Solving nuclear norm regularized and semidefinite matrix least squares problems with linear equality constraints

Kaifeng Jiang*, Defeng Sun†, and Kim-Chuan Toh‡

April 9, 2012

Abstract

We introduce a partial proximal point algorithm for solving nuclear norm regularized and semidefinite matrix least squares problems with linear equality constraints. For the inner subproblems, we show that the positive definiteness of the generalized Hessian of the objective function for the inner subproblems is equivalent to the constraint nondegeneracy of the corresponding primal problem, which is a key property for applying a semismooth Newton-CG method to solve the inner subproblems efficiently. Numerical experiments on large scale matrix least squares problems arising from low rank matrix approximation, as well as regularized kernel estimation and Euclidean distance matrix completion problems in molecular conformation show that our algorithm is efficient and robust.

1 Introduction

Let $\Re^{p\times q}$ be the space of all $p \times q$ matrices equipped with the standard trace inner product $\langle X, Y\rangle = \mathrm{Tr}(X^T Y)$ and its induced Frobenius norm $\|\cdot\|$. Without loss of generality, we assume $p \le q$ throughout this paper. For a given $X \in \Re^{p\times q}$, its nuclear norm $\|X\|_*$ is defined as the sum of all its singular values and its operator norm $\|X\|_2$ is the largest singular value. Let $\mathcal{S}^n$ be the space of all $n \times n$ symmetric matrices and $\mathcal{S}^n_+$ be the cone of symmetric positive semidefinite matrices. We use the notation $X \succeq 0$ to denote that X is a symmetric positive semidefinite matrix.

*Department of Mathematics, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076 (kaifengjiang@nus.edu.sg).

†Department of Mathematics and Risk Management Institute, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076 (matsundf@nus.edu.sg).

‡Department of Mathematics, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076 (mattohkc@nus.edu.sg); and Singapore-MIT Alliance, 4 Engineering Drive 3, Singapore 117576.

In this paper, we consider the following nuclear norm regularized matrix least squares problem with linear equality constraints:

$$\min_{X\in\Re^{p\times q}} \Big\{\frac{1}{2}\|\mathcal{A}(X)-b\|^2+\rho\|X\|_*+\langle C,X\rangle \ : \ \mathcal{B}(X)\ =\ d\Big\}, \tag{1}$$

where $\mathcal{A}:\Re^{p\times q}\to\Re^m$ and $\mathcal{B}:\Re^{p\times q}\to\Re^s$ are given linear maps, $C\in\Re^{p\times q}, b\in\Re^m, d\in\Re^s$, and ρ is a given positive parameter. Note that the nuclear norm has been a very popular regularizer which favors a low rank solution of (1) [5, 8, 9, 20]. The problem (1) arises in many applications when one needs to find a low rank approximation of a given matrix while preserving certain desired structures. In many data analysis problems, the collected empirical data, which are usually messy and incomplete, typically do not have the specified structure or the desired low rank. So it is important to find the nearest low rank approximation of the given matrix while maintaining the underlying structure of the original system. For example, in statistics, the regression matrix for the multiple regression model with a constant term has a column of all ones, and this column should not be perturbed during the low rank approximation.

When $C=0$ and either $\mathcal{A}$ or $\mathcal{B}$ is absent in (1), the problem (1) includes the well studied matrix completion problem if $\mathcal{A}$ or $\mathcal{B}$ is the projection onto the set of observed matrix entries. We should mention that many specialized first-order algorithms have been designed for various variants of the matrix completion problem; see for example [4, 18, 25, 15]. But as far as we are aware of, none of the specialized algorithms are designed to handle completion problems with additional structural constraints.

In this paper, we design a partial proximal point algorithm (PPA) proposed by Ha [12] for solving (1), in which only some of the variables appear in the quadratic proximal term. Given a sequence of parameters σ_k such that

$$0<\sigma_k\uparrow\sigma_\infty\le+\infty, \tag{2}$$

and an initial point $X^0\in\Re^{p\times q}$, the partial PPA for solving (1) generates a sequence $\{(u^k,X^k)\}\subseteq\Re^m\times\Re^{p\times q}$ via the following scheme:

$$(u^{k+1},X^{k+1})\approx\arg\min\Big\{f_\rho(u,X)+\frac{1}{2\sigma_k}\|X-X^k\|^2\ :\ \mathcal{A}(X)+u\ =\ b,\ \mathcal{B}(X)\ =\ d\Big\}, \tag{3}$$

where $f_\rho(u,X):=\dfrac{1}{2}\|u\|^2+\rho\|X\|_*+\langle C,X\rangle$. A key issue in the partial PPA which we must address is how to solve the partially regularized problem (3) efficiently. In our algorithm, we solve (3) via its dual, which is an unconstrained concave maximization problem whose objective function is continuously differentiable but not twice continuously differentiable. Because of the latter property, standard Newton's method cannot be used to solve the inner subproblem. However, we can show that the objective function is strongly semismooth due to the strong semismoothness of the soft thresholding operator [14, Theorem 2.1]. Thus we can apply the semismooth Newton method of Qi and Sun [19] to solve the inner subproblem. Recently Zhao, Sun and Toh [28] proposed a Newton-CG augmented Lagrangian (SDPNAL) method for solving SDP problems, in which the inner subproblems

are solved by using an inexact semismooth Newton-CG method. Their numerical results on a variety of large scale SDP problems demonstrated that the SDPNAL method is very efficient. This strongly motivated us to use a semismooth Newton-CG (SSNCG) method to solve the inner subproblems for achieving fast convergence. For our case, the global and fast local convergence of the SSNCG method is established under a constraint nondegeneracy condition, together with the strong semismoothness property of the soft thresholding operator.

The partial PPA which we will develop for solving (1) can easily be modified to solve the following semidefinite matrix least squares problem:

$$\min_{X\in\mathcal{S}^n}\Big\{\frac{1}{2}\|\mathcal{A}(X)-b\|^2+\langle C,\,X\rangle\;:\;\mathcal{B}(X)\,=\,d,\;X\succeq\,0\Big\},\tag{4}$$

where $\mathcal{A}:\mathcal{S}^n\to\Re^m$ and $\mathcal{B}:\mathcal{S}^n\to\Re^s$ are given linear maps, $b\in\Re^m, d\in\Re^s$, and $C\in\mathcal{S}^n$. Thus in this paper, we also design a partial PPA to solve (4).

For the partial PPA (with SSNCG method for solving the inner subproblems) we have designed and implemented, numerical experiments on large scale matrix least squares problems arising from low rank matrix approximation, as well as regularized kernel estimation and Euclidean distance matrix completion problems in molecular conformation show that our algorithm is efficient and robust.

The remaining parts of this paper are organized as follows. In section 2, we present some preliminaries about semismooth functions. In section 3, we describe how to use the partial PPA to solve (1) and introduce a SSNCG method for solving the inner subproblems. The convergence analysis of our proposed algorithm is also established. In section 4, we briefly explain how the SSNCG partial PPA for solving (1) can be modified to solve (4). In section 5, we report the numerical performance of our algorithm for solving the various classes of problems mentioned in the last paragraph. We conclude the paper in section 6.

2 Preliminaries

In this section, we give a brief introduction on some basic concepts such as the B-subdifferential and Clarke's generalized Jacobian of the soft-thresholding operator. These concepts and properties will be critical for us to develop a SSNCG method for solving the inner subproblems in our partial PPA.

Let $F:\Re^m\longrightarrow\Re^l$ be a locally Lipschitz function. By Rademacher's theorem, F is Fréchet differentiable almost everywhere. Let D_F denote the set of points where F is differentiable. The B-subdifferential of F at $x\in\Re^m$ is defined by

$$\partial_B F(x):=\{V:\;V=\lim_{k\to\infty}F'(x^k),\;x^k\longrightarrow x,\;x^k\in D_F\},$$

where $F'(x)$ denotes the Jacobian of F at $x\in D_F$. Then Clarke's [6] generalized Jacobian of F at $x\in\Re^m$ is defined as the convex hull of $\partial_B F(x)$, i.e., $\partial F(x)=\mathrm{conv}\{\partial_B F(x)\}$.

Let $Y\in\Re^{p\times q}$ admit the following singular value decomposition (SVD):

$$Y=U[\Sigma\;\;0]V^T,\tag{5}$$

where $U \in \Re^{p\times p}$ and $V \in \Re^{q\times q}$ are orthogonal matrices, $\Sigma = \mathrm{Diag}(\sigma_1, \cdots, \sigma_p)$ is the diagonal matrix of singular values of Y, with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$. Define $g_\rho : \Re \to \Re$ by

$$g_\rho(t) := (t-\rho)_+ - (-t-\rho)_+. \tag{6}$$

For each parameter $\rho > 0$, the soft-thresholding operator $D_\rho : \Re^{p\times q} \to \Re^{p\times q}$ is defined by

$$D_\rho(Y) = U[\Sigma_\rho \;\; 0]V^T, \tag{7}$$

where $\Sigma_\rho = \mathrm{Diag}(g_\rho(\sigma_1), \ldots, g_\rho(\sigma_p))$. From [14, Thorem 2.1], we know that $D_\rho(\cdot)$ is strongly semismooth everywhere in $\Re^{p\times q}$. Decompose $V \in \Re^{q\times q}$ into the form $V = [V_1 \; V_2]$, where $V_1 \in R^{q\times p}$ and $V_2 \in \Re^{q\times(q-p)}$. Let the orthogonal matrix $Q \in \Re^{(p+q)\times(p+q)}$ be defined by

$$Q := \frac{1}{\sqrt{2}} \begin{bmatrix} U & U & 0 \\ V_1 & -V_1 & \sqrt{2}V_2 \end{bmatrix}, \tag{8}$$

and $\Xi : \Re^{p\times q} \to \mathcal{S}^{p+q}$ be defined by

$$\Xi(Y) := \begin{bmatrix} 0 & Y \\ Y^T & 0 \end{bmatrix}, \quad Y \in \Re^{p\times q}. \tag{9}$$

Then, by [10, Section 8.6], we know that the symmetric matrix $\Xi(Y)$ has the following spectral decomposition:

$$\Xi(Y) \;=\; Q \begin{bmatrix} \Sigma & 0 & 0 \\ 0 & -\Sigma & 0 \\ 0 & 0 & 0 \end{bmatrix} Q^T, \tag{10}$$

i.e., the eigenvalues of $\Xi(Y)$ are $\pm\sigma_i, i = 1, \ldots, p$, and 0 of multiplicity $q-p$. For any $W = P\mathrm{Diag}(\lambda_1, \cdots, \lambda_{p+q})P^T \in \mathcal{S}^{p+q}$, define $G_\rho : \mathcal{S}^{p+q} \to \mathcal{S}^{p+q}$ by

$$G_\rho(W) := P\mathrm{Diag}(g_\rho(\lambda_1), \cdots, g_\rho(\lambda_{p+q}))P^T = (W-\rho I)_+ - (-W-\rho I)_+,$$

where $(\cdot)_+$ denotes the projection onto the cone of positive semidefinite matrices. By direct calculations, we have

$$\Psi(Y) \;:=\; G_\rho(\Xi(Y)) \;=\; Q \begin{bmatrix} \Sigma_\rho & 0 & 0 \\ 0 & -\Sigma_\rho & 0 \\ 0 & 0 & 0 \end{bmatrix} Q^T = \begin{bmatrix} 0 & D_\rho(Y) \\ D_\rho(Y)^T & 0 \end{bmatrix}. \tag{11}$$

