SIAM J. OPTIM. © 2019 Society for Industrial and Applied Mathematics
Vol. 29, No. 4, pp. 2785–2813

COMPUTING THE BEST APPROXIMATION OVER THE INTERSECTION OF A POLYHEDRAL SET AND THE DOUBLY NONNEGATIVE CONE*

YING CUI†, DEFENG SUN‡, AND KIM-CHUAN TOH§

Abstract. This paper introduces an efficient algorithm for computing the best approximation of a given matrix onto the intersection of linear equalities, inequalities, and the doubly nonnegative cone (the cone of all positive semidefinite matrices whose elements are nonnegative). In contrast to directly applying the block coordinate descent type methods, we propose an inexact accelerated (two-) block coordinate descent algorithm to tackle the four-block unconstrained nonsmooth dual program. The proposed algorithm hinges on the superlinearly convergent semismooth Newton method to solve each of the two subproblems generated from the (two-)block coordinate descent, which have no closed form solutions due to the merger of the original four blocks of variables. The $O(1/k^2)$ iteration complexity of the proposed algorithm is established. Extensive numerical results over various large scale semidefinite programming instances from relaxations of combinatorial problems demonstrate the effectiveness of the proposed algorithm.

Key words. semidefinite programming, doubly nonnegative cone, semismooth Newton, acceleration, complexity

AMS subject classifications. 90C06, 90C22, 90C25

DOI. 10.1137/18M1175562

1. Introduction.

In this paper, we are interested in solving the best approximation problem over the intersection of affine spaces defined by linear equalities, inequalities, and the doubly nonnegative cone. Mathematically, the optimization problem takes the form of

$$
\begin{array}{lll}
(1) & \underset{X\in\mathbb{S}^n}{\text{minimize}} & \frac{1}{2}\,\|\,X-G\,\|^2 \\
& \text{subject to} & \mathcal{A}X=b, \quad \mathcal{B}X\geq d, \quad X\in\mathbb{S}^n_+, \quad X\geq 0,
\end{array}
$$

where $\mathcal{A}:\mathbb{S}^n\to\mathbb{R}^{m_E}$, $\mathcal{B}:\mathbb{S}^n\to\mathbb{R}^{m_I}$ are linear operators, $G\in\mathbb{S}^n$, $b\in\mathbb{R}^{m_E}$, $d\in\mathbb{R}^{m_I}$ are given data, and $\mathbb{S}^n$ and $\mathbb{S}^n_+$ are the cones of all $n\times n$ symmetric matrices and positive semidefinite matrices, respectively.

Problem (1) is a special quadratic semidefinite programming that is challenging to solve when there is a large number of equality and inequality constraints. On one hand, pure first-order methods, such as the alternating direction method of multipliers, may need many iteration cycles until moderate accuracy can be reached. This is not desirable since within each cycle one needs to project a matrix onto the positive semidefinite cone that is presumably computationally intensive. On the other hand,

*Received by the editors March 14, 2018; accepted for publication (in revised form) September 4, 2019; published electronically October 31, 2019.
https://doi.org/10.1137/18M1175562
Funding: The second author's research is partially supported by a start-up research grant from the Hong Kong Polytechnic University. The third author's research is partially supported by the Ministry of Education of Singapore through ARF grant R-146-000-257-112.

†Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA 90089-0193 (yingcui@usc.edu).
‡Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong (defeng.sun@polyu.edu.hk).
§Department of Mathematics, National University of Singapore, Singapore, 119076 (mattohkc@nus.edu.sg).

the second-order interior point method suffers from the inherent ill-conditioning of the Newton system so that it is not effective to use the iterative algorithm to solve the corresponding large-scale linear equations in order to obtain the search direction. It is thus important to develop an efficient computational framework that properly integrates the first-order and second-order methods to solve the conic best approximation problem (1).

In this paper, by leveraging on the strong convexity of the best approximation problem, we take a dual approach to solve (1) based on the observation that its dual problem, with the form of

$$\underset{w}{\text{minimize}} \quad h(w) + \sum_{i=1}^{4} \varphi_i(w_i), \quad w \triangleq (w_1, w_2, w_3, w_4), \tag{2}$$

is a convex nonsmooth problem without coupling constraints between the blocks. Here h is a convex differentiable function whose gradient is Lipschitz continuous, and $\varphi_1, \ldots, \varphi_4$ are proper closed convex functions; see section 2 for the detailed derivations of this dual formulation. This nonsmooth-decoupling feature is a key reason for us to develop a sequential dual-updating scheme with convergence guarantee.

Naturally, one may consider the block coordinate (gradient) descent (BCD) method, whose computational complexity is at best $O(1/k)$, to solve the four block unconstrained problem (2). See the papers [55, 54, 6, 46, 27] and the recent survey [59] for extensive discussions of this method. A key factor for determining the efficiency of the BCD method is the number of blocks that are updated sequentially during one iteration, since there is always a trade-off between such a number of blocks and the difficulty for solving the subproblems for each block. One may notice that solving the conic program (1) is different in nature from solving those problems arising in computational statistics and machine learning. The properties of the latter problems, such as the low computational cost of calculating one component of the gradient or solving one subproblem, and the need for only low accuracy solutions, are conducive for the efficient implementations of a multiblock (usually at least hundreds-of-block) coordinate descent method. However, with the focus of solving the conic program (1) to a higher accuracy, we have the following issues to resolve:

- Treating (2) as a single block problem and solving all the variables simultaneously is difficult due to the degeneracy of this problem.
- Directly applying the four-block BCD method is inefficient as it potentially will need many more iteration cycles compared with those of three or fewer blocks. (This observation is indeed confirmed by the numerical experiments in section 5. As mentioned in the previous paragraph, within each iteration cycle, the algorithm may involve a computationally intensive step such as the projection onto the positive semidefinite cone.

To address the above issues, we propose to divide the four variables $(w_1, \ldots, w_4)$ into two groups and solve problem (2) via a two-block inexact majorized BCD method. The subproblems with regard to each group may be nondegenerate and relatively easy to solve by Newton-type methods. In order to make the overall algorithm converge fast, we also adopt Nesterov's acceleration technique for gradient type methods [39] in this alternating minimization scheme. In addition to incorporating the acceleration technique into the BCD framework, it is also important to allow the inexact computation of each block, which is essential if the resulting subproblems within each block are solved by iterative algorithms.

Related work. There are some pioneering papers on incorporating the techniques of blockwise updating and Nesterov's acceleration into solving multiblock convex optimization problems. In particular, Nesterov [40] studied the acceleration of the randomize BCD method for solving smooth convex problems. The dependence of the complexity bounds on the Lipschitz constant of the gradient for each block is further improved in [34, 36, 1, 41, 13]. With the presence of the separable nonsmooth functions in the objective, the accelerated randomized proximal coordinate gradient method is studied in [19, 20] and also allows parallel computation. The accelerated linear rates for the strongly convex case is further established in [35]. It has also been noticed in [45, 30] that proper incorporation of the second-order information in approximating the blocks can speed up the convergence rate of the randomized BCD method. When there are only two blocks of variables and the smooth part of the objective is a quadratic function, the accelerated BCD algorithm is discussed in [8]. Different from the other cited work in this paragraph, the updating order is deterministic in the latter work. It is worth mentioning that the deterministic (more precisely, cyclic) BCD method outperforms its random counterpart, both theoretically and numerically, for solving convex quadratic problems with a diagonally dominant Hessian matrix [25]. In fact, we have also observed similar superiority of the cyclic BCD over the randomized version in our numerical experiments.

Incorporating inexact updates into the accelerated (proximal) gradient methods is also an active research area. Both deterministic and stochastic types of error are considered in the existing literature. The former type is studied in [11] for smooth problems and in [49, 29] for nonsmooth problems. In addition, the paper [12] provides a comprehensive treatment of the inexact oracle that allows the application of the accelerated gradient method of smooth convex optimization to solve nonsmooth or weakly smooth convex problems. An additive stochastic noise model is considered in [33, 23, 24] and the optimal complexity bounds are derived for smooth convex problems, with or without the strong convexity of the objective function. A unified analysis of the deterministic and stochastic oracles for (constrained) smooth problems is conducted in [17, 10].

Our work is also closely related to the work of [51], which adopts an inexact accelerated BCD method to solve the doubly nonnegative best approximation problem. However, an additional regularization term related to the linear inequalities is added in [51], which leads to only *two* nonsmooth separable functions in the corresponding dual program. Furthermore, by Danskin's theorem [18, Theorem 10.2.1], one of the nonsmooth blocks can be solved implicitly and the accelerated proximal gradient (APG) method initiated by Nesterov [39] is thus applicable to the resulting problem. Different from [51], here we directly solve the original best approximation problem in (1) without adding any regularization term. This leads to a dual problem with *four* separable nonsmooth functions such that the algorithmic framework in [51] no longer applicable.

Our contributions. The key ingredient of our proposed algorithm is a combination of the *inexactness*, *deterministic blockwise-updating*, as well as the *Nesterov-type acceleration technique* to solve the dual of strongly convex conic optimization problems. The main contributions of our paper are as follows.

- Theoretically, we design an inexact two-block accelerated BCD method for solving (2) where h is a general smooth coupled function that is not necessarily quadratic. A key feature of the method is the inexactness framework that allows us to apply the two-block based method to solve the four-block problem (2) by viewing the four blocks as two larger blocks each containing two smaller blocks. Theoretically, we establish the $O(1/k^2)$ complexity of the proposed inexact accelerated BCD method. To the best of our knowledge, this is the first time that such optimal complexity is derived for the deterministic inexact (two)-block coordinate descent method. To achieve good practical performance, we suggest a proper way to merge several variables into one block so that only small proximal terms are added to the BCD subproblems. We also address the important issue of finding efficient algorithms to solve the BCD subproblems.

- Numerically, we provide an efficient solver based on the proposed method for solving the important class of best approximation problems of the form (1) that involves the positive semidefinite cone constraint and a large number of linear equality and inequality constraints. Our experiments on a large number of data instances from the Biq Mac Library maintained by A. Wiegele demonstrate that our solver is at least 3 to 4 times faster than other (accelerated) BCD-type methods.
- Finally, it is worth mentioning that besides being the dual formulation of the best approximation problem (1), problem (2) can be viewed as the dual of a more general class of strongly convex conic optimization problems,

$$
\begin{array}{ll} \underset{x\in\mathbb{X}}{\text{minimize}} & f(x)+\phi(x) \\ \text{subject to} & \mathcal{A}x=b, \quad g(x)\in\mathcal{C}, \quad x\in\mathcal{K}, \end{array} \tag{3}
$$

where $\mathbb{X}$, $\mathbb{Y}$, and $\mathbb{Z}$ are three finite-dimensional Euclidean spaces, $f:\mathbb{X}\to\mathbb{R}$ is a smooth and strongly convex function, $\phi:\mathbb{X}\to(-\infty,+\infty]$ is a closed proper convex function, $\mathcal{A}:\mathbb{X}\to\mathbb{Y}$ is a linear operator, $b\in\mathbb{Y}$ is the given data, $\mathcal{C}\subseteq\mathbb{Z}$ and $\mathcal{K}\subseteq\mathbb{X}$ are two closed convex cones, and $g:\mathbb{X}\to\mathbb{Z}$ is a smooth and $\mathcal{C}$-convex map for some closed convex set $\mathcal{C}$ (see, e.g., [47, Example 4′]) satisfying

$$
g(\alpha x_1+(1-\alpha)x_2) - [\,\alpha g(x_1)+(1-\alpha)g(x_2)\,] \in \mathcal{C} \quad \forall\, x_1,x_2 \in g^{-1}(\mathcal{C}),\ \forall\,\alpha\in(0,1).
$$

In order to make our discussions more general, we will directly derive the dual of (3) in the next section. In addition, many optimization problems themselves have the form of (2), such as the robust principle component analysis [58] and the robust matrix completion problem [31].

The rest of the paper is organized as follows. In the next section, we derive the dual form of problem (3) and propose the inexact majorized accelerated BCD method. Section 3 is devoted to the analysis of $O(1/k^2)$ iteration complexity of the proposed algorithm. In section 4, we describe the implementation of this inexact framework for solving the dual program (2). Newton-type algorithms for solving the subproblems are also discussed in this section. Numerical results are reported in section 5, where we show the effectiveness of our proposed algorithm via comparison with several variants of the BCD-type methods. We conclude our paper in section 6.

2. Formulation of the dual problem and the algorithmic framework.

By introducing an auxiliary variable $\tilde{x}=x$ and reformulating (3) as

$$
\begin{array}{ll} \underset{x,\tilde{x}\in\mathbb{X}}{\text{minimize}} & f(x)+\phi(\tilde{x}) \\ \text{subject to} & \mathcal{A}x=b, \quad g(x)\in\mathcal{C}, \quad x\in\mathcal{K}, \quad x=\tilde{x}, \end{array}
$$

we derive the following Lagrangian function associated with the dual variable $(y, \lambda, s, z) \in \mathbb{Y} \times \mathbb{Z} \times \mathbb{X} \times \mathbb{X}$:

$$\mathcal{L}(x, \tilde{x}; y, \lambda, s, z) \triangleq f(x) + \phi(\tilde{x}) - \langle y, \mathcal{A}x - b \rangle - \langle \lambda, g(x) \rangle - \langle s, x \rangle - \langle z, x - \tilde{x} \rangle,$$

which leads to the dual program

$$\begin{array}{ll} \underset{y,\lambda,s,z}{\text{maximize}} & \psi(\mathcal{A}^* y + s + z, \lambda) + \langle b, y \rangle - \phi^*(-z) \\ \text{subject to} & \lambda \in \mathcal{C}^*, \quad s \in \mathcal{K}^*. \end{array} \tag{4}$$

Here $\mathcal{A}^*$ is the adjoint of $\mathcal{A}$, ϕ^* is the conjugate function of ϕ, $\mathcal{C}^*$ and $\mathcal{K}^*$ are the dual cones of $\mathcal{C}$ and $\mathcal{K}$, and the function $\psi : \mathbb{X} \times \mathcal{C}^* \to \mathbb{R}$ is defined as

$$\psi(w, \lambda) \triangleq \inf_{x \in \mathbb{X}} \{ f(x) - \langle w, x \rangle - \langle \lambda, g(x) \rangle \}, \quad (w, \lambda) \in \mathbb{X} \times \mathcal{C}^*.$$

Since g is assumed to be $\mathcal{C}$-convex, the term $-\langle \lambda, g(x) \rangle$ is convex with respect to x for $\lambda \in \mathcal{C}^*$. The optimal solution of the above problem is thus a singleton by the strong convexity of f. In addition, the function ψ is concave and continuously differentiable with Lipschitz continuous gradient [18, Theorem 10.2.1]. It therefore follows that the dual problem (4) is a special case of (2).

In what follows, we introduce an inexact majorized accelerated two-block coordinate descent method (imABCD) to solve (2). By grouping the four blocks of variables (w_1, w_2, w_3, w_4) into two larger blocks, we can express (2) in the following two-block format:

$$\underset{u,v}{\text{minimize}} \quad \theta(w) \triangleq h(w) + p(u) + q(v), \; w \equiv (u, v), \; u \equiv (w_1, w_2) \in \mathbb{U}, \; v \equiv (w_3, w_4) \in \mathbb{V}, \tag{5}$$

where $p(u) = \varphi_1(w_1) + \varphi_2(w_2)$ and $q(v) = \varphi_3(w_3) + \varphi_4(w_4)$ are proper closed convex functions, and $\mathbb{U}$, $\mathbb{V}$ are two appropriately defined finite-dimensional Euclidean spaces. Since ∇h is assumed to be globally Lipschitz continuous, there exist two self-adjoint positive semidefinite linear operators $\mathcal{Q}$ and $\widehat{\mathcal{Q}} : \mathbb{U} \times \mathbb{V} \to \mathbb{U} \times \mathbb{V}$ such that

$$\begin{cases} h(w) \geq h(w') + \langle \nabla h(w'), w - w' \rangle + \frac{1}{2} \|w - w'\|^2_{\mathcal{Q}}, \\ h(w) \leq \widehat{h}(w; w') \triangleq h(w') + \langle \nabla h(w'), w - w' \rangle + \frac{1}{2} \|w - w'\|^2_{\widehat{\mathcal{Q}}}, \end{cases} \quad \forall\, w, w' \in \mathbb{U} \times \mathbb{V}. \tag{6}$$

The operators $\mathcal{Q}$ and $\widehat{\mathcal{Q}}$ may be decomposed into the following 2×2 block structures as

$$\mathcal{Q}w \equiv \begin{pmatrix} \mathcal{Q}_{11} & \mathcal{Q}_{12} \\ \mathcal{Q}_{12}^* & \mathcal{Q}_{22} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}, \quad \widehat{\mathcal{Q}}w \equiv \begin{pmatrix} \widehat{\mathcal{Q}}_{11} & \widehat{\mathcal{Q}}_{12} \\ \widehat{\mathcal{Q}}_{12}^* & \widehat{\mathcal{Q}}_{22} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}, \quad w \in \mathbb{U} \times \mathbb{V},$$

where $\mathcal{Q}_{11}, \widehat{\mathcal{Q}}_{11} : \mathbb{U} \to \mathbb{U}$ and $\mathcal{Q}_{22}, \widehat{\mathcal{Q}}_{22} : \mathbb{V} \to \mathbb{V}$ are self-adjoint positive semidefinite linear operators, and $\mathcal{Q}_{12}, \widehat{\mathcal{Q}}_{12} : \mathbb{V} \to \mathbb{U}$ are two linear mappings whose adjoints are given by $\mathcal{Q}_{12}^*$ and $\widehat{\mathcal{Q}}_{12}^*$, respectively. The following assumption is made in the subsequent discussions.

