Bregman primal–dual methods for conic linear optimization via exact-penalty and self-dual formulations

Xin Jiang* Justin Shao† Lieven Vandenberghe‡

November 15, 2026

Abstract

We study Bregman primal–dual proximal methods for conic linear programs (LPs). For problems involving non-polyhedral cones, Bregman distances lead to generalized projections that are often simpler than Euclidean projections. In this paper, we reformulate conic LPs as equivalent exact-penalty problems over truncated cones, and use self-dual embeddings to construct such formulations systematically. This approach yields Bregman proximal methods with simpler subproblems and convergence guarantees over bounded domains. We also study line searches for the Bregman primal–dual hybrid gradient and extragradient methods applied to self-dual problems. For logarithmically homogeneous kernels, the Bregman distances from the starting points are unbounded on the truncated cone, and the classical ergodic bounds do not apply. For these kernels, we derive convergence rates for the exact-penalty objective at the averaged iterates of both methods. Numerical experiments with nonnegative trigonometric polynomial optimization and sparse semidefinite optimization indicate the potential of this framework as a general-purpose approach for incorporating efficient Bregman projections in conic linear optimization.

1 Introduction

Primal–dual proximal methods are attractive for large-scale conic linear optimization because they require only matrix–vector products and projections on the primal or dual cones. Bregman variants of these methods are especially appealing because they can exploit problem geometry more effectively than Euclidean projections and thereby reduce the complexity of the proximal updates. However, many Bregman projections of interest are not naturally posed over the entire cone; they are simpler when the cone is truncated, *i.e.*, intersected with a bounding halfspace. Motivated by this observation, we reformulate conic LPs as equivalent exact-penalty problems over truncated cones, and use self-dual embeddings to construct these bounded formulations systematically. This yields Bregman primal–dual proximal methods with simpler subproblems and formulations that better reflect the underlying conic structure.

We discuss primal and dual conic linear programs (LPs) in the form

$$\begin{array}{lll} \text{Primal:} & \text{minimize} & c^T x \\ & \text{subject to} & Ax - b \in L \\ & & x \in K, \end{array} \tag{1a}$$

*Department of Industrial and Systems Engineering, University of Houston. Email: xinjiang@uh.edu

†Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California. Email: jmshao@usc.edu

‡Department of Electrical and Computer Engineering, University of California, Los Angeles. Email: vandenbe@ucla.edu

$$\begin{array}{rll} \text{Dual:} & \text{maximize} & b^T z \\ & \text{subject to} & -A^T z + c \in K^* \\ & & z \in L^*. \end{array} \tag{1b}$$

The variables are $x \in \mathbb{R}^n$ and $z \in \mathbb{R}^m$, and K and L are closed convex cones, not necessarily polyhedral or proper. Two important special cases of the primal problem (1a) are LPs in standard form (K is a proper cone, $L = \{0\}$) and LPs in inequality form with free variables ($K = \mathbb{R}^n$, L is a proper cone).

Exact-penalty formulations Primal–dual proximal algorithms for the conic LPs require at each iteration matrix–vector products with A and A^T, and projections on the cones K and L^*. They are well suited for large structured problems when these operations are inexpensive [7–9, 42, 45]. To reduce the complexity per iteration further, variants that use non-Euclidean projections associated with generalized (Bregman) distances are of interest. Two fundamental types of primal–dual proximal algorithms are the *primal–dual hybrid gradient method* (PDHG) [18, 19] and the *extragradient method* (EG) [41, 49], applied to the primal–dual optimality conditions or the associated saddle-point problem. For both classes of algorithms, extensions that use Bregman distances have been developed [10, 37, 40, 59, 64].

A convenient format for the study of proximal algorithms in convex optimization is the primal–dual pair

$$\begin{array}{rll} \text{Primal:} & \text{minimize} & c^T x + f(x) + g(b - Ax), \\ \text{Dual:} & \text{maximize} & b^T z - g^*(z) - f^*(A^T z - c), \end{array} \tag{2a, 2b}$$

where $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ are closed convex functions, with conjugates f^* and g^*. The primal and dual LPs (1) are a special case of (2) with

$$f = \delta_K, \qquad g = \delta_{-L}, \qquad f^* = \delta_{-K^*}, \qquad g^* = \delta_{L^*}, \tag{3}$$

where δ_C denotes the indicator function of the set C. For this choice of f and g, the primal–dual proximal algorithms require Euclidean or non-Euclidean projections on K and L^*. As an alternative to (3) one can apply the proximal algorithms to an exact-penalty formulation of (1). Assume for simplicity that K and L are proper cones. Suppose strong duality holds for (1), *i.e.*, the primal and dual optimal values are equal, and that there exist primal and dual solutions $x^\star, z^\star$ that satisfy $\hat{w}^T x^\star < 1$ and $\hat{s}^T z^\star < 1$, where $\hat{w} \in \text{int}(K^*)$ and $\hat{s} \in \text{int}(L)$. Then the primal and dual LPs can be shown to be equivalent to

$$\begin{array}{rll} \text{Primal:} & \text{minimize} & c^T x + \delta_S(x) + \delta_T^*(b - Ax), \\ \text{Dual:} & \text{maximize} & b^T z - \delta_T(z) - \delta_S^*(A^T z - c), \end{array} \tag{4a, 4b}$$

where δ_S and δ_T are the indicator functions of the truncated cones

$$S = \{x \in K \mid \hat{w}^T x \leq 1\}, \qquad T = \{z \in L^* \mid \hat{s}^T z \leq 1\}, \tag{5}$$

and δ_S^*, δ_T^* are their conjugates, *i.e.*, the support functions of S and T. The last terms in the objective functions of (4a) and (4b) are exact penalties for the constraints $Ax - b \in L$ and $-A^T z + c \in K^*$ in the LPs (1a) and (1b). The pair of problems (4) is a special case of (2) with

$$f = \delta_S, \qquad g = \delta_T^*, \qquad f^* = \delta_S^*, \qquad g^* = \delta_T. \tag{6}$$

The PDHG or EG algorithms applied to (4) require at each iteration Euclidean or non-Euclidean projections on the sets S and T. The choice (6) has advantages over (3) that are especially important for algorithms based on non-Euclidean distance functions.

- Generalized projections on the truncated cones S and T can be easier to define and compute than projections on K and L^*. When using generalized projections one must be able to guarantee that for every possible value of its arguments, the projection has a unique solution in the interior of the domain of the distance-generating function, as is the case for the squared Euclidean distance function and Euclidean projection on a closed convex set. For a non-Euclidean projection this property depends on the set and the generalized distance [12, 16], and is more difficult to satisfy when the set is unbounded. Specifically, for a conic LP with non-polyhedral cones K and L^*, it may be difficult to find Bregman distances for which projections on the cones K, L^* are uniquely defined and easy to compute. The task is simpler for the sets S, T in (5). The generalized projections discussed in [11, 15, 20, 34, 43] illustrate this. Consider for example the distance generated by a logarithmically homogeneous barrier function for a proper cone K. For this distance, projecting on S requires solving an optimization problem

$$\begin{array}{ll} \text{minimize} & a^T x + \phi(x) \\ \text{subject to} & \hat{w}^T x \leq 1, \end{array}$$

  which has a unique solution for every a. Without the inequality constraint, the projection only exists for $a \in \text{int}(K^*)$.

- The domains $\text{dom}\, f$ and $\text{dom}\, g^*$ in (6) are bounded. Explicit bounds on the feasible sets are an important assumption in the theory of first-order methods [36, p. 125]. For example, the convergence analysis of primal–dual first-order algorithms for (2) typically concludes with the result that

$$\mathcal{L}(x_k, z) - \mathcal{L}(x, z_k) \leq \frac{\alpha}{k}(\|x - x_0\|_2^2 + \|z - z_0\|_2^2) \quad \text{for all } x \in \text{dom}\, f,\ z \in \text{dom}\, g^*, \tag{7}$$

  where $\mathcal{L} = c^T x + f(x) + b^T z - g^*(z) - z^T A x$ and α is some constant (see, *e.g.*, [19] and Section 5). If $\text{dom}\, f$ and $\text{dom}\, g^*$ are bounded, one can maximize the two sides of the inequality over $x \in \text{dom}\, f$ and $z \in \text{dom}\, g^*$ to obtain an $O(1/k)$ convergence rate. However, if the domains are not bounded, it is more difficult to interpret (7) as establishing convergence.

- Most distance-generating functions for cones are strongly convex on bounded sets of the form (5), but not on the entire cone. This is important in convergence analyses of Bregman proximal methods because strong convexity of the distance-generating function is a common assumption [10, 19, 49]. For example, the negative entropy, which generates the relative entropy, is strongly convex on the unit simplex but not on the nonnegative orthant.

- The PDHG and EG algorithms applied to (4) generate primal and dual iterates x_k, z_k that are feasible, which simplifies stopping conditions. On the other hand, when the algorithms are applied to (2), the primal iterates x_k are in K but may not satisfy the constraint $Ax_k - b \in L$, and the dual iterates z_k are in L^* but may not satisfy $-A^T z_k + c \in K^*$.

Self-dual embedding The exact-penalty formulation (4) requires vectors $\hat{w} \in \text{int}(K^*)$, $\hat{s} \in \text{int}(L)$ that satisfy $\hat{w}^T x^\star < 1$ and $\hat{s}^T z^\star < 1$ for some optimal solutions $x^\star, z^\star$. For some applications such bounds can be derived from problem structure. In general, however, they are not available, and this limits the value of the method as a general-purpose approach. Moreover, overly loose bounds (with small vectors $\hat{w}$, $\hat{s}$) may slow down the convergence or cause numerical difficulties. These difficulties can be overcome by applying the exact-penalty formulation to a self-dual embedding of the LPs (1). A self-dual conic LP has the form

$$\begin{array}{ll} \text{minimize} & -b^T x \\ \text{subject to} & Ax - b \in K^* \\ & x \in K, \end{array} \tag{8}$$

where K is a closed convex cone and A is a skew-symmetric matrix. The self-dual property is evident if we note that (8) is a special case of (1) with $b = -c$, $L = K^*$, and $A = -A^T$.

Of the many interesting properties of self-dual LPs, the important one for our purpose is that the optimal set of a strictly consistent self-dual LP is nonempty and bounded, and that bounds on the optimal set are readily available. Assume for simplicity that the cone K in (8) is proper (a self-dual LP over a cone that is not proper can be written as an equivalent self-dual LP over a proper cone; see Appendix A.2), and that $\hat{x}$ is strictly feasible, *i.e.*, $\hat{x} \in \mathrm{int}(K)$ and $\hat{s} = A\hat{x} - b \in \mathrm{int}(K^*)$. Then it can be shown that optimal solutions $x^\star$ satisfy

$$\hat{s}^T x^\star \leq \hat{s}^T \hat{x}. \tag{9}$$

As a consequence, an equivalent exact-penalty formulation of the self-dual LP is

$$\text{minimize} \quad -b^T x + \delta_C(x) + \delta_C^*(b - Ax) \tag{10}$$

where $C = \{x \in K \mid \hat{s}^T x \leq \mu\}$ and $\mu > \hat{s}^T \hat{x}$. Problem (10) is self-dual and an example of the general self-dual convex problem

$$\text{minimize} \quad -b^T x + f(x) + f^*(b - Ax) \tag{11}$$

where f is a closed convex function and A is a skew-symmetric matrix.

An important source of self-dual LPs is the *self-dual embedding* of the conic LPs (1). In the Ye–Todd–Mizuno embedding [66] and its conic generalizations [22, 46], a conic LP is transformed into an equivalent self-dual problem with known strictly feasible points. These points are constructed explicitly from the data of the original problem and yield explicit bounds on the optimal set of the embedded problem. Therefore, by applying the exact-penalty reformulation developed above to the embedded problem, one obtains an equivalent formulation over truncated cones. This is the key property of the self-dual embedding used in this paper.

Contributions The main contribution of this paper is a systematic reformulation technique for applying Bregman primal–dual proximal methods to conic linear programs. We first show that, given suitable bounds, a conic LP can be replaced by an equivalent exact-penalty formulation. We then use self-dual embeddings to obtain the required bounds directly from the problem data, avoiding problem-specific estimates of primal and dual solution sets. The resulting reformulations can be solved by Bregman proximal methods and provide the bounded domains needed in the convergence analysis. The second contribution is a convergence analysis of Bregman PDHG and extragradient methods with line search, applied to self-dual problems. For logarithmically homogeneous kernels, the Bregman distances from the starting points are unbounded on the truncated cone, and the classical ergodic bounds do not apply. We show that the exact-penalty objective at the averaged iterates of both methods converges at the rate $O((\log k)/k)$ for these kernels, and at the rate $O(1/k)$ if the kernel is finite at a solution of the exact-penalty problem. Numerical results on nonnegative trigonometric polynomial optimization and sparse semidefinite optimization suggest the potential of the self-dual exact-penalty formulation as a general-purpose approach for cones with efficient Bregman projections.

Related work Primal–dual proximal algorithms are widely used in convex optimization; see, *e.g.*, [18, 19, 31, 47]. Their use in large-scale linear programming has recently attracted renewed interest, driven in part by the success of GPU implementations of PDHG-type methods [7–9, 45]. Recent work extends this direction to general conic linear optimization [42, 62]. These developments provide important motivation for this paper, but our focus is different: we consider non-polyhedral conic linear programs and use non-Euclidean proximal updates to exploit the structure of the relevant conic constraints.

More specifically, this paper is related to Bregman primal–dual proximal methods, which include the Bregman proximal point method, mirror-prox, and Bregman PDHG [10, 26, 29, 36, 37, 40, 49, 59, 64]. For conic linear optimization, generalized projections on truncated cones can often be computed efficiently, even when the Euclidean projection on the cone is expensive [15, 20, 34, 35]. The truncation of cones requires bounds on the primal and dual optimal sets, which in this paper are obtained from the problem data by a self-dual embedding.

Self-dual optimization problems have been studied since the 1950s [24, 25, 27, 30]. The paper by Duffin [25] is an early reference on self-dual conic LPs with non-polyhedral cones. Self-dual embeddings were introduced by Goldman and Tucker [27] and later became a standard tool for initialization and infeasibility detection in interior-point methods [22, 33, 46, 61, 63, 65, 66]. They have also been used with operator splitting methods and primal–dual proximal methods for conic optimization [44, 51, 55]. The present paper uses self-dual embeddings for a different purpose. Rather than using the embedding primarily to initialize an algorithm or to detect infeasibility, we use it to produce the bounded formulations needed by Bregman primal–dual proximal methods.

Adaptive stepsizes and line searches for primal–dual proximal methods have been studied in both Euclidean and Bregman settings. For Euclidean PDHG, such methods are proposed in [28, 47], and PDLP uses a similar adaptive rule heuristically [7, 8]. An extension of the Malitsky–Pock line search to a Bregman dual update is developed in [17]. Yazdandoost Hamedani and Aybat [64] analyze a line search for Bregman PDHG with separate primal and dual stepsizes, while the methods in [34, 35] combine a Bregman primal update with a Euclidean dual update. The line search for PDHG considered here uses a common kernel and stepsize in the primal and dual updates of a self-dual problem. For the extragradient method we use the line search of Tseng [59, section 5]. The relation between the line-search techniques is discussed in Section 5.1.

Barrier functions have also been used as distance-generating functions or regularizers in the barrier subgradient method [50], in online learning [1], and in (proximal) mirror descent methods [23]. The analysis in Section 5.3 follows [1, 50] in approximating boundary points by interior points.

Paper outline The rest of the paper is organized as follows. Section 2 reviews duality and optimality conditions for the problems (1) and (2), and their specialization to self-dual problems. Section 3 develops the bounded exact-penalty reformulation for conic LPs and establishes its equivalence to the original primal–dual pair. Section 4 discusses Bregman primal–dual proximal methods for the resulting formulations. Section 5 presents line searches for the Bregman PDHG and extragradient methods on self-dual problems, and analyzes their convergence for the exact-penalty formulation of a self-dual conic LP. Section 6 reports numerical experiments, and Section 7 concludes the paper. Appendix A extends the construction to mixed-form conic LPs, Appendix B derives the corresponding formulation from the Ye–Todd–Mizuno self-dual embedding, and Appendix C contains the proofs of the convergence results in Section 5.

2 Duality and optimality conditions

In this section we first review the duality and optimality theory for the problems (1) and (2). These duality results are covered in [53, section 31] [54, section 11.H]. We then apply the duality results to self-dual problems.

2.1 Duality

The primal problem (2a) is called *strongly consistent* if

$$b \in \operatorname{ri}(A \operatorname{dom} f + \operatorname{dom} g) = A \operatorname{ri}(\operatorname{dom} f) + \operatorname{ri}(\operatorname{dom} g), \tag{12}$$

where ri stands for relative interior. The equality in (12) follows from [53, theorem 6.6]. The problem is *strictly consistent* if it is strongly consistent and the set $A \operatorname{dom} f + \operatorname{dom} g$ has nonempty interior, *i.e.*, if

$$b \in \operatorname{int}(A \operatorname{dom} f + \operatorname{dom} g). \tag{13}$$

Note that $A \operatorname{int}(\operatorname{dom} f) + \operatorname{int}(\operatorname{dom} g) \subseteq \operatorname{int}(A \operatorname{dom} f + \operatorname{dom} g)$ but equality does not hold in general.

Theorem 1 ([53, corollary 31.2.1], [54, theorem 11.39])

- *If problem (2a) is strongly consistent, then strong duality holds: problems (2a) and (2b) have the same optimal value, which, for a strongly consistent problem, can be finite or $-\infty$. If problem (2a) is strongly consistent and the optimal value is finite, then the dual optimum is attained.*
- *The optimal set of the dual problem (2b) is nonempty and bounded if and only if the optimal value is finite and the primal problem (2a) is strictly consistent.*

Given the symmetry between (2a) and (2b), dual counterparts of these statements also hold. For the dual problem, strong consistency means that

$$c \in \operatorname{ri}(A^T \operatorname{dom} g^* - \operatorname{dom} f^*) = A^T \operatorname{ri}(\operatorname{dom} g^*) - \operatorname{ri}(\operatorname{dom} f^*)$$

and strict consistency that $c \in \operatorname{int}(A^T \operatorname{dom} g^* - \operatorname{dom} f^*)$.

For the LP (1a), which is the special case with $f = \delta_K$, $g = \delta_{-L}$, strict consistency (13) requires

$$b \in \operatorname{int}(AK - L). \tag{14}$$

This holds if and only if the following two conditions are satisfied.

1. The affine hull
$$\operatorname{span}(AK - L) = A \operatorname{span}(K) + \operatorname{span}(L)$$
has dimension m (the row dimension of A). For a cone C, the affine hull is $C + (-C)$, the subspace spanned by C, denoted here by $\operatorname{span}(C)$. The span of C is also the smallest subspace containing C. The subspace $A \operatorname{span}(K) + \operatorname{span}(L)$ has dimension m if and only if the orthogonal complement has dimension zero, *i.e.*, if and only if
$$\left.\begin{array}{l} y \in \operatorname{lin}(L^*) \\ A^T y \in \operatorname{lin}(K^*) \end{array}\right\} \quad \Longleftrightarrow \quad y = 0. \tag{15}$$
Here $\operatorname{lin}(C) = C \cap (-C)$ is the *lineality space* of a convex cone C, *i.e.*, the largest subspace contained in C. We used the property that $\operatorname{lin}(C) = \operatorname{span}(C^*)^\perp$.

2. *Strong consistency:* $b \in \operatorname{ri}(AK - L) = A \operatorname{ri}(K) - \operatorname{ri}(L)$. This means there exists an $\hat{x}$ that satisfies
$$\hat{x} \in \operatorname{ri}(K), \qquad A\hat{x} - b \in \operatorname{ri}(L).$$

For LPs in standard form or inequality form, strict consistency entails strict feasibility and a rank condition. Appendix A describes the details for a mixed-form LP, with cones of the form $K = K_1 \times \mathbb{R}^{n_2}$ and $L = L_1 \times \{0\}$ where K_1, L_1 are proper cones.

2.2 Optimality conditions

The algorithms discussed later in the paper are based on the following characterization of primal and dual optimal solutions.

Theorem 2 ([53, theorem 31.3], [54, theorem 11.39]) *Strong duality holds for the problems (2) and x, z are optimal for (2a) and (2b), respectively, if and only if*

$$A^T z - c \in \partial f(x), \qquad b - Ax \in \partial g^*(z). \tag{16}$$

Combined with Theorem 1 this theorem implies that if the primal problem (2a) is strongly consistent, then x is an optimal solution if and only if there exists a z that together with x satisfies (16).