Note that (11) provides an easy way for us to calculate the derivative (if it exists) of D_ρ at Y, as we shall see in Proposition 2.1. For later discussion, we define the following three index sets:

$$\alpha := \{1, \ldots, p\}, \quad \gamma := \{p+1, \ldots, 2p\}, \quad \beta := \{2p+1, \ldots, p+q\}. \tag{12}$$

For any $\lambda = (\lambda_1, \ldots, \lambda_{p+q})^T \in \Re^{p+q}$ and $\lambda_i \neq \pm\rho, i = 1, \ldots, p+q$, we denote by Ω the $(p+q)\times(p+q)$ first divided difference symmetric matrix of $g_\rho(\cdot)$ at λ [2] whose (i,j)th entry is given by

$$\Omega_{ij} = \begin{cases} \dfrac{g_\rho(\lambda_i) - g_\rho(\lambda_j)}{\lambda_i - \lambda_j} & \text{if } \lambda_i \neq \lambda_j, \\ g'_\rho(\lambda_i) & \text{if } \lambda_i = \lambda_j. \end{cases}$$

Proposition 2.1. *Let $Y \in \Re^{p\times q}$ admit the SVD in (5). If $\sigma_i \neq \rho, i = 1,\ldots,p$, then $D_\rho(\cdot)$ is differentiable at Y, and for any $H \in \Re^{p\times q}$, we have*

$$D'_\rho(Y)H = U\Big[\left(\Omega_{\alpha\alpha}\circ\left(\frac{H_1+H_1^T}{2}\right)+\Omega_{\alpha\gamma}\circ\left(\frac{H_1-H_1^T}{2}\right)\right)V_1^T+(\Omega_{\alpha\beta}\circ H_2)\,V_2^T\Big], \quad (13)$$

where $H_1 = U^THV_1$ and $H_2 = U^THV_2$.

Proof. For any $\lambda = (\lambda_1,\ldots,\lambda_{p+q})^T \in \Re^{p+q}$, let $\lambda_i = \sigma_i$ for $i \in \alpha$, $\lambda_i = -\sigma_{i-p}$ for $i \in \gamma$, and $\lambda_i = 0$ for $i \in \beta$. Since $\sigma_i \neq \rho, i = 1,\ldots,p$, from (10) and (11) we can obtain the first divided difference matrix for $g_\rho(\cdot)$ at λ:

$$\Omega = \begin{pmatrix} \Omega_{\alpha\alpha} & \Omega_{\alpha\gamma} & \Omega_{\alpha\beta} \\ \Omega_{\alpha\gamma}^T & \Omega_{\gamma\gamma} & \Omega_{\gamma\beta} \\ \Omega_{\alpha\beta}^T & \Omega_{\gamma\beta}^T & \Omega_{\beta\beta} \end{pmatrix}. \quad (14)$$

Since $g_\rho(\cdot)$ is an odd function, we have the following results:

$$\Omega_{\gamma\gamma} = \Omega_{\alpha\alpha}, \quad \Omega_{\alpha\gamma} = \Omega_{\alpha\gamma}^T, \quad \Omega_{\gamma\beta} = \Omega_{\alpha\beta}, \quad \Omega_{\beta\beta} = 0.$$

Now, by a result of Löwner [16], we have from (11) that for any $H \in \Re^{p\times q}$,

$$\Psi'(Y)H = G'_\rho(\Xi(Y))\Xi(H) = Q\big[\Omega\circ(Q^T\Xi(H)Q)\big]Q^T.$$

Since

$$Q^T\Xi(H)Q = \frac{1}{2}\begin{bmatrix} H_1+H_1^T & H_1^T-H_1 & \sqrt{2}H_2 \\ H_1-H_1^T & -(H_1+H_1^T) & \sqrt{2}H_2 \\ \sqrt{2}H_2^T & \sqrt{2}H_2^T & 0 \end{bmatrix}, \quad (15)$$

by simple algebraic calculations, we have that

$$\Psi'(Y)H = Q\big[\Omega\circ(Q^T\Xi(H)Q)\big]Q^T = \begin{bmatrix} 0 & M_{12} \\ M_{12}^T & 0 \end{bmatrix}, \quad (16)$$

where $M_{12} = U\Big[\left(\Omega_{\alpha\alpha}\circ\left(\frac{H_1+H_1^T}{2}\right)+\Omega_{\alpha\gamma}\circ\left(\frac{H_1-H_1^T}{2}\right)\right)V_1^T+(\Omega_{\alpha\beta}\circ H_2)\,V_2^T\Big]$. Since

$$\Psi'(Y)H \;=\; \begin{bmatrix} 0 & D'_\rho(Y)H \\ (D'_\rho(Y)H)^T & 0 \end{bmatrix},$$

we have from (16) that $D'_\rho(Y)H = M_{12}$ □

Next, we give a characterization of the generalized Jacobian of $D_\rho(\cdot)$, which was presented in [27, Lemma 2.3.6 and Proposition 2.3.7]. For any $\lambda = (\lambda_1,\ldots,\lambda_{p+q})^T \in \Re^{p+q}$, let $\lambda_i = \sigma_i$ for $i \in \alpha$, $\lambda_i = -\sigma_{i-p}$ for $i \in \gamma$, and $\lambda_i = 0$ for $i \in \beta$. For each threshold parameter $\rho > 0$, we decompose the index set α into the following three subindex sets:

$$\alpha_1 := \{i \,|\, \sigma_i(Y) > \rho, i \in \alpha\},\ \alpha_2 := \{i \,|\, \sigma_i(Y) = \rho, i \in \alpha\},\ \alpha_3 := \{i \,|\, \sigma_i(Y) < \rho, i \in \alpha\}. \quad (17)$$

Let Γ denote the following $(p+q)\times(p+q)$ symmetric matrix

$$\Gamma = \begin{pmatrix} \Gamma_{\alpha\alpha} & \Gamma_{\alpha\gamma} & \Gamma_{\alpha\beta} \\ \Gamma_{\alpha\gamma}^T & \Gamma_{\gamma\gamma} & \Gamma_{\gamma\beta} \\ \Gamma_{\alpha\beta}^T & \Gamma_{\gamma\beta}^T & \Gamma_{\beta\beta} \end{pmatrix}, \tag{18}$$

whose (i,j)th entry is given by

$$\Gamma_{ij} = \begin{cases} \dfrac{g_\rho(\lambda_i) - g_\rho(\lambda_j)}{\lambda_i - \lambda_j} & \text{if } \lambda_i \neq \lambda_j, \\ 1 & \text{if } \lambda_i = \lambda_j \text{ and } |\lambda_i| > \rho, \\ \in \partial g_\rho(\lambda_i) = [0,1] & \text{if } \lambda_i = \lambda_j \text{ and } |\lambda_i| = \rho, \\ 0 & \text{if } \lambda_i = \lambda_j \text{ and } |\lambda_i| < \rho. \end{cases} \tag{19}$$

Proposition 2.2. *Let $Y \in \Re^{p\times q}$ admit the SVD in (5). Then, for any $\mathcal{V} \in \partial_B \Psi(Y)$ and any $H \in \Re^{p\times q}$, we have*

$$\mathcal{V}(H) = Q(\Gamma \circ (Q^T \Xi(H) Q))Q^T. \tag{20}$$

Moreover, for any $\mathcal{W} \in \partial_B D_\rho(Y)$, we have

$$\mathcal{W}(H) = U\Big[\left(\Gamma_{\alpha\alpha} \circ \left(\frac{H_1 + H_1^T}{2} \right) + \Gamma_{\alpha\gamma} \circ \left(\frac{H_1 - H_1^T}{2} \right) \right) V_1^T + (\Gamma_{\alpha\beta} \circ H_2)\, V_2^T \Big], \tag{21}$$

where $H_1 = U^T H V_1$, $H_2 = U^T H V_2$, and

$$\Gamma_{\alpha\alpha} = \begin{pmatrix} \tau_{\alpha_1\alpha_1} & \tau_{\alpha_1\alpha_2} & \tau_{\alpha_1\alpha_3} \\ \tau_{\alpha_1\alpha_2}^T & \nu_{\alpha_2\alpha_2} & 0 \\ \tau_{\alpha_1\alpha_3}^T & 0 & 0 \end{pmatrix}, \quad \begin{array}{c} \tau_{ij} = 1, \text{ for } i \in \alpha_1,\ j \in \alpha_1 \cup \alpha_2, \\ \tau_{ij} = \dfrac{\sigma_i - \rho}{\sigma_i - \sigma_j}, \text{ for } i \in \alpha_1,\ j \in \alpha_3, \\ \nu_{ij} = \nu_{ji} \in [0,1], \text{ for } i,j \in \alpha_2, \end{array} \tag{22}$$

$$\Gamma_{\alpha\gamma} = \begin{pmatrix} \omega_{\alpha_1\alpha_1} & \omega_{\alpha_1\alpha_2} & \omega_{\alpha_1\alpha_3} \\ \omega_{\alpha_1\alpha_2}^T & 0 & 0 \\ \omega_{\alpha_1\alpha_3}^T & 0 & 0 \end{pmatrix},\ \omega_{ij} := \frac{(\sigma_i - \rho)_+ + (\sigma_j - \rho)_+}{\sigma_i + \sigma_j},\ \text{for } i \in \alpha_1,\ j \in \alpha, \tag{23}$$

$$\Gamma_{\alpha\beta} = \begin{pmatrix} \mu_{\alpha_1\bar{\beta}} \\ 0 \end{pmatrix},\ \bar{\beta} = \beta - 2p = \{1, \ldots, q-p\},\ \mu_{ij} = \frac{\sigma_i - \rho}{\sigma_i},\ \text{for } i \in \alpha_1,\ j \in \bar{\beta}. \tag{24}$$

Proof. See [27, Lemma 2.3.6 and Proposition 2.3.7]. □

Let the operator $\mathcal{W}^0 : \Re^{p\times q} \to \Re^{p\times q}$ be defined by

$$\mathcal{W}^0(H) = U\Big[\left(\Gamma^0_{\alpha\alpha} \circ \left(\frac{H_1 + H_1^T}{2} \right) + \Gamma_{\alpha\gamma} \circ \left(\frac{H_1 - H_1^T}{2} \right) \right) V_1^T + (\Gamma_{\alpha\beta} \circ H_2)\, V_2^T \Big], \tag{25}$$

where $\Gamma^0_{\alpha\alpha}$ is of the form (22) with $(\Gamma^0_{\alpha\alpha})_{\alpha_2\alpha_2} = 0$. Then we have that $\mathcal{W}^0$ is an element in $\partial_B D_\rho(Y)$.

3 A partial proximal point algorithm for matrix least squares problems

In this section, we will show how to use the partial proximal point algorithm (PPA) to solve the problem (1).