Assumption 1. There exist two self-adjoint positive semidefinite linear operators $\mathcal{D}_1 : \mathbb{U} \to \mathbb{U}$ and $\mathcal{D}_2 : \mathbb{V} \to \mathbb{V}$ such that

$$\widehat{\mathcal{Q}} \;=\; \mathcal{Q} + \mathrm{Diag}\,(\mathcal{D}_1, \mathcal{D}_2)\,.$$

Furthermore, $\widehat{\mathcal{Q}}$ satisfies that $\widehat{\mathcal{Q}}_{11} = \mathcal{Q}_{11} + \mathcal{D}_1 \succ 0$ and $\widehat{\mathcal{Q}}_{22} = \mathcal{Q}_{22} + \mathcal{D}_2 \succ 0$.

Below is our proposed algorithm for solving problem (2) via the two-block program (5).

---

imABCD: An inexact majorized accelerated block coordinate descent algorithm for (5)

Initialization. Choose an initial point $(u^1, v^1) = (\widetilde{u}^0, \widetilde{v}^0) \in \mathrm{dom}\, p \times \mathrm{dom}\, q$ and a nonnegative nonincreasing sequence $\{\varepsilon_k\}$. Let $t_1 = 1$. Perform the following steps in each iteration for $k \geq 1$.

Step 1. Compute

$$\left\{ \begin{array}{l} \widetilde{u}^k = \underset{u \in \mathbb{U}}{\mathrm{argmin}} \left\{ p(u) + \widehat{h}\left(u, v^k; w^k\right) + \left\langle \delta_u^k, u \right\rangle \right\}, \\ \widetilde{v}^k = \underset{v \in \mathbb{V}}{\mathrm{argmin}} \left\{ q(v) + \widehat{h}\left(\widetilde{u}^k, v; w^k\right) + \left\langle \delta_v^k, v \right\rangle \right\} \end{array} \right. \tag{7}$$

such that $(\delta_u^k, \delta_v^k) \in \mathbb{U} \times \mathbb{V}$ satisfies $\max\{\|\widehat{\mathcal{Q}}_{11}^{-1/2} \delta_u^k\|, \|\widehat{\mathcal{Q}}_{22}^{-1/2} \delta_v^k\|\} \leq \varepsilon_k$. Denote $\widetilde{w}^k = (\widetilde{u}^k, \widetilde{v}^k)$.

Step 2. Compute $\left\{ \begin{array}{l} t_{k+1} = \dfrac{1}{2}\left(1 + \sqrt{1 + 4t_k^2}\right), \\ w^{k+1} = \widetilde{w}^k + \dfrac{t_k - 1}{t_{k+1}}\left(\widetilde{w}^k - \widetilde{w}^{k-1}\right). \end{array} \right.$

The above imABCD algorithm can be taken as an accelerated as well as an inexact version of the alternating minimization method. When $\varepsilon_k \equiv 0$ for all $k \geq 0$, the proposed algorithm reduces to an exact version of the majorized accelerated block coordinate descent (mABCD) method. Since Nesterov's acceleration technique is able to improve the complexity of the gradient-type method from $O(1/k)$ to $O(1/k^2)$, it is interesting for us to investigate whether this acceleration technique can be extended to the two-block coordinate descent method without random selection of the updating blocks. In fact, extensive numerical experiments in the existing literature indicate that the acceleration technique may substantially improve the efficiency of the BCD algorithm; see, e.g., the numerical comparison in [51]. The study of this subject is thus critical for understanding the reasons behind this phenomenon.

Finally, we need to mention that the vectors δ_u^k, δ_v^k in (7) should be interpreted as the residual errors incurred when solving problem (7) without the terms $\langle \delta_u^k, u \rangle$ and $\langle \delta_v^k, v \rangle$. For ease of presentation, we will focus our subsequent discussion only on the u-block, but note that a similar discussion can be adapted for the v-block. Suppose an approximate optimal solution $\tilde{u}^k$ has been computed. Then we can find an appropriate residual vector δ_u^k such that $\delta_u^k \in \partial p(\tilde{u}^k) + \nabla \widehat{h}(\tilde{u}^k, v^k; w^k)$. Given the residual vector δ_u^k, the inexact criteria

$$\|\widehat{\mathcal{Q}}_{11}^{-1/2} \delta_u^k\| \;\leq \varepsilon_k \tag{8}$$

in Step 1 of the above imABCD algorithm can be checked numerically. Next we describe the precise mechanism to check the condition at a given trial approximate solution $\tilde{u}^{k;\text{try}}$ (obtained from an inexact algorithm for the u-block at the kth iteration without the term $\langle \delta_u^k, u\rangle$). First compute

$$\begin{cases} \hat{u}^k \triangleq \text{Prox}_p\left(\tilde{u}^{k;\text{try}} - \nabla\widehat{h}\left(\tilde{u}^{k;\text{try}}, v^k; w^k\right)\right), \\ \delta_u^k \triangleq \tilde{u}^{k;\text{try}} - \hat{u}^k + \nabla\widehat{h}\left(\hat{u}^k, v^k; w^k\right) - \nabla\hat{h}\left(\tilde{u}^{k;\text{try}}, v^k; w^k\right). \end{cases}$$

Then it holds that $\delta_u^k \in \partial p(\hat{u}^k) + \nabla\widehat{h}(\hat{u}^k, v^k; w^k)$. Thus if δ_u^k satisfies the required condition in (8), then we may set $\tilde{u}^k = \hat{u}^k$.

3. The $O(1/k^2)$ complexity of the objective values. In order to simplify the subsequent discussions, we introduce the positive semidefinite operator $\mathcal{H} : \mathbb{U} \times \mathbb{V} \to \mathbb{U} \times \mathbb{V}$ defined by

$$\mathcal{H} \triangleq \text{Diag}\,(\,\mathcal{D}_1\,,\ \mathcal{D}_2 + \mathcal{Q}_{22}\,). \tag{9}$$

We also write Ω as the optimal solution set of (2). We start by presenting the following simple lemma concerning the properties of the sequence $\{t_k\}$.

LEMMA 1. *The sequence $\{t_k\}_{k\geq 1}$ generated by the imABCD algorithm satisfies the following properties:*

(a) $1 - \dfrac{1}{t_{k+1}} = \dfrac{t_k^2}{t_{k+1}^2}$. (b) $\dfrac{k+1}{2} \leq t_k \leq \dfrac{5}{8}k + \dfrac{3}{8} \leq k$.

(c) *Assume any w^+, w' in $\mathbb{U}\times\mathbb{V}$. Consider $w = (1-\frac{1}{t_k})w^+ + \frac{1}{t_k}w'$. Then*

$$t_k^2\,[\,\theta(w) - \theta(w')\,] \leq t_{k-1}^2\,\left[\,\theta(w^+) - \theta(w')\,\right]. \tag{10}$$

Proof. By noting that $t_{k+1}^2 - t_{k+1} = t_k^2$, property (a) can be obtained directly. Property (b) can be derived by induction from the inequalities

$$\begin{aligned} t_{k+1} &= \frac{1 + 2t_k\sqrt{1+1/(4t_k^2)}}{2} \leq \frac{1+2t_k(1+1/(8t_k^2))}{2} \\ &= \frac{1}{2} + t_k + \frac{1}{8t_k} \leq \frac{5}{8} + t_k \leq \frac{5k}{8} + t_1 = \frac{5k}{8} + 1 \end{aligned}$$

and

$$t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2} \geq \frac{1+2t_k}{2} \geq \frac{k+2t_1}{2} = \frac{k+2}{2}.$$

(c) From the convexity of θ, we have that

$$t_k^2\,\theta(w) \leq \left(t_k^2 - t_k\right)\theta(w^+) + t_k\,\theta(w') = t_{k-1}^2\,\theta(w^+) + \left(t_k^2 - t_{k-1}^2\right)\theta(w').$$

From here, we get the desired inequality. ∎

In the following, we shall first provide the $O(1/k^2)$ complexity of the ABCD method with subproblems being solved exactly, and then extend the analysis to the inexact case.

3.1. The case where the subproblems are being solved exactly. The analysis in this subsection is partially motivated by the recent paper [8] in which the authors consider the $O(1/k^2)$ complexity of an accelerated BCD method for (2) where h is a special least-squares quadratic function. Here we extend this nice result to a more general setting where h is only required to be a smooth function.

The lemma below shows an important property of the objective values for the sequence generated by the mABCD algorithm, which is essential to prove the main global complexity result.

LEMMA 2. *Suppose that Assumption* 1 *holds. Let the sequences* $\{\widetilde{w}^k\} \triangleq \{(\widetilde{u}^k, \widetilde{v}^k)\}$ *and* $\{w^k\} = \{(u^k, v^k)\}$ *be generated by the mABCD algorithm. Then for any* $k \geq 1$, *it holds that*

$$\theta(\widetilde{w}^k) - \theta(w) \leq \frac{1}{2}\| w - w^k \|_{\mathcal{H}}^2 - \frac{1}{2}\| w - \widetilde{w}^k \|_{\mathcal{H}}^2 \quad \forall\, w \in \mathbb{U} \times \mathbb{V},$$

where the operator $\mathcal{H}$ *is defined in* (9). *In particular, if* w^* *is an optimal solution of* (2), *then*

$$\| w^* - \widetilde{w}^k \|_{\mathcal{H}} \leq \| w^* - w^k \|_{\mathcal{H}}.$$

Proof. By applying the optimality condition to the subproblems in (7), we derive that

$$\begin{cases} 0 \in \partial p\left(\widetilde{u}^k\right) + \nabla_u h\left(w^k\right) + \widehat{\mathcal{Q}}_{11}\left(\widetilde{u}^k - u^k\right), \\ 0 \in \partial q\left(\widetilde{v}^k\right) + \nabla_v h\left(w^k\right) + \mathcal{Q}_{12}^*\left(\widetilde{u}^k - u^k\right) + \widehat{\mathcal{Q}}_{22}\left(\widetilde{v}^k - v^k\right). \end{cases}$$

Thus, it follows from the convexity of $p(\cdot)$ and $q(\cdot)$ that

$$\begin{cases} p(u) \geq p\left(\widetilde{u}^k\right) + \left\langle u - \widetilde{u}^k,\ -\nabla_u h(w^k) - \widehat{\mathcal{Q}}_{11}(\widetilde{u}^k - u^k)\right\rangle \quad \forall\, u \in \mathbb{U}, \\ q(v) \geq q(\widetilde{v}^k) + \left\langle v - \widetilde{v}^k,\ -\nabla_v h(w^k) - \mathcal{Q}_{12}^*(\widetilde{u}^k - u^k) - \widehat{\mathcal{Q}}_{22}(\widetilde{v}^k - v^k)\right\rangle \quad \forall\, v \in \mathbb{V}. \end{cases} \tag{11}$$

Based on the inequalities in (6) that

$$\begin{cases} h(\widetilde{w}^k) \leq h(w^k) + \langle \nabla h(w^k), \widetilde{w}^k - w^k\rangle + \frac{1}{2}\| \widetilde{w}^k - w^k \|_{\widehat{\mathcal{Q}}}^2, \\ h(w) \geq h(w^k) + \langle \nabla h(w^k), w - w^k\rangle + \frac{1}{2}\| w - w^k \|_{\mathcal{Q}}^2, \end{cases}$$

we get

$$h(w) - h\left(\widetilde{w}^k\right) \geq \left\langle \nabla h\left(w^k\right), w - \widetilde{w}^k\right\rangle + \frac{1}{2}\| w - w^k \|_{\mathcal{Q}}^2 - \frac{1}{2}\| \widetilde{w}^k - w^k \|_{\widehat{\mathcal{Q}}}^2. \tag{12}$$

By the Cauchy–Schwarz inequality, we also have

$$\begin{aligned} &2\left\langle \widetilde{u}^k - u,\ \mathcal{Q}_{12}(\widetilde{v}^k - v^k)\right\rangle = 2\left\langle \mathcal{Q}\left(\widetilde{w}^k - w\right), \begin{pmatrix} 0 \\ \widetilde{v}^k - v^k \end{pmatrix}\right\rangle - 2\left\langle \mathcal{Q}_{22}\left(\widetilde{v}^k - v\right), \widetilde{v}^k - v^k\right\rangle \\ &\quad \leq \left(\| \widetilde{w}^k - w \|_{\mathcal{Q}}^2 + \| \widetilde{v}^k - v^k \|_{\mathcal{Q}_{22}}^2\right) - \left(\| \widetilde{v}^k - v \|_{\mathcal{Q}_{22}}^2 + \| \widetilde{v}^k - v^k \|_{\mathcal{Q}_{22}}^2 - \| v^k - v \|_{\mathcal{Q}_{22}}^2\right) \\ &\quad = \| \widetilde{w}^k - w \|_{\mathcal{Q}}^2 + \| v^k - v \|_{\mathcal{Q}_{22}}^2 - \| \widetilde{v}^k - v \|_{\mathcal{Q}_{22}}^2. \end{aligned} \tag{13}$$

Summing up the inequalities (11) and (12) and substituting the resulting inequality into (13), we obtain

$$\begin{aligned}
&2\left(\theta(w)-\theta\left(\widetilde{w}^k\right)\right)\\
&\quad\geq \left(\|w-w^k\|_{\mathcal{Q}}^2-\|\widetilde{w}^k-w^k\|_{\widehat{\mathcal{Q}}}^2\right)-2\left\langle w-\widetilde{w}^k,\widehat{\mathcal{Q}}\left(\widetilde{w}^k-w^k\right)\right\rangle-2\left\langle \widetilde{u}^k-u,\mathcal{Q}_{12}\left(\widetilde{v}^k-v^k\right)\right\rangle\\
&\quad\geq \|w-w^k\|_{\mathcal{Q}}^2-\|\widetilde{w}^k-w^k\|_{\widehat{\mathcal{Q}}}^2-\left(\|w-w^k\|_{\widehat{\mathcal{Q}}}^2-\|w-\widetilde{w}^k\|_{\widehat{\mathcal{Q}}}^2-\|\widetilde{w}^k-w^k\|_{\widehat{\mathcal{Q}}}^2\right)\\
&\qquad-\left(\|\widetilde{w}^k-w\|_{\mathcal{Q}}^2+\|v^k-v\|_{\mathcal{Q}_{22}}^2-\|\widetilde{v}^k-v\|_{\mathcal{Q}_{22}}^2\right)=\|w-\widetilde{w}^k\|_{\mathcal{H}}^2-\|w-w^k\|_{\mathcal{H}}^2,
\end{aligned}$$

where the last equality is due to Assumption 1. The stated inequality therefore follows. ∎

Based on the above lemma, we next show the $O(1/k^2)$ complexity for the sequence of objective values obtained by the mABCD algorithm.