The optimality conditions (16) can be expressed in several equivalent ways, which form the basis of different types of algorithms.

- *Monotone inclusion.* The conditions (16) form a *monotone inclusion*

  $$0 \in F(y) + \partial h(y) \tag{17}$$

  where $y = (x, z)$,

  $$F(y) = \begin{bmatrix} 0 & -A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} + \begin{bmatrix} c \\ -b \end{bmatrix}, \qquad h(y) = f(x) + g^*(z), \tag{18}$$

  and $\partial h(y) = \partial f(x) \times \partial g^*(z)$ is the subdifferential of the convex function h. The mapping F is monotone (affine with skew-symmetric linear term), but not the gradient of a convex function (unless $A = 0$).

- *Convex–concave saddle-point problem.* The primal and dual problems in (2) can be written as

  $$\text{Primal:} \qquad \text{minimize} \quad \sup_{\tilde{z}} \mathcal{L}(x, \tilde{z}) \tag{19a}$$

  $$\text{Dual:} \qquad \text{maximize} \quad \inf_{\tilde{x}} \mathcal{L}(\tilde{x}, z) \tag{19b}$$

  where $\mathcal{L}$ is the convex–concave function

  $$\mathcal{L}(x, z) = c^T x + f(x) + b^T z - g^*(z) - z^T A x.$$

  The inequality

  $$\sup_{\tilde{z}} \mathcal{L}(x, \tilde{z}) \geq \mathcal{L}(x, z) \geq \inf_{\tilde{x}} \mathcal{L}(\tilde{x}, z) \tag{20}$$

  holds for all $x \in \operatorname{dom} f$ and $z \in \operatorname{dom} g^*$, without assumptions on f and g. Strong duality holds for the problems (2) and x, z are optimal for (2a) and (2b), respectively, if and only if

  $$\sup_{\tilde{z}} \mathcal{L}(x, \tilde{z}) = \mathcal{L}(x, z) = \inf_{\tilde{x}} \mathcal{L}(\tilde{x}, z). \tag{21}$$

  This last condition is therefore equivalent to (16).

- *Generalized variational inequality.* In a *generalized variational inequality* problem one is given a closed convex function h and a mapping F, and seeks a vector y such that

$$h(\tilde{y}) - h(y) + F(y)^T(\tilde{y} - y) \geq 0 \quad \text{for all } \tilde{y} \in \operatorname{dom} h. \tag{22}$$

  This generalizes a (standard) variational inequality problem, in which h is the indicator function of a convex set C, and (22) can be written as

$$F(y)^T(\tilde{y} - y) \geq 0 \quad \text{for all } \tilde{y} \in C. \tag{23}$$

  The variational inequality problem is monotone if F is a monotone mapping.

  The condition (22) is the subgradient inequality that defines the subdifferential $\partial h(y)$ in (17). This shows that (16) is equivalent to a generalized variational inequality with F and h defined in (18). Using the skew-symmetric structure in F the generalized variational inequality problem can also be expressed as follows: find x, z such that

$$\mathcal{L}(\tilde{x}, z) - \mathcal{L}(x, \tilde{z}) \geq 0 \quad \text{for all } \tilde{x} \in \operatorname{dom} f, \tilde{z} \in \operatorname{dom} g^*,$$

  i.e., $\inf_{\tilde{x}} \mathcal{L}(\tilde{x}, z) \geq \sup_{\tilde{z}} \mathcal{L}(x, \tilde{z})$. Since strict inequality would contradict (20), this is the same as (21).

The optimality conditions for the LPs (1) follow by applying the general results to the cone indicator functions $f = \delta_K$ and $g^* = \delta_{L^*}$. The subdifferentials are $\partial f(x) = N_K(x)$ and $\partial g^*(z) = N_{L^*}(z)$, where N_C denotes the *normal cone* to a closed convex cone C, defined as

$$N_C(x) = \begin{cases} \emptyset & x \notin C \\ \{y \in -C^* \mid x^T y = 0\} & x \in C. \end{cases}$$

The optimality conditions (16) for the LPs (1) can therefore be expressed as a conic *linear complementarity problem*

$$\begin{bmatrix} 0 & -A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} + \begin{bmatrix} c \\ -b \end{bmatrix} = \begin{bmatrix} w \\ s \end{bmatrix}, \qquad x \in K, \quad z \in L^*, \quad s \in L, \quad w \in K^*, \qquad w^T x = s^T z = 0.$$

A convenient shorthand is

$$K \ni x \perp (c - A^T z) \in K^*, \qquad L \ni (Ax - b) \perp z \in L^*. \tag{24}$$

2.3 Self-dual problem

Problem (2a) is *self-dual* if $A = -A^T$, $b = -c$, and $g = f^*$. In that case the primal problem is formally the same as the dual (2b), except for the orientation of the objective. The LP (1a) is self-dual if $A = -A^T$, $b = -c$, and K, L are dual cones.

For a self-dual problem, the optimal value of the dual problem (2b) is the negative of the optimal value of the primal (2a). Therefore the optimal value is nonnegative (possibly $+\infty$ for an infeasible problem) and strong duality holds if and only if the optimal value is zero. We also note that if x is primal optimal, then one can take $z = x$ in the optimality conditions (16). The following corollary summarizes the implications of Theorems 1 and 2 for a self-dual problem.

Corollary 1 *Assume $A = -A^T$ in (11) and f is a closed convex function.*

- *If problem (11) is strongly consistent, then the optimal value is zero and the optimum is attained.*
- *The optimal set is nonempty and bounded if and only if the problem is strictly consistent.*
- *The optimal value is zero and $x, \tilde{x}$ are optimal if and only if*

$$b - Ax \in \partial f(\tilde{x}), \qquad b - A\tilde{x} \in \partial f(x).$$

- *The optimal value is zero and x is optimal if and only if*

$$b - Ax \in \partial f(x). \tag{25}$$

The optimality condition (25) is a monotone inclusion. The other equivalent optimality statements listed in Section 2.2 also simplify for a self-dual problem.

- *Convex–concave saddle-point problem.* For the self-dual problem,

$$\mathcal{L}(x, \tilde{x}) = b^T(\tilde{x} - x) + f(x) - f(\tilde{x}) - \tilde{x}^T A x. \tag{26}$$

  Note that $\mathcal{L}(x, \tilde{x}) = -\mathcal{L}(\tilde{x}, x)$ and $\mathcal{L}(x, x) = 0$. The optimal value of (11) is zero and x is optimal if and only if

$$\sup_{\tilde{x}} \mathcal{L}(x, \tilde{x}) = 0 = \inf_{\tilde{x}} \mathcal{L}(\tilde{x}, x). \tag{27}$$

- *Generalized variational inequality.* For the self-dual problem, the generalized variational inequality (22) simplifies to the condition

$$f(\tilde{x}) - f(x) - (b - Ax)^T(\tilde{x} - x) \geq 0 \quad \text{for all } \tilde{x}. \tag{28}$$

  This is the subgradient inequality that defines the subdifferential in (25). The optimal value of (11) is zero and x is optimal if and only if (28) holds.

An important example of a self-dual problem (11) is the LP (8) with $A = -A^T$. This corresponds to $f = \delta_K$ and $f^* = \delta_{-K^*}$ in (11), and

$$\mathcal{L}(x, \tilde{x}) = b^T(\tilde{x} - x) - \tilde{x}^T A x + \delta_K(x) - \delta_K(\tilde{x}). \tag{29}$$

The problem is strictly consistent if $b \in \operatorname{int}(AK - K^*)$. From the discussion in Section 2.1, this holds if and only if the problem is strongly consistent, *i.e.*, there exists an $\hat{x}$ that satisfies

$$\hat{x} \in \operatorname{ri}(K), \qquad A\hat{x} - b \in \operatorname{ri}(K^*), \tag{30}$$

and the subspace

$$\operatorname{span}(AK - K^*) = A \operatorname{span}(K) + \operatorname{span}(K^*) = A \operatorname{span}(K) + \operatorname{lin}(K)^\perp$$

has dimension n. Equivalently, specializing (15) to the self-dual LP,

$$\left.\begin{array}{l} y \in \operatorname{lin}(K) \\ Ay \in \operatorname{lin}(K^*) \end{array}\right\} \qquad \Longleftrightarrow \qquad y = 0. \tag{31}$$

A strictly consistent self-dual LP (8) satisfies the following properties.

- In the optimality condition (25), $\partial f(x)$ is the normal cone $N_K(x) = \{y \in -K^* \mid y^T x = 0\}$ to K at x. Hence x is optimal if and only if x and $s = Ax - b$ satisfy the conic linear complementarity conditions

$$x \in K, \qquad s \in K^*, \qquad s^T x = 0.$$

- The optimal value is zero and the optimal set is nonempty and bounded. Consider any x and define $s = Ax - b$. Let $\hat{x}$ be any point that satisfies (30) and define $\hat{s} = A\hat{x} - b$. Taking the inner product with $\hat{x} - x$ on both sides of the equation $A(\hat{x} - x) = \hat{s} - s$ shows that $(\hat{s} - s)^T(\hat{x} - x) = 0$. If x is optimal, $s^T x = 0$ and

$$\hat{s}^T x + \hat{x}^T s = \hat{s}^T \hat{x}. \tag{32}$$

  The right-hand side is positive because $\hat{x} \in \mathrm{ri}(K)$ and $\hat{s} \in \mathrm{ri}(K^*)$.

  The fact that all optimal solutions satisfy (32) gives a simple direct proof that the optimal set of (8) is bounded. We note the following fact. If C is a closed convex cone, then

$$\left.\begin{array}{l} u \in C \\ w \in \mathrm{ri}(C^*) \\ w^T u = 0 \end{array}\right\} \qquad \Longrightarrow \qquad u \in \mathrm{lin}(C).$$

  To show that the optimal set of (8) is bounded, suppose $x = x^\star + \alpha y$ is optimal for all $\alpha \geq 0$. Then (32) requires that

$$y \in K, \qquad Ay \in K^*, \qquad \hat{s}^T y = 0, \qquad \hat{x}^T A y = 0.$$

  The conditions $y \in K$, $\hat{s}^T y = 0$ imply $y \in \mathrm{lin}(K)$ because $\hat{s} \in \mathrm{ri}(K^*)$. The conditions $Ay \in K^*$, $\hat{x}^T A y = 0$ imply $Ay \in \mathrm{lin}(K^*)$ because $\hat{x} \in \mathrm{ri}(K)$. By (31), y is zero.

Strict consistency and its implications are especially simple if K is a proper cone. In that case, the problem is strictly consistent if and only if there exists a strictly feasible $\hat{x}$, *i.e.*, a point that satisfies $\hat{x} \in \mathrm{int}(K)$ with $A\hat{x} - b \in \mathrm{int}(K^*)$. Note that properness of K can be assumed without loss of generality. When the cone is not proper, one may make its lineality space and span explicit and reduce the problem to an equivalent self-dual LP over a proper cone; see Appendix A.

3 Exact-penalty formulations

We now apply the results of the previous section to derive the exact-penalty formulations (4) and (10). Exact-penalty formulations have a long history, reviewed in [14]. In the context of primal–dual proximal methods, they have appeared in convergence analyses and stopping criteria [34, 35, 56]. In this paper, they are used for equivalent problem formulations to which the proximal algorithms can be directly applied.

3.1 Minkowski gauges

The functions δ_T^* and δ_S^* in (4) are examples of *Minkowski gauges*. The Minkowski gauge of a proper cone C in $\mathbb{R}^n$ is the function

$$\begin{aligned} \sigma_C(y, a) &= \inf\{t \geq 0 \mid ta - y \in C\} && (33) \\ &= \sup\{y^T u \mid a^T u \leq 1,\ u \in C^*\} && (34) \end{aligned}$$

and is defined for all $a \in \mathrm{int}(C)$ and $y \in \mathbb{R}^n$. The point a is sometimes called the *pole* of the Minkowski gauge. The equivalence between the expressions (33) and (34) follows from conic duality. The first expression (33) shows that $\sigma_C(y,a) = 0$ if $-y \in C$. Otherwise, $1/\sigma_C(y,a)$ is the steplength from a to the boundary of the cone in the direction $-y$. The second expression (34) shows that $\sigma_C(\cdot,a)$ is the support function of the set $\{u \in C^* \mid a^T u \leq 1\}$. It is therefore also the conjugate of the indicator function of this set. The Minkowski gauge $\sigma_C(\cdot,a)$ is a nonnegative, convex, and positively homogeneous function of its first argument, with domain $\mathbb{R}^n$.

It is common in applications that C is a direct product of proper cones. Suppose $C = C_1 \times \cdots \times C_k$, where $C_1, \ldots, C_k$ are proper. Then the Minkowski gauge for C at $y = (y_1, \ldots, y_k)$ and $a = (a_1, \ldots, a_k) \in \mathrm{int}(C_1) \times \cdots \times \mathrm{int}(C_k)$ is

$$\begin{aligned}
\sigma_C(y,a) &= \inf\{t \geq 0 \mid ta_i - y_i \in C_i,\ i = 1, \ldots, k\} \\
&= \sup\{\sum_{i=1}^k y_i^T u_i \mid \sum_{i=1}^k a_i^T u_i \leq 1,\ u_i \in C_i^*,\ i = 1, \ldots, k\} \\
&= \max\{\sigma_{C_1}(y_1,a_1), \ldots, \sigma_{C_k}(y_k,a_k)\}.
\end{aligned}$$

We also note that if C is a proper cone and $a \in \mathrm{int}(C)$, then

$$\begin{aligned}
\max\{\sigma_C(y,a), \sigma_C(-y,a)\} &= \inf\{t \mid ta - y \in C,\ ta + y \in C\} \\
&= \sup\{y^T(u-v) \mid a^T(u+v) = 1,\ u, v \in C^*\}.
\end{aligned}$$

As a function of y, this is a norm. If we denote the corresponding unit norm ball by B_a, then $B_a + a$ is the largest convex set, symmetric about a and contained in C.

As an example, for $C = \mathbb{R}^m_+$ and a positive m-vector a,

$$\sigma_C(y,a) = \max\{0, \max_{i=1,\ldots,m} y_i/a_i\}, \qquad \max\{\sigma_C(y,a), \sigma_C(-y,a)\} = \max_{i=1,\ldots,m} |y_i|/a_i.$$

3.2 Exact penalties for conic linear optimization

Consider the conic LPs (1) with proper cones K and L. Assume that strong duality holds and that the primal and dual optima are attained. Also suppose we are given points $\hat{w} \in \mathrm{int}(K^*)$ and $\hat{s} \in \mathrm{int}(L)$ that satisfy $\hat{w}^T x^\star < 1$ and $\hat{s}^T z^\star < 1$ for some primal optimal $x^\star$ and dual optimal $z^\star$. Using the conjugate correspondence between Minkowski gauges and indicator functions, we can write the problems (4) as

$$\begin{array}{lll}
\text{Primal:} & \text{minimize} & c^T x + \sigma_L(b - Ax, \hat{s}) \\
& \text{subject to} & x \in K,\ \hat{w}^T x \leq 1
\end{array} \tag{35a}$$

$$\begin{array}{lll}
\text{Dual:} & \text{maximize} & b^T z - \sigma_{K^*}(A^T z - c, \hat{w}) \\
& \text{subject to} & z \in L^*,\ \hat{s}^T z \leq 1.
\end{array} \tag{35b}$$

We will show that these problems have the same optimal value as the primal and dual problems in (1), respectively. Moreover, every solution of (35a) is a solution of the primal LP in (1), and every solution of (35b) is a solution of the dual LP. Conversely, any optimal solution of the original primal or dual LP that satisfies the corresponding imposed bound is optimal for the associated penalized problem. The second term in the objective of (35a) therefore acts as an exact penalty for the constraint $Ax - b \in L$, and the second term in the objective of (35b) acts as an exact penalty for the constraint $-A^T z + c \in K^*$ in the dual LP.

Using the definition (33), we can express the two problems (35) as the pair of primal and dual conic LPs

$$\begin{array}{lll} \text{Primal:} & \text{minimize} & c^T x + \alpha \\ & \text{subject to} & Ax - b + \alpha\hat{s} \in L \\ & & x \in K,\ \hat{w}^T x \leq 1 \\ & & \alpha \geq 0 \end{array} \tag{36a}$$

$$\begin{array}{lll} \text{Dual:} & \text{maximize} & b^T z - \beta \\ & \text{subject to} & -A^T z + c + \beta\hat{w} \in K^* \\ & & z \in L^*,\ \hat{s}^T z \leq 1 \\ & & \beta \geq 0. \end{array} \tag{36b}$$

In the primal problem we added a scalar variable α and in the dual problem a scalar variable β. The equivalence between (35) and (36) follows directly from the definition of the Minkowski gauge. The following observations establish the exactness of these reformulations, (35) and (36), relative to the original conic LP pair (1).

- The problems (35a) and (35b) are duals, and a special case of (2a) and (2b) with $f = \delta_S$, $g = \delta_T^*$, where S and T are defined in (5). The support functions of these sets are $f^* = \delta_S^* = \sigma_{K^*}(\cdot, \hat{w})$ and $g = \delta_T^* = \sigma_L(\cdot, \hat{s})$.

- The problems (35a) and (35b) are strictly consistent. The primal problem (35a) is strictly consistent because $\operatorname{dom} \sigma_L(\cdot, \hat{s}) = \mathbb{R}^m$ and there exist points x that satisfy $x \in \operatorname{int}(K)$ and $\hat{w}^T x < 1$. The dual problem (35b) is strictly consistent because $\operatorname{dom} \sigma_{K^*}(\cdot, \hat{w}) = \mathbb{R}^n$ and there exist points z that satisfy $z \in \operatorname{int}(L^*)$ and $\hat{s}^T z < 1$. It follows from Theorem 1 that strong duality holds for (35a) and (35b), and that the optimum is attained in primal and dual. Theorem 2 provides necessary and sufficient optimality conditions that are satisfied by these optimal solutions.

  The same conclusions hold for the equivalent conic LPs (36). The optimality conditions for these problems are

  $$K \ni x \perp (\beta\hat{w} + c - A^T z) \in K^*, \qquad L \ni (\alpha\hat{s} + Ax - b) \perp z \in L^* \tag{37a}$$

  and

  $$0 \leq \begin{bmatrix} \alpha \\ 1 - \hat{w}^T x \end{bmatrix} \perp \begin{bmatrix} 1 - \hat{s}^T z \\ \beta \end{bmatrix} \geq 0. \tag{37b}$$

- Suppose x, z are optimal for (1) and satisfy the bounds $\hat{w}^T x \leq 1$ and $\hat{s}^T z \leq 1$. Since x is optimal for (1a), it is feasible for (35a) with objective value $c^T x + \sigma_L(b - Ax, \hat{s}) = c^T x$. Since z is optimal for (1b), it is feasible for (35b) with objective value $b^T z - \sigma_{K^*}(A^T z - c, \hat{w}) = b^T z$. By assumption, strong duality holds for (1), so $c^T x = b^T z$. Therefore the two points x, z are optimal for (35), and the optimal value for (35) is equal to the optimal value $c^T x = b^T z$ of (1).

- By assumption, there exists an optimal solution $x^\star$ of (1) that satisfies $\hat{w}^T x^\star < 1$, and an optimal solution $z^\star$ of the dual problem in (1) that satisfies $\hat{s}^T z^\star < 1$. Define $\alpha^\star = \beta^\star = 0$. The optimality conditions (37) show that $(x^\star, \alpha^\star)$ and $(z^\star, \beta^\star)$ are optimal solutions of (36).

  Now suppose x, z are any optimal solutions for (35), and define $\alpha = \sigma_L(b - Ax, \hat{s})$ and $\beta = \sigma_{K^*}(A^T z - c, \hat{w})$. Since (x, α) is optimal for the primal problem (36a), it must satisfy the complementarity conditions (37) with $(z^\star, \beta^\star)$. Hence $\alpha = 0$ and x is feasible for (1). If (z, β) is optimal for the dual

problem (36b), then it must satisfy the complementarity conditions (37) with $(x^\star, \alpha^\star)$. Hence $\beta = 0$ and z is feasible for (1). Moreover $c^T x = b^T z$ because x and z are optimal for (36) with $\alpha = \beta = 0$ and strong duality holds for (36). Since x and z are feasible in (1) and $c^T x = b^T z$ they are optimal for (1).

