Math. Program., Ser. A
DOI 10.1007/s10107-010-0350-1

FULL LENGTH PAPER

Convex optimization methods for dimension reduction and coefficient estimation in multivariate linear regression

Zhaosong Lu · Renato D. C. Monteiro · Ming Yuan

Received: 14 January 2008 / Accepted: 24 March 2010

Abstract In this paper, we study convex optimization methods for computing the nuclear (or, trace) norm regularized least squares estimate in multivariate linear regression. The so-called factor estimation and selection method, recently proposed by Yuan et al. (J Royal Stat Soc Ser B (Statistical Methodology) 69(3):329–346, 2007) 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 nuclear norm constraint. In this paper, we explore a variant due to Tseng of Nesterov's smooth method 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 variant of Nesterov's smooth method generally outperforms the interior point method implemented in SDPT3 version 4.0 (beta) (Toh et al. On the implementation and usage of sdpt3—a matlab software package for semidefinite-quadratic-linear programming, version 4.0. Manuscript, Department of

Zhaosong Lu was supported in part by SFU President's Research Grant and NSERC Discovery Grant. Renato D. C. Monteiro was supported in part by NSF Grants CCF-0430644, CCF-0808863 and CMMI-0900094 and ONR Grants N00014-05-1-0183 and N00014-08-1-0033. Ming Yuan was supported in part by NSF Grants DMS-0624841 and DMS-0706724.

Z. Lu (✉)
Department of Mathematics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada
e-mail: zhaosong@sfu.ca

R. D. C. Monteiro · M. Yuan
School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, USA
e-mail: monteiro@isye.gatech.edu

M. Yuan
e-mail: myuan@isye.gatech.edu

Published online: 18 April 2010

Mathematics, National University of Singapore (2006)) substantially. Moreover, the former method is much more memory efficient.

Keywords Cone programming · Smooth saddle point problem · First-order method · Multivariate linear regression · Nuclear or trace norm · Dimension reduction

Mathematics Subject Classification (2000) 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 l observations on q responses $\mathbf{b} = (b_1, \ldots, b_q)'$ and p explanatory variables $\mathbf{a} = (a_1, \ldots, a_p)'$, and

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

where $B = (\mathbf{b}^1, \ldots, \mathbf{b}^l)' \in \Re^{l\times q}$ and $A = (\mathbf{a}^1, \ldots, \mathbf{a}^l)' \in \Re^{l\times p}$ consists of the data of responses and explanatory variables, respectively, $U \in \Re^{p\times q}$ is the coefficient matrix, $E = (\mathbf{e}^1, \ldots, \mathbf{e}^l)' \in \Re^{l\times q}$ 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 $r \leq \min\{p, q\}$. 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 [9,10], reduced rank [1,11,18], principal components [14], partial least squares [24] and joint continuum regression [6] 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 [5]: small changes in the data can result in very different estimates.

Recently, Yuan et al. [25] 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 nuclear or 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:

$$\min_U \left\{ \mathrm{Tr}((B - AU)W(B - AU)') : \sum_{i=1}^{m} \sigma_i(U) \leq M \right\}, \tag{3}$$

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

$$\min_U \left\{ \|B - AU\|_F^2 : \sum_{i=1}^{m} \sigma_i(U) \leq M \right\}. \tag{4}$$

It is easy to see that, when the variable U in (4) is further required to be a diagonal matrix, problem (4) reduces to the well-known LASSO problem introduced in [19] and hence to a model that is known to induce sparsity on its optimal solution. More generally, it is shown in Yuan et al. [25] 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. Recently, Bach [2] further provided necessary and sufficient conditions for rank consistency of nuclear norm minimization with the square loss by considering the Lagrangian relaxation of (4). He also proposed a Newton-type method for finding an approximate solution to the latter problem, but his method is only suitable for the problems where p and q are not too large.

In addition, the nuclear norm relaxation has been used in literature for rank minimization problem. In particular, Fazel et al. [7] considered minimizing the rank of a matrix U subject to $U \in \mathcal{C}$, where $\mathcal{C}$ is a closed convex set. They proposed a convex relaxation to this problem by replacing the rank of U by the nuclear norm of U. Recently, Recht et al. [17] showed that under some suitable conditions, such a convex relaxation is tight when $\mathcal{C}$ is an affine manifold. The authors of Recht et al. [17] also

discussed some first- and second-order optimization methods for solving the nuclear norm relaxation problem.

The goal of this paper is to explore convex optimization methods, namely, a variant due to Tseng [21] of Nesterov's smooth method [15,16], 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 variant of Nesterov's smooth method generally outperforms the interior point method implemented in the code SDPT3 version 4.0 (beta) [20] substantially, and that the former method requires much less memory than the latter one.

The rest of this paper is organized as follows. In Sect. 1.1, we introduce the notation that is used throughout the paper. In Sect. 2, we present some technical results that are used in our presentation. In Sect. 3, we provide a simplification for problem (4), and present cone programming and smooth saddle point reformulations for it. In Sect. 4, we review a variant of Nesterov's smooth method [15,16,21] and discuss the details of its implementation for solving the aforementioned smooth saddle point reformulations of (4). In Sect. 5, we present computational results comparing a well-known second-order interior-point method applied to the aforementioned cone programming reformulations of (4) with the variant of Nesterov's smooth method for solving smooth saddle point reformulations of (4). Finally, we present some concluding remarks in Sect. 6 and state some additional technical results in the Appendix.

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 $i = 1, \ldots, n$, 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 $\sigma_i(X) = [\lambda_i(X^T X)]^{1/2} = [\lambda_i(X X^T)]^{1/2}$ 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

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

The following sets are used throughout the paper:

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

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 Some results on eigenvalues and singular values

In this subsection, we establish some technical results about eigenvalues and singular values which will be used in our presentation.

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

Lemma 2.1 *For any $Z \in \mathcal{S}^n$ and scalars $\alpha > 0$ and $\beta \in \Re$, the following statements hold:*

$$\lambda_{\max}(Z) = \max_{W\in\Delta_=^n(1)} Z \bullet W, \tag{8}$$

$$[\alpha\lambda_{\max}(Z) + \beta]^+ = \max_{W\in\Delta_\le^n(1)} \alpha Z \bullet W + \beta \mathrm{Tr}(W). \tag{9}$$

Proof Identity (8) is well-known. We have

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

where the third equality is due to (8) and the fourth equality is due to the fact that tW takes all possible values in $\Delta^n_\leq(1)$ 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 2.2 *Let $Z\in\mathcal{S}^n$ and integer $1\leq k\leq n$ be given. Then, the following statements hold:*

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

$$\sum_{i=1}^k \lambda_i(Z)\leq t \;\Leftrightarrow\; \begin{cases} t-ks-\mathrm{Tr}(Y) & \geq 0,\\ Y-Z+sI & \geq 0,\\ Y & \geq 0,\end{cases}$$

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

(b) *The following identities hold*:

$$\sum_{i=1}^k \lambda_i(Z) = \min_{Y\in\mathcal{S}^n_+}\max_{W\in\Delta^n_=(1)} k(Z-Y)\bullet W+\mathrm{Tr}(Y) \tag{10}$$
$$= \max_{W\in\mathcal{S}^n}\{Z\bullet W : \mathrm{Tr}(W)=k,\ 0\preceq W\preceq I\}. \tag{11}$$

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

$$\left[\alpha\sum_{i=1}^k\lambda_i(Z)+\beta\right]^+$$
$$= \min_{Y\in\mathcal{S}^n_+}\max_{W\in\Delta^n_\leq(1)} k(\alpha Z-Y)\bullet W+[\beta+\mathrm{Tr}(Y)]\,\mathrm{Tr}(W) \tag{12}$$
$$= \max_{W\in\mathcal{S}^n,\, t\in\Re}\{\alpha Z\bullet W+\beta t : \mathrm{Tr}(W)=tk,\ 0\preceq W\preceq tI,\ 0\leq t\leq 1\}. \tag{13}$$

Proof

(a) This statement is proved on pages 147–148 of Ben-Tal and Nemirovski [3].