THEOREM 2. *Suppose that Assumption* 1 *holds and the solution set* Ω *of the problem* (2) *is nonempty. Let* $w^* \triangleq (u^*, v^*) \in \Omega$. *Then the sequence* $\{\widetilde{w}^k\} \triangleq \{(\widetilde{u}^k, \widetilde{v}^k)\}$ *generated by the mABCD algorithm satisfies that*

$$\theta(\widetilde{w}^k)-\theta(w^*)\leq \frac{2\,\|\widetilde{w}^0-w^*\|_{\mathcal{H}}^2}{(k+1)^2}\quad \forall\, k\geq 1.$$

Proof. Letting $w=(1-\frac{1}{t_k})\widetilde{w}^{k-1}+\frac{1}{t_k}w^*$ in Lemma 2, we derive that, for any $k\geq 2$,

$$t_k^2\theta(w)-t_k^2\theta(\widetilde{w}^k)\geq\frac{1}{2}\left\|(t_k-1)\widetilde{w}^{k-1}+w^*-t_k\widetilde{w}^k\right\|_{\mathcal{H}}^2-\frac{1}{2}\left\|(t_k-1)\widetilde{w}^{k-1}+w^*-t_kw^k\right\|_{\mathcal{H}}^2.$$

By applying Lemma 1(c) with $w^+=\widetilde{w}^{k-1}$ and $w'=w^*$, we get

$$t_k^2\left[\theta(w)-\theta(w^*)\right]\leq t_{k-1}^2\left[\theta\left(\widetilde{w}^{k-1}\right)-\theta(w^*)\right].$$

By combining the above inequalities and noting that $w^k=\widetilde{w}^{k-1}+\frac{t_{k-1}-1}{t_k}(\widetilde{w}^{k-1}-\widetilde{w}^{k-2})$, we get for $k\geq 2$,

$$\begin{aligned}
&t_k^2\left[\theta(\widetilde{w}^k)-\theta(w^*)\right]+\frac{1}{2}\left\|t_k\widetilde{w}^k-w^*-(t_k-1)\widetilde{w}^{k-1}\right\|_{\mathcal{H}}^2\\
&\quad\leq t_{k-1}^2\left[\theta(\widetilde{w}^{k-1})-\theta(w^*)\right]+\frac{1}{2}\left\|t_{k-1}\widetilde{w}^{k-1}-w^*-(t_{k-1}-1)\widetilde{w}^{k-2}\right\|_{\mathcal{H}}^2.
\end{aligned}$$

By applying Lemma 2 again for $k=1$ and $w=w^*$, we get

$$\theta(\widetilde{w}^1)-\theta(w^*)\leq\frac{1}{2}\|w^1-w^*\|_{\mathcal{H}}^2-\frac{1}{2}\|\widetilde{w}^1-w^*\|_{\mathcal{H}}^2=\frac{1}{2}\|\widetilde{w}^0-w^*\|_{\mathcal{H}}^2-\frac{1}{2}\|\widetilde{w}^1-w^*\|_{\mathcal{H}}^2.$$

It therefore follows that, for all $k\geq 1$,

$$\begin{aligned}
&t_k^2\left[\theta\left(\widetilde{w}^k\right)-\theta(w^*)\right]+\frac{1}{2}\|t_k\widetilde{w}^k-w^*-(t_k-1)\widetilde{w}^{k-1}\|_{\mathcal{H}}^2\\
&\leq t_{k-1}^2\left[\theta\left(\widetilde{w}^{k-1}\right)-\theta(w^*)\right]+\frac{1}{2}\|t_{k-1}\widetilde{w}^{k-1}-w^*-(t_{k-1}-1)\widetilde{w}^{k-2}\|_{\mathcal{H}}^2\\
&\leq\cdots\\
&\leq t_1^2\left[\theta(\widetilde{w}^1)-\theta\left(w^*\right)\right]+\frac{1}{2}\|t_1\widetilde{w}^1-w^*-(t_1-1)\widetilde{w}^0\|_{\mathcal{H}}^2\leq\frac{1}{2}\|\widetilde{w}^0-w^*\|_{\mathcal{H}}^2.
\end{aligned}$$

The desired inequality of this theorem can thus be established since $t_k\geq\frac{k+1}{2}$ by Lemma 1(b). ∎

3.2. The case where the subproblems are being solved inexactly. Theorem 2 shows the $O(1/k^2)$ complexity of the objective values for the two-block majorized accelerated BCD algorithm for solving (2). However, there seems to have no easy extension of its proof to problems with three or more blocks. In this section, we consider allowing for inexactness when solving the subproblems. The introduction of the inexactness is crucial for efficiently solving the multiblock problems. We note that such an idea has already been incorporated into variants of the BCD and APG algorithms; see, e.g., [49, 29, 57, 53], but the analyses therein are not applicable to our setting.

We characterize the decreasing property of the objective values for the imABCD algorithm in the lemma below. Its proof can be derived similarly as the proof of the exact case in Proposition 2 by applying the optimality conditions at the iteration point $(\widetilde{u}^k, \widetilde{v}^k)$. We omit the details here for brevity.

LEMMA 3. *Suppose that Assumption* 1 *holds. Let the sequences* $\{\widetilde{w}^k\} \triangleq \{(\widetilde{u}^k, \widetilde{v}^k)\}$ *and* $\{w^k\} \triangleq \{(u^k, v^k)\}$ *be generated by the imABCD algorithm. Then for any* $k \geq 1$,

$$\theta\left(\widetilde{w}^k\right) - \theta(w) \leq \frac{1}{2}\| w - w^k \|_{\mathcal{H}}^2 - \frac{1}{2}\| w - \widetilde{w}^k \|_{\mathcal{H}}^2 + \varepsilon_k \| w - \widetilde{w}^k \|_{\mathrm{Diag}(\widehat{\mathcal{Q}}_{11}, \widehat{\mathcal{Q}}_{22})} \ \forall\, w \in \mathbb{U}\times\mathbb{V}.$$

For $k \geq 1$, we denote the exact solutions at the $(k+1)$th iteration as

$$\text{(14)}\quad \overline{u}^k \triangleq \underset{u\in\mathbb{U}}{\operatorname{argmin}} \left\{ p(u) + \widehat{h}\left(u, v^k; w^k\right) \right\}, \quad \overline{v}^k \triangleq \underset{v\in\mathbb{V}}{\operatorname{argmin}} \left\{ q(v) + \widehat{h}\left(\overline{u}^k, v; w^k\right) \right\}.$$

For consistency, we set $(\overline{u}^0, \overline{v}^0) = (\widetilde{u}^0, \widetilde{v}^0) = (u^1, v^1)$. Since $\widehat{\mathcal{Q}}_{11}$ and $\widehat{\mathcal{Q}}_{22}$ are assumed to be positive definite, the above two problems admit unique solutions and thus $\overline{u}^k$ and $\overline{v}^k$ are well defined for $k \geq 0$. The following lemma shows the gap between $(\overline{u}^k, \overline{v}^k)$ and $(\widetilde{u}^k, \widetilde{v}^k)$.

LEMMA 4. *Let the sequences* $\{(\widetilde{u}^k, \widetilde{v}^k)\}$ *and* $\{(u^k, v^k)\}$ *be generated by the imABCD algorithm, and let* $\{(\overline{u}^k, \overline{v}^k)\}$ *be given by* (14). *For any* $k \geq 1$, *the following inequalities hold:*

(a) $\| \overline{u}^k - u^* \|_{\widehat{\mathcal{Q}}_{11}}^2 \leq \| u^k - u^* \|_{\mathcal{D}_1}^2 + \| v^k - v^* \|_{\widehat{\mathcal{Q}}_{22}}^2 = \|w^k - w^*\|_{\mathcal{H}}^2$;
(b) $\| \widehat{\mathcal{Q}}_{11}^{1/2}(\widetilde{u}^k - \overline{u}^k) \| \leq \varepsilon_k, \quad \| \widehat{\mathcal{Q}}_{22}^{1/2}(\widetilde{v}^k - \overline{v}^k) \| \leq (1+\sqrt{2})\,\varepsilon_k$;
(c) $\| \widetilde{w}^k - \overline{w}^k \|_{\mathcal{H}} \leq \sqrt{7}\,\varepsilon_k$.

Proof. (a) By applying the optimality conditions to the problems in (14), we deduce that

$$0 \in \partial p\left(\overline{u}^k\right) + \nabla_u h(w^k) + \widehat{\mathcal{Q}}_{11}\left(\overline{u}^k - u^k\right) \quad \text{and} \quad 0 \in \partial p(u^*) + \nabla_u h(w^*).$$

By the monotonicity of ∂p, we get

$$\text{(15)}\qquad \left\langle \overline{u}^k - u^*,\ \nabla_u h(w^k) - \nabla_u h(w^*) + \widehat{\mathcal{Q}}_{11}\left(\overline{u}^k - u^k\right) \right\rangle \leq 0.$$

Since ∇h is assumed to be globally Lipschitz continuous, it is known from Clarke's mean-value theorem [9, Proposition 2.6.5] that there exists a self-adjoint and positive semidefinite operator $W^k \in \operatorname{conv}\left\{ \partial^2 h(\,[w^{k-1}, w^k]\,) \right\}$ such that

$$\nabla h\left(w^k\right) - \nabla h\left(w^{k-1}\right) = W^k\left(w^k - w^{k-1}\right),$$

where the set $\mathrm{conv}\{\,\partial^2 h(\,[w^{k-1}, w^k]\,)\,\}$ denotes the convex hull of all points in $\partial^2 h(z)$ for any $z \in [w^{k-1}, w^k]$. To proceed, we write $W^k \triangleq \begin{pmatrix} W_{11}^k & W_{12}^k \\ (W_{12}^k)^* & W_{22}^k \end{pmatrix}$, where $W_{11}^k : \mathbb{U} \to \mathbb{U}$, $W_{22}^k : \mathbb{V} \to \mathbb{V}$ are self-adjoint positive semidefinite linear operators and $W_{12}^k : \mathbb{U} \to \mathbb{V}$ is a linear operator. Since

$$\left\langle \begin{pmatrix} \bar{u}^k - u^* \\ v^k - v^* \end{pmatrix}, W^k \begin{pmatrix} \bar{u}^k - u^* \\ v^k - v^* \end{pmatrix} \right\rangle \geq 0 \quad \text{and} \quad \mathcal{Q} \preceq W^k \preceq \widehat{\mathcal{Q}}\,,$$

we can derive that

(16)

$$\begin{aligned}
&2\left\langle\, \overline{u}^k - u^*,\, \nabla_u h(w^k) - \nabla_u h(w^*) \right\rangle \,=\, 2\left\langle\, \overline{u}^k - u^*,\, W_{11}^k\left(u^k - u^*\right) + W_{12}^k\left(v^k - v^*\right)\,\right\rangle \\
&\quad\geq\, \left(\|\,\overline{u}^k - u^*\,\|_{W_{11}^k}^2 + \|\,u^k - u^*\,\|_{W_{11}^k}^2 - \|\,\overline{u}^k - u^k\,\|_{W_{11}^k}^2\right) \\
&\qquad - \left(\|\,\overline{u}^k - u^*\,\|_{W_{11}^k}^2 + \|\,v^k - v^*\,\|_{W_{22}^k}^2\right) \\
&\quad\geq\, \|\,u^k - u^*\,\|_{\mathcal{Q}_{11}}^2 - \|\,\overline{u}^k - u^k\,\|_{\widehat{\mathcal{Q}}_{11}}^2 - \|\,v^k - v^*\,\|_{\widehat{\mathcal{Q}}_{22}}^2.
\end{aligned}$$

From (15) and (16), we may obtain that

$$\begin{aligned}
0 &\geq \|\,u^k - u^*\,\|_{\mathcal{Q}_{11}}^2 - \|\,\overline{u}^k - u^k\,\|_{\widehat{\mathcal{Q}}_{11}}^2 - \|\,v^k - v^*\,\|_{\widehat{\mathcal{Q}}_{22}}^2 + 2\left\langle \overline{u}^k - u^*, \widehat{\mathcal{Q}}_{11}\left(\overline{u}^k - u^k\right)\right\rangle \\
&= \|\,u^k - u^*\,\|_{\mathcal{Q}_{11}}^2 - \|\,\overline{u}^k - u^k\,\|_{\widehat{\mathcal{Q}}_{11}}^2 - \|\,v^k - v^*\,\|_{\widehat{\mathcal{Q}}_{22}}^2 + \|\,\overline{u}^k - u^*\,\|_{\widehat{\mathcal{Q}}_{11}}^2 \\
&\qquad + \|\,\overline{u}^k - u^k\,\|_{\widehat{\mathcal{Q}}_{11}}^2 - \|\,u^k - u^*\,\|_{\widehat{\mathcal{Q}}_{11}}^2 \\
&= \|\,\overline{u}^k - u^*\,\|_{\widehat{\mathcal{Q}}_{11}}^2 - \|\,u^k - u^*\,\|_{\mathcal{D}_1}^2 - \|\,v^k - v^*\,\|_{\widehat{\mathcal{Q}}_{22}}^2,
\end{aligned}$$

which yields $\|\,\overline{u}^k - u^*\,\|_{\widehat{\mathcal{Q}}_{11}}^2 \,\leq\, \|\,u^k - u^*\,\|_{\mathcal{D}_1}^2 + \|\,v^k - v^*\|_{\widehat{\mathcal{Q}}_{22}}^2$. This completes the proof of part (a).

(b) In order to obtain bounds for $\|\,\widehat{\mathcal{Q}}_{11}^{1/2}\,(\widetilde{u}^k - \overline{u}^k)\|$ and $\|\,\widehat{\mathcal{Q}}_{22}^{1/2}\,(\widetilde{v}^k - \overline{v}^k)\|$, we apply the optimality conditions to the problems in (7) at $(\widetilde{u}^k, \widetilde{v}^k)$ and to the problems in (14) at $(\overline{u}^k, \overline{v}^k)$ to deduce that

$$\begin{aligned}
&\left\langle \widehat{\mathcal{Q}}_{11}\left(\widetilde{u}^k - \overline{u}^k\right) + \delta_u^k\,,\, \widetilde{u}^k - \overline{u}^k \right\rangle \,\leq\, 0 \quad \text{and} \\
&\left\langle \mathcal{Q}_{12}^*\,\left(\widetilde{u}^k - \overline{u}^k\right) + \widehat{\mathcal{Q}}_{22}\,\left(\widetilde{v}^k - \overline{v}^k\right) + \delta_v^k\,,\, \widetilde{v}^k - \overline{v}^k \right\rangle \,\leq\, 0.
\end{aligned}$$

The first inequality implies that

$$\|\,\widehat{\mathcal{Q}}_{11}^{1/2}\,(\widetilde{u}^k - \overline{u}^k)\,\| \,\leq\, \|\,\widehat{\mathcal{Q}}_{11}^{-1/2}\,\delta_u^k\,\| \,\leq\, \varepsilon_k.$$

The second inequality yields that

$$\begin{aligned}
\|\,\widetilde{v}^k - \overline{v}^k\,\|_{\widehat{\mathcal{Q}}_{22}}^2 \,&\leq\, \|\,\widehat{\mathcal{Q}}_{22}^{-1/2}\,\delta_v^k\,\|\,\|\,\widehat{\mathcal{Q}}_{22}^{1/2}\,(\widetilde{v}^k - \overline{v}^k)\,\| - \left\langle \mathcal{Q}_{12}^*\,(\widetilde{u}^k - \overline{u}^k), \widetilde{v}^k - \overline{v}^k \right\rangle \\
&\leq\, \|\,\widehat{\mathcal{Q}}_{22}^{-1/2}\,\delta_v^k\|\,\|\,\widehat{\mathcal{Q}}_{22}^{1/2}\,(\widetilde{v}^k - \overline{v}^k)\| + \frac{1}{2}\,\left(\|\,\widetilde{u}^k - \overline{u}^k\,\|_{\widehat{\mathcal{Q}}_{11}}^2 + \|\,\widetilde{v}^k - \overline{v}^k\,\|_{\widehat{\mathcal{Q}}_{22}}^2\right).
\end{aligned}$$

Hence,

$$\begin{aligned}
\|\,\widetilde{v}^k - \overline{v}^k\,\|_{\widehat{\mathcal{Q}}_{22}}^2 &\leq\, 2\,\|\,\widehat{\mathcal{Q}}_{22}^{-1/2}\delta_v^k\,\|\,\|\,\widehat{\mathcal{Q}}_{22}^{1/2}(\widetilde{v}^k - \overline{v}^k)\,\| + \|\,\widetilde{u}^k - \overline{u}^k\,\|_{\widehat{\mathcal{Q}}_{11}}^2 \\
&\leq\, 2\varepsilon_k\,\|\,\widehat{\mathcal{Q}}_{22}^{1/2}(\widetilde{v}^k - \overline{v}^k)\,\| + \varepsilon_k^2.
\end{aligned}$$

By solving this inequality, we obtain that

$$\| \widehat{\mathcal{Q}}_{22}^{1/2} (\widetilde{v}^k - \overline{v}^k) \| \le (1+\sqrt{2})\, \varepsilon_k.$$

(c) From parts (a) and (b), we have that

$$\| \overline{w}^k - \widetilde{w}^k \|_{\mathcal{H}}^2 = \| \overline{u}^k - \widetilde{u}^k \|_{\mathcal{D}_1}^2 + \| \overline{v}^k - \widetilde{v}^k \|_{\widehat{\mathcal{Q}}_{22}}^2 \le 7\varepsilon_k^2.$$

This completes the proof of this lemma. ∎

We are now ready to present the main theorem of this section on the $O(1/k^2)$ complexity of the imABCD algorithm.