It is easy to see that (1) can be rewritten as follows:

$$\min_{u\in\Re^m, X\in\Re^{p\times q}} \Big\{ f_\rho(u,X) := \frac{1}{2}\|u\|^2 + \rho\|X\|_* + \langle C, X\rangle \ : \ \mathcal{A}(X) + u = b, \mathcal{B}(X) = d \Big\}. \quad (26)$$

Note that the objective function $f_\rho(u,X)$ is strongly convex in u for all $X \in \Re^{p\times q}$. Let $l(u,X;\zeta,\xi) : \Re^m \times \Re^{p\times q} \times \Re^m \times \Re^s \to \Re$ be the Lagrangian function for (26):

$$l(u,X;\zeta,\xi) := f_\rho(u,X) + \langle \zeta, b - \mathcal{A}(X) - u\rangle + \langle \xi, d - \mathcal{B}(X)\rangle. \quad (27)$$

Then the essential objective function in (26) is

$$f(u,X) \ := \sup_{\zeta\in\Re^m,\ \xi\in\Re^s} l(u,X;\zeta,\xi) = \begin{cases} f_\rho(u,X) & \text{if } (u,X)\in\mathcal{F}_P, \\ +\infty & \text{if } (u,X)\notin\mathcal{F}_P, \end{cases} \quad (28)$$

where $\mathcal{F}_P = \{(u,X) \in \Re^m \times \Re^{p\times q} \mid \mathcal{A}(X) + u = b, \mathcal{B}(X) = d\}$ is the feasible set of (26). The dual problem of (26) is given by:

$$\max \Big\{ g_\rho(\zeta,\xi) \ : \ \mathcal{A}^*(\zeta) + \mathcal{B}^*(\xi) + Z = C,\ \|Z\|_2 \le \rho,\ \zeta\in\Re^m,\ \xi\in\Re^s,\ Z\in\Re^{p\times q} \Big\}, \quad (29)$$

where $g_\rho(\zeta,\xi) := -\frac{1}{2}\|\zeta\|^2 + \langle b,\zeta\rangle + \langle d,\xi\rangle$. Since $f(u,X)$ is strongly convex in u for all $X \in \Re^{p\times q}$, we apply the partial PPA proposed by Ha [12] to the maximal monotone operator $\mathcal{T}_f = \partial f$, in which only the variable X appears in the quadratic proximal term. Let $\Pi : \Re^m \times \Re^{p\times q} \to \Re^m \times \Re^{p\times q}$ be the orthogonal projector of $\Re^m \times \Re^{p\times q}$ onto $\{0\} \times \Re^{p\times q}$, i.e., $\Pi(u,X) = (0,X)$ and $P_\sigma := (\Pi + \sigma\mathcal{T}_f)^{-1}\Pi$ for a given positive parameter σ. From [14, Proposition 3.1], we know that the operator P_σ is single-valued. Given a starting point $(u^0,X^0) \in \Re^m \times \Re^{p\times q}$, the partial PPA for solving problem (26) can be expressed as follows:

$$(u^{k+1},X^{k+1}) \ \approx \ P_{\sigma_k}(u^k,X^k) := \operatorname*{argmin}_{u\in\Re^m, X\in\Re^{p\times q}} \left\{ f(u,X) + \frac{1}{2\sigma_k}\|X - X^k\|^2 \right\}, \quad (30)$$

where the sequence $\{\sigma_k\}$ satisfies (2). Note that for the standard PPA, the map Π in P_σ is replaced by the identity map.

Next we compute the partial quadratic regularization of f in (30), which plays a key role in the study of the partial PPA for solving (26). For a given parameter $\sigma > 0$, the partial quadratic regularization of f in (28) associated with σ is given by

$$F_\sigma(X) \ = \ \min_{u\in\Re^m, Y\in\Re^{p\times q}} \Big\{ f(u,Y) + \frac{1}{2\sigma}\|Y - X\|^2 \Big\}. \quad (31)$$

From [14, Section 3], we have that

$$F_\sigma(X) \ = \ \sup_{\zeta\in\Re^m,\ \xi\in\Re^s} \theta_\sigma(\zeta,\xi;X),$$

where

$$\theta_\sigma(\zeta,\xi;X) := -\frac{1}{2}\|\zeta\|^2 + \langle b,\zeta\rangle + \langle d,\xi\rangle + \frac{1}{2\sigma}\|X\|^2 - \frac{1}{2\sigma}\|D_{\rho\sigma}(W(\zeta,\xi;X))\|^2 \tag{32}$$

and $W(\zeta,\xi;X) = X - \sigma(C - \mathcal{A}^*\zeta - \mathcal{B}^*\xi)$. By the saddle point theorem [21, Theorem 28.3], we have that for any

$$(\zeta(X),\xi(X)) \in \operatorname*{argsup}_{\zeta\in\Re^m,\ \xi\in\Re^s} \theta_\sigma(\zeta,\xi;X),$$

the point $(\zeta(X), D_{\rho\sigma}(W(\zeta(X),\xi(X);X)))$ is the unique solution to (31).

Now we formally present the partial PPA for solving (26).

Algorithm 1. Given a tolerance $\varepsilon > 0$, $(u^0, X^0) \in \Re^m \times \Re^{p\times q}$, $\sigma_0 > 0$. Set $k = 0$. Iterate:

Step 1. Compute an approximate maximizer

$$(\zeta^{k+1},\xi^{k+1}) \approx \operatorname*{argsup}_{\zeta\in\Re^m,\,\xi\in\Re^s} \theta_{\sigma_k}(\zeta,\xi;X^k), \tag{33}$$

where $\theta_{\sigma_k}(\zeta,\xi;X^k)$ is defined in (32).

Step 2. Compute $W^{k+1} := W(\zeta^{k+1},\xi^{k+1};X^k)$. Set

$$u^{k+1} = \zeta^{k+1},\ X^{k+1} = D_{\rho\sigma_k}(W^{k+1}), \quad Z^{k+1} = \frac{1}{\sigma_k}(D_{\rho\sigma_k}(W^{k+1}) - W^{k+1}).$$

Step 3. If $\|(X^k - X^{k+1})/\sigma_k\| \le \varepsilon$; stop; else; update σ_k; end.

The global and local convergence of Algorithm 1 for solving (26) has been established in [14]. For details on the convergence analysis, we refer the reader to [14, Theorem 3.1 & Theorem 3.2].

3.1 A semismooth Newton-CG method for solving unconstrained inner subproblems

In this subsection, we introduce a semismooth Newton-CG (SSNCG) method for solving the unconstrained inner subproblem (33), which is the most expensive step in each PPA iteration. For later convenience, we let

$$\widehat{\mathcal{A}} = \begin{pmatrix}\mathcal{A}\\ \mathcal{B}\end{pmatrix},\ \widehat{b} = (b;d) \in \Re^{m+s},\ T = \begin{bmatrix} I_m & 0\\ 0 & 0\end{bmatrix} \in \Re^{(m+s)\times(m+s)},\ y = (\zeta;\xi) \in \Re^{m+s}. \tag{34}$$

For the convergence analysis, we assume that the following Slater condition holds:

$$\begin{cases} \mathcal{B} : \Re^{p\times q} \to \Re^s \text{ is onto,}\\ \exists X_0 \in \Re^{p\times q} \text{ such that } \mathcal{B}(X_0) \ = \ d. \end{cases} \tag{35}$$

In our proposed partial PPA, for some fixed $X \in \Re^{p\times q}$ and $\sigma > 0$, we need to solve an inner subproblem of the following form:

$$\min_{y\in\Re^{m+s}} \Big\{\varphi(y) := \frac{1}{2}\langle y,\, Ty\rangle + \frac{1}{2\sigma}\|D_{\rho\sigma}(W(y;X))\|^2 - \langle \widehat{b},\, y\rangle\Big\}, \tag{36}$$

where $W(y;X) = X - \sigma(C - \widehat{\mathcal{A}}^* y)$ and $\widehat{\mathcal{A}}^* = (\mathcal{A}^*,\, \mathcal{B}^*)$ is the adjoint of $\widehat{\mathcal{A}}$. The optimality condition for (36) is given by

$$\nabla\varphi(y) = Ty + \widehat{\mathcal{A}} D_{\rho\sigma}(W(y;X)) - \widehat{b} = 0. \tag{37}$$

Since the soft-thresholding operator $D_{\rho\sigma}(\cdot)$ is Lipschitz continuous with modulus 1 [15, 14], the mapping $\nabla\varphi(y)$ is Lipschitz continuous on $\Re^{m+s}$. Thus for any $y \in \Re^{m+s}$, the generalized Hessian of $\varphi(y)$ is well defined and it is given by

$$\partial^2\varphi(y) \;:=\; \partial(\nabla\varphi)(y), \tag{38}$$

where $\partial(\nabla\varphi)(y)$ is the Clarke's generalized Jacobian of $\nabla\varphi$ at y [6]. However, it is hard to express $\partial^2\varphi(y)$ exactly, so we define the following alternative for $\partial^2\varphi(y)$,

$$\hat{\partial}^2\varphi(y) \;:=\; T + \sigma\widehat{\mathcal{A}}\,\partial D_{\rho\sigma}(W(y;X))\widehat{\mathcal{A}}^*. \tag{39}$$

From [6, p.75], we have for any $h \in \Re^{m+s}$,

$$\partial^2\varphi(y)h \subseteq \hat{\partial}^2\varphi(y)h, \tag{40}$$

which implies that if all elements in $\hat{\partial}^2\varphi(y)$ are positive definite, so are those in $\partial^2\varphi(y)$.

Since the soft-thresholding operator $D_{\rho\sigma}(\cdot)$ is strongly semismooth, $\nabla\varphi(\cdot)$ is also strongly semismooth. We can solve the nonlinear equation (37) by using a SSNCG method for which the direction r at an iterate y is computed from the following linear system of equations:

$$\underbrace{(T + \sigma\widehat{\mathcal{A}}\mathcal{W}\widehat{\mathcal{A}}^*)}_{\mathcal{V}}\; r \;=\; -\nabla\varphi(y), \tag{41}$$

where $\mathcal{W}$ is any element in $\partial D_{\rho\sigma}(W(y;X))$. Note that if $s = 0$, i.e., the constraint $\mathcal{B}X = d$ is absent, then $\mathcal{V}$ is always positive definite due to the fact that all the elements in $\partial D_{\rho\sigma}(\cdot)$ are positive semidefinite [14, Proposition 2.1] and $T = I_m$.

