A first-order block-decomposition method for solving two-easy-block structured semidefinite programs

Renato D.C Monteiro* Camilo Ortiz† Benar F. Svaiter‡

July 27, 2012

Abstract

In this paper, we consider a first-order block-decomposition method for minimizing the sum of a convex differentiable function with Lipschitz continuous gradient, and two other proper closed convex (possibly, nonsmooth) functions with easily computable resolvents. The method presented contains two important ingredients from a computational point of view, namely: an adaptive choice of stepsize for performing an extragradient step; and the use of a scaling factor to balance the blocks. We then specialize the method to the context of conic semidefinite programming (SDP) problems consisting of two easy blocks of constraints. Without putting them in standard form, we show that four important classes of graph-related conic SDP problems automatically possess the above two-easy-block structure, namely: SDPs for θ-functions and θ_+-functions of graph stable set problems, and SDP relaxations of binary integer quadratic and frequency assignment problems. Finally, we present computational results on the aforementioned classes of SDPs showing that our method outperforms the three most competitive codes for large-scale conic semidefinite programs, namely: the boundary point (BP) method introduced by Povh et al., a Newton-CG augmented Lagrangian method, called SDPNAL, by Zhao et al., and a variant of the BP method, called the SPDAD method, by Wen et al.

1 Introduction

Let $\mathbb{R}$ denote the set of real numbers, $\mathbb{R}^n$ denote the n-dimensional Euclidean space, $\mathbb{R}^n_+$ denote the cone of nonnegative vectors in $\mathbb{R}^n$, $\mathcal{S}^n$ denote the set of all $n \times n$ symmetric matrices and $\mathcal{S}^n_+$ denote the cone of $n \times n$ symmetric positive semidefinite matrices. Let $\mathcal{X}$ and $\mathcal{W}$ be finite dimensional vector spaces and consider the conic programming problem

$$\min\{c(x) \ : \ \mathcal{A}x = b, \ x \in \mathcal{K}\}, \tag{1}$$

where $\mathcal{A} : \mathcal{X} \to \mathcal{W}$ and $c : \mathcal{X} \to \mathbb{R}$ are linear mappings, $b \in \mathcal{W}$, $c \in \mathcal{X}$ and $\mathcal{K} \subset \mathcal{X}$ is a closed convex cone. Several papers [8, 9, 20, 10, 19] in the literature discuss methods/codes for solving large-scale conic semidefinite programming problems, i.e., special cases of (1) in which

$$\mathcal{X} = \mathbb{R}^{n_u+n_l} \times \mathcal{S}^{n_s}, \qquad \mathcal{W} = \mathbb{R}^m, \qquad \mathcal{K} = \mathbb{R}^{n_u} \times \mathbb{R}^{n_l}_+ \times \mathcal{S}^{n_s}_+. \tag{2}$$

Presently, the most efficient methods/codes for solving large-scale conic SDP problems are the first-order projection-type discussed in [9, 20, 10, 19] (see also [14] for a slight variant of [9]).

*School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0205 (email: monteiro@isye.gatech.edu). The work of this author was partially supported by NSF Grants CCF-0808863 and CMMI-0900094 and ONR Grant ONR N00014-11-1-0062.

†School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0205 (email: camiort@gatech.edu).

‡IMPA, Estrada Dona Castorina 110, 22460-320 Rio de Janeiro, Brazil (email:benar@impa.br). The work of this author was partially supported by CNPq grants no. 303583/2008-8, 302962/2011-5, 480101/2008-6, 474944/2010-7, FAPERJ grants E-26/102.821/2008, E-26/102.940/2011.

More specifically, augmented Lagrangian approaches have been proposed for the dual formulation of (1) with $\mathcal{X}$, $\mathcal{W}$ and $\mathcal{K}$ as in (2) for the case when m, n_u and n_l are large (up to a few millions) and n_s is moderate (up to a few thousands). In [9, 14], a boundary point method for solving (1) is proposed which can be viewed as variants of the alternating direction method of multipliers introduced in [6, 7] applied to the dual formulation of (1). In [20], an inexact augmented Lagrangian method is proposed which solves a reformulation of the augmented Lagrangian subproblem via a semismooth Newton approach combined with the conjugate gradient method. Using the theory developed in [11], an implementation of a first-order block-decomposition (BD) algorithm, based on the hybrid proximal extragradient (HPE) method [17], for solving standard form conic SDP problems is discussed in [10], and numerical results are presented showing that it generally outperforms the methods of [9, 20]. In [19], an efficient variant of the BP method is discussed and numerical results are presented showing its impressive ability to solve important classes of large-scale graph-related SDP problems. It should be observed though that the implementation in [19] is very specific in the sense that a different code is developed for each SDP class taking advantage of its special structure, without bringing the SDPs to standard form as in the approaches of [9, 20, 10].

Our goal in this paper is to study the performance of a BD method based on the BD-HPE framework in [11] for solving conic optimization problems, not necessarily in standard form, with two "easy" blocks of constraints. We will simply say that these problems have the "two-easy-block" structure. We first present a first-order BD method for minimizing the sum of a convex differentiable function with Lipschitz continuous gradient, and two other proper closed convex (possibly, nonsmooth) functions with easily computable resolvents. The method presented contains two important ingredients from a computational point of view, namely: an adaptive choice of stepsize for performing an extragradient step; and the use of a scaling factor to balance the blocks. We discuss its specialization to the context of conic SDP problems possessing the "two-easy-block" structure. Then, we apply it to solve four important classes of graph-related conic SDP problems which have the two-easy-block structure, namely: SDPs for θ-functions and θ_+-functions of graph stable set problems, and SDP relaxations of binary integer quadratic and frequency assignment problems. Finally, we present computational results on several instances of the aforementioned classes of conic SDPs showing that our method substantially outperforms the codes in [10, 19, 20]. Since the code in this paper works directly in the conic optimization problem as given, and hence works with a formulation with less number of variables, it is not surprising that it also outperforms the BD method of [10], which in contrast requires as input an SDP problem in standard form.

Our paper is organized as follows. Section 2 reviews some facts about the ε-subdifferential of a convex function and the ε-enlargement of a monotone operator. Section 3 presents an adaptive block-decomposition HPE (A-BD-HPE) framework in the context of block-structured monotone inclusion problems, similar to the one presented in [11], but with an adaptive choice of stepsize for performing the extragradient step. Section 4 presents a first-order instance of the A-BD-HPE framework, and corresponding complexity results, for solving a minimization problem whose objective function is the sum of a finite everywhere convex function with Lipschitz continuous gradient and two proper closed convex (possibly, nonsmooth) functions with easily computable resolvents. Section 5 discusses the specialization of the method of Section 4 to the context of conic optimization problems with a two-easy-block structure. Section 6 describes a generic stopping criterion and a practical variant of the BD method of Section 5 which incorporates a dynamic update of the scaling factor to balance the blocks. Section 7 presents numerical results comparing the latter variant of the BD method to the method discussed in [19] . Section 8 briefly compares this variant of the BD method with the methods in [10] and [20]. Finally, Section 9 presents some final remarks.

2 The ε-subdifferential and ε-enlargement of monotone operators

In this section, we review some properties of the ε-subdifferential of a convex function and the ε-enlargement of a monotone operator.

Let $\mathcal{Z}$ denote a finite dimensional inner product space with inner product and associated norm denoted by $\langle \cdot, \cdot \rangle_{\mathcal{Z}}$ and $\|\cdot\|_{\mathcal{Z}}$. A point-to-set operator $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$ is a relation $T \subseteq \mathcal{Z} \times \mathcal{Z}$ and

$$T(z) = \{v \in \mathcal{Z} \mid (z, v) \in T\}.$$

Alternatively, one can consider T as a multi-valued function of $\mathcal{Z}$ into the family $\wp(\mathcal{Z}) = 2^{(\mathcal{Z})}$ of subsets of $\mathcal{Z}$. Regardless of the approach, it is usual to identify T with its graph defined as

$$Gr(T) = \{(z, v) \in \mathcal{Z} \times \mathcal{Z} \mid v \in T(z)\}.$$

The domain of T, denoted by Dom$T$, is defined as

$$\text{Dom}T := \{z \in \mathcal{Z} : T(z) \neq \emptyset\}.$$

An operator $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$ is *affine* if its graph is an affine manifold. An operator $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$ is *monotone* if

$$\langle v - \tilde{v}, z - \tilde{z}\rangle_{\mathcal{Z}} \geq 0, \qquad \forall (z, v), (\tilde{z}, \tilde{v}) \in Gr(T),$$

and T is *maximal monotone* if it is monotone and maximal in the family of monotone operators with respect to the partial order of inclusion, i.e., $S : \mathcal{Z} \rightrightarrows \mathcal{Z}$ monotone and $Gr(S) \supset Gr(T)$ implies that $S = T$.

In [1], Burachik, Iusem and Svaiter introduced the ε-enlargement of maximal monotone operators. In [12] this concept was extended to a generic point-to-set operator in $\mathcal{Z}$ as follows. Given $T : \mathcal{Z} \rightrightarrows \mathcal{Z}$ and a scalar ε, define $T^\varepsilon : \mathcal{Z} \rightrightarrows \mathcal{Z}$ as

$$T^\varepsilon(z) = \{v \in \mathcal{Z} \mid \langle z - \tilde{z}, v - \tilde{v}\rangle_{\mathcal{Z}} \geq -\varepsilon, \quad \forall \tilde{z} \in \mathcal{Z}, \forall \tilde{v} \in T(\tilde{z})\}, \quad \forall z \in \mathcal{Z}. \tag{3}$$

We now state a few useful properties of the operator T^ε that will be needed in our presentation.

Proposition 2.1. *Let $T, T' : \mathcal{Z} \rightrightarrows \mathcal{Z}$. Then,*

a) if $\varepsilon_1 \leq \varepsilon_2$, then $T^{\varepsilon_1}(z) \subseteq T^{\varepsilon_2}(z)$ for every $z \in \mathcal{Z}$;

b) $T^\varepsilon(z) + (T')^{\varepsilon'}(z) \subseteq (T + T')^{\varepsilon+\varepsilon'}(z)$ for every $z \in \mathcal{Z}$ and $\varepsilon, \varepsilon' \in \mathbb{R}$;

c) T is monotone if, and only if, $T \subseteq T^0$;

d) T is maximal monotone if, and only if, $T = T^0$;

We refer the reader to [2, 18] for further discussion on the ε-enlargement of a maximal monotone operator.

For a scalar $\varepsilon \geq 0$, the ε-subdifferential of a function $f : \mathcal{Z} \to [-\infty, +\infty]$ is the operator $\partial_\varepsilon f : \mathcal{Z} \rightrightarrows \mathcal{Z}$ defined as

$$\partial_\varepsilon f(z) = \{v \mid f(\tilde{z}) \geq f(z) + \langle \tilde{z} - z, v\rangle_{\mathcal{Z}} - \varepsilon, \ \forall \tilde{z} \in \mathcal{Z}\}, \quad \forall z \in \mathcal{Z}. \tag{4}$$

When $\varepsilon = 0$, the operator $\partial_\varepsilon f$ is simply denoted by ∂f and is referred to as the subdifferential of f. The operator ∂f is trivially monotone if f is proper. If f is a proper lower semi-continuous convex function, then ∂f is maximal monotone [16].

The conjugate f^* of f is the function $f^* : \mathcal{Z} \to [-\infty, \infty]$ defined as

$$f^*(v) = \sup_{z \in \mathcal{Z}} \langle v, z\rangle_{\mathcal{Z}} - f(z), \quad \forall v \in \mathcal{Z}.$$

The following result lists some useful properties about the ε-subdifferential of a proper convex function.

Proposition 2.2. *Let $f : \mathcal{Z} \to (-\infty, \infty]$ be a proper convex function. Then,*

a) $\partial_\varepsilon f(z) \subseteq (\partial f)^\varepsilon(z)$ for any $\varepsilon \geq 0$ and $z \in \mathcal{Z}$;

b) if f is closed, then $\partial_\varepsilon(f^*) = (\partial_\varepsilon f)^{-1}$ for any $\varepsilon \geq 0$;

c) if $v \in \partial f(z)$ and $f(\tilde{z}) < \infty$, then $v \in \partial_\varepsilon f(\tilde{z})$, where $\varepsilon := f(\tilde{z}) - [f(z) + \langle \tilde{z} - z, v\rangle]$.

The *indicator function* of a closed convex set $Z \subseteq \mathcal{Z}$ is the function $\delta_Z : \mathcal{Z} \to [0, \infty]$ defined as

$$\delta_Z(z) = \begin{cases} 0, & z \in Z, \\ \infty, & \text{otherwise.} \end{cases}$$

For a closed convex cone $\mathcal{K} \subseteq \mathcal{Z}$, we have the following characterization about the conjugate and ε-subdiferential of $\delta_{\mathcal{K}}$.

Proposition 2.3. *Let $\mathcal{K} \subseteq \mathcal{Z}$ be a (nonempty) closed convex cone. Then, the following statements hold:*

a) $(\delta_{\mathcal{K}})^* = \delta_{-\mathcal{K}^*}$, *where $\mathcal{K}^*$ is dual cone of $\mathcal{K}$ defined as*

$$\mathcal{K}^* := \{w \in \mathcal{Z} : \langle z, x\rangle \geq 0, \, \forall x \in \mathcal{K}\};$$

b) for any $\varepsilon \geq 0$, the pair $(z, w) \in \mathcal{Z} \times \mathcal{Z}$ satisfies $w \in -\partial_\varepsilon \delta_{\mathcal{K}}(z)$ if, and only if, $z \in \mathcal{K}$, $w \in \mathcal{K}^*$ and $\langle z, w\rangle_{\mathcal{Z}} \leq \varepsilon$.

We finish the section by stating the weak transportation formula for the ε-subdifferential whose proof can be found for example in Lemma 3.4 of [18].

Proposition 2.4. *Suppose that $f : \mathcal{Z} \rightrightarrows [-\infty, \infty]$ is a closed proper convex function. Let $z_i, v_i \in \mathcal{Z}$ and $\varepsilon_i, \alpha_i \in \mathbb{R}_+$, for $i = 1, \ldots, k$, be such that*

$$v_i \in \partial_{\varepsilon_i} f(z_i), \quad i = 1, \ldots, k, \qquad \sum_{i=1}^{k} \alpha_i = 1,$$

and define

$$z^a := \textstyle\sum_{i=1}^k \alpha_i z_i, \qquad v^a := \sum_{i=1}^k \alpha_i v_i,$$

$$\varepsilon^a := \textstyle\sum_{i=1}^k \alpha_i \left[\varepsilon_i + \langle z_i - z^a, v_i - v^a\rangle_{\mathcal{Z}}\right] = \sum_{i=1}^k \alpha_i \left[\varepsilon_i + \langle z_i - z^a, v_i\rangle_{\mathcal{Z}}\right].$$

Then, $\varepsilon^a \geq 0$ and $v^a \in \partial_{\varepsilon^a} f(z^a)$.

3 The A-BD-HPE framework

In this section, we review the A-BD-HPE framework with adaptive stepsize for solving a special type of monotone inclusion problem consisting of the sum of a continuous monotone map and a point-to-set maximal monotone operator with a separable two-block-structure.

Let $\mathcal{U}$ and $\mathcal{V}$ be finite dimensional inner product spaces with associated inner products denoted by $\langle \cdot, \cdot \rangle_{\mathcal{U}}$ and $\langle \cdot, \cdot \rangle_{\mathcal{V}}$, respectively, and associated norms denoted by $\| \cdot \|_{\mathcal{U}}$ and $\| \cdot \|_{\mathcal{V}}$, respectively. We endow the product space $\mathcal{U} \times \mathcal{V}$ with the canonical inner product $\langle \cdot, \cdot \rangle_{\mathcal{U},\mathcal{V}}$ and associated canonical norm $\| \cdot \|_{\mathcal{U},\mathcal{V}}$ defined as

$$\langle (u, v), (u', v')\rangle_{\mathcal{U},\mathcal{V}} := \langle u, u'\rangle_{\mathcal{U}} + \langle v, v'\rangle_{\mathcal{V}}, \quad \|(u, v)\|_{\mathcal{U},\mathcal{V}} := \sqrt{\langle (u, v), (u, v)\rangle_{\mathcal{U},\mathcal{V}}}, \tag{5}$$

for all $(u, v), (u', v') \in \mathcal{U} \times \mathcal{V}$.

Our problem of interest in this section is the monotone inclusion problem of finding $(u, v) \in \mathcal{U} \times \mathcal{V}$ such that

$$(0, 0) \in [F + H_1 \otimes H_2](u, v), \tag{6}$$

where

$$F(u, v) = (F_1(u, v), F_2(u, v)) \in \mathcal{U} \times \mathcal{V}, \qquad (H_1 \otimes H_2)(u, v) = H_1(u) \times H_2(v) \subseteq \mathcal{U} \times \mathcal{V}$$

and the following conditions are assumed:

A.1) $H_1 : \mathcal{U} \rightrightarrows \mathcal{U}$ and $H_2 : \mathcal{V} \rightrightarrows \mathcal{V}$ are maximal monotone;

A.2) $F : \text{Dom}\, F \subseteq \mathcal{U} \times \mathcal{V} \to \mathcal{U} \times \mathcal{V}$ is a continuous map such that $\text{Dom}F \supset \text{cl}(\text{Dom}\, H_1) \times \mathcal{V}$;

A.3) F is monotone on $\text{cl}(\text{Dom}\, H_1) \times \text{cl}(\text{Dom}\, H_2)$;

A.4) there exists $L_{uv} > 0$ such that

$$\|F_1(u, v') - F_1(u, v)\|_{\mathcal{U}} \le L_{uv}\|v' - v\|_{\mathcal{V}}, \quad \forall u \in \text{Dom}\, H_1, \;\; \forall v, v' \in \mathcal{V}. \tag{7}$$

It is trivial to check that $H_1 \otimes H_2$ is maximal monotone. Moreover, in view of the proof of Proposition A.1 of [13], it follows that $F + H_1 \otimes H_2$ is maximal monotone. Note that problem (6) is equivalent to

$$0 \in F_1(u, v) + H_1(u), \quad 0 \in F_2(u, v) + H_2(v),$$

We now state the A-BD-HPE framework.

Adaptive block-decomposition HPE (A-BD-HPE) framework:

0) Let $(u_0, v_0) \in \mathcal{U} \times \mathcal{V}$, $\sigma \in [0, 1]$, $\sigma_u, \sigma_v \in [0, 1)$ and $\tilde{\sigma}_u \in [0, \sigma_u]$ be given and set $k = 1$;

1) choose $\tilde{\lambda}_k > 0$ such that

$$\sigma_k := \left\{ \max\, \text{eig} \left(\begin{bmatrix} \sigma_u^2 & \tilde{\lambda}_k \tilde{\sigma}_u L_{uv} \\ \tilde{\lambda}_k \tilde{\sigma}_u L_{uv} & \sigma_v^2 + \tilde{\lambda}_k^2 L_{uv}^2 \end{bmatrix} \right) \right\}^{1/2} \le \sigma, \tag{8}$$

where max eig stands for the maximum eigenvalue;

2) compute $\tilde{u}_k, a_k \in \mathcal{U}$ and $\varepsilon'_k \ge 0$ such that

$$a_k \in H_1^{\varepsilon'_k}(\tilde{u}_k), \quad \|\tilde{\lambda}_k[F_1(\tilde{u}_k, v_{k-1}) + a_k] + \tilde{u}_k - u_{k-1}\|_{\mathcal{U}}^2 + 2\tilde{\lambda}_k \varepsilon'_k \le \sigma_u^2 \|\tilde{u}_k - u_{k-1}\|_{\mathcal{U}}^2, \tag{9}$$

$$\|\tilde{\lambda}_k[F_1(\tilde{u}_k, v_{k-1}) + a_k] + \tilde{u}_k - u_{k-1}\|_{\mathcal{U}} \le \tilde{\sigma}_u \|\tilde{u}_k - u_{k-1}\|_{\mathcal{U}}; \tag{10}$$

3) compute $\tilde{v}_k, b_k \in \mathcal{V}$ and $\varepsilon''_k \ge 0$ such that

$$b_k \in H_2^{\varepsilon''_k}(\tilde{v}_k), \quad \|\tilde{\lambda}_k[F_2(\tilde{u}_k, \tilde{v}_k) + b_k] + \tilde{v}_k - v_{k-1}\|_{\mathcal{V}}^2 + 2\tilde{\lambda}_k \varepsilon''_k \le \sigma_v^2 \|\tilde{v}_k - v_{k-1}\|_{\mathcal{V}}^2; \tag{11}$$

4) choose λ_k as the largest $\lambda > 0$ such that

$$\left\| \lambda \begin{pmatrix} F_1(\tilde{u}_k, \tilde{v}_k) + a_k \\ F_2(\tilde{u}_k, \tilde{v}_k) + b_k \end{pmatrix} + \begin{pmatrix} \tilde{u}_k \\ \tilde{v}_k \end{pmatrix} - \begin{pmatrix} u_{k-1} \\ v_{k-1} \end{pmatrix} \right\|_{\mathcal{U},\mathcal{V}}^2 + 2\lambda(\varepsilon'_k + \varepsilon''_k) \le \sigma^2 \left\| \begin{pmatrix} \tilde{u}_k \\ \tilde{v}_k \end{pmatrix} - \begin{pmatrix} u_{k-1} \\ v_{k-1} \end{pmatrix} \right\|_{\mathcal{U},\mathcal{V}}^2; \tag{12}$$

5) set

$$(u_k, v_k) = (u_{k-1}, v_{k-1}) - \lambda_k[F(\tilde{u}_k, \tilde{v}_k) + (a_k, b_k)], \tag{13}$$

$k \leftarrow k + 1$, and go to step 1.

end

The following result is proved in Proposition 3.1 of [11] (see also Proposition 3.1 of [10]).

Proposition 3.1. *Consider the sequences $\{\lambda_k\}$ and $\{\tilde{\lambda}_k\}$ generated by the A-BD-HPE framework. Then, for every $k \in \mathbb{N}$, $\lambda = \tilde{\lambda}_k$ satisfies* (12). *As a consequence $\lambda_k \ge \tilde{\lambda}_k$.*

We now review two convergence results (see Theorems 3.2 and 3.3 of [10]) for the A-BD-HPE framework. The first one, referred to as the pointwise convergence result, is about the sequence $\{(\tilde{u}_k, \tilde{v}_k)\}$, while the second one, referred to as the ergodic convergence result, is about an ergodic sequence obtained by averaging $\{(\tilde{u}_k, \tilde{v}_k)\}$ using the sequence of stepsizes $\{\lambda_k\}$ as weights.

