be suitably solved by Nesterov's smooth approximation scheme described in Subsection 4.1.2. In the following paragraphs we specify the norms $\|\cdot\|_{\mathcal{U}}$ and $\|\cdot\|_{\mathcal{V}}$ for $\mathcal{U}$ and $\mathcal{V}$, suitable prox-functions $p_U(\cdot)$ and $p_V(\cdot)$ for U and V and the subproblems that appear in Nesterov's smooth approximation scheme for solving (79).

The norms $\|\cdot\|_{\mathcal{U}}$ and $\|\cdot\|_{\mathcal{V}}$ for the spaces $\mathcal{U}$ and $\mathcal{V}$ are defined as

$$\begin{aligned} \|u\|_{\mathcal{U}} &:= (\xi_x \|X\|_F^2 + \xi_y \|Y\|_1^2)^{1/2}, \quad \forall u = (X, Y) \in \mathcal{U}, \\ \|v\|_{\mathcal{V}} &:= \|v\|_1, \quad \forall v \in \mathcal{V}, \end{aligned} \tag{80}$$

where ξ_x and ξ_y are some positive scalars defined below. We easily see that the dual norm of $\|\cdot\|_{\mathcal{U}}$ is given by

$$\|u\|_{\mathcal{U}}^* = (\xi_x^{-1}\|X\|_F^2 + \xi_y^{-1}\|Y\|_\infty^2)^{1/2}, \ \forall u = (X, Y) \in \mathcal{U}^*. \tag{81}$$

We will now specify the prox-functions for the sets U and V. We first define prox-functions $d_x(\cdot)$ and $d_y(\cdot)$ for the sets $\mathcal{B}_F^{p\times q}(1)$ and $\Delta_{=}^n(1)$ which appears in the definition of the set U as follows:

$$\begin{aligned} d_x(X) &= \|X\|_F^2/2, \quad \forall X \in \mathcal{B}_F^{p\times q}(1), \\ d_y(Y) &= \mathrm{Tr}(Y \log Y) + (\log n - 1)\mathrm{Tr}(Y) + 1, \quad \forall Y \in \Delta_{\leq}^n(1). \end{aligned}$$

We can easily see that $d_x(\cdot)$ (resp., $d_y(\cdot)$) is a strongly differentiable convex function with modulus $\sigma_x = 1$ (resp., $\sigma_y = 1$) with respect to the norm $\|\cdot\|_F$ (resp., $\|\cdot\|_1$). The prox-functions $p_U(\cdot)$ and $p_V(\cdot)$ for U and V, respectively, are then defined as

$$p_U(u) = \eta_x d_x(X) + \eta_y d_y(Y), \quad \forall u = (X, Y) \in U, \tag{82}$$

$$p_V(v) = \mathrm{Tr}(v \log v) + \log n, \quad \forall v \in \Delta_{=}^n(1), \tag{83}$$

where η_x and η_y are some positive scalars that will be specified below.

We will now develop an upper bound on the iteration-complexity bound (78) depending on the scalars ξ_x, ξ_y, η_x and η_y, which will then be chosen to minimize the derived upper bound. We easily see that $p_U(\cdot)$ (resp., $p_V(\cdot)$) is a strongly differentiable convex function on U (resp., V) with modulus

$$\sigma_U = \min\{\eta_x/\xi_x,\ \eta_y/\xi_y\} \quad (\text{resp., } \sigma_V = 1) \tag{84}$$

with respect to the norm $\|\cdot\|_{\mathcal{U}}$ (resp., $\|\cdot\|_{\mathcal{V}}$). Also, it is easy to verify that $\min\{p_U(u) : u \in \mathcal{U}\} = \min\{p_V(v) : v \in \mathcal{V}\} = 0$ and that

$$D_U \;:=\; \max_{u\in U} p_U(u) = \eta_x D_x + \eta_y D_y, \tag{85}$$

$$D_V \;:=\; \max_{v\in V} p_V(v) = \log n, \tag{86}$$

where

$$D_x \;:=\; \max\{d_x(X) : X \in \mathcal{B}_F^{p\times q}(1)\} = 1/2,$$

$$D_y \;:=\; \max_{Y\in\Delta^n_{\leq}(1)} d_y(Y) = \max\{1, \log n\}.$$

In addition, using (81), we can show that $\hat{f}(u)$ is $L_{\hat{f}}$-Lipschitz-differentiable on U with respect to $\|\cdot\|_{\mathcal{U}}$, where

$$L_{\hat{f}} = \max_{\{h=(h_x,h_y)\in\mathcal{U}:\ \|h\|_{\mathcal{U}}\leq 1\}} \left\| \begin{pmatrix} r_x^2\Lambda^2 h_x \\ 0 \end{pmatrix} \right\|_{\mathcal{U}}^* = \max_{h_x\in\mathcal{B}_F^{p\times q}(\xi_x^{-1/2})} \xi_x^{-1/2}\|r_x^2\Lambda^2 h_x\|_F = r_x^2\|\Lambda\|^2\xi_x^{-1}. \tag{87}$$