Define the operator $\mathcal{W}_y^0 : \Re^{p\times q} \to \Re^{p\times q}$ as in (25). To implement the above SSNCG method, we need to choose an explicit element $\mathcal{W}$ in $\partial D_{\rho\sigma}(W(y;X))$, which we take to be $\mathcal{W}_y^0$. With this specific choice, then the coefficient matrix in (41) is given by

$$\mathcal{V}_y^0 = T + \sigma\widehat{\mathcal{A}}\,\mathcal{W}_y^0\,\widehat{\mathcal{A}}^* \in \hat{\partial}^2\varphi(y). \tag{42}$$

Next, we shall study a certain constraint nondegeneracy condition and its connection to the positive definiteness of $\mathcal{V}_y \in \hat{\partial}^2\varphi(y)$. Suppose that the Slater condition (35) holds and $\overline{y} = (\overline{\zeta};\overline{\xi}) \in \Re^{m+s}$ is the optimal solution to problem (36). Let $W(\overline{y};X) = X - \sigma(C - \widehat{\mathcal{A}}^*\overline{y})$

and $\overline{X} = D_{\rho\sigma}(W(\overline{y}; X))$. Let $W(\overline{y}; X)$ admit the SVD as in (5). For the given threshold value $\rho\sigma$, we decompose the index set $\alpha = \{1, \ldots, p\}$ into the following three subindex sets:

$$\alpha_1 := \{i \,|\, \sigma_i(W) > \rho\sigma,\ i \in \alpha\},\ \alpha_2 := \{i \,|\, \sigma_i(W) = \rho\sigma,\ i \in \alpha\},\ \alpha_3 := \{i \,|\, \sigma_i(W) < \rho\sigma,\ i \in \alpha\}.$$

The constraint nondegeneracy condition is said to hold at $\overline{X}$ [14] if

$$\mathcal{B}(\mathcal{T}(\overline{X})) = \Re^s, \tag{43}$$

where the subspace $\mathcal{T}(\overline{X})$ of $\Re^{p\times q}$ is defined as

$$\mathcal{T}(\overline{X}) := \Big\{H \in \Re^{p\times q} \mid [U_{\alpha_2}\ U_{\alpha_3}]^T H [V_{\alpha_2}\ V_{\alpha_3}\ V_2] = 0\Big\}, \tag{44}$$

and its orthogonal complement is given by

$$\mathcal{T}^{\perp}(\overline{X}) = \Big\{H \in \Re^{p\times q} \mid U_{\alpha_1}^T H = 0,\ \ HV_{\alpha_1} = 0\Big\}. \tag{45}$$

The following lemma will be needed to analyze the connection between the constraint nondegeneracy condition at $\overline{X}$ and the positive definiteness of the elements of $\hat{\partial}^2\varphi(\overline{y})$.

Lemma 3.1. *Let $W(\overline{y}; X)$ admit the SVD as in (5). For any $\mathcal{W} \in \partial D_{\rho\sigma}(W(\overline{y}; X))$ and $H \in \Re^{p\times q}$ such that $\mathcal{W}H = 0$, it holds that*

$$H \in \mathcal{T}^{\perp}(\overline{X}). \tag{46}$$

Proof. Let $\mathcal{W} \in \partial D_{\rho\sigma}(W(\overline{y}; X))$ and $H \in \Re^{p\times q}$ be such that $\mathcal{W}H = 0$. Then we have

$$\begin{aligned} 0 &= \langle H,\, \mathcal{W}H\rangle = \frac{1}{2}\langle \Xi(H),\, \Xi(\mathcal{W}H)\rangle = \frac{1}{2}\langle \Xi(H),\, Q(\Gamma \circ (Q^T\Xi(H)Q))Q^T\rangle \\ &= \frac{1}{2}\langle Q^T\Xi(H)Q,\, \Gamma \circ (Q^T\Xi(H)Q)\rangle = \frac{1}{2}\langle \widetilde{H},\, \Gamma \circ \widetilde{H}\rangle, \end{aligned}$$

where $\Gamma \in \mathcal{S}^{p+q}$ is defined as in (18) and $\widetilde{H} = Q^T\Xi(H)Q$. Let $H_1 = U^THV_1, H_2 = U^THV_2, H_1^s = (H_1 + H_1^T)/2$ and $H_1^a = (H_1 - H_1^T)/2$. From (15) and (21), we have

$$0 = \frac{1}{2}\langle \widetilde{H},\, \Gamma \circ \widetilde{H}\rangle = \sum_{i\in\alpha}\sum_{j\in\alpha}\Gamma_{ij}(H_1^s)_{ij}^2 + \sum_{i\in\alpha}\sum_{j\in\gamma}\Gamma_{ij}(H_1^a)_{ij}^2 + \sum_{i\in\alpha}\sum_{j\in\beta}\Gamma_{ij}(H_2)_{ij}^2.$$

Since $\Gamma_{ij} \in [0, 1]$ for all $i, j = 1, \ldots, p+q$, it follows that

$$\sum_{i\in\alpha}\sum_{j\in\alpha}\Gamma_{ij}(H_1^s)_{ij}^2 = 0, \quad \sum_{i\in\alpha}\sum_{j\in\gamma}\Gamma_{ij}(H_1^a)_{ij}^2 = 0, \quad \sum_{i\in\alpha}\sum_{j\in\beta}\Gamma_{ij}(H_2)_{ij}^2 = 0.$$

Then from (22), (23) and (24), we have that

$$(H_1^s)_{\alpha_1\alpha} = 0,\ (H_1^s)_{\alpha\alpha_1} = 0,\ (H_1^a)_{\alpha_1\alpha} = 0,\ (H_1^a)_{\alpha\alpha_1} = 0,\ (H_2)_{\alpha_1\bar{\beta}} = 0,$$

where $\bar{\beta} = \{1, \ldots, q-p\}$. Since $H_1 = H_1^s + H_1^a$, we have that $(H_1)_{\alpha_1\alpha} = 0$ and $(H_1)_{\alpha\alpha_1} = 0$. From $H_1 = [U_{\alpha_1}\, U_{\alpha_2}\, U_{\alpha_3}]^T H [V_{\alpha_1}\, V_{\alpha_2}\, V_{\alpha_3}]$ and $H_2 = [U_{\alpha_1}\, U_{\alpha_2}\, U_{\alpha_3}]^T HV_2$, we obtain that

$$U_{\alpha_1}^T HV_1 = 0, \quad U_{\alpha_1}^T HV_2 = 0, \quad U^T HV_{\alpha_1} = 0.$$

Since both U and $V = [V_1\ V_2]$ are orthogonal matrices, we have $U_{\alpha_1}^T H = 0, HV_{\alpha_1} = 0$, which means that $H \in \mathcal{T}^{\perp}(\overline{X})$. □

Proposition 3.1. *Suppose that the Slater condition (35) is satisfied. Let $\overline{y}$ be the optimal solution to problem (36), $W(\overline{y}; X) = X - \sigma(C - \widehat{\mathcal{A}}^*\overline{y})$ admit the SVD as in (5), and $\overline{X} = D_{\rho\sigma}(W(\overline{y}; X))$. Then the following conditions are equivalent:*

(a) The constraint nondegeneracy condition (43) holds at $\overline{X}$.

(b) Every $\mathcal{V}_{\overline{y}} \in \hat{\partial}^2\varphi(\overline{y})$ is symmetric and positive definite.

(c) $\mathcal{V}^0_{\overline{y}} \in \hat{\partial}^2\varphi(\overline{y})$ is symmetric and positive definite.

Proof. "(a) $\Rightarrow$ (b)". Let $\mathcal{V}_{\overline{y}}$ be an arbitrary element in $\hat{\partial}^2\varphi(\overline{y})$. Then there exists an element $\mathcal{W}_{\overline{y}} \in \partial D_{\rho\sigma}(W(\overline{y}; X))$ such that

$$\mathcal{V}_{\overline{y}} = T + \sigma\widehat{\mathcal{A}}\mathcal{W}_{\overline{y}}\widehat{\mathcal{A}}^* = T + \sigma \begin{bmatrix} \mathcal{A}\mathcal{W}_{\overline{y}}\mathcal{A}^* & \mathcal{A}\mathcal{W}_{\overline{y}}\mathcal{B}^* \\ \mathcal{B}\mathcal{W}_{\overline{y}}\mathcal{A}^* & \mathcal{B}\mathcal{W}_{\overline{y}}\mathcal{B}^* \end{bmatrix}. \tag{47}$$

Since $\mathcal{W}_{\overline{y}}$ is self-adjoint and positive semidefinite [14, Proposition 2.1], we have that $\mathcal{V}_{\overline{y}}$ is self-adjoint and positive semidefinite. From (47) we obtain that $\mathcal{V}_{\overline{y}}$ is positive definite if only if $\mathcal{B}\mathcal{W}_{\overline{y}}\mathcal{B}^*$ is positive definite. Hence, it it enough to show the positive definiteness of $\mathcal{B}\mathcal{W}_{\overline{y}}\mathcal{B}^*$. Let $h \in \Re^s$ be such that $\mathcal{B}\mathcal{W}_{\overline{y}}\mathcal{B}^*h = 0$. Then we have

$$0 = \langle h,\, \mathcal{B}\,\mathcal{W}_{\overline{y}}\,\mathcal{B}^*h\rangle = \langle \mathcal{B}^*h,\, \mathcal{W}_{\overline{y}}\,\mathcal{B}^*h\rangle \geq \langle \mathcal{W}_{\overline{y}}\,\mathcal{B}^*h,\, \mathcal{W}_{\overline{y}}\,\mathcal{B}^*h\rangle,$$

where the last inequality follows from [14, Proposition 2.1], which implies that $\mathcal{W}_{\overline{y}}\,(\mathcal{B}^*h) = 0$. From Lemma 3.1, we have $\mathcal{B}^*h \in \mathcal{T}(\overline{X})^\perp$. Since the constraint nondegeneracy condition holds at $\overline{X}$, there exists a $Y \in \mathcal{T}(\overline{X})$ such that $\mathcal{B}Y = h$. Then, we have

$$\langle h,\, h\rangle = \langle h,\, \mathcal{B}Y\rangle = \langle \mathcal{B}^*h,\, Y\rangle = 0.$$

Thus $h = 0$, which implies that $\mathcal{B}\mathcal{W}_{\overline{y}}\mathcal{B}^*$ is positive definite. Hence, $\mathcal{V}_{\overline{y}}$ is positive definite.

"(b) $\Rightarrow$ (c)". This is obviously true since $\mathcal{V}^0_{\overline{y}} \in \hat{\partial}^2\varphi(\overline{y})$.

"(c) $\Rightarrow$ (a)". Suppose that the constraint nondegeneracy condition (43) does not hold at $\overline{X}$. Then there exists a non-zero $h \in [\mathcal{B}\mathcal{T}(\overline{X})]^\perp$. And we have

$$0 = \langle h,\, \mathcal{B}Y\rangle = \langle H,\, Y\rangle \quad \forall\, Y \in \mathcal{T}(\overline{X}),$$

where $H = \mathcal{B}^*h$, which implies that $H \in \mathcal{T}(\overline{X})^\perp$. From (45), we have $U_{\alpha_1}^T H = 0$ and $HV_{\alpha_1} = 0$. Then it follows that

$$U_{\alpha_1}^T HV = U_{\alpha_1}^T H[V_1\ V_2] = 0 \quad \text{and} \quad U^T HV_{\alpha_1} = 0. \tag{48}$$

Let $H_1 = U^T HV_1$ and $H_2 = U^T HV_2$. We have from (48) that

$$(H_1)_{\alpha_1\alpha} = 0, \quad (H_1)_{\alpha\alpha_1} = 0, \quad \text{and } (H_2)_{\alpha_1\bar{\beta}} = 0,$$

where $\bar{\beta} = \{1, \ldots, q-p\}$, from which we can further have that

$$(H_1^s)_{\alpha_1\alpha} = 0, \quad (H_1^s)_{\alpha\alpha_1} = 0, \quad (H_1^a)_{\alpha_1\alpha} = 0, \quad \text{and } (H_1^a)_{\alpha\alpha_1} = 0,$$

where $H_1^s = (H_1 + H_1^T)/2$ and $H_1^a = (H_1 - H_1^T)/2$. Then we have

$$\Gamma^0_{\alpha\alpha} \circ (H_1^s) = 0, \quad \Gamma_{\alpha\gamma} \circ (H_1^a) = 0, \quad \text{and } \Gamma_{\alpha\beta} \circ H_2 = 0.$$

From the definition of $\mathcal{W}^0_{\overline{y}}$ in (25), it follows that $\mathcal{W}^0_{\overline{y}}(H) = 0$, and hence

$$\langle h, \mathcal{B}\mathcal{W}^0_{\overline{y}}\mathcal{B}^* h\rangle = \langle H, \mathcal{W}^0_{\overline{y}}(H)\rangle = 0. \tag{49}$$

Since $\mathcal{V}^0_{\overline{y}}$ is positive definite, it follows from (47) that $\mathcal{B}\mathcal{W}^0_{\overline{y}}\mathcal{B}^*$ is also positive definite. Then (49) implies that $h = 0$, which contradicts the assumption that $h \neq 0$. Hence, we have that (a) holds. □

Now we present the steps of the SSNCG algorithm for solving (36).