(b) Statement (a) clearly implies that

$$\sum_{i=1}^k \lambda_i(Z) = \min_{s\in\Re,\, Y\in\mathcal{S}^n}\{ks+\mathrm{Tr}(Y):\ Y+sI\succeq Z,\ Y\succeq 0\}. \tag{14}$$

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

$$\sum_{i=1}^{k} \lambda_i(Z) = \min \{k\lambda_{\max}(Z - Y) + \mathrm{Tr}(Y) : Y \in \mathcal{S}_+^n\}. \tag{15}$$

This relation together with (8) clearly implies identity (10). Moreover, noting that the max problem (11) is the dual of min problem (14) and that they both have strictly feasible solutions, we conclude that identity (11) holds in view of a well-known strong duality result.

(c) Using (15), the fact that $\inf_{x\in X}[x]^+ = [\inf X]^+$ for any $X \subseteq \Re$ and (9), we obtain

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

from which (12) immediately follows. Moreover, using (11), 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., (13) holds.

□

Lemma 2.3 *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*

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

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 [3] and statement (b) is an immediate consequence of (a). □

The following result about the sum of the k largest singular values of a matrix follows immediately from Lemmas 2.2 and 2.3.

Proposition 2.4 *Let $X \in \Re^{p\times q}$ and integer $1 \leq k \leq \min\{p, q\}$ be given and set $n := p + q$. Then*:

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

$$\sum_{i=1}^{k} \sigma_i(X) \leq t \;\Leftrightarrow\; \begin{cases} t - ks - \mathrm{Tr}(Y) & \geq 0, \\ Y - \mathcal{G}(X) + sI & \succeq 0, \\ Y & \succeq 0, \end{cases}$$

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

(b) *The following identities hold:*

$$\sum_{i=1}^{k} \sigma_i(X) = \min_{Y\in\mathcal{S}^n_+} \max_{W\in\Delta^n_{=}(1)} k(\mathcal{G}(X) - Y) \bullet W + \mathrm{Tr}(Y) \tag{16}$$

$$= \max_{W\in\mathcal{S}^n} \{\mathcal{G}(X) \bullet W : \mathrm{Tr}(W) = k,\ 0 \preceq W \preceq I\}. \tag{17}$$

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

$$\left[\alpha \sum_{i=1}^{k} \sigma_i(X) + \beta\right]^{+} = \min_{Y\in\mathcal{S}^n_+} \max_{W\in\Delta^n_{=}(1)} k(\alpha\mathcal{G}(X) - Y) \bullet W + [\beta + \mathrm{Tr}(Y)]\,\mathrm{Tr}(W) \tag{18}$$

$$= \max_{W\in\mathcal{S}^n,\, t\in\Re} \{\alpha\mathcal{G}(X) \bullet W + \beta t : \mathrm{Tr}(W) = tk,\ 0 \preceq W \preceq tI,\ 0 \leq t \leq 1\}. \tag{19}$$

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 l, which is usually quite large in many applications. However, the size of the decision variable U in (4) does not depend on l. 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 l. 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 $A^TA = Q\Lambda^2Q^T$. Letting

$$X := Q^TU, \quad H := \Lambda^{-1}Q^TA^TB, \tag{20}$$

we have

$$\begin{aligned}\|B - AU\|_{\mathrm{F}}^2 - \|B\|_F^2 &= \|AU\|_F^2 - 2\,(AU) \bullet B \\ &= \mathrm{Tr}\left(U^TA^TAU\right) - 2\,\mathrm{Tr}\left(U^TA^TB\right) \\ &= \mathrm{Tr}\left(U^TQ\Lambda^2Q^TU\right) - 2\,\mathrm{Tr}\left(U^TQ\Lambda H\right) \\ &= \|\Lambda X\|_F^2 - 2(\Lambda X) \bullet H \;=\; \|\Lambda X - H\|_F^2 - \|H\|_F^2.\end{aligned}$$

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

$$\min_X \left\{ \frac{1}{2}\|\Lambda X - H\|_{\mathrm{F}}^2 : \sum_{i=1}^m \sigma_i(X) \le M \right\}, \tag{21}$$

where Λ and H are defined in (20).

A related problem to (21) is the one given by

$$\min_X \frac{1}{2}\|\Lambda X - H\|_{\mathrm{F}}^2 + \lambda \sum_{i=1}^m \sigma_i(X). \tag{22}$$

In view of Theorem 6.2, we observe that for any $\lambda \geq 0$ and $\epsilon \geq 0$, any ϵ-optimal solution X_ϵ of (22) is an ϵ-optimal solution of problem (21) with $M = \sum_{i=1}^{m} \sigma_i(X_\epsilon)$. Hence, as is the case in our statistics application mentioned in the beginning of Sect. 1, if one is interested in solving (21) for an arbitrary sequence of M values, then it suffices to solve (22) for a sequence of λ values. However, there are situations in which one is interested in solving (21) for one particular M value. This situation arises for example in the root-finding approach recently proposed by Friedlander and van den Berg [22,23] for solving

$$\min_{U} \left\{ \sum_{i=1}^{m} \sigma_i(U) : \|B - AU\|_F^2 \leq \tau \right\} \tag{23}$$

for some $\tau \geq 0$, in which subproblems of the form (4) (or equivalently, (21)) must be solved at every iteration. In their implementation, a projected gradient method is used to solve these subproblems, but we observe that one of the first-order methods discussed in Sect. 4 can alternatively be used to solve these subproblems.

More specifically, we will describe first-order algorithms for solving problem (21) in Sects. 3.3.2 and 4.3 and problem (22) in Sects. 3.3.1 and 4.2. In Sect. 5, we only present computational results for the method for solving problem (22) since it is quite comparable to the algorithm for solving (21) both in terms of theoretical complexity and computational efficiency.

Before ending this subsection, we provide bounds on the optimal solutions of problems (21) and (22).

Lemma 3.1 *For every $M > 0$, problem (21) has a unique optimal solution X_M^*. Moreover,*

$$\|X_M^*\|_F \leq \tilde{r}_x := \min \left\{ \frac{2\|\Lambda H\|_F}{\lambda_{\min}^2(\Lambda)}, M \right\}. \tag{24}$$

Proof Using the fact that Λ is a $p \times p$ positive diagonal matrix, it is easy to see that the objective function of (21) is a (quadratic) strongly convex function, from which we conclude that (21) has a unique optimal solution X_M^*. Since $\|H\|_F^2/2$ is the value of the objective function of (21) at $X = 0$, we have $\|\Lambda X_M^* - H\|_F^2/2 \leq \|H\|_F^2/2$, or equivalently $\|\Lambda X_M^*\|_F^2 \leq 2(\Lambda H) \bullet X_M^*$. Hence, we have

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

which implies that $\|X_M^*\|_F \leq 2\|\Lambda H\|_F/\lambda_{\min}^2(\Lambda)$. 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

$$\|X\|_F \leq \sum_{i=1}^{m} \sigma_i(X). \tag{25}$$

Since X_M^* is feasible for (21), it then follows from (25) that $\|X_M^*\|_F \leq M$. We have thus shown that inequality (24) holds. □

Lemma 3.2 *For every $\lambda > 0$, problem (22) has a unique optimal solution X_λ^*. Moreover,*

$$\|X_\lambda^*\|_F \leq \sum_{i=1}^m \sigma_i(X_\lambda^*) \leq r_x := \min\left\{\frac{\|H\|_F^2}{2\lambda}, \sum_{i=1}^m \sigma_i\left(\Lambda^{-1}H\right)\right\}. \tag{26}$$

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 (22) is strongly convex, from which we conclude that (22) has a unique optimal solution X_λ^*. Since $\|H\|_F^2/2$ is the value of the objective function of (22) at $X = 0$, we have

$$\lambda \sum_{i=1}^m \sigma_i(X_\lambda^*) \leq \frac{1}{2}\|\Lambda X_\lambda^* - H\|_F^2 + \lambda \sum_{i=1}^m \sigma_i(X_\lambda^*) \leq \frac{1}{2}\|H\|_F^2. \tag{27}$$

Also, considering the objective function of (22) at $X = \Lambda^{-1}H$, we conclude that

$$\lambda \sum_{i=1}^m \sigma_i(X_\lambda^*) \leq \frac{1}{2}\|\Lambda X_\lambda^* - H\|_F^2 + \lambda \sum_{i=1}^m \sigma_i(X_\lambda^*) \leq \lambda \sum_{i=1}^m \sigma_i(\Lambda^{-1}H). \tag{28}$$

Now, (26) follows immediately from (25), (27) and (28). □

3.2 Cone programming reformulations

In this subsection, we provide cone programming reformulations for problems (21) and (22), respectively.