In view of Lemma 2.6(a), we have $\|\mathcal{G}(X)\|_\infty = \sigma_1(X) \leq \|X\|_F$ for any $X \in \Re^{p\times q}$. Using this relation, (6), (80) and the fact $\|\cdot\|_{\mathcal{V}}^* = \|\cdot\|_\infty$, we further obtain that

$$\begin{aligned}
\|\mathcal{E}\|_{\mathcal{U},\mathcal{V}} &= \max\left\{\|\lambda m(r_x\mathcal{G}(X) - r_y Y)\|_{\mathcal{V}}^* :\ u = (X,Y)\in\mathcal{U}, \|u\|_{\mathcal{U}} \leq 1\right\}, \\
&\leq \lambda m \max\left\{r_x\|\mathcal{G}(X)\|_\infty + r_y\|Y\|_\infty :\ u = (X,Y)\in\mathcal{U}, \|u\|_{\mathcal{U}} \leq 1\right\}, \\
&\leq \lambda m \max\left\{r_x\|X\|_F + r_y\|Y\|_1 :\ u = (X,Y)\in\mathcal{U}, \|u\|_{\mathcal{U}} \leq 1\right\}, \\
&\leq \lambda m\sqrt{r_x^2/\xi_x + r_y^2/\xi_y},
\end{aligned} \tag{88}$$

where the last inequality follows from the Cauchy-Schwartz inequality. In view of (84)-(88), we see that iteration complexity (78) is bounded by

$$\Gamma(\xi_x,\xi_y,\eta_x,\eta_y) = \sqrt{\left(\frac{16\lambda^2 m^2(r_x^2\xi_x^{-1} + r_y^2\xi_y^{-1})D_V}{\epsilon^2} + \frac{8r_x^2\|\Lambda\|^2\xi_x^{-1}}{\epsilon}\right)\left(\frac{\eta_x D_x + \eta_y D_y}{\min(\eta_x/\xi_x, \eta_y/\xi_y)}\right)}.$$

It is now easy to see that the following choice of $(\xi_x,\xi_y,\eta_x,\eta_y)$ minimizes the above quantity:

$$\xi_x = \eta_x = r_x\sigma_x^{1/2}D_x^{-1/2}\left(\frac{\epsilon\sigma_V\|\Lambda\|^2}{2D_V} + \lambda^2 m^2\right)^{1/2}, \quad \xi_y = \eta_y = \lambda m r_y \sigma_y^{1/2} D_y^{-1/2}.$$

As a consequence of the above discussion and Theorem 4.4, we obtain the following result.

Theorem 4.5 *For a given $\epsilon > 0$, Nesterov's smooth approximation scheme applied to (43) finds an ϵ-optimal solution of problem (43) and its dual, and hence of problem (34), in a number of iterations which does not exceed*

$$2r_x \left[\epsilon^{-1/2}\|\Lambda\| + \lambda m \epsilon^{-1}\sqrt{\log n}\left(\sqrt{2} + \frac{2n}{n-m}\max\{1, \log n\}^{1/2}\right)\right], \tag{89}$$

where r_x is defined in (37).

Proof. By substituting the above ξ_x, ξ_y, η_x, η_y into $\Gamma(\xi_x, \xi_y, \eta_x, \eta_y)$, we can easily show that the iteration complexity of finding an ϵ-optimal solution to (43) and its dual by Nesterov's smooth approximation scheme does not exceed

$$2\sqrt{2} r_x \epsilon^{-1/2} D_x^{1/2}\|\Lambda\| + 4\lambda m \epsilon^{-1} D_V^{1/2}\left(r_x D_x^{1/2} + r_y D_y^{1/2}\right).$$

Recall from Theorem 3.5 that $r_y = nr_x/(n-m)$. Using this relation and the facts that $D_x = 1/2$, $D_y = \max\{1, \log n\}$ and $D_V = \log n$, we immediately observe that the above quantity equals the one given in (89), and hence the conclusion follows. ∎

In view of (37), we observe that the iteration complexity given in (89) is in terms of the transformed data Λ and H of problem (4). We next relate it to the original data A and B of problem (4).

Corollary 4.6 *For a given $\epsilon > 0$, Nesterov's smooth approximation scheme applied to (43) finds an ϵ-optimal solution of problem (43) and its dual, and hence of problem (34), in a number of iterations which does not exceed*

$$2\tau_\lambda \left[\epsilon^{-1/2}\|A\| + \lambda m \epsilon^{-1}\sqrt{\log n}\left(\sqrt{2} + \frac{2n}{n-m}\max\{1, \log n\}^{1/2}\right)\right],$$

where

$$\tau_\lambda := \min\left\{\frac{\|(A^TA)^{-1/2}A^TB\|_F^2}{2\lambda}, \sum_{i=1}^m \sigma_i\left((A^TA)^{-1}A^TB\right)\right\}. \tag{90}$$

Proof. We know from Subsection 3.1 that $A^TA = Q\Lambda^2Q^T$ and $H = \Lambda^{-1}Q^TA^TB$, where $Q \in \Re^{p\times p}$ is an orthonormal matrix. Using these relations, we obtain

$$\begin{aligned}
\|\Lambda\| &= \|Q\Lambda^2Q^T\|^{1/2} = \|A^TA\|^{1/2} = \|A\|, &(91)\\
\|H\|_F^2 &= \mathrm{Tr}(B^TAQ\Lambda^{-2}Q^TA^TB) = \mathrm{Tr}(B^TA(A^TA)^{-1}A^TB) = \|(A^TA)^{-1/2}A^TB\|_F^2, &(92)\\
\sigma_i(\Lambda^{-1}H) &= \left[\lambda_i(H^T\Lambda^{-2}H)\right]^{1/2} = \left[\lambda_i(B^TAQ\Lambda^{-4}Q^TA^TB)\right]^{1/2}\\
&= \left[\lambda_i(B^TA(A^TA)^{-2}A^TB)\right]^{1/2} = \sigma_i\left((A^TA)^{-1}A^TB\right). &(93)
\end{aligned}$$

In view of identities (92) and (93), we see that the quantity r_x defined in (37) coincides with the quantity τ_λ defined in (90). The conclusion of the corollary now follows from this observation, Theorem 4.5 and identity (91). ∎

