Applied Analysis and Optimization
Volume 3, Number 1, 2019, 1–28

Yokohama Publishers
ISSN 2189-1664 Online Journal

INTERIOR–PROXIMAL PRIMAL–DUAL METHODS

TUOMO VALKONEN

Abstract. We study preconditioned proximal point methods for a class of saddle point problems, where the preconditioner decouples the overall proximal point method into an alternating primal–dual method. This is akin to the Chambolle–Pock method or the ADMM. In our work, we replace the squared distance in the dual step by a barrier function on a symmetric cone, while using a standard (Euclidean) proximal step for the primal variable. We show that under non-degeneracy and simple linear constraints, such a hybrid primal–dual algorithm can achieve linear convergence on originally strongly convex problems involving the second-order cone in their saddle point form. On general symmetric cones, we are only able to show an $O(1/N)$ rate. These results are based on estimates of strong convexity of the barrier function, extended with a penalty to the boundary of the symmetric cone. The main contributions of the paper are these theoretical results.

1. Introduction

Interior point methods exhibit fast convergence on several non-smooth non-strongly-convex problems, including linear problems with symmetric cone constraints; see, e.g., [14,28,30,36]. The methods have had less success on large-scale problems with more complex structure. In particular, problems in image processing, inverse problems, and data science, can often be written in the form

$$\min_x G(x) + F(Kx) \tag{P}$$

for convex, proper, lower semicontinuous G and F, and a bounded linear operator K. Often, with G and F involving norms and linear operators, (P) can be converted into linear optimisation on symmetric cones. This is even automated by the disciplined convex programming approach of CVX [17, 18]. Nonetheless, the need to solve a very large scale and difficult Newton system on each step of the interior point method makes this approach seldom practical for real-world problems. Therefore, first-order splitting methods such as forward–backward splitting, ADMM (alternating directions method of multipliers) and their variants [2, 7, 16, 26] dominate these application areas. In our present work, we are curious *whether these two approaches—interior point and splitting methods—can be combined into an effective algorithm?*

The saddle point form of (P) is

$$\min_x \max_y \; G(x) + \langle Kx, y\rangle - F^*(y). \tag{S}$$

A popular algorithm for solving this problem is the primal–dual method of Chambolle and Pock [7]. As discovered in [21], the method can most concisely be written

2010 *Mathematics Subject Classification.* 90C25, 90C51.
Key words and phrases. primal–dual, interior point, proximal point.

as a *preconditioned proximal point method*, solving on each iteration for $u^{i+1} = (x^{i+1}, y^{i+1})$ the variational inclusion

$$0 \in H(u^{i+1}) + M_{i+1}(u^{i+1} - u^i), \tag{PP$_0$}$$

where the monotone operator

$$H(u) := \begin{pmatrix} \partial G(x) + K^* y \\ \partial F^*(y) - Kx \end{pmatrix}$$

encodes the optimality condition $0 \in H(\widehat{u})$ for (S). For the standard proximal point method [34], one would take $M_{i+1} = I$ the identity. With this choice, the system (PP$_0$) is generally difficult to solve. In the Chambolle–Pock method the *preconditioning* or step length operator is given for suitably chosen step length parameters $\tau_i, \sigma_{i+1}, \theta_i > 0$ by

$$M_{i+1} := \begin{pmatrix} \tau_i^{-1} I & -K^* \\ -\theta_i K & \sigma_{i+1}^{-1} I \end{pmatrix}.$$

This choice of M_{i+1} decouples the primal x and dual y updates, making the solution of (PP$_0$) feasible in a wide range of problems. If G is strongly convex, the step length parameters $\tau_i, \sigma_{i+1}, \theta_i$ can be chosen to yield $O(1/N^2)$ convergence rates of an ergodic duality gap and the squared distance $\|x^i - \widehat{x}\|^2$. If both G and F^* are strongly convex, then the method converges linearly. Without any strong convexity, only the ergodic duality gap converges at the rate $O(1/N)$, and the iterates weakly [40].

In our earlier work [39–41], we have modified M_{i+1} as well as the condition (PP$_0$) to still allow a level of mixed-rate acceleration when G is strongly convex only on sub-spaces or sub-blocks of the variable $x = (x_1, \ldots, x_m)$, and derived a corresponding doubly-stochastic block-coordinate descent method. As an extension of that work, our specific question now is:

(1.1) If F^* encodes the constraint $Ay = b$ and $y \in \mathcal{K}$

for a symmetric cone $\mathcal{K}$, *can we replace M_{i+1} in* (PP$_0$) *by a non-linear interior point preconditioner that yields tractable sub-problems and a fast, convergent algorithm?*

Our approach is motivated, firstly, by the fact that (1.1) frequently occurs in applications, in particular with $\mathcal{K}$ the *second-order cone* of elements $y = (y_0, \bar{y}) \in \mathbb{R}^{1+n}$ with $y_0 \geq \|\bar{y}\|$ and $Ay = y_0$. This can be used to model F^* that could otherwise be written as the constraint $F^*(y) = \delta_{B(0,b_0)}(\bar{y})$. Secondly, why we specifically want to try the interior point approach is that the standard and generic quadratic proximal term or preconditioner is not in any specific way adapted to the structure of the ball constraint or the cone $\mathcal{K}$: it is a penalty, but not a barrier approach. The logarithmic barrier, on the other hand, is exactly tuned to the structure of the problem. This suggests that it *might* be able to yield better performance.

Generalised proximal point methods motivated by interior point methods have been considered before in [9,23,27,37,43]. For iterates, which are generally shown to convergence, no convergence rates appear to be known. For function values, rates have been derived in [27, 37]. This is in contrast to the superlinear convergence of a gap functional in conventional interior point methods for linear programming on symmetric cones [14, 28, 30, 36]. The approach in the aforementioned works combining proximal point and interior point methods has essentially been to replace the squared distance in the proximal point method $x^{i+1} := \arg\min_{x\in\mathcal{K}} G(x)+\frac{1}{2\tau}\|x-x^i\|^2$ for $\min_{x\in\mathcal{K}} G(x)$ by a suitable Bregman distance supported on $\operatorname{int}\mathcal{K} \times \operatorname{int}\mathcal{K}$, typically $D(x, x') := \operatorname{tr}(x \circ \ln x - x \circ \ln x' + x' - x)$. In Section 4 of the present work, we will instead replace the squared distance in the proximal point step for the dual variable y by a more conventional barrier-based preconditioner $-\nabla \log\det(y)$. With this, we are able to obtain convergence rates *for the iterates* of the method: in general symmetric cones this is only $O(1/N)$ for the squared distance $\|x^N - \widehat{x}\|^2$ between the primal iterate and the primal solution. I the second-order cone under non-degeneracy and $A = \langle a, \cdot\rangle$ for $a \in \operatorname{int}\mathcal{K}$, this convergence becomes linear. We demonstrate these theoretical results by numerical experiments in Section 5.

The overall idea, how the theory works, is that the barrier-based preconditioner is strongly monotone on bounded subsets of $\operatorname{int}\mathcal{K}$, and "compatible" with ∂F^* on $\partial\mathcal{K}$ in such a way that these strong monotonicity estimates can, with some penalty term, be extended up to the boundary. This introduces some of the strong monotonicity that ∂F^* itself is missing.

Since the performance of the overall algorithm we derive does not improve upon existing methods, our main contributions are these theoretical results on symmetrical cones. An interesting question for future research is, whether the results for general cones can be improved, or whether the second-order cone is special? Nevertheless, our present theoretical results make progress towards closing the gap between direct methods for (P), and primal–dual methods for (S): among others, forward–backward splitting for (P) is known to obtain linear convergence with strong convexity assumptions on G alone [8], but primal–dual methods generally still require the strong convexity of F^* as well. For ADMM additional local estimates exist under quadratic [3, 20] or polyhedrality assumptions [22]. On the other hand, it has been recently established that forward–backward splitting converges at least locally linearly even under less restrictive assumptions than the strong convexity of G [4, 25].

Our convergence results depend on the convergence theory for non-linearly preconditioned proximal point methods from [40]. We quote the relevant aspects in Section 3. To use this theory, we need to compute estimates on the strong convexity of the barrier, with a penalty up to the boundary. This is the content of the latter half of Section 2, after introduction of the basic Jordan-algebraic machinery for interior point methods on symmetric cones.

2. Notation, concepts, and results on symmetric cones

We write $\mathcal{L}(X; Y)$ for space of bounded linear operators between Hilbert spaces X and Y. For any $A \in \mathcal{L}(X; Y)$ we write $\mathcal{N}(A)$ for the null-space, and $\mathcal{R}(A)$ for the range. Also for possibly non-self-adjoint $T \in \mathcal{L}(X; X)$, we introduce the inner product and norm-like notations

$$\langle x, z\rangle_T := \langle Tx, z\rangle, \quad \text{and} \quad \|x\|_T := \sqrt{\langle x, x\rangle_T}, \quad (x, z \in X). \tag{2.1}$$

With $\overline{\mathbb{R}} := [-\infty, \infty]$, we write $\mathcal{C}(X)$ for the space of convex, proper, lower semicontinuous functions from X to $\overline{\mathbb{R}}$. With $K \in \mathcal{L}(X; Y)$, $G \in \mathcal{C}(X)$ and $F^* \in \mathcal{C}(Y)$ on Hilbert spaces X and Y, we then wish to solve the minimax problem (S) assuming the existence of a solution $\widehat{u} = (\widehat{x}, \widehat{y})$ satisfying the optimality conditions

$0 \in H(\widehat{y})$, in other words

$$\text{(OC)} \qquad -K^*\widehat{y} \in \partial G(\widehat{x}), \quad \text{and} \quad K\widehat{x} \in \partial F^*(\widehat{y}).$$

For a function G, as above, ∂G stands the convex subdifferential [33]. For a set C, ∂C is the boundary. We denote by $N_C(x) = \partial\delta_C(x)$ the normal cone to any convex set C at $x \in C$, where δ_C is the indicator function of the set C in the sense of convex analysis.

In Section 4, we concentrate on F^* of the general form (2.2) in the next example.

Example 2.1 (From ball constraints to second-order cones)**.** Very often in (P), we have $F(z) = \sum_{i=1}^n \alpha_i \|z_i\|_2$, where the norm is the Euclidean norm on $\mathbb{R}^m$ and $z = (z_1, \ldots, z_n) \in \mathbb{R}^{mn}$. Then $F^*(\bar{y}) = \delta_{B(0,\alpha_i)}(\bar{y}_i)$ for $\bar{y} = (\bar{y}_1, \ldots, \bar{y}_n) \in \mathbb{R}^{mn}$. We may lift each $\bar{y}_i$ into $\mathbb{R}^{1+m}$ as $y_i = (y_{i,0}, \bar{y}_i)$, and replace F^* by

$$\hat{F}^*(y) := \sum_{i=1}^n \delta_{C_i}(y_i), \quad \text{where} \quad C_i := \{y_i \in \mathcal{K} \mid Ay = b\}, \tag{2.2}$$

where, the linear constraint is defined by $Ay := (y_{1,0}, \ldots, y_{n,0})$ and $b := (\alpha_1, \ldots, \alpha_n)$. The cone constraint is given by $\mathcal{K} = \mathcal{K}_{\text{soc}}^n$ for the *second-order cone*

$$\mathcal{K}_{\text{soc}} := \{y = (y^0, \bar{y}) \in \mathbb{R}^{1+m} \mid y_0 \geq \|\bar{y}\|\}.$$

In the following, we look at the Jordan-algebraic approach to analysis on the second-order cone and other *symmetric cones*.

2.1. **Euclidean Jordan algebras.** We now introduce the minimum amount of the theory of Jordan algebras necessary for our work. For further details, we refer to [13, 24].

Technically, a real *Jordan algebra* $\mathcal{J}$ is a real (additive) vector space together with a bilinear and commutative multiplication operator $\circ : \mathcal{J} \times \mathcal{J} \to \mathcal{J}$ that satisfies the associativity condition $x \circ (x^2 \circ y) = x^2 \circ (x \circ y)$. Here we define $x^2 := x \circ x$. The Jordan algebra $\mathcal{J}$ is *Euclidean* (or *formally real*) if $x^2 + y^2 = 0$ implies $x = y = 0$. We always assume that our Jordan algebras are Euclidean.

We will not directly need the last two technical definitions, but do rely on the very important consequence that $\mathcal{J}$ has a multiplicative unit element e: $x \circ e = x$ for all $x \in \mathcal{J}$. An element x of $\mathcal{J}$ is then called invertible, if there exists an element x^{-1}, such that $x \circ x^{-1} = x^{-1} \circ x = e$.