THEOREM 3. *Suppose that Assumption* 1 *holds and the solution set* Ω *of the problem* (2) *is nonempty. Let* $w^* \in \Omega$. *Assume that a. Then the sequence* $\{\widetilde{w}^k\} \triangleq \{(\widetilde{u}^k, \widetilde{v}^k)\}$ *generated by the imABCD algorithm satisfies that*

$$\theta(\widetilde{w}^k) - \theta(w^*) \le \frac{2\,\| \widetilde{w}^0 - w^* \|_{\mathcal{H}}^2 + c_0}{(k+1)^2} \quad \forall\, k \ge 1,$$

where c_0 *is a positive scalar (independent of* k*).*

Proof. By applying Lemma 2 for $w = w^*$ and $\overline{w}^1 = \widetilde{w}^1$, we get

$$\begin{aligned} 2\,[\,\theta(\overline{w}^1) - \theta(w^*)\,] &\le \| w^1 - w^* \|_{\mathcal{H}}^2 - \| \overline{w}^1 - w^* \|_{\mathcal{H}}^2 \\ &= \| \widetilde{w}^0 - w^* \|_{\mathcal{H}}^2 - \| \overline{w}^1 - w^* \|_{\mathcal{H}}^2 = \| \overline{w}^0 - w^* \|_{\mathcal{H}}^2 - \| \overline{w}^1 - w^* \|_{\mathcal{H}}^2 . \end{aligned}$$

For any $k \ge 2$, since $\overline{w}^k = (\overline{u}^k, \overline{v}^k)$ exactly solves the subproblem (7), we may take $\widetilde{w}^k = \overline{w}^k$ and $w = \frac{(t_k-1)\overline{w}^{k-1} + w^*}{t_k}$ in Lemma 2 to obtain the following inequality:

$$\theta(w) - \theta(\overline{w}^k) \ge \frac{1}{2} \left\| \frac{(t_k-1)\overline{w}^{k-1} + w^*}{t_k} - \overline{w}^k \right\|_{\mathcal{H}}^2 - \frac{1}{2} \left\| \frac{(t_k-1)\overline{w}^{k-1} + w^*}{t_k} - w^k \right\|_{\mathcal{H}}^2 .$$

By applying Lemma 1(c) with $w^+ = \overline{w}^{k-1}$ and $w' = w^*$, we get

$$t_k^2\,[\,\theta(w) - \theta(w^*)\,] \le t_{k-1}^2\,[\,\theta(\overline{w}^{k-1}) - \theta(w^*)\,].$$

By combining the above two inequalities and noting that $t_k w^k = t_k \widetilde{w}^{k-1} + (t_{k-1} - 1)(\widetilde{w}^{k-1} - \widetilde{w}^{k-2})$, we have for $k \ge 2$,

$$\begin{aligned} &2\,t_k^2 \left[\theta(\overline{w}^k) - \theta(w^*)\right] - 2\,t_{k-1}^2 \left[\theta(\overline{w}^{k-1}) - \theta(w^*)\right] \\ &\quad \le \| t_{k-1}\widetilde{w}^{k-1} - w^* - (t_{k-1}-1)\widetilde{w}^{k-2} - (t_k - 1)\left(\overline{w}^{k-1} - \widetilde{w}^{k-1}\right) \|_{\mathcal{H}}^2 \\ &\qquad - \| t_k \overline{w}^k - w^* - (t_k-1)\overline{w}^{k-1} \|_{\mathcal{H}}^2 \\ &\quad = \| \lambda^{k-1} \|_{\mathcal{H}}^2 - 2 \left\langle \mathcal{H}\lambda^{k-1},\, (t_{k-1}+t_k-1)\left(\overline{w}^{k-1} - \widetilde{w}^{k-1}\right) - (t_{k-1}-1)\left(\overline{w}^{k-2} - \widetilde{w}^{k-2}\right) \right\rangle \\ &\qquad + \| (t_{k-1}+t_k-1)\left(\overline{w}^{k-1} - \widetilde{w}^{k-1}\right) - (t_{k-1}-1)\left(\overline{w}^{k-2} - \widetilde{w}^{k-2}\right) \|_{\mathcal{H}}^2 - \| \lambda^k \|_{\mathcal{H}}^2 , \end{aligned} \tag{17}$$

where $\lambda^k \triangleq t_k \overline{w}^k - w^* - (t_k-1)\overline{w}^{k-1} = t_k(\overline{w}^k - w^*) - (t_k - 1)(\overline{w}^{k-1} - w^*)$. By Lemma 1(b), Lemma 4, and the nonincreasing property of $\{\varepsilon_k\}$, we derive that for all $k \ge 3$,

$$\begin{aligned} &\| (t_{k-1}+t_k-1)\left(\overline{w}^{k-1} - \widetilde{w}^{k-1}\right) - (t_{k-1}-1)\left(\overline{w}^{k-2} - \widetilde{w}^{k-2}\right) \|_{\mathcal{H}} \\ &\quad \le (t_{k-1}+t_k-1)\,\| \overline{w}^{k-1} - \widetilde{w}^{k-1} \|_{\mathcal{H}} + (t_{k-1}-1)\,\| \overline{w}^{k-2} - \widetilde{w}^{k-2} \|_{\mathcal{H}} \\ &\quad \le c_1(k-1)\,\varepsilon_{k-2}, \end{aligned}$$

where $c_1 = 5$. Note that in deriving the above inequality, we have used Lemma 4(c). For $k = 2$, we also have that

$$\begin{aligned}&\|\,(t_{k-1}+t_k-1)\left(\overline{w}^{k-1}-\widetilde{w}^{k-1}\right)-(t_{k-1}-1)(\overline{w}^{k-2}-\widetilde{w}^{k-2})\,\|_{\mathcal{H}} \;=\; t_2\|\,\left(\overline{w}^{1}-\widetilde{w}^{1}\right)\|_{\mathcal{H}}\\ &\quad\le c_1(k-1)\varepsilon_{k-2},\end{aligned}$$

where we set $\varepsilon_0 = \varepsilon_1$. It follows from (17) that for $k \ge 2$, we have

$$\begin{aligned}&2\,t_k^2\,\left[\,\theta(\overline{w}^k)-\theta(w^*)\,\right]+\|\,\lambda^k\,\|_{\mathcal{H}}^2\\ &\quad\le\; 2\,t_{k-1}^2\,\left[\,\theta\left(\overline{w}^{k-1}\right)-\theta(w^*)\,\right]+\|\,\lambda^{k-1}\,\|_{\mathcal{H}}^2+2c_1(k-1)\varepsilon_{k-2}\|\,\lambda^{k-1}\,\|_{\mathcal{H}}+c_1^2\,(k-1)^2\,\varepsilon_{k-2}^2\\ &\quad\le\;\cdots\\ &\quad\le\; 2\,t_1^2\,\left[\,\theta(\overline{w}^1)-\theta(w^*)\,\right]+\|\,\lambda^1\,\|_{\mathcal{H}}^2+2c_1\sum_{i=1}^{k-1} i\,\varepsilon_{i-1}\,\|\,\lambda^i\,\|_{\mathcal{H}}+c_1^2\sum_{i=1}^{k-1} i^2\,\varepsilon_{i-1}^2\\ &\quad\le\; \|\,\overline{w}^1-w^*\,\|_{\mathcal{H}}^2+2c_1\sum_{i=1}^{k-1} i\,\varepsilon_{i-1}\,\|\,\lambda^i\,\|_{\mathcal{H}}+\,c_1^2\sum_{i=1}^{k-1} i^2\,\varepsilon_{i-1}^2\,.\end{aligned}$$

Notice that by Lemma 2, $\|\,\overline{w}^1 - w^*\,\|_{\mathcal{H}}^2 \le \|\,w^1 - w^*\,\|_{\mathcal{H}}^2 = \|\,\overline{w}^0 - w^*\,\|_{\mathcal{H}}^2$. Next we show that the above inequality implies the boundedness of the sequence $\{\|\lambda^k\|_{\mathcal{H}}\}$. If $\|\lambda^k\|_{\mathcal{H}} \le 1$ for all $k \ge 1$, then we are done. Otherwise, for any given sufficiently large positive integer m, we have that

$$\|\,\lambda^{k_m}\,\|_{\mathcal{H}} \;:=\; \max\{\|\,\lambda^i\,\|_{\mathcal{H}} \mid 1 \le i \le m\} \ge 1.$$

Thus, for any $1 \le k \le m$, we have that

$$\begin{aligned}\|\,\lambda^k\|_{\mathcal{H}} \le \|\,\lambda^{k_m}\|_{\mathcal{H}} \;&\le\; \frac{1}{\|\,\lambda^{k_m}\|_{\mathcal{H}}}\left(\|\,\overline{w}^0-w^*\,\|_{\mathcal{H}}^2+2c_1\sum_{i=1}^{k_m-1} i\,\varepsilon_{i-1}\|\,\lambda^i\,\|_{\mathcal{H}}+c_1^2\sum_{i=1}^{k_m-1} i^2\,\varepsilon_{i-1}^2\right)\\ &\le\; \|\,\overline{w}^0-w^*\,\|_{\mathcal{H}}^2+2c_1\sum_{i=1}^{k_m-1} i\,\varepsilon_{i-1}\,\frac{\|\,\lambda^i\,\|_{\mathcal{H}}}{\|\,\lambda^{k_m}\,\|_{\mathcal{H}}}+\,c_1^2\sum_{i=1}^{k_m-1} i^2\,\varepsilon_{i-1}^2\\ &\le\; \|\,\overline{w}^0-w^*\,\|_{\mathcal{H}}^2+2c_1\sum_{i=1}^{k_m-1} i\,\varepsilon_{i-1}+\,c_1^2\sum_{i=1}^{k_m-1} i^2\,\varepsilon_{i-1}^2\\ &\le\; \|\,\overline{w}^0-w^*\,\|_{\mathcal{H}}^2+2c_1\sum_{i=1}^{\infty} i\,\varepsilon_{i-1}+\,c_1^2\sum_{i=1}^{\infty} i^2\,\varepsilon_{i-1}^2,\end{aligned}$$

where the third inequality follows from the definition of $\|\lambda^{k_m}\|_{\mathcal{H}}$. Thus by letting $m \to \infty$, we get

$$\|\,\lambda^k\|_{\mathcal{H}} \;\le\; c_2 \;\triangleq\; \max\left\{1,\;\|\,\overline{w}^0-w^*\,\|_{\mathcal{H}}^2+2c_1\sum_{i=1}^{\infty} i\,\varepsilon_{i-1}+\,c_1^2\sum_{i=1}^{\infty} i^2\,\varepsilon_{i-1}^2\right\}\quad\forall\,k\ge 1.$$

To estimate the bound for $\|\,\overline{w}^{k+1} - w^*\|_{\mathcal{H}}$, we set $w = w^*$ and $\widetilde{w}^k = \overline{w}^k$ in Lemma 3 and deduce that

$$\begin{aligned} t_{k+1} \| \overline{w}^{k+1} - w^* \|_{\mathcal{H}} &\leq t_{k+1} \| w^{k+1} - w^* \|_{\mathcal{H}} \\ &= \left\| (t_{k+1} + t_k - 1)\widetilde{w}^k - (t_k - 1)\widetilde{w}^{k-1} - t_{k+1} w^* \right\|_{\mathcal{H}} \\ &\leq (t_{k+1} - 1) \left\| \overline{w}^k - w^* \right\|_{\mathcal{H}} + \left\| t_k \overline{w}^k - (t_k - 1)\overline{w}^{k-1} - w^* \right\|_{\mathcal{H}} \\ &\quad + (t_{k+1} + t_k - 1) \left\| \widetilde{w}^k - \overline{w}^k \right\|_{\mathcal{H}} + (t_k - 1) \left\| \widetilde{w}^{k-1} - \overline{w}^{k-1} \right\|_{\mathcal{H}} \\ &\leq \frac{t_k^2}{t_{k+1}} \left\| \overline{w}^k - w^* \right\|_{\mathcal{H}} + c_2 + c_1 \varepsilon_{k-1} t_{k+1}, \end{aligned}$$

where the last inequality is obtained by Lemma 1(a). It follows that

$$\begin{aligned} \frac{t_k^2}{t_{k+1}^2} \| \overline{w}^k - w^* \|_{\mathcal{H}} &\leq \frac{t_k^2}{t_{k+1}^2} \left(\frac{t_{k-1}^2}{t_k^2} \left\| \overline{w}^{k-1} - w^* \right\|_{\mathcal{H}} + \frac{c_2}{t_k} + c_1 \varepsilon_{k-2} \right), \\ \frac{t_k^2}{t_{k+1}^2} \frac{t_{k-1}^2}{t_k^2} \| \overline{w}^{k-1} - w^* \|_{\mathcal{H}} &\leq \frac{t_k^2}{t_{k+1}^2} \frac{t_{k-1}^2}{t_k^2} \left(\frac{t_{k-2}^2}{t_{k-1}^2} \left\| \overline{w}^{k-2} - w^* \right\|_{\mathcal{H}} + \frac{c_2}{t_{k-1}} + c_1 \varepsilon_{k-3} \right), \\ \vdots \qquad &\leq \qquad \vdots \\ \frac{t_k^2}{t_{k+1}^2} \frac{t_{k-1}^2}{t_k^2} \cdots \frac{t_2^2}{t_3^2} \| (\overline{w}^2 - w^*) \|_{\mathcal{H}} &\leq \frac{t_k^2}{t_{k+1}^2} \frac{t_{k-1}^2}{t_k^2} \cdots \frac{t_2^2}{t_3^2} \left(\frac{t_1^2}{t_2^2} \left\| \overline{w}^1 - w^* \right\|_{\mathcal{H}} + \frac{c_2}{t_2} + c_1 \varepsilon_0 \right). \end{aligned}$$

Summing up the above inequalities, we obtain

$$\left\| \overline{w}^{k+1} - w^* \right\|_{\mathcal{H}} \leq \frac{t_1^2}{t_{k+1}^2} \left\| \overline{w}^1 - w^* \right\|_{\mathcal{H}} + c_2 \sum_{i=1}^k \frac{t_{i+1}}{t_{k+1}^2} + c_1 \sum_{i=1}^k \varepsilon_{i-1}. \tag{18}$$

By Lemma 1(b), we have

$$\sum_{i=1}^k \frac{t_{i+1}}{t_{k+1}^2} \leq \frac{(3+k)\,k}{2\left(\frac{1}{2}k+1\right)^2} \leq 2 \quad \forall\, k \geq 1.$$

Therefore, the inequality (18) implies that for all $k \geq 1$,

$$\begin{aligned} \left\| (\overline{w}^{k+1} - w^*) \right\|_{\mathcal{H}} &\leq \frac{4}{(k+2)^2} \left\| \overline{w}^1 - w^* \right\|_{\mathcal{H}} + 2c_2 + c_1 \sum_{i=1}^\infty \varepsilon_{i-1} \\ &\leq c_3 \triangleq \max \left(\left\| \overline{w}^1 - w^* \right\|_{\mathcal{H}}, \frac{4}{9} \left\| \overline{w}^1 - w^* \right\|_{\mathcal{H}} + 2c_2 + c_1 \sum_{i=1}^\infty \varepsilon_{i-1} \right). \end{aligned}$$

Notice that we have $\| \overline{w}^1 - w^* \|_{\mathcal{H}} \leq c_3$.

The next step is to prove the boundedness of the term $\| t_k \widetilde{u}^k - u^* - (t_k - 1)\widetilde{u}^{k-1} \|_{\widehat{\mathcal{Q}}_{11}}$. Before that, we need to first bound $\left\| \overline{u}^k - u^* \right\|_{\widehat{\mathcal{Q}}_{11}}^2$. By using Lemma 4, we have that for $k \geq 2$,

$$\begin{aligned} & t_k \left\| \overline{u}^k - u^* \right\|_{\widehat{\mathcal{Q}}_{11}} \leq t_k \left\| w^k - w^* \right\|_{\mathcal{H}} \\ &= \|(t_{k-1} + t_k - 1)(\widetilde{w}^{k-1} - w^*) - (t_{k-1} - 1)(\widetilde{w}^{k-2} - w^*)\|_{\mathcal{H}} \\ &\leq (t_k - 1)\|(\widetilde{w}^{k-1} - w^*)\|_{\mathcal{H}} + \|t_{k-1}(\widetilde{w}^{k-1} - w^*) - (t_{k-1} - 1)(\widetilde{w}^{k-2} - w^*)\|_{\mathcal{H}} \\ &\leq (t_k - 1) \left[\|\widetilde{w}^{k-1} - \overline{w}^{k-1}\|_{\mathcal{H}} + \|\overline{w}^{k-1} - w^*\|_{\mathcal{H}} \right] + \|t_{k-1}(\widetilde{w}^{k-1} - \overline{w}^{k-1}) \\ &\quad - (t_{k-1} - 1)(\widetilde{w}^{k-2} - \overline{w}^{k-2})\|_{\mathcal{H}} + \|t_{k-1}(\overline{w}^{k-1} - w^*) - (t_{k-1} - 1)(\overline{w}^{k-2} - w^*)\|_{\mathcal{H}} \\ &\leq (t_k - 1)[\sqrt{7}\varepsilon_{k-1} + c_3] + (2t_k - 1)\sqrt{7}\varepsilon_{k-2} + \|\gamma^{k-1}\|_{\mathcal{H}} \\ &\leq t_k[8\varepsilon_1 + c_2 + c_3]. \end{aligned}$$

For $k=1$, we also have that $\|\overline{u}^k - u^*\|_{\widehat{\mathcal{Q}}_{11}} \le \|\overline{w}^k - w^*\|_{\widehat{\mathcal{H}}} \le c_3$. Now we have for all $k \ge 2$,

$$
\begin{aligned}
&\| t_k\widetilde{u}^k - u^* - (t_k-1)\,\widetilde{u}^{k-1} \|_{\widehat{\mathcal{Q}}_{11}} \\
&\quad \le\ t_k \left(\| \overline{u}^k - \widetilde{u}^k \|_{\widehat{\mathcal{Q}}_{11}} + \| \overline{u}^k - u^* \|_{\widehat{\mathcal{Q}}_{11}} \right) \\
&\qquad + (t_k-1) \left(\| \overline{u}^{k-1} - u^* \|_{\widehat{\mathcal{Q}}_{11}} + \| \overline{u}^{k-1} - \widetilde{u}^{k-1} \|_{\widehat{\mathcal{Q}}_{11}} \right) \\
&\quad \le\ (2t_k-1)\,[\,8\,\varepsilon_1 + c_2 + c_3\,].
\end{aligned}
$$