Algorithm 2: A semismooth Newton-CG method

Given $y^0 \in \Re^{m+s}$, $\eta \in (0,1)$, $\tau \in (0,1]$, $\tau_1, \tau_2 \in (0,1)$, and $c \in (0,1/2)$, $\delta \in (0,1)$. Set $t = 0$. Iterate:

Step 1. Compute $\eta_t := \min\{\eta, \|\nabla\varphi(y^t)\|^{1+\tau}\}$. Apply the CG method to find an approximation solution r^t to

$$(\mathcal{V}_t + \varepsilon_t I)\, r = -\nabla\varphi(y^t), \tag{50}$$

where $\mathcal{V}_t \in \hat{\partial}^2\varphi(y^t)$ is defined in (42) and $\varepsilon_t = \min\{\tau_2, \tau_1\|\nabla\varphi(y^t)\|\}$, so that r^t satisfies the following condition:

$$\|(\mathcal{V}_t + \varepsilon_t I)r^t + \nabla\varphi(y^t)\| \le \eta_t. \tag{51}$$

Step 2. Set $\alpha_t = \delta^{m_t}$, where m_t is the first nonnegative integer m for which

$$\varphi(y^t + \delta^m r^t) \;\le\; \varphi(y^t) + c\delta^m\langle r^t,\, \nabla\varphi(y^t)\rangle.$$

Step 3. Set $y^{t+1} = y^t + \alpha_t r^t$.

In Algorithm 2, since $\mathcal{V}_t$ is always positive semidefinite, the matrix $\mathcal{V}_t + \varepsilon_t I$ is positive definite as long as $\nabla\varphi(y^t) \neq 0$. From [28, Lemma 3.1], we know that the generated search direction r^t is always a descent direction.

To analyze the global convergence of Algorithm 2, we assume that $\nabla\varphi(y^t) \neq 0$ for any $t \ge 0$. The global convergence and the rate of local convergence of Algorithm 2 can be derived similarly as in [28].

Theorem 3.1. *Suppose that the Slater condition (35) holds. Then Algorithm 2 is well defined and any accumulation point $\overline{y}$ of $\{y^t\}$ generated by Algorithm 2 is an optimal solution to the inner subproblem (36).*

Proof. See [28, Theorem 3.4]. □

Theorem 3.2. *Suppose that the Slater condition (35) holds. Let $\overline{y}$ be an accumulation point of the infinite sequence $\{y^t\}$ generated by Algorithm 2 for solving the inner subproblem (36). Suppose also that at each step $t \geq 0$, the inexact direction r^t satisfies the accuracy condition in (51). Assume that the constraint nondegeneracy condition (43) holds at $\overline{X} := D_{\rho\sigma}(W(\overline{y}; X))$. Then the whole sequence $\{y^t\}$ convergence to $\overline{y}$ and*

$$\|y^{t+1} - \overline{y}\| = O(\|y^t - \overline{y}\|^{1+\tau}). \tag{52}$$

Proof. See [28, Theorem 3.5]. □

4 Semidefinite matrix least squares problems

In this section, we show that the partial PPA developed for solving (26) can easily be adapted for solving the semidefinite matrix least squares problem (4). It is easy to see that (4) can be rewritten as follows:

$$\min_{u\in\Re^m, X\in\mathcal{S}^n} \Big\{\frac{1}{2}\|u\|^2 + \langle C,\, X\rangle \;:\; \mathcal{A}(X) + u \,=\, b,\; \mathcal{B}(X) \,=\, d,\; X \succeq\, 0\Big\}. \tag{53}$$

The dual problem of (53) is given by:

$$\max_{\zeta\in\Re^m, \xi\in\Re^s, Z\in\mathcal{S}^n} \Big\{ -\frac{1}{2}\|\zeta\|^2 + \langle b,\, \zeta\rangle + \langle d,\, \xi\rangle \;:\; \mathcal{A}^*(\zeta) + \mathcal{B}^*(\xi) + Z \,=\, C,\; Z \,\succeq\, 0\Big\}. \tag{54}$$

For some fixed $X \in \mathcal{S}^n$ and $\sigma > 0$, the partial quadratic regularization of problem (53) is given by:

$$\min_{u\in\Re^m, Y\in\mathcal{S}^n} \Big\{\frac{1}{2}\|u\|^2 + \langle C,\, Y\rangle + \frac{1}{2\sigma}\|Y - X\|^2 \;:\; \mathcal{A}(Y) + u \,=\, b,\; \mathcal{B}(Y) \,=\, d,\; Y \succeq\, 0\Big\}, \tag{55}$$

and the Lagrangian dual problem of (55) is given by

$$\max_{\zeta\in\Re^m, \xi\in\Re^s} \theta_\sigma(\zeta, \xi; X) := \inf_{u\in\Re^m, Y\succeq 0} L_\sigma(u, Y; \zeta, \xi, X)\,, \tag{56}$$

where

$$\begin{aligned} L_\sigma(u, Y; \zeta, \xi, X) &= \frac{1}{2}\|u\|^2 + \langle C,\, Y\rangle + \frac{1}{2\sigma}\|Y - X\|^2 + \langle \zeta,\, b - \mathcal{A}(Y) - u\rangle + \langle \xi,\, d - \mathcal{B}(Y)\rangle \\ &= \frac{1}{2}\|u\|^2 - \langle \zeta, u\rangle + \langle b, \zeta\rangle + \langle d, \xi\rangle + \frac{1}{2\sigma}\|Y - W(\zeta, \xi; X)\|^2 + \frac{1}{2\sigma}(\|X\|^2 - \|W(\zeta, \xi; X)\|^2), \end{aligned}$$

where $W(\zeta, \xi; X) = X - \sigma(C - \mathcal{A}^*\zeta - \mathcal{B}^*\xi)$. By minimizing $L_\sigma(u, Y; \zeta, \xi, X)$ over $Y \succeq 0$, we have

$$\theta_\sigma(\zeta, \xi; X) = -\frac{1}{2}\|\zeta\|^2 + \langle b, \zeta\rangle + \langle d, \xi\rangle + \frac{1}{2\sigma}\|X\|^2 - \frac{1}{2\sigma}\|\Pi_{\mathcal{S}^n_+}(W(\zeta, \xi; X))\|^2, \tag{57}$$

where $\Pi_{\mathcal{S}^n_+}(\cdot)$ is the metric projector of $\mathcal{S}^n$ onto $\mathcal{S}^n_+$. The problem (56) is an unconstrained continuously differentiable convex optimization problem, and it can be efficiently solved by

the SSNCG method developed in [28]. The SSNCG method for solving (56) is analogous to Algorithm 2 where for some fixed $X \in \mathcal{S}^n$ and $\sigma > 0$, the function φ is now given by

$$\varphi(y) = \frac{1}{2}\langle y,\, Ty\rangle + \frac{1}{2\sigma}\|\Pi_{\mathcal{S}^n_+}(W(y;X))\|^2 - \langle \widehat{b},\, y\rangle$$

and the operator $\mathcal{V}_t$ in (50) is replaced by

$$\mathcal{V}_t = T + \sigma\widehat{\mathcal{A}}\Pi'_{\mathcal{S}^n_+}(W(y^t;X))\widehat{\mathcal{A}}^*, \tag{58}$$

where $\Pi'_{\mathcal{S}^n_+}(W(y^t;X))$ denotes an element of $\partial\Pi_{\mathcal{S}^n_+}(W(y^t;X))$.

The fast local convergence of the SSNCG method for solving (56) can be established in a similar fashion as Theorem 3.2 where the positive definiteness of the element $\mathcal{V}_{\overline{y}}$ defined in (58) at the optimal solution $\overline{y}$ is again equivalent to a constraint nondegeneracy condition similar to (43) at $\overline{X} := \Pi_{\mathcal{S}^n_+}(W(\overline{y};X))$.

5 Numerical Results

In this section, we report some numerical results to demonstrate the efficiency of our SSNCG partial PPA. We implemented our algorithm in MATLAB 2011a (version 7.12), and the numerical experiments are run in MATLAB under a Linux operating system on an Intel Core 2 Duo 2.40GHz CPU with 2GB memory.

We measure the infeasibilities and optimality for the primal problem (26) and the dual problem (29) as follows:

$$R_P = \frac{\|\widehat{b} - (\zeta;0) - \widehat{\mathcal{A}}(X)\|}{1+\|\widehat{b}\|},\ R_D = \frac{\|C - \widehat{\mathcal{A}}^*y - Z\|}{1+\|\widehat{\mathcal{A}}^*\|},\ \text{relgap} = \frac{f_\rho(\zeta,X) - g_\rho(\zeta,\xi)}{1+|f_\rho(\zeta,X)| + |g_\rho(\zeta,\xi)|},$$

where $y = (\zeta;\xi), Z = (D_{\rho\sigma}(W) - W)/\sigma$ with $W = X - \sigma(C - \widehat{\mathcal{A}}^*y)$, and $f_\rho(\zeta,X)$ and $g_\rho(\zeta,\xi)$ are the objective functions of the primal and dual problems, respectively. The infeasibility of the condition $\|Z\|_2 \le \rho$ is not checked since it is satisfied up to machine precision throughout the algorithm. In our numerical experiments, we stop the partial PPA when

$$\max\{R_P, R_D\} \le \texttt{Tol}, \tag{59}$$

where `Tol` is a pre-specified accuracy tolerance. We use the alternating direction method of multipliers (ADMM) method [7] applied to (29) to generate a reasonably good starting point for our SSNCG partial PPA. Unless otherwise specified, we set the parameter ρ in (1) to be $\rho = 10^{-3}\|\mathcal{A}^*b\|_2$ and $\texttt{Tol} = 10^{-6}$ as the default.