Example 2.2 (The Jordan algebra of symmetric matrices)**.** To understand these and the following properties, it is helpful to think of the set of symmetric $m \times m$ matrices. They form a Jordan algebra endowed with the product $A \circ B := \frac{1}{2}(AB + BA)$. The inverse is the usual matrix inverse, as is the multiplicative identity. So are the properties discussed next.

An element c in a Jordan algebra $\mathcal{J}$ is an *idempotent* if $c \circ c = c$. It is *primitive*, if it is not the sum of other idempotents. A *Jordan frame* is a set of primitive idempotents $\{e_i\}_{i=1}^r$ such that $e_i \circ e_j = 0$ for $i \neq j$, and $\sum_{j=1}^r e_j = e$. The number r is the *rank* of $\mathcal{J}$. For each $x \in \mathcal{J}$, there indeed exist unique real numbers $\{\lambda_i\}_{i=1}^r$, and a Jordan frame $\{e_i\}_{i=1}^r$, satisfying $x = \sum_{j=1}^r \lambda_i e_i$. The numbers $\lambda_i(x) = \lambda_i$ are called the *eigenvalues* of x. If all the eigenvalues are positive, we write $x > 0$ and call x *positive definite*. Likewise we write $x \geq 0$ if the eigenvalues are non-negative, and call x *positive semi-definite*. With the eigenvalues, we can define

(1) Powers $x^\alpha := \sum_{j=1}^r \lambda_i^\alpha e_i$ when meaningful,
(i) The determinant $\det x := \prod_j \lambda_j$, and
(ii) The trace $\operatorname{tr} x := \sum_{j=1}^r \lambda_j$.
(iii) The inner product $\langle x, y\rangle := \operatorname{tr}(x \circ y)$, and the
(iv) Frobenius norm $\|x\| := \|x\|_F := \sqrt{\sum_{j=1}^r \lambda_j^2} = \sqrt{\langle x, x\rangle}$.

The inner product is positive-definite and associative, satisfying $\langle x \circ y, z\rangle = \langle y, x \circ z\rangle$. We also frequently write

$$\lambda_{\max}(x) := \max_{i=1,\ldots,r} \lambda_i(x) \quad \text{and} \quad \lambda_{\min}(x) := \min_{i=1,\ldots,r} \lambda_i(x).$$

For conciseness, we define for $x \in \mathcal{J}$ the operator $L(x)$ by $L(x)y := x \circ y$. The *quadratic presentation* of x—this is one of the most crucial concepts for us, as we will soon see when covering symmetric cones—is then defined as $Q_x := 2L(x)^2 - L(x^2)$. The invertibility of x is equivalent to the invertibility of Q_x. Other important properties include [13, 36]

(i) $Q_x^\alpha = Q_{x^\alpha}$ for $\alpha \in \mathbb{R}$,
(ii) $Q_{Q_x y} = Q_x Q_y Q_x$ (the *fundamental formula* of quadratic presentations),
(iii) $Q_x x^{-1} = x$,
(iv) $Q_x e = x^2$, and
(v) $\det(Q_x y) = \det(x^2)y = \det(x)^2 y$.

Moreover, Q_x is self-adjoint with respect to the inner product defined above, and the eigenvalues are products $\lambda_i(x)\lambda_j(x)$ [13, 24], so that

$$\lambda_{\min}^2(x)\|y\|^2 \leq \langle Q_x y, y\rangle \leq \lambda_{\max}(x)^2\|y\|^2 \quad \text{for all} \quad y \quad \text{when} \quad x \geq 0. \tag{2.3}$$

Example 2.3 (The Euclidean Jordan algebra of quadratic forms)**.** Let $\mathbb{E}_{1+m}$ denote the space of vectors $x = (x_0, \bar{x}) \in \mathbb{R}^{1+m}$ with x_0 scalar. Setting

$$x \circ y = (x^T y, x_0\bar{y} + y_0\bar{x}),$$

we make $(\mathbb{E}_{1+m}, \circ)$ into a Euclidean Jordan algebra. The identity element is $e = (1, 0)$, rank $r = 2$, and the inner product is

$$\langle x, y\rangle = 2x^T y. \tag{2.4}$$

Defining the diagonal mirroring operator $R := \left(\begin{smallmatrix} 1 & 0 \\ 0 & -I \end{smallmatrix}\right)$, we find that $\det x = x^T Rx = x_0^2 - \|\bar{x}\|^2$, and $x^{-1} = Rx/\det x$ when $\det x \neq 0$.

2.2. **Symmetric cones.** The *cone of squares* of a Euclidean Jordan algebra $\mathcal{J}$ is defined as

$$\mathcal{K} := \{x^2 \mid x \in \mathcal{J}\}.$$

The cones generated this way are precisely the so-called symmetric cones [13] $\mathcal{K}^* = -\mathcal{K}$, or the self-scaled cones of [30]. Their important properties include [13, 24]:

(i) $\operatorname{int} \mathcal{K} = \{x \in \mathcal{J} \mid x \text{ is positive-definite}\} = \{x \in \mathcal{J} \mid L(x) \text{ pos. def.}\}$.
(ii) $\langle x, y\rangle \geq 0$ for all $y \in \mathcal{K}$ iff $x \in \mathcal{K}$, and
(iii) $\langle x, y\rangle > 0$ for all $y \in \mathcal{K} \setminus \{0\}$ iff $x \in \operatorname{int} \mathcal{K}$.
(iv) Q_x for $x \in \operatorname{int} \mathcal{K}$ maps $\mathcal{K}$ onto itself.

(v) For $x, y \in \operatorname{int} \mathcal{K}$, there exists unique $a \in \operatorname{int} \mathcal{K}$, such that $x = Q_a y$.
(vi) For any $x, y \in \mathcal{K}$, $\langle x, y \rangle = 0$ iff $x \circ y = 0$ [15].

For application to interior point methods, and in particular for our work, the following properties are particularly important:

(i) The barrier function $B(x) := -\log(\det x)$ tends to infinity as x goes to $\operatorname{bd} \mathcal{K}$.
(ii) $\nabla B(x) = -x^{-1}$ and $\nabla^2 B(x) = Q_x^{-1}$ (differentiated wrt. the norm in $\mathcal{J}$).
(iii) The normal cone $N_{\mathcal{K}}(x) = -\{y \in \mathcal{K} \mid \langle y, x \rangle = 0\}$ for $x \in \mathcal{K}$ [38, Lemma 3.1].

Example 2.4 (The cone of symmetric positive definite matrices)**.** In the Jordan algebra of symmetric matrices from Example 2.2, the cone of squares is the set of positive semi-definite symmetric matrices.

Example 2.5 (The second order cone)**.** The cone of squares of the Jordan algebra $\mathbb{E}_{1+m}$ of quadratic forms is the second order cone that we have already seen in Example 2.1,

$$\mathcal{K} = \mathcal{K}_{\text{soc}} := \{x \in \mathbb{E}_{1+m} \mid x_0 \geq \|\bar{x}\|\}.$$

If $0 \neq x \in \operatorname{bd} \mathcal{K}$, we have $x^2 = 2x_0 x$. Rescaled, we get a primitive idempotent $c = x/\sqrt{2x_0}$. The only primitive idempotent orthogonal to c is $c' = Rx/\sqrt{2x_0}$. Therefore, the normal cone $N_{\mathcal{K}}(x) = \{-\alpha Rx \mid \alpha \geq 0\}$.

One has to be careful with the fact that the expressions for the barrier gradient and Hessian in (viii) are based on the inner product (2.4) in $\mathbb{E}_{1+m}$. This is scaled by the factor $r = 2$ with respect to the standard inner product on $\mathbb{R}^{1+m}$.

2.3. **Linear optimisation on symmetric cones.** Let $A \in \mathcal{L}(\mathcal{J}; \mathbb{R}^k)$ for an arbitrary Euclidean Jordan algebra $\mathcal{J}$ with the corresponding cone of squares $\mathcal{K}$. We will frequently make use of solutions $(y_\mu, d_\mu, z_\mu) \in \operatorname{int} \mathcal{K} \times \operatorname{int} \mathcal{K} \times \mathbb{R}^k$ to the system

$$\text{(SCLP}_\mu\text{)} \qquad Ay = b, \quad A^* z + c = d, \quad y \circ d = \mu e, \quad y, d \in \operatorname{int} \mathcal{K}.$$

These are meant to approximate solutions $(\widehat{y}, \widehat{d}, \widehat{z}) \in \mathcal{K} \times \mathcal{K} \times \mathbb{R}^k$ to the system

$$\text{(SCLP)} \qquad Ay = b, \quad A^* z + c = d, \quad y \circ d = 0, \quad y, d \in \mathcal{K}.$$

The system (SCLP) arises from primal–dual optimality conditions for linear optimisation on symmetric cones, specifically the problem

$$\min_{y \in \mathcal{K},\, Ay = b} \langle c, y \rangle.$$

The system (SCLP$_\mu$) arises from the introduction of the barrier in the problem

$$\min_{y \in \mathcal{K},\, Ay = b} \langle c, y \rangle - \mu \log \det(y). \tag{2.5}$$

The set of solutions to (SCLP$_\mu$) for varying $\mu > 0$ is called the *central path*. From [14, Theorem 2.2] we know that if there exists a primal–dual *interior feasible point*, i.e., some $(y^*, d^*, z^*) \in \operatorname{int} \mathcal{K} \times \operatorname{int} \mathcal{K} \times \mathbb{R}^k$ such that $Ay^* = b$ and $A^* z^* + c = d^*$, then there exists a solution (y_μ, d_μ, z_μ) to (SCLP$_\mu$) for every $\mu > 0$. In particular, if there exists a solution for some $\mu > 0$, there exist a solution for all $\mu > 0$. In fact, we have the following:

Lemma 2.6. *Suppose the primal feasible set $C := \{y \in \mathcal{K} \mid Ay = b\}$ is bounded, and that there exists a primal interior feasible point $y^* \in \operatorname{int} \mathcal{K} \cap C$. Then there exists a solution $(y_\mu, d_\mu, z_\mu) \in \operatorname{int} \mathcal{K} \times \operatorname{int} \mathcal{K} \times \mathbb{R}^k$ to* (SCLP$_\mu$) *for all $\mu > 0$.*

Proof. The article [14] considers a more general class of linear monotone complementarity problems (LMCPs) than our our SCLPs (symmetric cone linear programs). For the special case of SCLPs, our assumption on the existence of y^* implies that the feasible set in (2.5) non-empty and closed. Since the objective function is level-bounded, proper, and lower semicontinuous, the problem (2.5) has a solution y. This y has to satisfy (SCLP$_\mu$) for some d and z. Now [14, Theorem 2.2] applies. □

Practical methods [30, 36] for solving (SCLP) by closely following the central path are based on scaling the iterates (y^i, d^i) by Q_p for a suitable $p \in \operatorname{int}\mathcal{K}$. We will need this scaling for different purposes, and therefore recall the following basic properties.

Lemma 2.7. *Let $p \in \operatorname{int}\mathcal{K}$, and $y, d \in \mathcal{K}$. Define $\widetilde{y} := Q_p^{1/2} y$, and $\underset{\sim}{d} := Q_p^{-1/2} d$. Then*

(i) *$y \circ d = 0$ if and only if $\widetilde{y} \circ \underset{\sim}{d} = 0$.*
(ii) *If $y, d \in \operatorname{int}\mathcal{K}$ and $\mu > 0$, then $y \circ d = \mu e$ if and only if $\widetilde{y} \circ \underset{\sim}{d} = \mu e$.*
(iii) (SCLP) *(resp.* (SCLP$_\mu$)*) is satisfied for y and d if and only if it is satisfied for $\widetilde{y}$ and $\underset{\sim}{d}$ with A and c replaced by $\widetilde{A} := AQ_p^{-1/2}$ and $\underset{\sim}{c} := Q_p^{-1/2} c$.*

Proof. The claim (i) is a consequence of the properties Section 2.2 (iv) and (vi). The claim (iii) is the content of [36, Lemma 28]. Finally, to establish (iii), the remaining linear equations in (SCLP) and (SCLP$_\mu$) are obvious. □

As a last preparatory step, before starting to derive new results, we say that solutions $y, d \in \mathcal{K}$ to (SCLP) are *strictly complementary* if $y \circ d = 0$ and $y + d \in \operatorname{int}\mathcal{K}$. We say that y is *primal non-degenerate* if

(2.6) $$v = A^* z \text{ and } y \circ v = 0 \implies v = 0.$$

Likewise d is *dual non-degenerate* if

(2.7) $$Av = 0 \text{ and } d \circ v = 0 \implies v = 0.$$

2.4. **Convergence rate of the central path.** We now study convergence rates for the central path, which we will need to develop approximate strong monotonicity estimates. Some existing work can be found in [42], but overall the results in the literature are limited; more work can be found on the properties and mere existence of limits of the central path [5, 10, 19, 29, 31]. After all, in typical interior point methods, one is not interested in solving (SCLP$_\mu$) exactly; rather, one is interested in staying close to the central path while decreasing μ fast. So here we provide the result necessary for our work.