Theorem 3.2. *Assume that $\sigma < 1$ and consider the sequences $\{(\tilde{u}_k, \tilde{v}_k)\}, \{(a_k, b_k)\}$, $\{\lambda_k\}$ and $\{(\varepsilon'_k, \varepsilon''_k\}$ generated by the A-BD-HPE framework and let d_0 denote the distance of the initial point $(u_0, v_0) \in \mathcal{U} \times \mathcal{V}$ to the solution set of (6). Then, for every $k \in \mathbb{N}$, there exists $i \le k$ such that*

$$\|F(\tilde{u}_i, \tilde{v}_i) + (a_i, b_i)\|_{\mathcal{U},\mathcal{V}} \le d_0 \sqrt{\frac{1+\sigma}{1-\sigma}\left(\frac{1}{\lambda_i \sum_{j=1}^k \lambda_j}\right)}, \quad \varepsilon'_i + \varepsilon''_i \le \frac{\sigma^2 d_0^2}{2(1-\sigma^2)\sum_{j=1}^k \lambda_j}. \tag{14}$$

We now state the ergodic convergence result, which is specialized to the the case where the map F is affine.

Theorem 3.3. *In addition to conditions A.1-A.4, assume that F is an affine map. Let d_0 denote the distance of the initial point $(u_0, v_0) \in \mathcal{U} \times \mathcal{V}$ to the solution set of (6). Consider the sequences $\{(\tilde{u}_k, \tilde{v}_k)\}, \{(a_k, b_k)\}$, $\{\lambda_k\}$ and $\{(\varepsilon'_k, \varepsilon''_k)\}$ generated by the A-BD-HPE framework and define for every $k \in \mathbb{N}$:*

$$\Lambda_k := \sum_{i=1}^k \lambda_i, \quad (\tilde{u}_k^a, \tilde{v}_k^a) = \frac{1}{\Lambda_k}\sum_{i=1}^k \lambda_i (\tilde{u}_i, \tilde{v}_i), \quad (a_k^a, b_k^a) = \frac{1}{\Lambda_k}\sum_{i=1}^k \lambda_i (a_k, b_k) \tag{15}$$

and

$$\begin{aligned} \varepsilon_{1,k}^a &:= \frac{1}{\Lambda_k}\sum_{i=1}^k \lambda_i \left(\varepsilon'_i + \langle \tilde{u}_i - \tilde{u}_k^a, a_i\rangle_{\mathcal{U}}\right) \ge 0, \\ \varepsilon_{2,k}^a &:= \frac{1}{\Lambda_k}\sum_{i=1}^k \lambda_i \left(\varepsilon''_i + \langle \tilde{v}_i - \tilde{v}_k^a, b_i\rangle_{\mathcal{V}}\right) \ge 0. \end{aligned}$$

Then, for every $k \in \mathbb{N}$, $(a_k^a, b_k^a) \in H_1^{\varepsilon_{1,k}^a}(\tilde{u}_k^a) \times H_2^{\varepsilon_{2,k}^a}(\tilde{v}_k^a)$ and

$$\|F(\tilde{u}_k^a, \tilde{v}_k^a) + (a_k^a, b_k^a)\|_{\mathcal{U},\mathcal{V}} \le \frac{2d_0}{\Lambda_k}, \qquad \varepsilon_{1,k}^a + \varepsilon_{2,k}^a \le \frac{2d_0^2}{\Lambda_k}(1+\eta),$$

where

$$\eta := \frac{2\sqrt{2}\sigma}{1-\max\{\sigma_u, \sigma_v\}}\left(1 + \frac{1}{(1-\sigma_v)^2}\right)^{1/2}.$$

4 A BD algorithm for a class of structured convex optimization

This section presents a first-order BD algorithm, and corresponding complexity results, for solving a minimization problem whose objective function is the sum of a finite everywhere convex function with Lipschitz continuous gradient and two proper closed convex (possibly, nonsmooth) functions with easily computable resolvents.

Throughout this section, $\mathcal{X}$ denotes a finite dimensional inner product space with corresponding inner product and norm denoted by $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$, respectively. We are concerned with the optimization problem

$$\begin{array}{ll} \min & f(x) + h_1(x) + h_2(x) \\ \text{s.t.} & x \in \mathcal{X}, \end{array} \tag{16}$$

where:

B.1) $f, h_1, h_2 : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ are convex lower semicontinuous proper functions;

B.2) f is differentiable on $\mathcal{X}$ and its gradient is L-Lipschitz continuous, that is,

$$\|\nabla f(x) - \nabla f(x')\| \le L\,\|x - x'\|, \qquad \forall x, x' \in \mathcal{X};$$

B.3) the intersection of the relative interiors of the effective domains of h_1 and h_2 is non-empty.

In view of the above assumptions and [15, , Theorem 23.8], we have $\partial(f + h_1 + h_2) = \nabla f + \partial h_1 + \partial h_2$. Therefore, x^* is an optimal solution of (16) if, and only if,

$$0 \in \nabla f(x^*) + \partial h_1(x^*) + \partial h_2(x^*). \tag{17}$$

Using Proposition 2.2(b), it then follows that x^* is an optimal solution of (16) if, and only if, there exists $y^* \in \mathbb{R}^n$ such that

$$0 \in \nabla f(x^*) + \partial h_1(x^*) + y^*, \qquad 0 \in \partial h_2^*(y^*) - x^*.$$

It is interesting to note that the above inclusion problem is associated with the Lagrangian $\mathcal{L} : \mathcal{X} \times \mathcal{X} \to [-\infty, \infty]$ defined as

$$\mathcal{L}(x, y) = f(x) + h_1(x) + \langle x, y\rangle - h_2^*(y), \quad \forall (x, y) \in \mathcal{X} \times \mathcal{X},$$

in that it can be simply expressed as

$$0 \in \partial_x \mathcal{L}(x, y), \quad 0 \in \partial_y(-\mathcal{L})(x, y), \tag{18}$$

where the two partial derivatives are with respect to the same inner product $\langle\cdot,\cdot\rangle$ on $\mathcal{X}$. Although one can apply the A-BD-HPE framework directly to the above system with $H_1 = \partial(f + h_1)$ and $H_2 = \partial h_2^*$, and $F(x, y) = (y, -x)$ for all $(x, y) \in \mathcal{X} \times \mathcal{X}$, it is more efficient from a computational point of view to introduce a scale factor to balance the two inclusions in (18).

Indeed, let $\theta > 0$ be given and consider the scaled inner product $\langle\cdot,\cdot\rangle_\theta$ in $\mathcal{X}$ defined as

$$\langle x, x'\rangle_\theta := \theta^{-1}\langle x, x'\rangle, \qquad \forall x, x' \in \mathcal{X}, \tag{19}$$

and observe that the associated inner product norm, denoted by $\|\cdot\|_\theta$, satisfies

$$\|\cdot\|_\theta = \frac{1}{\sqrt{\theta}}\|\cdot\|. \tag{20}$$

Also, denote the gradient and ε-subdifferential of an arbitrary function $\phi : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$ with respect to $\langle\cdot,\cdot\rangle_\theta$ by $\nabla^\theta \phi$ and $\partial_\varepsilon^\theta \phi$, respectively. It is trivial to see that

$$\nabla^\theta \phi = \theta(\nabla \phi), \qquad \partial_\varepsilon^\theta \phi = \theta(\partial_\varepsilon \phi). \tag{21}$$

It turns out that the monotone inclusion problem (18) is equivalent to

$$0 \in \partial_x^\theta \mathcal{L}(x, y), \quad 0 \in \partial_y(-\mathcal{L}(x, y)), \tag{22}$$

or equivalently,

$$\begin{aligned} 0 &\in \theta\big(\nabla f(x) + \partial h_1(x) + y\big), \\ 0 &\in \partial h_2^*(y) - x. \end{aligned} \tag{23}$$

We note that the use of (18), or more generally (23), as a way of solving (17) is well known (see for example the methods described in [8, 4, 11]).

The above system is determined by $\mathcal{L}$ and the inner product norm on $\mathcal{X} \times \mathcal{X}$ defined as

$$\|(x, y)\|_{\theta,1} = \sqrt{\|x\|_\theta^2 + \|y\|^2}, \quad \forall (x, y) \in \mathcal{X} \times \mathcal{X}. \tag{24}$$

Note that this norm is the one given by (5) when $\mathcal{U} = \mathcal{V} = \mathcal{X}$ and $\|\cdot\|_\mathcal{U} = \|\cdot\|_\theta$ and $\|\cdot\|_\mathcal{V} = \|\cdot\|$. Note also that the inclusion system (23) is a special case of the monotone inclusion problem (6) with

$$\mathcal{U} = \mathcal{X}, \quad \langle\cdot,\cdot\rangle_\mathcal{U} = \langle\cdot,\cdot\rangle_\theta, \qquad \mathcal{V} = \mathcal{X}, \quad \langle\cdot,\cdot\rangle_\mathcal{V} = \langle\cdot,\cdot\rangle, \tag{25}$$

and $F : \mathcal{U} \times \mathcal{V} \to \mathcal{U} \times \mathcal{V}$, $H_1 : \mathcal{U} \rightrightarrows \mathcal{U}$, and $H_2 : \mathcal{V} \rightrightarrows \mathcal{V}$ defined as

$$F(x, y) := (\theta y, -x), \quad H_1(x) := \partial^\theta(f + h_1)(x) = \theta(\nabla f(x) + \partial h_1(x)), \quad H_2(y) := \partial h_2^*(y). \tag{26}$$

The following simple result summarizes the main properties of the scaled reformulation (23) (or equivalently, (22)) of (17).

Proposition 4.1. *The inner product spaces $\mathcal{U}$ and $\mathcal{V}$ defined by* (25)*, and the map F and operators H_1 and H_2 defined by* (26)*, satisfy conditions A.1-A.4 with $L_{uv} = \sqrt{\theta}$. Moreover, the inclusion problem* (23) *is equivalent to the inclusion problem*

$$(0,0) \in [F + H_1 \otimes H_2](x, y).$$

Our goal now will be to state an instance of the A-BD-HPE framework for solving (23) under the assumption that the resolvents of both ∂h_1 and ∂h_2, that is, the maps $(I + \lambda\partial h_i)^{-1}$ for every $\lambda > 0$ and $i = 1, 2$, can be computed exactly. In other words, we assume that minimization subproblems of the form

$$\min_{x\in\mathcal{X}} h_i(x) + \frac{1}{2\lambda}\|x - x_0\|^2$$

can be exactly solved for any $x_0 \in \mathcal{X}$, $\lambda > 0$ and $i = 1, 2$.

We now state the aforementioned instance of A-BD-HPE framework for solving (23).

Algorithm 1 : Scaled A-BD-HPE method for (16)

---

0) Let $(x_0, y_0) \in \mathcal{X} \times U$, $\theta > 0$, $\sigma_1 \in (0, 1)$ and $\sigma \in [\sigma_1, 1]$ be given, and set $k = 1$ and

$$\tilde{\lambda} := \min\left\{\frac{\sigma_1^2}{\theta L}, \frac{\sigma}{\sqrt{\theta}}\right\}; \tag{27}$$

1) set $\tilde{x}_k := \left(I + \tilde{\lambda}\theta\partial h_1\right)^{-1}\left(x_{k-1} - \tilde{\lambda}\theta\left(\nabla f(x_{k-1}) + y_{k-1}\right)\right)$;

2) set $\tilde{y}_k := (I + \tilde{\lambda}\partial h_2^*)^{-1}(y_{k-1} + \tilde{\lambda}\tilde{x}_k)$;

3) choose λ_k to be the largest $\lambda > 0$ such that

$$\|\lambda(\theta r_{1,k}, r_{2,k}) + (\tilde{x}_k, \tilde{y}_k) - (x_{k-1}, y_{k-1})\|^2_{\theta,1} + 2\lambda_k\varepsilon_k \le \sigma^2\|(\tilde{x}_k, \tilde{y}_k) - (x_{k-1}, y_{k-1})\|^2_{\theta,1}, \tag{28}$$

where

$$r_{1,k} := \frac{1}{\theta\tilde{\lambda}}\left(x_{k-1} - \tilde{x}_k\right) + \left(\tilde{y}_k - y_{k-1}\right), \quad r_{2,k} := \frac{1}{\tilde{\lambda}}(y_{k-1} - \tilde{y}_k), \quad \varepsilon_k := \frac{L}{2}\|\tilde{x}_k - x_{k-1}\|^2; \tag{29}$$

4) set $(x_k, y_k) = (x_{k-1}, y_{k-1}) - \lambda_k(\theta r_{1,k}, r_{2,k})$ and $k \leftarrow k + 1$, and go to step 1.

We now make a remark about Algorithm 1. Note that the formula for computing $\tilde{y}_k$ in step 2 of Algorithm 1 involves the resolvent of ∂h_2^*, instead of ∂h_2. Using Moreau's formula below and Proposition 2.2(b) with $f = h_2$ and $\varepsilon = 0$, it can be easily seen that $\tilde{y}_k$ can also be computed as

$$\tilde{y}_k = y_{k-1} + \tilde{\lambda}\tilde{x}_k - \tilde{\lambda}\left(I + \tilde{\lambda}^{-1}\partial h_2\right)^{-1}\left(\tilde{\lambda}^{-1}y_{k-1} + \tilde{x}_k\right). \tag{30}$$

Clearly, depending on the function h_2, one of these resolvents might be easier to compute than the other one, and hence is the better one for computing $\tilde{y}_k$. Using Moreau's formula below, it is also possible to give an expression for computing $\tilde{x}_k$ in terms of the resolvent of ∂h_1^*. Again, which one to use computationally will depend on the function h_1. We have chosen the formulae in steps 1 and 2 of Algorithm 1 due to their symmetry and the fact that they are more convenient for our theoretical presentation.

We now state Moreau's identity in the following result.

Lemma 4.2. *[Moreau's identity; see Lemma 6.3 in [11]] Let $\lambda > 0$, $a \in \mathbb{R}^n$ and $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a point to set maximal monotone operator. Then,*

$$a = (I + \lambda A)^{-1}(a) + \lambda\left(I + \lambda^{-1}A^{-1}\right)^{-1}\left(\lambda^{-1}a\right).$$

The following result shows that Algorithm 1 is a special instance of the A-BD-HPE framework applied to (22).

Lemma 4.3. *Let* $\sigma_u := \sigma_1$*, and* $\tilde{\sigma}_u := 0$ *and* $\sigma_v := 0$ *and let inner products* $\langle \cdot, \cdot \rangle_{\mathcal{U}}$ *and* $\langle \cdot, \cdot \rangle_{\mathcal{V}}$ *and operators* F*,* H_1 *and* H_2 *be defined according to* (25) *and* (26)*. Consider the sequences* $\{(x_k, y_k)\}$*,* $\{(\tilde{x}_k, \tilde{y}_k)\}$*,* $\{(r_{1,k}, r_{2,k})\}$ *and* $\{\varepsilon_k\}$ *generated by Algorithm 1 and, for every* $k \in \mathbb{N}$*, define* $\tilde{\lambda}_k := \tilde{\lambda}$*,*

$$u_k = x_k, \quad \tilde{u}_k := \tilde{x}_k, \quad a_k := \frac{x_{k-1} - \tilde{x}_k}{\tilde{\lambda}} - \theta y_{k-1}, \quad \varepsilon'_k := \varepsilon_k \tag{31}$$

and

$$v_k = y_k, \quad \tilde{v}_k := \tilde{y}_k, \quad b_k := \frac{y_{k-1} - \tilde{y}_k}{\tilde{\lambda}} + \tilde{x}_k, \quad \varepsilon''_k := 0. \tag{32}$$

Then, the following statements hold for every $k \in \mathbb{N}$*:*

a) $\tilde{\lambda}_k$ *satisfies* (8)*;*

b) $\tilde{\lambda}_k$*,* u_{k-1} *and the triple* $(\tilde{u}_k, a_k, \varepsilon'_k)$ *satisfies* (9) *and* (10) *and*

$$\theta^{-1} a_k = \nabla f(x_{k-1}) + \partial h_1(\tilde{x}_k) \in (\partial_{\varepsilon'_k} f + \partial h_1)(\tilde{x}_k); \tag{33}$$

c) $\tilde{\lambda}_k$*,* v_{k-1} *and the triple* $(\tilde{v}_k, b_k, \varepsilon''_k)$ *satisfies* (11)*, and*

$$b_k \in \partial h_2^*(\tilde{y}_k); \tag{34}$$

d) $(\theta r_{1,k}, r_{2,k}) = F(\tilde{x}_k, \tilde{y}_k) + (a_k, b_k)$*.*

Proof. Statement a) follows immediately from condition (27), the definition of σ_u, $\tilde{\sigma}_u$, σ_v and $\tilde{\lambda}_k$ in the statement of the lemma, and the fact that $L_{uv} = \sqrt{\theta}$ and $\sigma_1 \le \sigma$, in view of Proposition 4.1.

Now, it follows from (31), (32) and the definition of F in (26) that

$$\tilde{\lambda}_k[F_1(\tilde{u}_k, v_{k-1}) + a_k] + \tilde{u}_k - u_{k-1} = \tilde{\lambda}_k[\theta v_{k-1} + a_k] + \tilde{u}_k - u_{k-1} = \tilde{\lambda}[\theta y_{k-1} + a_k] + \tilde{x}_k - x_{k-1} = 0$$

and

$$\tilde{\lambda}_k[F_2(\tilde{u}_k, \tilde{v}_k) + b_k] + \tilde{v}_k - v_{k-1} = \tilde{\lambda}_k[-\tilde{u}_k + b_k] + \tilde{v}_k - v_{k-1} = \tilde{\lambda}[-\tilde{x}_k + b_k] + \tilde{y}_k - y_{k-1} = 0.$$

Clearly, the first identity and the fact that $\tilde{\sigma}_u = 0$ imply that $\tilde{\lambda}_k$, u_{k-1} and $(\tilde{u}_k, a_k, \varepsilon'_k)$ satisfy (10), and also the inequality in (9), due to the fact that the definition of ε'_k, ε_k, σ_u and $\tilde{\lambda}_k$, and relations (20) and (27), imply that

$$2\tilde{\lambda}_k \varepsilon'_k = 2\tilde{\lambda}\varepsilon_k = L\tilde{\lambda}\|\tilde{x}_k - x_{k-1}\|^2 \le \frac{\sigma_1^2}{\theta}\|\tilde{x}_k - x_{k-1}\|^2 = \sigma_u^2 \|\tilde{x}_k - x_{k-1}\|_\theta^2.$$

Moreover, the second identity and the fact that $\varepsilon''_k = 0$ and $\sigma_v = 0$ imply that $\tilde{\lambda}_k$, v_{k-1} and $(\tilde{v}_k, b_k, \varepsilon''_k)$ satisfy the inequality in (11). We will now show that the inclusions in (9) and (11) hold. Indeed, Assumption B.2 easily implies that

$$f(\tilde{x}_k) - f(x_{k-1}) - \langle \nabla f(x_{k-1}), \tilde{x}_k - x_{k-1} \rangle \le \frac{L}{2}\|\tilde{x}_k - x_{k-1}\|^2 = \varepsilon_k = \varepsilon'_k,$$

where the last two equalities follow from the definition of ε'_k and ε_k. Using the last conclusion, the fact that $\nabla f(x_{k-1}) \in \partial f(x_{k-1})$, Lemma 2.2(c) with $v = \nabla f(x_{k-1})$, $z = x_{k-1}$ and $\tilde{z} = \tilde{x}_k$, we then conclude that $\nabla f(x_{k-1}) \in \partial_{\varepsilon'_k} f(\tilde{x}_k)$. Now, using the definition of $\tilde{x}_k$ in step 1 of Algorithm 1, a_k in (31), and H_2 in (26), the last conclusion, relations (21) and (26), and Proposition 2.2(a), we conclude that

$$\begin{aligned} a_k &\in \theta[\nabla f(x_{k-1}) + \partial h_1(\tilde{x}_k)] \in \theta(\partial_{\varepsilon'_k} f + \partial h_1)(\tilde{x}_k) \subseteq \theta\left[\partial_{\varepsilon'_k}(f + h_1)(\tilde{x}_k)\right] \\ &= \partial^\theta_{\varepsilon'_k}(f + h_1)(\tilde{x}_k) \subseteq [\partial^\theta(f + h_1)]^{\varepsilon'_k}(\tilde{x}_k) = (H_1)^{\varepsilon'_k}(\tilde{x}_k), \end{aligned}$$

which shows that (33) and the inclusion in (9) hold. Also, the definition of $\tilde{y}_k$ in step 2 of Algorithm 1, b_k in (32), and H_1 in (26), the fact that $\varepsilon''_k = 0$, and Proposition 2.1(d), imply that

$$b_k \in \partial h_2^*(\tilde{y}_k) = H_2(\tilde{y}_k) = H_2^{\varepsilon''_k}(\tilde{y}_k),$$

which shows that (34) and the inclusion in (11) hold. We have thus shown statements b) and c).

Statement d) follow immediately from the definition of F, a_k and b_k in (26), (31) and (32), respectively, and the definition of $r_{1,k}$ and $r_{2,k}$ in (29). □

It follows from Lemma 4.3 that Algorithm 1 is a special instance of the A-BD-HPE framework. Hence, the convergence results described in Theorems 3.2 and 3.3 apply to it. In what follows, we will describe the implications of these two results towards the behavior of Algorithm 1.