3.3 Exact-penalty self-dual formulation

We now turn to the self-dual LP (8). Assume for simplicity that K is a proper cone, and that $\hat{x} \in \mathrm{int}(K)$ and $\hat{s} = A\hat{x} - b \in \mathrm{int}(K^*)$. Let $\mu > \hat{s}^T\hat{x}$. Then (32) shows that $\hat{s}^T x < \mu$ for all optimal solutions x. Therefore

$$\begin{array}{ll} \text{minimize} & -b^T x + \mu\sigma_{K^*}(b - Ax, \hat{s}) \\ \text{subject to} & x \in K, \;\; \hat{s}^T x \leq \mu, \end{array} \tag{38}$$

is an exact reformulation of the self-dual LP (8), *i.e.*, x is an optimal solution of (38) if and only if it is a solution of (8). Note that (38) is a self-dual problem of the form (11) and can be written as

$$\text{minimize} \quad -b^T x + \delta_C(x) + \delta_C^*(b - Ax)$$

where $C = \{x \in K \mid \hat{s}^T x \leq \mu\}$. Problem (38) is equivalent to the self-dual LP

$$\begin{array}{ll} \text{minimize} & -b^T x + \mu t \\ \text{subject to} & \left[\begin{array}{cc} A & \hat{s} \\ -\hat{s}^T & 0 \end{array}\right]\left[\begin{array}{c} x \\ t \end{array}\right] - \left[\begin{array}{c} b \\ -\mu \end{array}\right] \in K^* \times \mathbb{R}_+ \\ & (x, t) \in K \times \mathbb{R}_+. \end{array}$$

The objective function in (38) can also be written as

$$\begin{array}{rcl} -b^T x + \mu\sigma_{K^*}(b - Ax, \hat{s}) + \delta_C(x) & = & \sup\limits_{\tilde{x} \in C} \left(-b^T x + (b - Ax)^T \tilde{x} + \delta_C(x)\right) \\ & = & \begin{cases} \sup_{\hat{s}^T\tilde{x} \leq \mu} \mathcal{L}(x, \tilde{x}) & \hat{s}^T x \leq \mu \\ +\infty & \text{otherwise,} \end{cases} \end{array} \tag{39}$$

where $\mathcal{L}$ is defined in (29). The identity (39) will be used in Section 5.3 to show that the saddle-point gap of (38), restricted to C, is the sum of two exact-penalty objective values.

4 Primal–dual proximal methods

In this section we describe Bregman variants of the primal–dual hybrid gradient (PDHG) method and the extragradient (EG) method, and specialize them to the exact-penalty and self-dual formulations developed above. The algorithms themselves are classical or closely related to existing methods.

4.1 Bregman PDHG and extragradient methods

We discuss PDHG and EG applied to problems in the form (2). With fixed stepsizes, both algorithms can be interpreted as variations of the *Bregman proximal point method* [16, 26, 29], applied to the primal–dual optimality conditions (17). Let

$$d_{\mathrm{pd}}(y, \hat{y}) = \phi_{\mathrm{pd}}(y) - \phi_{\mathrm{pd}}(\hat{y}) - \nabla\phi_{\mathrm{pd}}(\hat{y})^T(y - \hat{y})$$

be a Bregman distance, generated by a differentiable convex function $\phi_{\text{pd}} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$. In the Bregman proximal point method applied to (17), the update $y_{k+1} = (x_{k+1}, z_{k+1})$ is defined as the solution of the inclusion problem

$$\begin{aligned} \frac{1}{t}(\nabla\phi_{\text{pd}}(y_k) - \nabla\phi_{\text{pd}}(y_{k+1})) &\in F(y_{k+1}) + \partial h(y_{k+1}) && (40)\\ &= \begin{bmatrix} c - A^T z_{k+1} \\ Ax_{k+1} - b \end{bmatrix} + \partial f(x_{k+1}) \times \partial g^*(z_{k+1}), && (41) \end{aligned}$$

where $y_k = (x_k, z_k)$ and t is a positive stepsize. The position of x_{k+1} and z_{k+1} in the first term on the right-hand side of (41) couples the two blocks in the equation and complicates the computation of y_{k+1}. In this section we discuss two classical approaches for simplifying the update. The two methods will involve primal and dual Bregman distances

$$d_{\text{p}}(x, x') = \phi_{\text{p}}(x) - \phi_{\text{p}}(x') - \nabla\phi_{\text{p}}(x')^T(x - x'), \qquad d_{\text{d}}(z, z') = \phi_{\text{d}}(z) - \phi_{\text{d}}(z') - \nabla\phi_{\text{d}}(z')^T(z - z'), \quad (42)$$

generated by convex functions $\phi_{\text{p}} : \mathbb{R}^n \to \mathbb{R}$ and $\phi_{\text{d}} : \mathbb{R}^m \to \mathbb{R}$, respectively.

Primal–dual hybrid gradient method The Bregman PDHG method can be interpreted as the Bregman proximal point method with a special choice of Bregman distance [19, 31, 34, 35]. Consider the proximal point method with a distance-generating function of the form

$$\phi_{\text{pd}}(x, z) = \phi_{\text{p}}(x) + \phi_{\text{d}}(z) \pm t\, z^T A x. \quad (43)$$

A condition for ϕ_{pd} to be convex is given in (53) below. The Bregman distance generated by ϕ_{pd} is

$$d_{\text{pd}}(y, y') = d_{\text{p}}(x, x') + d_{\text{d}}(z, z') \pm t\,(z - z')^T A(x - x'). \quad (44)$$

If we choose the positive sign for $\pm$ in (44), substitution in (41) gives

$$\begin{aligned} \tfrac{1}{t}(\nabla\phi_{\text{p}}(x_k) - \nabla\phi_{\text{p}}(x_{k+1})) + A^T z_k - c &\in \partial f(x_{k+1}) \\ \tfrac{1}{t}(\nabla\phi_{\text{d}}(z_k) - \nabla\phi_{\text{d}}(z_{k+1})) + b - A(2x_{k+1} - x_k) &\in \partial g^*(z_{k+1}). \end{aligned}$$

This gives a first PDHG variant:

$$\begin{aligned} x_{k+1} &= \operatorname*{argmin}_x \left((c - A^T z_k)^T x + f(x) + \tfrac{1}{t} d_{\text{p}}(x, x_k)\right) && (45a)\\ z_{k+1} &= \operatorname*{argmin}_z \left((A(2x_{k+1} - x_k) - b)^T z + g^*(z) + \tfrac{1}{t} d_{\text{d}}(z, z_k)\right). && (45b) \end{aligned}$$

The iteration (45) consists of a primal update followed by a dual update, and both updates require the evaluation of a generalized proximal operator. With the negative sign in (44), we get

$$\begin{aligned} \tfrac{1}{t}(\nabla\phi_{\text{p}}(x_k) - \nabla\phi_{\text{p}}(x_{k+1})) + A^T(2z_{k+1} - z_k) - c &\in \partial f(x_{k+1}) \\ \tfrac{1}{t}(\nabla\phi_{\text{d}}(z_k) - \nabla\phi_{\text{d}}(z_{k+1})) + b - Ax_k &\in \partial g^*(z_{k+1}). \end{aligned}$$

This is the second variant of PDHG:

$$z_{k+1} = \operatorname*{argmin}_z \left((Ax_k - b)^T z + g^*(z) + \tfrac{1}{t} d_{\text{d}}(z, z_k)\right) \quad (46a)$$

$$x_{k+1} = \operatorname*{argmin}_x \left((c - A^T(2z_{k+1} - z_k))^T x + f(x) + \tfrac{1}{t} d_{\mathrm{p}}(x, x_k)\right). \tag{46b}$$

It is also the first variant applied to the dual problem (2b). The algorithms (45) and (46) use the same stepsize parameter t in primal and dual updates. Different primal and dual stepsizes can be accommodated via the definition of d_{p} and d_{d}. For example, if quadratic kernels $\phi_{\mathrm{p}}(x) = \frac{1}{2\tau}(x^T x)$ and $\phi_{\mathrm{d}}(z) = \frac{1}{2\sigma}(z^T z)$ are used and $t = 1$, the two versions of PDHG are

$$\begin{aligned} x_{k+1} &= \mathbf{prox}_{\tau f}(x_k - \tau(c - A^T z_k)) \\ z_{k+1} &= \mathbf{prox}_{\sigma g^*}(z_k - \sigma(A(2x_{k+1} - x_k) - b)) \end{aligned}$$

and

$$\begin{aligned} z_{k+1} &= \mathbf{prox}_{\sigma g^*}(z_k - \sigma(Ax_k - b)) \\ x_{k+1} &= \mathbf{prox}_{\tau f}(x_k - \tau(c - A^T(2z_{k+1} - z_k))). \end{aligned}$$

If we apply the PDHG methods (45) and (46) to the exact-penalty conic LP formulation (35), we obtain

$$\begin{aligned} x_{k+1} &= \operatorname*{argmin}_{x \in K,\, \hat{w}^T x \leq 1} \left((c - A^T z_k)^T x + \tfrac{1}{t} d_{\mathrm{p}}(x, x_k)\right) \\ z_{k+1} &= \operatorname*{argmin}_{z \in L^*,\, \hat{s}^T z \leq 1} \left((A(2x_{k+1} - x_k) - b)^T z + \tfrac{1}{t} d_{\mathrm{d}}(z, z_k)\right) \end{aligned}$$

and

$$\begin{aligned} z_{k+1} &= \operatorname*{argmin}_{z \in L^*,\, \hat{s}^T z \leq 1} \left((Ax_k - b)^T z + \tfrac{1}{t} d_{\mathrm{d}}(z, z_k)\right) \\ x_{k+1} &= \operatorname*{argmin}_{x \in K,\, \hat{w}^T x \leq 1} \left((c - A^T(2z_{k+1} - z_k))^T x + \tfrac{1}{t} d_{\mathrm{p}}(x, x_k)\right). \end{aligned}$$

These iterations require Bregman projections on truncations of the cones K and L^*.

Extragradient method Korpelevich's extragradient method is a classical method for the monotone variational inequality problem (23) [41], [13, p. 285]. Extensions of the extragradient method have been developed that handle the generalized variational inequality (22) and use Bregman proximal operators. The following algorithm was proposed and analyzed in [49, section 3], [10, section 4], [59, section 5], [37, 57].

The Bregman extragradient algorithm can be interpreted as two iterations of a fixed-point iteration to compute y_{k+1} from (40). We start by writing (40) as a fixed-point condition for y_{k+1}:

$$y_{k+1} = \operatorname*{argmin}_y \left(h(y) + F(y_{k+1})^T y + \tfrac{1}{t} d_{\mathrm{pd}}(y, y_k)\right). \tag{47}$$

Two steps in the fixed-point iteration to compute the fixed point y_{k+1}, starting at y_k as initial point, are

$$\bar{y}_k = \operatorname*{argmin}_y \left(h(y) + F(y_k)^T y + \tfrac{1}{t} d_{\mathrm{pd}}(y, y_k)\right) \tag{48a}$$

$$y_{k+1} = \operatorname*{argmin}_y \left(h(y) + F(\bar{y}_k)^T y + \tfrac{1}{t} d_{\mathrm{pd}}(y, y_k)\right). \tag{48b}$$

If we use a separable distance function $d_{\mathrm{pd}}(y, y') = d_{\mathrm{p}}(x, x') + d_{\mathrm{d}}(z, z')$, where $y = (x, z)$ and $y' = (x', z')$, and substitute the expressions for F and h in (18), we obtain the algorithm

$$\bar{x}_k = \operatorname*{argmin}_x \left((c - A^T z_k)^T x + f(x) + \tfrac{1}{t} d_{\mathrm{p}}(x, x_k)\right) \tag{49a}$$

$$\bar{z}_k = \operatorname*{argmin}_z \left((Ax_k - b)^T z + g^*(z) + \tfrac{1}{t} d_{\mathrm{d}}(z, z_k)\right) \tag{49b}$$

$$x_{k+1} = \operatorname*{argmin}_x \left((c - A^T \bar{z}_k)^T x + f(x) + \tfrac{1}{t} d_{\mathrm{p}}(x, x_k)\right) \tag{49c}$$

$$z_{k+1} = \operatorname*{argmin}_z \left((A\bar{x}_k - b)^T z + g^*(z) + \tfrac{1}{t} d_{\mathrm{d}}(z, z_k)\right). \tag{49d}$$

With $t = 1$ and quadratic kernels $\phi_{\mathrm{p}}(x) = \frac{1}{2\tau}(x^T x)$ and $\phi_{\mathrm{d}}(z) = \frac{1}{2\sigma}(z^T z)$, we get

$$\bar{x}_k = \mathbf{prox}_{\tau f}(x_k - \tau(c - A^T z_k)) \tag{50a}$$

$$\bar{z}_k = \mathbf{prox}_{\sigma g^*}(z_k - \sigma(Ax_k - b)) \tag{50b}$$

$$x_{k+1} = \mathbf{prox}_{\tau f}(x_k - \tau(c - A^T \bar{z}_k)) \tag{50c}$$

$$z_{k+1} = \mathbf{prox}_{\sigma g^*}(z_k - \sigma(A\bar{x}_k - b)). \tag{50d}$$

Applied to the exact-penalty conic LP formulation, the extragradient method takes the following four steps:

$$\bar{x}_k = \operatorname*{argmin}_{x \in K,\, \hat{w}^T x \le 1} \left((c - A^T z_k)^T x + \tfrac{1}{t} d_{\mathrm{p}}(x, x_k)\right)$$

$$\bar{z}_k = \operatorname*{argmin}_{z \in L^*,\, \hat{s}^T z \le 1} \left((Ax_k - b)^T z + \tfrac{1}{t} d_{\mathrm{d}}(z, z_k)\right)$$

$$x_{k+1} = \operatorname*{argmin}_{x \in K,\, \hat{w}^T x \le 1} \left((c - A^T \bar{z}_k)^T x + \tfrac{1}{t} d_{\mathrm{p}}(x, x_k)\right)$$

$$z_{k+1} = \operatorname*{argmin}_{z \in L^*,\, \hat{s}^T z \le 1} \left((A\bar{x}_k - b)^T z + \tfrac{1}{t} d_{\mathrm{d}}(z, z_k)\right).$$

Each of these steps is a Bregman projection on a truncated cone.

Stepsize condition In Section 5 we will extend the two methods to incorporate adaptive stepsizes and analyze their convergence on self-dual problems. Here we mention standard results on allowable values of the fixed stepsize t in (45), (46), and (49).

We assume that ϕ_{p} and ϕ_{d} are strongly convex with respect to norms $\|\cdot\|_{\mathrm{p}}$ and $\|\cdot\|_{\mathrm{d}}$, normalized so that

$$d_{\mathrm{p}}(x, x') \ge \frac{1}{2}\|x - x'\|_{\mathrm{p}}^2, \qquad d_{\mathrm{d}}(z, z') \ge \frac{1}{2}\|z - z'\|_{\mathrm{d}}^2 \tag{51}$$

for all $x, x' \in \operatorname{dom} f \cap \operatorname{dom} \phi_{\mathrm{p}}$ with $x' \in \operatorname{int}(\operatorname{dom} \phi_{\mathrm{p}})$, and all $z, z' \in \operatorname{dom} g^* \cap \operatorname{dom} \phi_{\mathrm{d}}$ with $z' \in \operatorname{int}(\operatorname{dom} \phi_{\mathrm{d}})$. The corresponding matrix norm is

$$\|A\| = \sup_{x,y \ne 0} \frac{y^T A x}{\|x\|_{\mathrm{p}} \|y\|_{\mathrm{d}}} = \sup_{x \ne 0} \frac{\|Ax\|_{\mathrm{d}*}}{\|x\|_{\mathrm{p}}} = \sup_{y \ne 0} \frac{\|A^T y\|_{\mathrm{p}*}}{\|y\|_{\mathrm{d}}}, \tag{52}$$

where $\|\cdot\|_{\mathrm{p}*}$ and $\|\cdot\|_{\mathrm{d}*}$ are the dual norms of $\|\cdot\|_{\mathrm{p}}$ and $\|\cdot\|_{\mathrm{d}}$. In the two PDHG variants and the extragradient method, a fixed stepsize t satisfying $t\|A\| \le 1$ can be used [19, 34, 49, 59]. For PDHG, this condition guarantees convexity of the kernel ϕ_{pd} in (43), since (51) and (52) imply

$$d_{\mathrm{pd}}(y, y') \ge \frac{1 - t\|A\|}{2}(\|x - x'\|_{\mathrm{p}}^2 + \|z - z'\|_{\mathrm{d}}^2) \ge 0. \tag{53}$$

For the extragradient method, $t\|A\| \le 1$ guarantees the stepsize condition used in the analysis of [59, section 5]; see [59, lemma 1], which extends [49, theorem 3.1].

4.2 Specialization to self-dual problems

For a self-dual problem ($b = -c$, $g = f^*$, and $A = -A^T$), with equal primal and dual distances $d_{\rm p} = d_{\rm d} = d$, the primal and dual variants of PDHG (45) and (46) are equivalent up to interchanging the two sequences. The first version is

$$x_{k+1} = \operatorname*{argmin}_x \left((Az_k - b)^T x + f(x) + \tfrac{1}{t} d(x, x_k)\right) \tag{54a}$$

$$z_{k+1} = \operatorname*{argmin}_z \left((A(2x_{k+1} - x_k) - b)^T z + f(z) + \tfrac{1}{t} d(z, z_k)\right). \tag{54b}$$

This is identical to the second variant if we switch the notation for the two sequences x_k and z_k. For the extragradient method (49) applied to the self-dual problem (11), if we choose $d_{\rm p} = d_{\rm d} = d$ and start the iteration (49) at $x_0 = z_0$, then $\bar{x}_k = \bar{z}_k$ and $x_k = z_k$ for all k, and the four steps in the iteration collapse to two:

$$\bar{x}_k = \operatorname*{argmin}_x \left((Ax_k - b)^T x + f(x) + \tfrac{1}{t} d(x, x_k)\right) \tag{55a}$$

$$x_{k+1} = \operatorname*{argmin}_x \left((A\bar{x}_k - b)^T x + f(x) + \tfrac{1}{t} d(x, x_k)\right). \tag{55b}$$

This is also the extragradient method applied directly to the simplified optimality conditions (25).

For the exact-penalty formulation (38) of the self-dual conic LP (8), the PDHG and extragradient methods are

$$x_{k+1} = \operatorname*{argmin}_{x \in K,\, \hat{s}^T x \le \mu} \left((Az_k - b)^T x + \tfrac{1}{t} d(x, x_k)\right) \tag{56a}$$

$$z_{k+1} = \operatorname*{argmin}_{z \in K,\, \hat{s}^T z \le \mu} \left((A(2x_{k+1} - x_k) - b)^T z + \tfrac{1}{t} d(z, z_k)\right), \tag{56b}$$

and

$$\bar{x}_k = \operatorname*{argmin}_{x \in K,\, \hat{s}^T x \le \mu} \left((Ax_k - b)^T x + \tfrac{1}{t} d(x, x_k)\right) \tag{57a}$$

$$x_{k+1} = \operatorname*{argmin}_{x \in K,\, \hat{s}^T x \le \mu} \left((A\bar{x}_k - b)^T x + \tfrac{1}{t} d(x, x_k)\right). \tag{57b}$$

5 Adaptive stepsizes for self-dual problems

In this section we study adaptive stepsizes selected by a line search for PDHG (54) and the extragradient method (55), applied to the self-dual problem (11). The line search technique removes the need for an estimate of $\|A\|$ in the condition $t\|A\| \le 1$ of Section 4.1, and can potentially allow stepsizes larger than $1/\|A\|$. Throughout the section, the two methods are treated in parallel. Section 5.1 presents a line search for each method. The relation of the Bregman PDHG algorithm with line search (Algorithm 1) to existing line searches for Euclidean and Bregman PDHG is discussed after Algorithm 1. The line search for the Bregman extragradient method (Algorithm 2) was proposed in [59, section 5], and is specialized here to the self-dual problem.

Section 5.2 presents ergodic convergence bounds for the two methods on bounded effective domains, following standard arguments for Bregman proximal methods [19, 49, 59]. The resulting bounds on the gap function require a finite Bregman radius, *i.e.*, the Bregman distances from the starting points must be

bounded on the effective domain. Section 5.3 specializes the results to the exact-penalty formulation of self-dual LPs. In this case the gap function reduces to the exact-penalty objective. We remove the assumption of finite Bregman radius by allowing distances generated by logarithmically homogeneous functions, including logarithmically homogeneous barriers. For such distances, Theorem 5 shows that the exact-penalty objective at the averaged iterates converges to zero at the rate $O((\log k)/k)$, even though the Bregman radius is infinite. Theorem 6 shows that the rate improves to $O(1/k)$ if the kernel is finite at a solution. Both results hold for the extragradient method, and for PDHG if the stepsize increases are summable.