Lemma 2.8. *Let $\widehat{y}, \widehat{d} \in \mathcal{K}$ and $\widehat{z} \in \mathbb{R}^k$ solve* (SCLP). *Also let $y_\mu, d_\mu \in \operatorname{int}\mathcal{K}$ and $z_\mu \in \mathbb{R}^k$ solve* (SCLP$_\mu$) *for some $\mu > 0$. If $\widehat{y}$ and $\widehat{d}$ are strictly complementary, and both primal and dual non-degenerate, then*

(2.8) $$\|y_\mu - \widehat{y}\| \leq \frac{2\mu\sqrt{r}}{\lambda_{\min}(M_{\widehat{y},\widehat{d}})},$$

where $\lambda_{\min}(M_{y,d}) > 0$ *is the minimal eigenvalue of the the linear operator* $M_{y,d} \in \mathcal{L}(\mathcal{J};\mathcal{J})$ *defined at* $y, d \in \mathcal{J}$ *for* $\eta \in \mathcal{N}(A)$ *and* $\xi \in \mathcal{R}(A^*)$ *by*

$$M_{y,d}(\xi + \eta) := L(y)\xi + L(d)\eta.$$

Proof. Observe that (y_μ, d_μ, z_μ) solves (SCLP$_\mu$) if and only if $y_\mu = \widehat{y} + \Delta y$ and $d_\mu = \widehat{d} + \Delta d$ with

$$\Delta y \in \mathcal{N}(A), \quad \Delta d \in \mathcal{R}(A^*), \quad \text{and} \quad M_{\widehat{y},\widehat{d}}(\Delta y + \Delta d) = \mu e - \Delta y \circ \Delta d.$$

Here we have used the fact that $\widehat{y} \circ \widehat{d} = 0$. We may rearrange the final condition as

$$\frac{1}{2} M_{\widehat{y},\widehat{d}}(\Delta y + \Delta d) = \mu e - \frac{1}{2}(\widehat{y} + \Delta y) \circ \Delta d - \frac{1}{2}\Delta y \circ (\widehat{d} + \Delta d).$$

This simply says that

$$\frac{1}{2}\left(M_{\widehat{y},\widehat{d}} + M_{y_\mu,d_\mu}\right)(\Delta y + \Delta d) = \mu e.$$

From [14, Corollary 4.9] we know that the operator $M_{\widehat{y},\widehat{d}}$ is invertible when the solution $(\widehat{y}, \widehat{d})$ is strictly complementary and both primal and dual non-degenerate. Moreover, for (y_μ, d_μ) satisfying (SCLP$_\mu$), we know from [14, Corollary 4.6] that M_{y_μ,d_μ} is invertible. In fact, both M_{y_μ,d_μ} and $M_{\widehat{y},\widehat{d}}$ are positive definite: in both cases, $(y, d) = (\widehat{y}, \widehat{d})$, and $(y, d) = (y_\mu, d_\mu)$, the map $m(\zeta) := \langle \zeta, M_{y,d}\zeta\rangle$ is continuous on $\mathcal{J}$, while $m(\eta) > 0$ and $m(\xi) > 0$ for all $\eta \in \mathcal{N}(A)$ and $\xi \in \mathcal{R}(A^*)$. For $(y, d) = (\widehat{y}, \widehat{d})$ the positivity follows from the assumed primal and dual non-degeneracy, as the operators $L(\widehat{y})$ and $L(\widehat{d})$ are positive semi-definite. For $(y, d) = (y_\mu, d_\mu) \in \operatorname{int}\mathcal{K} \times \operatorname{int}\mathcal{K}$, the operators $L(y_\mu)$ and $L(d_\mu)$ are positive definite; see Section 2.2(i).By an interpolation argument, a contradiction to invertibility would therefore be reached if $M_{y,d}$ were not positive semi-definite on the whole space; cf. [36, proof of Lemma 32].

As a sum of invertible positive definite operators, it now follows that $M_{\widehat{y},\widehat{d}} + M_{y_\mu,d_\mu}$ is invertible. Consequently we estimate

$$\begin{aligned}\|\Delta y\| \leq \|\Delta y + \Delta d\| &= 2\mu\|e\|\|(M_{\widehat{y},\widehat{d}} + M_{y_\mu,d_\mu})^{-1}\| \\ &\leq \frac{2\mu\sqrt{r}}{\lambda_{\min}(M_{\widehat{y},\widehat{d}} + M_{y_\mu,d_\mu})} \leq \frac{2\mu\sqrt{r}}{\lambda_{\min}(M_{\widehat{y},\widehat{d}})},\end{aligned}$$

where the first inequality holds by the orthogonality of Δy and Δd. The claim follows. ☐

2.5. **Strong monotonicity of the barrier.** If the barrier function $B(y) = -\log(\det y)$ is as in Section 2.2, then in the next lemma $d = -\nabla B(y)$. Therefore, the lemma provides an estimate of strong monotonicity of the gradient of the barrier.

Lemma 2.9. *Let* $y, y' \in \operatorname{int}\mathcal{K}$, *and denote* $d := y^{-1}$, *and* $d' := (y')^{-1}$. *Then*

$$-\langle d' - d, y' - y\rangle \geq \frac{1}{\lambda_{\max}(y')\lambda_{\max}(y)}\|y' - y\|^2. \tag{2.9}$$

Proof. There exists a unique $w \in \operatorname{int}\mathcal{K}$ s.t. $d' = Q_w^{-1} y$ and $d = Q_w^{-1} y'$; see, e.g., [30, Corollary 3.1]. We thus see (2.9) to hold if

$$Q_w^{-1} \geq \frac{1}{\lambda_{\max}(y')\lambda_{\max}(y)}. \tag{2.10}$$

In fact, w is given by the Nesterov–Todd direction

$$w = \left(Q_{y^{-1/2}}(Q_{y^{1/2}} d')^{1/2}\right)^{-1}. \tag{2.11}$$

Indeed, using the fundamental formula for quadratic presentations (Section 2.1(vii), we see

$$Q_w^{-1} = Q_{w^{-1}} = Q_{Q_{y^{-1/2}}(Q_{y^{1/2}} d')^{1/2}} = Q_{y^{-1/2}} Q^{1/2}_{Q_{y^{1/2}} d'} Q_{y^{-1/2}}. \tag{2.12}$$

Following [1, p.42], from this we quickly compute

$$Q_w^{-1} y = Q_{y^{-1/2}} Q^{1/2}_{Q_{y^{1/2}} d'} e = Q_{y^{-1/2}} Q_{y^{1/2}} d' = d'.$$

Inverting $d' = Q_w^{-1} y$, we get $(d')^{-1} = y' = (Q_v^{-1} y)^{-1} = Q_v y^{-1} = Q_v d$. Hence $d = Q_v^{-1} y$. This establishes the claimed properties of w.

Continuing from (2.12), we also have

$$Q_w^{-1} = Q_{y^{-1/2}} [Q_{y^{1/2}} Q_{d'} Q_{y^{1/2}}]^{1/2} Q_{y^{-1/2}} \tag{2.13}$$

From Section 2.1(i)and (2.3), we observe that $Q_{d'} = Q_{y'}^{-1} \geq \lambda_{\max}(y')^{-2} I$. Thus

$$\begin{aligned} Q_w^{-1} &\geq \frac{1}{\lambda_{\max}(y')} Q_{y^{-1/2}} [Q_y]^{1/2} Q_{y^{-1/2}} = \frac{1}{\lambda_{\max}(y')} Q_{y^{-1/2}} \\ &\geq \frac{1}{\lambda_{\max}(y')\lambda_{\max}(y)}. \end{aligned} \tag{2.14}$$

This proves (2.10) and consequently (2.9). □

We now extend the estimate to the boundary of $\mathcal{K}$ with a penalty using the approximations form Section 2.4.

Lemma 2.10. *Let $y, d \in \operatorname{int}\mathcal{K}$ and $\widehat{y}, \widehat{d} \in \mathcal{K}$ with $d = y^{-1}$, and $\widehat{y} \circ \widehat{d} = 0$. Suppose there exist $y', d' \in \mathcal{K}$ such that*

$$\langle \widehat{d} - d', y - \widehat{y}\rangle = 0 \quad \textit{and} \quad y' \circ d' = e. \tag{2.15}$$

Then for any $\alpha \in (0,1)$ and any $a \in \operatorname{int}\mathcal{K}$ holds

$$-\langle d - \widehat{d}, y - \widehat{y}\rangle \geq \frac{1-\alpha}{\lambda_{\max}(\widetilde{y})\lambda_{\max}(\widetilde{y}')} \|y - \widehat{y}\|^2_{Q_a} - \frac{\lambda_{\max}(d)\lambda_{\max}(d')}{4\alpha} \|y' - \widehat{y}\|^2, \tag{2.16}$$

where $\widetilde{y} := Q_a^{1/2} y$, and $\widetilde{y}' := Q_a^{1/2} y'$.

Proof. Let Q_w be as in the proof of Lemma 2.9.

$$\begin{aligned} -\langle d - \widehat{d}, y - \widehat{y}\rangle &\overset{(2.15)}{=} -\langle d - d', y - \widehat{y}\rangle = \langle y - y', y - \widehat{y}\rangle_{Q_w^{-1}} \\ &= \langle y - \widehat{y}, y - \widehat{y}\rangle_{Q_w^{-1}} + \langle \widehat{y} - y', y - \widehat{y}\rangle_{Q_w^{-1}} \\ &\geq (1-\alpha)\|y - \widehat{y}\|^2_{Q_w^{-1}} - \tfrac{1}{4\alpha}\|y' - \widehat{y}\|^2_{Q_w^{-1}}. \end{aligned} \tag{2.17}$$

In the final step we have used Cauchy's inequality.

Let $\underset{\sim}{w} := Q_{a^{1/2}}w$. By the fundamental formula of quadratic presentations (Section 2.1(vii),

$$Q_w^{-1} = Q_a^{1/2} Q_{Q_a^{1/2} w}^{-1} Q_a^{1/2} = Q_a^{1/2} Q_{\underset{\sim}{w}}^{-1} Q_a^{1/2}.$$

We also observe using fundamental formula of quadratic presentations that $\underset{\sim}{w}$ is w from (2.11) computed with the transformed variables $\widetilde{y} = Q_a^{1/2} y$ and $\underset{\sim}{d}' = Q_{a^{-1/2}} d'$. We therefore estimate $Q_{\widetilde{w}}^{-1}$ as in (2.14). Since (2.13) implies

$$Q_w^{-1} = Q_{d^{1/2}} [Q_{d^{-1/2}} Q_{d'} Q_{d^{-1/2}}]^{1/2} Q_{d^{1/2}},$$

we also estimate $Q_w^{-1} \le \lambda_{\max}(d')\lambda_{\max}(d)$. Thus (2.16) follows from (2.17). □

Lemma 2.11. *Let $y, d \in \operatorname{int} \mathcal{K}$ and $\widehat{y}, \widehat{d} \in \mathcal{K}$ with $u \circ d = \mu e$ for some $\mu > 0$, and $\widehat{y} \circ \widehat{d} = 0$. Suppose there exist $y', d' \in \mathcal{K}$ such that $\langle \widehat{d} - d', y - \widehat{y} \rangle = 0$ and $y' \circ d' = \mu e$. Then for any $\alpha \in (0, 1)$ holds*

$$\begin{aligned} -\langle d - \widehat{d}, y - \widehat{y} \rangle \ge & \frac{(1-\alpha)\mu}{\lambda_{\max}(\widetilde{y})\lambda_{\max}(\widetilde{y}')} \|y - y'\|_{Q_a}^2 \\ & - \frac{\lambda_{\max}(d)\lambda_{\max}(d')}{4\alpha\mu} \|y' - \widehat{y}\|^2. \end{aligned} \tag{2.18}$$

Proof. We apply Lemma 2.10 with $\widehat{d}$, d, and d' replaced by $\widehat{d}/\mu$, d/μ, and d'/μ. This causes the right-hand-side of the estimate (2.16) to be multiplied by μ, along with both $\lambda_{\max}(d)$ and $\lambda_{\max}(d')$ to be divided by μ. □

Applied to solutions of (SCLP_μ), we can estimate $\lambda_{\max}(y)$ and $\lambda_{\max}(y')$.