Proposition 3.3 *Problem (22) can be reformulated as the following cone programming:*

$$\begin{array}{ll} \min\limits_{r,s,t,X,Y} & 2r + \lambda t \\ s.t. & \begin{pmatrix} r+1 \\ r-1 \\ \mathrm{vec}(\Lambda X - H) \end{pmatrix} \in \mathcal{L}^{pq+2}, \\ & Y - \mathcal{G}(X) + sI \succeq 0, \\ & ms + \mathrm{Tr}(Y) - t \leq 0,\ Y \succeq 0, \end{array} \tag{29}$$

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 (22) is equivalent to

$$\begin{array}{ll} \min\limits_{r,X} & 2r + \lambda t \\ \text{s.t.} & \|\Lambda X - H\|_{\mathrm{F}}^2 \le 4r \\ & \sum\limits_{i=1}^{m} \sigma_i(X) - t \le 0. \end{array} \tag{30}$$

Using Lemma 2.3 and the following relation

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

for any $v \in \Re^k$ and $r \in \Re$, we easily see that (30) is equivalent to (29) □

The following proposition can be similarly established.

Proposition 3.4 *Problem* (21) *can be reformulated as the following cone programming:*

$$\begin{array}{ll} \min\limits_{r,s,X,Y} & 2r \\ s.t. & \begin{pmatrix} r+1 \\ r-1 \\ \mathrm{vec}(\Lambda X - H) \end{pmatrix} \in \mathcal{L}^{pq+2}, \\ & Y - \mathcal{G}(X) + sI \succeq 0, \\ & ms + \mathrm{Tr}(Y) \le M, \quad Y \succeq 0, \end{array} \tag{31}$$

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 (21) and (22).

3.3.1 Smooth saddle point reformulations for (22)

In this subsection, we reformulate (22) into a smooth saddle point problem that can be suitably solved by a variant of Nesterov's smooth method [15,16,21] as described in Sects. 4.1 and 4.2.

We start by introducing the following notation. For every $t \ge 0$, we let Ω_t denote the set defined as

$$\Omega_t := \{W \in \mathcal{S}^{p+q} : 0 \preceq W \preceq tI/m, \mathrm{Tr}(W) = t\}. \tag{32}$$

Theorem 3.5 *For some $\epsilon \geq 0$, assume that X_ϵ is an ϵ-optimal solution of the smooth saddle point problem*

$$\min_{X \in \mathcal{B}_F^{p\times q}(r_x)} \max_{W \in \Omega_1} \left\{ \frac{1}{2}\|\Lambda X - H\|_F^2 + \lambda m \mathcal{G}(X) \bullet W \right\}, \tag{33}$$

where $\mathcal{G}(X)$ and r_x are defined in (5) and (26), respectively. Then, X_ϵ is an ϵ-optimal solution of problem (22).

Proof This result follows immediately from Lemma 3.2 and relations (17) with $k = m$, (22) and (32) with $t = 1$. □

In addition to the saddle point (min-max) reformulation (33), it is also possible to develop an alternative saddle point reformulation based on the identity (16). These two reformulations can in turn be solved by a suitable method, namely Nesterov's smooth approximation scheme [16], for solving these min-max type problems, which we will not describe in this paper. In our computational experiments, we found that, among these two reformulations, the first one is computationally superior than the latter one. Details of the computational comparison of these two approaches can be found in the technical report (see [13]), which this paper originated from.

A more efficient method than the ones outlined in the previous paragraph for solving (22) is based on solving the dual of (33), namely the problem

$$\max_{W \in \Omega_1} \min_{X \in \mathcal{B}_F^{p\times q}(r_x)} \left\{ \frac{1}{2}\|\Lambda X - H\|_F^2 + \lambda m \mathcal{G}(X) \bullet W \right\}, \tag{34}$$

whose objective function has the desirable property that it has Lipschitz continuous gradient (see Sect. 4.2 for specific details). In Sects. 4.1 and 4.2, we describe an algorithm, namely, a variant of Nesterov's smooth method [15,16,21], for solving (34) which, as a by-product, yields a pair of primal and dual nearly-optimal solutions, and hence a nearly-optimal solution of (33). Finally, Sect. 5 only reports computational results for the approach outlined in this paragraph since it is far superior than the other two approaches outlined in the previous paragraph.

3.3.2 Smooth saddle point reformulations for (21)

In this subsection, we will provide a smooth saddle point reformulation for (21) that can be suitably solved by a variant of Nesterov's smooth method [15,16,21] as described in Sect. 4.1.

By directly applying Theorem 6.1 to problem (21), we obtain the following result.

Lemma 3.6 *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*

$$\bar{\gamma} = \frac{\|\Lambda \bar{X} - H\|_{\mathrm{F}}^2/2}{M - \sum_{i=1}^{m} \sigma_i(\bar{X})}. \tag{35}$$

Then, the following statements hold:

(a) *The optimal values of* (21) *and the penalized problem*

$$\min_{X\in\Re^{p\times q}}\left\{\frac{1}{2}\|\Lambda X-H\|_{\mathrm{F}}^2+\gamma\left[\sum_{i=1}^{m}\sigma_i(X)-M\right]^+\right\} \tag{36}$$

coincide, and the optimal solution solution X_M^* *of* (21) *is an optimal solution of* (36)*;*

(b) *if* $\epsilon\geq 0$ *and* X_ϵ *is an* ϵ*-optimal solution of problem* (36)*, then the point* X^ϵ *defined as*

$$X^\epsilon:=\frac{X_\epsilon+\theta\bar{X}}{1+\theta},\quad \textit{where}\quad \theta:=\frac{\left[\sum_{i=1}^{m}\sigma_i(X_\epsilon)-M\right]^+}{M-\sum_{i=1}^{m}\sigma_i(\bar{X})}, \tag{37}$$

is an ϵ*-optimal solution of* (21)*.*

We next provide a smooth saddle point reformulation for problem (21).

Theorem 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 defined in* (35)*. For some* $\epsilon\geq 0$*, assume that* X_ϵ *is an* ϵ*-optimal solution of the problem*

$$\min_{X\in\mathcal{B}_F^{p\times q}(\tilde{r}_x)}\ \max_{(t,W)\in\tilde{\Omega}}\left\{\frac{1}{2}\|\Lambda X-H\|_{\mathrm{F}}^2+\gamma(m\mathcal{G}(X)\bullet W-Mt)\right\}, \tag{38}$$

where $\tilde{r}_x$ *is defined in* (24) *and* $\tilde{\Omega}$ *is defined as*

$$\tilde{\Omega}:=\{(t,W)\in\Re\times\mathcal{S}^{p+q}:W\in\Omega_t,\ 0\leq t\leq 1\}. \tag{39}$$

Let X^ϵ *be defined in* (37)*. Then,* X^ϵ *is an* ϵ*-optimal solution of* (21)*.*

Proof Let X_M^* denote the unique optimal solution of (21). Then, X_M^* is also an optimal solution of (36) in view of Lemma 3.6(a), and X_M^* satisfies $X_M^*\in\mathcal{B}_F^{p\times q}(\tilde{r}_x)$ due to Lemma 3.1. Also, relation (19) with $\alpha=1,\beta=-M$ and $k=m$ implies that the objective functions of problems (36) and (38) are equal to each other over the whole space $\Re^{p\times q}$. The above observations then imply that X_M^* is also an optimal solution of (38) and that problems (36) and (38) have the same optimal value. Since by assumption X_ϵ is an ϵ-optimal solution of (38), it follows that X_ϵ is also an ϵ-optimal solution of problem (36). The latter conclusion together with Lemma 3.6(b) immediately yields the conclusion of the theorem. □

The saddle point (min-max) reformulation (38) can be solved by a suitable method, namely, Nesterov's smooth approximation scheme [16], which we will not describe in this paper. A more efficient method for solving (21) is based on solving the dual of (38), namely the problem

$$\max_{(t,W)\in\tilde{\Omega}}\ \min_{X\in\mathcal{B}_F^{p\times q}(\tilde{r}_x)}\left\{\frac{1}{2}\|\Lambda X-H\|_{\mathrm{F}}^2+\gamma(m\mathcal{G}(X)\bullet W-Mt)\right\}, \tag{40}$$

whose objective function has the desirable property that it has Lipschitz continuous gradient (see Sect. 4.3 for specific details). In Sects. 4.1 and 4.3, we describe an algorithm, namely a variant of Nesterov's smooth method [15,16,21], for solving (40) which, as a by-product, yields a pair of primal and dual nearly-optimal solutions, and hence a nearly-optimal solution of (38).