However, we first make some observations regarding the distance of the initial point (x_0, y_0) to the solution set of (22) with respect to the norm $\|(\cdot,\cdot)\|_{\theta,1}$. First observe that the solution sets of (18) and (22) are the same. Second, by the saddle-point theory, this set is of the form $X^* \times Y^* \subseteq \mathcal{X} \times \mathcal{X}$. Third, the distance d_0^θ of the initial point (x_0, y_0) to the solution set of (22) with respect to the norm $\|(\cdot,\cdot)\|_{\theta,1}$ can be expressed as

$$d_0^\theta := \sqrt{\theta^{-1} d_{1,0}^2 + d_{2,0}^2}, \tag{35}$$

where

$$d_{1,0} := \min\{\|x - x_0\| : x \in X^*\}, \quad d_{2,0} := \min\{\|y - y_0\| : y \in Y^*\}.$$

Theorem 4.4. *Consider the sequences $\{(x_k, y_k)\}$, $\{(\tilde{x}_k, \tilde{y}_k)\}$, $\{(r_{1,k}, r_{2,k})\}$ and $\{\varepsilon_k\}$ generated by Algorithm 1 under the assumption that $\sigma < 1$. Then, for every $k \in \mathbb{N}$,*

$$r_{1,k} \in \nabla f(x_{k-1}) + \partial h_1(\tilde{x}_k) + \tilde{y}_k \in \partial_{\varepsilon_k} f(\tilde{x}_k) + \partial h_1(\tilde{x}_k) + \tilde{y}_k, \tag{36}$$

$$r_{2,k} \in \partial h_2^*(\tilde{y}_k) - \tilde{x}_k, \tag{37}$$

and there exists $i \le k$ such that

$$\sqrt{\theta\|r_{1,i}\|^2 + \|r_{2,i}\|^2} \le \max\left\{\frac{1}{\sigma}, \frac{L\sqrt{\theta}}{\sigma_1^2}\right\}\left(\frac{\sqrt{\theta}}{\sqrt{k}}\right)\sqrt{\left(\frac{1+\sigma}{1-\sigma}\right)\left(\theta^{-1}d_{1,0}^2 + d_{2,0}^2\right)},$$

$$\varepsilon_i \le \max\left\{\frac{1}{\sigma}, \frac{L\sqrt{\theta}}{\sigma_1^2}\right\}\left(\frac{\sqrt{\theta}}{k}\right)\frac{\sigma^2(\theta^{-1}d_{1,0}^2 + d_{2,0}^2)}{2(1-\sigma^2)}.$$

Proof. Consider the sequences $\{a_k\}$ and $\{b_k\}$ defined in (31) and (32), respectively. It follows from the definition of $r_{1,k}$ and $r_{2,k}$ in (29) that

$$r_{1,k} = \tilde{y}_k + \theta^{-1}a_k = \theta^{-1}[F_1(\tilde{x}_k, \tilde{y}_k) + a_k], \qquad r_{2,k} = -\tilde{x}_k + b_k = F_2(\tilde{x}_k, \tilde{y}_k) + b_k. \tag{38}$$

Now, (36) and (37) follow from the above two identities and relations (33) and (34). Moreover, the above two identities together with Theorem 3.2 imply the existence of $i \le k$ such that

$$\|(\theta r_{1,i}, r_{2,i})\|_{\theta,1} \le d_0\sqrt{\frac{1+\sigma}{1-\sigma}\left(\frac{1}{\lambda_i\sum_{j=1}^k \lambda_j}\right)} \le \sqrt{\frac{1+\sigma}{1-\sigma}}\frac{d_0}{\tilde{\lambda}\sqrt{k}}$$

and

$$\varepsilon_i \le \frac{\sigma^2 d_0^2}{2(1-\sigma^2)\sum_{j=1}^k \lambda_j} \le \frac{\sigma^2 d_0^2}{2(1-\sigma^2)\tilde{\lambda}k},$$

where the second and fourth inequality follow from Proposition 3.1 and the fact that $\tilde{\lambda}_k = \tilde{\lambda}$ for all $k \in \mathbb{N}$. Using the definition of $\|\cdot\|_\theta$ and $\|(\cdot,\cdot)\|_{\theta,1}$ in (20) and (24), respectively, identity (35), and the fact that (27) implies that

$$\tilde{\lambda}^{-1} = \sqrt{\theta}\max\left\{\frac{1}{\sigma}, \frac{\sqrt{\theta}L}{\sigma_1^2}\right\}, \tag{39}$$

we easily see that the last two estimates imply the two bounds in the conclusion of the theorem. □

Theorem 4.5. *Consider the sequences $\{(x_k, y_k)\}$, $\{\tilde{x}_k, \tilde{y}_k)\}$ and $\{\varepsilon_k\}$ generated by Algorithm 1 and the sequences of residuals $\{r_{1,k}\}$ and $\{r_{1,k}\}$ defined in* (29)*. For every $k \in \mathbb{N}$, define*

$$\Lambda_k = \sum_{i=1}^k \lambda_i, \quad (\tilde{x}_k^a, \tilde{y}_k^a) = \Lambda_k^{-1}\sum_{i=1}^k \lambda_i(\tilde{x}_i, \tilde{y}_i), \quad (r_{1,k}^a, r_{2,k}^a) := \Lambda_k^{-1}\sum_{i=1}^k \lambda_i(r_{1,k}, r_{2,k}) \tag{40}$$

and

$$\varepsilon_{1,k}^a := \Lambda_k^{-1}\sum_{i=1}^k \lambda_i[\varepsilon_k + \langle r_{1,i} - \tilde{y}_i, \tilde{x}_i - \tilde{x}_k^a\rangle], \quad \varepsilon_{2,k}^a := \Lambda_k^{-1}\sum_{i=1}^k \lambda_i\langle r_{2,i} + \tilde{x}_i, \tilde{y}_i - \tilde{y}_k^a\rangle, \quad \varepsilon_k^a := \varepsilon_{1,k}^a + \varepsilon_{2,k}^a. \tag{41}$$

Then, for every $k \in \mathbb{N}$,

$$\begin{aligned}(r_{1,k}^a, r_{2,k}^a) &\in \left[\partial_{\varepsilon_{1,k}^a}\left(f + h_1 + \langle \tilde{y}_k^a, \cdot\rangle\right)(\tilde{x}_k^a)\right] \times \left[\partial_{\varepsilon_{2,k}^a}\left(h_2^* - \langle \tilde{x}_k^a, \cdot\rangle\right)(\tilde{y}_k^a)\right] \\ &\subseteq \partial_{\varepsilon_k^a}[\mathcal{L}(\cdot, \tilde{y}_k^a) - \mathcal{L}(\tilde{x}_k^a, \cdot)](\tilde{x}_k^a, \tilde{y}_k^a)\end{aligned} \tag{42}$$

and

$$\sqrt{\theta\|r_{1,k}^a\|^2 + \|r_{2,k}^a\|^2} \leq \max\left\{\frac{1}{\sigma}, \frac{\sqrt{\theta}L}{\sigma_1^2}\right\}\left(\frac{2\sqrt{\theta}}{k}\right)\sqrt{\theta^{-1}d_{1,0}^2 + d_{2,0}^2}, \tag{43}$$

$$\varepsilon_k^a \leq \max\left\{1, \frac{\sqrt{\theta}L\sigma}{\sigma_1^2}\right\}\left[\frac{8\sqrt{\theta}}{(1-\sigma_1)k}\right]\left(\theta^{-1}d_{1,0}^2 + d_{2,0}^2\right). \tag{44}$$

Proof. Consider the sequences $\{a_k\}$ and $\{b_k\}$ defined in (31) and (32), respectively, and the sequences $\{(a_k^a, b_k^a)\}$ defined in (15). Note that by (38) and the definition of $\langle\cdot,\cdot\rangle_\theta$, we have

$$\varepsilon_{1,k}^a := \Lambda_k^{-1}\sum_{i=1}^k \lambda_i[\varepsilon_k + \langle a_i, \tilde{x}_i - \tilde{x}_k^a\rangle_\theta], \quad \varepsilon_{2,k}^a := \Lambda_k^{-1}\sum_{i=1}^k \lambda_i\langle b_i, \tilde{y}_i - \tilde{y}_k^a\rangle.$$

Hence, it follows from Proposition 3.1, Theorem 3.3, and relations (26) and (38), that

$$\|(\theta r_{1,k}^a, r_{2,k}^a)\|_{\theta,1} = \|(\theta\tilde{y}_k, -\tilde{x}_k) + (\tilde{a}_k^a, \tilde{b}_k^a)\|_{\theta,1} = \|F(\tilde{x}_k^a, \tilde{y}_k^a) + (\tilde{a}_k^a, \tilde{b}_k^a)\|_{\theta,1} \leq 2\frac{d_0^\theta}{\Lambda_k} \leq 2\frac{d_0^\theta}{k\tilde{\lambda}},$$

and

$$\varepsilon_k^a = \varepsilon_{1,k}^a + \varepsilon_{2,k}^a \leq \left(\frac{8\sigma}{1-\sigma_1}\right)\frac{(d_0^\theta)^2}{\Lambda_k} \leq \left(\frac{8\sigma}{1-\sigma_1}\right)\frac{(d_0^\theta)^2}{k\tilde{\lambda}}$$

Using the definition of $\|(\cdot,\cdot)\|_{\theta,1}$, identities (35) and (39), we easily see that the above two inequalities imply (43) and (44). Now, (36), (37), (41) and Proposition 2.4 imply that

$$r_{1,k}^a \in \partial_{\varepsilon_{1,k}^a}(f + h_1)(\tilde{x}_k^a) + \tilde{y}_k^a, \qquad r_{2,k}^a \in \partial_{\varepsilon_{2,k}^a}(h_2^*)(\tilde{y}_k^a) - \tilde{x}_k^a.$$

and hence that

$$r_{1,k}^a \in (\partial_{x,\varepsilon_{1,k}^a}\mathcal{L})(\tilde{x}_k^a, \tilde{y}_k^a), \qquad r_{2,k}^a \in (\partial_y\mathcal{L})(\tilde{x}_k^a, \tilde{y}_k^a).$$

These inclusions are easily seen to imply (42). □

5 Specialization of Algorithm 1 to conic optimization

In this section, we discuss the specialization of Algorithm 1 to the context of conic optimization problems possessing the two-easy-block structure.

More specifically, let $\mathcal{X}$ be as in Section 4 and, for $i = 1, 2$, let $\mathcal{W}_i$ be an inner product space whose inner product and associated norm is denoted by $\langle \cdot, \cdot \rangle_{\mathcal{W}_i}$ and $\|\cdot\|_{\mathcal{W}_i}$. We consider the conic optimization problem of the form

$$\begin{aligned} z_P^* := \min \ & \langle c, x \rangle \\ \text{s.t.} \ & \mathcal{A}_1 x - b_1 \in \mathcal{K}_1 \\ & \mathcal{A}_2 x - b_2 \in \mathcal{K}_2, \end{aligned} \tag{45}$$

where $c \in \mathcal{X}$, $b_1 \in \mathcal{W}_1$, $b_2 \in \mathcal{W}_2$, $\mathcal{A}_1 : \mathcal{X} \to \mathcal{W}_1$ and $\mathcal{A}_2 : \mathcal{X} \to \mathcal{W}_2$ are linear maps, and $\mathcal{K}_1 \subseteq \mathcal{W}_1$ and $\mathcal{K}_2 \subseteq \mathcal{W}_2$ are nonempty closed convex cones. Observe that (45) is a special of (16) in which

$$f(\cdot) = \langle c, \cdot \rangle, \qquad h_i(\cdot) = \delta_{\mathcal{M}_i}(\cdot) = \delta_{\mathcal{K}_i}(\mathcal{A}_i(\cdot) - b_i), \quad i = 1, 2, \tag{46}$$

and

$$\mathcal{M}_i := \{x \in \mathcal{X} : \mathcal{A}_i x - b_i \in \mathcal{K}_i\}, \quad i = 1, 2. \tag{47}$$

Throughout this section, we make the following assumptions on (45):

C.1) (45) has an optimal solution, and hence $z_P^* \in \mathbb{R}$;

C.2) (45) has a Slater point, i.e., there exists $x \in \mathcal{X}$ such that $\mathcal{A}_i x - b_i \in \operatorname{ri} \mathcal{K}_i$ for $i = 1, 2$;

C.3) the resolvent of $h_i = \delta_{\mathcal{M}_i}$, or equivalently, the projection onto $\mathcal{M}_i$, is easy to evaluate for $i = 1, 2$.

The dual of (45) is the conic optimization problem given by

$$\begin{aligned} z_D^* := \max \ & \langle b_1, w_1 \rangle_{\mathcal{W}_1} + \langle b_2, w_2 \rangle_{\mathcal{W}_2} \\ \text{s.t.} \ & \mathcal{A}_1^* w_1 + \mathcal{A}_2^* w_2 = c \\ & w_1 \in \mathcal{K}_1^*, \ w_2 \in \mathcal{K}_2^*, \end{aligned} \tag{48}$$

where $\mathcal{A}_i^*$ is the adjoint of $\mathcal{A}_i$ and $\mathcal{K}_i^*$ is the dual cone of $\mathcal{K}_i$, $i = 1, 2$. It is well-known that assumptions C.1 and C.2 imply that the dual of (45) has an optimal solution and that $z_P^* = z_D^*$.

Although not apparent in the discussion of Section 4, it will be shown in Theorem 5.2 below that Algorithm 1 applied to (45) generates a dual sequence $\{(w_{1,k}, w_{2,k})\}$ which solves the dual problem (48) in the limit. We start by stating the following result which shows that assumption C.3, i.e., the ability to compute the projection onto $\mathcal{M}_i$, provides meaningful dual information.

Lemma 5.1. *Let $h_i : U \to [-\infty, \infty]$ be defined as in (46), and assume that $\mathcal{A}_i : \mathcal{X} \to \mathcal{W}_i$ is such that $\mathcal{A}_i(\mathcal{X}) \cap \operatorname{ri} \mathcal{K}_i \neq \emptyset$. Then, the following statements hold:*

a) for any $\varepsilon \geq 0$, $(x, y_i) \in \mathcal{X} \times \mathcal{X}$ satisfies $y_i \in \partial_\varepsilon h_i(x)$ if, and only if, $x \in \mathcal{M}_i$ and there exists $w_i \in \mathcal{W}_i$ such that

$$w_i \in \mathcal{K}_i^*, \quad \langle w_i, \mathcal{A}_i x - b_i \rangle_{\mathcal{W}_i} \leq \varepsilon, \quad \mathcal{A}_i^* w_i = -y_i;$$

b) for any $\lambda > 0$, $(x, p_i) \in \mathcal{X} \times \mathcal{X}$ satisfies $x = (I + \lambda \partial h_i)^{-1} p_i$ if, and only if, $x \in M_i$ and there exists $w_i \in \mathcal{W}_i$ such that

$$w_i \in \mathcal{K}_i^*, \quad \langle w_i, \mathcal{A}_i x - b_i \rangle_{\mathcal{W}_i} = 0, \quad \mathcal{A}_i^* w_i = \frac{x - p_i}{\lambda}. \tag{49}$$

Proof. We first prove a). Since $h_i = \delta_{\mathcal{K}_i}(\mathcal{A}_i(\cdot) - b_i)$, it follows from a well-known property of the subdifferential and the assumption $\mathcal{A}_i(\mathcal{X}) \cap \operatorname{ri} \mathcal{K}_i \neq \emptyset$ that $\partial h_i(x) = \mathcal{A}_i^* \partial \delta_{\mathcal{K}_i}(\mathcal{A}_i x - b_i)$. Hence, $y_i \in \partial h_i(x)$ if, and only if, there exists $w_i \in -\partial \delta_{\mathcal{K}_i}(\mathcal{A}_i x - b_i)$ such that $\mathcal{A}_i^* w_i = -y_i$. The lemma now follows Proposition 2.3(b). Statement b) follows trivially from a) with $\varepsilon = 0$. $\square$

We observe that the condition $\mathcal{A}_i(\mathcal{X}) \cap \operatorname{ri} \mathcal{K}_i \neq \emptyset$ in the statement of Lemma 5.1 is implied by Assumption C.2.

The following result shows precisely how Algorithm 1 is solving the dual problem (48) in the limit.

Theorem 5.2. *Consider the sequences $\{(\tilde{x}_k, \tilde{y}_k)\}$ and $\{(x_k, y_k)\}$ generated by Algorithm 1 with f, h_1 and h_2 given by* (46) *and under the assumption that $\sigma < 1$. Then, for every $k \in \mathbb{N}$, the following statements hold:*

a) $\mathcal{A}_1\tilde{x}_k - b_1 \in \mathcal{K}_1$, $\mathcal{A}_2(\tilde{x}_k + r_{2,k}) - b_2 \in \mathcal{K}_2$ and there exists $(w_{1,k}, w_{2,k}) \in \mathcal{K}_1^* \times \mathcal{K}_2^*$ such that

$$c - \mathcal{A}_1^* w_{1,k} - \mathcal{A}_2^* w_{2,k} = r_{1,k}, \quad \mathcal{A}_2^* w_{2,k} = -\tilde{y}_k,$$

$$\langle w_{1,k}, \mathcal{A}_1\tilde{x}_k - b_1\rangle_{\mathcal{W}_1} = 0, \quad \langle w_{2,k}, \mathcal{A}_2(\tilde{x}_k + r_{2,k}) - b_2\rangle_{\mathcal{W}_2} = 0;$$

b) the duality gap $dg_k := \langle c, \tilde{x}_k\rangle - (\langle b_1, w_{1,k}\rangle_{\mathcal{W}_1} + \langle b_2, w_{2,k}\rangle_{\mathcal{W}_2})$ can be alternatively computed as

$$dg_k = \langle r_{1,k}, \tilde{x}_k\rangle + \langle r_{2,k}, \tilde{y}_k\rangle;$$

c) there exists $i \leq k$ such that

$$\max\left\{\sqrt{\theta}\|r_{1,i}\|\,,\, \|r_{2,i}\|\right\} \leq \frac{\sqrt{\theta}}{\sigma\sqrt{k}}\sqrt{\left(\frac{1+\sigma}{1-\sigma}\right)(\theta^{-1}d_{1,0}^2 + d_{2,0}^2)}.$$

Proof. In view of step 1 of Algorithm 1 and Lemma 5.1(b) with $i = 1$, $\lambda = \tilde{\lambda}\theta$, $x = \tilde{x}_k$ and $p_1 = x_{k-1} - \tilde{\lambda}\theta(c + y_{k-1})$, we conclude that $\mathcal{A}_1\tilde{x}_k - b_1 \in \mathcal{K}_1$ and there exists $w_{1,k} \in \mathcal{K}_1^*$ such that

$$\langle w_{1,k}, \mathcal{A}_1\tilde{x}_k - b_1\rangle_{\mathcal{W}_1} = 0, \quad \mathcal{A}_1^* w_{1,k} = \frac{\tilde{x}_k - [x_{k-1} - \lambda\theta(c + y_{k-1})]}{\tilde{\lambda}\theta} = c - r_{1,k} + \tilde{y}_k,$$

where the last equality follows from (29). Moreover, it follows from (30) and (29) that

$$\tilde{x}_k + r_{2,k} \in (I + \tilde{\lambda}^{-1}\partial h_2)^{-1}(\tilde{\lambda}^{-1}y_{k-1} + \tilde{x}_k).$$

Hence, it follows from Lemma 5.1(b) with $i = 2$, $\lambda = \tilde{\lambda}^{-1}$, $x = \tilde{x}_k + r_{2,k}$ and $p_2 = \tilde{\lambda}^{-1}y_{k-1} + \tilde{x}_k$ that $\mathcal{A}_2(\tilde{x}_k + r_{2,k}) - b_2 \in \mathcal{K}_2$ and there exists $w_{2,k} \in \mathcal{K}_2^*$ such that

$$\langle w_{2,k}, \mathcal{A}_2(\tilde{x}_k + r_{2,k}) - b_2\rangle_{\mathcal{W}_2} = 0, \quad \mathcal{A}_2^* w_{2,k} = \tilde{\lambda} r_{2,k} - y_{k-1} = -\tilde{y}_k,$$

where the last equality is due to (29). We have thus shown a) and b). Statement c) follows from Theorem 4.4 and the fact that now $L = 0$. □

We now make some observations about Theorem 5.2. First, although Theorem 5.2 shows how to generate a dual sequence $\{(w_{1,k}, w_{2,k})\}$ which solves the dual problem (48) in the limit, it is important to note that Algorithm 1 applied to (45) can be implemented without ever generating such a sequence. Second, Theorem 5.2(a) shows that $\tilde{x}_k$ and its perturbation $\tilde{x}_k + r_{2,k}$ exactly satisfy the first and second blocks $\mathcal{A}_1 x - b_1 \in \mathcal{K}_1$ and $\mathcal{A}_2 x - b_2 \in \mathcal{K}_2$, respectively. Third, in regards to the first observation, statements a) and b) of Theorem 5.2 show that the quantities $c - \mathcal{A}_1^* w_{1,k} - \mathcal{A}_2^* w_{2,k}$ and $\langle c, \tilde{x}_k\rangle - (\langle b_1, w_{1,k}\rangle_{\mathcal{W}_1} + \langle b_2, w_{2,k}\rangle_{\mathcal{W}_2})$, commonly used in stopping criteria presented in the literature, can be computed in terms of $\tilde{x}_k$ and $\tilde{y}_k$, and hence their computation do not require $(w_{1,k}, w_{2,k})$. In view of the latter observation, we can check whether an iterate of Algorithm 1 applied to (45) satisfies the usual stopping criteria in terms of the duality gap measure and violations to the constraints in (45) and (48) without the need of generating $(w_{1,k}, w_{2,k})$. Fourth, the inner products $\langle \cdot, \cdot\rangle_{\mathcal{W}_1}$ and $\langle \cdot, \cdot\rangle_{\mathcal{W}_2}$ play no role in the actual implementation of Algorithm 1 applied to (45). In fact, they are only used to construct the dual problem (48) and, if necessary, the dual sequence $\{(w_{1,k}, w_{2,k})\}$. Fifth, Theorem 5.2(c) sheds light on how the scaling parameter θ might affect the sizes of the residuals $r_{1,k}$

and $r_{2,k}$. Roughly speaking, viewing all quantities in the bound of Theorem 5.2(c), with the exception of θ, as constants, we see that

$$\|r_{1,k}\| = \mathcal{O}\left(\max\left\{1, \theta^{-1/2}\right\}\right), \qquad \|r_{2,k}\| = \mathcal{O}\left(\max\left\{1, \theta^{1/2}\right\}\right).$$

Hence, the ratio $\|r_{2,k}\|/\|r_{1,k}\|$ can grow significantly as $\theta \to \infty$, while it can become very small as $\theta \to 0$. This suggests that this ratio increases (resp., decreases) as θ increases (resp., decreases). In fact, we have observed in our computational experiments that this ratio behaves just as described.

6 A practical dynamically scaled BD method

In this section, we describe three measures that quantify the optimality of an approximate solution of (45), namely: the primal infeasibility measure; the dual infeasibility measure; and the relative duality gap. We also describe two important refinements of Algorithm 1 for solving (45), whose goal is to balance the magnitudes of the primal and dual infeasibility measures. More specifically, we describe: i) a scheme for choosing the initial scaling parameter θ; and ii) a procedure for dynamically updating the scaling parameter θ to balance the sizes of the primal and dual infeasibility measures as the algorithm progresses.