After having completely specified all the ingredients required by Nesterov's smooth approximation scheme for solving (79), we now discuss some of the computational technicalities involved in the actual implementation of the method. First, for a given $u \in U$, the optimal solution for the maximization subproblem appearing in (72) with $\eta = \epsilon/(2D_V)$ needs to be found in order to compute the gradient of $f_\eta(u)$. Using (83) and the facts that $V = \Delta^n_=(1)$ and $\phi(v) = 0$ for all $v \in V$, we see that the maximization subproblem in (72) is equivalent to

$$\max_{v \in \Delta^n_=(1)} g \bullet v - \mathrm{Tr}(v \log v)$$

for some $g \in \mathcal{S}^n$. The optimal solution for this problem has a closed-form expression which requires the computation of an eigenvalue decomposition of g (see [12]). In addition, the two subproblems (63) and (64) with $f = f_\eta$ can be reduced to form

$$\min_{u \in U} \langle h, u \rangle + p_U(u),$$

for some $h = (h_x, h_y) \in \mathcal{U}$, where $\langle h, u \rangle := h_x \bullet X + h_y \bullet Y$ for every $u = (X, Y) \in \mathcal{U}$. In view of (82) and the fact that $U = \mathcal{B}^{p\times q}_F(1) \times \Delta^n_\leq(1)$, it is easy to see that the solution of the above problem has a closed-form expression which requires the computation of an eigenvalue decomposition of h_y (see [12]).

Finally, to terminate Nesterov's smooth approximation scheme, we need to properly evaluate the primal and dual objective functions of problem (79) at any given point. In view of (15), the primal objective function $f(u)$ of (79) can be computed as

$$f(u) = \|r_x \Lambda X - H\|_F^2/2 + \lambda m \lambda_{\max}(r_x \mathcal{G}(X) - r_y Y) + \lambda r_y \mathrm{Tr}(Y), \quad \forall u = (X, Y) \in U.$$

Now, the dual objective function $g(v)$ of (79) is

$$g(v) = \min_{u=(X,Y)\in U} \left\{ \frac{1}{2}\|r_x \Lambda X - H\|_F^2 + \lambda m (r_x \mathcal{G}(X) - r_y Y) \bullet v + \lambda r_y \mathrm{Tr}(Y) \right\}, \quad \forall v \in V.$$

Recalling that $U = \mathcal{B}^{p\times q}_F(1) \times \Delta^n_\leq(1)$, we easily see that

$$g(v) = \min_{X \in \mathcal{B}^{p\times q}_F(1)} \left\{ \frac{1}{2}\|r_x \Lambda X - H\|_F^2 + \lambda m r_x \mathcal{G}(X) \bullet v \right\} + \lambda r_y \min\{\lambda_{\min}(I - mv), 0\}, \tag{94}$$

for every $v \in V$. In view of (5), the minimization problem in (94) can be written as

$$\min_{X \in \mathcal{B}^{p\times q}_F(1)} \frac{1}{2}\|r_x \Lambda X - H\|_F^2 + G \bullet X \tag{95}$$

for some $G \in \Re^{p\times q}$. We now briefly discuss how to solve (95). For $\xi \geq 0$, let

$$X(\xi) = (r_x^2 \Lambda^2 + \xi I)^{-1}(r_x \Lambda H - G), \qquad \Psi(\xi) = \|X(\xi)\|_F^2 - 1$$

for $\xi \geq 0$. If $\Psi(0) \leq 0$, then clearly $X(0)$ is the optimal solution of problem (95). Otherwise, the optimal solution of problem (95) is equal to $X(\xi^*)$, where ξ^* is the root of the equation $\Psi(\xi) = 0$.

4.2.2 Implementation details of Nesterov's smooth approximation scheme for (45)

The implementation details of Nesterov's smooth approximation scheme (see Subsection 4.1.2) for solving formulation (45) are addressed in this subsection. In particular, we specify in the context of this formulation the norms, prox-functions and subproblems that appear in Nesterov's smooth approximation scheme.

For the purpose of our implementation, we reformulate problem (45) into the problem

$$\min_{X\in\mathcal{B}_F^{p\times q}(1)} \max_{W\in\Omega} \left\{ \frac{1}{2}\|r_x\Lambda X - H\|_F^2 + \lambda m r_x \mathcal{G}(X)\bullet W \right\} \tag{96}$$

obtained by scaling the variables X of (45) as $X \leftarrow X/r_x$. From now on, we will focus on formulation (96) rather than (45).

Let $n := p+q$, $u := X$, $v := W$ and define

$$\begin{aligned} U &:= \mathcal{B}_F^{p\times q}(1) \subseteq \Re^{p\times q} =: \mathcal{U}, \\ V &:= \Omega \subseteq \mathcal{S}^n =: \mathcal{V} \end{aligned}$$

and

$$\hat{f}(u) := \frac{1}{2}\|r_x\Lambda u - H\|_F^2, \quad \mathcal{E}u := \lambda m r_x \mathcal{G}(u), \quad \phi(v) = 0, \quad \forall u \in U, \ \forall v \in V,$$

where Ω is defined in (46). Then, in view of these definitions, it follows that problem (96) can be expressed in the form (70), and thus be suitably solved by Nesterov's smooth approximation scheme described in Subsection 4.1.2. In the following paragraphs we specify the norms $\|\cdot\|_{\mathcal{U}}$ and $\|\cdot\|_{\mathcal{V}}$ for $\mathcal{U}$ and $\mathcal{V}$, suitable prox-functions $p_U(\cdot)$ and $p_V(\cdot)$ for U and V and the subproblems that appear in Nesterov's smooth approximation scheme for solving (96).