4 Numerical methods

In this section, we discuss numerical methods for solving problem (22). More specifically, Sect. 4.1 reviews a variant of Nesterov's smooth method [15,16,21], for solving a convex minimization problem over a relatively simple set with a smooth objective function that has Lipschitz continuous gradient. In Sects. 4.2 and 4.3, we present the implementation details of the variant of Nesterov's smooth method for solving the reformulations (34) of problem (22) and (40) of problem (21), respectively.

The implementation details of the other formulations discussed in the paper, more specifically, the reformulations (33) of problem (22) and (38) of problem (21) will not be presented here. The implementation details of some other reformulations of problems (22) and (21) can be found in Section 4.2 of [13].

4.1 Review of a variant of Nesterov's smooth method

In this subsection, we review a variant of Nesterov's smooth first-order method [15, 16,21] for solving a class of smooth convex programming (CP) problems.

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 a variant of Nesterov's smooth first-order method for solving the class of CP problems

$$\min_{u\in U} f(u) \tag{41}$$

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

$$f(u) := \max_{v\in V} \phi(u, v), \quad \forall u \in U, \tag{42}$$

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

$$\nabla f(u) = \nabla_u \phi(u, v(u)), \quad \forall u \in U, \tag{43}$$

where $v(u)$ denotes the unique solution of (42) (see for example Proposition B.25 of [4]). Moreover, problem (41) and its dual, namely:

$$\max_{v \in V} \{g(v) := \min_{u \in U} \phi(u, v)\}, \tag{44}$$

both have optimal solutions u^* and v^* such that $f(u^*) = g(v^*)$. Finally, using Assumption B.3, Lu [12] recently showed that problem (41-42) and its dual problem (44) can be suitably solved by Nesterov's smooth method [16], simultaneously. However, we note that Nesterov's smooth method [16] requires solving two prox-type subproblems per iteration. More recently, Tseng [21] proposed a variant of Nesterov's smooth method which solves one prox subproblem per iteration only.

We will now describe the aforementioned variant. 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.,

$$p_U(u) \geq p_U(\tilde{u}) + \langle \nabla p_U(\tilde{u}), u - \tilde{u} \rangle + \frac{\sigma_U}{2} \|u - \tilde{u}\|_{\mathcal{U}}^2, \quad \forall u, \tilde{u} \in U. \tag{45}$$

Let u_0 be defined as

$$u_0 = \arg\min\{p_U(u) : u \in U\}. \tag{46}$$

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

$$d_{p_U}(u; \tilde{u}) = p_U(u) - l_{p_U}(u; \tilde{u}), \quad \forall u, \tilde{u} \in U, \tag{47}$$

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 will now explicitly state Tseng's variant of Nesterov's smooth method for solving problem (41)–(42) and its dual problem (44). It uses a sequence $\{\alpha_k\}_{k \geq 0}$ of scalars satisfying the following condition:

$$0 < \alpha_k \leq \left(\sum_{i=0}^{k} \alpha_i \right)^{1/2}, \quad \forall k \geq 0. \tag{48}$$

Clearly, (48) implies that $\alpha_0 \in (0, 1]$.

Variant of Nesterov's smooth algorithm:

Let $u_0 \in U$ and $\{\alpha_k\}_{k \geq 0}$ satisfy (46) and (48), respectively.

Set $u_0^{sd} = u_0$, $v_0 = 0 \in \mathcal{V}$, $\tau_0 = 1$ 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

$$\begin{aligned} v_k &\equiv (1-\tau_{k-1})v_{k-1} + \tau_{k-1} v(u_{k-1}) \\ u_k^{ag} &\equiv \operatorname{argmin}\left\{ \frac{L}{\sigma_U} d_{p_U}(u; u_0) + \sum_{i=0}^{k-1} \alpha_i \, l_f(u; u_i) : u \in U \right\} \\ u_k^{sd} &\equiv (1-\tau_{k-1})u_{k-1}^{sd} + \tau_{k-1} u_k^{ag}. \end{aligned} \tag{49}$$

(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 for the above variant of Nesterov's smooth algorithm. Its proof is given in Corollary 3 of Tseng [21].

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

$$0 \le f(u_k^{sd}) - g(v_k) \le \frac{L D_U}{\sigma_U (\sum_{i=0}^{k-1} \alpha_i)}, \quad \forall k \ge 1, \tag{50}$$

where

$$D_U = \max\{p_U(u) : u \in U\}. \tag{51}$$

A typical sequence $\{\alpha_k\}$ satisfying (48) 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.

Corollary 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 the variant of Nesterov's smooth algorithm satisfies*

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

where D_U is defined in (51). Thus, the iteration-complexity of finding an ϵ-optimal solution to (41) and its dual (44) by the variant of Nesterov's smooth algorithm does not exceed $2[(L D_U)/(\sigma_U \epsilon)]^{1/2}$.

Before ending this subsection, we state sufficient conditions for the function ϕ to satisfy Assumptions B.1–B.3. The proof of the following result can be found in Theorem 1 of [16].

Proposition 4.3 *Let a norm $\|\cdot\|_{\mathcal{V}}$ on $\mathcal{V}$ be given. Assume that $\phi : U \times V \to \Re$ has the form*

$$\phi(u, v) = \theta(u) + \langle u, \mathcal{E} v \rangle - h(v), \quad \forall (u, v) \in U \times V, \tag{52}$$

where $\mathcal{E} : \mathcal{V} \to \mathcal{U}^*$ is a linear map, $\theta : U \to \Re$ is L_θ-Lipschitz-differentiable in U with respect to $\|\cdot\|_{\mathcal{U}}$, and $h : V \to \Re$ is a differentiable strongly convex function with modulus $\sigma_V > 0$ with respect to $\|\cdot\|_{\mathcal{V}}$. Then, the function f defined by (42) is $(L_\theta + \|\mathcal{E}\|^2_{\mathcal{V},\mathcal{U}}/\sigma_V)$-Lipschitz-differentiable in U with respect to $\|\cdot\|_{\mathcal{U}}$. As a consequence, ϕ satisfies Assumptions B.1–B.3 with norm $\|\cdot\|_{\mathcal{U}}$ and $L = L_\theta + \|\mathcal{E}\|^2_{\mathcal{V},\mathcal{U}}/\sigma_V$.

We will see in Sect. 4 that all saddle-point reformulations (41)–(42) of problems (21) and (22) studied in this paper have the property that the corresponding function ϕ can be expressed as in (52).

4.2 Implementation details of the variant of Nesterov's smooth method for (34)

The implementation details of the variant of Nesterov's smooth method (see Sect. 4.1) for solving formulation (34) (that is, the dual of (33) are addressed in this subsection. In particular, we describe in the context of this formulation the prox-function, the Lipschitz constant L and the subproblem (49) used by the variant of Nesterov's smooth algorithm of Sect. 4.1.

For the purpose of our implementation, we reformulate problem (34) into the problem

$$\min_{W\in\Omega_1} \max_{X\in\mathcal{B}_F^{p\times q}(1)} \left\{ -\lambda m r_x \mathcal{G}(X) \bullet W - \frac{1}{2}\|r_x \Lambda X - H\|_F^2 \right\} \tag{53}$$

obtained by scaling the variable X of (34) as $X \leftarrow X/r_x$, and multiplying the resulting formulation by -1. From now on, we will focus on formulation (53) rather than (34).