The analysis in this section is independent of the proximal point method interpretation in Section 4.

5.1 Bregman PDHG and EG with line search

We first consider the PDHG iteration (54) with a variable stepsize t_k.

Algorithm 1 *(Bregman PDHG with line search).* Choose starting points $x_0, z_0 \in \operatorname{dom} f \cap \operatorname{int}(\operatorname{dom} \phi)$, parameters $\delta \in (0,1]$, $\beta \in (0,1)$, $\bar{t} > 0$, an initial stepsize $t_{-1} \in (0, \bar{t}\,]$, and growth factors $\gamma_k \geq 1$, $k = 0, 1, \ldots$. At iteration k, set $t := \min\{\bar{t}, \gamma_k t_{k-1}\}$ and repeat the following two steps.

1. Compute
$$x^+ = \operatorname*{argmin}_x \left((Az_k - b)^T x + f(x) + \tfrac{1}{t} d(x, x_k)\right) \tag{58a}$$
$$z^+ = \operatorname*{argmin}_z \left((A(2x^+ - x_k) - b)^T z + f(z) + \tfrac{1}{t} d(z, z_k)\right). \tag{58b}$$
2. If
$$-t\,(z^+ - z_k)^T A(x^+ - x_k) \leq \delta\left(d(x^+, x_k) + d(z^+, z_k)\right), \tag{59}$$
accept the stepsize and set $t_k := t$, $x_{k+1} := x^+$, and $z_{k+1} := z^+$. Otherwise, set $t := \beta t$ and return to step 1.

Each trial in Algorithm 1 requires the two generalized projections in (58). The acceptance test in step 2 requires one inner product and the evaluation of the Bregman distances $d(x^+, x_k)$ and $d(z^+, z_k)$. Note that if $(z^+ - z_k)^T A(x^+ - x_k)$ is nonnegative, the trial stepsize is accepted automatically.

For the squared Euclidean distance, the test coincides with [28, eq. (29)] for equal primal and dual stepsizes, and is closely related to the adaptive rule used in PDLP [7]. Goldstein et al. [28] establish an $O(1/k)$ ergodic rate under a summability condition on the relative stepsize changes, whereas the PDLP rule is used heuristically. Malitsky and Pock [47, eq. (8)] propose a different Euclidean line search. In the notation used here, their test compares $\sqrt{\tau_k \sigma_k}\|A^T(z_{k+1} - z_k)\|$ with $\|z_{k+1} - z_k\|$, and only the dual update is recomputed during backtracking.

For Bregman distances, the test is a pointwise version of the Chambolle–Pock condition [19, eq. (13)], specialized to equal stepsizes and problems with smooth terms; their condition is imposed for all pairs of points. Yazdandoost Hamedani and Aybat [64] consider general, possibly nonlinear coupling functions and use Bregman distances in both updates. In their merely convex setting, the accepted stepsizes are nonincreasing and satisfy $\sigma_k = \gamma_0 \tau_k$. The methods in [34, eq. (36)] and [35, eq. (49)] instead use a Bregman distance in the primal update and a Euclidean distance in the dual update. The self-dual setting considered here uses the same distance and stepsize in the two updates, so both updates can be evaluated by the same projection oracle.

An adaptive stepsize can be incorporated in the extragradient method (55) as follows.

Algorithm 2 *(Bregman EG with line search).* Choose a starting point $x_0 \in \operatorname{dom} f \cap \operatorname{int}(\operatorname{dom} \phi)$, parameters $\beta \in (0,1)$, $\bar{t} > 0$, an initial stepsize $t_{-1} \in (0, \bar{t}\,]$, and growth factors $\gamma_k \geq 1$, $k = 0, 1, \ldots$. At iteration k, set $t := \min\{\bar{t}, \gamma_k t_{k-1}\}$ and repeat the following two steps.

1. Compute

$$\bar{x} = \operatorname*{argmin}_x \left((Ax_k - b)^T x + f(x) + \tfrac{1}{t} d(x, x_k) \right) \tag{60a}$$

$$x^+ = \operatorname*{argmin}_x \left((A\bar{x} - b)^T x + f(x) + \tfrac{1}{t} d(x, x_k) \right). \tag{60b}$$

2. If

$$f(\bar{x}) - b^T \bar{x} \leq (A\bar{x} - b)^T x^+ + f(x^+) + \frac{1}{t} d(x^+, x_k) \tag{61}$$

accept the stepsize and set $t_k := t$, $\bar{x}_k := \bar{x}$, and $x_{k+1} := x^+$. Otherwise, set $t := \beta t$ and return to step 1.

Each trial in Algorithm 2 also requires two generalized projections. The acceptance test requires the function values $f(\bar{x})$ and $f(x^+)$, one inner product, and the Bregman distance $d(x^+, x_k)$. The test (61) is the condition [59, eq. (45)] specialized to (55); see also [49, eq. (3.4)]. Alternatively, Nemirovski [49, section 6] adjusts the stepsize adaptively in his numerical experiments. Armijo-type line searches for Euclidean extragradient-type methods are studied in [32, 38, 58].

5.2 Convergence analysis

In this section we analyze the convergence of Algorithms 1 and 2 for the self-dual problem (11). We make the following assumptions.

A1. f is closed convex and $A = -A^T$.

A2. The set $D = \operatorname{dom} f \cap \operatorname{dom} \phi$ is bounded, and $D^\circ = \operatorname{dom} f \cap \operatorname{int}(\operatorname{dom} \phi)$ is nonempty.

A3. The distance-generating kernel ϕ is strongly convex on D: there exists a norm $\|\cdot\|$ such that

$$d(u, v) \geq \frac{1}{2}\|u - v\|^2 \quad \text{for all } u \in D,\ v \in D^\circ. \tag{62}$$

A4. The minimization problems in (60) and in (58) have unique solutions in D° for every stepsize $t > 0$ and all $x_k, z_k \in D^\circ$.

We define the constants

$$\rho_1 = \sup_{x \in D} d(x, x_0) + \sup_{z \in D} d(z, z_0), \qquad \rho_2 = \sup_{u,v,p,q \in D} |(u - v)^T A(p - q)|. \tag{63}$$

We refer to the constant ρ_1 as the *Bregman radius*. The Bregman radius ρ_1 may be infinite. By Assumption A2, the constant ρ_2 is finite: $\rho_2 \leq \|A\|_2 (\operatorname{diam} D)^2 < \infty$, where $\|A\|_2$ is the spectral norm of A and $\operatorname{diam} D$ is the Euclidean diameter of D.

Termination of the line searches Let $\|A\|$ be the matrix norm induced by the norm in A3,

$$\|A\| = \sup_{x,y\neq 0} \frac{y^T A x}{\|x\|\|y\|} = \sup_{x\neq 0} \frac{\|Ax\|_*}{\|x\|}.$$

For the trial points x^+, z^+ in Algorithm 1,

$$\begin{aligned} \left|(z^+ - z_k)^T A(x^+ - x_k)\right| &\leq \|A\|\|z^+ - z_k\|\|x^+ - x_k\| \\ &\leq \frac{\|A\|}{2}\left(\|x^+ - x_k\|^2 + \|z^+ - z_k\|^2\right) \\ &\leq \|A\|\left(d(x^+, x_k) + d(z^+, z_k)\right) \end{aligned}$$

by (62). Hence the test (59) is satisfied whenever $t \leq \delta/\|A\|$. Similarly, the test (61) in Algorithm 2 is satisfied whenever $t \leq 1/\|A\|$ [59, lemma 1]. (The argument in [59] only needs (62) at the points $\bar{x}$, x^+, and x_k.) Since in each unsuccessful trial t is multiplied by β, only finitely many trial stepsizes are rejected. The trial value at the start of iteration k is at least t_{k-1}, because $\gamma_k \geq 1$ and $t_{k-1} \leq \bar{t}$. If the first trial is successful, $t_k \geq t_{k-1}$. Otherwise, the last rejected value t_k/β exceeds $\delta/\|A\|$, so $t_k > \delta\beta/\|A\|$. By induction, the accepted stepsizes in Algorithm 1 satisfy

$$t_k \geq t_{\min} := \min\{t_{-1}, \delta\beta/\|A\|\}, \quad k = 0, 1, \ldots. \tag{64}$$

The same argument shows that the accepted stepsizes in Algorithm 2 satisfy (64) with $\delta = 1$. In particular, if $\gamma_k = 1$, $t_{-1} = \bar{t} = t \leq 1/\|A\|$ (and $\delta = 1$ in Algorithm 1), the first trial stepsize is accepted in every iteration, and the two algorithms reduce to the methods of Section 4.2 with fixed stepsize t.

Growth factors The factors γ_k control the increase of the trial stepsize. Define

$$\Gamma_k = \sum_{i=1}^{k} (t_i - t_{i-1})_+, \tag{65}$$

where $(a)_+ = \max\{a, 0\}$. Since $t_i \leq \min\{\bar{t}, \gamma_i t_{i-1}\}$, we have $\Gamma_k \leq \bar{t}\sum_{i=1}^{k}(\gamma_i - 1)$. The following choices give useful bounds on Γ_k.

(a) If $\gamma_k = 1$, the stepsizes are nonincreasing and $\Gamma_k = 0$.

(b) If $\gamma_k = 1 + \alpha_k/t_{k-1}$, the first trial stepsize is $\min\{\bar{t}, t_{k-1} + \alpha_k\}$ and $\Gamma_k \leq \sum_{i=1}^{k} \alpha_i$. Hence Γ_k is bounded if $\sum_k \alpha_k < \infty$.

(c) If $\gamma_k = 1 + \alpha_k$, then $\Gamma_k \leq \bar{t}\sum_{i=1}^{k} \alpha_i$. Hence Γ_k is bounded if $\sum_k \alpha_k < \infty$.

(d) If $\gamma_k = 1 + \alpha(k+1)^{-p}$, where $\alpha > 0$ and $p > 0$, then

$$\Gamma_k = \begin{cases} O(k^{1-p}) & 0 < p < 1 \\ O(\log k) & p = 1 \\ O(1) & p > 1. \end{cases}$$

The choice $p = 0.6$ is used in PDLP [7].

Convergence results For $u, v \in D$, define the gap function

$$\eta(u, v) = \sup_{z \in D} \mathcal{L}(u, z) - \inf_{x \in D} \mathcal{L}(x, v), \tag{66}$$

where $\mathcal{L}$ is the convex–concave function (26) associated with (11). Clearly, $\eta(u, v) \geq 0$ for $u, v \in D$, since $\mathcal{L}(u, u) = \mathcal{L}(v, v) = 0$. Moreover, $\mathcal{L}(x, z) = -\mathcal{L}(z, x)$, so

$$\eta(u, u) = 2 \sup_{z \in D} \mathcal{L}(u, z). \tag{67}$$

The next theorem is the main convergence result for Algorithm 1. For $k \geq 0$, we define the weighted averages of the iterates of Algorithm 1 as

$$x_k^{\text{avg}} = \frac{1}{T_k} \sum_{i=0}^{k} t_i x_{i+1}, \qquad z_k^{\text{avg}} = \frac{1}{T_k} \sum_{i=0}^{k} t_i z_{i+1}, \qquad T_k = \sum_{i=0}^{k} t_i. \tag{68}$$

Theorem 3 *Suppose Assumptions A1–A4 hold. For all $x, z \in D$, the averaged iterates (68) of Algorithm 1 satisfy*

$$\mathcal{L}(x_k^{\text{avg}}, z) - \mathcal{L}(x, z_k^{\text{avg}}) \leq \frac{d(x, x_0) + d(z, z_0) + 2\rho_2(t_0 + \Gamma_k)}{T_k}, \quad k = 0, 1, \ldots. \tag{69}$$

If $\rho_1 < \infty$, then

$$\eta(x_k^{\text{avg}}, z_k^{\text{avg}}) \leq \frac{\rho_1 + 2\rho_2(t_0 + \Gamma_k)}{T_k}, \quad k = 0, 1, \ldots. \tag{70}$$

The proof is given in Appendix C.1. Since $T_k \geq (k+1)t_{\min}$, the right-hand side of (70) is $O(1/k)$ if $\Gamma_k = 0$ (case (a)), and for the growth factors in (b) and (c) if $\sum_k \alpha_k < \infty$, since then $\Gamma_k = O(1)$. With the growth factors in (d), it is $O(k^{-p})$ for $0 < p < 1$, $O((\log k)/k)$ for $p = 1$, and $O(1/k)$ for $p > 1$. For the growth schedule of PDLP ($p = 0.6$), it is $O(k^{-0.6})$.

For Algorithm 2, the averaged iterate is defined as

$$x_k^{\text{avg}} = \frac{1}{T_k} \sum_{i=0}^{k} t_i \bar{x}_i, \tag{71}$$

with T_k defined in (68). The next theorem is the ergodic bound of [59, proposition 4(a)] (see also [10, proposition 4.1]), specialized to the self-dual problem and written in a form that parallels Theorem 3, using the identity $\mathcal{L}(x, z) = -\mathcal{L}(z, x)$.

Theorem 4 ([59, proposition 4(a)]) *Suppose Assumptions A1–A4 hold. For $x, z \in D$, the averaged iterate (71) of Algorithm 2 satisfies*

$$\mathcal{L}(x_k^{\text{avg}}, z) - \mathcal{L}(x, x_k^{\text{avg}}) \leq \frac{d(x, x_0) + d(z, x_0)}{T_k}, \quad k = 0, 1, \ldots. \tag{72}$$

If $\rho_1 < \infty$, then

$$\eta(x_k^{\text{avg}}, x_k^{\text{avg}}) \leq \frac{\rho_1}{T_k}, \quad k = 0, 1, \ldots. \tag{73}$$

In (73), ρ_1 is the constant in (63) with $z_0 = x_0$.

For completeness, a proof is given in Appendix C.2. Since $T_k \geq (k+1)t_{\min}$, the right-hand side of (73) is $O(1/k)$. The bound (72) holds for any sequence of stepsizes that satisfy (61), and does not depend on how the stepsizes change from one iteration to the next. This is the main difference from Theorem 3.

The preceding results concern the weighted averages of the iterates and do not establish convergence of the iterates. Point convergence requires additional assumptions on the kernel, such as boundedness of Bregman sublevel sets and suitable continuity properties; see, *e.g.*, [34, section 3.3.4].

5.3 Exact-penalty formulation of a self-dual conic LP

We now specialize the results to the exact-penalty formulation (38) of the self-dual LP (8), *i.e.*, to $f = \delta_C$ with

$$C = \{x \in K \mid \hat{s}^T x \leq \mu\}. \tag{74}$$

We make the following assumptions.

B1. $A = -A^T$, K is a proper cone, $\hat{x} \in \operatorname{int}(K)$, $\hat{s} = A\hat{x} - b \in \operatorname{int}(K^*)$, and $\mu > \hat{s}^T\hat{x}$.

B2. The kernel ϕ is closed and convex, and $\operatorname{int}(K) \subseteq \operatorname{dom}\phi$.

B3. The kernel ϕ is strongly convex on $C \cap \operatorname{dom}\phi$: there exists a norm $\|\cdot\|$ such that

$$d(u, v) \geq \frac{1}{2}\|u - v\|^2 \quad \text{for all } u \in C \cap \operatorname{dom}\phi,\ v \in C \cap \operatorname{int}(\operatorname{dom}\phi).$$

B4. For every $a \in \mathbb{R}^n$, $t > 0$, and $y \in C \cap \operatorname{int}(\operatorname{dom}\phi)$, the generalized projection

$$\operatorname*{argmin}_{x \in C} \left(a^T x + \tfrac{1}{t} d(x, y)\right)$$

exists, is unique, and lies in $\operatorname{int}(\operatorname{dom}\phi)$.

Under B1–B4, Theorems 3 and 4 apply to (38) with $D = C \cap \operatorname{dom}\phi$. The PDHG and extragradient methods for this problem are given in (56) and (57). Each update is a generalized projection on the truncated cone C. Hence B3 is required only on the truncated cone. This is important in applications that involve distances that are strongly convex on the truncated cone but not on the entire cone, as is the case for the examples in Section 6.

For $f = \delta_C$, the function $\mathcal{L}$ is bilinear on $C \times C$,

$$\mathcal{L}(x, z) = b^T(z - x) - z^T A x \tag{75}$$

for $x, z \in C$, hence continuous. By B2, $\operatorname{cl} D = C$. The supremum and infimum in (66) are therefore unchanged if D is replaced by C. We use the same notation for the resulting extension to $C \times C$:

$$\eta(u, v) = \sup_{z \in C} \mathcal{L}(u, z) - \inf_{x \in C} \mathcal{L}(x, v)$$

for $u, v \in C$. By (39) and the identity $\mathcal{L}(x, z) = -\mathcal{L}(z, x)$,

$$\eta(u, v) = q(u) + q(v) \tag{76}$$

for $u, v \in C$, where $q(x) = -b^T x + \mu\sigma_{K^*}(b - Ax, \hat{s})$ is the objective function of the exact-penalty problem (38). Since $q(u) = \sup_{z \in C} \mathcal{L}(u, z) \geq \mathcal{L}(u, u) = 0$, both terms on the right-hand side of (76) are nonnegative. The optimal value is zero, so $q(u) = 0$ if and only if u is optimal.

In the analyses of Euclidean PDHG [7, 28], restricting the gap to a bounded set provides the finite diameter needed in the convergence bound. For the exact-penalty formulation, the restricted gap also has a direct interpretation: by (76), $\eta(u, v) = 0$ if and only if both u and v solve the self-dual LP (8). This interpretation motivates analyzing the exact-penalty formulation rather than (8).

Distances with finite Bregman radius Suppose B1–B4 hold. If $\rho_1 < \infty$, Theorems 3 and 4, combined with (76), apply directly. The averaged iterates are in D, and $T_k \geq (k+1)t_{\min}$. Therefore, Algorithm 1 applied to (38) satisfies

$$q(x_k^{\text{avg}}) + q(z_k^{\text{avg}}) \leq \frac{\rho_1 + 2\rho_2(t_0 + \Gamma_k)}{(k+1)\, t_{\min}}, \quad k = 0, 1, \ldots. \tag{77}$$

Algorithm 2 applied to (38) satisfies

$$2q(x_k^{\text{avg}}) \leq \frac{\rho_1}{(k+1)\, t_{\min}}, \quad k = 0, 1, \ldots, \tag{78}$$

with $z_0 = x_0$ in the definition of ρ_1 and $\delta = 1$ in $t_{\min}$. In particular, the exact-penalty objective values at the averaged iterates of Algorithm 2 converge to zero at the rate $O(1/k)$. For Algorithm 1, they converge to zero if $\Gamma_k = o(k)$, at the rates discussed after Theorem 3.

Distances generated by logarithmically homogeneous kernels The next theorem relaxes the condition $\rho_1 < \infty$ under the additional assumption that the distance is generated by a *logarithmically homogeneous* function.

B5. ϕ is ν-logarithmically homogeneous with respect to K, where $\nu > 0$, *i.e.*,

$$\text{int}(K) \subseteq \text{dom}\, \phi \subseteq K, \qquad \phi(\lambda x) = \phi(x) - \nu \log \lambda \quad \text{for all } x \in \text{int}(K) \text{ and } \lambda > 0. \tag{79}$$

Examples include the logarithmically homogeneous barrier functions used in interior-point methods (in this case, $\text{dom}\, \phi = \text{int}(K)$). The Itakura–Saito distance used in Section 6.2 is generated by a logarithmically homogeneous function that is not a barrier function (in this case $\text{dom}\, \phi = K \setminus \{0\}$).

Theorem 5 *Suppose Assumptions B1–B5 hold. Define*

$$\rho_3 = 2\phi(\hat{x}) - \phi(x_0) - \phi(z_0) + \sup_{u,v \in C} \left(\nabla\phi(x_0)^T (x_0 - u) + \nabla\phi(z_0)^T (z_0 - v) \right). \tag{80}$$

Then Algorithm 1 applied to (38) satisfies

$$q(x_k^{\text{avg}}) + q(z_k^{\text{avg}}) \leq \frac{2\nu \log(k+2) + \rho_3 + 2\rho_2(t_0 + \Gamma_k)}{(k+1)\, t_{\min}} + \frac{2(\hat{s}^T \hat{x} + \rho_2)}{k+2}, \quad k = 0, 1, \ldots, \tag{81}$$

and Algorithm 2 applied to (38) satisfies

$$2q(x_k^{\text{avg}}) \leq \frac{2\nu \log(k+2) + \rho_3}{(k+1)\, t_{\min}} + \frac{2(\hat{s}^T \hat{x} + \rho_2)}{k+2}, \quad k = 0, 1, \ldots, \tag{82}$$

with $z_0 = x_0$ in the definition of ρ_3.