Let $\mathcal{X}$ be as in Section 4. For the purpose of describing a generic stopping criterion for Algorithm 1, let $\|\cdot\|'_{\mathcal{W}_i}$ be a given norm in the inner product space $\mathcal{W}_i$ and define the distance $d_i(\cdot)$ as

$$d_i(w) = \min\left\{\|w - \tilde{w}\|'_{\mathcal{W}_i} : \tilde{w} \in \mathcal{K}_i\right\} \qquad \forall w \in \mathcal{W}_i.$$

We can define the primal infeasibility measure as

$$\epsilon_P(x) := \sqrt{d_1\left(\mathcal{A}_1 x - b_1\right)^2 + d_2\left(\mathcal{A}_2 x - b_2\right)^2}, \qquad \forall x \in \mathcal{X}. \tag{50}$$

Also, for a pre-specified scalar $\mu > 0$, define the dual infeasibility measure as

$$\epsilon_D(w_1, w_2) := \frac{1}{\mu}\left\|c - \mathcal{A}_1^* w_1 - \mathcal{A}_2^* w_2\right\|, \qquad \forall (w_1, w_2) \in \mathcal{W}_1 \times \mathcal{W}_2. \tag{51}$$

Note that, in view of (20), $\epsilon_D(w_1, w_2)$ is the magnitude of the dual residual $c - \mathcal{A}_1^* w_1 - \mathcal{A}_2^* w_2$ in terms of the norm $\|\cdot\|_\theta$, where $\theta = \mu^2$. Clearly, an arbitrary norm on $\mathcal{X}$ could be used in place of the latter norm to define $\epsilon_D(w_1, w_2)$, but this norm suffices for the sake of our discussion. Finally, define the relative duality gap as

$$\epsilon_G(x, w_1, w_2) := \frac{\langle c, x\rangle - (\langle b_1, w_1\rangle + \langle b_2, w_2\rangle)}{|\langle c, x\rangle| + |\langle b_1, w_1\rangle + \langle b_2, w_2\rangle| + 1}, \qquad \forall x \in \mathcal{X},\ \forall (w_1, w_2) \in \mathcal{W}_1 \times \mathcal{W}_2. \tag{52}$$

Observe that if $\epsilon_{P,k}$, $\epsilon_{D,k}$ and $\epsilon_{G,k}$ denote the values of (50), (51) and (52) evaluated at $x = \tilde{x}_k$ and $(w_1, w_2) = (w_{1,k}, w_{2,k})$, respectively, then, in view of Theorem 5.2, we have

$$\epsilon_{P,k} = d_2\left(\mathcal{A}_2 \tilde{x}_k - b_2\right) \le \left\|(\mathcal{A}_2 \tilde{x}_k - b_2) - (\mathcal{A}_2\left(\tilde{x}_k + r_{2,k}\right) - b_2)\right\|'_{\mathcal{W}_2} \le \left\|\mathcal{A}_2 r_{2,k}\right\|'_{\mathcal{W}_2}, \tag{53}$$

$$\epsilon_{D,k} = \frac{1}{\mu}\left\|r_{1,k}\right\|, \tag{54}$$

$$\epsilon_{G,k} = \frac{\langle r_{1,k}, \tilde{x}_k\rangle + \langle r_{2,k}, \tilde{y}_k\rangle}{|\langle c, \tilde{x}_k\rangle| + |\langle r_{1,k}, \tilde{x}_k\rangle + \langle r_{2,k}, \tilde{y}_k\rangle - \langle c, \tilde{x}_k\rangle| + 1}. \tag{55}$$

Given a tolerance $\bar{\epsilon} > 0$, a suitable stopping criterion for Algorithm 1 applied to (45) would be

$$\max\left\{\epsilon_{P,k}, \epsilon_{D,k}, \epsilon_{G,k}\right\} \le \bar{\epsilon}. \tag{56}$$

We now discuss two important refinements of Algorithm 1 for solving (45), whose goal is to balance the magnitudes of the primal and dual infeasibility measures $\epsilon_{P,k}$ and $\epsilon_{D,k}$. First note that (53) and (54) implies that $\epsilon_{P,k}/\epsilon_{D,k} = \mathcal{O}(\|r_{2,k}\|/\|r_{1,k}\|)$. Hence, in view of the last observation in the paragraph immediately after Theorem 5.2, the latter ratio can grow significantly as $\theta \to \infty$, while it can become very small as $\theta \to 0$. This suggests that this ratio increases (resp., decreases) as θ increases (resp., decreases). Indeed, our computational experiments indicate that the ratio $\epsilon_{P,k}/\epsilon_{D,k}$ behaves in this manner.

In the following, let θ_k denotes the dynamic value of θ at the kth iteration of Algorithm 1. Observe that, in view of (53), (54) and (29), the measures $\epsilon_{P,k}$ and $\epsilon_{D,k}$ depend on $\tilde{x}_k$ and $\tilde{y}_k$, whose values in turn depend on the choice of θ_k, in view of steps 1 and 2 of Algorithm 1. Hence, these two measures are indeed functions of θ, which are denoted as $\epsilon_{P,k}(\theta)$ and $\epsilon_{D,k}(\theta)$.

We first describe a scheme for choosing the initial scaling parameter θ_1. Let a constant $\rho > 1$ be given and set $\theta = 1$. If $\epsilon_{P,1}(\theta)/\epsilon_{D,1}(\theta) > \rho$ (resp., $\epsilon_{P,1}(\theta)/\epsilon_{D,1}(\theta) < \rho^{-1}$), we successively divide (resp., successively multiply) the current value of θ by 2 until $\epsilon_{P,1}(\theta)/\epsilon_{D,1}(\theta) \le \rho$ (resp., $\epsilon_{P,1}(\theta)/\epsilon_{D,1}(\theta) \ge \rho^{-1}$) is satisfied, and set $\theta_1 = \theta_1^*$, where θ_1^* is the last value of θ. Since there is no guarantee that this procedure will terminate, we specify an upper bound on the number of times that θ can be updated. In our implementation, we set this upper bound to be 20.

We next describe a procedure for dynamically updating the scaling parameter θ to balance the sizes of the two measures $\epsilon_{P,k}(\theta)$ and $\epsilon_{D,k}(\theta)$ as the algorithm progresses. As in [10], we use the heuristic of changing θ every time a specified number of iterations have been performed. More specifically, given an integer $\bar{k} \ge 1$, and scalars $\gamma > 1$ and $0 < \tau < 1$, we use the following dynamic scaling procedure for updating θ_k,

$$\theta_k = \begin{cases} \theta_{k-1}, & k \not\equiv 0 \bmod \bar{k} \text{ or } \gamma^{-1} \le \bar{\epsilon}_{P,k-1}/\bar{\epsilon}_{D,k-1} \le \gamma \\ \tau^2\theta_{k-1}, & k \equiv 0 \bmod \bar{k} \text{ and } \bar{\epsilon}_{P,k-1}/\bar{\epsilon}_{D,k-1} > \gamma \\ \tau^{-2}\theta_{k-1}, & k \equiv 0 \bmod \bar{k} \text{ and } \bar{\epsilon}_{P,k-1}/\bar{\epsilon}_{D,k-1} < \gamma^{-1} \end{cases}, \qquad \forall k \ge 2, \tag{57}$$

where

$$\bar{\epsilon}_{P,k-1} = \left(\prod_{i=k-\bar{k}}^{k-1} \epsilon_{P,i} \right)^{1/\bar{k}}, \qquad \bar{\epsilon}_{D,k-1} = \left(\prod_{i=k-\bar{k}}^{k-1} \epsilon_{D,i} \right)^{1/\bar{k}}, \qquad \forall k > \bar{k}. \tag{58}$$

Roughly speaking, the above dynamic scaling procedure changes the value of θ at most a single time in the right direction, so as to balance the sizes of the residuals based on the information provided by their values at the previous $\bar{k}$ iterations. We observe that a dynamic scaling procedure using $\epsilon_{P,k-1}$ and $\epsilon_{D,k-1}$ in place of $\bar{\epsilon}_{P,k-1}$ and $\bar{\epsilon}_{D,k-1}$ in (57), respectively, is proposed in [10]. However, the more conservative procedure based on the aggregated measures in (58) have performed better in our computational experiments.

In our computational experiments, we will refer to the variant of Algorithm 1 in which the two schemes described above are incorporated into as the *two-easy-block-decomposition HPE (2EBD-HPE)* method. To illustrate the importance of the above two schemes, we have chosen an instance of a conic optimization problem to compare the performance of the *2EBD-HPE* method against the performance of the *2EBD-HPE* method without incorporating exactly one of the schemes above. Figure 1 compares the performance of the *2EBD-HPE* method against its variant (VAR1) in which θ_1 is initialized as 1 instead of θ_1^*. Figure 2 compares the performance of the *2EBD-HPE* method against the its variant (VAR2) in which dynamic scaling is removed (i.e., θ_k set to θ_1^* , for every $k \ge 1$). In addition, in order to illustrate the importance of making an adaptive choice of stepsize in Algorithm 1, Figure 3 compares the performance of the *2EBD-HPE* method against the its variant (VAR3) in which the stepsize λ_k is chosen to be $\tilde{\lambda} = \frac{\sigma}{\sqrt{\theta_k}}$ for every $k \ge 1$. Figure 4 compares the performance of the *2EBD-HPE* method against the following three variants: i) (VAR2) the one that removes the dynamic scaling (i.e., set $\theta_k = \theta_1^*$, for every $k \ge 1$); ii) (VAR4) the one that removes the dynamic scaling and the initialization scheme for θ_1 (i.e., set $\theta_k = 1$, for every $k \ge 1$); and iii) (VAR5) the one that removes these latter two refinements and the use of adaptive stepsize (i.e., set $\theta_k = 1$ and $\lambda_k = \tilde{\lambda} = \sigma$, for every $k \ge 1$).

7 Numerical results: part I

In this section, we compare the 2EBD-HPE method described in Section 6 with a variant of the boundary point method, namely SDPAD, presented in [19]. More specifically, we compare these two methods on four

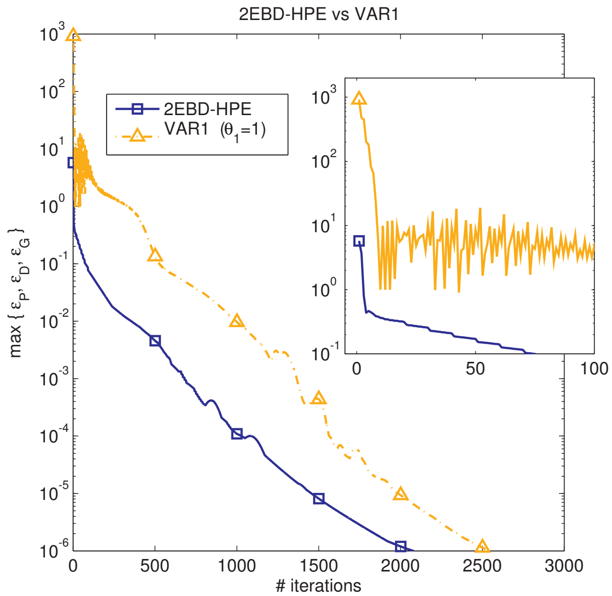

Figure 1: This example (BIQ-be200.8.8) illustrates how the scheme for choosing the initial scaling parameter θ_1 can help Algorithm 1 to start with an error at least 2 orders of magnitude smaller.

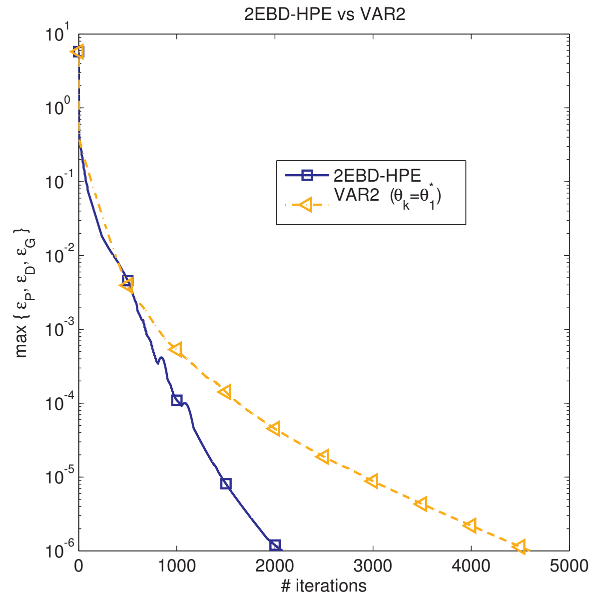

Figure 2: This example (BIQ-be200.8.8) illustrates how the dynamic scaling improves the performance of Algorithm 1 considerably.

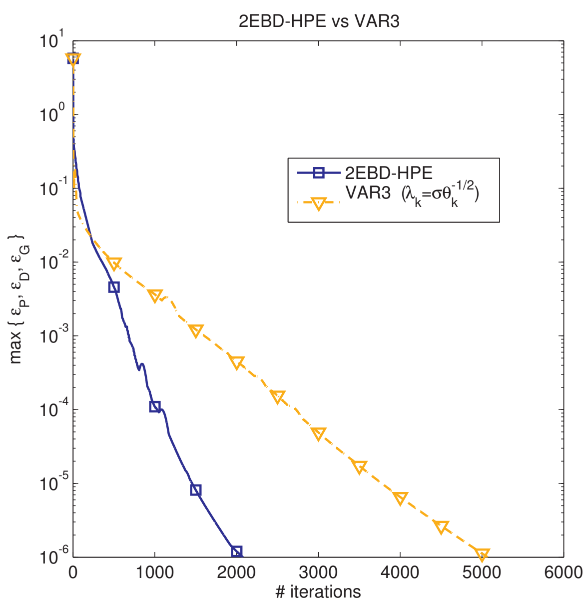

Figure 3: This example (BIQ-be200.8.8) illustrates how the adaptive stepsize improves the performance of Algorithm 1 considerably.

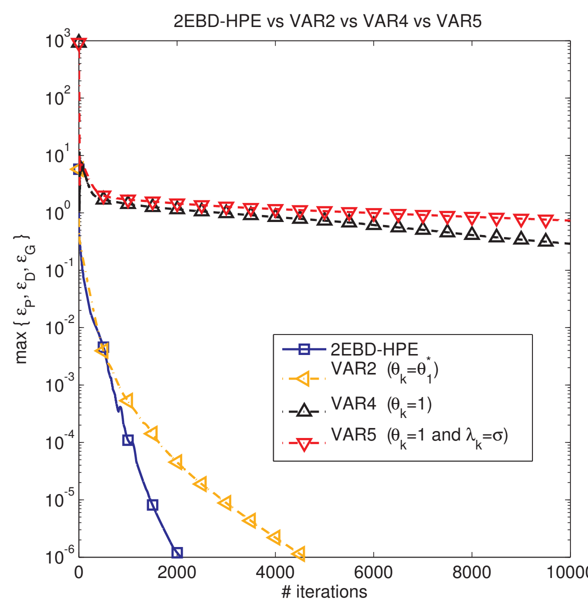

Figure 4: This example (BIQ-be200.8.8) illustrates how all the refinements made in the application of the BD-HPE framework to conic optimization helped improve the performance of the algorithm.

important classes of graph-related SDP problems, namely: SDP relaxations of binary integer quadratic (BIQ) and frequency assignment (FAP) problems, and SDPs for θ-functions and θ_+-functions of graph stable set problems. This section contains three subsections. The first subsection considers SDP relaxations of BIQ problems, the second one deals with SDP relaxations of FAPs, and the third one discusses SDPs corresponding to the θ-functions and θ_+-functions of graph stable set problems.

We have implemented the 2EBD-HPE method for solving (45) in a Matlab code which requires the user to provide their own projection subroutines onto the sets $\mathcal{M}_i$, $i = 1, 2$, defined in (47).

For the 2EBD-HPE method, the computational results for the SDP relaxations of BIQs and FAPs were obtained on a server with 2 Xeon X5460 processors at 3.16GHz and 32GB RAM, and the ones for the SDPs corresponding to the θ-functions and θ_+-functions of graph stable set problems were obtained on a single core of a server with 2 Xeon X5520 processors at 2.27GHz and 48GB RAM. On the other hand, for the SDPAD method we were able to obtain computational results for all instances of the above SDP classes only on a laptop with an Intel Core 2 Duo processor and 4GB RAM, since it was the only machine we had at our disposal which was compatible with this code. The use of two different machines is not an issue due to the following reasons: i) our benchmark is based solely on the total number of iterations; and ii) the work per iteration for the two methods are almost identical. Moreover, we should note that the stopping criteria used by our method for solving the different classes of conic SDP problems are the same as the ones used by SDPAD. The stopping criterion used for each one of the above classes of conic SDP problems will be described separately in each one of the subsections below. Even though we have our own reservations in regards to the stopping criteria used for the SDP relaxations of BIQ and the θ_+-function SDP problems in that they omit the amount of violation to the nonnegative orthant, we have decided to adopt them anyways for the purpose of performing this benchmark. Thus, we have decided to preserve the integrity of SDPAD instead of modify its stopping criterion and possibly its dynamic update of the penalty parameter. The only change we have made on SDPAD is that we have eliminated the possibility of its stopping based on stagnation, i.e., little progress from one iteration to the next one (see Subsection 3.4 of [19]).

We now make some general remarks about how the results are reported on the tables given below. Tables 1, 2, 5, 7 and 9 compare our method against SDPAD. For every instance, we stop both methods when they have reached an accuracy of 10^{-6} (i.e., $\bar{\epsilon} = 10^{-6}$ in relation (56)) according to the stopping criteria used for its corresponding problem class. The number of iterations performed by any of the two methods for any particular instance is marked in red, and also with an asterisk (*), whenever it cannot solve the instance by the required accuracy, in which case the residual (i.e., the maximum between the infeasibility measures and the relative duality gap) reported is the one obtained at the last iteration of the method. Also, the number of iterations marked in blue in a row is the best one among the ones listed in that row with the convention that, when a method cannot solve the instance, the corresponding number of iterations is assumed to be ∞. Tables 3, 4, 6, 8 and 10 report more detailed computational results obtained by our method 2EBD-HPE.

Finally, Figures 5, 6, 7 and 8 plot the performance profiles (see [5]) of 2EBD-HPE and SDPAD methods for each of the four problem classes. We recall the following definition of a performance profile. For a given instance, a method A is said to be at most x times slower than method B, if the number of iterations performed by method A is at most x times the number of iterations performed by method B. A point (x, y) is in the performance profile curve of a method if it can solve exactly $(100y)\%$ of all the tested instances x times slower than any other competing method.

7.1 Binary integer quadratic problems

This subsection compares the performance of our method 2EBD-HPE with that of SDPAD on a collection of SDP relaxations of BIQ problems.

The SDP relaxation of the BIQ problem can be described as follows (see for example Section 7 in [20]). Given an $n \times n$ symmetric matrix Q, the BIQ problem can be formulated as

$$\min \left\{ z^T Q z : z \in \{0,1\}^n \right\}.$$

By representing the binary set $\{0,1\}^n$ as $\left\{ z \in \mathbb{R}^n | z_i^2 - z_i = 0 \right\}$, we obtain the following SDP relaxation

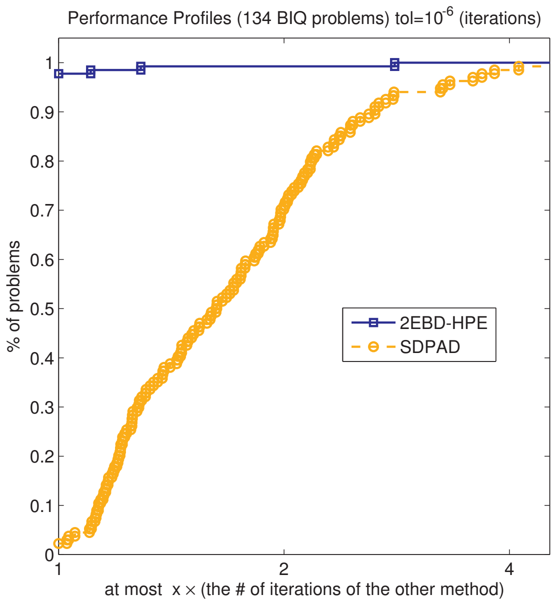

Figure 5: Performance profiles of 2EBD-HPE and SDPAD for solving 134 SDP relaxations of BIQ problems with accuracy $\bar{\epsilon} = 10^{-6}$.

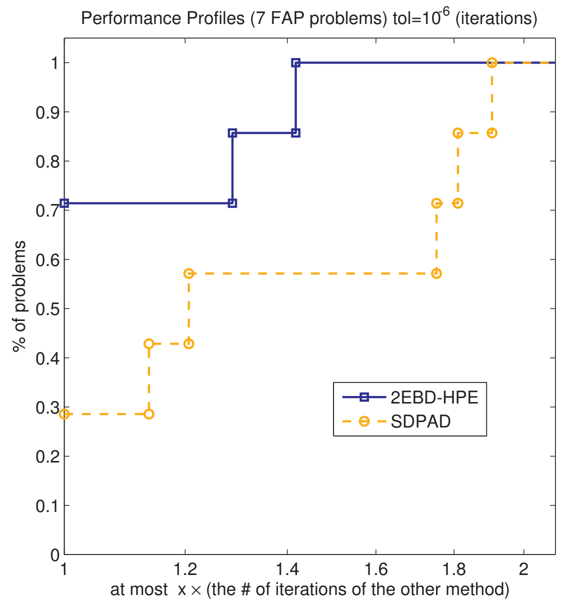

Figure 6: Performance profiles of 2EBD-HPE and SDPAD for solving 7 SDP relaxations of FAPs with accuracy $\bar{\epsilon} = 10^{-6}$.

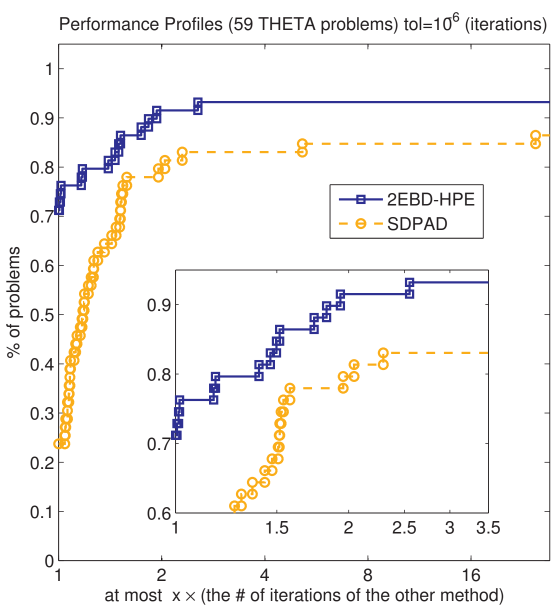

Figure 7: Performance profiles of 2EBD-HPE and SDPAD for solving 59 $\theta(G)$ problems with accuracy $\bar{\epsilon} = 10^{-6}$.