Proposition 2.12. *Suppose $Ay = b$ implies $\langle a, y \rangle = b_0$ for some $a \in \operatorname{int} \mathcal{K}$ and $b_0 > 0$. Fix $\mu > 0$, and let $(y, d, z) \in \operatorname{int} \mathcal{K} \times \operatorname{int} \mathcal{K} \times \mathbb{R}^k$ solve (SCLP_μ). Likewise, suppose $(y_\mu, d_\mu, z_\mu) \in \operatorname{int} \mathcal{K} \times \operatorname{int} \mathcal{K} \times \mathbb{R}^k$ solves (SCLP_μ) for $c = \hat{c}$, where $(\widehat{y}, \widehat{d}, \widehat{z})$ solves (SCLP) for $c = \hat{c}$. If $\widehat{y}$ and $\widehat{d}$ are strictly complementary, $\widehat{d}$ dual non-degenerate, and $\widehat{y}$ primal non-degenerate, then for any $\alpha \in (0, 1)$ holds*

$$-\langle d - \widehat{d}, y - \widehat{y} \rangle \ge \frac{(1-\alpha)\mu}{b_0^2} \|y - \widehat{y}\|_{Q_a}^2 - \frac{C_{c,\mu} C_{\hat{c},\mu} r}{\alpha \lambda_{\min}(M_{\widehat{y},\widehat{d}})^2} \mu, \tag{2.19}$$

where for some fixed $y^* \in \operatorname{int} \mathcal{K}$ with $Ay^* = b$ the constants

$$C_{c,\mu} := \frac{\mu r + 2b_0 \|c\|_{Q_a^{-1}}}{\lambda_{\min}(y^*)}. \tag{2.20}$$

Proof. We begin by applying Lemma 2.11 with (y', d') equal to the μ-approximation (y_μ, d_μ) to $(\widehat{y}, \widehat{d})$ provided by Lemma 2.8. Inserting (2.8) into (2.18), we therefore obtain

$$-\langle d - \widehat{d}, y - \widehat{y} \rangle \ge \frac{(1-\alpha)\mu}{\lambda_{\max}(\widetilde{y})\lambda_{\max}(\widetilde{y}_\mu)} \|y - y_\mu\|_{Q_a}^2 - \frac{\mu \lambda_{\max}(d)\lambda_{\max}(d_\mu) r}{\alpha \lambda_{\min}(M_{\widehat{y},\widehat{d}})^2}. \tag{2.21}$$

It remains to estimate the eigenvalues in this expression.

First of all, we easily derive the necessary bounds on $\lambda_{\max}(\widetilde{y})$ and $\lambda_{\max}(y')$ as

$$\lambda_{\max}(\widetilde{y}) \le \operatorname{tr}(\widetilde{y}) = \langle e, \widetilde{y} \rangle = \langle a, y \rangle = b_0. \tag{2.22}$$

Secondly, regarding the estimate on $\lambda_{\max}(d)$, we fix some $y^* \in \operatorname{int}\mathcal{K}$ satisfying $Ay^* = b$. Such a point exist by our assumption of there existing solutions to (SCLP_μ); see also Lemma 2.6. Since $d = A^*z + c$ for some $z \in \mathbb{R}^k$, and $d \circ y = \mu e$, we then derive

$$\begin{aligned}\lambda_{\min}(y^*)\lambda_{\max}(d) &\le \lambda_{\min}(y^*)\langle e, d\rangle \le \langle y^*, d\rangle = \langle \widetilde{y}^*, \underset{\sim}{d}\rangle \\ &= \langle \widetilde{y}, \underset{\sim}{d}\rangle + \langle \widetilde{y}^* - \widetilde{y}, \underset{\sim}{d}\rangle = \mu r + \langle \widetilde{y}^* - \widetilde{y}, \underset{\sim}{c}\rangle \\ &\le \mu r + \|\underset{\sim}{c}\|(\lambda_{\max}(\widetilde{y}) + \lambda_{\max}(\widetilde{y}^*)) \le \mu r + 2b_0\|\underset{\sim}{c}\|.\end{aligned}$$

In the last inequality we have used (2.22) for both $\widetilde{y}$ and $\widetilde{y}^*$. Since $y^* \in \operatorname{int}\mathcal{K}$, so that $\lambda_{\min}(y^*) > 0$, and $\|\underset{\sim}{c}\| = \|c\|_{Q_a^{-1}}$, this gives the claimed bounds on $\lambda_{\max}(d)$ and $\lambda_{\max}(d')$. ☐

Remark 2.13. In Proposition 2.12, the assumption that $Ay = b$ implies $\langle a, y\rangle = b_0$ for some $a \in \operatorname{int}\mathcal{K}$ was only used to derive the bound (2.22) on the maximum eigenvalues of the transformed variable $\widetilde{y} = Q_a^{1/2}y$. If we did not have this assumption, we could still bound the eigenvalues of the untransformed variable y in a local neighbourhood of $\widehat{y}$. Since the factor in front of $\|y - \widehat{y}\|^2_{Q_a}$ in particular would now depend on $\widehat{y}$, doing so would, however, require a more local convergence analysis in Section 4.

2.6. **Strong monotonicity of the barrier in the second-order cone.** In the second-order cone $\mathcal{K} = \mathcal{K}_{\mathrm{soc}} \subset \mathbb{E}_{1+m}$, under suitable constraints $Ay = b$, we have a stronger result.

Lemma 2.14. *Suppose $y, y', d, d' \in \operatorname{int}\mathcal{K}_{\mathrm{soc}}$ with $y \circ d = y' \circ d' = \mu e$ for given $\mu > 0$. Then*

$$-\langle d - d', y - y'\rangle_{\mathcal{J}} \ge \frac{\det(d) + \det(d')}{\mu}\|y - y'\|^2_{-R}, \tag{2.23}$$

where $\|y - y'\|^2_{-R} := \|\overline{y} - \overline{y}'\|^2_{\mathbb{R}^m} - (y_0 - y_0')^2 = -\det(y - y')$.

Proof. We have $d = \mu Ry/\det(y) = \mu^{-1}\det(d)Ry$ and $d' = \mu^{-1}\det(d')Ry'$. We write for brevity $\beta := \mu^{-1}\det(d)$ and $\beta' := \mu^{-1}\det(d')$. Then

$$-\langle d - d', y - y'\rangle_{\mathcal{J}} = -\langle \beta Ry - \beta' Ry', y - y'\rangle_{\mathcal{J}} = 2\langle \beta y - \beta' y', y - y'\rangle_{-R},$$

where the second "inner product" is $\langle x, y\rangle_{-R} := -\langle Rx, y\rangle_{\mathbb{R}^{1+m}}$. We can thus write

$$-\langle d - d', y - y'\rangle_{\mathcal{J}} = 2\beta\|y - y'\|^2_{-R} + 2(\beta - \beta')\langle y', y - y'\rangle_{-R}$$

as well as

$$-\langle d - d', y - y'\rangle_{\mathcal{J}} = 2\beta'\|y - y'\|^2_{-R} + 2(\beta - \beta')\langle y, y - y'\rangle_{-R}.$$

Summing these two expressions we deduce

$$-\langle d - d', y - y'\rangle_{\mathcal{J}} = (\beta + \beta')\|y - y'\|^2_{-R} + (\beta - \beta')(\|y\|^2_{-R} - \|y'\|^2_{-R}). \tag{2.24}$$

Now observe that

$$\|y\|^2_{-R} = y_0^2 - \|\overline{y}\|^2 = -\det(y) = -\mu^2/\det(d).$$

Thus

$$
\begin{aligned}
(\beta-\beta')(\|y\|_{-R}^2-\|y'\|_{-R}^2) &= \mu(\det(d)-\det(d'))(\det(d')^{-1}-\det(d)^{-1}) \\
&= \mu(\det(d')-\det(d))^2/(\det(d)\det(d')) > 0.
\end{aligned}
$$

This and (2.24) immediately prove the claim. □

For solutions of (SCLP_μ) with one-dimensional linear constraints, we can extend the estimate to the boundary with some penalty. For this, we first bound the determinant with the distance

$$
D_F(w,d) := \|Q_w^{1/2}d-\mu_{w,d}e\| \quad \text{for} \quad \mu_{w,d}=\langle w,d\rangle/r, \quad (w,d\in\mathcal{K}).
$$

This distance is typically used to define the so-called short-step neighbourhood of the central path; see, e.g., [36].

Lemma 2.15. *Suppose $y,d\in\operatorname{int}\mathcal{K}_{\mathrm{soc}}$ with $y\circ d=\mu e$ and $\langle a,y\rangle=b_0$ for some $\mu,b_0>0$ and $a\in\operatorname{int}\mathcal{K}_{\mathrm{soc}}$. Then*

$$
\frac{2\mu^2+\sqrt{2}b_0D_F(a^{-1},d)\mu}{b_0^2\det(a)} \le \det(d) \le \frac{4\mu^2+\sqrt{2}b_0D_F(a^{-1},d)\mu}{b_0^2\det(a)}. \tag{2.25}
$$

Proof. We define $\widetilde{y}:=Q_a^{1/2}y$, and $\underset{\sim}{d}:=Q_a^{-1/2}d$. Then $\langle e,\widetilde{y}\rangle=\langle a,y\rangle=b_0$, and by [36, Lemma 28], $\widetilde{y}\circ\underset{\sim}{d}=\mu e$. These conditions expand to $\widetilde{y}_0\underset{\sim}{d}_0+\bar{\widetilde{y}}^T\bar{\underset{\sim}{d}}=\mu$, $\widetilde{y}_0\bar{\underset{\sim}{d}}+\underset{\sim}{d}_0\bar{\widetilde{y}}=0$, and $2\widetilde{y}_0=b_0$. (In the latter, recall that the $\mathbb{E}_{1+m}$-inner product satisfies $\langle e,\widetilde{y}\rangle=2e^T\widetilde{y}$.) We reduce this system to $\underset{\sim}{d}_0^2-\|\bar{\underset{\sim}{d}}\|^2-2\underset{\sim}{d}_0\mu/b_0=0$, from where we solve

$$
\underset{\sim}{d}_0=\frac{\mu+\sqrt{\mu^2+b_0^2\|\bar{\underset{\sim}{d}}\|^2}}{b_0}. \tag{2.26}
$$

Thus

$$
\det(\underset{\sim}{d})=\underset{\sim}{d}_0^2-\|\bar{\underset{\sim}{d}}\|^2=\frac{2\mu^2+2\mu\sqrt{\mu^2+b_0^2\|\bar{\underset{\sim}{d}}\|^2}}{b_0^2},
$$

from which we easily estimate

$$
\frac{2\mu^2+2\mu b_0\|\bar{\underset{\sim}{d}}\|}{b_0^2} \le \det(\underset{\sim}{d}) \le \frac{4\mu^2+2\mu b_0\|\bar{\underset{\sim}{d}}\|}{b_0^2}. \tag{2.27}
$$

To finish deriving (2.25), we recall from Section 2.1(v)

$$
\det(\underset{\sim}{d})=\det(a)\det(d).
$$

We also have $r\underset{\sim}{d}_0=\langle\underset{\sim}{d},e\rangle=\langle d,a^{-1}\rangle$ for the rank $r=2$, so

$$
\sqrt{2}\|\bar{\underset{\sim}{d}}\|_{\mathbb{R}^n}=\|\underset{\sim}{d}-\underset{\sim}{d}_0e\|_{\mathcal{J}}=\|Q_a^{-1/2}d-\mu_{a^{-1},d}e\|_{\mathcal{J}}=D_F(a^{-1},d), \tag{2.28}
$$

where we emphasise the standard Euclidean norm on $\bar{\underset{\sim}{d}}\in\mathbb{R}^n$ versus the $\sqrt{2}$-scaled standard norm on $\mathcal{J}$. With this, (2.27) gives (2.25). □

If $D_F(a^{-1},\widehat{d})>0$, or alternatively $\det(\widehat{y})>0$, then the next proposition shows local strong monotonicity of the barrier for d close to $\widehat{d}$ and $\mu>0$ small. Moreover, if $D_F(a^{-1},\widehat{d})>0$, the factor of strong monotonicity does not vanish as $\mu\searrow 0$.