(Here the averaged iterates are defined in (68) and (71), and $t_{\min}$ in (64), with $\delta = 1$ for Algorithm 2.) The constant ρ_3 is finite, because C is bounded and $\hat{x}, x_0, z_0 \in \text{int}(\text{dom}\, \phi)$. It is nonnegative, because the choice $u = v = \hat{x}$ in the argument of the supremum in (80) shows that $\rho_3 \geq d(\hat{x}, x_0) + d(\hat{x}, z_0)$.

For Algorithm 1, the theorem shows that $q(x_k^{\text{avg}}) + q(z_k^{\text{avg}}) = O((\nu \log k + \rho_2 \Gamma_k)/k)$, which is $O((\log k)/k)$ if $\Gamma_k = O(1)$. For Algorithm 2, it shows that $q(x_k^{\text{avg}}) = O((\log k)/k)$.

The proof of Theorem 5 is given in Appendix C.3 and uses a technique from the analysis of a barrier subgradient method [50], which was later applied to online learning with self-concordant barrier regularizers [1]. For mirror descent and proximal mirror descent with logarithmic barrier kernels, convergence rates of order $O((\log k)/k)$ are derived in a recent paper [23].

The next theorem establishes that the factor $\log(k+2)$ in (81) and (82) can be removed if ϕ is finite at a solution of (38). This is useful for applications involving logarithmically homogeneous entropic kernels that are finite on most of the boundary of K. (See Section 6.2 for an example.)

Theorem 6 *Suppose Assumptions B1–B5 hold and problem (38) has a solution $x^\star$ with $\phi(x^\star) < \infty$. Define*

$$\rho_4 = 2\phi(x^\star) - \phi(x_0) - \phi(z_0) + \sup_{u,v \in C} \left(\nabla\phi(x_0)^T(x_0 - u) + \nabla\phi(z_0)^T(z_0 - v)\right). \tag{83}$$

Then Algorithm 1 applied to (38) satisfies

$$q(x_k^{\text{avg}}) + q(z_k^{\text{avg}}) \leq \frac{2\,(2\nu\log 2 + \rho_4 + 2\rho_2(t_0 + \Gamma_k))}{(k+1)\,t_{\min}}, \quad k = 0, 1, \ldots, \tag{84}$$

and Algorithm 2 applied to (38) satisfies

$$2q(x_k^{\text{avg}}) \leq \frac{2\,(2\nu\log 2 + \rho_4)}{(k+1)\,t_{\min}}, \quad k = 0, 1, \ldots, \tag{85}$$

with $z_0 = x_0$ in the definition of ρ_4.

The constant ρ_4 is finite because C is bounded and $\phi(x^\star) < \infty$. It is nonnegative, because the choice $u = v = x^\star$ in the argument of the supremum in (83) shows that $\rho_4 \geq d(x^\star, x_0) + d(x^\star, z_0)$. The right-hand side of (84) is $O(1/k)$ if $\Gamma_k = O(1)$. The right-hand side of (85) is $O(1/k)$.

The proof of Theorem 6 is given in Appendix C.4.

6 Numerical experiments

We illustrate the reformulation and algorithmic framework on two classes of conic linear programs. Starting from a conic LP, we first form the self-dual embedding described in Appendix B. We then apply the exact-penalty reformulation to obtain a problem over truncated cones. Finally, we solve the resulting problem with the Bregman primal–dual methods of Sections 4 and 5. The two examples differ only in the cone and in the Bregman projection oracle: the first uses the cone of nonnegative trigonometric polynomials with the Itakura–Saito distance, and the second uses a sparse positive semidefinite cone with the distance generated by a logarithmic barrier function. The experiments are intended to demonstrate that the proposed reformulation can be used as a practical route from a generic conic LP to Bregman first-order methods when efficient generalized projections are available.

6.1 Implementation details

We describe the implementation for a conic LP in standard primal–dual form

$$\begin{array}{lll} \text{minimize} & c^T x & \\ \text{subject to} & Ax = b & \\ & x \in K, & \end{array} \qquad\qquad \begin{array}{lll} \text{maximize} & b^T z & \\ \text{subject to} & c - A^T z \in K^* & \end{array}$$

where K is a proper cone. The first step is to form the self-dual embedding

$$\begin{array}{ll} \text{minimize} & \theta \\ \text{subject to} & \begin{bmatrix} 0 & -A^T & c & q_x \\ A & 0 & -b & q_z \\ -c^T & b^T & 0 & q_\tau \\ -q_x^T & -q_z^T & -q_\tau & 0 \end{bmatrix} \begin{bmatrix} x \\ z \\ \tau \\ \theta \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} w \\ s \\ \kappa \\ \eta \end{bmatrix} \\ & (x, z, \tau, \theta) \in K \times \mathbb{R}^m \times \mathbb{R}_+ \times \mathbb{R} \\ & (w, s, \kappa, \eta) \in K^* \times \{0\}^m \times \mathbb{R}_+ \times \{0\}. \end{array} \tag{86}$$

This is a self-dual conic LP in the form of (8). We then apply Bregman PDHG and extragradient methods to its exact-penalty reformulation, as described in Section 4.

The parameters in the embedding (86) are defined as

$$\begin{bmatrix} q_x \\ q_z \\ q_\tau \end{bmatrix} = \frac{1}{\hat\theta} \left(\begin{bmatrix} \hat w \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 & -A^T & c \\ A & 0 & -b \\ -c^T & b^T & 0 \end{bmatrix} \begin{bmatrix} \hat x \\ \hat z \\ 1 \end{bmatrix} \right),$$

where $\hat\theta = \hat w^T \hat x + 1$ and $\hat x \in \operatorname{int}(K)$, $\hat z \in \mathbb{R}^m$, $\hat w \in \operatorname{int}(K^*)$. The points $\hat x, \hat z, \hat w$ are also used to initialize the Bregman proximal methods. The choice of $\hat x, \hat z, \hat w$ in our experiments uses the heuristic in CVXOPT [5], which is adapted from the initialization strategy of [48], and can be summarized as follows. We assume that A has full row rank and compute the least-norm primal estimate $x_{\text{ls}} = A^T(AA^T)^{-1}b$. If this point is not in the interior of the cone, it is shifted in a fixed interior direction:

$$\hat x = \begin{cases} x_{\text{ls}} & \alpha_{\text{p}} < 0 \\ x_{\text{ls}} + (1 + \alpha_{\text{p}}) e_{\text{p}} & \text{otherwise,} \end{cases}$$

where $e_{\text{p}} \in \operatorname{int}(K)$ and $\alpha_{\text{p}} = \inf\{\alpha \mid x_{\text{ls}} + \alpha e_{\text{p}} \in K\}$. Similarly, on the dual side, we compute

$$z_{\text{ls}} = \operatorname*{argmin}_z \|c - A^T z\|_2^2, \qquad w_{\text{ls}} = c - A^T z_{\text{ls}},$$

and set $\hat z = z_{\text{ls}}$ together with

$$\hat w = \begin{cases} w_{\text{ls}} & \alpha_{\text{d}} < 0 \\ w_{\text{ls}} + (1 + \alpha_{\text{d}}) e_{\text{d}} & \text{otherwise,} \end{cases}$$

where $e_{\text{d}} \in \operatorname{int}(K^*)$ and $\alpha_{\text{d}} = \inf\{\alpha \mid w_{\text{ls}} + \alpha e_{\text{d}} \in K^*\}$. The remaining components of the reference point are chosen as $\hat s = 0$, $\hat\tau = \hat\kappa = 1$, $\hat\theta = \hat w^T \hat x + 1$, and $\hat\eta = 0$.

The exact-penalty formulation of the embedded problem contains the bounded constraint

$$\hat w^T x + \tau \le \mu, \qquad x \in K, \qquad \tau \ge 0, \tag{87}$$

where $\mu > \hat w^T \hat x + 1$. Direct application of Bregman PDHG or extragradient methods couples the x- and τ-updates through (87). In the implementation we replace (87) by the weaker constraints

$$\hat w^T x \le \mu, \qquad 0 \le \tau \le \mu.$$

Consequently, the x-update is separated from the τ-update and becomes the Bregman projection

$$\Pi(a, \hat w, \mu) = \operatorname*{argmin}_{\hat w^T x \le \mu,\, x \in K} (-a^T x + \phi(x)), \tag{88}$$

where ϕ is the distance-generating function.

PDHG uses the line search of Algorithm 1 and the extragradient method uses Algorithm 2. In both methods the trial value at the start of iteration k is $t = \min\{\bar{t}, \gamma_k t_{k-1}\}$, with $\bar{t} = t_{-1} = 1$ and $\gamma_k = 1+(k+1)^{-1.1}$. The backtracking factor is $\beta = 0.8$, and the parameter in the line-search test (59) is $\delta = 1$.

At each iteration with $\tau > 0$, we form the recovered primal–dual variables as described at the end of Appendix B.1:

$$\bar{x} = x/\tau, \qquad \bar{z} = z/\tau, \qquad \bar{w} = w/\tau.$$

The residuals are measured for the original conic LP:

$$r_{\mathrm{p}} = \frac{\|A\bar{x} - b\|_2}{1 + \|b\|_2}, \qquad r_{\mathrm{d}} = \frac{\|A^T\bar{z} + \bar{w} - c\|_2}{1 + \|c\|_2}, \qquad r_{\mathrm{gap}} = \frac{|c^T\bar{x} - b^T\bar{z}|}{1 + |c^T\bar{x}| + |b^T\bar{z}|}.$$

The algorithm is terminated when $r_{\max} = \max\{r_{\mathrm{p}}, r_{\mathrm{d}}, r_{\mathrm{gap}}\} \leq 10^{-3}$ or when the number of iterations reaches 10^6.

All experiments are carried out in Python 3.14 on a Mac with an Apple M2 chip and 16 GB of RAM.

6.2 Nonnegative trigonometric polynomials

We first consider conic LPs over the cone of nonnegative trigonometric polynomials of degree at most p:

$$K = \{x \in \mathbb{R}^{p+1} \mid F_x(\omega) \geq 0 \text{ for all } \omega\},$$

where $F_x(\omega) = x_1 + 2\sum_{k=2}^{p+1} x_k \cos((k-1)\omega)$. Optimization problems over the cone of nonnegative trigonometric polynomials or its dual cone, the cone of positive semidefinite Toeplitz matrices, are common in signal processing and system identification. As discussed in [20], a useful Bregman distance on K is the Itakura–Saito distance generated by the negative Burg entropy

$$\phi(x) = -\frac{1}{2\pi}\int_0^{2\pi} \log F_x(\omega)\, d\omega.$$

In [20], an efficient method is described to compute a generalized projection (88). The key observation is that the optimality conditions for this projection reduce to a one-dimensional nonlinear equation [20, section 5]. Each Newton step for this scalar nonlinear equation requires the solution of a positive definite Toeplitz linear system, which can be carried out with complexity $O(p^2)$ by the Levinson–Durbin algorithm, or with complexity $O(p(\log p)^2)$ using superfast Toeplitz solvers. Hence the cost of evaluating the Bregman projection is roughly quadratic in the degree p of the trigonometric polynomial, with a very small number of Newton steps in practice. This should be contrasted with the Euclidean projection on the same truncated cone, which is a significantly more difficult convex optimization problem and, through its semidefinite representation, is tied to eigenvalue decomposition or SDP subproblems of order $p + 1$.

The Itakura–Saito kernel is 1-logarithmically homogeneous with $\operatorname{dom}\phi = K \setminus \{0\}$, *i.e.*, ϕ is finite on the boundary of K, except at the origin. Theorem 6 therefore applies if the exact-penalty problem has a nonzero solution. With the growth factors in Section 6.1, $\Gamma_k = O(1)$, so the exact-penalty objective converges at the rate $O(1/k)$ for both methods.

Numerical results Table 1 reports results for randomly generated linear programs over the cone of nonnegative trigonometric polynomials. The table shows that the per-iteration complexity remains well controlled as the problem size increases. The bound $\hat{w}^T x \leq \mu$ is active at only a small fraction of the projections,

					Projection statistics	
p	m	Method	# iter.	Time/iter. (s)	active	Newton
1000	200	EG	7553	0.01325	7	5.71
		PDHG	5728	0.01931	8	7.89
2000	500	EG	4710	0.03573	12	5.25
		PDHG	3957	0.03817	21	4.91
5000	1000	EG	10918	0.2164	6	5.83
		PDHG	11948	0.2917	11	5.76
6000	1200	EG	9759	0.4967	5	6.20
		PDHG	7562	0.5194	6	5.98

Table 1: Results for randomly generated LPs over the cone of nonnegative trigonometric polynomials. p is the degree and m is the number of equality constraints. The column *active* gives the number of projections for which $\hat{w}^T x = \mu$ at the solution (the total number of projections is twice # iter.). *Newton* gives the average number of Newton steps (averaged over the active projections). Inactive projections are available in closed form.

and at those projections five to eight Newton steps suffice, with no apparent dependence on p. The truncation introduced by the exact-penalty reformulation is therefore essentially free: it bounds the domain, as the convergence theory requires, without materially changing the cost of the proximal step. The time per iteration grows roughly quadratically with p, consistent with the cost of the projection routine for the cone of nonnegative trigonometric polynomials [20].

Figure 1 compares three choices of the reference point $(\hat{x}, \hat{z}, \hat{w})$ on a representative instance, using the Bregman extragradient method with line search. The reference point enters the embedding (86) via the definition of q_x, q_z, q_τ, and is also used to initialize the algorithm. The first choice is the *least-squares reference point* described in Section 6.1, obtained from the least-norm primal and least-squares dual estimates and shifted into the cone. The second choice is a *random reference point*. The third is the *unit reference point*

$$\hat{x} = (1, 0, \ldots, 0) \in \operatorname{int}(K), \qquad \hat{w} = (1, 0, \ldots, 0) \in \operatorname{int}(K^*), \qquad \hat{s} = 0,$$

with a randomly chosen $\hat{z}$. The vector $(1, 0, \ldots, 0)$ corresponds to the constant polynomial $F_x(\omega) = 1$. The figure shows the best-so-far KKT residuals versus iteration number. Although the least-squares reference point is only a heuristic, it gives substantially better performance in this example and is also more reliable in most of our experiments. We therefore use it in all the tests.

6.3 Sparse semidefinite programming

We next consider linear programs over sparse semidefinite cones [4, 34, 60]. Let E be an $n \times n$ chordal sparsity pattern, and let $\mathbb{S}^n_E$ denote the subspace of symmetric matrices with zeros outside E. We consider the primal–dual pair of cones

$$K = \Pi_E(\mathbb{S}^n_+), \qquad K^* = \mathbb{S}^n_+ \cap \mathbb{S}^n_E,$$

where Π_E denotes projection onto the entries in E. The primal cone K consists of partial symmetric matrices with sparsity pattern E that admit a positive semidefinite completion, while the dual cone K^* consists of positive semidefinite matrices with the same sparsity pattern. In the implementation, the pattern E is obtained

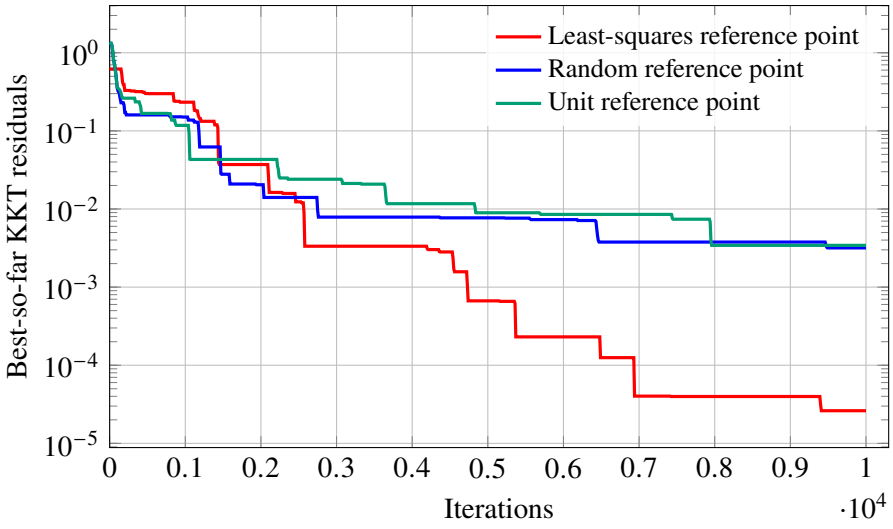


Figure 1: Best-so-far KKT residuals for the Bregman extragradient method, using three different reference points $(\hat{x}, \hat{z}, \hat{w})$ in the self-dual embedding (86).

from a chordal extension of the aggregate sparsity pattern of the problem, so that computations with K and K^* can be carried out using sparse Cholesky factorizations.

We use the logarithmic barrier kernel associated with K^*: $\phi_*(S) = -\log\det(S)$. The kernel on K is the (negative) conjugate

$$\phi(x) = \sup_{s \in \mathrm{int}(K^*)} \{-s^T x - \phi_*(s)\}.$$

This choice is well adapted to sparse semidefinite cones because its gradient and directional Hessian, as well as the corresponding Bregman projection (88), can be evaluated efficiently via sparse Cholesky factorization.

The kernel ϕ satisfies (79) with $\nu = n$ and $\mathrm{dom}\,\phi = \mathrm{int}(K)$, and is infinite on the boundary of K. Theorem 5 applies, but not Theorem 6, and the exact-penalty objective converges at the rate $O((\log k)/k)$ for both methods.

Numerical results The sparse SDP instances are generated with the function `rand_SDP` from SMCP [3]. As in [3], the aggregate sparsity patterns are taken from the SuiteSparse Matrix Collection [39]. Chordal extensions are computed with CHOMPACK [6] using the AMD ordering [2]. CHOMPACK is also used for the sparse Cholesky factorizations and related operations on the chordal embedding. The remaining sparse matrix computations are implemented using CVXOPT [5].

Table 2 reports the results for the sparse SDP instances. As in Table 1, the bound $\hat{w}^T x \leq \mu$ is active at fewer than 30 of the more than 10^4 projections in each run, and those projections take two to six Newton steps on average. The main observation is that the cost per iteration is dominated by the Cholesky factorizations and related sparse matrix operations in the Bregman projection oracle. This keeps the per-iteration cost inexpensive compared with methods based on dense eigenvalue decomposition, and allows the first-order methods to handle matrix variables with dimensions in the thousands. Iteration counts of this size (about 10^4 in Table 2) are common for first-order methods for SDPs, even when only moderate accuracy is required.

							Projection statistics	
Instance	n	density	m	Method	# iter.	Time/iter. (s)	active	Newton
rs35	2003	20.9%	2000	EG	21759	8.3520	16	5.36
				PDHG	22984	8.6134	27	5.29
rs200	3025	0.23%	2000	EG	9578	0.7911	6	2.95
				PDHG	11038	0.9183	11	3.10
rs228	1919	0.88%	2000	EG	5713	0.7622	8	3.04
				PDHG	8327	0.7513	7	2.99
rs365	4704	0.47%	2000	EG	17474	3.7162	2	2.50
				PDHG	15037	3.8123	3	3.00
rs1555	7479	0.12%	2000	EG	11759	1.4712	7	5.83
				PDHG	9187	1.5198	11	5.51

Table 2: Results for randomly generated sparse SDP instances. The dimension n is the matrix order and m is the number of equality constraints. The column *active* gives the number of projections for which $\hat{w}^T x = \mu$ at the solution (the total number of projections is twice # iter.). *Newton* gives the average number of Newton steps (averaged over the active projections). Inactive projections are available in closed form.

7 Conclusion

We developed a bounded exact-penalty framework for applying Bregman primal–dual proximal methods to conic linear programs. The main idea is to replace the original conic LP by an equivalent formulation over truncated cones. Generalized projections on truncated cones are often easier to define and compute than projections on the full cones, and the truncation to bounded domains simplifies the convergence analysis.

For strictly feasible self-dual conic LPs, the required bounds are available explicitly. Combined with self-dual embeddings, this observation gives a systematic procedure for obtaining bounded reformulations of general conic LPs, without requiring problem-specific bounds on the primal or dual optimal sets. The resulting problems can be solved by Bregman PDHG or Bregman extragradient methods, with subproblems posed on truncated cones rather than on the full cones.