5.1 Example 1

We consider the low rank matrix approximation problem in which certain specified entries of the matrix are fixed. In [11], Golub, Hoffman and Stewart derived an explicit formula for finding the nearest lower-rank approximation of the target matrix while certain specified

$p/q/\tau$	r	m	it.|itsub|cg	R_p | R_D | relgap	MSE	#sv	time
500/500/0.0	10	99189	9.0 | 31.0 | 19.3	2.44e-7 | 4.59e-7 | -8.83e-5	1.34e-3	10	48
	50	250000	8.2 | 9.2 | 10.0	2.64e-7 | 7.26e-7 | -6.47e-5	1.42e-3	50	17
	100	250000	9.0 | 9.4 | 7.0	2.27e-7 | 4.75e-7 | -1.28e-5	1.65e-3	100	17
1000/1000/0.0	10	199104	10.2 | 30.2 | 18.8	2.18e-7 | 8.05e-7 | -1.57e-4	1.32e-3	10	4:48
	50	974891	8.6 | 10.8 | 22.2	2.12e-7 | 3.84e-7 | -2.48e-5	1.31e-3	50	2:30
	100	1000000	9.0 | 10.0 | 8.0	6.85e-8 | 1.69e-7 | -6.69e-6	1.44e-3	100	1:38
1500/1500/0.0	10	299272	13.0 | 32.4 | 16.9	1.93e-7 | 1.92e-7 | -1.69e-5	1.30e-3	10	16:11
	50	1474562	9.0 | 24.0 | 27.1	1.24e-7 | 2.89e-7 | -1.78e-5	1.34e-3	50	15:31
	100	2250000	9.0 | 9.6 | 12.0	1.17e-7 | 1.95e-7 | -4.20e-6	1.37e-3	100	5:56
500/500/0.1	10	99189	20.0 | 42.6 | 10.0	1.78e-7 | 6.74e-7 | -2.96e-5	8.32e-2	10	1:00
	50	250000	9.0 | 11.2 | 9.8	2.00e-7 | 4.52e-7 | -1.40e-5	9.67e-2	50	22
	100	250000	9.8 | 11.6 | 7.9	8.44e-8 | 2.55e-7 | -6.45e-6	9.77e-2	100	26
1000/1000/0.1	10	199104	20.0 | 46.4 | 10.5	4.97e-7 | 7.95e-7 | -8.79e-5	7.75e-2	10	6:15
	50	974891	18.0 | 21.4 | 11.2	2.30e-7 | 6.04e-7 | -7.86e-7	9.50e-2	50	4:24
	100	1000000	14.0 | 14.0 | 6.2	1.68e-9 | 5.46e-7 | -1.95e-5	9.67e-2	100	2:42
1500/1500/0.1	10	299272	21.4 | 48.6 | 10.7	1.17e-7 | 8.67e-7 | -1.46e-4	7.50e-2	10	20:51
	50	1474562	21.8 | 40.8 | 11.1	2.35e-7 | 6.45e-7 | -3.32e-6	8.90e-2	50	21:43
	100	2250000	9.0 | 11.2 | 13.1	1.18e-7 | 5.60e-7 | -1.18e-5	9.59e-2	100	8:06
100/5000/0.0	10	500000	8.0 | 10.0 | 12.8	1.47e-7 | 2.06e-7 | -4.58e-5	1.25e-3	10	58
100/5000/0.1	10	500000	9.0 | 12.0 | 8.7	6.39e-8 | 1.16e-7 | -9.34e-6	9.64e-2	10	1:04
100/10000/0.0	10	1000000	8.0 | 10.0 | 14.6	8.67e-8 | 1.96e-7 | -4.55e-5	1.25e-3	10	2:30
100/10000/0.1	10	1000000	9.0 | 12.0 | 9.2	8.31e-8 | 1.01e-7 | -8.31e-6	9.64e-2	10	2:36
100/20000/0.0	10	2000000	8.0 | 10.0 | 15.4	8.11e-8 | 2.01e-7 | -4.71e-5	1.25e-3	10	5:11
100/20000/0.1	10	2000000	13.0 | 14.0 | 6.5	8.41e-10 | 2.87e-7 | -3.24e-5	9.64e-2	10	5:05

Table 1: Numerical performance of the partial PPA on (60). In the table, $m = 10d$ and $d = r(p + q - r)$.

columns of the matrix are fixed. In our numerical experiments, we assume that only partial information of the original matrix is available and the specified fixed entries can be in any random position of the original matrix. For each triplet (p, q, r), we first generate a random matrix $M \in \Re^{p\times q}$ by setting $M = M_1 M_2^T$ where $M_1 \in \Re^{p\times r}$, $M_2 \in \Re^{q\times r}$ have i.i.d. Gaussian entries. Then we sample a subset $\mathcal{E}$ of m entries of M uniformly at random, and generate a random matrix $N_{\mathcal{E}} \in \Re^{p\times q}$ with sparsity pattern $\mathcal{E}$ and i.i.d Gaussian entries. Then we assume that the observed data is given by $\widetilde{M_{\mathcal{E}}} = M_{\mathcal{E}} + \tau N_{\mathcal{E}} \|M_{\mathcal{E}}\|/\|N_{\mathcal{E}}\|$, where τ is the noise factor. The minimization problem which we solve can be stated as follows:

$$\min_{X\in\Re^{p\times q}} \Big\{ \frac{1}{2}\|X_{\mathcal{E}} - \widetilde{M_{\mathcal{E}}}\|_F^2 + \rho\|X\|_* \ : \ X_{i_t,j_t} = M_{i_t,j_t}, 1 \le t \le k \Big\}, \tag{60}$$

where $(i_1, j_1), \ldots, (i_k, j_k)$ are distinct pairs. In our numerical experiments, we set $k = \lceil 10^{-3} pq \rceil$, which is the number of prescribed entries selected uniformly at random, the noise level $\tau = 0$, 0.1 and the number of sampled entries to be $m = 10d$, where $d = r(p + q - r)$ is the degree of freedom in an $p \times q$ matrix of rank r.

For each triplet $(p, q, r), m$ and τ, we generate 5 random instances. In Table 1, we report the average number of the following quantities: number of sampled entries (m); number of outer iterations (it); total number of inner iterations (itsub); number of CG steps taken to

solve each linear system in (50) (cg); infeasibilities (R_p, R_D); relative duality gap (relgap); relative mean square error MSE := $\|X - M\|/\|M\|$; numerical rank of X (#sv); and the CPU time taken. Here we report the numerical rank of X defined as follows:

$$\#\mathrm{sv}(X) := \max\{k \,:\, \sigma_k(X) \geq \max\{10^{-8}, \tau\}\sigma_1(X)\}. \tag{61}$$

We may observe from the table that our algorithm is very efficient for solving (60). For the problem where p is moderate but q is large, e.g., $p = 100$ and $q = 20000$, it takes about 5 minutes to solve the last instance to achieve the tolerance of 10^{-6} while the MSE is reasonably small.

5.2 Example 2

In the Euclidean metric embedding problem, we are given an incomplete, possibly noisy, dissimilarity matrix $B \in \mathcal{S}^n$ with $\mathrm{Diag}(B) = 0$ and sparsity pattern specified by the set of indices $\mathcal{E} = \{(i,j) \mid B_{ij} \neq 0, 1 \leq i < j \leq n\}$. The goal is to find an Euclidean distance matrix (EDM) [1] that is nearest to B. If the measure of nearness is in the Frobenius norm, then the mathematical formulation of the problem is as follows:

$$\min\Big\{\frac{1}{2}\sum_{(i,j)\in\mathcal{E}} W_{ij}(D_{ij} - B_{ij})^2 + \frac{\rho}{2n}\langle E,\, D\rangle \,:\, D \text{ is an EDM}\Big\}, \tag{62}$$

where $W_{ij} > 0, (i,j) \in \mathcal{E}$, are given weights, $E \in \mathcal{S}^n$ is the matrix of all ones and $\rho > 0$ is a regularization parameter. Here we add the term $\frac{\rho}{2n}\langle E,\, D\rangle$ to encourage a sparse solution. Recall that a standard characterization [1] of an EDM D is that $D = \mathrm{Diag}(X)e^T + e\,\mathrm{Diag}(X)^T - 2X$ for some $X \succeq 0$ with $Xe = 0$, where $e \in \Re^n$ is the vector of all ones. Thus the problem (62) can be rewritten as:

$$\min\Big\{\frac{1}{2}\sum_{(i,j)\in\mathcal{E}} W_{ij}(\langle A_{ij},\, X\rangle - B_{ij})^2 + \rho\langle I,\, X\rangle \,:\, \langle E,\, X\rangle = 0,\; X \succeq 0\Big\}, \tag{63}$$

where $A_{ij} = (e_i - e_j)(e_i - e_j)^T$ and e_i is the i-th standard unit vector in $\Re^n$. Note that under the condition $X \succeq 0$, the constraint $Xe = 0$ is equivalent to $\langle E,\, X\rangle = 0$. It is interesting to note that desiring sparsity in the EDM D leads to the regularization term $\rho\langle I,\, X\rangle$, which is a proxy for desiring a low-rank X.

Let $m = |\mathcal{E}|$. The linear maps $\mathcal{A} : \mathcal{S}^n \to \Re^m$ and $\mathcal{B} : \mathcal{S}^n \to \Re$ for the problem (63) are given as follows:

$$(\mathcal{A}(X))_{(i,j)} = \langle\sqrt{W_{ij}}A_{ij},\, X\rangle,\; \forall\,(i,j) \in \mathcal{E}, \quad \mathcal{B}(X) = \langle E,\, X\rangle.$$

Note that the components of a vector in $\Re^m$ are enumerated based on the elements in $\mathcal{E}$. And the operator $\mathcal{V}_t$ in (58) is given as follows:

$$\mathcal{V}_t = \sigma\Big(\frac{1}{\sigma}T + \begin{bmatrix}\mathcal{A}\\ \mathcal{B}\end{bmatrix}\Pi'_{\mathcal{S}^n_+}(W(y^t; X))\,[\mathcal{A}^* \;\; \mathcal{B}^*]\Big). \tag{64}$$

For the EDM problem (63), the condition number of $\mathcal{V}_t$ can be quite large and it is important to find a good preconditioner for $\mathcal{V}_t$ so that the CG method can have a reasonable convergence speed when solving the linear system of equations associated with $\mathcal{V}_t$. Let $\mathbf{A}, \mathbf{B}$ and $\mathbf{S}$ be the matrix representations of $\mathcal{A}, \mathcal{B}$ and $\Pi'_{\mathcal{S}^n_+}(W(y^t; X))$ with respect to the standard basis of $\mathcal{S}^n$ and $\Re^m$, respectively. Let $\mathbf{h} \in \Re^{n(n+1)/2}$ and $\Gamma = \{(i,j) \,|\, 1 \le i \le j \le n\}$. Suppose Diag($\mathbf{h}$) is a positive definite diagonal approximation of $\mathbf{S}$. (In our implementation, we choose $\mathbf{h}$ to be the approximate diagonal of $\Pi'_{\mathcal{S}^n_+}(W(y^t; X))$ considered in [24].) Let $H \in \mathcal{S}^n$ be the matrix such that $H_{ij} = H_{ji} = \mathbf{h}_{(i,j)}$ for all $(i,j) \in \Gamma$. We consider the following approximation of $\frac{1}{\sigma}\mathcal{V}_t$:

$$\mathcal{M} = \begin{bmatrix} M & q \\ q^T & \alpha \end{bmatrix}, \tag{65}$$

where $q = \mathcal{A}(H) \in \Re^m$, $\alpha = \langle E,\, H \rangle$, and

$$M = \frac{1}{\sigma} I_m + \mathbf{A}\mathrm{Diag}(\mathbf{h})\mathbf{A}^T \in \Re^{m \times m}. \tag{66}$$