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

$$\begin{aligned} U &:= \Omega_1 \subseteq \mathcal{S}^n =: \mathcal{U}, \\ V &:= \mathcal{B}_F^{p\times q}(1) \subseteq \Re^{p\times q} =: \mathcal{V}, \end{aligned}$$

and

$$\phi(u, v) := -\lambda m r_x \mathcal{G}(v) \bullet u - \frac{1}{2}\|r_x \Lambda v - H\|_F^2, \quad \forall (u, v) \in U \times V, \tag{54}$$

where Ω_1 is defined in (32). Also, assume that the norm on $\mathcal{U}$ is chosen as

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

Our aim now is to show that ϕ satisfies Assumptions B.1–B.3 with $\|\cdot\|_{\mathcal{U}}$ as above and some Lipschitz constant $L > 0$, and hence that the variant of Nesterov's method can be applied to the corresponding saddle-point formulation (53). This will be done with the help of Proposition 4.3. Indeed, the function ϕ is of the form (52) with $\theta \equiv 0$

and the functions $\mathcal{E}$ and h given by

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

Assume that we fix the norm on $\mathcal{V}$ to be the Frobenius norm, i.e., $\|\cdot\|_{\mathcal{V}} = \|\cdot\|_F$. Then, it is easy to verify that the above function h is strongly convex with modulus $\sigma_V := r_x^2/\|\Lambda^{-1}\|^2$ with respect to $\|\cdot\|_{\mathcal{V}} = \|\cdot\|_F$. Now, using (6), we obtain

$$\begin{aligned} \|\mathcal{E}\|_{\mathcal{V},\mathcal{U}} &= \max\left\{\|\lambda m r_x \mathcal{G}(v)\|_{\mathcal{U}}^* : \ v \in \mathcal{V}, \|v\|_{\mathcal{V}} \le 1\right\}, \\ &= \lambda m r_x \max\left\{\|\mathcal{G}(v)\|_F : \ v \in \mathcal{V}, \|v\|_F \le 1\right\}, \\ &= \lambda m r_x \max\left\{\sqrt{2}\|v\|_F : \ v \in \mathcal{V}, \|v\|_F \le 1\right\} = \sqrt{2}\lambda m r_x. \end{aligned} \tag{55}$$

Hence, by Proposition 4.3, we conclude that ϕ satisfies Assumptions B.1–B.3 with $\|\cdot\|_{\mathcal{U}} = \|\cdot\|_F$ and

$$L = \|\mathcal{E}\|_{\mathcal{V},\mathcal{U}}^2/\sigma_V = 2\lambda^2 m^2 \|\Lambda^{-1}\|^2.$$

The prox-function $p_U(\cdot)$ for the set U used in the variant of Nesterov's algorithm is defined as

$$p_U(u) = \mathrm{Tr}(u \log u) + \log n, \quad \forall u \in U = \Omega_1. \tag{56}$$

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 $\|\cdot\|_{\mathcal{U}} = \|\cdot\|_F$. Also, it is easy to verify that $\min\{p_U(u) : u \in U\} = 0$ and that

$$\begin{aligned} u_0 &:= \arg\min_{u \in U} p_U(u) = I/n, \\ D_U &:= \max_{u \in U} p_U(u) = \log(n/m). \end{aligned} \tag{57}$$

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

Theorem 4.4 *For a given $\epsilon > 0$, the variant of Nesterov's smooth method applied to (34) finds an ϵ-optimal solution of problem (34) and its dual, and hence of problem (22), in a number of iterations which does not exceed*

$$\left\lceil \frac{2\sqrt{2}\lambda\|\Lambda^{-1}\|}{\sqrt{\epsilon}} \sqrt{m \log(n/m)} \right\rceil. \tag{58}$$

We observe that the iteration-complexity given in (58) is in terms of the transformed data of problem (4). We next relate it to the original data of problem (4).

Corollary 4.5 *For a given $\epsilon > 0$, the variant of Nesterov's smooth method applied to (34) finds an ϵ-optimal solution of problem (34) and its dual, and hence of problem (22), in a number of iterations which does not exceed*

$$\left\lceil \frac{2\sqrt{2}\lambda \|(A^T A)^{-1/2}\|}{\sqrt{\epsilon}} \sqrt{m \log(n/m)} \right\rceil.$$

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

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

The conclusion immediately follows from this identity and Theorem 4.4. □

It is interesting to note that the iteration-complexity of Corollary 4.5 depends on the data matrix A but not on B. Based on the discussion below, the arithmetic operation cost per iteration of the variant of Nesterov's smooth method when applied to problem (40) is bounded by $\mathcal{O}(mpq)$ where $m = \min(p, q)$, due to the fact that its most expensive operation consists of finding a partial singular value decomposition of a $p \times q$ matrix h as in (61). Thus, the overall arithmetic-complexity of the variant of Nesterov's smooth method when applied to (34) is

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

After having completely specified all the ingredients required by the variant of Nesterov's smooth method for solving (53), we now discuss some of the computational technicalities involved in the actual implementation of the method.

First, recall that, for a given $u \in U$, the optimal solution for the maximization subproblem (42) needs to be found in order to compute the gradient of $\nabla f(u)$. Using (54) and the fact that $V = \mathcal{B}_F^{p\times q}(1)$, we see that the maximization problem (42) is equivalent to

$$\min_{v\in\mathcal{B}_F^{p\times q}(1)} \frac{1}{2}\|r_x \Lambda v - H\|_F^2 + G \bullet v, \tag{59}$$

where $G := \mathcal{G}^*(u) \in \Re^{p\times q}$. We now briefly discuss how to solve (59). For any $\xi \geq 0$, let

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

If $\Psi(0) \leq 0$, then clearly $v(0)$ is the optimal solution of problem (59). Otherwise, the optimal solution of problem (59) is equal to $v(\xi^*)$, where ξ^* is the root of the equation $\Psi(\xi) = 0$. The latter can be found by well-known root finding schemes specially tailored for solving the above equation.

In addition, each iteration of the variant of Nesterov's smooth method requires solving subproblem (49). In view of (43) and (54), it is easy to see that for every $u \in U$, we have $\nabla f(u) = \mathcal{G}(v)$ for some $v \in \Re^{p\times q}$. Also, $\nabla p_U(u_0) = (1 - \log n)I$ due to (56) and (57). These remarks together with (47) and (56) imply that subproblem (49) is of the form

$$\min_{u\in\Omega_1} \ (\varsigma I + \mathcal{G}(h)) \bullet u + \mathrm{Tr}(u \log u) \tag{60}$$

for some real scalar ς and $h \in \Re^{p\times q}$, where Ω_1 is given by (32).

We now present an efficient approach for solving (60) which, instead of finding the eigenvalue factorization of the $(p+q)$-square matrix $\varsigma I + \mathcal{G}(h)$, computes the singular value decomposition of the smaller $p \times q$-matrix h. 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

$$\tilde{U}^T\tilde{U} = I, \quad \Sigma = \mathrm{Diag}(\sigma_1(h), \ldots, \sigma_m(h)), \quad \tilde{V}^T\tilde{V} = I, \tag{61}$$

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. Using (5), it is easy to see that

$$f^i = \frac{1}{\sqrt{2}}\begin{pmatrix}\eta_i\\ \xi_i\end{pmatrix}, \quad i = 1, \ldots, m; \quad f^{m+i} = \frac{1}{\sqrt{2}}\begin{pmatrix}\eta_i\\ -\xi_i\end{pmatrix}, \quad i = 1, \ldots, m, \tag{62}$$

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 [3]). Thus, we obtain the following eigenvalue decomposition of $\varsigma I + \mathcal{G}(h)$:

$$\begin{aligned}\varsigma I + \mathcal{G}(h) = F\mathrm{Diag}(a)F^T, \quad a = \varsigma e + (\sigma_1(h), \ldots,\\ \sigma_m(h), -\sigma_1(h), \ldots, -\sigma_m(h), 0, \ldots, 0)^T.\end{aligned}$$

Using this relation and (32) with $t = 1$, it is easy to see that the optimal solution of (60) is $v^* = F\mathrm{Diag}(w^*)F^T$, where $w^* \in \Re^n$ is the unique optimal solution of the problem

$$\begin{array}{ll}\min & a^Tw + w^T\log w\\ \text{s.t.} & e^Tw = 1,\\ & 0 \le w \le e/m.\end{array} \tag{63}$$

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

$$w_{2m+1}^* = w_{2m+2}^* = \cdots = w_n^* = \vartheta. \tag{64}$$

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 (64), we see that the optimal solution v^* of (60) can be efficiently computed as

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

where the scalar ϑ is defined above and the vectors $\{f^i : i = 1, \ldots 2m\}$ are given by (62).

Finally, to terminate the variant of Nesterov's smooth method, we need to evaluate the primal and dual objective functions of problem (53). As mentioned above, the primal objective function $f(u)$ of (53) can be computed by solving a problem of the form (59). Additionally, in view of (17) and (32), the dual objective function $g(v)$ of (53) can be computed as

$$g(v) = -\frac{1}{2}\|r_x \Lambda v - H\|_F^2 - \lambda r_x \sum_{i=1}^{m} \sigma_i(v), \quad \forall v \in V.$$

4.3 Implementation details of the variant of Nesterov's smooth method for (40)

The implementation details of the variant of Nesterov's smooth method (see Sect. 4.1) for solving formulation (40) (that is, the dual of (38) are addressed in this subsection. In particular, we describe in the context of this formulation the prox-function, the Lipschitz constant L and the subproblem (49) used by the variant of Nesterov's smooth algorithm of Sect. 4.1.