We also analyzed the convergence of the Bregman PDHG and extragradient methods with line search, applied to self-dual convex problems. For self-dual conic LPs, the bounds are stated directly in terms of the exact-penalty objective. They extend to distances generated by logarithmically homogeneous kernels, for which the Bregman radius on the truncated cone is infinite, and give an $O((\log k)/k)$ rate for the averaged iterates. The rate improves to $O(1/k)$ if the kernel is finite at a solution of the self-dual conic LP.

Numerical experiments on nonnegative trigonometric polynomial optimization and sparse semidefinite optimization indicate that the approach is promising for cone families with efficient Bregman projections. Further algorithmic improvements are needed before it can serve as a practical general-purpose first-order solver for structured conic LPs. Two directions are restarting [8, 9, 42] and preconditioning [7, 52]. Both have been highly effective in first-order methods for LPs and conic LPs. It remains to determine how these techniques should be adapted to Bregman primal–dual proximal methods, and how they interact with the exact-penalty formulations studied here.

A Mixed-form conic LPs and exact penalties

This appendix records the mixed-form variants of the constructions used in the main text. The main text presents the exact-penalty reformulation for conic LPs written without explicit free variables or equality blocks. Here we give the corresponding formulation for mixed-form conic LPs, which include free variables and explicit equality constraints. This form is also useful for self-dual LPs whose cones are not proper and for the Ye–Todd–Mizuno embedding in Appendix B.

Consider the primal–dual pair

$$\begin{aligned} \text{Primal:} \quad & \text{minimize} && c_1^T x_1 + c_2^T x_2 \\ & \text{subject to} && A_{11}x_1 + A_{12}x_2 - b_1 \in L_1 \\ & && A_{21}x_1 + A_{22}x_2 - b_2 = 0 \\ & && x_1 \in K_1, \end{aligned} \tag{89a}$$

$$\begin{aligned} \text{Dual:} \quad & \text{maximize} && b_1^T z_1 + b_2^T z_2 \\ & \text{subject to} && -A_{11}^T z_1 - A_{21}^T z_2 + c_1 \in K_1^* \\ & && -A_{12}^T z_1 - A_{22}^T z_2 + c_2 = 0 \\ & && z_1 \in L_1^*. \end{aligned} \tag{89b}$$

Here $K_1 \subseteq \mathbb{R}^{n_1}$ and $L_1 \subseteq \mathbb{R}^{m_1}$ are proper cones, $x_2 \in \mathbb{R}^{n_2}$ and $z_2 \in \mathbb{R}^{m_2}$ are free variables, and the submatrices A_{ij} have size $m_i \times n_j$.

A.1 Strict consistency

The primal problem (89a) is a special case of (1a) with $K = K_1 \times \mathbb{R}^{n_2}$ and $L = L_1 \times \{0\}$. Hence

$$K^* = (K_1 \times \mathbb{R}^{n_2})^* = K_1^* \times \{0\}, \qquad L^* = (L_1 \times \{0\})^* = L_1^* \times \mathbb{R}^{m_2},$$

and

$$\operatorname{lin}(K^*) = \operatorname{lin}(K_1^*) \times \{0\} = \{0\} \times \{0\}, \qquad \operatorname{lin}(L^*) = \operatorname{lin}(L_1^*) \times \mathbb{R}^{m_2} = \{0\} \times \mathbb{R}^{m_2}.$$

The condition (15) therefore means that the matrix $\begin{bmatrix} A_{21} & A_{22} \end{bmatrix}$ has full row rank m_2. The primal LP (89a) is strictly consistent if and only if this rank condition holds and there exists $\hat{x} = (\hat{x}_1, \hat{x}_2) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ such that

$$\hat{x}_1 \in \operatorname{int}(K_1), \qquad A_{11}\hat{x}_1 + A_{12}\hat{x}_2 - b_1 \in \operatorname{int}(L_1), \qquad A_{21}\hat{x}_1 + A_{22}\hat{x}_2 - b_2 = 0.$$

Similarly, the dual LP (89b) is strictly consistent if and only if $\operatorname{rank}\left(\begin{bmatrix} A_{12}^T & A_{22}^T \end{bmatrix}\right) = n_2$, and there exists $\hat{z} = (\hat{z}_1, \hat{z}_2) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$ such that

$$\hat{z}_1 \in \operatorname{int}(L_1^*), \qquad -A_{11}^T\hat{z}_1 - A_{21}^T\hat{z}_2 + c_1 \in \operatorname{int}(K_1^*), \qquad -A_{12}^T\hat{z}_1 - A_{22}^T\hat{z}_2 + c_2 = 0.$$

A.2 The self-dual case

A useful special case of (89) occurs when it is self-dual: $b = -c$, $L_1 = K_1^*$, and $A = -A^T$. That is, the problem (89) has the form

$$\begin{aligned} & \text{minimize} && -b_1^T x_1 - b_2^T x_2 \\ & \text{subject to} && A_{11}x_1 + A_{12}x_2 - b_1 \in K_1^*, \\ & && -A_{12}^T x_1 + A_{22}x_2 - b_2 = 0, \\ & && x_1 \in K_1, \end{aligned}$$

where K_1 is proper, $A_{11} = -A_{11}^T$, and $A_{22} = -A_{22}^T$.

Splitting the free variable x_2 as $x_2 = u_2 - v_2$ gives the equivalent self-dual LP over the proper cone $K_1 \times \mathbb{R}^{n_2}_+ \times \mathbb{R}^{n_2}_+$:

$$\begin{array}{ll} \text{minimize} & -b_1^T x_1 - b_2^T u_2 + b_2^T v_2 \\ \text{subject to} & \begin{bmatrix} A_{11} & A_{12} & -A_{12} \\ -A_{12}^T & A_{22} & -A_{22} \\ A_{12}^T & -A_{22} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ u_2 \\ v_2 \end{bmatrix} - \begin{bmatrix} b_1 \\ b_2 \\ -b_2 \end{bmatrix} \in K_1^* \times \mathbb{R}^{n_2}_+ \times \mathbb{R}^{n_2}_+, \\ & (x_1, u_2, v_2) \in K_1 \times \mathbb{R}^{n_2}_+ \times \mathbb{R}^{n_2}_+. \end{array}$$

The cone $K_1 \times \mathbb{R}^{n_2}_+ \times \mathbb{R}^{n_2}_+$ is proper. Thus a self-dual LP over a nonproper cone can be represented as an equivalent self-dual LP over a proper cone after making the lineality space explicit.

A.3 Exact penalties for mixed-form LPs

The exact-penalty construction in Section 3 extends to the mixed-form pair (89). The conic constraint blocks are penalized by the corresponding Minkowski gauges, whereas the equality blocks are penalized by dual norms.

Assume that the primal and dual optima are attained, and that there exist optimal solutions $x^\star = (x_1^\star, x_2^\star)$ and $z^\star = (z_1^\star, z_2^\star)$ satisfying

$$\hat{w}_1^T x_1^\star < 1, \qquad \|x_2^\star\| < 1, \qquad \hat{s}_1^T z_1^\star < 1, \qquad \|z_2^\star\|' < 1,$$

where $\|\cdot\|$ is a norm on $\mathbb{R}^{n_2}$ and $\|\cdot\|'$ is a norm on $\mathbb{R}^{m_2}$. Then the following penalized problems have the same optimal values as the mixed-form primal and dual LPs. Every optimal solution of a penalized problem is optimal for the corresponding original LP, and conversely every original optimal solution satisfying the imposed bounds is optimal for the corresponding penalized problem:

$$\begin{array}{lll} \text{Primal:} & \text{minimize} & c_1^T x_1 + c_2^T x_2 + \sigma_{L_1}(b_1 - A_{11}x_1 - A_{12}x_2, \hat{s}_1) + \|b_2 - A_{21}x_1 - A_{22}x_2\|'_* \quad (90a) \\ & \text{subject to} & x_1 \in K_1, \;\; \hat{w}_1^T x_1 \le 1 \\ & & \|x_2\| \le 1, \\ \text{Dual:} & \text{maximize} & b_1^T z_1 + b_2^T z_2 - \sigma_{K_1^*}(A_{11}^T z_1 + A_{21}^T z_2 - c_1, \hat{w}_1) - \|-A_{12}^T z_1 - A_{22}^T z_2 + c_2\|_* \quad (90b) \\ & \text{subject to} & z_1 \in L_1^*, \;\; \hat{s}_1^T z_1 \le 1 \\ & & \|z_2\|' \le 1. \end{array}$$

In these problems, $\|\cdot\|_*$ and $\|\cdot\|'_*$ are the dual norms of $\|\cdot\|$ and $\|\cdot\|'$. The problems (90) are a special case of (2) with $f = \delta_S$, the indicator function of the set

$$S = \{(x_1, x_2) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \mid x_1 \in K_1, \; \hat{w}_1^T x_1 \le 1, \; \|x_2\| \le 1\},$$

and $g = \delta_T^*$, the support function of the set

$$T = \{(z_1, z_2) \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \mid z_1 \in L_1^*, \; \hat{s}_1^T z_1 \le 1, \; \|z_2\|' \le 1\}.$$

B Ye–Todd–Mizuno self-dual embedding

This appendix presents the Ye–Todd–Mizuno self-dual embedding for conic LPs

$$\begin{aligned} &\text{Primal:} && \text{minimize} && c^T x && \text{(91a)}\\ &&& \text{subject to} && Ax - b = s \\ &&&&& x \in K, \quad s \in L, \\ &\text{Dual:} && \text{maximize} && b^T z && \text{(91b)}\\ &&& \text{subject to} && -A^T z + c = w \\ &&&&& z \in L^*, \quad w \in K^*. \end{aligned}$$

The embedding converts the primal–dual pair into a single self-dual conic LP with known strictly feasible points and explicit bounds on the optimal set of the embedded problem.

We assume the cones K and L are nonempty closed convex cones and $A \in \mathbb{R}^{m\times n}$. If the cones are not proper, one can make the linear hulls and lineality spaces of K and L explicit, which leads to a pair of equality-constrained primal and dual problems of the form

$$\begin{aligned} &\text{Primal:} && \text{minimize} && c_1^T x_1 + c_2^T x_2 && \text{(92a)}\\ &&& \text{subject to} && \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \\ &&&&& (x_1, x_2) \in K_1 \times \mathbb{R}^{n_2}, \quad (s_1, s_2) \in L_1 \times \{0\}, \\ &\text{Dual:} && \text{maximize} && b_1^T z_1 + b_2^T z_2 && \text{(92b)}\\ &&& \text{subject to} && -\begin{bmatrix} A_{11}^T & A_{21}^T \\ A_{12}^T & A_{22}^T \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \\ &&&&& (z_1, z_2) \in L_1^* \times \mathbb{R}^{m_2}, \quad (w_1, w_2) \in K_1^* \times \{0\}, \end{aligned}$$

where K_1, L_1 are proper cones in $\mathbb{R}^{n_1}$ and $\mathbb{R}^{m_1}$, respectively, and $A_{ij} \in \mathbb{R}^{m_i \times n_j}$ for $i, j = 1, 2$. Important special cases are LPs in standard form ($n_2 = m_1 = 0$) and LPs in inequality form ($n_1 = m_2 = 0$). It is useful to note that a problem in the form (92) can be converted to the simpler (91) with proper cones K, L, by the familiar procedure of splitting free variables into nonnegative parts and writing each equality as two opposing inequalities, for example, as

$$\begin{aligned} &\text{Primal:} && \text{minimize} && c_1^T x_1 + c_2^T x_{2+} - c_2^T x_{2-} && \text{(93a)}\\ &&& \text{subject to} && \begin{bmatrix} A_{11} & A_{12} & -A_{12} \\ A_{21} & A_{22} & -A_{22} \\ -A_{21} & -A_{22} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_{2+} \\ x_{2-} \end{bmatrix} - \begin{bmatrix} b_1 \\ b_2 \\ -b_2 \end{bmatrix} = \begin{bmatrix} s_1 \\ s_{2+} \\ s_{2-} \end{bmatrix} \\ &&&&& (x_1, x_{2+}, x_{2-}) \in K_1 \times \mathbb{R}_+^{n_2} \times \mathbb{R}_+^{n_2} \\ &&&&& (s_1, s_{2+}, s_{2-}) \in L_1 \times \mathbb{R}_+^{m_2} \times \mathbb{R}_+^{m_2}, \\ &\text{Dual:} && \text{maximize} && b_1^T z_1 + b_2^T z_{2+} - b_2^T z_{2-} && \text{(93b)}\\ &&& \text{subject to} && -\begin{bmatrix} A_{11}^T & A_{21}^T & -A_{21}^T \\ A_{12}^T & A_{22}^T & -A_{22}^T \\ -A_{12}^T & -A_{22}^T & A_{22}^T \end{bmatrix} \begin{bmatrix} z_1 \\ z_{2+} \\ z_{2-} \end{bmatrix} + \begin{bmatrix} c_1 \\ c_2 \\ -c_2 \end{bmatrix} = \begin{bmatrix} w_1 \\ w_{2+} \\ w_{2-} \end{bmatrix} \\ &&&&& (z_1, z_{2+}, z_{2-}) \in L_1^* \times \mathbb{R}_+^{m_2} \times \mathbb{R}_+^{m_2} \\ &&&&& (w_1, w_{2+}, w_{2-}) \in K_1^* \times \mathbb{R}_+^{n_2} \times \mathbb{R}_+^{n_2}. \end{aligned}$$

B.1 The embedding

The Ye–Todd–Mizuno self-dual embedding of (91) is

$$\begin{array}{ll} \text{minimize} & \theta \\ \text{subject to} & \begin{bmatrix} 0 & -A^T & c & q_x \\ A & 0 & -b & q_z \\ -c^T & b^T & 0 & q_\tau \\ -q_x^T & -q_z^T & -q_\tau & 0 \end{bmatrix} \begin{bmatrix} x \\ z \\ \tau \\ \theta \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} w \\ s \\ \kappa \\ \eta \end{bmatrix} \\ & (x, z, \tau, \theta) \in K \times L^* \times \mathbb{R}_+ \times \mathbb{R} \\ & (w, s, \kappa, \eta) \in K^* \times L \times \mathbb{R}_+ \times \{0\}. \end{array} \tag{94}$$

The parameters q_x, q_z, q_τ are determined by the condition that given points

$$\hat{x} \in \operatorname{ri}(K), \qquad \hat{z} \in \operatorname{ri}(L^*), \qquad \hat{w} \in \operatorname{ri}(K^*), \qquad \hat{s} \in \operatorname{ri}(L),$$

are feasible with $\hat{\tau} = \hat{\kappa} = 1$, $\hat{\eta} = 0$, and $\hat{\theta}$ chosen so that the embedding equations are satisfied:

$$\hat{\theta} = \hat{w}^T\hat{x} + \hat{z}^T\hat{s} + 1, \qquad \begin{bmatrix} q_x \\ q_z \\ q_\tau \end{bmatrix} = \frac{1}{\hat{\theta}} \left(\begin{bmatrix} \hat{w} \\ \hat{s} \\ 1 \end{bmatrix} - \begin{bmatrix} 0 & -A^T & c \\ A & 0 & -b \\ -c^T & b^T & 0 \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{z} \\ 1 \end{bmatrix} \right).$$

The embedding (94) is of the general form (8). It can be written as

$$\begin{array}{ll} \text{minimize} & -\tilde{b}^T\tilde{x} \\ \text{subject to} & \tilde{A}\tilde{x} - \tilde{b} = \tilde{s} \\ & \tilde{x} \in \tilde{K}, \quad \tilde{s} \in \tilde{K}^*, \end{array}$$

with

$$\tilde{x} = \begin{bmatrix} x \\ z \\ \tau \\ \theta \end{bmatrix}, \qquad \tilde{s} = \begin{bmatrix} w \\ s \\ \kappa \\ \eta \end{bmatrix}, \qquad \tilde{b} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \end{bmatrix}, \qquad \tilde{A} = \begin{bmatrix} 0 & -A^T & c & q_x \\ A & 0 & -b & q_z \\ -c^T & b^T & 0 & q_\tau \\ -q_x^T & -q_z^T & -q_\tau & 0 \end{bmatrix} \tag{95}$$

and

$$\tilde{K} = K \times L^* \times \mathbb{R}_+ \times \mathbb{R}, \qquad \tilde{K}^* = K^* \times L \times \mathbb{R}_+ \times \{0\}. \tag{96}$$

The self-dual embedding of (92) is

$$\begin{array}{ll} \text{minimize} & \theta \\ \text{subject to} & \begin{bmatrix} 0 & 0 & -A_{11}^T & -A_{21}^T & c_1 & q_{x1} \\ 0 & 0 & -A_{12}^T & -A_{22}^T & c_2 & q_{x2} \\ A_{11} & A_{12} & 0 & 0 & -b_1 & q_{z1} \\ A_{21} & A_{22} & 0 & 0 & -b_2 & q_{z2} \\ -c_1^T & -c_2^T & b_1^T & b_2^T & 0 & q_\tau \\ -q_{x1}^T & -q_{x2}^T & -q_{z1}^T & -q_{z2}^T & -q_\tau & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ z_1 \\ z_2 \\ \tau \\ \theta \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ s_1 \\ s_2 \\ \kappa \\ \eta \end{bmatrix} \\ & (x_1, x_2, z_1, z_2, \tau, \theta) \in K_1 \times \mathbb{R}^{n_2} \times L_1^* \times \mathbb{R}^{m_2} \times \mathbb{R}_+ \times \mathbb{R} \\ & (w_1, w_2, s_1, s_2, \kappa, \eta) \in K_1^* \times \{0\} \times L_1 \times \{0\} \times \mathbb{R}_+ \times \{0\}. \end{array} \tag{97}$$

The parameters $q_{x1}, q_{x2}, q_{z1}, q_{z2}, q_\tau$ are determined by the condition that given points

$$(\hat{x}_1, \hat{x}_2) \in \operatorname{int}(K_1) \times \mathbb{R}^{n_2}, \quad (\hat{z}_1, \hat{z}_2) \in \operatorname{int}(L_1^*) \times \mathbb{R}^{m_2}, \quad (\hat{w}_1, \hat{w}_2) \in \operatorname{int}(K_1^*) \times \{0\}, \quad (\hat{s}_1, \hat{s}_2) \in \operatorname{int}(L_1) \times \{0\}$$

are feasible with $\hat{\tau} = \hat{\kappa} = 1$, $\hat{\eta} = 0$, and $\hat{\theta} = \hat{w}_1^T\hat{x}_1 + \hat{s}_1^T\hat{z}_1 + 1$. Here necessarily $\hat{w}_2 = 0$ and $\hat{s}_2 = 0$. Problem (97) is also a self-dual problem of the form (8).

Strict consistency The self-dual embedding (94) satisfies the strong consistency condition (30) by construction, since

$$\begin{array}{rcl}(\hat{x},\hat{z},1,\hat{\theta}) & \in & \mathrm{ri}(\tilde{K}) \\ & = & \mathrm{ri}(K)\times\mathrm{ri}(L^*)\times\mathbb{R}_{++}\times\mathbb{R}, \\ (\hat{w},\hat{s},1,\hat{\eta}) & \in & \mathrm{ri}(\tilde{K}^*) \\ & = & \mathrm{ri}(K^*)\times\mathrm{ri}(L)\times\mathbb{R}_{++}\times\{0\}\end{array}$$

where $\tilde{K}$ is defined in (96). The additional condition (31) needed for strict consistency,

$$\left.\begin{array}{l} y\in\mathrm{lin}(\tilde{K}) \\ \tilde{A}y\in\mathrm{lin}(\tilde{K}^*)\end{array}\right\} \quad\Longleftrightarrow\quad y=0, \tag{98}$$

generally requires additional assumptions on A, K, L in (91).