The norms $\|\cdot\|_{\mathcal{U}}$ and $\|\cdot\|_{\mathcal{V}}$ for the spaces $\mathcal{U}$ and $\mathcal{V}$ are defined as

$$\begin{aligned} \|u\|_{\mathcal{U}} &:= \|u\|_F, \quad \forall u \in \mathcal{U}, \\ \|v\|_{\mathcal{V}} &:= \|v\|_F, \quad \forall v \in \mathcal{V}, \end{aligned}$$

and the prox-functions $p_U(\cdot)$ and $p_V(\cdot)$ for the sets U and V are defined as

$$p_U(u) \ := \ \|u\|_F^2/2, \quad \forall u \in U = \mathcal{B}_F^{p\times q}(1), \tag{97}$$

$$p_V(v) \ := \ \mathrm{Tr}(v \log v) + \log n, \quad \forall v \in V = \Omega. \tag{98}$$

We can easily see that $p_U(\cdot)$ (resp., $p_V(\cdot)$) is a strongly differentiable convex function on U (resp., V) with modulus $\sigma_U = 1$ (resp., $\sigma_V = m$) with respect to the norm $\|\cdot\|_{\mathcal{U}}$ (resp., $\|\cdot\|_{\mathcal{V}}$). Also, it is easy to verify that $\min\{p_U(u) : u \in \mathcal{U}\} = \min\{p_V(v) : v \in \mathcal{V}\} = 0$ and that

$$\begin{aligned} D_U &:= \max_{u\in U} p_U(u) = 1/2, \\ D_V &:= \max_{v\in V} p_V(v) = \log(n/m). \end{aligned}$$

Recall that Nesterov's smooth approximation scheme consists of applying Nesterov's optimal method to problem (75). Moreover, by Theorem 4.3(b), the objective function f_η of (75) is L_η-Lipschitz differentiable, where L_η is given by (74). We will now show how to compute L_η, which can then be used as input for the optimal method applied to (75). First, we observe that for any u, $h \in \mathcal{U}$,

$$\|\nabla \hat{f}(u+h) - \nabla \hat{f}(u)\|_{\mathcal{U}}^* = r_x^2 \|\Lambda^2 h\|_F \leq r_x^2 \|\Lambda\|^2 \|h\|_F = r_x^2 \|\Lambda\|^2 \|h\|_{\mathcal{U}},$$

and hence, $\hat{f}(u)$ is $L_{\hat{f}}$-Lipschitz-differentiable on U with respect to $\|\cdot\|_{\mathcal{U}}$, where

$$L_{\hat{f}} := r_x^2 \|\Lambda\|^2. \tag{99}$$

Using (6) and the relation $\|\mathcal{G}(u)\|_\infty = \sigma_1(u)$ for every $u \in \Re^{p \times q}$, we have

$$\begin{aligned}
\|\mathcal{E}\|_{\mathcal{U},\mathcal{V}} &= \max\left\{\|\lambda m r_x \mathcal{G}(u)\|_{\mathcal{V}}^* : \ u \in \mathcal{U}, \|u\|_{\mathcal{U}} \leq 1\right\}, \\
&= \lambda m r_x \max\left\{\|\mathcal{G}(u)\|_F : \ u \in \mathcal{U}, \|u\|_F \leq 1\right\}, \\
&= \lambda m r_x \max\left\{\sqrt{2}\|u\|_F : \ u \in \mathcal{U}, \|u\|_F \leq 1\right\} = \sqrt{2}\lambda m r_x.
\end{aligned} \tag{100}$$

Relations (74), (99), (100) and the fact that $\sigma_V = m$ then imply that $L_\eta = r_x^2(\|\Lambda\|^2 + 2\lambda^2\eta^{-1}m)$.

As a consequence of the above discussion and Theorem 4.4, we obtain the following result.

Theorem 4.7 *For a given $\epsilon > 0$, Nesterov's smooth approximation scheme applied to (45) finds an ϵ-optimal solution of problem (45) and its dual, and hence of problem (34), in a number of iterations which does not exceed*

$$2r_x\sqrt{\frac{4\lambda^2 m \log(n/m)}{\epsilon^2} + \frac{\|\Lambda\|^2}{\epsilon}}, \tag{101}$$

where r_x is defined in (37).

Note that the iteration-complexity of Theorem 4.7 is always smaller than the one derived in Theorem 4.5. Observe also that the iteration complexity (101) is expressed in terms of the transformed data Λ and H of problem (4). We next relate it to the original data A and B of problem (4).

Corollary 4.8 *For a given $\epsilon > 0$, Nesterov's smooth approximation scheme applied to (45) finds an ϵ-optimal solution of problem (45) and its dual, and hence of problem (34), in a number of iterations which does not exceed*

$$2\tau_\lambda\sqrt{\frac{2\lambda^2 m \log(n/m)}{\epsilon^2} + \frac{\|A\|^2}{\epsilon}},$$

where τ_λ is given by (90).

Proof. The proof is similar to that of Corollary 4.6. ∎

After having completely specified all the ingredients required by Nesterov's smooth approximation scheme for solving (96), we now discuss some of the computational technicalities involved in the actual implementation of the method. First, for a given $u \in U$, the optimal solution for the maximization subproblem appearing in (72) with $\eta = \epsilon/(2D_V)$ needs to be found in order to compute the gradient of $f_\eta(u)$. Using (98) and the facts that $V = \Delta^n_=(1)$ and $\phi(v) = 0$ for all $v \in V$, we see that the maximization subproblem in (72) is equivalent to

$$\max_{v \in \Omega} \mathcal{G}(h) \bullet v - \mathrm{Tr}(v \log v) \tag{102}$$

for some $h \in \Re^{p\times q}$, where Ω is given in (46).