In addition, for $k=1$, we have

$$
\begin{aligned}
&\| t_1\widetilde{u}^1 - u^* - (t_1-1)\widetilde{u}^0 \|_{\widehat{\mathcal{Q}}_{11}} \ =\ \| \widetilde{u}^1 - u^* \|_{\widehat{\mathcal{Q}}_{11}} \ \le\ \| \overline{u}^1 - u^* \|_{\widehat{\mathcal{Q}}_{11}} + \| \widetilde{u}^1 - \overline{u}^1 \|_{\widehat{\mathcal{Q}}_{11}} \\
&\quad \le\ \| w^1 - w^* \|_{\mathcal{H}} + \varepsilon_1 \ \le\ c_2 + \varepsilon_1 \ \le\ (2t_k-1)\,[\,8\,\varepsilon_1 + c_2 + c_3\,].
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
&\| t_k\widetilde{v}^k - v^* - (t_k-1)\,\widetilde{v}^{k-1} \|_{\widehat{\mathcal{Q}}_{22}} \ =\ \| t_k(\widetilde{v}^k - v^*) - (t_k-1)\,(\widetilde{v}^{k-1} - v^*) \|_{\widehat{\mathcal{Q}}_{22}} \\
&\le \| t_k(\overline{v}^k - v^*) - (t_k-1)(\overline{v}^{k-1} - v^*) \|_{\widehat{\mathcal{Q}}_{22}} + \| t_k(\widetilde{v}^k - \overline{v}^k) - (t_k-1)\,(\widetilde{v}^{k-1} - \overline{v}^{k-1}) \|_{\widehat{\mathcal{Q}}_{22}} \\
&\le \|\lambda^k\|_{\mathcal{H}} + t_k\|\,(\widetilde{v}^k - \overline{v}^k)\|_{\widehat{\mathcal{Q}}_{22}} + (t_k-1)\|(\widetilde{v}^{k-1} - \overline{v}^{k-1})\,\|_{\widehat{\mathcal{Q}}_{22}} \\
&\le c_2 + (2t_k-1)(1+\sqrt{2})\varepsilon_{k-1} \ \le\ (2t_k-1)[c_2 + 3\varepsilon_1].
\end{aligned}
$$

Finally, by applying Lemma 3 at $w = \frac{(t_k-1)\widetilde{w}^{k-1} + w^*}{t_k}$ and using Lemma 1(c), we see that

$$
\begin{aligned}
&t_k^2\left[\,\theta\left(\widetilde{w}^k\right) - \theta(w^*)\,\right] + \frac{1}{2}\,\| t_k\widetilde{w}^k - w^* - (t_k-1)\widetilde{w}^{k-1} \|_{\mathcal{H}}^2 \\
&\quad \le t_{k-1}^2\left[\,\theta\left(\widetilde{w}^{k-1}\right) - \theta(w^*)\,\right] + \frac{1}{2}\,\| t_{k-1}\widetilde{w}^{k-1} - w^* - (t_{k-1}-1)\widetilde{w}^{k-2} \|_{\mathcal{H}}^2 \\
&\qquad + \varepsilon_k\,\| t_k\widetilde{w}^k - w^* - (t_k-1)\widetilde{w}^{k-1} \|_{\mathrm{Diag}(\widehat{\mathcal{Q}}_{11},\,\widehat{\mathcal{Q}}_{22})} \\
&\quad \le \cdots \\
&\quad \le\ t_1^2\left[\,\theta(\widetilde{w}^1) - \theta(w^*)\,\right] + \frac{1}{2}\,\| t_1\widetilde{w}^1 - w^* - (t_1-1)\widetilde{w}^0 \|_{\mathcal{H}}^2 \\
&\qquad + \sum_{i=2}^{k} \varepsilon_i\,\| t_i\widetilde{w}^i - w^* - (t_i-1)\widetilde{w}^{i-1} \|_{\mathrm{Diag}(\widehat{\mathcal{Q}}_{11},\,\widehat{\mathcal{Q}}_{22})} \\
&\quad \le \frac{1}{2}\,\| \widetilde{w}^0 - w^* \|_{\mathcal{H}}^2 + \sum_{i=1}^{k} \varepsilon_i\,\| t_i\widetilde{w}^i - w^* - (t_i-1)\widetilde{w}^{i-1} \|_{\mathrm{Diag}(\widehat{\mathcal{Q}}_{11},\,\widehat{\mathcal{Q}}_{22})} \\
&\quad \le \frac{1}{2}\,\| \widetilde{w}^0 - w^* \|_{\mathcal{H}}^2 + \sum_{i=1}^{k} (2t_i-1)\,[\,11\varepsilon_1 + 2c_2 + c_3\,]\,\varepsilon_i \\
&\quad \le \frac{1}{2}\,\| \widetilde{w}^0 - w^* \|_{\mathcal{H}}^2 + \frac{1}{4}c_0,
\end{aligned}
$$

where

$$
c_0 \ \triangleq\ 4\sum_{i=1}^{\infty} (2t_i-1)\,[\,11\varepsilon_1 + 2c_2 + c_3\,]\,\varepsilon_i
$$

is a finite value since $\sum_{i=1}^{\infty} t_i\,\varepsilon_i < \infty$. Since $t_k \geq \frac{k+1}{2}$, we complete the proof of this theorem. ∎

In order to guarantee that $\sum_{i=1}^{\infty} i\varepsilon_i < \infty$ in the above theorem, one may stop the ith subproblem of the imABCD algorithm with $\varepsilon_i = c/i^{2.1}$ for some constant c.

4. Solving the best approximation problem (1). In this section, we discuss an application of the imABCD framework to solve the dual of the best approximation problem (1). The the best approximation problems with only equality constraints have been studied for more than three decades [28, 2, 14, 15, 16]. The best approximation problem with the positive semidefinite cone constraint was studied recently; see, e.g., [37, 43, 7, 22, 51].

The dual of (1) is given by

(19)
$$\underset{y,S,z,Z}{\text{minimize}} \quad \frac{1}{2}\|\mathcal{A}^*y + S + \mathcal{B}^*z + Z + G\,\|^2 - \langle b, y\rangle - \langle d, z\rangle + \delta_{\mathbb{S}^n_+}(S) + \delta_{\geq 0}(z) + \delta_{\geq 0}(Z),$$

where $\delta_C(\cdot)$ denotes the indicator function of a given set C, i.e., $\delta_C(x) = 0$ if $x \in C$, and $\delta_C(x) = \infty$ otherwise. The notation $\delta_{\geq 0}(\cdot)$ is used to denote the indicator function over a nonnegative orthant. To implement the two-block imABCD algorithm, we take (y, S) as one block and (z, Z) as the other block. The operator $\mathcal{Q}$ and its majorization $\widehat{\mathcal{Q}}$ in Assumption 1 are taken as

$$\mathcal{Q} = \left(\begin{array}{cc|cc} \mathcal{A}\mathcal{A}^* & \mathcal{A} & \mathcal{A}\mathcal{B}^* & \mathcal{A} \\ \mathcal{A}^* & \mathcal{I} & \mathcal{B}^* & \mathcal{I} \\ \hline \mathcal{A}^*\mathcal{B} & \mathcal{B} & \mathcal{B}\mathcal{B}^* & \mathcal{B} \\ \mathcal{A}^* & \mathcal{I} & \mathcal{B}^* & \mathcal{I} \end{array}\right) \quad \text{and} \quad \widehat{\mathcal{Q}} = \mathcal{Q} + \left(\begin{array}{cc|cc} c\mathcal{I} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & c\mathcal{I} & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

for some scalar $c > 0$. Notice that the additional proximal term associated with the linear operator $c\mathcal{I}$ is added for the block of y and z to make the conditions $\mathcal{D}_1 \succ 0$ and $\mathcal{D}_2 \succ 0$ in Assumption 1 hold. There are two important reasons for us to merge the four blocks in such a way. First, based on our previous experiences from developing several solvers with similar types of constraints, in particular for linear and least-squares semidefinite programming in [60, 51], we find that, compared to the linear inequalities and the nonnegative cone constraints, the linear equalities and the semidefinite cone constraints are more challenging to satisfy numerically. Putting the corresponding multipliers y (for the linear equalities constraints) and S (for the positive semidefinite cone constraint) in one group and solving them simultaneously by the inexact semismooth Newton method may help these two types of constraints to achieve a high accuracy simultaneously. Second, putting S and z (corresponding to the linear inequalities constraints) or Z (corresponding to the entrywise nonnegative constraints) in one group often leads to a degenerate Newton system when the semismooth Newton method is applied to solve the resulting subproblem.

4.1. Solving the subproblems by the Newton-type methods. With fixed (z, Z) and a properly defined matrix $G_1 \in \mathbb{S}^n$, the first workhorse of the imABCD for solving (1) is the following convex program:

$$\underset{y,S}{\text{minimize}} \quad \frac{1}{2}\,\|\mathcal{A}^*y + S + G_1\|^2 - \langle b, y\rangle + \frac{c}{2}\|y - y_0\|^2 \quad \text{subject to} \quad S \in \mathbb{S}^n_+,\ y \in \mathbb{R}^m.$$

Let $\Pi_{\mathbb{S}^+_n}(\cdot)$ denote the projection onto the cone $\mathbb{S}^n_+$. Since the optimal solution $(\bar{y}, \overline{S})$ always satisfies

$$\overline{S} \,=\, \Pi_{\mathbb{S}^n_+}(-G_1 - \mathcal{A}^*\bar{y}), \tag{20}$$

we may solve y first via the unconstrained minimization problem

$$\begin{aligned} \underset{y}{\text{minimize}}\ \xi(y) &\triangleq \tfrac{1}{2}\,\|\mathcal{A}^*y + G_1 + \Pi_{\mathbb{S}^n_+}(-G_1 - \mathcal{A}^*y)\|^2 - \langle b, y\rangle + \frac{c}{2}\|y - y_0\|^2 \\ &= \tfrac{1}{2}\,\|\Pi_{\mathbb{S}^n_+}(G_1 + \mathcal{A}^*y)\|^2 - \langle b, y\rangle + \frac{c}{2}\|y - y_0\|^2 \end{aligned} \tag{21}$$

and then substitute the solution into (20) to obtain $\overline{S}$. Notice that ξ is a continuously differentiable function and its gradient

$$\nabla\xi(y) \,=\, \mathcal{A}\Pi_{\mathbb{S}^n_+}(G_1 + \mathcal{A}^*y) - b + c\,(y - y_0)$$

is strongly semismooth [50]. Therefore, the semismooth Newton-CG algorithm with line search, which is proven to converge globally and local superlinearly/quadratically [32, 44, 43, 61], can be applied to solve the above unconstrained problem (21). The details of this algorithm are given below.

SNCG: A semismooth Newton-CG method for solving (21)

Initialization. Given $\mu \in (0, 1/2)$, $\eta \in (0, 1)$, $\tau \in (0, 1]$ and $\rho \in (0, 1)$. Iterate the following steps for $j \geq 0$.

Step 1. Choose $V^j \in \partial\Pi_{\mathbb{S}^n_+}(G_1 + \mathcal{A}^*y^j)$. Solve the following linear system to find d^j by the conjugate gradient (CG) method:

$$(\mathcal{A}V^j\mathcal{A}^* + c\mathcal{I})\,d + \nabla\xi(y^j) = 0$$

such that d^j satisfies the accuracy condition that

$$\|\,\big(\mathcal{A}V^j\mathcal{A}^* + c\mathcal{I}\big)\ d + \nabla\xi(y^j)\,\| \,\leq\, \min\big\{\,\eta, \|\,\nabla\xi(y^j)\,\|^{1+\tau}\,\big\}\,.$$

Step 2. (Line search) Set $\alpha_j \,=\, \rho^{m_j}$, where m_j is the first nonnegative integer m for which

$$\xi\left(y^j + \rho^m d_j\right) \,\leq\, \xi(y^j) + \mu\,\rho^m\ \left\langle\nabla\xi\left(y^j\right), d_j\right\rangle.$$

Step 3. Set $y^{j+1} \,=\, y^j + \alpha_j d^j$.

To solve the second subproblem involving (z, Z) in (7), we need to deal with the program

(22)

$$\underset{z,Z}{\text{minimize}}\quad \frac{1}{2}\,\|\,\mathcal{B}^*z + Z + G_2\,\|^2 - \langle d, z\rangle + \frac{c}{2}\,\|\,z - z_0\,\|^2 \quad \text{subject to } z \geq 0, \quad Z \geq 0,$$

for some properly defined matrix $G_2 \in \mathbb{S}^n$. Similar to the first subproblem, the optimal solution $(\bar{z}, \overline{Z})$ satisfies that

$$\overline{Z} \,=\, \Pi_{\geq 0}(-\mathcal{B}^*\bar{z} - G_2).$$

We therefore need to solve the problem

$$\text{(23)} \quad \underset{z}{\text{minimize}} \quad \frac{1}{2}\,\|\,\Pi_{\geq 0}(\mathcal{B}^* z + G_2)\,\|^2 - \langle d, z\rangle + \frac{c}{2}\,\|\, z - z_0\,\|^2 \quad \text{subject to } z \geq 0.$$

Different from (21), the above problem is a constrained SC^1 problem (i.e., the objective function is continuously differentiable with a semismooth gradient). Though we may still apply a globally convergent semismooth Newton algorithm as proposed in [42], a strictly convex quadratic programming problem has to be solved in each step, which itself may be challenging. In fact, the problem (23) is a special case of the general unconstrained nonsmooth convex program

$$\text{(24)} \qquad \underset{x}{\text{minimize}} \;\; \zeta(x) + \psi(x),$$

where $\zeta : \mathbb{X} \to (-\infty, \infty)$ is a strongly convex and smooth function, and $\psi : \mathbb{X} \to (-\infty, +\infty]$ is a convex but possibly nonsmooth function. One can apply Nesterov's APG method [39], which converges globally and linearly, to solve such a strongly convex problem [49]. Alternatively, the solution of (24) can be obtained via the nonsmooth equation

$$F(x) \triangleq x - \mathrm{Prox}_{\psi}\,(\, x - \nabla\zeta(x)\,) = 0,$$

where $\mathrm{Prox}_{\psi}(x) \triangleq \operatorname{argmin}\left\{\psi(x') + \frac{1}{2}\,\|\, x' - x\,\|^2 \mid\; x' \in \mathbb{X}\right\}$ denotes the proximal mapping of ψ at x; see, e.g., [48, Definition 1.22]. Since the composition of semismooth functions is semismooth [21], it follows that F is semismooth at x if $\nabla\zeta$ is semismooth at x and $\mathrm{Prox}_{\psi}(\cdot)$ is semismooth at $x - \nabla\zeta(x)$. We may then apply the semismooth Newton-CG method locally to solve the above nonsmooth equation for a faster convergence rate. Therefore, a convergent and efficient way to solve (23) may be a hybrid of the APG algorithm and the semismooth Newton-CG method. We present this algorithm below, where the positive scalar L_ζ denotes the Lipschitz constant of $\nabla\zeta$.

APG-SNCG: A hybrid of the APG algorithm and the SNCG method for solving (24)

Choose an initial point $x^1 \in \mathbb{X}$, positive constants $\eta, \gamma \in (0,1)$, $\rho \in (0, 1/2)$, and a positive integer $m_0 > 0$. Iterate the following steps for $j \geq 0$.

Step 1: Select $V^j \in \partial F(x^j)$, the generalized Jacobian of F at x^j, and apply the CG method to find an approximate solution d^j to

$$\text{(25)} \qquad V^j d + F(x^j) = 0$$

such that

$$\text{(26)} \qquad R^j \triangleq V^j d^j + F(x^j) \quad \text{and} \quad \|\, R^j\,\| \,\leq\, \eta_j\,\|\, F(x^j)\,\|,$$

where $\eta_j \triangleq \min\{\,\eta, \|\, F(x^j)\,\|\,\}$. If (26) is achievable, go to Step 2. Otherwise, go to Step 3.

Step 2: Let $m_j \leq m_0$ be the smallest nonnegative integer m such that

$$\|\, F\left(x^j + \rho^m d^j\right)\,\| \,\leq\, \gamma\,\|\, F\left(x^j\right)\,\|.$$

If the above inequality is achievable, set $t_j = \rho^{m_j}$ and $x^{j+1} = x^j + t_j d^j$. Replace j by $j+1$ and go to Step 1. Otherwise (i.e., the above inequality fails for all $m \le m_0$) go to Step 3.

Step 3: Set $x^{j_1} = \tilde{x}^{j_0} = x^j$, $\beta_{j_1} = 1$ and $i = 1$, compute

$$\begin{cases} \tilde{x}^{j_i} = \mathrm{Prox}_{\psi/L_\zeta}\left(x^{j_i} - \nabla\zeta\left(x^{j_i}\right)/L_\zeta\right), \\ \beta_{j_{i+1}} = \frac{1}{2}\left(1 + \sqrt{1 + 4\beta_{j_i}^2}\right), \\ x^{j_{i+1}} = \tilde{x}^{j_i} + \frac{\beta_{j_i} - 1}{\beta_{j_{i+1}}}\left(\tilde{x}^{j_i} - \tilde{x}^{j_{i-1}}\right). \end{cases}$$

If $\| F(x^{j_{i+1}}) \| \le \gamma \| F(x^j) \|$, set $x^{j+1} = x^{j_{i+1}}$. Replace j by $j+1$ and go to Step 1. Otherwise, set $i = i+1$ and continue the above iteration.