- The embedding is strictly consistent if K is pointed and L has nonempty interior. Then

  $$\mathrm{lin}(\tilde{K})=\{0\}\times\{0\}\times\{0\}\times\mathbb{R}, \qquad \mathrm{lin}(\tilde{K}^*)=\mathrm{lin}(K^*)\times\mathrm{lin}(L)\times\{0\}\times\{0\},$$

  so the condition (98) requires that

  $$\begin{bmatrix} 0 & -A^T & c & q_x \\ A & 0 & -b & q_z \\ -c^T & b^T & 0 & q_\tau \\ -q_x^T & -q_z^T & -q_\tau & 0 \end{bmatrix}\begin{bmatrix} 0\\0\\0\\\theta\end{bmatrix}=\begin{bmatrix} w\\s\\0\\0\end{bmatrix}, \qquad w\in\mathrm{lin}(K^*), \quad s\in\mathrm{lin}(L)$$

  holds only if $\theta=0$. To show this, we take the inner product with $(\hat{x},\hat{z},\hat{\tau},\hat{\theta})$ on both sides of the equation:

  $$\hat{x}^Tw+\hat{z}^Ts=\begin{bmatrix}\hat{x}\\\hat{z}\\\hat{\tau}\\\hat{\theta}\end{bmatrix}^T\begin{bmatrix} 0 & -A^T & c & q_x \\ A & 0 & -b & q_z \\ -c^T & b^T & 0 & q_\tau \\ -q_x^T & -q_z^T & -q_\tau & 0 \end{bmatrix}\begin{bmatrix} 0\\0\\0\\\theta\end{bmatrix}=-\begin{bmatrix}\hat{w}\\\hat{s}\\\hat{\kappa}\\-1\end{bmatrix}^T\begin{bmatrix} 0\\0\\0\\\theta\end{bmatrix}=\theta.$$

  The two inner products on the left-hand side are zero because $w\in\mathrm{lin}(K^*)=\mathrm{span}(K)^\perp$, $\hat{x}\in K$ and $s\in\mathrm{lin}(L)=\mathrm{span}(L^*)^\perp$, $\hat{z}\in L^*$.

- If K is not pointed or L has empty interior, the problem can be reformulated as the LP (92), where K_1 and L_1 are proper cones and the dimension of A_{ij} is $m_i\times n_j$. The embedding (97) is strictly consistent if

  $$\mathrm{rank}\begin{bmatrix} A_{21} & A_{22}\end{bmatrix}=m_2, \qquad \mathrm{rank}\begin{bmatrix} A_{12}\\A_{22}\end{bmatrix}=n_2. \tag{99}$$

  To see this, first note that the embedding is strongly consistent by construction, with known feasible points

  $$(\hat{x}_1,\hat{x}_2,\hat{z}_1,\hat{z}_2,\hat{\tau},\hat{\theta})\in\mathrm{ri}(K_1)\times\mathbb{R}^{n_2}\times\mathrm{ri}(L_1^*)\times\mathbb{R}^{m_2}\times\mathbb{R}_{++}\times\mathbb{R},$$

  and

  $$(\hat{w}_1,\hat{w}_2,\hat{s}_1,\hat{s}_2,\hat{\kappa},\hat{\eta})\in\mathrm{ri}(K_1^*)\times\{0\}\times\mathrm{ri}(L_1)\times\{0\}\times\mathbb{R}_{++}\times\{0\},$$

where $\hat\theta = \hat w_1^T \hat x_1 + \hat s_1^T \hat z_1 + 1$. The condition (98) requires that

$$\begin{bmatrix} 0 & 0 & -A_{11}^T & -A_{21}^T & c_1 & q_{x1} \\ 0 & 0 & -A_{12}^T & -A_{22}^T & c_2 & q_{x2} \\ A_{11} & A_{12} & 0 & 0 & -b_1 & q_{z1} \\ A_{21} & A_{22} & 0 & 0 & -b_2 & q_{z2} \\ -c_1^T & -c_2^T & b_1^T & b_2^T & 0 & q_\tau \\ -q_{x1}^T & -q_{x2}^T & -q_{z1}^T & -q_{z2}^T & -q_\tau & 0 \end{bmatrix} \begin{bmatrix} 0 \\ x_2 \\ 0 \\ z_2 \\ 0 \\ \theta \end{bmatrix} = 0 \tag{100}$$

implies $x_2 = 0$, $z_2 = 0$, $\theta = 0$. We first show that it implies $\theta = 0$. Taking the inner product of (100) with $(\hat x_1, \hat x_2, \hat z_1, \hat z_2, 1, \hat\theta)$, we see that

$$0 = \begin{bmatrix} \hat w_1 \\ \hat w_2 \\ \hat s_1 \\ \hat s_2 \\ 1 \\ -1 \end{bmatrix}^T \begin{bmatrix} 0 \\ x_2 \\ 0 \\ z_2 \\ 0 \\ \theta \end{bmatrix} = \hat w_2^T x_2 + \hat s_2^T z_2 - \theta = -\theta.$$

If $\theta = 0$ in (100), then

$$\begin{bmatrix} 0 & -A_{21}^T \\ 0 & -A_{22}^T \\ A_{12} & 0 \\ A_{22} & 0 \\ -c_2^T & b_2^T \\ -q_{x2}^T & -q_{z2}^T \end{bmatrix} \begin{bmatrix} x_2 \\ z_2 \end{bmatrix} = 0,$$

and the rank conditions (99) imply that $x_2 = 0$, $z_2 = 0$.

- The primal problem in (91) is in the standard LP form if $L = \{0\}$ and K is a proper cone. The self-dual embedding is strictly consistent if A has full row rank m. This is a special case of the rank conditions (99) with $m_1 = 0$, $m_2 = m$, $n_1 = n$, $n_2 = 0$.
- The primal problem in (91) is in inequality form if $K = \mathbb{R}^n$ and L is a proper cone. The self-dual embedding is strictly consistent if A has rank n. This is a special case of the rank conditions (99) with $m_1 = m$, $m_2 = 0$, $n_1 = 0$, $n_2 = n$.

Optimal solution The Ye–Todd–Mizuno embedding is a standard technique for initialization and infeasibility detection in implementations of interior-point methods. In this paper, it is used as a tool for constructing an equivalent conic LP with simple, known bounds on the optimal set. With this in mind, it is useful to note that if the primal and dual LPs (91) are feasible, then κ must be zero at the optimum of (94). If at the optimum $\kappa > 0$, then $\tau = 0$ and the optimal solution x, z of the embedded problem satisfies

$$-A^T z \in \mathrm{ri}(K^*), \qquad Ax \in \mathrm{ri}(L), \qquad -c^T x + b^T z > 0, \qquad x \in K, \qquad z \in L^*.$$

If $b^T z > 0$ this shows that the primal problem in (91) is infeasible. If $c^T x < 0$ it shows the dual problem is infeasible. If $\tau > 0$ at the optimum of (94), it is easily verified that x/τ, s/τ, w/τ, z/τ are optimal for (91).

B.2 Exact-penalty formulation

If the embedding is strictly consistent, then its optimal value is zero and the optimal set is nonempty and bounded, as explained in Section 2.3. Optimal solutions (x, z, τ, θ) and (w, s, κ, η) of (94) satisfy $\theta = 0$ and

$$\begin{aligned} \hat{w}^T x + \hat{s}^T z + \tau &= \hat{w}^T x + \hat{s}^T z + \hat{\kappa}\tau + \hat{\eta}\theta \\ &= \hat{w}^T \hat{x} + \hat{s}^T \hat{z} + \hat{\kappa}\hat{\tau} + \hat{\eta}\hat{\theta} - w^T \hat{x} - s^T \hat{z} - \kappa\hat{\tau} - \eta\hat{\theta} \\ &\leq \hat{w}^T \hat{x} + \hat{s}^T \hat{z} + \hat{\kappa}\hat{\tau} + \hat{\eta}\hat{\theta} \\ &= \hat{w}^T \hat{x} + \hat{s}^T \hat{z} + 1 \\ &= \hat{\theta}. \end{aligned} \tag{101}$$

Hence, optimal x, z, τ satisfy

$$x \in K, \qquad z \in L^*, \qquad \hat{w}^T x \leq \hat{\theta}, \qquad \hat{s}^T z \leq \hat{\theta}, \qquad 0 \leq \tau \leq \hat{\theta}.$$

Assume now that K and L^* are proper cones. Then $\hat{w} \in \operatorname{int}(K^*)$ and $\hat{s} \in \operatorname{int}(L)$. For any $\mu > \hat{\theta}$, the following exact-penalty formulation has the same optimal value as (94) and preserves all embedded optimal solutions satisfying the displayed bounds:

$$\begin{array}{ll} \text{minimize} & \theta + \mu\sigma_{K^*}(A^T z - \tau c - \theta q_x, \hat{w}) + \mu\sigma_L(-Ax + \tau b - \theta q_z, \hat{s}) \\ & \quad + \mu \max\{0, c^T x - b^T z - \theta q_\tau\} + |q_x^T x + q_z^T z + q_\tau \tau - 1| \\ \text{subject to} & x \in K, \ \hat{w}^T x \leq \mu \\ & z \in L^*, \ \hat{s}^T z \leq \mu \\ & 0 \leq \tau \leq \mu \\ & |\theta| \leq 1, \end{array}$$

where $\mu > \hat{\theta} = \hat{w}^T \hat{x} + \hat{z}^T \hat{s} + 1$.

C Proofs of the convergence results

This appendix contains the proofs of the results in Section 5.

The proofs use the following property of Bregman proximal operators; see, *e.g.*, [21,35], [59, property 1]. Let $y \in \operatorname{int}(\operatorname{dom}\phi)$ and $a \in \mathbb{R}^n$, and suppose

$$u = \operatorname*{argmin}_x \left(f(x) + a^T x + d(x, y)\right)$$

exists, is unique, and lies in $\operatorname{int}(\operatorname{dom}\phi)$. Then, for all $x \in \operatorname{dom} f \cap \operatorname{dom}\phi$,

$$f(u) - f(x) + a^T(u - x) \leq d(x, y) - d(x, u) - d(u, y). \tag{102}$$

C.1 Proof of Theorem 3

We first derive an inequality for one iteration of Algorithm 1. Fix $x, z \in D$ and $k \geq 0$. The iterates x_{k+1} and z_{k+1} are the points (58) computed with $t = t_k$. Applying (102) to (58a) gives

$$t_k \left(f(x_{k+1}) - f(x) + (Az_k - b)^T (x_{k+1} - x)\right) \leq d(x, x_k) - d(x, x_{k+1}) - d(x_{k+1}, x_k). \tag{103}$$

Similarly, applying (102) to (58b) gives

$$t_k \left(f(z_{k+1}) - f(z) + (A(2x_{k+1} - x_k) - b)^T (z_{k+1} - z) \right) \leq d(z, z_k) - d(z, z_{k+1}) - d(z_{k+1}, z_k). \tag{104}$$

Using the definition of $\mathcal{L}$ in (26) and $A^T = -A$, we write

$$\begin{aligned}
& t_k \left(\mathcal{L}(x_{k+1}, z) - \mathcal{L}(x, z_{k+1}) \right) \\
&\quad = t_k \big(f(x_{k+1}) - f(x) + f(z_{k+1}) - f(z) - b^T (x_{k+1} - x) - b^T (z_{k+1} - z) - z^T A x_{k+1} + z_{k+1}^T A x \big) \\
&\quad = t_k \big(f(x_{k+1}) - f(x) + (Az_k - b)^T (x_{k+1} - x) + f(z_{k+1}) - f(z) + (A(2x_{k+1} - x_k) - b)^T (z_{k+1} - z) \big) \\
&\qquad + t_k (z - z_k)^T A (x - x_k) - t_k (z - z_{k+1})^T A (x - x_{k+1}) - t_k (z_{k+1} - z_k)^T A (x_{k+1} - x_k) \\
&\quad \leq d(x, x_k) + d(z, z_k) + t_k (z - z_k)^T A (x - x_k) - d(x, x_{k+1}) - d(z, z_{k+1}) - t_k (z - z_{k+1})^T A (x - x_{k+1}) \\
&\qquad - d(x_{k+1}, x_k) - d(z_{k+1}, z_k) - t_k (z_{k+1} - z_k)^T A (x_{k+1} - x_k).
\end{aligned}$$

The second equality can be verified by expanding the inner products and using the skew-symmetry of A. The inequality follows from (103) and (104). Finally, the line-search condition (59), which holds for the accepted stepsize t_k and points x_{k+1}, z_{k+1}, gives

$$\begin{aligned}
& t_k \left(\mathcal{L}(x_{k+1}, z) - \mathcal{L}(x, z_{k+1}) \right) \\
&\quad \leq d(x, x_k) + d(z, z_k) + t_k (z - z_k)^T A (x - x_k) - d(x, x_{k+1}) - d(z, z_{k+1}) - t_k (z - z_{k+1})^T A (x - x_{k+1}) \\
&\qquad - (1 - \delta) \left(d(x_{k+1}, x_k) + d(z_{k+1}, z_k) \right).
\end{aligned} \tag{105}$$

Replacing k by i in (105) and summing from $i = 0$ to $i = k$, and omitting the nonpositive terms with coefficient $-(1 - \delta)$, gives

$$\begin{aligned}
& \sum_{i=0}^{k} t_i \left(\mathcal{L}(x_{i+1}, z) - \mathcal{L}(x, z_{i+1}) \right) \\
&\quad \leq \sum_{i=0}^{k} \big(d(x, x_i) + d(z, z_i) + t_i (z - z_i)^T A (x - x_i) - d(x, x_{i+1}) - d(z, z_{i+1}) - t_i (z - z_{i+1})^T A (x - x_{i+1}) \big) \\
&\quad = d(x, x_0) + d(z, z_0) + t_0 (z - z_0)^T A (x - x_0) - d(x, x_{k+1}) - d(z, z_{k+1}) - t_k (z - z_{k+1})^T A (x - x_{k+1}) \\
&\qquad + \sum_{i=1}^{k} (t_i - t_{i-1}) (z - z_i)^T A (x - x_i).
\end{aligned} \tag{106}$$

Since $x, z, x_i, z_i \in D$, each of the inner products in (106) is bounded in absolute value by ρ_2. Moreover,

$$t_0 + t_k + \sum_{i=1}^{k} |t_i - t_{i-1}| = 2t_0 + 2 \sum_{i=1}^{k} (t_i - t_{i-1})_+ = 2(t_0 + \Gamma_k),$$

because $t_k - t_0 = \sum_{i=1}^{k} (t_i - t_{i-1})$. Hence the inner product terms in (106) add up to at most $2\rho_2(t_0 + \Gamma_k)$. Omitting the terms $-d(x, x_{k+1}) - d(z, z_{k+1}) \leq 0$, we obtain

$$\sum_{i=0}^{k} t_i \left(\mathcal{L}(x_{i+1}, z) - \mathcal{L}(x, z_{i+1}) \right) \leq d(x, x_0) + d(z, z_0) + 2\rho_2(t_0 + \Gamma_k). \tag{107}$$

Since $\mathcal{L}(\cdot, z)$ is convex and $\mathcal{L}(x, \cdot)$ is concave, and x_k^{avg}, z_k^{avg} are convex combinations of $x_1, \ldots, x_{k+1}$ and $z_1, \ldots, z_{k+1}$ with weights t_i / T_k,

$$\mathcal{L}(x_k^{\text{avg}}, z) - \mathcal{L}(x, z_k^{\text{avg}}) \leq \frac{1}{T_k} \sum_{i=0}^{k} t_i \left(\mathcal{L}(x_{i+1}, z) - \mathcal{L}(x, z_{i+1}) \right).$$

Combined with (107) this gives (69).

The bound (70) follows by taking the supremum over $z \in D$ and the infimum over $x \in D$ on both sides of the inequality (69).

C.2 Proof of Theorem 4

Fix $x \in D$ and $i \geq 0$. Applying (102) to the second update (60b) in Algorithm 2 at iteration i (so that $t = t_i$, $\bar{x} = \bar{x}_i$, and $x^+ = x_{i+1}$), we obtain

$$\begin{aligned} f(x) + (A\bar{x}_i - b)^T x + \frac{1}{t_i} d(x, x_i) &\geq f(x_{i+1}) + (A\bar{x}_i - b)^T x_{i+1} + \frac{1}{t_i} d(x_{i+1}, x_i) + \frac{1}{t_i} d(x, x_{i+1}) \\ &\geq f(\bar{x}_i) - b^T \bar{x}_i + \frac{1}{t_i} d(x, x_{i+1}). \end{aligned}$$

The second inequality is the stepsize condition (61). Rearranging and using the definition of $\mathcal{L}$ in (26) gives

$$\mathcal{L}(\bar{x}_i, x) = f(\bar{x}_i) - f(x) + b^T(x - \bar{x}_i) - x^T A\bar{x}_i \leq \frac{1}{t_i} \left(d(x, x_i) - d(x, x_{i+1}) \right).$$

Multiplying by t_i and summing from $i = 0$ to $i = k$ gives

$$\sum_{i=0}^{k} t_i \mathcal{L}(\bar{x}_i, x) \leq d(x, x_0) - d(x, x_{k+1}) \leq d(x, x_0).$$

Since $\mathcal{L}(\cdot, x)$ is convex and x_k^{avg} is a convex combination of $\bar{x}_0, \ldots, \bar{x}_k$ with weights t_i / T_k, this implies

$$\mathcal{L}(x_k^{\text{avg}}, x) \leq \frac{d(x, x_0)}{T_k} \quad \text{for all } x \in D.$$

Adding this inequality and the same inequality at $z \in D$, and using $\mathcal{L}(x, x_k^{\text{avg}}) = -\mathcal{L}(x_k^{\text{avg}}, x)$, gives (72).

The bound (73) follows by taking the supremum over $z \in D$ and the infimum over $x \in D$ on both sides of the inequality (72).

C.3 Proof of Theorem 5

We first note some consequences of (79). These properties are well known for logarithmic barrier functions ($\operatorname{dom} \phi = \operatorname{int}(K)$).

- $\nabla\phi(x)^T x = -\nu$ for all $x \in \operatorname{int}(K)$. This can be seen by differentiating (79) with respect to λ at $\lambda = 1$.
- $-\nabla\phi(x) \in K^*$ for all $x \in \operatorname{int}(K)$. This can be shown as follows. If $-\nabla\phi(x) \notin K^*$, there exists a $y \in \operatorname{int}(K)$ with $\nabla\phi(x)^T y > 0$. By convexity and logarithmic homogeneity,

$$\begin{aligned} \phi(ty) &\geq \phi(x) + \nabla\phi(x)^T (ty - x) \\ &= \phi(x) + t\nabla\phi(x)^T y + \nu \\ &> \phi(x) + \nu, \end{aligned}$$

  which contradicts $\phi(ty) = \phi(y) - \nu \log t$ for all $t > 0$.

- The identity in (79) holds for all $x \in \operatorname{dom}\phi$: if $x \in \operatorname{dom}\phi$ and $\lambda > 0$, then $\lambda x \in \operatorname{dom}\phi$ and $\phi(\lambda x) = \phi(x) - \nu \log\lambda$. To see this, define $x_\epsilon = (1-\epsilon)x + \epsilon\hat{x}$ for $\epsilon \in (0,1]$. Since $x \in \operatorname{dom}\phi \subseteq K$ and $\hat{x} \in \operatorname{int}(K)$, the point x_ϵ is in $\operatorname{int}(K)$, so (79) gives $\phi(\lambda x_\epsilon) = \phi(x_\epsilon) - \nu\log\lambda$. The points $\hat{x}$ and $\lambda\hat{x}$ are in $\operatorname{int}(K)$, which is the interior of $\operatorname{dom}\phi$. Since ϕ is closed, $\phi(x_\epsilon) \to \phi(x)$ and $\phi(\lambda x_\epsilon) \to \phi(\lambda x)$ as $\epsilon \to 0$ [53, theorem 7.5].