We now present an efficient approach for solving (102). First, we compute a singular value decomposition of h, i.e., $h = \tilde{U}\Sigma\tilde{V}^T$, where $\tilde{U} \in \Re^{p\times m}$, $\tilde{V} \in \Re^{q\times m}$ and Σ are such that

$$\tilde{U}^T\tilde{U} = I, \quad \Sigma = \mathrm{Diag}(\sigma_1(h), \ldots, \sigma_m(h)), \quad \tilde{V}^T\tilde{V} = I,$$

where $\sigma_1(h), \ldots, \sigma_m(h)$ are the $m = \min(p,q)$ singular values of h. Let ξ_i and η_i denote the ith column of $\tilde{U}$ and $\tilde{V}$, respectively. It is easy to see that

$$f^i = \frac{1}{\sqrt{2}}\begin{pmatrix}\eta_i\\ \xi_i\end{pmatrix}, \ i = 1, \ldots, m; \quad f^{m+i} = \frac{1}{\sqrt{2}}\begin{pmatrix}\eta_i\\ -\xi_i\end{pmatrix}, \ i = 1, \ldots, m, \tag{103}$$

are orthonormal eigenvectors of $\mathcal{G}(h)$ with eigenvalues $\sigma_1(h), \ldots, \sigma_m(h), -\sigma_1(h), \ldots, -\sigma_m(h)$, respectively. Now let $f^i \in \Re^n$ for $i = 2m+1, \ldots, n$ be such that the matrix $F := (f^1, f^2, \ldots, f^n)$ satisfies $F^TF = I$. It is well-known that the vectors $f^i \in \Re^n$, $i = 2m+1, \ldots, n$, are eigenvectors of $\mathcal{G}(h)$ corresponding to the zero eigenvalue (e.g., see [2]). Thus, we obtain the following eigenvalue decomposition of $\mathcal{G}(h)$:

$$\mathcal{G}(h) = F\mathrm{Diag}(a)F^T, \quad a = (\sigma_1(h), \ldots, \sigma_m(h), -\sigma_1(h), \ldots, -\sigma_m(h), 0, \ldots, 0)^T.$$

Using this relation and (46), it is easy to see that the optimal solution of (102) is $v^* = F\mathrm{Diag}(w^*)F^T$, where $w^* \in \Re^n$ is the unique optimal solution of the problem

$$\begin{array}{ll} \max & a^Tw - w^T\log(w) \\ \text{s.t.} & e^Tw = 1, \\ & 0 \le w \le e/m. \end{array} \tag{104}$$

It can be easily shown that $w_i^* = \min\{\exp(a_i - 1 - \xi^*), 1/m\}$, where ξ^* is the unique root of the equation

$$\sum_{i=1}^n \min\{\exp(a_i - 1 - \xi), 1/m\} - 1 = 0.$$

Let $\vartheta := \min\{\exp(-1-\xi^*), 1/m\}$. In view of the above formulas for a and w^* , we immediately see that

$$w^*_{2m+1} = w^*_{2m+2} = \cdots = w^*_n = \vartheta. \tag{105}$$

Further, using the fact that $FF^T = I$, we have

$$\sum_{i=2m+1}^n f^i(f^i)^T = I - \sum_{i=1}^{2m} f^i(f^i)^T.$$

Using this result and (105), we see that the optimal solution v^* of (102) can be efficiently computed as

$$v^* = F\mathrm{Diag}(w^*)F^T = \sum_{i=1}^n w_i^* f^i(f^i)^T = \vartheta I + \sum_{i=1}^{2m}(w_i^* - \vartheta)f^i(f^i)^T,$$

where the scalar ϑ is defined above and the vectors $\{f^i : i = 1, \ldots 2m\}$ are given by (103).

In addition, the two subproblems (63) and (64) with $f = f_\eta$ can be reduced to form

$$\min_{u \in U} h \bullet u + p_U(u),$$

for some $h \in \mathcal{U}$. In view of (97) and the fact that $U = \mathcal{B}_F^{p\times q}(1)$, it is easy to see that the solution of the above problem has a closed-form expression which can be readily found.

Finally, to terminate Nesterov's smooth approximation scheme, we need to properly evaluate the primal and dual objective functions of problem (96) at any given point. In view of (29) and (46), the primal objective function $f(u)$ of (96) can be computed as

$$f(u) = \frac{1}{2}\|r_x \Lambda u - H\|_F^2 + \lambda r_x \sum_{i=1}^{m} \sigma_i(u), \quad \forall u \in U.$$

Now, the dual objective function $g(v)$ of (96) is

$$g(v) = \min_{u \in \mathcal{B}_F^{p\times q}(1)} \frac{1}{2}\|r_x \Lambda u - H\|_F^2 + \lambda m r_x \mathcal{G}(u) \bullet v, \quad \forall v \in V. \tag{106}$$

The minimization problem in (106) can be solved in the same way as the one which appears in (94).

4.2.3 Implementation details of Nesterov's smooth optimal method for (47)

The implementation details of Nesterov's smooth method (see Subsection 4.1.1) for solving formulation (47) (that is, the dual of (45)) are addressed in this subsection. In particular, we specify in the context of this formulation the norms, prox-functions and subproblems that appear in Nesterov's smooth method.

For the purpose of our implementation, we reformulate problem (47) into the problem

$$\min_{W \in \Omega} \max_{X \in \mathcal{B}_F^{p\times q}(1)} \left\{ -\lambda m r_x \mathcal{G}(X) \bullet W - \frac{1}{2}\|r_x \Lambda X - H\|_F^2 \right\}, \tag{107}$$

obtained by scaling the variable X of (47) by setting $X \leftarrow X/r_x$, and premultiplying the resulting formulation by negative one. From now on, we will focus on formulation (107) rather than (47).