Note that the rows and columns of M are enumerated based on the elements of $\mathcal{E}$. We have that

$$M_{(i,j),(s,t)} = \begin{cases} 0 & \text{if } i \ne s,\ j \ne t, \\ \sqrt{W_{ij}W_{st}}\, H_{ss} & \text{if } i = s,\ j \ne t, \\ \sqrt{W_{ij}W_{st}}\, H_{tt} & \text{if } i \ne s,\ j = t, \\ 1/\sigma + \sqrt{W_{ij}W_{st}}(H_{ss} + H_{tt} + 2H_{st}) & \text{if } i = s,\ j = t. \end{cases}$$

Let $\bar{\mathbf{h}}_{(i,j)} = W_{ij}H_{ij}$ for all $(i,j) \in \mathcal{E}$. Then we know that M has the following structure

$$M = \mathbf{D} + JJ^T, \tag{67}$$

where $\mathbf{D} = \frac{1}{\sigma} I_m + 2\mathrm{Diag}(\bar{\mathbf{h}})$, and $J \in \Re^{m \times n}$ is the weighted arc-node incidence matrix where the entry at the (s,t)-th row and k-th column is given by

$$J_{(s,t),k} = \begin{cases} \sqrt{W_{st}H_{ss}} & \text{if } k = s, \\ \sqrt{W_{st}H_{tt}} & \text{if } k = t, \\ 0 & \text{otherwise.} \end{cases}$$

To use $\mathcal{M}$ as a preconditioner for $\mathcal{V}_t$, we need the inverse of $\mathcal{M}$, which is given in the following expression:

$$\mathcal{M}^{-1} = \begin{bmatrix} S^{-1} & -\alpha^{-1}S^{-1}q \\ -\alpha^{-1}q^TS^{-1} & \alpha^{-1} + \alpha^{-2}q^TS^{-1}q \end{bmatrix}, \tag{68}$$

where

$$S = M - \alpha^{-1}qq^T = \mathbf{D} + \underbrace{[J,\, q]}_{\widehat{J}} \begin{bmatrix} I_n & 0 \\ 0 & -\alpha^{-1} \end{bmatrix} \begin{bmatrix} J^T \\ q^T \end{bmatrix}.$$

By using the Sherman-Morrison-Woodbury formula [10], we have that

$$S^{-1} = \mathbf{D}^{-1} - \mathbf{D}^{-1}\widehat{J}(\Lambda + \widehat{J}^T\mathbf{D}^{-1}\widehat{J})^{-1}\widehat{J}^T\mathbf{D}^{-1}, \tag{69}$$

where $\Lambda = [I_n, 0; \, 0, -\alpha]$. Here we assume that $\Lambda + \widehat{J}^T\mathbf{D}^{-1}\widehat{J}$ is nonsingular; otherwise we may consider the following block diagonal approximation of $\frac{1}{\sigma}\mathcal{V}_t$:

$$\mathcal{M}^d = \begin{bmatrix} M & 0 \\ 0 & \alpha \end{bmatrix}, \tag{70}$$

where the inverse of M can also be computed via the Sherman-Morrison-Woodbury formula.

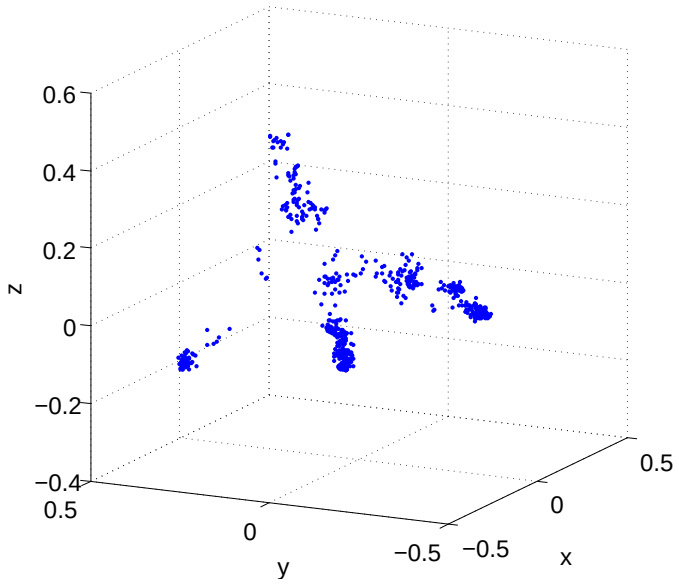

Figure 1: A 3D representation of the sequence space for 630 proteins.

problem	n	m	ρ	it.|itsub|cg	R_p | R_D | relgap	#sv	time
RKE630	630	198136	5.07e-1	6 | 36| 10.4	2.67e-7 | 2.26e-8 | -1.54e-6	388	1:22
PDB25	1898	1646031	1.84e+0	23 | 68 | 22.0	1.73e-7 | 6.28e-7 | -1.33e-6	1371	57:56

Table 2: Numerical results on the RKE problem arising from protein clustering.

The Euclidean metric problem arises in many applications. For the regularized kernel estimation (RKE) problem in statistics [17], we are given a set of n objects and dissimilarity measures d_{ij} for certain object pairs $(i, j) \in \mathcal{E}$. The goal is to estimate a positive semidefinite kernel matrix $X \in \mathcal{S}_+^n$ such that the fitted squared distances induced by X between the objects satisfy the following condition:

$$X_{ii} + X_{jj} - 2X_{ij} = \langle A_{ij}, \, X \rangle \;\approx\; d_{ij}^2 \quad \forall \, (i, j) \in \mathcal{E}.$$

Formally, one version of the RKE problem proposed in [17] is the SDP problem (63).

In our numerical experiments, the data d_{ij} are normalized to be in the interval [0, 1], and $\mathcal{E} = \{(i,j) : 1 \le i < j \le 630\}$. We set $W_{ij} = 1$ for all $(i,j) \in \mathcal{E}$. In [17], due to the prohibitive computational load encountered by the standard interior-point solver (such as SDPT3 or SeDuMi) used to solve (63), a subset of 280 globin proteins were selected from the entire set of 630 proteins for the numerical results reported in [17]. And for each of the selected proteins, 55 dissimilarities were randomly selected out of the total of 279. Here we are able to consider the entire set of 630 proteins and the dissimilarities among all the pairs of proteins.

As mentioned in [17], the RKE methodology can provide an efficient way to represent each protein sequence by a feature vector in an appropriate coordinate system using the pairwise dissimilarity between protein sequences. Specifically, we project the computed solution $\overline{X}$ onto a 3D space corresponding to the largest three eigenvalues. Figure 1 displays a 3D representation of the sequence space for 630 proteins. There are at least 4 classes visually identifiable in the data set of 630 proteins, which is consistent with the observations in [17]. The numerical results for solving (63) are reported in Table 2, where #sv is the number of positive eigenvalues of $\overline{X}$. For the computed solution $\overline{X}$, we have $\langle \overline{X}, E\rangle = 4.46 \times 10^{-13}$ and $\langle \overline{X}, I\rangle = 1.85 \times 10^{2}$.

We also conducted numerical experiments on a much larger protein data set to evaluate the performance of our algorithm. We used the `PDB_SELECT 25` data set [13], a representative subset of the Protein Data Bank database, which contains 1898 protein chains. The numerical results for the `PDB_SELECT 25` data set are reported in Table 2. For the computed solution $\overline{X}$, we have $\langle \overline{X}, E\rangle = -1.94 \times 10^{-14}$ and $\langle \overline{X}, I\rangle = 8.76 \times 10^{2}$.

5.3 Example 3: Molecular conformation problems

The molecular conformation problem for a molecule with n atoms is the problem of determining the positions $x_1, \ldots, x_n$ of the atoms, given estimated inter-atomic distances d_{ij} between some pairs of atoms. The estimated distances could be information derived from covalent bond lengths or measured from nuclear magnetic resonance (NMR) experiments. Let $\mathcal{E}$ be the set of pairs of indices (i,j) $(i<j)$ for which estimates on the distances $\|x_i - x_j\|$ are available. The molecular conformation problem can mathematically be formulated as follows:

$$\min \Big\{ \frac{1}{2} \sum_{(i,j)\in\mathcal{E}} W_{ij}(\|x_i - x_j\|^2 - d_{ij}^2)^2 - \frac{\rho}{2n} \sum_{i,j=1}^{n} \|x_i - x_j\|^2 \;\mid\; \sum_{i=1}^{n} x_i = 0 \Big\}, \tag{71}$$

where $W \in \mathcal{S}^n$ is a weight matrix with positive entries. The second term in the objective function is added to maximize the pairwise separations between the atoms. The equality constraint in (71) is included to set the center of mass of the molecule at the origin. The inclusion of weights is useful for differentiating data that are derived from different sources, and hence of different reliability. For example, distance data which are derived from covalent bond lengths are usually much more accurate than data which are derived from the NMR experiments.

Let $X = [x_1, \ldots, x_n] \in \Re^{3\times n}$ and $A_{ij} = e_{ij}e_{ij}^T$, where $e_{ij} = e_i - e_j$. Then we have $x_i - x_j = Xe_{ij}$ and hence $\|x_i - x_j\|^2 = \langle X^TX, A_{ij}\rangle$. Let $Y = X^TX$. The constraint $\sum_{i=1}^n x_i = 0$ can equivalently be replaced by $\langle E, Y\rangle = 0$. Note that under the latter constraint, it is easy to see that $\sum_{i,j=1}^n \|x_i - x_j\|^2 = 2n\langle I, Y\rangle$. By relaxing the non-convex constraint $Y = X^TX$ to $Y \succeq 0$ in (71), we get the following SDP problem:

$$\min\Big\{\frac{1}{2}\sum_{(i,j)\in\mathcal{E}} W_{ij}(\langle A_{ij}, Y\rangle - d_{ij}^2)^2 - \rho\langle I, Y\rangle \;\mid\; \langle E, Y\rangle = 0, Y \succeq 0\Big\}. \tag{72}$$

Observe that under the constraints $\langle E, Y\rangle = 0$ and $Y \succeq 0$, we have $\langle I, Y\rangle = \frac{1}{2n}\langle E, D\rangle$, where $D = \mathrm{Diag}(Y)e^T + e\,\mathrm{Diag}(Y)^T - 2Y$ and $e \in \Re^n$ is a vector of all ones. Thus (72) seeks an Euclidean distance matrix D which is encouraged to have as many nonzero entries as possible. Note that in the maximum variance unfolding problem [26], one also consider a problem that has exactly the same form as (72).