Proposition 2.16. *Let $\mathcal{K} = \mathcal{K}_{\mathrm{soc}}$, and suppose $Ay = b$ implies $\langle a, y\rangle = b_0$ for some $a \in \operatorname{int} \mathcal{K}$ and $b_0 > 0$. Let $(y, d, z) \in \operatorname{int} \mathcal{K} \times \operatorname{int} \mathcal{K} \times \mathbb{R}^k$ solve* (SCLP$_\mu$)*, and likewise that $(\widehat{y}, \widehat{d}, \widehat{z}) \in \mathcal{K} \times \mathcal{K} \times \mathbb{R}^k$ solve* (SCLP) *for $c = \widehat{c}$. Then*

$$(2.29)\quad -\langle d - \widehat{d}, y - \widehat{y}\rangle \geq \frac{\mu + 2^{-1/2} b_0 [D_F(a^{-1}, d) + D_F(a^{-1}, \widehat{d})]}{b_0^2/2} \|y - \widehat{y}\|^2_{Q_a} - \mu + \frac{\mu + 2^{-1/2} b_0 D_F(a^{-1}, d)}{b_0^2/2} \det(Q_a^{1/2} \widehat{y}).$$

Proof. We have

$$(2.30)\quad 0 = \widehat{y} \circ \widehat{d} = (\widehat{y}_0 \widehat{d}_0 + \widehat{\bar{y}}^T \widehat{\bar{d}}, \widehat{y}_0 \widehat{\bar{d}} + \widehat{d}_0 \widehat{\bar{y}}).$$

Since $\langle a, \widehat{y}\rangle = b_0 > 0$, and $\widehat{y} \in \mathcal{K}$, necessarily $\widehat{y}_0 > 0$. Since, moreover, $\widehat{y} \neq 0$, we cannot have $\widehat{d} \in \operatorname{int} \mathcal{K}$ for $\widehat{y} \circ \widehat{d} = 0$ to hold. Therefore $0 = \det(\widehat{d}) = \widehat{d}_0^2 - \|\widehat{\bar{d}}\|^2$. It follows from (2.30) that $\widehat{d} = \widehat{\beta} R \widehat{y}$ for

$$(2.31)\quad \widehat{\beta} = -\frac{\widehat{\bar{y}}^T \widehat{\bar{d}}}{\widehat{y}_0^2} = \frac{\widehat{d}_0}{\widehat{y}_0} = \frac{\|\widehat{\bar{d}}\|_{\mathbb{R}^m}}{\widehat{y}_0} \geq 0.$$

We may therefore repeat the steps of Lemma 2.14 until (2.24) to obtain

$$(2.32)\quad -\langle d - \widehat{d}, y - \widehat{y}\rangle = (\beta + \widehat{\beta}) \|y - \widehat{y}\|^2_{-R} + (\beta - \widehat{\beta})(\|y\|^2_{-R} - \|\widehat{y}\|^2_{-R}).$$

We have $\det(\widehat{y}) = -\|\widehat{y}\|^2_{-R} = \widehat{y}_0^2 - \|\widehat{\bar{y}}\|^2 \geq 0$. If this is non-zero, $\widehat{y} \in \operatorname{int} \mathcal{K}$. But in that case $\widehat{y} \circ \widehat{d} = 0$ implies $\widehat{d} = 0$, and consequently $\widehat{\beta} = 0$. Thus $\widehat{\beta} \|\widehat{y}\|^2_{-R} = 0$ whether or not $\|\widehat{y}\|^2_{-R} = 0$. Using $\|y\|^2_{-R} = -\det(y) = -\mu^2/\det(d)$ and $\beta = \det(d)/\mu$, we therefore obtain from (2.32) that

$$(2.33)\quad -\langle d - \widehat{d}, y - \widehat{y}\rangle = (\mu^{-1} \det(d) + \widehat{\beta}) \|y - \widehat{y}\|^2_{-R} - \mu + \frac{\widehat{\beta} \mu^2}{\det(d)} + \frac{\det(d) \det(\widehat{y})}{\mu}.$$

If $a = e$, we have $y_0 = \widehat{y}_0 = b_0/2$, so that $2\|y - \widehat{y}\|^2_{-R} = \|y - \widehat{y}\|^2_{\mathcal{J}}$. Reasoning as in (2.28), (2.31) gives $\widehat{\beta} = \sqrt{2} D_F(a^{-1}, \widehat{d})/b_0 = \sqrt{2} D_F(e, \widehat{d})/b_0$. With the help of Lemma 2.15, (2.33) thus yields

$$(2.34)\quad -\langle d - \widehat{d}, y - \widehat{y}\rangle \geq \frac{2\mu + \sqrt{2} b_0 [D_F(e, d) + D_F(e, \widehat{d})]}{b_0^2} \|y - \widehat{y}\|^2 - \mu + \frac{2\mu + \sqrt{2} b_0 D_F(e, d)}{b_0^2} \det(\widehat{y}),$$

where we have entirely eliminated the term $\widehat{\beta} \mu^2/\det(d) \geq 0$. Since $\lambda_{\min}(e) = \det(e) = 1$, the estimate (2.29) is immediate in the case $a = e$.

If $a \neq e$, we define $\widetilde{y} := Q_a^{1/2} y$, and $\underset{\sim}{d} := Q_a^{-1/2} d$ as in Lemma 2.15. Then $(\widetilde{y}, \underset{\sim}{d}, z)$ continues to satisfy (SCLP$_\mu$) with A replaced by $\widetilde{A} := A Q_a^{-1/2}$ and $\underset{\sim}{c} := Q_a^{-1/2} c$. The same holds with (SCLP) for $\widetilde{\widehat{y}} := Q_a^{1/2} \widehat{y}$ and $\underset{\sim}{\widehat{d}} := Q_a^{-1/2} \widehat{d}$. Therefore, (2.34) holds for these transformed variables. Since $D_F(e, \underset{\sim}{d}) = D_F(a^{-1}, d)$, as well as $\|\widetilde{y} - \widetilde{\widehat{y}}\|^2 = \|y - \widehat{y}\|^2_{Q_a}$, and $-\langle d - d', y - y'\rangle = -\langle \underset{\sim}{d} - \underset{\sim}{\widehat{d}}, \widetilde{y} - \widetilde{\widehat{y}}\rangle$, we obtain the claim. □

Corollary 2.17. *Let $\mathcal{K} = \mathcal{K}_{\mathrm{soc}}$, and suppose $A = \langle a, \cdot\rangle$ for some $a \in \operatorname{int}\mathcal{K}$. Suppose moreover that $\langle a^{-1}, c\rangle = \langle a^{-1}, \widehat{c}\rangle = 0$. Let $(y, d, z) \in \operatorname{int}\mathcal{K} \times \operatorname{int}\mathcal{K} \times \mathbb{R}^k$ solve* (SCLP$_\mu$)*, and likewise that $(\widehat{y}, \widehat{d}, \widehat{z}) \in \mathcal{K} \times \mathcal{K} \times \mathbb{R}^k$ solve* (SCLP) *for $c = \widehat{c}$. If $\widehat{c} \neq 0$, then*

$$-\langle d - \widehat{d}, y - \widehat{y}\rangle \geq \frac{\mu + 2^{-1/2} b_0 [\|c\|_{Q_a^{-1}} + \|\widehat{c}\|_{Q_a^{-1}}]}{b_0^2/2} \|y - \widehat{y}\|_{Q_a}^2 - \mu. \tag{2.35}$$

Otherwise, if $\widehat{c} = 0$ with $\widehat{y} = ba^{-1}/2$, then

$$-\langle d - \widehat{d}, y - \widehat{y}\rangle \geq \frac{\mu + 2^{-1/2} b_0 \|c\|_{Q_a^{-1}}}{b_0^2/2} \|y - \widehat{y}\|_{Q_a}^2. \tag{2.36}$$

We say that (2.35) is strong monotonicity of the barrier "with a penalty", μ.

Proof. We do not until the very end of the proof use the assumption $A = \langle a, \cdot\rangle$. For now, we use the weaker assumption that $Ay = b$ implies $\langle a, y\rangle = b_0$. We apply Proposition 2.16. This gives

$$-\langle d - \widehat{d}, y - \widehat{y}\rangle \geq \frac{\mu + 2^{-1/2} b_0 [D_F(a^{-1}, d) + D_F(a^{-1}, \widehat{d})]}{b_0^2/2} \|y - \widehat{y}\|_{Q_a}^2 - \mu + \frac{\mu + 2^{-1/2} b_0 D_F(a^{-1}, d)}{b_0^2/2} \det(Q_a^{1/2} \widehat{y}). \tag{2.37}$$

If $D_F(a^{-1}, \widehat{d}) = 0$, by assumption $\widehat{y} = 2b_0 a^{-1}$. This implies $\det(Q_a^{1/2}\widehat{y}) = b_0/2$. Consequently

$$\frac{\mu + 2^{-1/2} b_0 D_F(a^{-1}, d)}{b_0^2/2} \det(Q_a^{1/2} \widehat{y}) \geq \mu.$$

Therefore no penalty is imposed, and (2.37) reduces to

$$-\langle d - \widehat{d}, y - \widehat{y}\rangle \geq \frac{\mu + 2^{-1/2} b_0 D_F(a^{-1}, d)}{b_0^2/2} \|y - \widehat{y}\|_{Q_a}^2. \tag{2.38}$$

Suppose then that $D_F(a^{-1}, \widehat{d}) > 0$. On the right hand side of (2.37), only the term $-\mu$ is negative. Thus the condition holds if

$$-\langle d - \widehat{d}, y - \widehat{y}\rangle \geq \frac{\mu + 2^{-1/2} b_0 [D_F(a^{-1}, d) + D_F(a^{-1}, \widehat{d})]}{b_0^2/2} \|y - \widehat{y}\|_{Q_a}^2 - \mu. \tag{2.39}$$

Finally, using our assumptions that $A = \langle a, \cdot\rangle$ and $\langle a^{-1}, c\rangle = 0$, we have $d = za + c$ and $\mu_{a^{-1},d} = \langle a^{-1}, d\rangle / r = z$ for some $z \in \mathbb{R}$. Thus

$$D_F(a^{-1}, d) = \|Q_a^{-1/2}(d - za)\| = \|c\|_{Q_a^{-1}}. \tag{2.40}$$

Likewise $D_F(a^{-1}, \widehat{d}) = \|\widehat{c}\|_{Q_a^{-1}}$. Therefore, the cases $D_F(a^{-1}, \widehat{d}) > 0$ and $D_F(a^{-1}, \widehat{d}) = 0$ are equivalent to the cases on $\|\widehat{c}\|$ in the statement of the corollary. Inserting (2.40) into (2.38) consequently yields the claimed estimates. □

Remark 2.18. Recall Remark 2.13 on removing the assumption on the existence of $a \in \operatorname{int}\mathcal{K}$ such that $\langle a, y\rangle = b_0$. In the proof of Proposition 2.16, this assumption was not used until the derivation of (2.34) from (2.33). At that point, we used this fact to ensure that $\langle a, y - \widehat{y}\rangle = 0$ and, in particular, that $\|y - \widehat{y}\|^2_{-R} = \|y - \widehat{y}\|^2 \geq 0$ when $a = e$ after transformation. Could we still get our overall estimates without this assumption?

On the two-dimensional Jordan algebra $\mathbb{E}_{1+1}$, pick $a = (a_0, \bar{a}) \notin \mathcal{K}$, $b \in \mathbb{R}$, and set $Ay := a_0 y_0 + \bar{a}\bar{y}$. Without loss of generality, by negating both a and b. Assume that $\bar{a} < 0$. Then y satisfying $Ay = b$ has the form $y = \theta v + (b/a_0)e$ for $v = (-\bar{a}, a_0)$ and some $\theta \in \mathbb{R}$. Since $a \notin \mathcal{K}$ and $\bar{a} < 0$, we have $-\bar{a} = |\bar{a}| > a_0$. Consequently, $v \in \operatorname{int} \mathcal{K}$.