It is known that for the sequence $\{x^{j_i}\}_{i=1}^{\infty}$ generated by the APG algorithm, $[\zeta(x^{j_i}) + \psi(x^{j_i})] - [\zeta(x^*) + \psi(x^*)] \to 0$ as $i \to \infty$, where x^* is the unique solution of the problem (24); see, e.g., [5, Theorem 4.4]. By the strong convexity of ζ, this further implies that $\|x^{j_i} - x^*\| \to 0$. Therefore, $F(x^{j_i}) \to 0$ as $i \to \infty$ by the continuity of the proximal mapping. That being said, the sequence generated by the APG algorithm can be viewed as a safeguard for the global convergence of $\{x^j\}$ in the above framework; the condition $\| F(x^{j_{i+1}}) \| \le \gamma \| F(x^j) \|$ in Step 3 is always achievable for sufficiently large i. We make two further remarks on this algorithm below.

Remark 4. It is known from Rademacher's theorem that the Lipschitz continuous function F is differentiable almost everywhere. Assume that (26) is achievable at a differentiable point x^j and $\|F(x^j)\| \neq 0$; then $\|F(x)\|^2$ is differentiable at x^j and

$$\begin{aligned} \| F(x^j + td^j) \|^2 &= \| F(x^j) + t\,[\,R^j - F(x^j)\,] + o(t) \|^2 \\ &= \| F(x^j) \|^2 + t\,\langle F(x^j), R^j - F(x^j) \rangle + o(t) \\ &\le \| F(x^j) \|^2 + t\,(\eta_j - 1)\,\| F(x^j) \|^2 + o(t). \end{aligned}$$

Since $\eta_j \le \eta < 1$, we have $\| F(x^j + t\,d^j) \| < \| F(x^j) \|$ for t sufficiently small such that d^j is a descent direction of $\| F(x) \|$ at x^j. This yields that the direction obtained by (25) is a descent direction of $\| F(x) \|$ at x^j with probability 1.

Remark 5. Equation (25) may not be a symmetric linear system. If this occurs, one may use the BiCGStab iterative solver (e.g., van der Vorst [56]) to solve the corresponding equation.

4.2. Decomposing (22) into smaller decoupled problems. For some best approximation problems, the number of inequalities in $\mathcal{B}X \ge d$ may be ultra large, and that can make the subproblem (22) extremely expensive to solve. Fortunately, by the design of imABCD, one can add an appropriate proximal term in (7) to make the subproblem involving (z, Z) easier to solve, in particular, by decomposing (22) into smaller decoupled problems.

A practical way to achieve the decomposition of (22) is to add a proximal term of the form $\frac{1}{2}\left\|\binom{z}{Z} - \binom{z_0}{Z_0}\right\|_{\mathcal{D}_2}^2$, where the positive semidefinite linear operator $\mathcal{D}_2$ is constructed based on dividing the operator $\mathcal{B}$ and the dual variable z into $q \ge 1$ parts as

$$\begin{cases} \mathcal{B}X \equiv \begin{pmatrix} \mathcal{B}_1 X \\ \vdots \\ \mathcal{B}_q X \end{pmatrix} \text{ with } \mathcal{B}_i : \mathbb{S}^n \to \mathbb{R}^{m_i}, \quad X \in \mathbb{S}^n, \\ \mathcal{B}^* z \equiv (\mathcal{B}_1^* z_1, \mathcal{B}_2^* z_2, \ldots, \mathcal{B}_q^* z_q),\ z \equiv (z_1, z_2, \ldots z_q) \in \mathbb{R}^{m_{i_1}} \times \mathbb{R}^{m_{i_2}} \times \cdots \times \mathbb{R}^{m_{i_q}} = \mathbb{R}^{m_I}. \end{cases}$$

By observing that for any given matrix $X \in \mathbb{R}^{m\times n}$,

$$\begin{pmatrix} & X \\ X^T & \end{pmatrix} \preceq \begin{pmatrix} (XX^T)^{\frac{1}{2}} & \\ & (X^T X)^{\frac{1}{2}} \end{pmatrix},$$

we may derive that

$$\mathcal{B}\mathcal{B}^* \preceq \mathcal{M} \triangleq \operatorname{Diag}(\mathcal{M}_1, \ldots, \mathcal{M}_q),$$

where

$$\mathcal{M}_i \triangleq \mathcal{B}_i \mathcal{B}_i^* + \sum_{j=1,\ldots,q,\, j\neq i} (\mathcal{B}_i \mathcal{B}_j^* \mathcal{B}_j \mathcal{B}_i^*)^{1/2}, \quad i = 1, \ldots q.$$

It therefore follows that

$$\begin{pmatrix} 2\mathcal{M} & \\ & 2I \end{pmatrix} \succeq \begin{pmatrix} 2\mathcal{B}\mathcal{B}^* & \\ & 2I \end{pmatrix} \succeq \begin{pmatrix} \mathcal{B} \\ I \end{pmatrix} \begin{pmatrix} \mathcal{B}^* & I \end{pmatrix}.$$

By choosing $\mathcal{D}_2 = \operatorname{diag}(2\mathcal{M}, 2I) - (\mathcal{B}^*,\ I)^*(\mathcal{B}^*,\ I) \succeq 0$, we can show that

$$\begin{aligned} &\frac{1}{2}\left\|(\mathcal{B}^*,\ I)\begin{pmatrix} z \\ Z \end{pmatrix} + G_2\right\|^2 - \langle d, z\rangle + \frac{c}{2}\,\|\, z - z_0\,\|^2 + \frac{1}{2}\left\|\begin{pmatrix} z \\ Z \end{pmatrix} - \begin{pmatrix} z_0 \\ Z_0 \end{pmatrix}\right\|_{\mathcal{D}_2}^2 \\ &= \langle z, \mathcal{M}z\rangle + \frac{c}{2}\|z\|^2 - \langle z, h\rangle + \langle Z, Z\rangle + \langle Z, G_2 - Z_0 + \mathcal{B}^* z_0\rangle + \kappa \\ &= \kappa + \langle Z, Z\rangle + \langle Z, G_2 - Z_0 + \mathcal{B}^* z_0\rangle + \sum_{i=1}^q \frac{1}{2}\langle (2\mathcal{M}_i + c\,I) z_i,\, z_i\rangle - \langle z_i, h_i\rangle, \end{aligned}$$

where $h = d + cz_0 + 2\mathcal{M}z_0 - \mathcal{B}(G_2 + Z_0 + \mathcal{B}^* z_0)$ and $\kappa = \frac{c}{2}\|z_0\|^2 + \frac{1}{2}\|\left(\begin{smallmatrix} z_0 \\ Z_0 \end{smallmatrix}\right)\|_{\mathcal{D}_2}^2 + \frac{1}{2}\|G_2\|^2$. Thus, we can obtain the following decomposition for solving (22) with an additional proximal term:

$$\begin{aligned} &\underset{z,Z}{\text{minimize}} \left\{ \frac{1}{2}\,\|\,\mathcal{B}^* z + Z + G_2\,\|^2 - \langle d, z\rangle + \frac{c}{2}\,\|\, z - z_0\,\|^2 \right. \\ &\qquad \left. + \frac{1}{2}\left\|\begin{pmatrix} z \\ Z \end{pmatrix} - \begin{pmatrix} z_0 \\ Z_0 \end{pmatrix}\right\|_{\mathcal{D}_2}^2 \ \middle|\ z \geq 0,\ Z \geq 0 \right\} \\ &\quad\Longleftrightarrow \underset{Z}{\text{minimize}} \left\{ \langle Z, Z\rangle + \langle Z, G_2 - Z_0 + \mathcal{B}^* z_0\rangle \mid Z \geq 0 \right\} \\ &\quad + \sum_{i=1}^q \underset{z_i}{\text{minimize}} \left\{ \frac{1}{2}\langle (2\mathcal{M}_i + cI) z_i,\, z_i\rangle - \langle z_i, h_i\rangle \ \middle|\ z_i \geq 0 \right\}. \end{aligned}$$

Observe that the subproblem with respect to Z can be solved analytically. For each i, the decoupled subproblem with respect to z_i is simple convex quadratic program of the form (24) that can be solved by the APG-SNCG method.

5. Numerical experiments. In this section, we test our imABCD algorithm on solving the dual problem (19). The equality and inequality constraints are generated from the doubly nonnegative relaxation of a binary integer nonconvex quadratic (ex-BIQ) programming that was considered in [51]:

$$\begin{array}{ll} \underset{x,Y,X}{\text{minimize}} & \frac{1}{2}\langle Q, Y\rangle + \langle c, x\rangle \\ \text{subject to} & \mathrm{Diag}(Y) = x, \quad \alpha = 1, \quad X = \begin{pmatrix} Y & x \\ x^T & \alpha \end{pmatrix}, \quad X \in \mathbb{S}^n_+, \quad X \geq 0, \\ & -Y_{ij} + x_i \geq 0, \quad -Y_{ij} + x_j \geq 0, \quad Y_{ij} - x_i - x_j \geq -1, \\ & \quad \forall\, i < j,\ j = 2, \ldots, n-1. \end{array}$$

The test data for Q and c are taken from the Biq Mac Library maintained by Wiegele, which is available at http://biqmac.uni-klu.ac.at/biqmaclib.html.

We set $G = -\frac{1}{2}\left(\begin{smallmatrix} Q & c \\ c & 0 \end{smallmatrix}\right)$ in (1). Under Slater's condition, the KKT optimality conditions of the problem (1) are

$$\begin{cases} X = G + \mathcal{A}^* y + \mathcal{B}^* z + S + Z, \quad \mathcal{A}X = b, \\ \mathcal{B}X - d = \Pi_{\geq 0}(\mathcal{B}X - d - z), \quad X = \Pi_{\mathbb{S}^n_+}(X - S), \quad X = \Pi_{\geq 0}(X - Z). \end{cases}$$

We measure the accuracy of an approximate dual solution (y, z, S, Z) by the relative residual of the KKT system $\eta \triangleq \max\{\eta_1, \eta_2, \eta_3, \eta_4\}$, where, by letting $X = G + \mathcal{A}^* y + \mathcal{B}^* z + S + Z$,

$$\begin{cases} \eta_1 \triangleq \dfrac{\|\mathcal{A}X - b\|}{1 + \|b\|}, \quad \eta_2 \triangleq \dfrac{\|\mathcal{B}X - d - \Pi_{\geq 0}(\mathcal{B}X - d - z)\|}{1 + \|d\|}, \\ \eta_3 \triangleq \dfrac{\|X - \Pi_{\mathbb{S}^n_+}(X - S)\|}{1 + \|X\| + \|S\|}, \quad \eta_4 \triangleq \dfrac{\|X - \Pi_{\geq 0}(X - Z)\|}{1 + \|X\| + \|Z\|}. \end{cases}$$

We also report the duality gap defined by

$$\eta_{\text{gap}} \triangleq \frac{\text{obj}_{\text{p}} - \text{obj}_{\text{d}}}{1 + |\text{obj}_{\text{p}}| + |\text{obj}_{\text{d}}|},$$

where $\text{obj}_{\text{p}} \triangleq \frac{1}{2}\|X - G\|^2$ and $\text{obj}_{\text{d}} \triangleq -\frac{1}{2}\|\mathcal{A}^* y + \mathcal{B}^* z + S + Z + G\|^2 + \langle b, y\rangle + \langle d, z\rangle + \frac{1}{2}\|G\|^2$. We stop the algorithms if $\eta < \varepsilon$ for some prescribed accuracy ε. Throughout the numerical experiments, we initialize all the algorithms by setting the dual variables to be zero.

In order to demonstrate the importance for incorporating the second-order information when solving the subproblems, we compare our imABCD method with the two-block accelerated block coordinate gradient descent algorithm proposed by Chambolle and Pock in [8]. For a fair comparison, the two blocks are also taken as (y, S) and (z, Z). Let $\lambda_{\max}(\mathcal{B}\mathcal{B}^*)$ be the largest eigenvalue of $\mathcal{B}\mathcal{B}^*$. The algorithm adapted from [8] is given as follows.

ABCGD: An accelerated block coordinate gradient descent algorithm for solving (19).

Initialization. Choose an initial point $W^1 = \widetilde{W}^0 \in \mathbb{W}$. Let $t_1 = 1$.

Step 1. Compute $R^{k+\frac{1}{2}} = \mathcal{A}^* y^k + \mathcal{B}^* z^k + S^k + Z^k + G$ and

$$\begin{cases} \widetilde{y}^k = y^k - \frac{1}{2}(\mathcal{A}\mathcal{A}^*)^{-1}\left(\mathcal{A}R^{k+\frac{1}{2}} - b\right), \\ \widetilde{S}^k = \Pi_{\mathbb{S}^n_+}\left(S^k - \frac{1}{2}R^{k+\frac{1}{2}}\right). \end{cases}$$

Step 2. Compute $R^k = \mathcal{A}^* \widetilde{y}^k + \mathcal{B}^* z^k + \widetilde{S}^k + Z^k + G$ and

$$\begin{cases} \widetilde{z}^k = \Pi_{\geq 0}\left(z^k - \frac{1}{2\lambda_{\max}(\mathcal{B}\mathcal{B}^*)}\left(\mathcal{B}R^k - d\right)\right), \\ \widetilde{Z}^k = \Pi_{\geq 0}\left(Z^k - \frac{1}{2}R^k\right). \end{cases}$$

Step 3. Compute

$$\begin{cases} t_{k+1} = \frac{1}{2}\left(1 + \sqrt{1 + 4t_k^2}\right), \\ W^{k+1} = \widetilde{W}^k + \frac{t_k - 1}{t_{k+1}}\left(\widetilde{W}^k - \widetilde{W}^{k-1}\right). \end{cases}$$

We note that as the adapted algorithm from [8] does not cater for inexact computations of the subproblems, we have adopted appropriate majorizations so that the associated subproblems can be solved exactly. In fact, the above ABCGD algorithm is a special case of our imABCD algorithm by taking the majorized operator in Assumption 1 as $\widehat{\mathcal{Q}}_{11} = \mathrm{Diag}(2\mathcal{A}\mathcal{A}^*, 2\mathcal{I})$ and $\widehat{\mathcal{Q}}_{22} = \mathrm{Diag}(2\lambda_{\max}(\mathcal{B}\mathcal{B}^*)\mathcal{I}, 2\mathcal{I})$.

Figure 1 shows the performance profile of the imABCD and ABCGD algorithms for some large-scale ex-BIQ problems with $\varepsilon = 10^{-6}$. A point (x, y) is on the performance profile curve of a method if and only if it can solve exactly $(100y)\%$ of all the tested problems at most x times slower than any other methods. The detailed numerical results are presented in Table 1. The first four columns list the problem names, the dimension of the variable y (m_E), z (m_I), and the size of the matrix G (n_s), respectively. The last several columns report the number of iterations, the

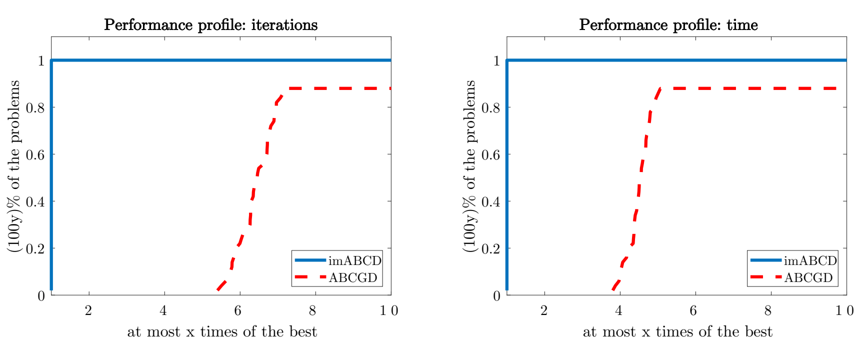

FIG. 1. *Performance profile of imABCD and ABCGD with* $\varepsilon = 10^{-6}$.

TABLE 1
The performance of imABCD and ABCGD with accuracy $\varepsilon = 10^{-6}$.