For the purpose of our implementation, we reformulate problem (40) into the problem

$$\min_{(t,W)\in\tilde{\Omega}} \max_{X\in\mathcal{B}_F^{p\times q}(1)} \left\{-\gamma[m\tilde{r}_x\mathcal{G}(X)\bullet W - Mt] - \frac{1}{2}\|\tilde{r}_x\Lambda X - H\|_{\mathrm{F}}^2\right\} \tag{65}$$

obtained by scaling the variables X of (40) as $X \leftarrow X/\tilde{r}_x$, and multiplying the resulting formulation by -1. From now on, our discussion in this subsection will focus on formulation (65) rather than (40).

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

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

and

$$\phi(u, v) := -\gamma[m\tilde{r}_x\mathcal{G}(v)\bullet W - Mt] - \frac{1}{2}\|\tilde{r}_x\Lambda v - H\|_{\mathrm{F}}^2, \quad \forall (u, v) \in U \times V, \tag{66}$$

where $\tilde{\Omega}$ is defined in (39). Also, assume that the norm on $\mathcal{U}$ is chosen as

$$\|u\|_{\mathcal{U}} := \left(\xi t^2 + \|W\|_F^2\right)^{1/2}, \quad \forall u = (t, W) \in \mathcal{U},$$

where ξ is a positive scalar that will be specified later. Our aim now is to show that ϕ satisfies Assumptions B.1-B.3 with $\|\cdot\|_{\mathcal{U}}$ as above and some Lipschitz constant $L > 0$, and hence that the variant of Nesterov's method can be applied to the corresponding saddle-point formulation (65). This will be done with the help of Proposition 4.3. Indeed, the function ϕ is of the form (52) with θ, $\mathcal{E}$ and h given by

$$\begin{aligned} \theta(u) &:= \gamma Mt, \quad \forall u = (t, W) \in \mathcal{U}, \\ \mathcal{E}v &:= (0, -\gamma m\tilde{r}_x\mathcal{G}(v)), \quad \forall v \in \mathcal{V}, \\ h(v) &:= \frac{1}{2}\|\tilde{r}_x\Lambda v - H\|_F^2, \quad \forall v \in \mathcal{V}. \end{aligned} \tag{67}$$

Clearly, θ is a linear function, and thus it is a 0-Lipschitz-differentiable function on U with respect to $\|\cdot\|_{\mathcal{U}}$. Now, assume that we fix the norm on $\mathcal{V}$ to be the Frobenius norm, i.e., $\|\cdot\|_{\mathcal{V}} = \|\cdot\|_F$. Then, it is easy to verify that the above function h is strongly convex with modulus $\sigma_V := \tilde{r}_x^2/\|\Lambda^{-1}\|^2$ with respect to $\|\cdot\|_{\mathcal{V}} = \|\cdot\|_F$. Now, using (6), (67) and the fact that

$$\|u\|_{\mathcal{U}}^* = \left(\xi^{-1}t^2 + \|W\|_F^2\right)^{1/2}, \quad \forall u = (t, W) \in \mathcal{U}^* = \mathcal{U}, \tag{68}$$

we obtain

$$\begin{aligned}\|\mathcal{E}\|_{\mathcal{V},\mathcal{U}} &= \max\left\{\|(0, -\gamma m\tilde{r}_x \mathcal{G}(v))\|_{\mathcal{U}}^* : \ v \in \mathcal{V}, \|v\|_{\mathcal{V}} \le 1\right\}, \\ &= \gamma m\tilde{r}_x \max\left\{\|\mathcal{G}(v)\|_F : \ v \in \mathcal{V}, \|v\|_F \le 1\right\}, \\ &= \gamma m\tilde{r}_x \max\left\{\sqrt{2}\|v\|_F : \ v \in \mathcal{V}, \|v\|_F \le 1\right\} = \sqrt{2}\gamma m\tilde{r}_x. \end{aligned} \tag{69}$$

Hence, by Proposition 4.3, we conclude that ϕ satisfies Assumptions B.1–B.3 with $\|\cdot\|_{\mathcal{U}} = \|\cdot\|_F$ and

$$L = L_\theta + \|\mathcal{E}\|_{\mathcal{V},\mathcal{U}}^2/\sigma_V = 2\gamma^2 m^2 \|\Lambda^{-1}\|^2. \tag{70}$$

We will now specify the prox-function p_U for the set U used in the variant of Nesterov's algorithm. We let

$$p_U(u) = \mathrm{Tr}(W \log W) + at \log t + bt + c, \quad \forall u = (t, W) \in U, \tag{71}$$

where

$$a := \log \frac{n}{m}, \quad b := \log n - a - 1 = \log m - 1, \quad c := a + 1. \tag{72}$$

For a fixed $t \in [0, 1]$, it is easy to see that

$$\min_{W \in \Omega_t} p_U(t, W) = \psi(t) := t \log \frac{t}{n} + at \log t + bt + c,$$

and that the minimum is achieved at $W = tI/n$. Now,

$$\psi'(1) = \log \frac{1}{n} + 1 + a(\log 1 + 1) + b = 1 - \log n + a + b = 0,$$

where the last equality follows from the second identity in (72). These observations together with (39) allow us to conclude that

$$\arg\min_{u \in U} p_U(u) = u_0 := (1, I/n), \tag{73}$$

$$\min_{u \in U} p_U(u) = \psi(1) = -\log n + b + c = 0, \tag{74}$$

where the last equality is due to second and third identities in (72). Moreover, it is easy to see that

$$\begin{aligned} D_U := \max_{u \in U} p_U(u) &= \max_{t \in [0,1]} t \log \frac{t}{m} + at \log t + bt + c \\ &= c + \max\{0, b - \log m\} = 1 + \log \frac{n}{m}, \end{aligned} \tag{75}$$

where the last identity is due to (72). Also, we easily see that $p_U(\cdot)$ is a strongly differentiable convex function on U with modulus

$$\sigma_U = \min(a/\xi,\ m) \tag{76}$$

with respect to the norm $\|\cdot\|_{\mathcal{U}}$.

In view of (70), (75), (76) and Corollary 4.2, it follows that the iteration-complexity of the variant of Nesterov's smooth method for finding an ϵ-optimal solution of (65) and its dual is bounded by

$$\Gamma(\xi) = \left\lceil \frac{2\gamma m\|\Lambda^{-1}\|}{\sqrt{\epsilon}} \sqrt{\frac{2[1+\log(n/m)]}{\min(a/\xi,\ m)}} \right\rceil.$$

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

Theorem 4.6 *For a given $\epsilon > 0$, the variant of Nesterov's smooth method, with prox-function defined by (71)–(72), L given by (70) and σ_U given by (76) with $\xi = a/m$, applied to (65), finds an ϵ-optimal solution of problem (65) and its dual in a number of iterations which does not exceed*

$$\left\lceil \frac{2\sqrt{2}\gamma\|\Lambda^{-1}\|\sqrt{m}}{\sqrt{\epsilon}} \sqrt{1+\log(n/m)} \right\rceil. \tag{77}$$

Proof We have seen in the discussion preceding this theorem that the iteration-complexity of the variant of Nesterov's smooth method for finding an ϵ-optimal solution of (65) and its dual is bounded by $\Gamma(\xi)$ for any $\xi > 0$. Taking $\xi = a/m$, we obtain the iteration-complexity bound (77). □

We observe that the iteration-complexity given in (77) is in terms of the transformed data of problem (4). We next relate it to the original data of problem (4). The proof of the following corollary is similar to that of Corollary 4.5.

Corollary 4.7 *For a given $\epsilon > 0$, the variant of Nesterov's smooth method, with prox-function defined by (71)–(72), L given by (70) and σ_U given by (76) with $\xi = a/m$, applied to (40) finds an ϵ-optimal solution of problem (40) and its dual in a number of iterations which does not exceed*

$$\left\lceil \frac{2\sqrt{2}\gamma\|(A^TA)^{-1/2}\|\sqrt{m}}{\sqrt{\epsilon}} \sqrt{\log(n/m)+1} \right\rceil. \tag{78}$$

Observe that, in view of Lemma 3.6 with $\bar{X} = 0$ and Theorem 3.7, (78) is also an iteration-complexity bound for finding an ϵ-optimal solution of problem (21) whenever

$$\gamma = \frac{\|H\|_F^2}{M} = \frac{\|(A^T A)^{-1/2} A^T B\|_F^2}{M}, \tag{79}$$

where the latter equality is due to (20).