Now, with $y = \theta v + (b/a_0)e$ and $\widehat{y} = \widehat{\theta} v + (b/a_0)e$ with $\theta \neq \theta^*$, we have

$$\|y - \widehat{y}\|^2_{-R} = \|(\theta - \theta^*)v\|^2_{-R} = -(\theta - \theta^*)^2 \det(v) < 0.$$

This implies that the first term in (2.33) is negative for all the feasible points in every neighbourhood of $\widehat{y}$. This seems at first a negative result. If, however $\det(\widehat{y}) > 0$, then also $\det(\widehat{d}) > 0$, so in a neighbourhood of $(\widehat{y}, \widehat{d})$, the last term of (2.33) will be bounded away from zero. We can therefore still, *locally*, obtain quadratic estimates like those in Corollary 2.17.

On the other hand, if $\det(\widehat{y}) = 0$, we can run into difficulties. Consider $b = 0$, so that $y = \theta v$ and $\widehat{y} = 0$. Then also $\widehat{\beta} = 0$, so the right-hand-side of (2.33) is negative, and we do not get the quadratic-penalised estimate. The solution $\widehat{y} = 0$ would, however, be primal degenerate. Indeed, in the general non-degenerate strictly complementary case, Proposition 2.12 and Remark 2.13 still guarantee a local estimate with worse constants than the more fine-grained approach of Proposition 2.12 might provide.

3. An abstract preconditioned proximal point iteration

In this section, we recall some of the core results from [40]. We start by setting

$$H(u) := \begin{pmatrix} \partial G(x) + K^* y \\ \partial F^*(y) - Kx \end{pmatrix}, \tag{3.1}$$

and for some $\tau_i, \phi_i, \sigma_{i+1}, \psi_{i+1} > 0$, defining the step length and "testing" operators

$$W_{i+1} := \begin{pmatrix} \tau_i I & 0 \\ 0 & \sigma_{i+1} I \end{pmatrix}, \quad \text{and} \quad Z_{i+1} := \begin{pmatrix} \phi_i I & 0 \\ 0 & \psi_{i+1} I \end{pmatrix}. \tag{3.2}$$

We also let $V_{i+1} : X \times Y \rightrightarrows X \times Y$ for each $i \in \mathbb{N}$ be an abstract non-linear preconditioner, dependent on the current iterate u^i. Then we consider the generalised proximal point method, which involves solving

$$0 \in W_{i+1} H(u^{i+1}) + V_{i+1}(u^{i+1}) \tag{PP}$$

for the unknown next iterate u^{i+1}. To obtain convergence rates for the resulting method, the idea from [40, 41] will be to analyse the inclusion obtained after multiplying (PP) by the testing operator Z_{i+1}.

Assuming G to be (strongly) convex with factor $\gamma > 0$, we also introduce

$$\Xi_{i+1}(\gamma) := \begin{pmatrix} 2\gamma\tau_i I & 2\tau_i K^* \\ -2\sigma_{i+1} K & 0 \end{pmatrix},$$

which is an operator measure of strong monotonicity of H.

The next lemma, which is relatively trivial to prove [40], forms the basis from which our work proceeds.

Theorem 3.1. *Let us be given $K \in \mathcal{L}(X;Y)$, $G \in \mathcal{C}(X)$, and $F^* \in \mathcal{C}(Y)$ on Hilbert spaces X and Y. For each $i \in \mathbb{N}$, for some $V'_{i+1} : X \times Y \rightrightarrows X \times Y$ and $M_{i+1} \in \mathcal{L}(X \times Y; X \times Y)$, take*

$$V_{i+1}(u) := V'_{i+1}(u) + M_{i+1}(u - u^i). \tag{3.3}$$

Assume that (PP) *is solvable, $Z_{i+1}M_{i+1}$ is self-adjoint, and G is (strongly) convex with factor $\gamma \geq 0$. If for all $i \in \mathbb{N}$ the estimate*

$$\begin{aligned}
\text{(C0-}\Gamma\text{)}\quad & \underbrace{\frac{1}{2}\|u^{i+1} - u^i\|^2_{Z_{i+1}M_{i+1}}}_{\text{step length in local metric}} + \underbrace{\frac{1}{2}\|u^{i+1} - \widehat{u}\|^2_{Z_{i+1}(\Xi_{i+1}(\gamma)+M_{i+1})-Z_{i+2}M_{i+2}}}_{\text{linear preconditioner update discrepancy}} \\
&+ \underbrace{\langle \partial F^*(y^{i+1}) - \partial F^*(\widehat{y}), y^{i+1} - \widehat{y}\rangle_{\Psi_{i+1}\Sigma_{i+1}}}_{\text{variably useful remainder from } H} \\
&+ \underbrace{\langle Z_{i+1}V'_{i+1}(u^{i+1}), u^{i+1} - \widehat{u}\rangle}_{\text{from non-linear preconditioner}} \geq -\Delta_{i+1}
\end{aligned}$$

holds, then

$$\frac{1}{2}\|u^N - \widehat{u}\|^2_{Z_{N+1}M_{N+1}} \leq \frac{1}{2}\|u^0 - \widehat{u}\|^2_{Z_1M_1} + \sum_{i=0}^{N-1} \Delta_{i+1}, \quad (N \geq 1). \tag{3.4}$$

Proof. This is [40, Theorem 3.1] specialised to scalar step length and testing operators $T_i = \tau_i I$, $\Phi_i = \phi_i I$, $\Sigma_{i+1} = \sigma_{i+1} I$, and $\Psi_{i+1} = \psi_{i+1} I$, as well as $\widetilde{\Gamma} = \gamma I$. □

It is possible to extend this theorem to provide an estimate on an ergodic duality gap; see [40, Theorem 4.6]. For the sake of conciseness, we have however decided against including such estimates in the present work. For this reason, in the following, we concentrate on strongly convex G.

4. A primal–dual method with a barrier preconditioner

Let $F(y) := \delta_{\{A\,\cdot\,=b\}}(y) + \delta_{\mathcal{K}}(y)$ for some $A \in \mathcal{L}(\mathcal{J}; Z)$, where $\mathcal{J}$ and Z are Hilbert spaces, $\mathcal{J}$ also a Euclidean Jordan algebra. Let $\mathcal{K}$ be the cone of squares of $\mathcal{J}$. We suppose there exists some $y \in \operatorname{int}\mathcal{K}$ with $Ay = b$. Then the subdifferential sum formula (see, e.g., [33]) applies, so that

$$\partial F^*(y) = \begin{cases} \{A^*z \mid z \in Z\} + N_{\mathcal{K}}(y), & Ay = b \text{ and } y \in \mathcal{K}, \\ \emptyset, & \text{otherwise.} \end{cases} \tag{4.1}$$

In particular, if $y \in \operatorname{int}\mathcal{K}$ with $Ay = b$, then $\partial F^*(y) = \{A^*z \mid z \in Z\}$. Note from Section 2.2(iii) and (vi) that

$$N_{\mathcal{K}}(y) = \{-d \mid d \in \mathcal{K},\ p \circ d = 0\} \quad (y \in \mathcal{K}). \tag{4.2}$$

Inserting (4.1) into $0 \in H(\widehat{u})$, the latter expands as

$$0 \in \partial G(\widehat{x}) + K^*\widehat{y}, 0 \in A^*\widehat{z} + N_{\mathcal{K}}(\widehat{y}) - K\widehat{x}, Ay = b, y \in \mathcal{K}$$

for some $\widehat{z} \in Z$. Based on (4.2), this may also be written as the existence of $(\widehat{x}, \widehat{y}, \widehat{d}, \widehat{z}) \in X \times \mathcal{K} \times \mathcal{K} \times Z$ with

$$-K^*\widehat{y} \in \partial G(\widehat{x}), \quad A\widehat{y} = b, \quad A^*\widehat{z} - K\widehat{x} = \widehat{d}, \quad \widehat{y} \circ \widehat{d} = 0. \tag{IOC}$$

In the following, we develop an algorithm for solving this system, incorporating a barrier-based nonlinear preconditioner for dual updates. As mentioned after Theorem 3.1, for conciseness we limit our attention to strongly convex G, and only analyse the convergence of iterates, not the gap. The theory from [40] could be used to extend the analysis to the gap. Moreover, following the approach of [39], it would be possible to extend our work to stochastic and "spatially-adaptive" updates.

4.1. **A general estimate for dual barrier preconditioning.** To construct algorithms with the help of the theory from Section 3, we have to construct the preconditioner $V_{i+1}(u^{i+1}) := V'_{i+1}(u^{i+1}) + M_{i+1}(u^{i+1} - u^i)$. We specifically take

$$M_{i+1} = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}, \quad \text{and} \quad V'_{i+1}(u^{i+1}) = (0, \sigma_{i+1}[K(x^{i+1} - x^i) - d^{i+1}]), \tag{4.3}$$

where $d^{i+1} \in \operatorname{int} \mathcal{K}$ is defined to satisfy $y^{i+1} \circ d^{i+1} = \mu_{i+1} e$ for some $\mu_{i+1} > 0$. The term $\sigma_{i+1} K(x^{i+1} - x^i)$ in V'_{i+1} decouples the primal and dual updates so that(PP) may be written as the system

$$0 \in \tau_i \partial G(x^{i+1}) + \tau_i K^* y^{i+1} + (x^{i+1} - x^i), \tag{4.4a}$$

$$0 \in \sigma_{i+1}[A^* z^{i+1} - K x^i - d^{i+1}], \quad \text{as well as} \tag{4.4b}$$

$$y^{i+1} \circ d^{i+1} = \mu_{i+1} e \quad \text{and} \quad A y^{i+1} = b \quad \text{with} \quad y^{i+1}, d^{i+1} \in \operatorname{int} \mathcal{K}. \tag{4.4c}$$

For this general setup, we have the following lemma:

Lemma 4.1. *Let F^* have the structure* (4.1)*. Take M_{i+1} and V'_{i+1} according to* (4.3)*. Suppose for some $\omega_{i+1}, \delta_{i+1} \in \mathbb{R}$ for all $i \in \mathbb{N}$ that*

$$-\langle d^{i+1} - \widehat{d}, y^{i+1} - \widehat{y} \rangle \geq \omega_{i+1} \|y^{i+1} - \widehat{y}\|^2 - \delta_{i+1}, \tag{4.5a}$$

$$\psi_{i+1} \sigma_{i+1} = \phi_i \tau_i, \tag{4.5b}$$

$$2\omega_{i+1} \geq \tau_i \|K\|^2, \quad \text{and} \tag{4.5c}$$

$$\phi_{i+1} \leq \phi_i (1 + 2\tau_i \widetilde{\gamma}). \tag{4.5d}$$

Then (C0-Γ) *holds with $\Delta_{i+1} = \psi_{i+1} \sigma_{i+1} \delta_{i+1}$, and $Z_{i+1} M_{i+1}$ is self-adjoint with*

$$Z_{i+1} M_{i+1} = \begin{pmatrix} \phi_i I & 0 \\ 0 & 0 \end{pmatrix} \geq 0. \tag{4.6}$$

Proof. The condition (C0-Γ) now reads

$$\begin{aligned}(4.7)\quad &\underbrace{\frac{1}{2}\|u^{i+1}-u^i\|^2_{Z_{i+1}M_{i+1}}}_{\text{step in local norm}}+\underbrace{\frac{1}{2}\|u^{i+1}-\widehat{u}\|^2_{D_{i+2}}}_{\text{lin. precond. upd. d.}}\\ &\qquad+\underbrace{\psi_{i+1}\sigma_{i+1}\langle K(x^{i+1}-x^i),y^{i+1}-\widehat{y}\rangle}_{\text{de-coupling term from }V'}\\ &+\underbrace{\psi_{i+1}\sigma_{i+1}\langle A^*(z^{i+1}-\widehat{z}),y^{i+1}-\widehat{y}\rangle-\psi_{i+1}\sigma_{i+1}\langle d^{i+1}-\widehat{d},y^{i+1}-\widehat{y}\rangle}_{F^*\text{ term from (C0-}\Gamma\text{) as well as }d^{i+1}\text{ from }V'}\geq-\Delta_{i+1}\end{aligned}$$

with the linear preconditioner update discrepancy

$$D_{i+2}:=Z_{i+1}(\Xi_{i+1}(\gamma)+M_{i+1})-Z_{i+2}M_{i+2}.$$

The expansion and estimate (4.6) are trivially verified along with the self-adjointness of $Z_{i+1}M_{i+1}$. This expansion allows us to write

$$D_{i+2}=\begin{pmatrix}\phi_i(1+2\tau_i\gamma)I-\phi_{i+1}I & 2\phi_i\tau_iK^*\\ -2\psi_{i+1}\sigma_{i+1}K & 0\end{pmatrix}.$$