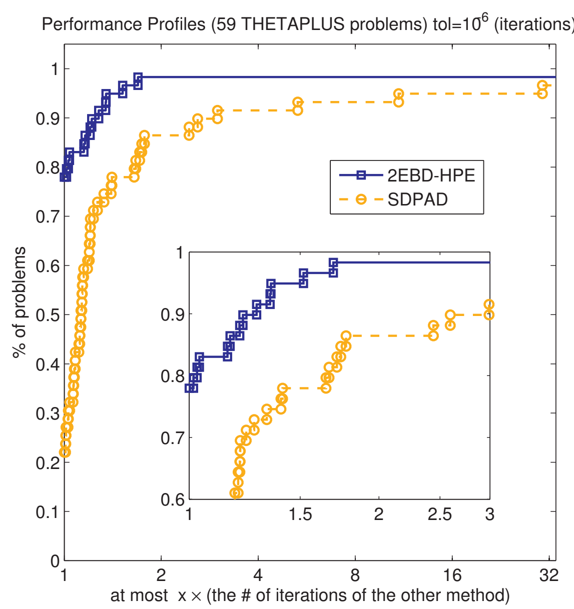

Figure 8: Performance profiles of 2EBD-HPE and SDPAD for solving 59 $\theta_+(G)$ problems with accuracy $\bar{\epsilon} = 10^{-6}$.

$$
\begin{aligned}
\min \quad & Q \bullet Z \\
\text{s.t.} \quad & x := \begin{bmatrix} Z & z \\ z^T & \alpha \end{bmatrix} \succeq 0, && \text{(59a)} \\
& \operatorname{diag}(Z) - z = 0,\ \alpha = 1,\ Z \geq 0,\ z \geq 0, && \text{(59b)}
\end{aligned}
$$

where $Z \in \mathcal{S}^n$, $z \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$.

There is more than one way of viewing (59) as a special case of the two-easy-block structure formulation (45). In our current implementation, we considered the following formulation. Let $\mathcal{X} = \mathcal{W}_1 := \mathcal{S}^{n+1}$, $\mathcal{W}_2 = \mathbb{R}^n \times \mathbb{R} \times \mathcal{S}^n \times \mathbb{R}^n$, $\mathcal{K}_1 = \mathcal{S}^{n+1}_+$ and $\mathcal{K}_2 = \mathbf{0}_n \times \mathbf{0}_1 \times \mathbb{R}^{n(n+1)/2}_+ \times \mathbb{R}^n_+$, where $\mathbf{0}_n$ denotes an n dimensional vector of all zeros. Also, endow $\mathcal{X}$ with the Frobenius inner product. With these definitions, we can easily see that we can view (59) as having the two-easy-block structure (45) if we let (59a) as $\mathcal{M}_1$ and (59b) as $\mathcal{M}_2$.

In order to agree with the scaling of the data and error measures adopted by SDPAD for (59), we measure the primal infeasibility measure ϵ_P as

$$
\epsilon_P(x) = \frac{\sqrt{\frac{2}{3}\left\|\operatorname{diag}(Z) - z\right\|_F^2 + (\alpha - 1)^2}}{2}, \tag{60}
$$

and the dual infeasibility measure ϵ_D as in (51) with $\mu = 1 + 2\left\|C\right\|_F$, where $\|\cdot\|_F$ is the Frobenius norm defined as

$$
\|A\|_F := \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2} \qquad \forall A \in \mathbb{R}^{m\times n}. \tag{61}
$$

Note that the above primal infeasibility measure is not a special case of (50) since it does not take into consideration the violations with respect to the constraints $Z \geq 0$, $z \geq 0$ and $x \succeq 0$. We observe also that, in view of the first inclusion in Theorem 5.2(a), the constraint $x \succeq 0$ is always satisfied by 2EBD-HPE, while SDPAD approaches it in the limit. Also, with respect to the other constraints $Z \geq 0$ and $z \geq 0$, both methods approach them only in the limit.

Tables 1 and 2 compare the two methods on a collection of 134 SDP relaxations of BIQ problems using the tolerance $\bar{\epsilon} = 10^{-6}$. For the purpose of this comparison, we considered 2EBD-HPE with $\sigma = 0.99$ and the values of γ, τ and $\bar{k}$ in the dynamic scaling rule (57) set to $\gamma = 1.5$, $\tau = 0.9$ and $\bar{k} = 10$.

Tables 3 and 4 give more detailed computational results obtained by our method 2EBD-HPE, such as the primal and dual objective function values, number of iterations, the primal and dual infeasibility measures as described above, and the relative duality gap. Since our implementation of 2EBD-HPE is based on ϵ_P in order to conform with the stopping criterion of SDPAD and, as observed above, ϵ_P does not take into consideration the violations with respect to the constraints $Z \geq 0$, $z \geq 0$, we also include in Tables 3 and 4 a column with the minimum value of all entries of x. Figure 5 plots the performance profiles of both methods.

Note that 2EBD-HPE solves 132 (out of a total of 134) problems faster than SDPAD. Moreover, 2EBD-HPE solves about 7 of them at least 3.5 times faster than SDPAD. Note also that 2EBD-HPE is able to solve all instances while SDPAD fails to solve one of them, namely `gka9b`.

7.2 Frequency assignment problems

This subsection compares the performance of our method 2EBD-HPE with that of SDPAD on a collection of SDP relaxations of FAPs.

The SDP relaxation of the FAP can be described as follows (see for example Subsection 2.4 in [3]). Given a network represented by a graph G with n nodes and an edge-weight matrix W, the frequency assignment

Table 1: Comparison of the methods on BIQ problems

Problem		$\max\{\epsilon_P, \epsilon_D, \epsilon_G\}$		Iterations	
Instance	$n_s \mid m$	**2EBD-HPE**	**SDPAD**	**2EBD-HPE**	**SDPAD**
be100.1	101|5252	9.70 -7	9.98 -7	1511	3200
be100.10	101|5252	9.63 -7	9.96 -7	1111	1394
be100.2	101|5252	9.69 -7	9.98 -7	1200	2945
be100.3	101|5252	9.91 -7	9.99 -7	1531	3630
be100.4	101|5252	9.99 -7	9.95 -7	1500	2879
be100.5	101|5252	9.98 -7	9.98 -7	1215	2508
be100.6	101|5252	9.95 -7	9.97 -7	1470	2578
be100.7	101|5252	9.94 -7	9.98 -7	1279	2619
be100.8	101|5252	9.98 -7	9.27 -7	1246	2027
be100.9	101|5252	9.93 -7	9.99 -7	1168	1615
be120.3.1	121|7502	9.96 -7	9.98 -7	1535	3038
be120.3.10	121|7502	9.97 -7	9.93 -7	1373	2080
be120.3.2	121|7502	9.99 -7	1.00 -6	1721	3491
be120.3.3	121|7502	9.95 -7	9.25 -7	1527	1765
be120.3.4	121|7502	9.90 -7	9.47 -7	1660	2287
be120.3.5	121|7502	9.99 -7	1.00 -6	2015	5094
be120.3.6	121|7502	9.99 -7	9.98 -7	2047	4457
be120.3.7	121|7502	9.99 -7	1.00 -6	3229	7395
be120.3.8	121|7502	9.99 -7	9.98 -7	2405	6362
be120.3.9	121|7502	9.96 -7	1.00 -6	2370	7724
be120.8.1	121|7502	9.92 -7	9.97 -7	1165	1754
be120.8.10	121|7502	9.97 -7	1.00 -6	1513	3254
be120.8.2	121|7502	1.00 -6	1.00 -6	2246	5942
be120.8.3	121|7502	9.94 -7	9.99 -7	1632	3276
be120.8.4	121|7502	1.00 -6	9.98 -7	1802	2843
be120.8.5	121|7502	9.97 -7	1.00 -6	1662	4076
be120.8.6	121|7502	9.98 -7	9.99 -7	1357	3161
be120.8.7	121|7502	9.97 -7	9.24 -7	1682	2252
be120.8.8	121|7502	9.99 -7	9.53 -7	1278	1429
be120.8.9	121|7502	9.91 -7	9.14 -7	1196	1371
be150.3.1	151|11627	9.98 -7	1.00 -6	2104	2894
be150.3.10	151|11627	9.99 -7	1.00 -6	2573	5580
be150.3.2	151|11627	9.96 -7	1.00 -6	2063	3175
be150.3.3	151|11627	9.95 -7	9.98 -7	1715	2240
be150.3.4	151|11627	9.98 -7	9.99 -7	2007	2885
be150.3.5	151|11627	9.99 -7	9.96 -7	1878	2259
be150.3.6	151|11627	9.93 -7	9.99 -7	1619	3384
be150.3.7	151|11627	9.98 -7	9.96 -7	1997	2905
be150.3.8	151|11627	9.97 -7	1.00 -6	2294	4535
be150.3.9	151|11627	9.96 -7	9.95 -7	1323	1753
be150.8.1	151|11627	9.99 -7	9.75 -7	1477	1856
be150.8.10	151|11627	9.97 -7	9.99 -7	1720	2947
be150.8.2	151|11627	9.98 -7	9.96 -7	1431	2451
be150.8.3	151|11627	1.00 -6	9.91 -7	1768	2266
be150.8.4	151|11627	9.97 -7	9.99 -7	1691	3564
be150.8.5	151|11627	9.96 -7	9.98 -7	1930	3131
be150.8.6	151|11627	9.99 -7	9.97 -7	1649	4075
be150.8.7	151|11627	1.00 -6	9.99 -7	2184	3819
be150.8.8	151|11627	9.99 -7	1.00 -6	2137	5844
be150.8.9	151|11627	9.99 -7	9.96 -7	2066	5793
be200.3.1	201|20502	1.00 -6	9.98 -7	2023	2383
be200.3.10	201|20502	9.99 -7	9.99 -7	2287	4532
be200.3.2	201|20502	9.98 -7	9.97 -7	2304	2803
be200.3.3	201|20502	9.97 -7	1.00 -6	3167	7002
be200.3.4	201|20502	9.98 -7	9.99 -7	2316	3574
be200.3.5	201|20502	9.98 -7	1.00 -6	2882	5825
be200.3.6	201|20502	9.99 -7	9.99 -7	2126	2443
be200.3.7	201|20502	9.98 -7	9.82 -7	2689	2827
be200.3.8	201|20502	9.97 -7	9.94 -7	2265	2323
be200.3.9	201|20502	9.98 -7	1.00 -6	3787	6942
be200.8.1	201|20502	9.98 -7	9.99 -7	2620	5078
be200.8.10	201|20502	9.97 -7	9.99 -7	2074	2639
be200.8.2	201|20502	1.00 -6	9.97 -7	1769	2264
be200.8.3	201|20502	9.99 -7	1.00 -6	2309	4468
be200.8.4	201|20502	1.00 -6	9.93 -7	2104	2436
be200.8.5	201|20502	1.00 -6	9.99 -7	1965	3648
be200.8.6	201|20502	9.78 -7	9.91 -7	2451	3071
be200.8.7	201|20502	9.97 -7	9.87 -7	2118	2506
be200.8.8	201|20502	9.99 -7	9.96 -7	2079	2420
be200.8.9	201|20502	9.97 -7	1.00 -6	2310	4097
be250.1	251|31877	9.98 -7	9.99 -7	3638	5923
be250.10	251|31877	9.99 -7	1.00 -6	4220	10957
be250.2	251|31877	1.00 -6	1.00 -6	3643	5404
be250.3	251|31877	9.97 -7	9.99 -7	3297	3647
be250.4	251|31877	1.00 -6	1.00 -6	5779	12310
be250.5	251|31877	9.99 -7	9.98 -7	3560	6326
be250.6	251|31877	9.99 -7	1.00 -6	3570	4319
be250.7	251|31877	9.98 -7	9.98 -7	3321	6136
be250.8	251|31877	9.99 -7	9.91 -7	3502	4292
be250.9	251|31877	9.97 -7	9.98 -7	3947	6531

Table 2: Comparison of the methods on BIQ problems

Problem		$\max\{\epsilon_P, \epsilon_D, \epsilon_G\}$		Iterations	
Instance	$n_s \| m$	**2EBD-HPE**	**SDPAD**	**2EBD-HPE**	**SDPAD**
bqp100-1	101|5252	9.98 -7	9.98 -7	1291	1667
bqp100-10	101|5252	9.99 -7	9.96 -7	1977	6585
bqp100-2	101|5252	9.98 -7	1.00 -6	2742	5004
bqp100-3	101|5252	1.00 -6	1.00 -6	3505	14426
bqp100-4	101|5252	9.98 -7	1.00 -6	2596	6059
bqp100-5	101|5252	1.00 -6	1.00 -6	2248	8264
bqp100-6	101|5252	9.97 -7	9.95 -7	1237	2393
bqp100-7	101|5252	9.95 -7	9.82 -7	1775	1831
bqp100-8	101|5252	9.94 -7	9.99 -7	2307	7462
bqp100-9	101|5252	9.99 -7	9.99 -7	2156	5770
bqp250-1	251|31877	1.00 -6	9.99 -7	3484	5572
bqp250-10	251|31877	1.00 -6	1.00 -6	2922	3326
bqp250-2	251|31877	9.98 -7	9.99 -7	3591	6082
bqp250-3	251|31877	9.91 -7	9.98 -7	3877	3513
bqp250-4	251|31877	1.00 -6	9.97 -7	2972	3465
bqp250-5	251|31877	1.00 -6	1.00 -6	4962	9768
bqp250-6	251|31877	9.98 -7	1.00 -6	3663	7066
bqp250-7	251|31877	9.98 -7	9.99 -7	3509	4398
bqp250-8	251|31877	9.98 -7	1.00 -6	2729	3364
bqp250-9	251|31877	9.99 -7	1.00 -6	3530	5097
bqp500-1	501|126252	9.99 -7	1.00 -6	6542	7964
bqp500-10	501|126252	9.99 -7	1.00 -6	6967	7833
bqp500-2	501|126252	9.99 -7	1.00 -6	7129	8611
bqp500-3	501|126252	9.99 -7	9.99 -7	6947	7795
bqp500-4	501|126252	9.99 -7	1.00 -6	6991	7741
bqp500-5	501|126252	9.99 -7	1.00 -6	6553	7411
bqp500-6	501|126252	9.98 -7	1.00 -6	6561	7875
bqp500-7	501|126252	9.99 -7	1.00 -6	6858	8302
bqp500-8	501|126252	1.00 -6	9.99 -7	7357	9184
bqp500-9	501|126252	1.00 -6	9.99 -7	6443	7708
gka10b	126|8127	9.98 -7	9.95 -7	2633	2045
gka10d	101|5252	9.97 -7	9.99 -7	1344	1963
gka1d	101|5252	9.99 -7	9.98 -7	1982	5543
gka1e	201|20502	9.99 -7	1.00 -6	3227	5657
gka1f	501|126252	1.00 -6	1.00 -6	6680	7548
gka2d	101|5252	9.99 -7	9.97 -7	1421	3081
gka2e	201|20502	9.99 -7	9.99 -7	2531	4375
gka2f	501|126252	9.99 -7	9.99 -7	7024	9190
gka3d	101|5252	9.92 -7	1.00 -6	2643	9450
gka3e	201|20502	9.93 -7	9.98 -7	2466	4134
gka3f	501|126252	9.99 -7	9.99 -7	6528	7177
gka4d	101|5252	1.00 -6	9.99 -7	1427	3399
gka4e	201|20502	9.99 -7	9.98 -7	3023	6043
gka4f	501|126252	9.98 -7	9.99 -7	6233	7402
gka5d	101|5252	1.00 -6	9.73 -7	1240	1747
gka5e	201|20502	9.99 -7	9.98 -7	2917	5486
gka5f	501|126252	9.99 -7	9.99 -7	6255	8483
gka6d	101|5252	9.93 -7	9.57 -7	1468	2143
gka7c	101|5252	9.97 -7	1.00 -6	3193	7008
gka7d	101|5252	9.66 -7	9.60 -7	1212	1805
gka8a	101|5252	9.94 -7	9.58 -7	15111	5377
gka8d	101|5252	9.98 -7	9.99 -7	2151	8219
gka9b	101|5252	7.58 -7	9.96 -4*	681	25000*
gka9d	101|5252	9.91 -7	9.92 -7	1175	1872