Given a $\lambda > 0$, by Theorem 6.2, the optimal solution X_λ^* of the λ-problem (22) is also an optimal solution of problem (21) with $M = M_\lambda := \sum_{i=1}^m \sigma_i(X_\lambda^*)$. The latter problem, which we refer to as the M_λ-problem, is hence equivalent to the λ-problem. By Lemma 3.2, $M_\lambda \leq \|H\|_F^2/(2\lambda)$, so that γ given by (79) with $M = M_\lambda$ satisfies $\gamma \geq 2\lambda$. Hence, the complexity of solving the M_λ-problem is at least on the same order as that of solving the equivalent λ-problem. In practice, the computational complexities of solving these two problems are about the same.

If one is interested in solving a single M-problem, one could try to use a search scheme to find a scalar λ such that $M_\lambda \approx M$, and then solve the corresponding λ-problem. However, we believe that a more promising approach is to solve the M-problem directly by using the method described in this subsection.

Based on the discussion below and in Sect. 4.2, the arithmetic operation cost per iteration of the variant of Nesterov's smooth method when applied to problem (40) is bounded by $\mathcal{O}(mpq)$ where $m = \min(p, q)$, due to the fact that its most expensive operation consists of finding a partial singular value decomposition of a $p \times q$ matrix h as in (61). Thus, the overall arithmetic-complexity of the variant of Nesterov's smooth method when applied to (40) is

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

After having completely specified all the ingredients required by the variant of Nesterov's smooth method for solving (65), 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 (42) needs to be found in order to compute the gradient of $\nabla f(u)$. The details here are similar to the corresponding ones described in Sect. 4.2 (see the paragraph containing relation (59)).

In addition, each iteration of the variant of Nesterov's smooth method requires solving subproblem (49). In view of (43) and (66), it is easy to observe that for every $u = (t, W) \in U$, we have $\nabla f(u) = (\eta, \mathcal{G}(v))$ for some $\eta \in \Re$ and $v \in \Re^{p\times q}$. Also, by (71), (72) and (73), we easily see that $\nabla p_U(u_0) = (\log n - 1)(1, -I)$. Using these results along with (47) and (56), we easily see that subproblem (49) is equivalent to one of the form

$$\min_{(t,W)\in\tilde{\Omega}} \{(\varsigma I + \mathcal{G}(h)) \bullet W + \alpha t + \mathrm{Tr}(W \log W) + at \log t\} \tag{80}$$

for some $\alpha, \varsigma \in \Re$ and $h \in \Re^{p\times q}$, where a and $\tilde{\Omega}$ are given by (72) and (39), respectively.

We now discuss how the above problem can be efficiently solved. First, note that by (39), we have $(t, W) \in \tilde{\Omega}$ if, and only if, $W = tW'$ for some $W' \in \Omega_1$.

This observation together with the fact that $\mathrm{Tr}\, W' = 1$ for every $W' \in \Omega_1$ allows us to conclude that problem (80) is equivalent to

$$\min_{W' \in \Omega_1,\, t \in [0,1]} \left\{ t\,(\varsigma I + \mathcal{G}(h)) \bullet W' + \alpha t + t\left[\mathrm{Tr}(W' \log W') + (\log t)\mathrm{Tr}(W')\right] + at \log t \right\} \tag{81}$$

$$= \min_{t \in [0,1]} \alpha t + (a+1) t \log t + td, \tag{82}$$

where

$$d := \min_{W' \in \Omega_1} (\varsigma I + \mathcal{G}(h)) \bullet W' + \mathrm{Tr}(W' \log W'). \tag{83}$$

Moreover, if W' is the optimal solution of (83) and t is the optimal solution of (82), then $W = tW'$ is the optimal solution of (81). Problem (83) is of the form (60) where an efficient scheme for solving it is described in Sect. 4.2. It is easy to see that the optimal solution of (82) is given by

$$t = \min\left[1,\ \exp\left(-1 - \frac{\alpha + d}{a+1}\right)\right].$$

5 Computational results

In this section, we report the results of our computational experiment which compares the performance of the variant of Nesterov's smooth method discussed in Sect. 4.2 for solving problem (22) with the interior point method implemented in SDPT3 version 4.0 (beta) [20] on a set of randomly generated instances. We do not report computational results on the performance of the variant of Nesterov's smooth method for solving the M-problem (21) since the latter approach is quite similar to the one for solving the λ-problem (22) both in terms of theoretical complexity and computational efficiency (see Sect. 4.3 for a discussion regarding this issue).

The random instances of (22) used in our experiments were generated as follows. We first randomly generated matrices $A \in \Re^{l \times p}$ and $B \in \Re^{l \times q}$, where $p = 2q$ and $l = 10q$, with entries uniformly distributed in [0, 1] for different values of q. We then computed H and Λ for (22) according to the procedures described in Sect. 3.1 and set the parameter λ in (22) to one. In addition, all computations were performed on an Intel Xeon 5320 CPU (1.86 GHz) and 12 GB RAM running Red Hat Enterprise Linux 4 (kernel 2.6.9).

In this experiment, we compared the performance of the variant of Nesterov's smooth method (labeled as VNS) discussed in Sect. 4.2 for solving problem (22) with the interior point method implemented in SDPT3 version 4.0 (beta) [20] for solving the cone programming reformulation (29). The code for VNS is written in C, and the initial point for this method is set to be $u_0 = I/(p+q)$. It is worth mentioning that the code SDPT3 uses MATLAB as interface to call several C subroutines to handle all its

Table 1 Comparison of VNS and SDPT3

Problem (p, q)	Iter		Obj		Time		Memory	
	VNS	SDPT3	VNS	SDPT3	VNS	SDPT3	VNS	SDPT3
(20, 10)	36,145	17	4.066570508	4.066570512	16.6	5.9	2.67	279
(40, 20)	41,786	15	8.359912031	8.359912046	55.7	77.9	2.93	483
(60, 30)	35,368	15	13.412029944	13.412029989	96.7	507.7	3.23	1,338
(80, 40)	36,211	15	17.596671337	17.596671829	182.9	2209.8	3.63	4,456
(100, 50)	33,602	19	22.368563640	22.368563657	272.6	8916.1	4.23	10,445
(120, 60)	33,114	N/A	26.823206950	N/A	406.6	N/A	4.98	>16,109

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 VNS 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 VNS and SDPT3 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 VNS and SDPT3 are given in columns two and three, and the objective function values are given in columns four and five, CPU times (in seconds) are given in columns six to seven, and the amount of memory (in mega bytes) used by VNS 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 VNS, namely, the variant of Nesterov's smooth method, generally outperforms SDPT3 substantially even for relatively small-scale problems. Moreover, VNS requires much less memory than SDPT3. For example, for the instance with $(p, q) = (100, 50)$, SDPT3 needs $10,445$ mega ($\approx$ 10.2 giga) bytes of memory, but VNS 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 VNS only requires about 4.98 mega bytes of memory.

6 Concluding remarks

In this paper, we studied convex optimization methods for computing the nuclear norm regularized least squares estimate in multivariate linear regression. In particular, we explore a variant of Nesterov's smooth method 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 showed that the variant of Nesterov's smooth method generally substantially outperforms the interior point method implemented in SDPT3 version 4.0 (beta) [20]. Moreover, the former method is much more memory efficient.

In Sect. 3.1 we provided an approach for simplifying problem (4) which changes the variable U, in addition to the data A and B. A drawback of this approach is that

it can not handle extra constraints (not considered in this paper) on U. It turns out that there exists an alternative scheme for simplifying problem (4), i.e. one that eliminates the dependence of the data on the (generally, large) dimension l, which does not change U. Indeed, by performing either a QR factorization of A or a Cholesky factorization of $A^T A$, compute an upper triangular matrix R such that $R^T R = A^T A$. Letting $G := R^{-T} A^T B$, it is straightforward to show that problem (4) can be reduced to

$$\min_U \left\{ \|G - RU\|_F^2 : \sum_{i=1}^{m} \sigma_i(U) \leq M \right\}. \tag{84}$$

Clearly, in contrast to reformulation (21), the above one does not change the variable U and hence extra constraints on U can be easily handled. On the other hand, a discussion similar to that in Sect. 4.2 shows that each iteration of the variant of Nesterov's smooth method applied to (84), or its Lagrangian relaxation version, needs to solve subproblem (59) with Λ replaced by R. Since R is an upper triangular matrix and Λ is a diagonal matrix, the latter subproblems are much harder to solve than subproblems of the form (59). For this reason, we have opted to use reformulation (21) rather than (84) in this paper.