We use (4.5b) to cancel the off-diagonals of D_{i+2} in (4.7). Then we use the fact that $A(y^{i+1}-\widehat{y})=0$ to cancel the first term on the second line of (4.7). Finally, we use $\Delta_{i+1}=\psi_{i+1}\sigma_{i+1}\delta_{i+1}$ and (4.5a) to estimate the second term on the second line of (4.7). This gives the condition

$$\begin{aligned}(4.8)\quad &\frac{\phi_i}{2}\|x^{i+1}-x^i\|^2+\frac{\psi_{i+1}\sigma_{i+1}\omega_{i+1}}{2}\|y^{i+1}-\widehat{y}\|^2\\ &\qquad+\frac{\phi_i(1+2\gamma\tau_i)-\phi_{i+1}}{2}\|x^{i+1}-\widehat{x}\|^2\\ &\qquad\qquad+\psi_{i+1}\sigma_{i+1}\langle K(x^{i+1}-x^i),y^{i+1}-\widehat{y}\rangle\geq 0.\end{aligned}$$

Application of (4.5d), as well as Cauchy's inequality to the inner product term, shows that (4.8) and consequently (C0-Γ) is satisfied if

$$\psi_{i+1}\sigma_{i+1}\omega_{i+1}\geq\frac{1}{2}\phi_i^{-1}\psi_{i+1}^2\sigma_{i+1}^2KK^*.$$

This follows from (4.5b) and (4.5c). □

We define τ_i through (4.5c) for a lower bound $\omega_{*,i+1}$ of ω_{i+1}. Likewise, we take (4.5d) as an equality as the definition of ϕ_{i+1}. We observe that σ_{i+1} and ψ_{i+1} are irrelevant to the algorithm in (4.4), as will be the specific choice of $\phi_0>0$ to the satisfaction of (4.5). Taking $\phi_0=1$, we obtain Algorithm 1 from (4.4).

Algorithm 1 (Barrier-preconditioned primal–dual method)**.**

Require: Linear operator $K \in \mathcal{L}(X; \mathcal{J})$, strongly convex $G \in \mathcal{C}(X)$, and $F^* \in \mathcal{C}(\mathcal{J})$ of the form (46). Factor $\gamma > 0$ of the strong convexity of G. Rules for $\mu_i, \omega_{*,i} > 0$.

1: Choose initial iterates $x^0 \in X$ and $y^0 \in Y$.
2: Set initial testing parameter $\phi_0 := 1$.
3: **repeat**
4: Calculate μ_i, $\omega_{*,i}$, and step length

$$\tau_i := 2\omega_{*,i+1}/\|K\|^2.$$

5: Update testing parameter

$$\phi_{i+1} := \phi_i(1 + 2\gamma\tau_i).$$

6: Dual update: for $(y^{i+1}, d^{i+1}, z^{i+1}) \in \operatorname{int} \mathcal{K} \times \operatorname{int} \mathcal{K} \times Z$ solve the system

$$Ay^{i+1} = b, \quad A^* z^{i+1} - Kx^i = d^{i+1}, \quad \text{and} \quad y^{i+1} \circ d^{i+1} = \mu_{i+1} e.$$

7: Primal update

$$x^{i+1} := (I + \tau_i \partial G)^{-1}(x^i - \tau_i K^* y^{i+1}).$$

8: **until** a stopping criterion is satisfied.

Remark 4.2 (Solution of Line 6 of Algorithm 1)**.** The system on Line 6 is a standard (SCLP$_\mu$). In the second-order cone with $A = \langle e, \,\cdot\, \rangle$ and $\langle e, \mathcal{R}(K) \rangle = \{0\}$, it is easy to solve. Indeed, $(0, \vec{d}^{i+1}) = -Kx^i$ while d_0^{i+1} is given by the expression in (2.26). Finally

$$y^{i+1} = \mu_{i+1}(d^{i+1})^{-1} = \frac{\mu_{i+1} R d^{i+1}}{\det(d^{i+1})} = \frac{\mu_{i+1} R d^{i+1}}{(d_0^{i+1})^2 - \|\vec{d}^{i+1}\|^2}.$$

More general cases $A = \langle a, \,\cdot\, \rangle$ and $\langle a^{-1}, \mathcal{R}(K) \rangle = \{0\}$ follow by scaling.

We leave the solution of more general problems than the easy one considered in Remark 4.2 for future research. In particular, we would expect to combine the overall algorithm with a path-following interior point method in order to not have to solve the sub-problem exactly in each step, but to merely take a single step of the path-following method towards its solution. Such an approach may yield a primal–dual version of the work in [37].

4.2. **Convergence rates in general symmetric cones.** We still need to specify μ_{i+1}, verify (4.5a), and produce convergence rates. In general symmetric cones, we have:

Theorem 4.3. *With $\mathcal{K}$ an arbitrary symmetric cone, and $Z = \mathbb{R}^k$, let the requirements of Algorithm 1 be satisfied. Assuming that $Ay = b$ implies $\langle a, y \rangle = b_0$ for some $a \in \operatorname{int} \mathcal{K}$ and $b_0 > 0$, suppose there exists a solution $(\widehat{x}, \widehat{y}, \widehat{d}, \widehat{z}) \in X \times \mathcal{K} \times \mathcal{K} \times Z$ to* (IOC) *with $\widehat{y}$ and $\widehat{d}$ strictly complementary, $\widehat{d}$ dual non-degenerate, and $\widehat{y}$ primal non-degenerate. Suppose further that* $\operatorname{dom} G$ *is bounded, or that the primal iterates $\{x^i\}_{i \in \mathbb{N}}$ of Algorithm 1 stay bounded through other means. For some constant $\theta > 0$ and $\zeta \in (0, b_0^{-2})$, take*

$$\mu_{i+1} := \theta \phi_i^{-1/2}, \quad \text{and} \quad \omega_{*,i+1} := \zeta \lambda_{\min}(a) \mu_{i+1}. \tag{4.9}$$

Then $\|x^N - \widehat{x}\|^2 = O(1/N)$.

Remark 4.4. The assumption $Z = \mathbb{R}^k$ is merely for the simplicity of application of Proposition 2.12 and later Corollary 2.17. There would be nothing stopping us from applying the results on uncountable products of symmetric cones, for example.

Proof. We use Proposition 2.12, which verifies (4.5a) with

$$\delta_{i+1} \le \hat{C} C_{-Kx^i,\mu_{i+1}} C_{-K\widehat{x},\mu_{i+1}} \mu_{i+1} \quad \text{and} \quad \omega_{i+1} = \omega_{*,i+1} = \zeta\lambda_{\min}(a)\mu_{i+1}$$

for $C_{-Kx^i,\mu_{i+1}}$, $C_{-K\widehat{x},\mu_{i+1}}$ defined in (2.20), and some $\hat{C} > 0$. From (2.20) we see that the former constants are bounded as long as $\{\mu_i\}_{i\in\mathbb{N}}$ is non-increasing, and the sequence $\{\|Kx^i\|\}_{i\in\mathbb{N}}$ bounded. The latter is guaranteed by our assumptions, and the former by our construction of μ_{i+1} in (4.9) and Line 5 of the algorithm. Therefore $\delta_{i+1} \le C\mu_{i+1}$ for some constant $C > 0$. From (4.5b) and (4.9) it now follows

$$\Delta_{i+1} := \psi_{i+1}\sigma_{i+1}\delta_{i+1} \le C\tau_i\phi_i\mu_{i+1} = C\theta\tau_i\phi_i^{1/2}. \tag{4.10}$$

Next we use Theorem 3.1 and Lemma 4.1. For $C_0 := \frac{1}{2}\|u^0 - \widehat{u}\|^2_{Z_1M_1}$, (3.4), (4.6), and (4.10) give the combined estimate

$$\frac{\phi_N}{2}\|x^N - \widehat{x}\|^2 \le C_0 + C\theta\sum_{i=0}^{N-1}\tau_i\phi_i^{1/2}, \quad (N \ge 1). \tag{4.11}$$

Inserting $\omega_{*,i+1}$ and μ_{i+1} from (4.9), Line 4 and Line 5 of the algorithm say

$$\phi_{i+1} = \phi_i + \gamma\nu\phi_i^{1/2} \quad \text{and} \quad \tau_i = \phi_i^{-1/2}\nu/\|K\|^2 \quad \text{for} \quad \nu := 2\zeta\lambda_{\min}(a)\theta.$$

It follows (see [41]) that $\phi_N = \Theta(N^2)$, while $\sum_{i=0}^{N-1}\tau_i\phi_i^{1/2} = N\nu/\|K\|^2$. Inserting these estimates into (4.11), we verify the $O(1/N)$ rate. □

4.3. **Convergence rates in the second-order cone.** In the second-order cone, we obtain linear convergence under dual non-degeneracy, $K\widehat{x} = 0$. In image processing example such as those we consider in Section 5, we would have $Kx = (0, \nabla x)$, lifting a discretised gradient to the second-order cone (or a pointwise product cone). Therefore $K\widehat{x} = 0$ means that the solution image cannot be flat.

Theorem 4.5. *For* $\mathcal{K} = \mathcal{K}_{\mathrm{soc}}$ *the second-order cone,* $Z = \mathbb{R}^k$*, and* $A = \langle a, \cdot\rangle$ *for some* $a \in \operatorname{int}\mathcal{K}$ *with* $\langle a^{-1}, \mathcal{R}(K)\rangle = \{0\}$*, let the requirements of Algorithm 1 be satisfied. Suppose there exists a solution* $(\widehat{x}, \widehat{y}, \widehat{d}, \widehat{z}) \in X \times \mathcal{K} \times \mathcal{K} \times Z$ *to* (IOC)*. If* $K\widehat{x} = 0$*, take* $\widehat{y} = ba^{-1}/2$ *and* $\widehat{d} = 0$*. For some* $\theta > 0$ *and* $\zeta \in (0, 2b_0^{-2}]$*, take*

$$\mu_{i+1} := \theta\phi_i^{-1/2} \quad \text{and} \quad \omega_{*,i+1} := (\mu_{i+1}\zeta + 2^{-1/2}b_0^{-1}\|Kx^i\|_{Q_a^{-1}})\lambda_{\min}(a). \tag{4.12}$$

Suppose further that $\operatorname{dom} G$ *is bounded, or that the primal iterates* $\{x^i\}_{i\in\mathbb{N}}$ *of Algorithm 1 stay bounded through other means. Then for some* $C, \varepsilon > 0$ *holds*

$$\|x^N - \widehat{x}\|^2 \le \begin{cases} C(1+\varepsilon)^{-N}, & K\widehat{x} \ne 0, \\ C/N^2, & K\widehat{x} = 0. \end{cases}$$

Proof. From Line 5 of the algorithm and (4.12), we expand

$$\tau_i := 2(\zeta\theta\phi_i^{-1/2} + \widetilde{\ell}_{i+1})\lambda_{\min}(a)/\|K\|^2 \quad \text{for} \quad \widetilde{\ell}_{i+1} := 2^{-1/2}b_0^{-1}\|Kx^i\|_{Q_a^{-1}}. \tag{4.13}$$

From (4.13) and Line 5, we estimate

$$\phi_N \geq \phi_0 + 2\gamma\zeta\theta \sum_{i=0}^{N-1} \phi_i^{1/2}. \tag{4.14}$$

It follows from (4.13) that $\sup_i \tau_i \leq C_\tau$ for some constant $C_\tau > 0$. From (4.12), we also obtain $\mu_{i+1} \searrow 0$.

We then use Corollary 2.17, which verifies (4.5a) with $\omega_{i+1} := (\mu_{i+1}\zeta + \ell_{i+1})\lambda_{\min}(a)$ and

$$\begin{cases} \ell_{i+1} := \frac{\|Kx^i\|_{Q_a^{-1}}}{b_0/\sqrt{2}}, \text{ and } \delta_{i+1} := 0, & \text{if } K\widehat{x} = 0, \\ \ell_{i+1} = \frac{\|K\widehat{x}\|_{Q_a^{-1}} + \|Kx^i\|_{Q_a^{-1}}}{b_0/\sqrt{2}}, \text{ and } \delta_{i+1} = \mu_{i+1}, & \text{if } K\widehat{x} \neq 0. \end{cases}$$

Setting $\ell := \sqrt{2}\|K\widehat{x}\|_{Q_a^{-1}}/b_0 > 0$, we have $\ell_{i+1} = \widetilde{\ell}_{i+1} + \ell$.