Let $n := p + q$, $u := W$, $v := X$ and define

$$U := \Omega \subseteq \mathcal{S}^n =: \mathcal{U},$$
$$V := \mathcal{B}_F^{p\times q}(1) \subseteq \Re^{p\times q} =: \mathcal{V},$$

and

$$\phi(u, v) := -\lambda m r_x \mathcal{G}(v) \bullet u - \frac{1}{2}\|r_x \Lambda v - H\|_F^2, \quad \forall (u, v) \in U \times V,$$

where Ω is defined in (46). Invoking that $\Lambda \succ 0$, we observe that $\phi(u, v)$ is strictly concave in v for any fixed u. Then, in view of this fact and the above definitions, it follows that problem (107) can be expressed in the form (54)-(55) and satisfies all assumptions imposed on (54). Thus it can be suitably solved by Nesterov's smooth method described in Subsection 4.1.1. In the following paragraphs we specify the norms $\|\cdot\|_{\mathcal{U}}$ and $\|\cdot\|_{\mathcal{V}}$ for $\mathcal{U}$ and $\mathcal{V}$, suitable prox-function $p_U(\cdot)$ for U and the subproblems that appear in Nesterov's smooth method for solving (107).

The norms $\|\cdot\|_{\mathcal{U}}$ and $\|\cdot\|_{\mathcal{V}}$ for the spaces $\mathcal{U}$ and $\mathcal{V}$ are defined as

$$\|u\|_{\mathcal{U}} := \|u\|_F, \quad \forall u \in \mathcal{U},$$
$$\|v\|_{\mathcal{V}} := \|v\|_F, \quad \forall v \in \mathcal{V},$$

and the prox-function $p_U(\cdot)$ for the set U is defined as

$$p_U(u) = \mathrm{Tr}(u \log u) + \log n, \quad \forall u \in U = \Omega. \tag{108}$$

We can easily see that $p_U(\cdot)$ is a strongly differentiable convex function on U with modulus $\sigma_U = m$ with respect to the norm $\|\cdot\|_{\mathcal{U}}$. Also, it is easy to verify that $\min\{p_U(u) : u \in \mathcal{U}\} = 0$ and that

$$D_U := \max_{u \in U} p_U(u) = \log(n/m).$$

Recall that the Lipschitz constant L of $\nabla f(u)$ is used as input for Nesterov's smooth method applied to (54), where $f(u)$ is the objective function of (54). In the context of formulation (107), $f(u)$ is given by

$$f(u) = \max_{v \in V} -\lambda m r_x \mathcal{G}(v) \bullet u - \frac{1}{2}\|r_x \Lambda v - H\|_F^2, \ \forall u \in U. \tag{109}$$

We will now show how to compute L. Let us define

$$\begin{aligned} \mathcal{E}v \quad &:= \quad -\lambda m r_x \mathcal{G}(v), \quad \forall v \in \mathcal{V}, \\ p_V(v) \quad &:= \quad \frac{1}{2}\|r_x \Lambda v - H\|_F^2, \quad \forall v \in \mathcal{V}. \end{aligned}$$

As shown in (100), we know that $\|\mathcal{E}\|_{\mathcal{U},\mathcal{V}} = \lambda m r_x$. Also, we have

$$\langle \nabla^2 p_V(v) h, h\rangle = r_x^2 \|\Lambda h\|_F^2 \geq r_x^2 \min_j \Lambda_{jj}^2 \|h\|_F^2 = r_x^2 \min_j \Lambda_{jj}^2 \|h\|_{\mathcal{V}}^2, \quad \forall h \in \mathcal{V},$$

and hence, $p_V(\cdot)$ is a strongly convex function on V with modulus $\sigma_V = r_x^2 \min_j \Lambda_{jj}^2$ with respect to the norm $\|\cdot\|_{\mathcal{V}}$. Using these results and Theorem 4.3 (b), we conclude that $f(u)$ is L-Lipschitz-differentiable, where

$$L = \|\mathcal{E}\|_{U,V}^2/\sigma_V = (\sqrt{2}\lambda m r_x)^2/(r_x^2 \min_j \Lambda_{jj}^2) = 2\lambda^2 m^2 \|\Lambda^{-1}\|^2.$$

As a consequence of the above discussion and Theorem 4.2, we obtain the following result.

Theorem 4.9 *For a given $\epsilon > 0$, Nesterov's smooth method applied to (47) finds an ϵ-optimal solution of problem (47) and its dual, and hence of problem (34), in a number of iterations which does not exceed*

$$\frac{2\sqrt{2}\lambda\|\Lambda^{-1}\|}{\sqrt{\epsilon}}\sqrt{m \log(n/m)}. \tag{110}$$

Before proceeding, we compare the iteration-complexities obtained in Theorems 4.5, 4.7 and 4.9. First, note that the first and the second ones are $\mathcal{O}(\epsilon^{-1})$ while the latter one is $\mathcal{O}(\epsilon^{-1/2})$. Hence, Nesterov's smooth method applied to (47) should be superior than the other two methods for small values of ϵ. In fact, we will see later in Section 5 that this is indeed the case. Second, note that as the parameter λ approaches zero, the iteration-complexity (110) converges to zero while the iteration-complexities (89) and (101) remains bounded away from zero. Third, as the parameter λ grows to infinity, the iteration-complexity (110) converges to infinity while the iteration-complexities (89) and (101) remains bounded.

We observe that the iteration complexity given in (110) is in terms of the transformed data of problem (4). We next relate it to the original data of problem (4).