		Iterations	η	η_{gap}	Time
Problem	$m_E; m_I \mid n_s$	imabcd|abcgd	imabcd|abcgd	imabcd|abcgd	imabcd|abcgd
be120.3.1	121 ; 21420 | 121	4141 | 27781	9.9-7 | 9.9-7	-2.3-7 | -5.6-8	23 | 1:56
be120.3.2	121 ; 21420 | 121	4114 | 23809	9.9-7 | 9.9-7	-2.1-7 | -6.3-8	23 | 1:40
be120.3.3	121 ; 21420 | 121	3599 | 21867	9.9-7 | 9.9-7	-2.3-8 | -6.6-8	20 | 1:30
be120.3.4	121 ; 21420 | 121	4917 | 31783	9.9-7 | 9.9-7	-3.5-7 | -6.7-8	28 | 2:13
be120.3.5	121 ; 21420 | 121	5646 | 31076	9.9-7 | 9.9-7	-5.4-8 | -7.1-8	33 | 2:10
be120.3.6	121 ; 21420 | 121	3951 | 26558	9.9-7 | 9.9-7	-1.8-7 | -4.8-8	22 | 1:52
be120.3.7	121 ; 21420 | 121	4170 | 26176	9.9-7 | 9.9-7	-2.7-7 | -6.5-8	23 | 1:50
be120.3.8	121 ; 21420 | 121	3791 | 23796	9.9-7 | 9.9-7	-1.6-7 | -3.8-8	21 | 1:40
be120.3.9	121 ; 21420 | 121	4932 | 28518	9.9-7 | 9.9-7	-2.2-7 | -5.2-8	29 | 2:00
be120.3.10	121 ; 21420 | 121	4254 | 24803	9.9-7 | 9.9-7	-3.4-7 | -5.3-8	24 | 1:44
be120.8.1	121 ; 21420 | 121	5660 | 32200	9.9-7 | 9.9-7	-3.0-7 | -8.1-8	33 | 2:14
be120.8.2	121 ; 21420 | 121	5857 | 35336	9.9-7 | 9.9-7	-3.1-7 | -7.1-8	35 | 2:26
be120.8.3	121 ; 21420 | 121	5236 | 33259	9.9-7 | 9.9-7	-4.9-7 | -9.2-8	30 | 2:19
be120.8.4	121 ; 21420 | 121	6507 | 40964	9.9-7 | 9.9-7	-4.3-7 | -8.0-8	39 | 2:51
be120.8.5	121 ; 21420 | 121	5985 | 41263	9.9-7 | 9.9-7	-3.1-7 | -6.3-8	36 | 2:50
be120.8.6	121 ; 21420 | 121	4714 | 29524	9.9-7 | 9.9-7	-4.7-7 | -7.8-8	28 | 2:00
be120.8.7	121 ; 21420 | 121	4452 | 29722	9.9-7 | 9.9-7	-3.9-7 | -6.5-8	25 | 2:02
be120.8.8	121 ; 21420 | 121	5982 | 35240	9.9-7 | 9.9-7	-2.8-7 | -6.4-8	35 | 2:26
be120.8.9	121 ; 21420 | 121	5799 | 37397	9.9-7 | 9.9-7	-3.3-7 | -8.5-8	34 | 2:35
be120.8.10	121 ; 21420 | 121	5630 | 35274	9.9-7 | 9.9-7	-3.4-7 | -8.0-8	33 | 2:26
be250.1	251 ; 93375 | 251	3675 | 25038	9.9-7 | 9.9-7	-5.8-7 | -4.9-8	57 | 4:28
be250.2	251 ; 93375 | 251	4213 | 29313	9.9-7 | 9.9-7	-3.7-7 | -6.8-8	1:04 | 5:15
be250.3	251 ; 93375 | 251	4011 | 27211	9.9-7 | 9.8-7	-4.5-7 | -4.4-8	1:03 | 4:52
be250.4	251 ; 93375 | 251	4059 | 28985	9.9-7 | 9.9-7	-3.6-7 | -5.8-8	1:03 | 5:14
be250.5	251 ; 93375 | 251	4361 | 29277	9.9-7 | 9.9-7	-3.9-7 | -5.2-8	1:08 | 5:20
bqp100-1	101 ; 14850 | 101	7277 | 50000	9.9-7 | 1.1-6	-1.2-7 | -1.1-7	38 | 2:46
bqp100-2	101 ; 14850 | 101	3793 | 24170	9.9-7 | 9.9-7	-1.4-7 | -6.7-8	17 | 1:20
bqp100-3	101 ; 14850 | 101	3627 | 22570	9.9-7 | 9.9-7	6.9-8 | -5.1-8	17 | 1:14
bqp100-4	101 ; 14850 | 101	4297 | 27893	9.9-7 | 9.9-7	-2.2-7 | -5.9-8	20 | 1:32
bqp100-5	101 ; 14850 | 101	5095 | 34243	9.9-7 | 9.9-7	-1.2-7 | -4.8-8	25 | 1:53
bqp500-1	501 ; 374250 | 501	6523 | 50000	9.9-7 | 1.3-6	-1.4-6 | -1.2-7	13:35 | 54:18
bqp500-2	501 ; 374250 | 501	7106 | 50000	9.9-7 | 1.7-6	-1.3-6 | -1.6-7	15:02 | 54:23
bqp500-3	501 ; 374250 | 501	6067 | 50000	9.9-7 | 1.1-6	-1.1-6 | -9.1-8	12:56 | 54:34
bqp500-4	501 ; 374250 | 501	5822 | 50000	9.9-7 | 1.2-6	-1.1-6 | -8.0-8	12:51 | 54:53
bqp500-5	501 ; 374250 | 501	7203 | 50000	9.9-7 | 1.8-6	-1.6-6 | -1.7-7	15:55 | 54:46
gka1e	201 ; 59700 | 201	5293 | 37861	9.9-7 | 9.9-7	-2.6-7 | -4.8-8	1:15 | 5:56
gka2e	201 ; 59700 | 201	4623 | 29338	9.9-7 | 9.9-7	-6.8-7 | -7.1-8	1:03 | 4:35
gka3e	201 ; 59700 | 201	6033 | 40016	9.9-7 | 9.9-7	-3.7-7 | -6.0-8	1:25 | 6:20
gka4e	201 ; 59700 | 201	6760 | 47779	9.9-7 | 9.9-7	-5.5-7 | -6.9-8	1:36 | 7:37
gka5e	201 ; 59700 | 201	6247 | 42175	9.9-7 | 9.9-7	-5.3-7 | -7.8-8	1:28 | 6:43

relative residual η, the relative duality gap $\eta_{\rm gap}$, and the computation time in the format of "hours:minutes:seconds." One can see from the performance profile that the ABCGD algorithm requires at least 5 times the number of iterations taken by imABCD and is about 4 times slower than the imABCD algorithm. In particular, the ABCGD method cannot solve all the large-scale bdq500 problems within 50000 iterations, whereas our imABCD can obtain satisfactory solutions after 6000 iterations. This indicates that even though the computational cost for each cycle of the imABCD method is larger than that of the ABCGD method, this cost is compensated by taking many fewer iterations. In fact, the Newton system is well-conditioned in this case such that it takes only one or two CG iterations to compute a satisfactory Newton direction.

We also compare our imABCD algorithm with some other BCD-type methods. The first one is a direct four-block BCD method. In this case, the block z is solved by the APG-SNCG algorithm, while other blocks have analytical solutions. The second one is an enhanced version of the four-block inexact randomized ABCD method (denoted as eRABCD) that is modified from [35], where we use the proximal terms $\frac{1}{2}\|y-y^k\|^2_{\mathcal{A}\mathcal{A}^*}$ instead of $\frac{\lambda_{\max}(\mathcal{A}\mathcal{A}^*)}{2}\|y-y^k\|^2$ when updating the block y^{k+1} and $\frac{1}{2}\|z-z^k\|^2_{\mathcal{B}\mathcal{B}^*+\|\mathcal{B}\|_2\mathcal{I}}$ when updating the block z^{k+1}. (Note that the addition of the term $\|B\|_2\mathcal{I}$ is to make the associated subproblem strongly convex so that it is easier to solve.) A similar modification has also been used in [51] when the randomized BCD algorithm is used to solve a class of positive semidefinite feasibility problems. The detailed steps of the eRABCD are given below.

---

eRABCD: A four-block inexact enhanced randomized ABCD algorithm for solving (19)

Initialization. Choose an initial point $\widetilde{W}^0=\widehat{W}^0\in\mathbb{W}$. Set $k=1$ and $\alpha_0=\frac{1}{4}$. Let $\{\varepsilon_k\}$ be a given summable sequence of error tolerance such that the error vector $\delta_z^k\in\mathbb{R}^{m_I}$ satisfies $\|\delta_z^k\|\le\varepsilon_k$.

Step 1. Denote $\widehat{R}^k=\mathcal{A}^*\hat{y}^k+\mathcal{B}^*\hat{z}^k+\widehat{S}^k+\widehat{Z}^k+G$. Choose $i_k\in\{1,2,3,4\}$ uniformly at random and update $\widetilde{W}^{k+1}_{i_k}$ according to the following rule if the kth block is selected:

$$\begin{cases} i_k=1: & \widetilde{y}^{k+1}=(\mathcal{A}\mathcal{A}^*)^{-1}((b-\mathcal{A}\widehat{R}^k)/(4\alpha_k)+\mathcal{A}\mathcal{A}^*\widetilde{y}^k),\\ i_k=2: & \widetilde{z}^{k+1}\approx\underset{z\ge0}{\operatorname{argmin}}\left\{\langle\nabla_z h(\widehat{W}^{k+1}),z\rangle+\dfrac{4\alpha_k}{2}\|z-\widetilde{z}^k\|^2_{\mathcal{B}\mathcal{B}^*+\|\mathcal{B}\|_2\mathcal{I}}+\langle z,\delta_z^k\rangle\right\},\\ i_k=3: & \widetilde{Z}^{k+1}=\Pi_{\ge0}(\widetilde{Z}^k-\widehat{R}^k/(4\alpha_k)),\\ i_k=4: & \widetilde{S}^{k+1}=\Pi_{\mathbb{S}^n_+}(\widetilde{S}^k-\widehat{R}^k/(4\alpha_k)).\end{cases}$$

Set $\widetilde{W}^{k+1}_i=\widetilde{W}^k_i$ for all $i\ne i_k$, $k=1,2,3,4$.

Step 2. Set $W_i^{k+1}=\begin{cases}\widehat{W}_i^k+4\alpha_k(\widetilde{W}_i^{k+1}-\widetilde{W}_i^k), & i=i_k,\\ \widehat{W}_i^k, & i\ne i_k,\end{cases}\qquad i=1,2,3,4.$

Step 3. Compute $\begin{cases}\alpha_{k+1}=\dfrac{1}{2}\left(\sqrt{\alpha_k^4+4\alpha_k^2}-\alpha_k^2\right),\\ \widehat{W}^{k+1}=(1-\alpha_{k+1})W^{k+1}+\alpha_{k+1}\widetilde{W}^{k+1}.\end{cases}$

TABLE 2
The performance of eRABCD, meRABCD, BCD, and mBCD with accuracy $\varepsilon = 10^{-5}$.

	Iteration	η	Time
Problem	erabcd|merabcd|bcd|mbcd	erabcd|merabcd|bcd|mbcd	erabcd|merabcd|bcd|mbcd
bqp50-1	5168 | 20172 | 110848 | 500000	9.9-6 | 9.9-6 | 9.9-6 | **7.6-3**	19 | 40 | 8:02 | 18:02
bqp100-1	7789 | 39167 | 203733 | 500000	9.3-6 | 9.0-6 | 9.9-6 | **1.1-2**	48 | 2:19 | 33:24 | 52:57

In order to know whether our proposed APG-SNCG method could universally improve the efficiency for different outer loops, we also test two variants of the BCD and eRABCD, where the block z is updated by the proximal gradient step. They are named mBCD and meRABCD. The numerical performance of two selected test examples is shown in Table 2. Note that in the table, one iteration of the (m)BCD method refers to one sweep of all the four blocks, while that of the (m)eRABCD method refers to one sweep of the three steps presented in the algorithm. One can see that the mBCD and meRABCD perform much worse than their inexact counterparts. These numerical results may indicate that if one of the blocks is computationally intensive (such as the block S in (19) that requires the eigenvalue decomposition for each update), a small proximal term is always preferred for the other blocks in order to reduce the the number of iterations taken by the algorithm, which is also the number of updates required by the difficult block. In fact, it has also been noted in [45] that smaller number of blocks can speed up the performance of the stochastic dual Newton method.

Table 3 lists the numerical performance of the imABCD, eRABCD and BCD methods, with the performance profile given in Figure 2. One can see that the BCD algorithm is much less efficient than the other algorithms, as all the test examples cannot be solved to the required accuracy within 50000 iteration steps (we thus do not include its performance in the performance profile). This phenomenon has clearly demonstrated the power of the acceleration technique. Observe that the imABCD method is about 3 times faster than the eRABCD method.

Based on the above numerical results, we may conclude that the efficiency of the imABCD algorithm can be attributed to the double acceleration procedure: the outer acceleration of the two-block coordinate descent method, and the inner acceleration by the proper incorporation of the second-order information through solving the subproblems in each iteration by Newton type methods.

6. Conclusions.

In this paper, for the purpose of overcoming the potential degeneracy of the matrix best approximation problem (1), we have proposed a two-block imABCD method with each block solved by the Newton type methods. Extensive numerical results demonstrated the efficiency and robustness of our algorithm in solving various instances of the large-scale matrix approximation problems. We believe that our algorithmic framework is a powerful tool to deal with degenerate problems and might be adapted to other convex matrix optimization problems in the future.

TABLE 3
The performance of imABCD, eRABCD, and BCD with accuracy $\varepsilon = 10^{-6}$.