Acknowledgments The authors would like to thank two anonymous referees and the associate editor for numerous insightful comments and suggestions, which have greatly improved the paper.

Appendix

In this section, we discuss some technical results that are used in our presentation. More specifically, we discuss two ways of solving a constrained nonlinear programming problem based on some unconstrained nonlinear programming reformulations.

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

$$f^* = \inf \{ f(x) : x \in X,\ h_i(x) \leq 0,\ i = 1, \ldots, k \}. \tag{85}$$

The first reformulation of (85) is based on the exact penalty approach, which consists of solving the exact penalization problem

$$f_\gamma{}^* = \inf \{ f_\gamma(x) := f(x) + \gamma [g(x)]^+ : x \in X \}, \tag{86}$$

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 (85):

(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 $\epsilon \geq 0$, we say that x_ϵ is an ϵ-optimal solution of this problem if $x_\epsilon \in \mathcal{F}$ and $f(x_\epsilon) \leq \epsilon + \bar{f}$.

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

Theorem 6.1 *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

$$z(x) := \frac{x + \theta(x)x^0}{1 + \theta(x)}, \quad \text{where} \quad \theta(x) := \frac{[g(x)]^+}{|g(x^0)|}. \tag{87}$$

Then, the following statements hold:

(a) *for every $x \in X$, the point $z(x)$ is a feasible solution of (85);*
(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 (85) is also an ϵ-optimal solution of (86);*
(d) *if $\gamma \geq \bar{\gamma}, \epsilon \geq 0$ and x_ϵ^γ is an ϵ-optimal solution of (86), then the point $z(x_\epsilon^\gamma)$ is an ϵ-optimal solution of (85).*
(e) *if $\gamma > \bar{\gamma}, \epsilon \geq 0$ and x_ϵ^γ is an ϵ-optimal solution of (86), 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

$$\begin{aligned} f_\gamma(x) - f^* &= f(x) + \gamma[g(x)]^+ - f^* \\ &\geq (1 + \theta(x))f(z(x)) - \theta(x)f(x^0) + \gamma[g(x)]^+ - f^* \\ &= (1 + \theta(x))(f(z(x)) - f^*) - \theta(x)(f(x^0) - f^*) + \gamma[g(x)]^+ \\ &= (1 + \theta(x))(f(z(x)) - f^*) + (\gamma - \bar{\gamma})[g(x)]^+. \end{aligned} \tag{88}$$

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 (86). Then, statement (b) and inequality (88) imply that

$$\epsilon \geq f_\gamma(x_\epsilon^\gamma) - f_\gamma{}^* \geq (1 + \theta(x_\epsilon^\gamma))(f(z(x_\epsilon^\gamma)) - f^*) + (\gamma - \bar{\gamma})[g(x_\epsilon^\gamma)]^+. \quad (89)$$

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 (85) in view of statement (a). Hence, statement (d) follows. Moreover, if $\gamma > \bar{\gamma}$, we also conclude from (89) that $[g(x_\epsilon^\gamma)]^+ \leq \epsilon/(\gamma - \bar{\gamma})$. Also, the first inequality of (89) 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 (86) as $\gamma := (f(x^0) - f_l)/|g(x^0)|$ guarantees that an ϵ-optimal solution x_ϵ^γ of (86) yields the ϵ-optimal solution $z(x_\epsilon^\gamma)$ of (85), in view of Theorem 6.1(c).

The following result, which is a slight variation of a result due to H. Everett (see for example pages 147 and 163 of [8]), shows that approximate optimal solutions of Lagrangian subproblems associated with (85) yield approximate optimal solutions of a perturbed version of (85).

Theorem 6.2 (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*

$$f_\lambda^* = \inf \left\{ f(x) + \sum_{i=1}^k \lambda_i h_i(x) : x \in X \right\}. \quad (90)$$

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

$$f_{\epsilon\lambda}^* = \inf \left\{ f(x) : x \in X,\ h_i(x) \leq h_i(x_\epsilon^\lambda),\ i = 1, \ldots, k \right\}. \quad (91)$$

Proof Let $\tilde{x}$ be a feasible solution of (91). Since x_ϵ^λ is an ϵ-optimal solution of (90), 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 (90) 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 (91). Since the latter inequality holds for every feasible solution $\tilde{x}$ of (91), we conclude that $f(x_\epsilon^\lambda) \leq f_{\epsilon\lambda}^* + \epsilon$, and hence that x_ϵ^λ is an ϵ-optimal solution of (91). □

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 (90) for many different Lagrange multipliers $\lambda \in \Re^k_+$. We note that this idea is specially popular in statistics for the case when $k = 1$.

References

1. Anderson, T.W.: Estimating linear restriction on regression coefficients for multivariate normal distributions. Ann. Appl. Probab. **22**, 327–351 (1951)
2. Bach, F.: Consistency of trace norm minimization. J. Mach. Learn. Res. **8**, 1019–1048 (2008)
3. Ben-Tal, A., Nemirovski, A.: Lectures on Modern Convex Optimization: Analysis, algorithms, Engineering Applications. MPS-SIAM Series on Optimization. SIAM, Philadelphia (2001)
4. Bertsekas, D.: Nonlinear Programming. 2nd edn. Athena Scientific, New York (1999)
5. Breiman, L.: Heuristics of instability and stabilization in model selection. Ann. Stat. **24**, 2350–2383 (1996)
6. Brooks, R., Stone, M.: Joint continuum regression for multiple predictands. J. Am. Stat. Assoc. **89**, 1374–1377 (1994)
7. Fazel, M., Hindi, H., Boyd, S.P.: A rank minimization heuristic with application to minimum order system approximation. In: Proceedings American Control Conference, vol. 6, pp. 4734–4739 (2001)
8. Hiriart-Urruty, J.-B., Lemaréchal, C.: Convex Analysis and Minimization algorithms I, volume 305 of Comprehensive Study in Mathematics. Springer, New York (1993)
9. Hotelling, H.: The most predictable criterion. J. Educational Psychol. **26**, 139–142 (1935)
10. Hotelling, H.: Relations between two sets of variables. Biometrika **28**, 321–377 (1936)
11. Izenman, A.: Reduced-rank regression for the multivariate linear model. J. Multivar. Anal. **5**, 248–264 (1975)
12. Lu, Z.: Smooth optimization approach for covariance selection. SIAM J. Optim. **19**, 1807–1827 (2009)
13. Lu, Z., Monteiro, R.D.C., Yuan, M.: Convex optimization methods for dimension reduction and coefficient estimation in multivariate linear regression. Technical report, Department of Mathematics, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada, January 2008
14. Massy, W.: Principle components regression with exploratory statistical research. J. Am. Stat. Assoc. **60**, 234–246 (1965)
15. Nesterov, Y. E.: 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
16. Nesterov, Y.E.: Smooth minimization of nonsmooth functions. Math. Program. **103**, 127–152 (2005)
17. Recht, B., Fazel, M., Parrilo, P.A. (2007) Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. Technical report arXiv:0706.4138v1, arXiv
18. Reinsel, G., Velu, R.: Multivariate Reduced-rank Regression: Theory and Application. Springer, New York (1998)
19. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. Royal. Statist. Soc. B **58**, 267–288 (1996)
20. Toh, K.C., Tütüncü, R.H., Todd, M.J.: 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
21. Tseng, P.: On accelerated proximal gradient methods for convex–concave optimization. Manuscript, Department of Mathematics, University of Washington, May 2008
22. van den Berg, E., Friedlander, M.: A root-finding approach for sparse recovery. Talk given at ISMP, Chicago, 23–28 August 2009
23. van den Berg, E., Friedlander, M.: In pursuit of a root. Working paper, Department of Computer Science, University of British Columbia, November 2009
24. Wold, H.: 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)
25. Yuan, M., Ekici, A., Lu, Z., Monteiro, R.D.C.: Dimension reduction and coefficient estimation in multivariate linear regression. J. R. Stat. Soc. Series B Stat. Methodol. **69**(3), 329–346 (2007)