In this example, we focus on molecular conformation problems with noisy and sparse distance data. In our numerical experiments, we set $W_{ij} = 1$ for all $(i, j) \in \mathcal{E}$, $\rho = 10^{-2}\|\mathcal{A}^*b\|_2$ and $\texttt{Tol} = 10^{-5}$. For each molecule, we generated the partial inter-atomic distance matrix as follows. If the distance between two atoms is less than 6Å (1Å = 10^{-8}cm), which is nearly the maximal distance that the NMR experiment can measure between two atoms, the distance is chosen; otherwise no distance information about the pair is known. Since not all the distances below 6Å are known from NMR experiments, we randomly choose 30% of all the distance below 6Å in our experiment. For realistic molecular conformation problems, in which the exact distances are not known and only the lower bounds $\underline{d}_{ij}$ and upper bounds $\bar{d}_{ij}$ on distances are provided, we use the mean $d_{ij} = (\bar{d}_{ij} + \underline{d}_{ij})/2$ as the estimated distances. After selecting 30% of inter-atomic distances, we add certain amount of normal noise or uniform noise to the distances to generate the lower and upper bounds. Suppose that $\hat{d}_{ij}$ is the exact distance between atom i and atom j, we set

$$\underline{d}_{ij} = \max(1, (1 - |\underline{\varepsilon}_{ij}|)\hat{d}_{ij}), \quad \bar{d}_{ij} = (1 + |\bar{\varepsilon}_{ij}|)\hat{d}_{ij}.$$

Let τ be a given noise level. In the normal noise model, $\underline{\varepsilon}_{ij}, \bar{\varepsilon}_{ij} \sim \mathcal{N}(0, \pi\tau^2/2)$ are independent normal random variables. In the uniform noise model, $\underline{\varepsilon}_{ij}, \bar{\varepsilon}_{ij}$ are independent uniform random variables in the interval $[0, 2\tau]$. We said that the distances are corrupted by 20% noise if $\tau = 0.2$.

In Table 3 and Table 4, we report the numerical results on molecular conformation problems under the normal noise model and uniform noise model, respectively, where the root mean square deviation (RMSD) is used to measure the accuracy of the estimated positions. The RMSD is defined by the following formula:

$$\mathrm{RMSD} := \frac{1}{\sqrt{n}}\Big(\sum_{i=1}^n \|x_i - \hat{x}_i\|^2\Big)^{1/2}, \tag{73}$$

where x_i is the estimated position and $\hat{x}_i$ is the actual position. We can observe from Table 3 and Table 4 that the SDP solution Y of (72) is typically in a higher dimensional space, and a simple projection of Y onto the 3-dimensional space does not yield correct atomic

Molecule	$n; m+s$	it.|itsub|cg	R_p | R_D | relgap	RMSD	time
1GM2	166; 1119	29 | 69 | 25.8	4.13e-6 | 5.78e-6 | 8.82e-5	3.54 (1.06)	12
1PBM	388; 3145	29 | 86 | 33.3	4.82e-6 | 6.58e-6 | -2.01e-4	4.34 (1.83)	52
1PTQ	402; 2182	29 | 113 | 47.6	4.40e-6 | 6.05e-6 | -3.02e-4	4.73 (1.43)	1:21
1CTF	487; 2630	33 | 128 | 55.1	3.03e-6 | 7.47e-6 | -6.60e-4	4.92 (2.25)	2:31
1AU6	506; 4767	27 | 85 | 55.1	3.97e-6 | 7.27e-6 | -5.45e-4	4.46 (0.65)	1:55
1HOE	558; 3083	30 | 130 | 60.7	8.23e-6 | 8.54e-6 | -4.65e-4	5.03 (1.40)	3:36
1PHT	814; 5239	35 | 162 | 85.2	3.08e-6 | 7.66e-6 | -6.00e-4	5.47 (2.23)	12:54
1POA	914; 5045	32 | 196 | 106.0	9.85e-6 | 5.48e-6 | -6.65e-4	5.73 (1.95)	23:18
1AX8	1003; 5563	31 | 188 | 109.2	1.58e-6 | 6.18e-6 | -6.83e-4	5.56 (1.73)	28:08

Table 3: Numerical results on molecular conformation problems with 30% distances ≤ 6Å, which are corrupted by 20% normal noise.

Molecule	$n; m+s$	it.|itsub|cg	R_p | R_D | relgap	RMSD	time
1GM2	166; 1119	29 | 71 | 26.1	3.91e-6 | 5.23e-6 | 8.43e-5	3.53 (0.92)	11
1PBM	388; 3145	30 | 86 | 33.9	4.18e-6 | 5.78e-6 | -1.72e-4	4.32 (1.83)	50
1PTQ	402; 2182	29 | 112 | 45.5	5.17e-6 | 6.07e-6 | -2.99e-4	4.70 (1.67)	1:14
1CTF	487; 2630	33 | 136 | 55.9	2.92e-6 | 7.64e-6 | -6.69e-4	4.89 (2.23)	2:41
1AU6	506; 4767	28 | 89 | 53.0	3.67e-6 | 7.15e-6 | -5.32e-4	4.45 (1.51)	2:00
1HOE	558; 3083	30 | 121 | 60.0	6.59e-6 | 8.67e-6 | -4.75e-4	5.02 (1.53)	3:22
1PHT	814; 5239	35 | 164 | 87.2	4.11e-6 | 7.47e-6 | -5.86e-4	5.47 (2.75)	13:19
1POA	914; 5045	32 | 200 | 101.1	4.02e-6 | 5.47e-6 | -6.62e-4	5.73 (1.94)	22:56
1AX8	1003; 5563	35 | 199 | 96.3	3.97e-6 | 6.29e-6 | -6.99e-4	5.57 (1.40)	26:51

Table 4: Numerical results on molecular conformation problems with 30% distances ≤ 6Å, which are corrupted by 20% uniform noise.

positions with RMSD ≥ 3.5Å. As proposed in [3], the positions estimated can be further refined by applying a gradient descent method as a postprocessing step to the following problem:

$$\min \sum_{(i,j)\in\mathcal{E}} (\|x_i - x_j\| - d_{ij})^2 - \beta \sum_{i=1}^{n} \|x_i\|^2, \tag{74}$$

where β is a positive regularization parameter. In general, the gradient descent method is a local optimization method which does not deliver the global optimal solution of a non-convex problem, unless a good starting point is available. Fortunately, the SDP solution estimated from (72) can serve as a good initial point to the gradient descent method. For each molecule, the RMSD after refinement by applying the gradient descent method to (74) is also reported in the parenthesis next to the RMSD of the SDP computed positions. We can observe from the tables that after refinement, the estimated positions are fairly accurate with RMSD ≈ 2Å.

6 Conclusion

In this paper, we introduced a partial PPA for solving nuclear norm regularized and semidefinite matrix least squares problems with linear equality constraints. The inner subproblems are solved inexactly by a semismooth Newton-CG method, whose convergence analysis is established under a constraint nondegeneracy condition, together with the strong semismoothness property of the soft-thresholding operator and the metric projector $\Pi_{\mathcal{S}^n_+}$. Numerical experiments conducted on nuclear norm regularized matrix least squares problems, regularized kernel estimation problems and molecular conformation problems demonstrated that our algorithm is efficient and robust.

References

[1] A.Y. Alfakih, A. Khandani, H. Wolkowicz, *Solving Euclidean distance matrix completion problems via semidefinite programming*, Computational Optimization and Applications, 12 (1999), 13–30.

[2] R. Bhatia, Matrix Analysis, Springer-Verlag, NewYork, 1997.

[3] P. Biswas, K.C. Toh, and Y. Ye, *A distributed SDP approach for large scale noisy anchor-free graph realization with applications to molecular conformation*, SIAM Journal on Scientific Computing, 30 (2008), 1251–1277.

[4] J. F. Cai, E. J. Candès and Z. Shen, *A singular value thresholding algorithm for matrix completion*, SIAM J. on Optimization 20 (2010), 1956–1982.

[5] E. J. Candès and B. Recht, *Exact matrix completion via convex optimization*, Foundations of Computational Mathematics, 9 (2009), 717–772.

[6] F. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley and Sons, New York, 1983.

[7] J. Eckstein and D. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Mathematical Programming, 55 (1992), pp. 293-318.

[8] M. Fazel, Matrix Rank Minimization with Applications. PhD thesis, Stanford University, 2002.

[9] M. Fazel, H. Hindi, and S. Boyd, *A rank minimization heuristic with application to minimum order system approximation*. In Proceedings of the American Control Conference, 2001.

[10] G.H. Golub and C.F. van Loan, Matrix Computations, The Johns Hopkins University Press, Baltimore, USA, Third Edition, 1996.

[11] G.H. Golub, A. Hoffman, and G.W. Stewart, *A generalization of the Eckart-Young-Mirsky matrix approximation theorem*, Linear Algebra and its applications, 88 (1987), 317–327.

[12] C. Ha, *A generalization of the proximal point algorithm*, SIAM Journal on Control and Optimization, 28 (1990), 503–512.

[13] J. Hou, S. R. Jun, C. Zhang, and S. H. Kim, *Global mapping of the protein structure space and application in structure-based inference of protein function*, Proceedings of the National Academy of Sciences of the United States of America, 102 (2005), 3651–3656.

[14] K. F. Jiang, D. F. Sun, and K. C. Toh, *A partial proximal point algorithm for nuclear norm regularized matrix least squares problems with polyhedral constraints*, preprint, 2012.

[15] Y.J. Liu, D.F. Sun, and K.C. Toh, *An implementable proximal point algorithmic framework for nuclear norm minimization*, Mathematical Programming, accepted, 2011.

[16] K. Löwner, *Über monotone matrixfunktionen*, Mathematische Zeitschrift, 38 (1934), 177–216.

[17] F. Lu, S. Keleş, S. Wright, and G. Wahba, *Framework for kernel regularization with application to protein clustering*, Proceedings of the National Academy of Sciences of the United States of America, 102 (2005), 12332–12337.

[18] Ma, S., Goldfarb, D., and Chen, L., *Fixed point and bregman iterative methods for matrix rank minimization*, Mathematical Programming, 128 (2011), 321–353.

[19] L. Qi and J. Sun, A nonsmooth version of Newton's method, Mathematical Programming, 58 (1993), 353–367.

[20] B. Recht, M. Fazel and P.A. Parrilo, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM Review, 52 (2010), 471–501.

[21] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, 1970.

[22] R.T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. on Control and Optimization, 14 (1976), 877–898.

[23] R.T. Rockafellar, *Augmented Lagrangains and applications of the proximal point algorithm in convex programming*, Mathematics of Operation Research, 1 (1976), 97–116.

[24] Y. Gao and D. Sun, *Calibrating least squares semidefinite programming with equality and inequality constraints*, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 14321457.

[25] K.C. Toh and S.W. Yun,*An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems*, Pacific J. of Optimization 6 (2010), 615–640.

[26] K. Q. Weinberger, F. Sha, Q. Zhu, and L. K. Saul, *Graph Laplacian regularization for large-scale semidefinite programming*, in B. Schoelkopf, J. Platt, and T. Hofmann (eds.), Advances in Neural Information Processing Systems 19 (2007), 1489–1496, MIT Press, Cambridge, MA.

[27] Z. Yang, *A study on nonsymmetric matrix-valued functions*, master's thesis, Master thesis, Department of Mathematics, National University of Singapore, 2009.

[28] X. Y. Zhao, D. Sun, and K. C. Toh, *A Newton-CG augmented Lagrangian method for semidefinite programming*, SIAM Journal on Optimization, 20 (2010), 1737–1765.