- ϕ is nonincreasing with respect to the cone order:

$$\phi(v) \geq \phi(u) \tag{108}$$

  if $u, v \in \operatorname{int}(K)$ and $u - v \in K$, and also if $v \in \operatorname{dom}\phi$ and $u - v \in \operatorname{dom}\phi$. In the first case, the inequality follows from convexity and $-\nabla\phi(u) \in K^*$: $\phi(v) \geq \phi(u) + \nabla\phi(u)^T(v-u) \geq \phi(u)$. In the second case, let $w = u - v$, and define $\lambda_1 = \exp(-\phi(v)/\nu)$ and $\lambda_2 = \exp(-\phi(w)/\nu)$. By the previous property, $\phi(v/\lambda_1) = \phi(w/\lambda_2) = 0$. The point $u/(\lambda_1+\lambda_2)$ is a convex combination of v/λ_1 and w/λ_2, so $\phi(u/(\lambda_1+\lambda_2)) \leq 0$ by convexity. Therefore,

$$\phi(u) = \phi\big(u/(\lambda_1+\lambda_2)\big) - \nu\log(\lambda_1+\lambda_2) \leq -\nu\log(\lambda_1+\lambda_2) < -\nu\log\lambda_1 = \phi(v).$$

Proof Fix $x, z \in C$ and $\xi \in (0,1]$, and define $x_\xi = (1-\xi)x + \xi\hat{x}$ and $z_\xi = (1-\xi)z + \xi\hat{x}$. Since $\hat{x} \in \operatorname{int}(K)$ and $\xi > 0$, these points are in $C \cap \operatorname{int}(K) \subseteq D$. Since $x_\xi - \xi\hat{x} = (1-\xi)x \in K$, combining (108) and (79) gives

$$\phi(x_\xi) \leq \phi(\xi\hat{x}) = \phi(\hat{x}) + \nu\log(1/\xi),$$

and similarly $\phi(z_\xi) \leq \phi(\hat{x}) + \nu\log(1/\xi)$. Since $x_\xi, z_\xi \in D$ and $\operatorname{cl} D = C$, the definition (80) of ρ_3 gives

$$\begin{aligned} d(x_\xi, x_0) + d(z_\xi, z_0) &= \phi(x_\xi) + \phi(z_\xi) - \phi(x_0) - \phi(z_0) + \nabla\phi(x_0)^T(x_0 - x_\xi) + \nabla\phi(z_0)^T(z_0 - z_\xi) \\ &\leq 2\nu\log(1/\xi) + \rho_3. \end{aligned}$$

Applying (69) at (x_ξ, z_ξ) therefore gives

$$\mathcal{L}(x_k^{\text{avg}}, z_\xi) - \mathcal{L}(x_\xi, z_k^{\text{avg}}) \leq \frac{2\nu\log(1/\xi) + \rho_3 + 2\rho_2(t_0 + \Gamma_k)}{T_k}, \tag{109}$$

where $\mathcal{L}$ is defined in (75). Since $\mathcal{L}$ is affine in each of its arguments on C,

$$\begin{aligned} \mathcal{L}(x_k^{\text{avg}}, z) &= \mathcal{L}(x_k^{\text{avg}}, z_\xi) + \xi\big(\mathcal{L}(x_k^{\text{avg}}, z) - \mathcal{L}(x_k^{\text{avg}}, \hat{x})\big) \\ &= \mathcal{L}(x_k^{\text{avg}}, z_\xi) + \xi(z - \hat{x})^T(b - Ax_k^{\text{avg}}), \\ -\mathcal{L}(x, z_k^{\text{avg}}) &= -\mathcal{L}(x_\xi, z_k^{\text{avg}}) + \xi\big(\mathcal{L}(\hat{x}, z_k^{\text{avg}}) - \mathcal{L}(x, z_k^{\text{avg}})\big) \\ &= -\mathcal{L}(x_\xi, z_k^{\text{avg}}) + \xi(x - \hat{x})^T(b - Az_k^{\text{avg}}). \end{aligned}$$

Using $A\hat{x} - b = \hat{s}$ and $A^T = -A$ (Assumption B1), we can write the last terms as

$$\begin{aligned} (z - \hat{x})^T(b - Ax_k^{\text{avg}}) &= \hat{s}^T(\hat{x} - z) + (z - \hat{x})^T A(\hat{x} - x_k^{\text{avg}}), \\ (x - \hat{x})^T(b - Az_k^{\text{avg}}) &= \hat{s}^T(\hat{x} - x) + (x - \hat{x})^T A(\hat{x} - z_k^{\text{avg}}). \end{aligned}$$

Since $x, z \in K$ and $\hat{s} \in K^*$, the first terms on the right-hand sides are at most $\hat{s}^T\hat{x}$. The second terms are at most ρ_2, since $\operatorname{cl} D = C$. Hence

$$\mathcal{L}(x_k^{\text{avg}}, z) \leq \mathcal{L}(x_k^{\text{avg}}, z_\xi) + \xi(\hat{s}^T\hat{x} + \rho_2),$$

$$-\mathcal{L}(x, z_k^{\text{avg}}) \quad \leq \quad -\mathcal{L}(x_\xi, z_k^{\text{avg}}) + \xi(\hat{s}^T\hat{x} + \rho_2).$$

Adding the two inequalities and combining with (109) gives

$$\mathcal{L}(x_k^{\text{avg}}, z) - \mathcal{L}(x, z_k^{\text{avg}}) \leq \frac{2\nu\log(1/\xi) + \rho_3 + 2\rho_2(t_0 + \Gamma_k)}{T_k} + 2\xi(\hat{s}^T\hat{x} + \rho_2).$$

The right-hand side does not depend on x and z. Taking the supremum over $z \in C$ and the infimum over $x \in C$ on the left-hand side, and using (76), we obtain

$$q(x_k^{\text{avg}}) + q(z_k^{\text{avg}}) = \eta(x_k^{\text{avg}}, z_k^{\text{avg}}) \leq \frac{2\nu\log(1/\xi) + \rho_3 + 2\rho_2(t_0 + \Gamma_k)}{T_k} + 2\xi(\hat{s}^T\hat{x} + \rho_2).$$

The bound (81) follows by taking $\xi = 1/(k+2)$ and using $T_k \geq (k+1)t_{\min}$.

For Algorithm 2, the same argument applies with $z_k^{\text{avg}} = x_k^{\text{avg}}$, $z_0 = x_0$, and (72) in place of (69). Using (76), we obtain

$$2q(x_k^{\text{avg}}) = \eta(x_k^{\text{avg}}, x_k^{\text{avg}}) \leq \frac{2\nu\log(1/\xi) + \rho_3}{T_k} + 2\xi(\hat{s}^T\hat{x} + \rho_2).$$

The bound (82) follows by taking $\xi = 1/(k+2)$ and using $T_k \geq (k+1)t_{\min}$.

C.4 Proof of Theorem 6

Fix $x, z \in D$, and define $\tilde{x} = (x + x^\star)/2$ and $\tilde{z} = (z + x^\star)/2$. These points are in C. Since $x, x^\star \in \operatorname{dom}\phi$, we have $x/2, x^\star/2 \in \operatorname{dom}\phi$ and $\phi(x^\star/2) = \phi(x^\star) + \nu\log 2$. Inequality (108), together with $\tilde{x} - x^\star/2 = x/2$, therefore gives

$$\phi(\tilde{x}) \leq \phi(x^\star/2) = \phi(x^\star) + \nu\log 2,$$

and similarly $\phi(\tilde{z}) \leq \phi(x^\star) + \nu\log 2$. Hence $\tilde{x}, \tilde{z} \in D$, and the definition (83) of ρ_4 gives

$$d(\tilde{x}, x_0) + d(\tilde{z}, z_0) \leq 2\nu\log 2 + \rho_4.$$

Applying (69) at $(\tilde{x}, \tilde{z})$ gives

$$\mathcal{L}(x_k^{\text{avg}}, \tilde{z}) - \mathcal{L}(\tilde{x}, z_k^{\text{avg}}) \leq \frac{2\nu\log 2 + \rho_4 + 2\rho_2(t_0 + \Gamma_k)}{T_k}. \tag{110}$$

Since $\mathcal{L}$ is affine in each of its arguments on C,

$$\mathcal{L}(x_k^{\text{avg}}, \tilde{z}) = \frac{1}{2}\big(\mathcal{L}(x_k^{\text{avg}}, z) + \mathcal{L}(x_k^{\text{avg}}, x^\star)\big), \qquad \mathcal{L}(\tilde{x}, z_k^{\text{avg}}) = \frac{1}{2}\big(\mathcal{L}(x, z_k^{\text{avg}}) + \mathcal{L}(x^\star, z_k^{\text{avg}})\big).$$

The point $x^\star$ is optimal, so

$$\mathcal{L}(x^\star, u) \leq \sup_{v \in C} \mathcal{L}(x^\star, v) = q(x^\star) = 0$$

for all $u \in C$. In particular, $\mathcal{L}(x^\star, z_k^{\text{avg}}) \leq 0$ and $\mathcal{L}(x_k^{\text{avg}}, x^\star) = -\mathcal{L}(x^\star, x_k^{\text{avg}}) \geq 0$. Therefore,

$$\begin{aligned}\mathcal{L}(x_k^{\text{avg}}, z) - \mathcal{L}(x, z_k^{\text{avg}}) \quad &\leq \quad 2\big(\mathcal{L}(x_k^{\text{avg}}, \tilde{z}) - \mathcal{L}(\tilde{x}, z_k^{\text{avg}})\big) \\ &\leq \quad \frac{2\,(2\nu\log 2 + \rho_4 + 2\rho_2(t_0 + \Gamma_k))}{T_k}.\end{aligned}$$

Taking the supremum over $z \in D$ and the infimum over $x \in D$, and using (76) and $T_k \geq (k+1)t_{\min}$, gives (84).

For Algorithm 2, the same argument applies with $z_k^{\text{avg}} = x_k^{\text{avg}}$, $z_0 = x_0$, and (72) in place of (69), and gives (85).

References

[1] Jacob D. Abernethy, Elad Hazan, and Alexander Rakhlin. Interior-point methods for full-information and bandit online learning. *IEEE Transactions on Information Theory*, 58(7):4164–4175, 2012.

[2] Patrick R. Amestoy, Timothy A. Davis, and Iain S. Duff. An approximate minimum degree ordering algorithm. *SIAM Journal on Matrix Analysis and Applications*, 17(4):886–905, 1996.

[3] Martin S. Andersen, Joachim Dahl, and Lieven Vandenberghe. Implementation of nonsymmetric interior-point methods for linear optimization over sparse matrix cones. *Mathematical Programming Computation*, 2:167–201, 2010.

[4] Martin S. Andersen, Joachim Dahl, and Lieven Vandenberghe. Logarithmic barriers for sparse matrix cones. *Optimization Methods and Software*, 28(3):396–423, 2013.

[5] Martin S. Andersen, Joachim Dahl, and Lieven Vandenberghe. CVXOPT: A Python package for convex optimization, version 1.2.4, 2020.

[6] Martin S. Andersen and Lieven Vandenberghe. CHOMPACK: A Python package for chordal matrix computations, version 2.2.1, 2015.

[7] David Applegate, Mateo Díaz, Oliver Hinder, Haihao Lu, Miles Lubin, Brendan O'Donoghue, and Warren Schudy. Practical large-scale linear programming using primal–dual hybrid gradient. *Advances in Neural Information Processing Systems*, 34:20243–20257, 2021.

[8] David Applegate, Mateo Díaz, Oliver Hinder, Haihao Lu, Miles Lubin, Brendan O'Donoghue, and Warren Schudy. PDLP: A practical first-order method for large-scale linear programming. *Mathematical Programming Computation*, 2026.

[9] David Applegate, Oliver Hinder, Haihao Lu, and Miles Lubin. Faster first-order primal–dual methods for linear programming using restarts and sharpness. *Mathematical Programming*, 201(1):133–184, 2023.

[10] Alfred Auslender and Marc Teboulle. Interior projection-like methods for monotone variational inequalities. *Mathematical Programming*, 104(1):39–68, 2005.

[11] Heinz H. Bauschke, Jérôme Bolte, and Marc Teboulle. A descent lemma beyond Lipschitz gradient continuity: First-order methods revisited and applications. *Mathematics of Operations Research*, 42(2):330–348, 2017.

[12] Heinz H. Bauschke, Jonathan M. Borwein, and Patrick L. Combettes. Bregman monotone optimization algorithms. *SIAM Journal on Control and Optimization*, 42(2):596–636, 2003.

[13] Dimitri P. Bertsekas and John N. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, Belmont, MA, 1997.

[14] James V. Burke. An exact penalization viewpoint of constrained optimization. *SIAM Journal on Control and Optimization*, 29(4):968–998, 1991.

[15] Daniel Cederberg. First-order methods for nonnegative trigonometric matrix polynomials. *Journal of Optimization Theory and Applications*, 204:32, 2025.

[16] Yair Censor and Stavros A. Zenios. *Parallel Optimization: Theory, Algorithms, and Applications*. Numerical Mathematics and Scientific Computation. Oxford University Press, New York, 1997.

[17] Antonin Chambolle and Juan Pablo Contreras. Accelerated Bregman primal–dual methods applied to optimal transport and Wasserstein barycenter problems. *SIAM Journal on Mathematics of Data Science*, 4(4):1369–1395, 2022.

[18] Antonin Chambolle and Thomas Pock. A first-order primal–dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40:120–145, 2011.

[19] Antonin Chambolle and Thomas Pock. On the ergodic convergence rates of a first-order primal–dual algorithm. *Mathematical Programming*, 159:253–287, 2016.

[20] Hsiao-Han Chao and Lieven Vandenberghe. Entropic proximal operators for nonnegative trigonometric polynomials. *IEEE Transactions on Signal Processing*, 66(18):4826–4838, 2018.

[21] Gong Chen and Marc Teboulle. Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM Journal on Optimization*, 3(3):538–543, 1993.

[22] Etienne de Klerk, Cornelis Roos, and Tamás Terlaky. Initialization in semidefinite programming via a self-dual skew-symmetric embedding. *Operations Research Letters*, 20(5):213–221, 1997.

[23] Alberto De Marchi, Yura Malitsky, and Adrien B. Taylor. Mirror descent algorithms with logarithmic barriers. *arXiv e-prints*, arXiv:2608.22834, 2026.

[24] W. S. Dorn. Self-dual quadratic programs. *Journal of the Society for Industrial and Applied Mathematics*, 9(1):51–54, 1961.

[25] Richard J. Duffin. Infinite programs. In Harold W. Kuhn and Albert W. Tucker, editors, *Linear Inequalities and Related Systems*, pages 157–170. Princeton University Press, 1956.

[26] Jonathan Eckstein. Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. *Mathematics of Operations Research*, 18(1):202–226, 1993.

[27] Alan J. Goldman and Albert W. Tucker. Theory of linear programming. In Harold W. Kuhn and Albert W. Tucker, editors, *Linear Inequalities and Related Systems*, pages 55–97. Princeton University Press, 1956.

[28] Tom Goldstein, Min Li, Xiaoming Yuan, Ernie Esser, and Richard Baraniuk. Adaptive primal–dual hybrid gradient methods for saddle-point problems. *arXiv e-prints*, arXiv:1305.0546, 2015.

[29] Osman Güler. Ergodic convergence in proximal point algorithms with Bregman functions. In Ding-Zhu Du and Jie Sun, editors, *Advances in Optimization and Approximation*, pages 155–165. Springer, 1994.

[30] M. A. Hanson. Duality and self-duality in mathematical programming. *Journal of the Society for Industrial and Applied Mathematics*, 12(2):446–449, 1964.

[31] Bingsheng He and Xiaoming Yuan. Convergence analysis of primal–dual algorithms for a saddle-point problem: From contraction perspective. *SIAM Journal on Imaging Sciences*, 5(1):119–149, 2012.

[32] Alfredo N. Iusem and Benar F. Svaiter. A variant of Korpelevich's method for variational inequalities with a new search strategy. *Optimization*, 42:309–321, 1997.

[33] Benjamin Jansen, Cornelis Roos, and Tamás Terlaky. The theory of linear programming: Skew symmetric self-dual problems and the central path. *Optimization*, 29(3):225–233, 1994.

[34] Xin Jiang and Lieven Vandenberghe. Bregman primal–dual first-order method and application to sparse semidefinite programming. *Computational Optimization and Applications*, 81(1):127–159, 2022.

[35] Xin Jiang and Lieven Vandenberghe. Bregman three-operator splitting methods. *Journal of Optimization Theory and Applications*, 196(3):936–972, 2023.

[36] Anatoli Juditsky and Arkadi Nemirovski. First-order methods for nonsmooth convex large-scale optimization, I: General purpose methods. In Suvrit Sra, Sebastian Nowozin, and Stephen J. Wright, editors, *Optimization for Machine Learning*, pages 121–148. MIT Press, 2012.

[37] Anatoli Juditsky and Arkadi Nemirovski. First-order methods for nonsmooth convex large-scale optimization, II: Utilizing problem's structure. In Suvrit Sra, Sebastian Nowozin, and Stephen J. Wright, editors, *Optimization for Machine Learning*, pages 149–183. MIT Press, 2012.

[38] E. N. Khobotov. Modification of the extra-gradient method for solving variational inequalities and certain optimization problems. *USSR Computational Mathematics and Mathematical Physics*, 27(5):120–127, 1987.

[39] Scott P. Kolodziej, Mohsen Aznaveh, Matthew Bullock, Jarrett David, Timothy A. Davis, Matthew Henderson, Yifan Hu, and Read Sandstrom. The SuiteSparse matrix collection website interface. *Journal of Open Source Software*, 4(35):1244, 2019.

[40] Oliver Kolossoski and Renato D. C. Monteiro. An accelerated non-Euclidean hybrid proximal extragradient-type algorithm for convex–concave saddle-point problems. *Optimization Methods and Software*, 32(6):1244–1272, 2017.

[41] G. M. Korpelevich. The extragradient method for finding saddle points and other problems. *Ekonomika i Matematicheskie Metody*, 12(4):747–756, 1976.

[42] Zhenwei Lin, Zikai Xiong, Dongdong Ge, and Yinyu Ye. PDCS: A primal–dual large-scale conic programming solver with GPU enhancements. *arXiv e-prints*, arXiv:2505.00311, 2025.

[43] Haihao Lu, Robert M. Freund, and Yurii Nesterov. Relatively smooth convex optimization by first-order methods, and applications. *SIAM Journal on Optimization*, 28(1):333–354, 2018.

[44] Haihao Lu and Bo Tang. H-PDLP: GPU-accelerated linear programming via generalized PDHG and homogeneous self-dual embedding. *arXiv e-prints*, arXiv:2610.17695, 2026.

[45] Haihao Lu and Jinwen Yang. An overview of GPU-based first-order methods for linear programming and extensions. *arXiv e-prints*, arXiv:2506.02174, 2025.

[46] Zhi-Quan Luo, Jos F. Sturm, and Shuzhong Zhang. Conic convex programming and self-dual embedding. *Optimization Methods and Software*, 14:169–218, 2000.

[47] Yura Malitsky and Thomas Pock. A first-order primal–dual algorithm with linesearch. *SIAM Journal on Optimization*, 28(1):411–432, 2018.

[48] Sanjay Mehrotra. On the implementation of a primal–dual interior point method. *SIAM Journal on Optimization*, 2(4):575–601, 1992.

[49] Arkadi Nemirovski. Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex–concave saddle point problems. *SIAM Journal on Optimization*, 15(1):229–251, 2004.

[50] Yurii Nesterov. Barrier subgradient method. *Mathematical Programming*, 127(1):31–56, 2011.

[51] Brendan O'Donoghue, Eric Chu, Neal Parikh, and Stephen Boyd. Conic optimization via operator splitting and homogeneous self-dual embedding. *Journal of Optimization Theory and Applications*, 169(3):1042–1068, 2016.

[52] Thomas Pock and Antonin Chambolle. Diagonal preconditioning for first order primal–dual algorithms in convex optimization. In *2011 IEEE International Conference on Computer Vision*, pages 1762–1769. IEEE, 2011.

[53] Ralph T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[54] Ralph T. Rockafellar and Roger J.-B. Wets. *Variational Analysis*. Springer, 1998.

[55] Justin Muyot Shao. First-order methods for self-dual linear programs. MS thesis, Electrical and Computer Engineering, University of California, Los Angeles, 2024.

[56] Ron Shefi and Marc Teboulle. Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization. *SIAM Journal on Optimization*, 24(1):269–297, 2014.

[57] Quoc Tran-Dinh and Nghia Nguyen-Trung. Revisiting extragradient-type methods: Part 1—generalizations and sublinear convergence rates. *Computational Optimization and Applications*, 94(1):141–210, 2026.

[58] Paul Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM Journal on Control and Optimization*, 38:431–446, 2000.

[59] Paul Tseng. On accelerated proximal gradient methods for convex–concave optimization. Unpublished manuscript, 2008.

[60] Lieven Vandenberghe and Martin S. Andersen. Chordal graphs and semidefinite optimization. *Foundations and Trends in Optimization*, 1(4):241–433, 2015.

[61] Stephen J. Wright. *Primal–Dual Interior-Point Methods*. SIAM, Philadelphia, 1997.

[62] Zikai Xiong and Robert M. Freund. Level-set geometry and the theoretical performance of PDHG for conic linear optimization. *arXiv e-prints*, arXiv:2609.19097, 2026.

[63] Xiaojie Xu, Pi-Fang Hung, and Yinyu Ye. A simplified homogeneous and self-dual linear programming algorithm and its implementation. *Annals of Operations Research*, 62:151–172, 1996.

[64] Erfan Yazdandoost Hamedani and Necdet S. Aybat. A primal–dual algorithm with line search for general convex–concave saddle point problems. *SIAM Journal on Optimization*, 31(2):1299–1329, 2021.

[65] Yinyu Ye. *Interior Point Algorithms: Theory and Analysis*. Discrete Mathematics and Optimization. Wiley, New York, 1997.

[66] Yinyu Ye, Michael J. Todd, and Shinji Mizuno. An $O(\sqrt{n}L)$-iteration homogeneous and self-dual linear programming algorithm. *Mathematics of Operations Research*, 19(1):53–67, 1994.