Table 3: 2EBD-HPE results on BIQ problems

INSTANCE	$n\|m$	$\langle c, x\rangle$	$\langle b, w\rangle$	**Time**	$\min_i x_i$	ϵ_P	ϵ_D	ϵ_G	**Iterations**
be100.1	101|5252	-2.002135 +4	-2.002135 +4	10	-2.36 -7	9.70 -7	9.01 -7	-1.15 -6	1511
be100.10	101|5252	-1.640851 +4	-1.640851 +4	9	-9.21 -8	1.43 -7	9.63 -7	-8.46 -7	1111
be100.2	101|5252	-1.798872 +4	-1.798872 +4	10	-1.79 -6	9.69 -7	8.02 -7	+3.55 -7	1200
be100.3	101|5252	-1.823106 +4	-1.823106 +4	13	-1.60 -6	9.91 -7	6.99 -7	-2.86 -7	1531
be100.4	101|5252	-1.984180 +4	-1.984180 +4	9	-1.66 -6	9.99 -7	5.14 -7	-2.96 -7	1500
be100.5	101|5252	-1.688871 +4	-1.688871 +4	12	-2.15 -6	9.98 -7	7.02 -7	-4.16 -7	1215
be100.6	101|5252	-1.814822 +4	-1.814822 +4	13	-9.80 -7	9.95 -7	6.75 -7	-8.45 -7	1470
be100.7	101|5252	-1.970087 +4	-1.970087 +4	10	-1.19 -6	9.73 -7	9.94 -7	-4.66 -7	1279
be100.8	101|5252	-1.994637 +4	-1.994637 +4	8	-8.32 -8	6.46 -7	9.98 -7	-5.86 -7	1246
be100.9	101|5252	-1.426338 +4	-1.426338 +4	8	-5.91 -7	8.39 -7	9.93 -7	-7.46 -7	1168
be120.3.1	121|7502	-1.380356 +4	-1.380356 +4	16	-1.87 -6	9.96 -7	6.61 -7	-8.65 -7	1535
be120.3.10	121|7502	-1.293087 +4	-1.293087 +4	14	-6.21 -7	9.97 -7	9.59 -7	-2.08 -6	1373
be120.3.2	121|7502	-1.362663 +4	-1.362663 +4	13	-1.10 -6	9.99 -7	7.11 -7	-1.36 -6	1721
be120.3.3	121|7502	-1.298791 +4	-1.298791 +4	16	-3.31 -7	9.38 -7	9.95 -7	-1.21 -6	1527
be120.3.4	121|7502	-1.451126 +4	-1.451126 +4	18	-2.18 -7	9.90 -7	8.96 -7	-1.50 -6	1660
be120.3.5	121|7502	-1.199191 +4	-1.199191 +4	21	-2.63 -6	9.99 -7	5.24 -7	+2.70 -8	2015
be120.3.6	121|7502	-1.343206 +4	-1.343206 +4	22	-2.84 -6	9.99 -7	6.79 -7	-4.58 -7	2047
be120.3.7	121|7502	-1.456411 +4	-1.456411 +4	32	-1.51 -6	9.99 -7	7.13 -7	-3.30 -7	3229
be120.3.8	121|7502	-1.530302 +4	-1.530302 +4	22	-1.13 -6	9.99 -7	9.25 -7	-1.74 -7	2405
be120.3.9	121|7502	-1.124132 +4	-1.124132 +4	24	-4.52 -7	9.96 -7	7.10 -7	-4.99 -7	2370
be120.8.1	121|7502	-2.019395 +4	-2.019395 +4	9	-5.18 -7	5.21 -7	9.92 -7	+7.84 -7	1165
be120.8.10	121|7502	-2.002401 +4	-2.002401 +4	16	-1.49 -7	3.57 -7	9.97 -7	-5.88 -7	1513
be120.8.2	121|7502	-2.007413 +4	-2.007413 +4	26	-2.24 -6	1.00 -6	8.29 -7	-6.92 -7	2246
be120.8.3	121|7502	-2.050590 +4	-2.050590 +4	17	-1.17 -6	9.94 -7	6.83 -7	-4.56 -10	1632
be120.8.4	121|7502	-2.177981 +4	-2.177981 +4	19	-1.99 -6	8.88 -7	1.00 -6	-1.47 -6	1802
be120.8.5	121|7502	-2.131628 +4	-2.131628 +4	14	-1.85 -6	9.97 -7	7.06 -7	-2.04 -7	1662
be120.8.6	121|7502	-1.967697 +4	-1.967697 +4	14	-2.86 -6	9.98 -7	6.40 -7	-1.37 -6	1357
be120.8.7	121|7502	-2.373240 +4	-2.373240 +4	18	-3.36 -7	8.82 -7	9.97 -7	-7.11 -7	1682
be120.8.8	121|7502	-2.120476 +4	-2.120476 +4	10	-1.96 -7	9.32 -7	9.99 -7	-5.33 -7	1278
be120.8.9	121|7502	-1.928441 +4	-1.928441 +4	12	-1.65 -7	9.82 -7	9.91 -7	-2.72 -7	1196
be150.3.1	151|11627	-1.984919 +4	-1.984919 +4	26	-3.37 -7	8.56 -7	9.98 -7	-1.86 -6	2104
be150.3.10	151|11627	-1.923092 +4	-1.923092 +4	35	-1.38 -6	9.99 -7	6.67 -7	-3.47 -7	2573
be150.3.2	151|11627	-1.886485 +4	-1.886485 +4	25	-5.78 -7	9.96 -7	6.92 -7	-7.15 -7	2063
be150.3.3	151|11627	-1.804372 +4	-1.804372 +4	22	-4.03 -7	9.16 -7	9.95 -7	-1.09 -6	1715
be150.3.4	151|11627	-2.065267 +4	-2.065267 +4	24	-8.37 -7	8.71 -7	9.98 -7	-9.51 -7	2007
be150.3.5	151|11627	-1.776865 +4	-1.776865 +4	22	-3.83 -6	9.99 -7	7.10 -7	-9.66 -7	1878
be150.3.6	151|11627	-1.805069 +4	-1.805069 +4	21	-3.22 -6	9.93 -7	8.33 -7	-6.60 -7	1619
be150.3.7	151|11627	-1.910131 +4	-1.910131 +4	30	-5.30 -7	9.98 -7	7.34 -7	-6.81 -7	1997
be150.3.8	151|11627	-1.969807 +4	-1.969807 +4	27	-1.97 -6	7.96 -7	9.97 -7	-1.10 -6	2294
be150.3.9	151|11627	-1.410338 +4	-1.410338 +4	17	-1.72 -7	9.96 -7	8.79 -7	-1.55 -6	1323
be150.8.1	151|11627	-2.914369 +4	-2.914369 +4	19	-2.79 -7	9.41 -7	9.99 -7	-5.72 -7	1477
be150.8.10	151|11627	-3.004798 +4	-3.004798 +4	22	-1.72 -6	7.32 -7	9.97 -7	-1.37 -6	1720
be150.8.2	151|11627	-2.882111 +4	-2.882111 +4	20	-5.65 -7	9.98 -7	9.53 -7	-1.88 -6	1431
be150.8.3	151|11627	-3.106038 +4	-3.106038 +4	25	-5.01 -7	9.10 -7	1.00 -6	-7.56 -7	1768
be150.8.4	151|11627	-2.872931 +4	-2.872931 +4	24	-1.61 -6	9.97 -7	6.75 -7	-1.29 -6	1691
be150.8.5	151|11627	-2.948208 +4	-2.948208 +4	24	-1.42 -6	9.96 -7	6.73 -7	-1.12 -6	1930
be150.8.6	151|11627	-3.143723 +4	-3.143723 +4	21	-3.01 -6	9.99 -7	5.63 -7	-4.85 -7	1649
be150.8.7	151|11627	-3.325211 +4	-3.325211 +4	28	-1.42 -6	1.00 -6	9.59 -7	-1.61 -6	2184
be150.8.8	151|11627	-3.160000 +4	-3.160000 +4	27	-3.38 -6	8.09 -7	9.99 -7	-1.10 -6	2137
be150.8.9	151|11627	-2.711073 +4	-2.711073 +4	27	-2.59 -6	9.99 -7	7.04 -7	-1.55 -6	2066
be200.3.1	201|20502	-2.771610 +4	-2.771610 +4	47	-1.90 -7	9.25 -7	1.00 -6	-2.04 -6	2023
be200.3.10	201|20502	-2.576069 +4	-2.576069 +4	53	-1.22 -6	9.99 -7	7.52 -7	-7.36 -7	2287
be200.3.2	201|20502	-2.676079 +4	-2.676079 +4	55	-3.80 -7	7.78 -7	9.98 -7	-1.22 -6	2304
be200.3.3	201|20502	-2.947864 +4	-2.947864 +4	73	-7.33 -7	9.08 -7	9.97 -7	-1.08 -6	3167
be200.3.4	201|20502	-2.910622 +4	-2.910622 +4	54	-3.47 -7	9.98 -7	9.71 -7	-1.49 -6	2316
be200.3.5	201|20502	-2.807299 +4	-2.807299 +4	70	-1.23 -6	9.71 -7	9.98 -7	-1.13 -6	2882
be200.3.6	201|20502	-2.792835 +4	-2.792835 +4	47	-2.61 -7	7.64 -7	9.99 -7	-9.35 -7	2126
be200.3.7	201|20502	-3.162051 +4	-3.162051 +4	61	-2.16 -7	8.10 -7	9.98 -7	-1.08 -6	2689
be200.3.8	201|20502	-2.924430 +4	-2.924430 +4	44	-4.89 -7	9.13 -7	9.97 -7	-1.82 -6	2265
be200.3.9	201|20502	-2.643705 +4	-2.643705 +4	88	-2.39 -7	9.47 -7	9.98 -7	-1.50 -6	3787
be200.8.1	201|20502	-5.086950 +4	-5.086950 +4	60	-5.15 -7	9.98 -7	8.35 -7	-1.31 -6	2620
be200.8.10	201|20502	-4.574308 +4	-4.574308 +4	46	-5.69 -7	9.57 -7	9.97 -7	-1.60 -6	2074
be200.8.2	201|20502	-4.433606 +4	-4.433606 +4	40	-1.58 -7	9.30 -7	1.00 -6	-1.34 -6	1769
be200.8.3	201|20502	-4.625398 +4	-4.625398 +4	49	-9.78 -7	9.99 -7	7.90 -7	-9.74 -7	2309
be200.8.4	201|20502	-4.662126 +4	-4.662126 +4	46	-2.74 -7	9.00 -7	1.00 -6	-1.38 -6	2104
be200.8.5	201|20502	-4.427125 +4	-4.427125 +4	44	-1.00 -6	1.00 -6	9.24 -7	-8.21 -7	1965
be200.8.6	201|20502	-5.121888 +4	-5.121888 +4	56	-2.05 -6	5.30 -7	9.78 -7	-5.35 -7	2451
be200.8.7	201|20502	-4.935283 +4	-4.935283 +4	47	-4.25 -8	4.24 -7	9.97 -7	-1.24 -6	2118
be200.8.8	201|20502	-4.768917 +4	-4.768917 +4	46	-2.61 -6	9.99 -7	8.11 -7	-3.74 -7	2079
be200.8.9	201|20502	-4.549560 +4	-4.549560 +4	52	-4.31 -6	9.97 -7	7.04 -7	-9.25 -7	2310
be250.1	251|31877	-2.511947 +4	-2.511947 +4	136	-1.65 -6	9.93 -7	9.98 -7	-7.49 -7	3638
be250.10	251|31877	-2.435502 +4	-2.435502 +4	150	-1.94 -6	9.99 -7	9.58 -7	-6.56 -7	4220
be250.2	251|31877	-2.368150 +4	-2.368150 +4	117	-6.31 -7	1.00 -6	6.76 -7	-1.32 -6	3643
be250.3	251|31877	-2.400002 +4	-2.400002 +4	118	-7.38 -7	9.97 -7	8.80 -7	-2.07 -6	3297
be250.4	251|31877	-2.572032 +4	-2.572032 +4	215	-9.30 -7	1.00 -6	5.24 -7	-6.93 -7	5779
be250.5	251|31877	-2.237471 +4	-2.237471 +4	120	-8.94 -7	9.99 -7	6.91 -7	-1.05 -6	3560
be250.6	251|31877	-2.401885 +4	-2.401885 +4	125	-3.23 -7	8.36 -7	9.99 -7	-1.06 -6	3570
be250.7	251|31877	-2.511896 +4	-2.511896 +4	121	-2.53 -6	9.98 -7	6.87 -7	-3.44 -7	3321
be250.8	251|31877	-2.502041 +4	-2.502041 +4	127	-8.06 -7	7.06 -7	9.99 -7	-1.94 -6	3502
be250.9	251|31877	-2.139707 +4	-2.139707 +4	139	-8.25 -7	8.66 -7	9.97 -7	-2.15 -6	3947

Table 4: 2EBD-HPE results on BIQ problems

INSTANCE	$n\|m$	$\langle c, x \rangle$	$\langle b, w \rangle$	Time	$\min_i x_i$	ϵ_P	ϵ_D	ϵ_G	Iterations
bqp100-1	101|5252	-8.380387 +3	-8.380387 +3	8	-6.98 -7	9.98 -7	7.27 -7	-3.34 -7	1291
bqp100-10	101|5252	-1.298027 +4	-1.298027 +4	17	-4.99 -6	9.99 -7	6.81 -7	-5.06 -7	1977
bqp100-2	101|5252	-1.148926 +4	-1.148926 +4	19	-7.09 -7	9.90 -7	9.98 -7	-1.12 -6	2742
bqp100-3	101|5252	-1.315318 +4	-1.315318 +4	26	-7.97 -7	9.30 -7	1.00 -6	-6.28 -7	3505
bqp100-4	101|5252	-1.073189 +4	-1.073189 +4	20	-1.03 -6	9.98 -7	9.06 -7	-4.87 -7	2596
bqp100-5	101|5252	-9.487028 +3	-9.487028 +3	16	-2.38 -6	1.00 -6	5.17 -7	-4.42 -7	2248
bqp100-6	101|5252	-1.082476 +4	-1.082476 +4	7	-3.44 -7	8.40 -7	9.97 -7	+2.53 -7	1237
bqp100-7	101|5252	-1.068915 +4	-1.068915 +4	11	-1.94 -7	9.95 -7	9.82 -7	-1.23 -6	1775
bqp100-8	101|5252	-1.176999 +4	-1.176999 +4	18	-1.47 -6	8.88 -7	9.94 -7	-2.77 -7	2307
bqp100-9	101|5252	-1.173325 +4	-1.173325 +4	18	-8.36 -7	8.54 -7	9.99 -7	+2.26 -7	2156
bqp250-1	251|31877	-4.766312 +4	-4.766312 +4	112	-9.51 -7	8.94 -7	1.00 -6	-2.08 -6	3484
bqp250-10	251|31877	-4.301453 +4	-4.301453 +4	112	-8.83 -8	7.70 -7	1.00 -6	-2.24 -6	2922
bqp250-2	251|31877	-4.722238 +4	-4.722238 +4	128	-5.98 -7	9.98 -7	6.70 -7	-1.01 -6	3591
bqp250-3	251|31877	-5.107680 +4	-5.107680 +4	136	-1.21 -7	9.91 -7	8.61 -7	-1.99 -6	3877
bqp250-4	251|31877	-4.331257 +4	-4.331257 +4	89	-3.73 -7	8.09 -7	1.00 -6	-1.54 -6	2972
bqp250-5	251|31877	-5.000433 +4	-5.000433 +4	184	-2.58 -7	8.18 -7	1.00 -6	-1.85 -6	4962
bqp250-6	251|31877	-4.366886 +4	-4.366886 +4	139	-3.99 -7	9.98 -7	6.34 -7	-1.14 -6	3663
bqp250-7	251|31877	-4.892176 +4	-4.892176 +4	119	-3.21 -7	9.98 -7	9.24 -7	-2.80 -6	3509
bqp250-8	251|31877	-3.877955 +4	-3.877955 +4	104	-9.76 -8	6.83 -7	9.98 -7	-1.02 -6	2729
bqp250-9	251|31877	-5.149755 +4	-5.149755 +4	131	-2.73 -6	9.99 -7	7.98 -7	-7.48 -7	3530
bqp500-1	501|126252	-1.259642 +5	-1.259642 +5	821	-2.37 -8	7.56 -7	9.99 -7	-2.77 -6	6542
bqp500-10	501|126252	-1.385345 +5	-1.385345 +5	938	-1.89 -8	7.32 -7	9.99 -7	-3.22 -6	6967
bqp500-2	501|126252	-1.360111 +5	-1.360111 +5	919	-9.53 -8	7.05 -7	9.99 -7	-2.10 -6	7129
bqp500-3	501|126252	-1.384535 +5	-1.384535 +5	877	-1.87 -8	7.26 -7	9.99 -7	-3.12 -6	6947
bqp500-4	501|126252	-1.393284 +5	-1.393284 +5	946	-2.58 -8	8.01 -7	9.99 -7	-2.80 -6	6991
bqp500-5	501|126252	-1.340922 +5	-1.340922 +5	848	-4.49 -7	7.06 -7	9.99 -7	-2.70 -6	6553
bqp500-6	501|126252	-1.307644 +5	-1.307644 +5	885	-6.48 -8	8.26 -7	9.98 -7	-1.74 -6	6561
bqp500-7	501|126252	-1.314915 +5	-1.314915 +5	923	-4.96 -8	6.96 -7	9.99 -7	-2.27 -6	6858
bqp500-8	501|126252	-1.334899 +5	-1.334899 +5	976	-6.95 -8	8.00 -7	1.00 -6	-1.48 -6	7357
bqp500-9	501|126252	-1.302883 +5	-1.302883 +5	872	-2.18 -7	7.14 -7	1.00 -6	-2.24 -6	6443
gka10b	126|8127	-1.555750 +2	-1.555750 +2	25	-5.31 -8	9.98 -7	2.65 -8	-1.30 -5	2633
gka10d	101|5252	-2.010856 +4	-2.010856 +4	9	-4.66 -7	8.60 -7	9.97 -7	+9.90 -7	1344
gka1d	101|5252	-6.528430 +3	-6.528430 +3	16	-6.31 -6	8.71 -7	9.99 -7	-3.90 -7	1982
gka1e	201|20502	-1.706982 +4	-1.706982 +4	76	-7.86 -7	9.99 -7	7.53 -7	-4.94 -7	3227
gka1f	501|126252	-6.555910 +4	-6.555910 +4	911	-1.67 -7	7.16 -7	1.00 -6	-2.03 -6	6680
gka2d	101|5252	-6.990710 +3	-6.990710 +3	9	-1.31 -7	5.66 -7	9.99 -7	-3.91 -7	1421
gka2e	201|20502	-2.491764 +4	-2.491764 +4	59	-5.50 -7	9.99 -7	6.81 -7	-1.09 -6	2531
gka2f	501|126252	-1.079318 +5	-1.079318 +5	1023	-1.96 -7	7.92 -7	9.99 -7	-3.02 -6	7024
gka3d	101|5252	-9.734332 +3	-9.734332 +3	19	-1.59 -6	9.87 -7	9.92 -7	-2.04 -9	2643
gka3e	201|20502	-2.689874 +4	-2.689874 +4	58	-3.32 -6	9.93 -7	8.63 -7	-6.91 -7	2466
gka3f	501|126252	-1.501511 +5	-1.501511 +5	815	-1.04 -7	6.85 -7	9.99 -7	-2.27 -6	6528
gka4d	101|5252	-1.127841 +4	-1.127841 +4	12	-4.75 -7	4.54 -7	1.00 -6	-5.76 -7	1427
gka4e	201|20502	-3.722515 +4	-3.722515 +4	72	-4.21 -7	6.44 -7	9.99 -7	-1.27 -6	3023
gka4f	501|126252	-1.870880 +5	-1.870880 +5	879	-9.55 -7	9.73 -7	9.98 -7	-2.27 -6	6233
gka5d	101|5252	-1.239886 +4	-1.239886 +4	12	-1.98 -6	1.00 -6	6.01 -7	+6.52 -8	1240
gka5e	201|20502	-3.800231 +4	-3.800231 +4	68	-2.23 -6	7.29 -7	9.99 -7	-7.77 -7	2917
gka5f	501|126252	-2.069143 +5	-2.069143 +5	879	-8.21 -8	8.04 -7	9.99 -7	-1.46 -6	6255
gka6d	101|5252	-1.492937 +4	-1.492937 +4	11	-1.84 -6	9.47 -7	9.93 -7	-4.16 -7	1468
gka7c	101|5252	-7.316449 +3	-7.316449 +3	26	-6.22 -7	9.97 -7	9.85 -7	-6.45 -7	3193
gka7d	101|5252	-1.537583 +4	-1.537583 +4	10	-2.95 -7	9.66 -7	8.97 -7	-7.00 -7	1212
gka8a	101|5252	-1.119723 +4	-1.119723 +4	143	-2.00 -8	8.70 -7	9.94 -7	-2.45 -6	15111
gka8d	101|5252	-1.700536 +4	-1.700536 +4	16	-1.27 -6	9.98 -7	8.96 -7	-7.41 -7	2151
gka9b	101|5252	-1.370000 +2	-1.370000 +2	5	-9.24 -7	7.58 -7	6.71 -8	+2.57 -7	681
gka9d	101|5252	-1.653390 +4	-1.653390 +4	8	-3.07 -8	2.68 -7	9.91 -7	+9.30 -7	1175

problem on G can be formulated as a κ-cut problem

$$\begin{array}{ll} \max & \left[\left(\frac{\kappa-1}{2\kappa}\right) L(G,W) - \frac{1}{2}\mathrm{Diag}(We)\right] \bullet X \\ \text{s.t.} & -E^{ij} \bullet X \leq 2/(k\kappa - 1), \quad \forall\ (i,j), \\ & -E^{ij} \bullet X = 2/(\kappa - 1), \quad \forall\ (i,j) \in U \subseteq E, \\ & \mathrm{diag}(X) = e,\ X \succeq 0,\ \mathrm{rank}(X) = \kappa, \end{array}$$

where $\kappa > 1$ is an integer, $L(G,W) := \mathrm{Diag}(We) - W$ is the Laplacian matrix, $E^{ij} = e_i e_j^T + e_j e_i^T$ with $e_i \in \mathbb{R}^n$ the vector with all zeros except in the ith position and $e \in \mathbb{R}^n$ is the vector with all ones. An SDP relaxation of the problem above is obtained by dropping the rank restriction and the inequality constraint for the non-edges to obtain the following formulation

$$\begin{array}{lll} \max & \left[\left(\frac{\kappa-1}{2\kappa}\right) L(G,W) - \frac{1}{2}\mathrm{Diag}(We)\right] \bullet X & \\ \text{s.t.} & X \succeq 0, & \text{(62a)} \\ & -E^{ij} \bullet X \leq 2/(\kappa-1) \quad \forall\ (i,j) \in E \setminus U, & \text{(62b)} \\ & -E^{ij} \bullet X = 2/(\kappa-1) \quad \forall\ (i,j) \in U \subseteq E,\ \mathrm{diag}(X) = e, & \text{(62c)} \end{array}$$

where $X \in \mathcal{S}^n$.

There is more than one way of viewing (62) as a special case of formulation (45). In our current implementation, we considered the following two-easy-block structure formulation. Let $\mathcal{X} = \mathcal{W}_1 := \mathcal{S}^n$, $\mathcal{W}_2 = \mathbb{R}^{|E \setminus U|} \times \mathbb{R}^{|U|} \times \mathbb{R}^n$, $\mathcal{K}_1 = \mathcal{S}^n_+$ and $\mathcal{K}_2 = \mathbb{R}^{|E \setminus U|}_+ \times \mathbf{0}_{|U|} \times \mathbf{0}_n$, where $\mathbf{0}_n$ denotes an n dimensional vector of all zeros. Also, endow $\mathcal{X}$ with the Frobenius inner product. With these definitions, we can easily see that we can view (62) as having the two-easy-block structure (45) if we let (62a) as $\mathcal{M}_1$, and (62b) and (62c) as $\mathcal{M}_2$.

In order to agree with the scaling of the data and error measures adopted by SDPAD for (59), we measure the primal infeasibility measure ϵ_P as

$$\epsilon_P(x) = \frac{\sqrt{\frac{1}{2}\sum_{(i,j)\in E\setminus U} \|\min\{0, E^{ij} \bullet X + 2/(\kappa-1)\}\|_F^2 + \frac{1}{2}\sum_{(i,j)\in U} \|E^{ij} \bullet X + 2/(\kappa-1)\|_F^2 + \|\mathrm{diag}(X) - e\|_F^2}}{1 + \sqrt{n + 2|E|/(\kappa-1)^2}}, \tag{63}$$

and the dual infeasibility measure ϵ_D as in (51) with $\mu = 1 + \|C\|_1$, where $\|\cdot\|_F$ is the Frobenius norm defined in (61) and $\|\cdot\|_1$ is the matrix 1-norm defined as

$$\|A\|_1 = \max_j \sum_{i=1}^m |A_{ij}|, \qquad \forall A \in \mathbb{R}^{m\times n}.$$

Note that the above primal infeasibility measure is not a special case of (50) since it does not take into consideration the violations with respect to the constraint $X \succeq 0$. We observe also that, in view of the first inclusion in Theorem 5.2(a), the constraint $X \succeq 0$ is always satisfied by 2EBD-HPE , while SDPAD approaches it in the limit.

Table 5 compares the two methods on a collection of 7 SDP relaxations of FAPs using the tolerance $\bar{\epsilon} = 10^{-6}$. For the purpose of this comparison, we considered 2EBD-HPE with $\sigma = 0.99$ and the values of γ, τ and $\bar{k}$ in the dynamic scaling rule (57) set to $\gamma = 1.5$, $\tau = 0.75$ and $\bar{k} = 5$.

Table 6 give more detailed computational results obtained by our method 2EBD-HPE, such as the primal and dual objective function values, number of iterations, the primal and dual infeasibility measures as described above, and the relative duality gap. Figure 5 plots the performance profiles of both methods.

Note that 2EBD-HPE solves 5 (out of a total of 7) problems faster than SDPAD. Moreover, 2EBD-HPE solves about 3 of them almost 2 times faster than SDPAD. Note also that 2EBD-HPE performs better than SDPAD on large SDP relaxations of FAPs (i.e., `fap25` and `fap36`).

Table 5: Comparison of the methods on FAPs

Problem		$\max\{\epsilon_P, \epsilon_D, \epsilon_G\}$		Iterations	
Instance	$n_s \vert m$	**2EBD-HPE**	**SDPAD**	**2EBD-HPE**	**SDPAD**
fap08	120|7260	9.29 -7	9.90 -7	924	717
fap09	174|15225	9.57 -7	9.98 -7	716	505
fap10	183|14479	9.98 -7	9.92 -7	2480	4728
fap11	252|24292	9.98 -7	9.99 -7	2419	2749
fap12	369|26462	9.99 -7	9.93 -7	2991	3609
fap25	2118|322924	1.00 -6	9.96 -7	4389	7948
fap36	4110|1154467	9.99 -7	9.92 -7	3609	6328

Table 6: 2EBD-HPE results on FAPs

Instance	$n \vert m$	$\langle c, x\rangle$	$\langle b, w\rangle$	**Time**	ϵ_P	ϵ_D	ϵ_G	**Iterations**
fap08	120|7260	+2.436300 +0	+2.436300 +0	8	9.29 -7	8.62 -7	+5.77 -7	924
fap09	174|15225	+1.079770 +1	+1.079770 +1	10	9.57 -7	9.35 -7	-3.73 -6	716
fap10	183|14479	+9.643400 -3	+9.643400 -3	41	9.98 -7	5.84 -7	-8.08 -5	2480
fap11	252|24292	+2.964500 -2	+2.964500 -2	66	9.98 -7	7.08 -7	-2.38 -4	2419
fap12	369|26462	+2.731000 -1	+2.731000 -1	142	9.99 -7	4.33 -7	-1.79 -4	2991
fap25	2118|322924	+1.287750 +1	+1.287750 +1	9074	8.72 -7	1.00 -6	-7.70 -5	4389
fap36	4110|1154467	+6.985640 +1	+6.985640 +1	63291	9.99 -7	9.87 -7	-1.99 -5	3609

7.3 SDPs arising from relaxation of maximum stable set problems

This subsection compares the performance of our method 2EBD-HPE with that of SDPAD on a collection of SDPs corresponding to θ-functions and θ_+-functions of graph stable set problems.