			Iterations			η			η_{gap}			Time		
Problem	$m_E; m_I$	n_s	imabcd	erabcd	bcd	imabcd	erabcd	bcd	imabcd	erabcd	bcd	imabcd	erabcd	bcd
be150.3.1	151 ; 33525	151	5920	28593	50000	9.9-7	9.7-7	3.9-5	-1.8-7	-8.8-7	-1.2-5	47	2:02	7:24
be150.3.2	151 ; 33525	151	4576	24433	50000	9.9-7	9.7-7	3.9-5	-5.2-7	-1.3-6	-1.4-5	35	1:42	7:21
be150.3.3	151 ; 33525	151	4522	23251	50000	9.9-7	9.9-7	4.0-5	-4.9-7	-9.5-7	-1.3-5	34	1:37	7:19
be150.3.4	151 ; 33525	151	4374	23088	50000	9.9-7	9.6-7	3.7-5	-3.9-7	-8.1-7	-1.3-5	33	1:36	7:14
be150.3.5	151 ; 33525	151	5997	29844	50000	9.9-7	9.9-7	3.8-5	-3.0-7	-1.0-6	-9.8-6	46	2:04	7:23
be150.3.6	151 ; 33525	151	4131	22946	50000	9.9-7	9.9-7	3.9-5	-2.8-7	-6.3-7	-1.7-5	31	1:33	7:53
be150.3.7	151 ; 33525	151	4929	24820	50000	9.9-7	9.9-7	4.0-5	-4.9-7	-1.2-6	-1.1-5	37	1:43	7:19
be150.3.8	151 ; 33525	151	4424	23018	50000	9.9-7	9.9-7	3.6-5	-4.5-7	-7.5-7	-1.4-5	33	1:34	7:12
be150.3.9	151 ; 33525	151	5489	27988	50000	9.9-7	9.9-7	3.9-5	-3.7-7	-6.7-7	-1.1-5	44	1:57	7:17
be150.3.10	151 ; 33525	151	5646	28607	50000	9.9-7	9.6-7	3.9-5	-2.7-7	-7.1-7	-1.2-5	44	1:59	7:15
be150.8.1	151 ; 33525	151	6006	30134	50000	9.9-7	9.9-7	3.7-5	-4.0-7	-9.5-7	-9.0-6	46	2:01	7:18
be150.8.2	151 ; 33525	151	6377	33413	50000	9.9-7	9.9-7	3.7-5	-5.4-7	-9.8-7	-8.3-6	50	2:17	7:22
be150.8.3	151 ; 33525	151	6935	31222	50000	9.9-7	9.9-7	4.0-5	-5.4-7	-1.2-6	-8.3-6	54	2:10	7:22
be150.8.4	151 ; 33525	151	6091	29849	50000	9.9-7	9.9-7	3.8-5	-4.0-7	-1.1-6	-8.0-6	46	2:03	7:18
be150.8.5	151 ; 33525	151	6492	33640	50000	9.9-7	9.9-7	4.2-5	-4.5-7	-8.3-7	-7.6-6	52	2:22	7:25
be150.8.6	151 ; 33525	151	5905	29849	50000	9.9-7	9.9-7	3.6-5	-3.3-7	-1.2-6	-9.7-6	44	2:04	7:14
be150.8.7	151 ; 33525	151	5918	31215	50000	9.9-7	9.9-7	3.7-5	-4.7-7	-1.3-6	-1.0-5	44	2:08	7:13
be150.8.8	151 ; 33525	151	6554	33624	50000	9.9-7	9.9-7	3.9-5	-5.5-7	-9.2-7	-7.9-6	52	2:21	7:23
be150.8.9	151 ; 33525	151	6170	28607	50000	9.9-7	9.8-7	4.0-5	-2.6-7	-8.4-7	-7.4-6	46	1:57	7:21
be150.8.10	151 ; 33525	151	6232	29847	50000	9.9-7	9.9-7	4.0-5	-3.4-7	-1.1-6	-9.4-6	47	2:01	7:17
be200.3.1	201 ; 59700	201	6159	28607	50000	9.9-7	9.6-7	4.1-5	-3.0-7	-1.1-6	-1.5-5	1:12	3:02	11:37
be200.3.2	201 ; 59700	201	5463	28590	50000	9.9-7	9.6-7	4.3-5	-6.4-7	-1.1-6	-1.6-5	1:04	3:06	13:36
be200.3.3	201 ; 59700	201	6182	30123	50000	9.9-7	9.8-7	4.3-5	-4.7-7	-1.4-6	-1.3-5	1:14	3:16	12:01
be200.3.4	201 ; 59700	201	5922	28607	50000	9.9-7	9.8-7	4.4-5	-5.8-7	-1.3-6	-1.3-5	1:11	3:03	12:07
be200.3.5	201 ; 59700	201	6511	33535	50000	9.9-7	9.8-7	4.1-5	-5.8-7	-1.3-6	-1.2-5	1:19	3:44	12:08
be200.3.6	201 ; 59700	201	5492	28211	50000	9.9-7	9.9-7	4.3-5	-1.8-7	-9.8-7	-1.6-5	1:08	3:03	13:56
be200.3.7	201 ; 59700	201	5403	28593	50000	9.9-7	9.8-7	4.2-5	-5.4-7	-1.1-6	-1.6-5	1:05	3:08	12:41
be200.3.8	201 ; 59700	201	6108	30028	50000	9.9-7	9.9-7	4.1-5	-5.2-7	-1.3-6	-1.4-5	1:14	3:22	12:05
be200.3.9	201 ; 59700	201	6095	29909	50000	9.9-7	9.9-7	4.2-5	-3.9-7	-1.2-6	-1.4-5	1:15	3:12	11:56
be200.3.10	201 ; 59700	201	5554	28211	50000	9.9-7	9.9-7	4.0-5	-6.0-7	-1.2-6	-1.6-5	1:06	3:04	11:51
be200.8.1	201 ; 59700	201	7009	34304	50000	9.9-7	9.8-7	4.0-5	-7.2-7	-1.4-6	-7.9-6	1:24	3:50	12:16
be200.8.2	201 ; 59700	201	6295	33897	50000	9.9-7	9.9-7	4.0-5	-7.9-7	-1.5-6	-1.0-5	1:15	3:47	12:05
be200.8.3	201 ; 59700	201	7310	35879	50000	9.9-7	9.9-7	4.1-5	-5.9-7	-1.2-6	-8.3-6	1:30	4:00	12:22
be200.8.4	201 ; 59700	201	7726	34628	50000	9.9-7	9.7-7	4.2-5	-6.4-7	-1.2-6	-6.7-6	1:40	3:58	12:36
be200.8.5	201 ; 59700	201	8082	37110	50000	9.9-7	9.9-7	4.1-5	-7.1-7	-1.0-6	-6.8-6	1:46	4:13	12:43
be200.8.6	201 ; 59700	201	7058	34304	50000	9.9-7	9.9-7	3.9-5	-6.8-7	-1.4-6	-7.6-6	1:27	3:47	12:23
be200.8.7	201 ; 59700	201	5642	31222	50000	9.9-7	9.9-7	3.9-5	-7.3-7	-1.7-6	-1.5-5	1:07	3:27	12:12
be200.8.8	201 ; 59700	201	6047	30134	50000	9.9-7	9.9-7	3.9-5	-5.8-7	-1.4-6	-1.1-5	1:13	3:20	12:15
be200.8.9	201 ; 59700	201	7445	34304	50000	9.9-7	9.7-7	4.0-5	-7.4-7	-1.3-6	-7.4-6	1:35	3:50	12:23
be200.8.10	201 ; 59700	201	7028	34304	50000	9.9-7	9.3-7	3.9-5	-7.6-7	-1.4-6	-8.3-6	1:25	3:49	12:13
be250.1	251 ; 93375	251	3675	19265	50000	9.9-7	9.8-7	5.2-5	-5.8-7	-1.5-6	-3.0-5	57	2:34	18:42
be250.2	251 ; 93375	251	4213	22835	50000	9.9-7	9.9-7	5.1-5	-3.7-7	-9.6-7	-2.8-5	1:04	2:58	19:55
be250.3	251 ; 93375	251	4011	23125	50000	9.9-7	9.8-7	5.1-5	-4.5-7	-1.1-6	-3.0-5	1:03	3:12	19:53
be250.4	251 ; 93375	251	4059	23125	50000	9.9-7	9.5-7	4.7-5	-3.6-7	-1.0-6	-2.4-5	1:03	3:09	17:13
be250.5	251 ; 93375	251	4361	23125	50000	9.9-7	9.8-7	5.2-5	-3.9-7	-9.5-7	-2.7-5	1:08	3:07	20:24
be250.6	251 ; 93375	251	4122	22708	50000	9.9-7	9.9-7	5.2-5	-2.0-7	-6.9-7	-2.9-5	1:08	3:30	19:53
be250.7	251 ; 93375	251	5756	28225	50000	9.9-7	9.8-7	4.9-5	-3.4-7	-1.1-6	-2.1-5	1:37	4:21	17:32
be250.8	251 ; 93375	251	3398	18562	50000	9.9-7	9.8-7	5.5-5	-5.0-7	-10.0-7	-3.3-5	57	2:40	20:17
be250.9	251 ; 93375	251	5299	27990	50000	9.9-7	9.9-7	4.9-5	-5.8-7	-1.0-6	-2.4-5	1:29	4:18	17:33
be250.10	251 ; 93375	251	4678	24443	50000	9.9-7	9.9-7	5.0-5	-3.5-7	-1.0-6	-2.6-5	1:22	3:40	20:02
bqp250-1	251 ; 93375	251	7546	34304	50000	9.9-7	9.6-7	4.7-5	-7.6-7	-1.6-6	-1.4-5	2:07	5:37	18:28
bqp250-2	251 ; 93375	251	6788	34304	50000	9.9-7	9.9-7	4.8-5	-6.8-7	-1.7-6	-1.5-5	1:57	5:22	18:02
bqp250-3	251 ; 93375	251	5433	28225	50000	9.9-7	9.9-7	4.8-5	-8.1-7	-1.4-6	-2.1-5	1:29	4:09	20:32
bqp250-4	251 ; 93375	251	6997	34064	50000	9.9-7	9.9-7	5.0-5	-5.6-7	-1.4-6	-1.3-5	1:58	5:18	18:28
bqp250-5	251 ; 93375	251	7495	34304	50000	9.9-7	9.4-7	4.7-5	-8.3-7	-1.6-6	-1.5-5	2:09	5:26	18:27
bqp250-6	251 ; 93375	251	6362	34131	50000	9.9-7	9.9-7	4.9-5	-7.8-7	-1.5-6	-1.4-5	1:50	5:26	19:17
bqp250-7	251 ; 93375	251	5950	33578	50000	9.9-7	9.9-7	5.0-5	-5.4-7	-9.7-7	-2.1-5	1:40	5:18	20:21
bqp250-8	251 ; 93375	251	7693	34304	50000	9.9-7	9.9-7	5.0-5	-6.8-7	-1.4-6	-1.1-5	2:20	5:39	18:57
bqp250-9	251 ; 93375	251	5625	33431	50000	9.9-7	9.9-7	4.8-5	-8.9-7	-1.3-6	-1.9-5	1:37	5:10	18:52
bqp250-10	251 ; 93375	251	6924	34187	50000	9.9-7	9.9-7	4.6-5	-6.1-7	-1.7-6	-1.7-5	1:59	5:25	20:13

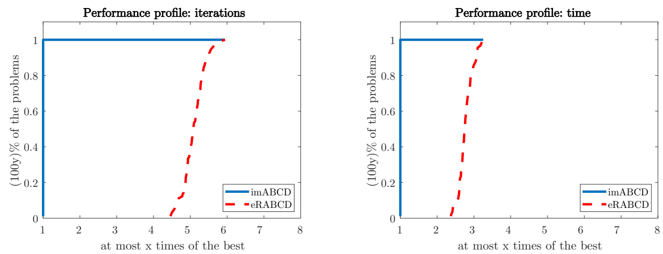

FIG. 2. *Performance profile of imABCD and eRABCD for with accuracy* $\varepsilon = 10^{-6}$.

REFERENCES

[1] Z. Allen-Zhu, Z. Qu, P. Richtárik, and Y. Yuan, *Even faster accelerated coordinate descent using non-uniform sampling*, in Proceedings of the International Conference on Machine Learning, 2016, pp. 1110–1119.

[2] L. E. Andersson and T. Elfving, *An algorithm for constrained interpolation*, SIAM J. Sci. Stat. Comput., 8 (1987), pp. 1012–1025.

[3] H. H. Bauschke and J. M. Borwein, *On projection algorithms for solving convex feasibility problems*, SIAM Rev., 38 (1996), pp. 367–426.

[4] H. H. Bauschke, J. M. Borwein, and W. Li, *Strong conical hull intersection property, bounded linear regularity, Jameson's property* (G), *and error bounds in convex optimization*, Math. Program., 86 (1999), pp. 135–160.

[5] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.

[6] A. Beck and L. Tetruashvili, *On the convergence of block coordinate descent type methods*, SIAM J. Optim., 23 (2013), pp. 2037–2060.

[7] S. Boyd and L. Xiao, *Least-squares covariance matrix adjustment*, SIAM J. Matrix Anal. Appl., 27 (2005), pp. 532–546.

[8] A. Chambolle and T. Pock, *A remark on accelerated block coordinate descent for computing the proximity operators of a sum of convex functions*, SIAM J. Comput. Math., 1 (2015), pp. 29–54.

[9] F. H. Clarke. *Optimization and Nonsmooth Analysis*, Classics in Appl. Math. 5, SIAM Philadelphia, 1990.

[10] M. B. Cohen, J. Diakonikolas and L. Orecchia, *On acceleration with noise-corrupted gradients*, in Proceedings of ICML, 2018.

[11] A. d'Aspremont, *Smooth optimization with approximate gradient*, SIAM J. Optim., 19 (2008), pp. 1171–1183.

[12] O. Devolder, F. Glineur, and Y. Nesterov, *First-order methods of smooth convex optimization with inexact oracle*, Math. Program., 146 (2014), pp. 37–75.

[13] J. Diakonikolas and L. Orecchia. *Alternating randomized block coordinate descent*, in Proceedings of ICML, 2018.

[14] A. L. Dontchev and B. D. Kalchev, *Duality and well-posedness in convex interpolation*, Numer. Funct. Anal. Optim., 10 (1989), pp. 673–689.

[15] A. L. Dontchev, H. Qi, and L. Qi, *Convergence of Newton's method for convex best interpolation*, Numer. Math., 87 (2001), pp. 435–456.

[16] A. L. Dontchev, H. Qi, and L. Qi, *Quadratic convergence of Newton's method for convex interpolation and smoothing*, Constr. Approx., 19 (2003).

[17] P. Dvurechensky and A. Gasnikov, *Stochastic intermediate gradient method for convex problems with stochastic inexact oracle*, J. Optim. Theory Appl., 171 (2016), pp. 121–145.

[18] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems: Volume* I, Springer, New York, 2003.

[19] O. Fercoq and P. Richtárik, *Accelerated, parallel, and proximal coordinate descent*, SIAM J. Optim., 25 (2015), pp. 1997–2023.

[20] O. Fercoq and P. Richtárik, *Optimization in high dimensions via accelerated, parallel, and proximal coordinate descent*, SIAM Rev., 58 (2016), pp. 739–771.

[21] A. Fischer, *Solution of monotone complementarity problems with locally Lipschitzian functions*, Math. Program., 76 (1997), pp. 513–532.

[22] Y. Gao and D. F. Sun, *Calibrating least squares semidefinite programming with equality and inequality constraints*, SIAM J. Matrix Anal. Appl., 31 (2009), pp. 1432–1457.

[23] S. Ghadimi and G. Lan, *Optimal stochastic approximation algorithms for strongly convex stochastic composite optimization* I*: A generic algorithmic framework*, SIAM J. Optim., 22 (2012), pp. 1469–1492.

[24] S. Ghadimi and G. Lan, *Optimal stochastic approximation algorithms for strongly convex stochastic composite optimization* II*: Shrinking procedures and optimal algorithms*, SIAM J. Optim., 23 (2013), pp. 2061–2089.

[25] M. Gurbuzbalaban, A. Ozdaglar, N. Vanli, and S. Wright, *Randomness and Permutations in Coordinate Descent Methods*, arXiv:1803.08200, 2018.

[26] J. B. Hiriart-Urruty and C. Lemaéchal, *Convex Analysis and Minimization Algorithms,* Springer, Berlin, 1993.

[27] M. Hong, X. Wang, M. Razaviyayn, and Z. Q. Luo, *Iteration complexity analysis of block coordinate descent methods*, Math. Program., 163 (2017), pp. 85–114.

[28] L. D. Irvine, S. P. Marin, and P. W. Smith, *Constrained interpolation and smoothing*, Constr. Approx., 2 (1986), pp. 129–151.

[29] K. F. Jiang, D. F. Sun, and K.-C. Toh, *An inexact accelerated proximal gradient method for large scale linearly constrained convex SDP*, SIAM J. Optim., 22 (2012), pp. 1042–1064.

[30] S. Karimireddy, S. Stich, and M. Jaggi, *Adaptive balancing of gradient and update computation times using global geometry and approximate subproblems*, in Proceedings of the 21st International Conference on Artificial Intelligence and Statistics, 2018, pp. 1204–1213.

[31] O. Klopp, K, Lounici, and A. B. Tsybakov, *Robust matrix completion*, Probab. Theory Relat. Fields, 169 (2017), pp. 523–564.

[32] B. Kummer, *Newton's method for non-differentiable functions*, Adv. Math. Optim., 45 (1988), pp. 114–125.

[33] G. Lan, *An optimal method for stochastic composite optimization*, Math. Program., 133 (2012), pp. 365–397.

[34] Y. Lee and A. Sidford, *Efficient accelerated coordinate descent methods and faster algorithms for solving linear systems*, in Proceedings of the 54th Annual Symposium on Foundations of Computer Science, 2013, pp. 147–156.

[35] Q. Lin, Z. Lu, and L. Xiao, *An accelerated proximal coordinate gradient method*, in Proceedings of Advances in Neural Information Processing Systems, 2014, pp. 3059–3067.

[36] Z. Lu and L. Xiao, *On the complexity analysis of randomized block-coordinate descent methods*, Math. Program., 152 (2015), pp. 615–642.

[37] J. Malick, *A dual approach to semidefinite least-squares problems*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 272–284.

[38] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization*, SIAM J. Control Optim., 15 (1977), pp. 959–972.

[39] Y. Nesterov, *A method of solving a convex programming problem with convergence rate* $O(1/k^2)$, Soviet Math. Dokl., 27 (1983), pp. 372–376.

[40] Y. Nesterov, *Efficiency of coordinate descent methods on huge-scale optimization problems*, SIAM J. Optim., 22 (2012), pp. 341–362.

[41] Y. Nesterov and S. Stich, *Efficiency of the accelerated coordinate descent method on structured optimization problems*, SIAM J. Optim., 27 (2017), pp. 110–123.

[42] J.-S. Pang and L. Qi. *A globally convergent Newton method for convex SC^1 minimization problems*, J. Optim. Theory Appl., 85 (1995) pp. 633–648.

[43] H. Qi and D. F. Sun, *A quadratically convergent Newton method for computing the nearest correlation matrix*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 360–385.

[44] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Program., 58 (1993), pp. 353–367.

[45] Z. Qu, R. Peter, T. Martin, and F. Olivier, *SDNA: Stochastic dual Newton ascent for empirical risk minimization*, in Proceedings of the International Conference on Machine Learning, 2016, pp. 1823–1832.

[46] P. Richtárik and M. Takáč, *Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function*, Math. Program., 144 (2014), pp. 1–38.

[47] R. T. Rockafellar, *Conjugate Duality and Optimization*, CBMS-NEF Regional Conf. Ser. in Appl. Math. 16, SIAM, Philadelphia, 1974.

[48] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer, New York, 1998.

[49] M. Schmidt, N. L. Roux, and F. R. Bach, *Convergence rates of inexact proximal-gradient methods for convex optimization*, in Proceedings of Advances in Neural Information Processing Systems, 2011, pp. 1458–1466.

[50] D. F. Sun and J. Sun, *Semismooth matrix valued functions*, Math. Oper. Res., 27 (2002), pp. 150–169.

[51] D. F. Sun, K.-C. Toh, and L. Q. Yang, *An efficient inexact ABCD method for least squares semidefinite programming*, SIAM J. Optim., 26 (2016) pp. 1072–1100.
[52] J. Sun and S. Zhang, *A modified alternating direction method for convex quadratically constrained quadratic semidefinite programs*, European J. Oper. Res., 207 (2010), pp. 1210–1220.
[53] R. Tappenden, P. Richtárik, and J. Gondzio, *Inexact coordinate descent: Complexity and preconditioning*, J. Optim. Theory Appl., 170 (2016), pp. 144–176.
[54] P. Tseng, *Convergence of a block coordinate descent method for nondifferentiable minimization*, J. Optim. Theory Appl., 109 (2001), pp. 475–494.
[55] P. Tseng and S. Yun, *A coordinate gradient descent method for nonsmooth separable minimization*, Math. Program., 117 (2009), pp. 387–423.
[56] H. A. Van der Vorst, *Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput., 13 (1992), pp. 631–644.
[57] S. Villa, S. Salzo, L. Baldassarre, and A. Verri, *Accelerated and inexact forward-backward algorithms*, SIAM J. Optim., 23 (2013), pp. 1607–1633.
[58] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, *Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization*, in Proceedings of Advances in Neural Information Processing Systems, 2009, pp. 2080–2088.
[59] S. J. Wright, *Coordinate descent algorithms*, Math. Program., 151 (2015), pp. 3–34.
[60] L. Q. Yang, D. F. Sun, and K.-C. Toh, *SDPNAL+: A majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints*, Math. Program. Comput., 7 (2015), pp. 1–36.
[61] X. Y. Zhao, D. F. Sun, and K.-C. Toh, *A Newton-CG augmented Lagrangian method for semidefinite programming*, SIAM J. Optim., 20 (2010), pp. 1737–1765.