Corollary 4.10 *For a given $\epsilon > 0$, Nesterov's smooth method applied to (47) finds an ϵ-optimal solution of problem (47) and its dual, and hence of problem (34), in a number of iterations which does not exceed*

$$\frac{2\sqrt{2}\lambda\|(A^TA)^{-1/2}\|}{\sqrt{\epsilon}}\sqrt{m\log(n/m)}.$$

Proof. We know from Subsection 3.1 that $A^TA = Q\Lambda^2Q^T$, where $Q \in \Re^{p\times p}$ is an orthonormal matrix. Using this relation, we have

$$\|\Lambda^{-1}\| = \|\Lambda^{-2}\|^{1/2} = \|(A^TA)^{-1}\|^{1/2} = \|(A^TA)^{-1/2}\|.$$

The conclusion immediately follows from this identity, (110) and Theorem 4.9. ∎

It is interesting to note that the iteration-complexity of Corollary 4.10 depends on the data matrix A but not on B. Moreover, as seen from the discussion below, the arithmetic operation cost per iteration of Nesterov's smooth method applied to (47) is $\mathcal{O}(n^3)$ dominated by the eigenvalue decomposition of an $n \times n$ matrix. Thus, the overal arithmetic operation cost for Nesterov's smooth method applied to (47) is

$$\mathcal{O}\left(\frac{\lambda\|(A^TA)^{-1/2}\|}{\sqrt{\epsilon}}n^3\sqrt{m\log(n/m)}\right).$$

After having completely specified all the ingredients required by Nesterov's smooth method for solving (107), we now discuss some of the computational technicalities involved in the actual implementation of the method. First, for a given $u \in U$, the optimal solution for the maximization subproblem appearing in (109) needs to be found in order to compute the gradient of $\nabla f(u)$. Using the fact that $V = \mathcal{B}_F^{p\times q}(1)$, the maximization problem in (109) can be solved in a similar way as the minimization problem appearing in (94).

In addition, each iteration of Nesterov's smooth method requires solving the two subproblems (63) and (64). In view of (109), it is easy to observe that for every $u \in U$, $\nabla f(u) = \mathcal{G}(h)$ for some $h \in \Re^{p\times q}$. Using this result and (108), we can see that subproblem (63) is equivalent to (102), and thus it can be efficiently solved by the approach proposed in Subsection 4.2.2. Also, using (60) and (108), we observe that subproblem (64) is equivalent to

$$\min_{u\in\Omega} g \bullet u + \mathrm{Tr}(u\log u)$$

for some $g \in \mathcal{S}^n$. In view of (46), it is easy to see that the solution of the above problem has a closed-form expression which requires the computation of an eigenvalue decomposition of g.

Finally, to terminate Nesterov's smooth method described in 4.1.1, we need to properly evaluate the primal and dual objective functions of problem (107) at any given point. Since problem (107) is the dual of (96), the primal and dual objective functions of (107) can be evaluated in the same way as those of (96).

Table 1: Comparison of SA1, SA2 and SM

Problem	Iter			Obj			Time		
(p, q)	SA1	SA2	SM	SA1	SA2	SM	SA1	SA2	SM
(20, 10)	5622	39	2	4.11	4.09	4.09	11.89	0.01	0.00
(40, 20)	23658	92	1	8.40	8.41	8.44	234.66	0.11	0.01
(60, 30)	56449	145	1	13.45	13.46	13.49	1566.82	0.38	0.03
(80, 40)	100252	211	1	17.64	1.765	17.68	5679.14	1.02	0.05
(100, 50)	160067	271	1	22.41	22.42	22.46	16488.88	2.21	0.10
(120, 60)	231510	330	1	26.86	26.87	26.91	88866.22	4.15	0.15

5 Computational results

In this section, we report the results of our computational experiments which compare the performance of the three approaches discussed in Subsections 4.2.1, 4.2.2 and 4.2.3 for solving problem (34). We also compare the performance of the best of these three approaches with the interior point method implemented in SDPT3 version 4.0 (beta) [15] on a set of randomly generated instances.

The random instances of (34) used in our experiments were generated as follows. We first randomly generated matrices $A \in \Re^{n\times p}$ and $B \in \Re^{n\times q}$, where $p = 2q$ and $n = 10q$, with entries uniformly distributed in $[0, 1]$ for different values of q. We then computed H and Λ for (34) according to the procedures described in Subsection 3.1 and set the parameter λ in (34) to one. In addition, all computations were performed on an Intel Xeon 5320 CPU (1.86GHz) and 12GB RAM running Red Hat Enterprise Linux 4 (kernel 2.6.9).

In the first experiment, we compared the performance of the three approaches discussed in Subsections 4.2.1, 4.2.2 and 4.2.3 for solving problems (79), (96) and (107), respectively. For convenience of presentation, we label these three approaches as SA1, SA2 and SM, respectively. The codes for them were written in C. These three methods terminate once the duality gap is less than $\epsilon = 0.1$. The initial points for these methods are set to be $u_0 = (0, I/n)$, $u_0 = 0$, and $u_0 = I/n$, respectively.

The performance of methods SA1, SA2 and SM for our randomly generated instances are presented in Table 1. The problem size (p, q) is given in column one. The numbers of iterations of SA1, SA2 and SM are given in columns two to four, and the objective function values are given in columns five to seven, and the CPU times (in seconds) are given in the last three columns, respectively. From Table 1, we conclude that methods SA2 and SM substantially outperforms method SA1.

In the second experiment, we compared the performance of methods SA2 and SM for solving problems (96) and (107) on randomly generated instances of larger sizes than those used in the first experiment. Both methods were terminated once the duality gap is less than $\epsilon = 0.1$. The initial points for these methods were set to be $u_0 = 0$, and $u_0 = I/n$, respectively. The performance of these two methods are presented in Table 2. The problem size (p, q) is given in column one. The numbers of iterations of SA2 and SM are given in columns two to three, and the objective function values are given in columns four to six, and the CPU times (in seconds) are given in the last two columns, respectively. From Table 2, we conclude that SM, namely, Nesterov's smooth method applied to (96), substantially outperforms SA2, that is, Nesterov's smooth approximation approach applied to (107).