The SDPs for θ-functions and θ_+-functions of graph stable set problems can be described as follows. Given a graph G with n nodes and an edge set E, the SDP relaxations $\theta(G)$ and $\theta_+(G)$ of the maximum stable set problem are defined as

$$
\begin{aligned}
\theta(G) := \max\, & C \bullet X & \qquad \theta_+(G) := \max\, & C \bullet X & \\
\text{s.t } & X \succeq 0, & \text{s.t } & X \succeq 0, & \quad (64a)\\
& I \bullet X = 1, & & I \bullet X = 1, & \quad (64b)\\
& X_{ij} = 0,\ (i,j) \in E, & & X_{ij} = 0,\ (i,j) \in E,\ X \geq 0, & \quad (64c)
\end{aligned}
$$

where $C = ee^T$, $X \in \mathcal{S}^n$ and $e \in \mathbb{R}^n$ is the vector with all ones.

There is more than one way of viewing the $\theta(G)$ and $\theta_+(G)$ problems as special cases of formulation (45). In our current implementation, we considered the following two-easy-block structure formulations. For the case of the $\theta(G)$ (resp. $\theta_+(G)$) problem, we let $\mathcal{X} = \mathcal{S}^n$, $\mathcal{W}_1 := \mathcal{S}^n \times \mathbb{R}$, $\mathcal{W}_2 = \mathbb{R} \times \mathbb{R}^{|E|}$, $\mathcal{K}_1 = \mathcal{S}^n_+ \times \mathbf{0}_1$ and $\mathcal{K}_2 = \mathbf{0}_1 \times \mathbf{0}_{|E|}$ (resp. $\mathcal{K}_2 = \mathbf{0}_1 \times \mathbf{0}_{|E|} \times \mathbb{R}^{n(n+1)/2}_+$). Also, endow $\mathcal{X}$ with the Frobenius inner product. With these definitions, we can easily see that we can view the $\theta(G)$ and $\theta_+(G)$ problems as having the two-easy-block structure (45) if we let $\mathcal{M}_1$ (resp. $\mathcal{M}_2$) to be the set of $X \in \mathcal{S}^n$ satisfying (64a) and (64b) (resp. (64b) and (64c)). Note that (64b) is used to define both $\mathcal{M}_1$ and $\mathcal{M}_2$.

In order to agree with the scaling of the data and error measures adopted by SDPAD for the $\theta(G)$ and $\theta_+(G)$ problems, we measure the primal infeasibility measure ϵ_P as

$$
\epsilon_P(x) = \frac{\sqrt{\frac{1}{2} \sum\limits_{(i,j)\in E} X_{ij}^2 + \frac{1}{n} \left(I \bullet X - 1\right)^2}}{1 + \sqrt{\frac{1}{n}}}, \tag{65}
$$

and the dual infeasibility measure ϵ_D as in (51) with $\mu = 1 + 2\|C\|_F$, where $\|\cdot\|_F$ is the Frobenius norm defined in (61). Note that the above primal infeasibility measure is not a special case of (50) since it does not take into consideration the violations with respect to the constraint $X \succeq 0$ (and also $X \geq 0$ for the case of the $\theta_+(G)$ problem). We observe also that, in view of the first inclusion in Theorem 5.2(a), the constraints $X \succeq 0$

and $I \bullet X = 1$ are always satisfied by 2EBD-HPE , while SDPAD approaches them in the limit. Note also that both methods approach the constraint $X \geq 0$ for the $\theta_+(G)$ problem only in the limit.

Tables 7 and 9 compare the two methods on a collection of 59 $\theta(G)$ and $\theta_+(G)$ problems using the tolerance $\bar{\epsilon} = 10^{-6}$. For the purpose of this comparison, we considered 2EBD-HPE with $\sigma = 0.9$ and the values of γ, τ and $\bar{k}$ in the dynamic scaling rule (57) set to $\gamma = 1.5$, $\tau = 0.75$ and $\bar{k} = 5$. For the $\theta(G)$ problems, we stopped the dynamic scaling rule when an accuracy of $\bar{\epsilon} = 10^{-5}$ is achieved.

Tables 8 and 10 give more detailed computational results obtained by our method 2EBD-HPE, such as the primal and dual objective function values, number of iterations, the primal and dual infeasibility measures as described above, and the relative duality gap. Since our implementation of 2EBD-HPE is based on ϵ_P in order to conform with the stopping criterion of SDPAD and, as observed above for the $\theta_+(G)$ problem, ϵ_P does not take into consideration the violations with respect to the constraint $X \geq 0$, we also include in Tables 10 a column with the minimum value of all entries of X. Figures 7 and 8 plots the performance profiles of both methods for solving $\theta(G)$ and $\theta_+(G)$ instances, respectively.

Note that 2EBD-HPE solves 46 (out of a total of 59) $\theta(G)$ and 47 (out of a total of 59) $\theta_+(G)$ problems faster than SDPAD. Moreover, 2EBD-HPE solves about 5 $\theta(G)$ and 11 $\theta_+(G)$ problems at least 2 times faster than SDPAD. Note also that 2EBD-HPE fails to solve 4 $\theta(G)$ (`1et.2048`, `1tc.2048`, `G52` and `G53`) and 1 $\theta_+(G)$ (`1et.256`) instances while SDPAD fails to solve 8 $\theta(G)$ (`1dc.128`, `1dc.512`, `1tc.1024`, `1tc.2048`, `2dc.1024`, `2dc.2048`, `G52` and `G53`) and 2 $\theta_+(G)$ (`G52` and `G53`) instances.

8 Numerical results: part II

In this section, we briefly compare 2EBD-HPE with the SDPNAL method presented in [20] and a BD method presented in [10], namely DSA-BD. For this comparison with use the same classes of conic optimization problems mentioned in Section 7.

In contrast to 2EBD-HPE, the methods DSA-BD and SDPNAL always require as input a conic optimization problem given in standard form, i.e., as in (1). Hence, for the latter two codes it is necessary (except for the θ-function SDP problems) to add additional variables to the original conic optimization problem (45) in order to obtain a standard form formulation. Thus, the number of variables handled by the latter two codes are usually larger than the number of variables handled by 2EBD-HPE. As our computational results of this section show, this has a negative effect on the performance of DSA-BD and SDPNAL compared to 2EBD-HPE.

For the 2EBD-HPE, DSA-BD and SDPNAL methods, the computational results for the SDP relaxations of BIQs and FAPs were obtained on a server with 2 Xeon X5460 processors at 3.16GHz and 32GB RAM, and the ones corresponding to the SDPs for θ-functions and θ_+-functions of graph stable set problems were obtained on a single core of a server with 2 Xeon X5520 processors at 2.27GHz and 48GB RAM. For every problem class, we endow $\mathcal{X}$ with the Frobenius inner product.

We use the same stopping criterion adopted by both DSA-BD and SDPNAL. More specifically, we measure the primal and dual infeasibility measures ϵ_P and ϵ_D as in (50) and (51), respectively, where the norm $\|\cdot\|'_{\mathcal{W}_i}$ is defined as

$$\|\cdot\|'_{\mathcal{W}_i} = \frac{\|\cdot\|_F}{\left(1 + \sqrt{\|b_1\|_F^2 + \|b_2\|_F^2}\right)}, \qquad i = 1, 2,$$

and the parameter μ is defined as $\mu = 1 + \|C\|_F$. For $\bar{\epsilon} = 10^{-6}$, we stop the three methods whenever

$$\max\{\epsilon_{P,k}, \epsilon_{D,k}\} \leq \bar{\epsilon}. \tag{66}$$

For the sake of shortness, we only report the performance profiles and exclude the detailed tables as the ones reported in Section 7. Figures 9, 10 and 11 plot the performance profiles of 2EBD-HPE, DSA-BD and SDPNAL for the SDP relaxations of BIQ problems, the SDP relaxations of FAPs, and the SDPs for θ-functions and θ_+-functions of graph stable set problems, respectively. Note that based on these performance profiles, 2EBD-HPE outperforms DSA-BD and SDPNAL in every problem class.

Table 7: Comparison of the methods on $\theta(G)$

Problem		$\max\{\epsilon_P, \epsilon_D, \epsilon_G\}$		Iterations	
Instance	$n_s\|m$	**2EBD-HPE**	**SDPAD**	**2EBD-HPE**	**SDPAD**
1dc.1024	1024|24064	1.00 -6	9.99 -7	13209	9458
1dc.128	128|1472	8.56 -7	9.23 -6*	13260	20000*
1dc.2048	2048|58368	1.00 -6	9.98 -7	16540	10905
1dc.256	256|3840	9.58 -7	9.43 -7	2601	3969
1dc.512	512|9728	1.00 -6	2.17 -6*	11041	20000*
1et.1024	1024|9601	1.00 -6	1.00 -6	9150	7857
1et.128	128|673	9.69 -7	9.71 -7	580	333
1et.2048	2048|22529	3.12 -3*	1.00 -6	20000*	10968
1et.256	256|1665	9.99 -7	9.70 -7	1758	2188
1et.512	512|4033	9.72 -7	9.71 -7	3147	3302
1tc.1024	1024|7937	1.00 -6	2.17 -6*	12797	20000*
1tc.128	128|513	9.88 -7	9.12 -7	391	898
1tc.2048	2048|18945	2.85 -5*	2.08 -6*	20000*	20000*
1tc.256	256|1313	9.99 -7	9.99 -7	5079	4329
1tc.512	512|3265	1.00 -6	1.00 -6	11860	15022
1zc.1024	1024|16641	8.71 -7	9.72 -7	1176	608
1zc.128	128|1121	8.18 -7	9.96 -7	311	330
1zc.256	256|2817	9.05 -7	8.13 -7	316	211
1zc.512	512|6913	9.70 -7	9.56 -7	510	349
2dc.1024	1024|169163	1.00 -6	5.58 -6*	15667	20000*
2dc.2048	2048|504452	1.00 -6	5.78 -6*	12399	20000*
2dc.512	512|54896	9.82 -7	1.00 -6	7688	8622
G43	1000|9991	9.97 -7	9.96 -7	726	1102
G44	1000|9991	9.91 -7	7.77 -7	750	1131
G45	1000|9991	1.00 -6	9.90 -7	746	1149
G46	1000|9991	9.99 -7	9.61 -7	750	1136
G47	1000|9991	9.91 -7	9.92 -7	772	1103
G51	1000|5910	9.99 -7	9.99 -7	4183	5110
G52	1000|5917	2.73 -6*	2.72 -6*	20000*	20000*
G53	1000|5915	3.67 -6*	6.40 -6*	20000*	20000*
G54	1000|5917	9.99 -7	1.00 -6	2953	3843
brock200-1	200|5067	9.59 -7	9.74 -7	212	240
brock200-4	200|6812	9.76 -7	9.69 -7	196	208
brock400-1	400|20078	9.63 -7	9.75 -7	214	254
c-fat200-1	200|18367	9.72 -7	1.00 -6	258	302
hamming-10-2	1024|23041	9.26 -7	9.10 -7	1153	630
hamming-7-5-6	128|1793	8.86 -7	9.01 -7	252	515
hamming-8-3-4	256|16129	9.09 -7	9.00 -7	191	188
hamming-8-4	256|11777	8.91 -7	7.58 -7	347	136
hamming-9-5-6	512|53761	8.04 -7	8.73 -7	212	1092
hamming-9-8	512|2305	8.87 -7	9.55 -7	1340	2622
keller4	171|5101	1.00 -6	9.87 -7	252	249
p-hat300-1	300|33918	9.97 -7	9.97 -7	709	705
san200-0.7-1	200|5971	8.41 -7	9.52 -7	152	3752
sanr200-0.7	200|6033	9.78 -7	9.98 -7	204	219
theta10	500|12470	9.96 -7	9.84 -7	291	460
theta102	500|37467	9.44 -7	9.83 -7	215	253
theta103	500|62516	9.91 -7	9.42 -7	216	257
theta104	500|87245	9.97 -7	9.72 -7	240	260
theta12	600|17979	9.96 -7	9.79 -7	297	404
theta123	600|90020	9.65 -7	9.71 -7	237	263
theta32	150|2286	9.94 -7	9.89 -7	278	300
theta4	200|1949	9.96 -7	9.92 -7	379	395
theta42	200|5986	9.72 -7	9.85 -7	226	242
theta6	300|4375	9.81 -7	9.88 -7	299	440
theta62	300|13390	9.67 -7	9.90 -7	219	229
theta8	400|7905	9.65 -7	9.88 -7	292	369
theta82	400|23872	9.60 -7	9.71 -7	210	243
theta83	400|39862	9.79 -7	9.89 -7	216	233

Table 8: 2EBD-HPE results on $\theta(G)$

Instance	$n\|m$	$\langle c, x\rangle$	$\langle b, w\rangle$	Time	ϵ_P	ϵ_D	ϵ_G	Iterations
1dc.1024	1024|24064	-9.598550 +1	-9.598550 +1	14841	1.00 -6	2.87 -7	-2.96 -6	13209
1dc.128	128|1472	-1.684180 +1	-1.684180 +1	135	6.57 -7	6.63 -7	+8.56 -7	13260
1dc.2048	2048|58368	-1.747310 +2	-1.747310 +2	104725	1.00 -6	2.89 -7	-4.61 -6	16540
1dc.256	256|3840	-3.000000 +1	-3.000000 +1	107	6.02 -7	9.58 -7	+4.09 -7	2601
1dc.512	512|9728	-5.303100 +1	-5.303100 +1	1913	1.00 -6	4.37 -7	-2.86 -6	11041
1et.1024	1024|9601	-1.842271 +2	-1.842271 +2	9638	1.00 -6	5.49 -7	-2.26 -6	9150
1et.128	128|673	-2.923090 +1	-2.923090 +1	6	9.69 -7	8.87 -7	-2.48 -7	580
1et.2048	2048|22529	-3.983327 +2	-3.456929 +2	167053	3.12 -3	2.92 -3	-7.07 -2	20000
1et.256	256|1665	-5.511440 +1	-5.511440 +1	58	9.99 -7	6.34 -7	-1.40 -6	1758
1et.512	512|4033	-1.044244 +2	-1.044244 +2	554	9.42 -7	9.72 -7	-1.93 -6	3147
1tc.1024	1024|7937	-2.063051 +2	-2.063051 +2	13894	1.00 -6	9.50 -7	-1.87 -6	12797
1tc.128	128|513	-3.800000 +1	-3.800000 +1	4	9.88 -7	9.56 -7	+9.24 -7	391
1tc.2048	2048|18945	-3.747874 +2	-3.746651 +2	148639	2.85 -5	2.79 -5	-1.63 -4	20000
1tc.256	256|1313	-6.339990 +1	-6.339990 +1	197	9.99 -7	6.77 -7	-4.40 -7	5079
1tc.512	512|3265	-1.134006 +2	-1.134006 +2	2170	9.92 -7	1.00 -6	-1.87 -6	11860
1zc.1024	1024|16641	-1.286700 +2	-1.286700 +2	1233	8.35 -7	8.71 -7	-1.40 -5	1176
1zc.128	128|1121	-2.066690 +1	-2.066690 +1	3	6.78 -7	8.18 -7	-4.02 -6	311
1zc.256	256|2817	-3.800050 +1	-3.800050 +1	9	8.39 -7	9.05 -7	-6.35 -6	316
1zc.512	512|6913	-6.875130 +1	-6.875130 +1	87	9.70 -7	9.42 -7	-1.03 -5	510
2dc.1024	1024|169163	-1.863870 +1	-1.863870 +1	17880	1.00 -6	5.56 -7	-1.91 -5	15667
2dc.2048	2048|504452	-3.067430 +1	-3.067430 +1	88042	1.00 -6	2.30 -7	-2.22 -5	12399
2dc.512	512|54896	-1.176790 +1	-1.176790 +1	1491	6.74 -7	9.82 -7	-3.04 -6	7688
G43	1000|9991	-2.806257 +2	-2.806257 +2	817	9.97 -7	8.68 -7	-1.54 -6	726
G44	1000|9991	-2.805840 +2	-2.805840 +2	808	9.91 -7	7.87 -7	-1.11 -6	750
G45	1000|9991	-2.801858 +2	-2.801858 +2	800	1.00 -6	8.01 -7	-8.52 -7	746
G46	1000|9991	-2.798371 +2	-2.798371 +2	717	9.99 -7	7.51 -7	-2.17 -7	750
G47	1000|9991	-2.818943 +2	-2.818943 +2	757	9.91 -7	7.07 -7	-4.46 -7	772
G51	1000|5910	-3.490001 +2	-3.490001 +2	4167	9.99 -7	4.67 -7	-1.77 -7	4183
G52	1000|5917	-3.483936 +2	-3.483877 +2	24146	2.73 -6	1.08 -6	-1.18 -5	20000
G53	1000|5915	-3.483595 +2	-3.483505 +2	19737	3.67 -6	1.91 -6	-1.29 -5	20000
G54	1000|5917	-3.410001 +2	-3.410001 +2	3430	9.99 -7	3.38 -7	-2.09 -7	2953
brock200-1	200|5067	-2.745670 +1	-2.745670 +1	5	9.59 -7	8.54 -7	-6.45 -7	212
brock200-4	200|6812	-2.129350 +1	-2.129350 +1	5	9.76 -7	9.08 -7	-7.21 -7	196
brock400-1	400|20078	-3.970200 +1	-3.970200 +1	24	9.20 -7	9.63 -7	-8.17 -7	214
c-fat200-1	200|18367	-1.200000 +1	-1.200000 +1	5	9.72 -7	7.38 -7	-2.75 -6	258
hamming-10-2	1024|23041	-1.024037 +2	-1.024037 +2	1107	8.07 -7	9.26 -7	-1.69 -5	1153
hamming-7-5-6	128|1793	-4.266680 +1	-4.266680 +1	2	6.96 -7	8.86 -7	-2.42 -6	252
hamming-8-3-4	256|16129	-2.559990 +1	-2.559990 +1	6	2.11 -7	9.09 -7	+4.00 -7	191
hamming-8-4	256|11777	-1.600010 +1	-1.600010 +1	12	2.76 -7	8.91 -7	-4.14 -6	347
hamming-9-5-6	512|53761	-8.533390 +1	-8.533390 +1	31	8.04 -7	7.81 -7	-3.79 -6	212
hamming-9-8	512|2305	-2.240020 +2	-2.240020 +2	192	8.87 -7	8.65 -7	-3.68 -6	1340
keller4	171|5101	-1.401230 +1	-1.401230 +1	4	1.00 -6	8.77 -7	-1.03 -6	252
p-hat300-1	300|33918	-1.006800 +1	-1.006800 +1	43	9.97 -7	2.63 -7	-1.31 -6	709
san200-0.7-1	200|5971	-3.000000 +1	-3.000000 +1	4	8.41 -7	6.85 -7	-8.64 -8	152
sanr200-0.7	200|6033	-2.383620 +1	-2.383620 +1	5	9.78 -7	8.98 -7	-5.32 -7	204
theta10	500|12470	-8.380610 +1	-8.380610 +1	52	9.96 -7	9.83 -7	-7.89 -7	291
theta102	500|37467	-3.839060 +1	-3.839060 +1	42	9.11 -7	9.44 -7	-1.03 -6	215
theta103	500|62516	-2.252860 +1	-2.252860 +1	47	9.13 -7	9.91 -7	-1.42 -6	216
theta104	500|87245	-1.333620 +1	-1.333620 +1	50	8.86 -7	9.97 -7	-1.99 -6	240
theta12	600|17979	-9.280180 +1	-9.280180 +1	86	8.97 -7	9.96 -7	-7.39 -7	297
theta123	600|90020	-2.466870 +1	-2.466870 +1	68	9.65 -7	6.39 -7	-1.69 -6	237
theta32	150|2286	-2.757160 +1	-2.757160 +1	5	9.94 -7	5.50 -7	-1.94 -7	278
theta4	200|1949	-5.032130 +1	-5.032130 +1	9	9.96 -7	6.59 -7	-4.95 -7	379
theta42	200|5986	-2.393170 +1	-2.393170 +1	6	9.72 -7	6.76 -7	-4.01 -7	226
theta6	300|4375	-6.347720 +1	-6.347720 +1	19	9.81 -7	9.02 -7	-4.81 -7	299
theta62	300|13390	-2.964130 +1	-2.964130 +1	13	9.67 -7	6.18 -7	-8.22 -7	219
theta8	400|7905	-7.395370 +1	-7.395370 +1	32	9.65 -7	8.52 -7	-6.46 -7	292
theta82	400|23872	-3.436700 +1	-3.436700 +1	25	9.41 -7	9.60 -7	-8.93 -7	210
theta83	400|39862	-2.030190 +1	-2.030190 +1	28	9.31 -7	9.79 -7	-1.27 -6	216

Table 9: Comparison of the methods on $\theta_+(G)$

Problem		$\max\{\epsilon_P, \epsilon_D, \epsilon_G\}$		Iterations	
Instance	$n_s \vert m$	**2EBD-HPE**	**SDPAD**	**2EBD-HPE**	**SDPAD**
1dc.1024	1024|24064	1.00 -6	9.99 -7	2514	2790
1dc.128	128|1472	9.99 -7	9.99 -7	870	926
1dc.2048	2048|58368	1.00 -6	9.95 -7	5444	6171
1dc.256	256|3840	8.93 -7	9.74 -7	348	3798
1dc.512	512|9728	1.00 -6	9.98 -7	1767	1991
1et.1024	1024|9601	9.97 -7	9.90 -7	1701	2392
1et.128	128|673	9.22 -7	9.78 -7	477	392
1et.2048	2048|22529	9.99 -7	1.00 -6	3322	3547
1et.256	256|1665	9.53 -5*	1.00 -6	20000*	1133
1et.512	512|4033	9.99 -7	9.89 -7	1055	1401
1tc.1024	1024|7937	9.99 -7	9.96 -7	5150	5003
1tc.128	128|513	8.27 -7	9.90 -7	201	1066
1tc.2048	2048|18945	9.99 -7	9.98 -7	4937	5283
1tc.256	256|1313	9.99 -7	9.99 -7	1881	2264
1tc.512	512|3265	9.99 -7	9.94 -7	2535	3032
1zc.1024	1024|16641	9.11 -7	7.18 -7	1141	849
1zc.128	128|1121	6.96 -7	9.75 -7	217	161
1zc.256	256|2817	8.83 -7	8.13 -7	209	180
1zc.512	512|6913	8.84 -7	9.80 -7	550	769
2dc.1024	1024|169163	1.00 -6	1.00 -6	1974	1716
2dc.2048	2048|504452	9.94 -7	1.00 -6	1421	1802
2dc.512	512|54896	9.99 -7	1.00 -6	1809	2179
G43	1000|9991	9.99 -7	9.59 -7	622	1103
G44	1000|9991	9.93 -7	9.70 -7	660	1100
G45	1000|9991	9.94 -7	8.74 -7	658	1145
G46	1000|9991	9.94 -7	9.98 -7	645	1106
G47	1000|9991	9.86 -7	9.99 -7	651	1073
G51	1000|5910	9.99 -7	9.99 -7	4475	13379
G52	1000|5917	9.97 -7	3.00 -4*	6260	20000*
G53	1000|5915	1.00 -6	1.93 -4*	14407	20000*
G54	1000|5917	9.93 -7	9.91 -7	2561	6645
brock200-1	200|5067	9.60 -7	9.80 -7	228	237
brock200-4	200|6812	9.56 -7	9.84 -7	210	212
brock400-1	400|20078	9.76 -7	9.98 -7	240	258
c-fat200-1	200|18367	9.42 -7	9.98 -7	248	293
hamming-10-2	1024|23041	9.00 -7	9.39 -7	882	581
hamming-7-5-6	128|1793	8.73 -7	9.45 -7	611	508
hamming-8-3-4	256|16129	9.09 -7	9.00 -7	191	188
hamming-8-4	256|11777	9.49 -7	6.51 -7	193	114
hamming-9-5-6	512|53761	9.98 -7	9.27 -7	554	433
hamming-9-8	512|2305	8.37 -7	9.88 -7	1053	2569
keller4	171|5101	9.42 -7	9.41 -7	435	419
p-hat300-1	300|33918	9.98 -7	9.96 -7	566	677
san200-0.7-1	200|5971	8.11 -7	9.68 -7	121	3689
sanr200-0.7	200|6033	9.68 -7	9.57 -7	222	228
theta10	500|12470	9.72 -7	9.80 -7	317	382
theta102	500|37467	9.68 -7	9.97 -7	235	263
theta103	500|62516	9.84 -7	9.52 -7	231	262
theta104	500|87245	9.99 -7	9.55 -7	246	266
theta12	600|17979	9.87 -7	9.77 -7	319	393
theta123	600|90020	9.97 -7	9.43 -7	239	267
theta32	150|2286	9.78 -7	9.99 -7	273	276
theta4	200|1949	9.70 -7	9.78 -7	355	405
theta42	200|5986	9.88 -7	9.72 -7	235	242
theta6	300|4375	9.93 -7	9.75 -7	322	370
theta62	300|13390	9.87 -7	9.61 -7	232	235
theta8	400|7905	9.98 -7	9.89 -7	318	358
theta82	400|23872	9.74 -7	9.75 -7	231	250
theta83	400|39862	9.70 -7	9.69 -7	225	256

Table 10: 2EBD-HPE results on $\theta_+(G)$

Instance	$n\|m$	$\langle c, x\rangle$	$\langle b, w\rangle$	**Time**	$\min_i x_i$	ϵ_P	ϵ_D	ϵ_G	**Iterations**
1dc.1024	1024|24064	-9.555170 +1	-9.555170 +1	3299	-2.02 -8	1.00 -6	7.97 -7	-2.68 -6	2514
1dc.128	128|1472	-1.667840 +1	-1.667840 +1	10	-4.37 -8	9.72 -7	9.99 -7	-3.92 -6	870
1dc.2048	2048|58368	-1.742592 +2	-1.742592 +2	46243	-6.66 -9	1.00 -6	8.91 -7	-4.62 -6	5444
1dc.256	256|3840	-3.000010 +1	-3.000010 +1	12	-1.40 -8	6.28 -7	8.93 -7	-8.25 -6	348
1dc.512	512|9728	-5.269530 +1	-5.269530 +1	305	-4.04 -8	1.00 -6	6.88 -7	-1.86 -6	1767
1et.1024	1024|9601	-1.820731 +2	-1.820731 +2	2034	-4.64 -9	9.74 -7	9.97 -7	-4.48 -6	1701
1et.128	128|673	-2.923090 +1	-2.923090 +1	5	-5.07 -8	9.22 -7	7.91 -7	+3.26 -7	477
1et.2048	2048|22529	-3.381690 +2	-3.381690 +2	25791	-3.42 -9	9.99 -7	8.15 -7	-5.66 -6	3322
1et.256	256|1665	-5.448150 +1	-5.447670 +1	861	-1.06 -4	7.92 -5	9.53 -5	+4.42 -5	20000
1et.512	512|4033	-1.035501 +2	-1.035501 +2	160	-1.04 -8	9.99 -7	9.35 -7	-3.87 -6	1055
1tc.1024	1024|7937	-2.042054 +2	-2.042054 +2	6548	-3.30 -8	9.99 -7	9.87 -7	-3.43 -6	5150
1tc.128	128|513	-3.799990 +1	-3.799990 +1	2	-1.58 -7	8.25 -7	8.27 -7	-7.85 -7	201
1tc.2048	2048|18945	-3.704916 +2	-3.704916 +2	40227	-2.13 -8	9.20 -7	9.99 -7	-4.23 -6	4937
1tc.256	256|1313	-6.324050 +1	-6.324050 +1	64	-9.33 -8	9.99 -7	7.02 -7	-1.46 -6	1881
1tc.512	512|3265	-1.125343 +2	-1.125343 +2	418	-4.98 -8	9.99 -7	7.31 -7	-2.65 -6	2535
1zc.1024	1024|16641	-1.280043 +2	-1.280043 +2	1201	-1.18 -9	7.79 -7	9.11 -7	-1.79 -5	1141
1zc.128	128|1121	-2.066660 +1	-2.066660 +1	2	-6.76 -9	2.49 -7	6.96 -7	+4.40 -7	217
1zc.256	256|2817	-3.733370 +1	-3.733370 +1	6	-1.66 -8	8.83 -7	7.85 -7	-5.92 -6	209
1zc.512	512|6913	-6.800050 +1	-6.800050 +1	82	-1.37 -8	8.84 -7	6.20 -7	-4.14 -6	550
2dc.1024	1024|169163	-1.771020 +1	-1.771020 +1	2630	-3.91 -9	7.72 -7	1.00 -6	-8.60 -6	1974
2dc.2048	2048|504452	-2.878750 +1	-2.878750 +1	14410	-3.96 -9	8.80 -7	9.94 -7	-1.44 -5	1421
2dc.512	512|54896	-1.138370 +1	-1.138370 +1	330	-4.55 -9	9.99 -7	8.65 -7	-1.23 -5	1809
G43	1000|9991	-2.797368 +2	-2.797368 +2	764	-3.24 -8	9.99 -7	6.77 -7	-1.35 -6	622
G44	1000|9991	-2.797466 +2	-2.797466 +2	762	-4.32 -8	8.01 -7	9.93 -7	-7.53 -7	660
G45	1000|9991	-2.793182 +2	-2.793182 +2	719	-4.58 -8	9.94 -7	6.64 -7	-1.08 -6	658
G46	1000|9991	-2.790329 +2	-2.790329 +2	711	-1.05 -7	7.93 -7	9.94 -7	-6.34 -7	645
G47	1000|9991	-2.808921 +2	-2.808921 +2	669	-5.30 -8	8.34 -7	9.86 -7	-6.03 -7	651
G51	1000|5910	-3.490001 +2	-3.490001 +2	5599	-3.52 -7	9.99 -7	7.86 -7	-3.27 -7	4475
G52	1000|5917	-3.483865 +2	-3.483865 +2	7951	-2.93 -7	8.68 -7	9.97 -7	-8.46 -7	6260
G53	1000|5915	-3.482135 +2	-3.482135 +2	20045	-7.63 -8	1.00 -6	7.07 -7	-3.04 -6	14407
G54	1000|5917	-3.410003 +2	-3.410003 +2	3044	-1.54 -7	9.93 -7	9.43 -7	-1.13 -6	2561
brock200-1	200|5067	-2.719680 +1	-2.719680 +1	6	-4.59 -8	9.60 -7	8.12 -7	-7.39 -7	228
brock200-4	200|6812	-2.112110 +1	-2.112110 +1	5	-6.21 -8	9.56 -7	8.20 -7	-9.87 -7	210
brock400-1	400|20078	-3.933100 +1	-3.933100 +1	26	-3.42 -8	9.76 -7	9.31 -7	-1.15 -6	240
c-fat200-1	200|18367	-1.200000 +1	-1.200000 +1	4	-3.81 -8	9.42 -7	9.02 -7	+1.60 -8	248
hamming-10-2	1024|23041	-8.533720 +1	-8.533720 +1	1041	-6.68 -11	8.06 -7	9.00 -7	-2.36 -5	882
hamming-7-5-6	128|1793	-3.600060 +1	-3.600060 +1	5	-3.14 -10	8.73 -7	7.52 -7	-7.91 -6	611
hamming-8-3-4	256|16129	-2.559990 +1	-2.559990 +1	5	-1.25 -9	2.11 -7	9.09 -7	+4.00 -7	191
hamming-8-4	256|11777	-1.600030 +1	-1.600030 +1	5	-1.97 -9	9.49 -7	3.73 -7	-1.09 -5	193
hamming-9-5-6	512|53761	-5.866730 +1	-5.866730 +1	80	-3.92 -10	5.32 -7	9.98 -7	-6.31 -6	554
hamming-9-8	512|2305	-2.240014 +2	-2.240014 +2	137	+4.13 -9	6.22 -7	8.37 -7	-3.92 -6	1053
keller4	171|5101	-1.346590 +1	-1.346590 +1	7	-3.90 -8	7.56 -7	9.42 -7	-6.18 -7	435
p-hat300-1	300|33918	-1.002020 +1	-1.002020 +1	34	-6.61 -8	9.98 -7	6.71 -7	-1.24 -6	566
san200-0.7-1	200|5971	-3.000000 +1	-3.000000 +1	3	-1.25 -8	1.81 -7	8.11 -7	-1.14 -6	121
sanr200-0.7	200|6033	-2.363330 +1	-2.363330 +1	6	-5.84 -8	9.68 -7	7.90 -7	-7.21 -7	222
theta10	500|12470	-8.314920 +1	-8.314920 +1	56	-4.55 -8	9.72 -7	8.88 -7	-9.28 -7	317
theta102	500|37467	-3.806640 +1	-3.806640 +1	43	-1.44 -8	9.68 -7	9.24 -7	-1.42 -6	235
theta103	500|62516	-2.237750 +1	-2.237750 +1	43	-8.48 -9	9.84 -7	9.80 -7	-1.87 -6	231
theta104	500|87245	-1.328270 +1	-1.328270 +1	50	-5.12 -9	9.26 -7	9.99 -7	-2.32 -6	246
theta12	600|17979	-9.209110 +1	-9.209110 +1	92	-4.67 -8	9.01 -7	9.87 -7	-9.64 -7	319
theta123	600|90020	-2.449520 +1	-2.449520 +1	78	-6.76 -9	8.85 -7	9.97 -7	-1.81 -6	239
theta32	150|2286	-2.729160 +1	-2.729160 +1	4	-1.15 -7	8.45 -7	9.78 -7	-3.43 -7	273
theta4	200|1949	-4.986910 +1	-4.986910 +1	8	-1.99 -7	8.62 -7	9.70 -7	-5.65 -7	355
theta42	200|5986	-2.373820 +1	-2.373820 +1	6	-1.15 -7	9.88 -7	6.86 -7	-6.18 -7	235
theta6	300|4375	-6.296190 +1	-6.296190 +1	18	-1.05 -7	9.93 -7	7.69 -7	-6.13 -7	322
theta62	300|13390	-2.937800 +1	-2.937800 +1	13	-3.26 -8	8.59 -7	9.87 -7	-9.08 -7	232
theta8	400|7905	-7.340800 +1	-7.340800 +1	36	-5.61 -8	9.98 -7	7.80 -7	-8.48 -7	318
theta82	400|23872	-3.406440 +1	-3.406440 +1	25	-2.09 -8	9.74 -7	9.15 -7	-1.25 -6	231
theta83	400|39862	-2.016720 +1	-2.016720 +1	25	-1.44 -8	9.70 -7	9.67 -7	-1.63 -6	225

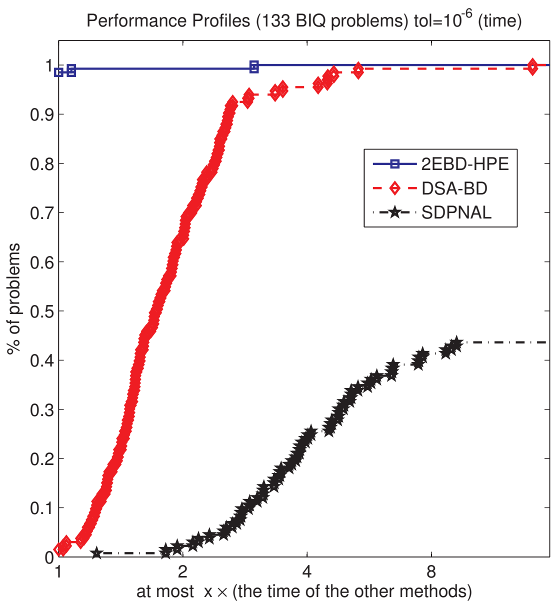

Figure 9: Performance profiles of 2EBD-HPE, the BD method in [10] and SDPNAL for solving 133 SDP relaxations of BIQ problems with accuracy $\bar{\epsilon} = 10^{-6}$.

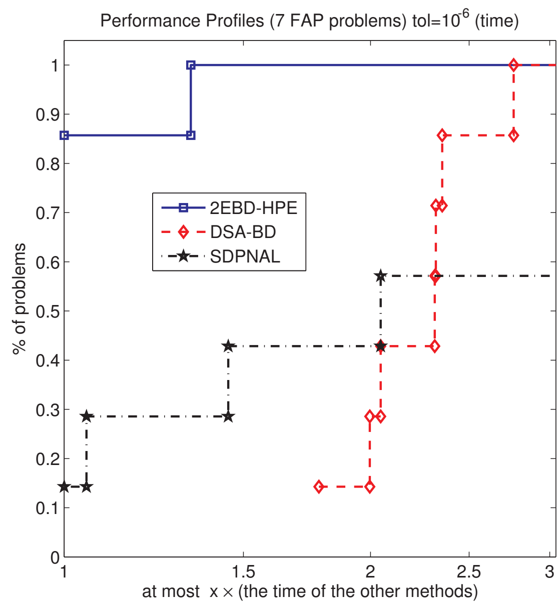

Figure 10: Performance profiles of 2EBD-HPE, the BD method in [10] and SDPNAL for solving 7 SDP relaxations of FAPs with accuracy $\bar{\epsilon} = 10^{-6}$.

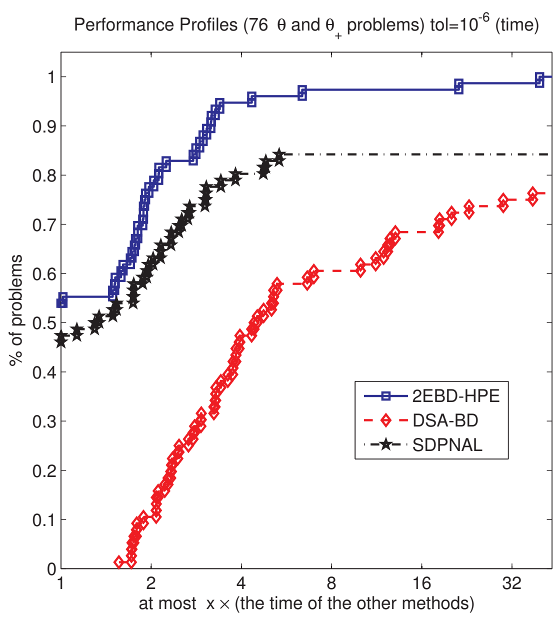

Figure 11: Performance profiles of 2EBD-HPE, the BD method in [10] and SDPNAL for solving 76 $\theta(G)$ and $\theta_+(G)$ problems with accuracy $\bar{\epsilon} = 10^{-6}$.

9 Concluding remarks

Note that when applying the A-BD-HPE framework to (23), it is necessary to first specify the first and second blocks, namely $0 \in F_1(x, y) + A(x)$ and $0 \in F_2(x, y) + B(x)$, respectively. We have seen that Algorithm 1 corresponds to applying the A-BD-HPE framework to (23) by choosing the first and second blocks to be the first and second inclusions in (23), respectively. Clearly, a variant of Algorithm 1 can be obtained by changing the choice of the first and second blocks to be the second and first inclusions in (23), respectively. The resulting method can be easily shown to possess similar convergence properties as those of Algorithm 1. We observe that $\tilde{\lambda}$ for this variant should be chosen as

$$\tilde{\lambda} := \min\left\{\frac{\sigma_1^2}{\theta L}, \frac{(\sigma^2 - \sigma_1^2)^{1/2}}{\sqrt{\theta}}\right\}.$$

The approach in Section 4 can be easily extended to the convex problem

$$\begin{array}{ll} \min & f(x) + \sum_{i=0}^{m} h_i(x) \\ \text{s.t.} & x \in \mathcal{X}, \end{array} \tag{67}$$

which is equivalent to solving the inclusion problem

$$\begin{aligned} &0 \in \nabla f(x) + \partial h_0(x) + \sum_{i=1}^{m} y_i, \\ &0 \in \theta_i[-x + \partial h_i^*(y_i)], \quad i = 1, \dots, m, \end{aligned}$$

where $\theta_i > 0$, $i = 1, \dots, m$, are scaling factors. Even though, this inclusion system has $m + 1$ blocks of inclusions, it can be viewed as having two blocks for the purpose of applying the A-BD-HPE framework to it. Indeed, the first block would be the first inclusion and the second block would consist of the other m inclusions. Note that once $\tilde{x}_k$ is obtained from the proximal equation associated with the first block, it can be updated in the proximal equations corresponding to the other inclusions, and the $\tilde{y}_{i,k}$ can all be computed simultaneously. Convergence results similar to the ones obtained in Section 4 can be derived for (67).

References

[1] R. S. Burachik, Alfredo N. Iusem, and B. F. Svaiter. Enlargement of monotone operators with applications to variational inequalities. *Set-Valued Analysis*, 5:159–180, 1997. 10.1023/A:1008615624787. URL: `http://dx.doi.org/10.1023/A:1008615624787`.

[2] R. S. Burachik and B. F. Svaiter. Maximal monotone operators, convex functions and a special family of enlargements. *Set-Valued Analysis*, 10:297–316, 2002. 10.1023/A:1020639314056. URL: `http://dx.doi.org/10.1023/A:1020639314056`.

[3] S. Burer, R. D. C. Monteiro, and Y. Zhang. A computational study of a gradient-based log-barrier algorithm for a class of large-scale SDPs. *Mathematical Programming*, 95:359–379, 2003. 10.1007/s10107-002-0353-7. URL: `http://dx.doi.org/10.1007/s10107-002-0353-7`.

[4] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40:120–145, 2011. 10.1007/s10851-010-0251-1. URL: `http://dx.doi.org/10.1007/s10851-010-0251-1`.

[5] E. D. Dolan and J. J. Moré. Benchmarking optimization software with performance profiles, January 2002. URL: `http://dx.doi.org/http://dx.doi.org/doi:10.1007/s101070100263`, http://dx.doi.org/http://dx.doi.org/doi:10.1007/s101070100263 `doi:http://dx.doi.org/doi:10.1007/s101070100263`.

[6] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers Mathematics with Applications*, 2(1):17 – 40, 1976. URL: http://www.sciencedirect.com/science/article/pii/0898122176900031, doi:10.1016/0898-1221(76)90003-1.

[7] R. Glowinski and A. Marrocco. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par penalisation-dualité, d'une classe de problèmes de dirichlet non linéaires. *RAIRO Anal. Numér.*, 2:41 – 76, 1975.

[8] S. Ma, W. Yin, Y. Zhang, and A. Chakraborty. An efficient algorithm for compressed mr imaging using total variation and wavelets. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1 –8, june 2008. doi:10.1109/CVPR.2008.4587391.

[9] J. Malick, J. Povh, F. Rendl, and A. Wiegele. Regularization methods for semidefinite programming. *SIAM J. on Optimization*, 20(1):336–356, April 2009. URL: http://dx.doi.org/10.1137/070704575, http://dx.doi.org/10.1137/070704575 doi:10.1137/070704575.

[10] R. D. C. Monteiro, C. Ortiz, and B. F. Svaiter. Implementation of a block-decomposition algorithm for solving large-scale conic semidefinite programming problems. *Optimization-online preprint 3032*, pages 1–32, 2011. URL: http://www.optimization-online.org/DB_HTML/2011/05/3032.html.

[11] R. D. C. Monteiro and B. F. Svaiter. Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers. *Optimization-online preprint 2713*, pages 1–28, 2010. URL: http://www.optimization-online.org/DB_HTML/2010/08/2713.html.

[12] R. D. C. Monteiro and B. F. Svaiter. On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. *SIAM Journal on Optimization*, 20(6):2755–2787, 2010. URL: http://link.aip.org/link/?SJE/20/2755/1, http://dx.doi.org/10.1137/090753127 doi:10.1137/090753127.

[13] R. D. C. Monteiro and B. F. Svaiter. Complexity of variants of tseng's modified f-b splitting and korpelevich's methods for hemivariational inequalities with applications to saddle-point and convex optimization problems. *SIAM Journal on Optimization*, 21(4):1688–1720, 2011. URL: http://link.aip.org/link/?SJE/21/1688/1, http://dx.doi.org/10.1137/100801652 doi:10.1137/100801652.

[14] J. Povh, F. Rendl, and A. Wiegele. A boundary point method to solve semidefinite programs. *Computing*, 78:277–286, 2006. 10.1007/s00607-006-0182-2. URL: http://dx.doi.org/10.1007/s00607-006-0182-2.

[15] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, NJ, 1970.

[16] R. T. Rockafellar. On the maximal monotonicity of subdifferential mappings. *Pacific J. Math.*, 33:209–216, 1970.

[17] M. V. Solodov and B. F. Svaiter. A hybrid approximate extragradient–proximal point algorithm using the enlargement of a maximal monotone operator. *SetValued Analysis*, 7(4):323–345, 1999.

[18] B. F. Svaiter. A family of enlargements of maximal monotone operators. *Set-Valued Analysis*, 8:311–328, 2000. 10.1023/A:1026555124541. URL: http://dx.doi.org/10.1023/A:1026555124541.

[19] Z. Wen, D. Goldfarb, and W. Yin. Alternating direction augmented lagrangian methods for semidefinite programming. *Mathematical Programming Computation*, 2:203–230, 2010. 10.1007/s12532-010-0017-1. URL: http://dx.doi.org/10.1007/s12532-010-0017-1.

[20] X.-Y. Zhao, D. Sun, and K.-C. Toh. A Newton-CG augmented lagrangian method for semidefinite programming. *SIAM Journal on Optimization*, 20(4):1737–1765, 2010. URL: http://link.aip.org/link/?SJE/20/1737/1, http://dx.doi.org/10.1137/080718206 doi:10.1137/080718206.