The above two experiments show that SM substantially outperforms the other two first-order methods, namely, SA1 and SA2 for solving problem (34). In the third experiment, we compare the performance of method SM with the interior point method implemented in SDPT3 version 4.0

Table 2: Comparison of SA2 and SM

Problem	Iter		Obj		Time	
(p, q)	SA2	SM	SA2	SM	SA2	SM
(200, 100)	610	1	45.78	45.82	29.60	0.91
(400, 200)	1310	1	91.80	91.83	432.92	8.36
(600, 300)	2061	1	137.04	137.07	2155.76	31.23
(800, 400)	2848	1	183.65	183.68	7831.09	76.75
(1000, 500)	3628	1	229.66	229.69	21128.70	156.68
(1200, 600)	4436	1	275.76	275.78	47356.32	276.64
(1400, 700)	5280	1	321.14	321.16	98573.73	456.61
(1600, 800)	6108	1	367.13	367.16	176557.49	699.47

Table 3: Comparison of SM and SDPT3

Problem	Iter		Obj		Time		Memory	
(p, q)	SM	SDPT3	SM	SDPT3	SM	SDPT3	SM	SDPT3
(20, 10)	26928	17	4.066570508	4.066570512	26.7	5.9	2.67	279
(40, 20)	28702	15	8.359912031	8.359912046	122.7	77.9	2.93	483
(60, 30)	22209	15	13.412029944	13.412029989	242.7	507.7	3.23	1338
(80, 40)	21923	15	17.596671337	17.596671829	496.7	2209.8	3.63	4456
(100, 50)	19933	19	22.368563640	22.368563657	813.5	8916.1	4.23	10445
(120, 60)	18302	N/A	26.823206950	N/A	1216.4	N/A	4.98	> 16109

(beta) [15] for solving the cone programming reformulation (40). It is worth mentioning that code SDPT3 uses Matlab as interface to call several C subroutines to handle all its heavy computational tasks. SDPT3 can be suitably applied to solve a standard cone programming with the underlying cone represented as a Cartesian product of nonnegative orthant, second-order cones, and positive semidefinite cones. The method SM terminates once the duality gap is less than $\epsilon = 10^{-8}$, and SDPT3 terminates once the relative accuracy is less than 10^{-8}.

The performance of SM and SDPT3 for our randomly generated instances are presented in Table 3. The problem size (p, q) is given in column one. The numbers of iterations of SM and SDPT3 are given in columns two to three, and the objective function values are given in columns four to six, CPU times (in seconds) are given in columns seven to eight, and the amount of memory (in mega bytes) used by SM and SDPT3 are given in the last two columns, respectively. The symbol "N/A" means "not available". The computational result of SDPT3 for the instance with $(p, q) = (120, 60)$ is not available since it ran out of the memory in our machine (about 15.73 giga bytes). We conclude from this experiment that the method SM, namely, Nesterov's smooth method, substantially outperforms SDPT3. Moreover, SDPT3 requires much more memory than SM. For example, for the instance with $(p, q) = (100, 50)$, SDPT3 needs 10445 mega ($\approx$ 10.2 giga) bytes of memory, but SM only requires about 4.23 mega bytes of memory; for the instance with $(p, q) = (120, 60)$, SDPT3 needs at least 16109 mega ($\approx$ 15.73 giga) bytes of memory, but SM only requires about 4.98 mega bytes of memory.

References

[1] T. W. Anderson. Estimating linear restriction on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics*, 22:327–351, 1951.

[2] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization: Analysis, algorithms, Engineering Applications*. MPS-SIAM Series on Optimization. SIAM, Philadelphia, 2001.

[3] D. Bertsekas. *Nonlinear Programming*. Athena Scientific, New York, second edition, 1999.

[4] L. Breiman. Heuristics of instability and stabilization in model selection. *Annals of Statistics*, 24:2350–2383, 1996.

[5] R. Brooks and M. Stone. Joint continuum regression for multiple predictands. *Journal of American Statistics Association*, 89:1374–1377, 1994.

[6] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization algorithms I*, volume 305 of *Comprehensive Study in Mathematics*. Springer-Verlag, New York, 1993.

[7] H. Hotelling. The most predictable criterion. *Journal of Educational Psychology*, 26:139–142, 1935.

[8] H. Hotelling. Relations between two sets of variables. *Biometrika*, 28:321–377, 1936.

[9] A. Izenman. Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis*, 5:248–264, 1975.

[10] Z. Lu. Smooth optimization approach for covariance selection. Manuscript, Department of Mathematics, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada, June 2007.

[11] W. Massy. Principle components regression with exploratory statistical research. *Journal of American Statistics Association*, 60:234–246, 1965.

[12] Y. E. Nesterov. A method for unconstrained convex minimization problem with the rate of convergence $O(1/k^2)$. *Doklady AN SSSR*, 269:543–547, 1983. translated as Soviet Math. Docl.

[13] Y. E. Nesterov. Smooth minimization of nonsmooth functions. *Mathematical Programming*, 103:127–152, 2005.

[14] G. Reinsel and R. Velu. *Multivariate Reduced-rank Regression: Theory and Application*. Springer, New York, 1998.

[15] K.C. Toh, R. H. Tütüncü, and M. J. Todd. On the implementation and usage of sdpt3 - a matlab software package for semidefinite-quadratic-linear programming, version 4.0. Manuscript, Department of Mathematics, National University of Singapore, July 2006.

[16] H. Wold. Soft modeling by latent variables: the nonlinear iterative partial least squares approach. In *In Perspectives in Probability and Statistics: Papers in Honor of M. S. Bartlett*. Academic Press, New York, 1975.

[17] M. Yuan, A. Ekici, Z. Lu, and R. D. C. Monteiro. Dimension reduction and coefficient estimation in multivariate linear regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(3):329–346, 2007.