Next we use Theorem 3.1 and Lemma 4.1. Recalling (4.5b) and that $\Delta_{i+1} = \psi_{i+1}\sigma_{i+1}\delta_{i+1}$ in Lemma 4.1, setting $C_0 := \frac{1}{2}\|u^0 - \widehat{u}\|^2_{Z_1M_1}$, (3.4) and (4.6) yield

$$\frac{\phi_N}{2}\|x^N - \widehat{x}\|^2 \leq C_0 + D_N \quad \text{for} \quad D_N := \sum_{i=0}^{N-1} \tau_i\phi_i\delta_{i+1} \quad (N \geq 1). \tag{4.15}$$

In the case $K\widehat{x} = 0$, we have $\delta_{i+1} = 0$. As in the proof of Theorem 4.3, by a standard analysis [39, 41], it follows from (4.14) that $\phi_N \geq CN^2$ for some $C > 0$. We therefore get from (4.15) the claimed $O(1/N^2)$ rate.

Consider then the case $K\widehat{x} \neq 0$. We estimate

$$D_N = \sum_{i=0}^{N-1} \tau_i\phi_i\mu_{i+1} \leq C_\tau \sum_{i=0}^{N-1} \phi_i\mu_{i+1} \tag{4.16}$$

By Line 4 and Line 5 of the algorithm, $\phi_N \geq \phi_0 + 2\gamma\zeta\|K\|^{-2}\sum_{i=0}^{N-1} \phi_i\mu_{i+1}$. Using these estimates in (4.15), it follows that $\|x^N - \widehat{x}\|$ is bounded. If $\widetilde{\ell}_{i+1} \searrow 0$, (4.13) and (4.14) shows that also $\tau_i \searrow 0$. Restarting our analysis from a later iteration, we can therefore make $C_\tau > 0$ arbitrarily small. Consequently, for any $\epsilon > 0$, for large enough N holds $\|x^N - \widehat{x}\| \leq \epsilon$. Since $\ell > 0$, this is in contradiction to $\widetilde{\ell}_{i+1} \searrow 0$. We may therefore assume that $\widetilde{\ell}_{i+1} \geq \widetilde{\epsilon}$ for some $\widetilde{\epsilon} > 0$, at least for large i. Since our claims are asymptotical, we may without loss of generality assume this for all i.

From (4.13), we now estimate $\tau_i \geq \widetilde{\epsilon}\lambda_{\min}(a)/\|K\|^2 =: \tau_* > 0$. From Line 5 consequently

$$\phi_{i+1} \geq \phi_i(1 + 2\gamma\tau_*). \tag{4.17}$$

This shows that $\phi_N \geq \Theta((1+\gamma\tau_*)^N)$ grows exponentially, predicting (4.15) to yield linear rates if we can control the penalty D_N.

Continuing form (4.16), by Hlder's inequality, since the conjugate exponent of $1/(1-p)$ is $1/p$, for any $p \in (0,1)$ holds

$$D_N \le C_\tau \theta \sum_{i=0}^{N-1} \phi_i^{1-p} \phi_i^{p-1/2} \le C_\tau \theta \left(\sum_{i=0}^{N-1} \phi_i \right)^{1-p} \left(\sum_{i=0}^{N-1} \phi_i^{1-1/(2p)} \right)^p .$$

By (4.17), the second sum on the right is bounded if $1 - 1/(2p) < 0$, that is $p \in (0, 1/2)$. From Line 5 of the algorithm

$$\phi_N - \phi_0 = 2\gamma \sum_{i=0}^{N-1} \phi_i \tau_i \ge 2\gamma \tau_* \sum_{i=0}^{N-1} \phi_i .$$

For some constant $C' > 0$ we therefore get

$$D_N \le C'(\phi_N - \phi_0)^{1-p} \le C' \phi_N^{1-p}.$$

Minding (4.15) and (4.17), this shows the claimed linear rate. □

5. Numerical demonstrations

We study the performance of the proposed algorithm on two image processing problems, total variation (TV) denoising, and H^1 denoising. These can be written as

$$\min_{x \in \mathbb{R}^{n_1 n_2}} \frac{1}{2} \|z - x\|_2^2 + \alpha R(x), \tag{5.1}$$

where $n_1 \times n_2$ is the image size in pixels, and z the noisy image as a vector in $\mathbb{R}^{n_1 n_2}$. The parameter $\alpha > 0$ is a regularisation parameter, and R a regularisation term. For TV regularisation, it is $R(x) = \|Dx\|_{2,1}$, and for H^1 regularisation, it is $R(x) = \|Dx\|_2$. Here $D \in \mathbb{R}^{2n_1 n_2 \times n_1 n_2}$ is a matrix for a discretisation of the gradient, and $\|g\|_{2,1} := \sum_{i=1}^{n_1 n_2} \sqrt{g_{i,1} + g_{i,2}}$ for $g = (g_{\cdot,1}, g_{\cdot,2}) \in \mathbb{R}^{2n_1 n_2}$. We specifically take D as forward-differences with Neumann boundary conditions.

The problem (5.1) can in both cases be written in the saddle point form

$$\min_{x \in \mathbb{R}^{n_1 n_2}} \max_{y \in \mathcal{J}} \frac{1}{2} \|z - x\|_2^2 + \langle Kx, y \rangle - \delta_{\mathcal{K} \cap A^{-1} b}(y),$$

where for H^1 denoising

$$\mathcal{J} = \mathbb{E}_{1+2n_1 n_2}, \qquad Kx = (0, Dx),$$
$$Ay = y_0, \qquad b = \alpha,$$

and for TV denoising

$$\mathcal{J} = (\mathbb{E}_{1+2})^{n_1 n_2}, \qquad [Kx]_i = (0, [Dx]_{i,1}, [Dx]_{i,2}) \qquad (i = 1, \ldots, n_1 n_2),$$
$$Ay = ((y_1)_0, \ldots (y_{n_1 n_2})_0), \qquad b = (\alpha, \ldots, \alpha).$$

In the latter case, Line 6 of Algorithm 1 splits into $n_1 n_2$ parallel problems of the form covered by Remark 4.2. The remark therefore shows how to efficiently solve the step for both example problems.

While TV denoising [35] is a fundamental benchmark in mathematical image processing, we have to emphasise here that H^1 denoising is not an approach of practical importance. It blurs images unlike TV denoising. Nevertheless, it forms a non-trivial optimisation problem, as we do not square the norm of the gradient. (The optimality conditions in that case would be linear: in the continuous setting the heat equation.)

5.1. **Remarks on convergence rates.** The linear convergence results for the second-order cone in Section 4.3 apply to H^1 denoising, but they do not apply to TV denoising. In the latter case, $\mathcal{K} = \mathcal{K}_{\text{soc}}^{n_1 n_2}$ is a product of second-order cones, but not a second-order cone. It would be possible to extend the analysis of Section 4.3 to product cones. Due to the coupling through (4.5b), a straightforward approach would yield linear convergence when $\min_i \|[K\widehat{x}]_i\| > 0$. From the structure of the TV denoising problem, it is however easy to see that it can often happen that $[K\widehat{x}]_i = 0$. This is the case when the solution image is locally flat. This happens in total variation denoising more often than one might expect, due to the characteristic staircasing effect of the approach [32]. Therefore, there is little hope to obtain linear convergence on practical TV denoising problems using this approach.

5.2. **Numerical setup.** We performed some numerical experiments on the parrot image (#23) from the free Kodak image suite photo.[1] We used the image, converted to greyscale, both at the original resolution of $n_1 \times n_2 = 768 \times 512$, and scaled down to $n_1 \times n_2 = 192 \times 128$ pixels. Together with the dual variable, the problem dimensions are therefore $768 \cdot 512 \cdot 3 = 1179648 \simeq 10^6$ and $128 \cdot 128 \cdot 3 = 49152 \approx 4 \cdot 10^4$. To the high-resolution test image, we added Gaussian noise with standard deviation 29.6 (12dB). In the downscaled image, this becomes 6.15 (25.7dB). With the low-resolution image, we used regularisation parameter $\alpha = 0.01$ for TV denoising, and $\alpha = 5$ for H^1 denoising. We scale these up to $\alpha/0.25$ for the high-resolution image [11].

We compared our algorithm (denoted PEDI, *Primal Euclidean–Dual Interior*) to the accelerated Chambolle–Pock method (PDHGM, *Primal–Dual Hybrid Gradient method, Modified* [12]) on the saddle-point problem, as well as forward–backward splitting on the dual problem (Dual FB). For Dual FB we took as the basic step size $\tau = 1/L^2$, where $L := \sqrt{8} \geq \|K\|$ [6]. For the PDHGM, we took $\tau_0 \approx 0.52/L$ and $\sigma_0 = 1.9/L$, using the strong convexity parameter $\gamma = 0.9 < 1$ for acceleration. For our method, we took $\zeta = 0.9/b_0^2$ and $\theta = 1/\zeta$, keeping τ_0 and γ unchanged from the PDHGM. For the initial iterates we always took $x^0 = 0$ and $y^0 = 0$. The hardware we used was a MacBook Pro with 16GB RAM and a 2.8 GHz Intel Core i5 CPU. The codes were written in MATLAB+C-MEX.

For our reporting, we computed a target optimal solution $\widehat{x}$ by taking one million iterations of the basic PDHGM. In Figure 1 and Table 1 for TV denoising, and Figure 2 and Table 2 for H^1 denoising, we report the following: the distance to $\widehat{x}$ in decibels $10\log_{10}(\|x^i - \widehat{x}\|^2/\|\widehat{x}\|^2)$, the primal objective value $\text{val}(x) := G(x) + F(Kx)$ relative to the target $10\log_{10}((\text{val}(x) - \text{val}(\hat{x}))^2/\text{val}(\hat{x})^2)$, as well as the duality gap $10\log_{10}(\text{gap}^2/\text{gap}_0^2)$, again in decibels relative to the initial iterate. For forward–backward splitting, to compute the duality gap, we solve the primal variable x^i from the primal optimality condition $K^* y^i = \nabla G(x^i) = x^i - z$.

[1] At the time of writing online at `http://r0k.us/graphics/kodak/`.

TABLE 1. TV denoising performance: CPU time and number of iterations (at a resolution of 10) to reach given duality gap, distance to target, or primal objective value.

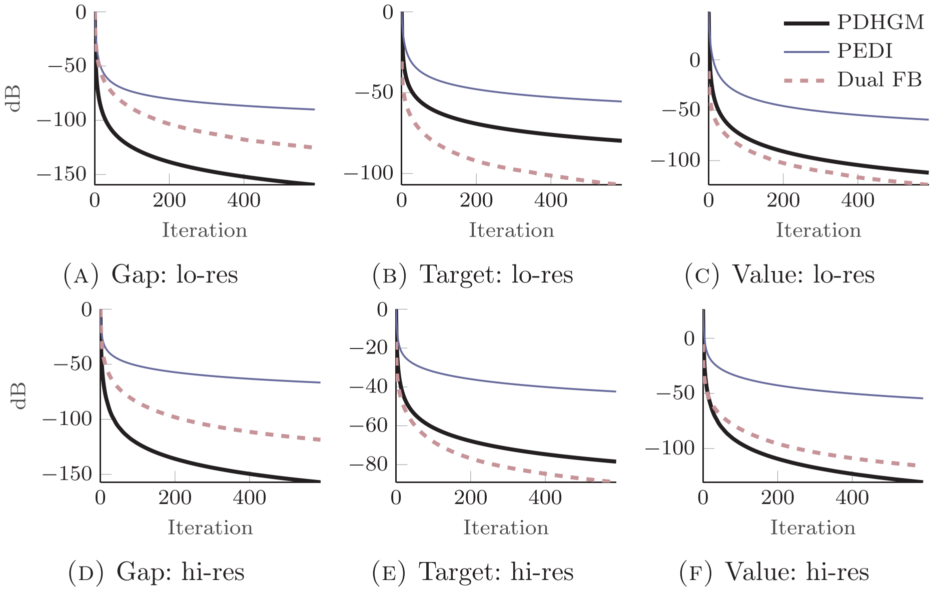
	low resolution						high resolution					
	gap ≤ -50dB		tgt ≤ -50dB		val ≤ -50dB		gap ≤ -50dB		tgt ≤ -50dB		val ≤ -50dB	
Method	iter	time	iter	time	iter	time	iter	time	iter	time	iter	time
PDHGM	4	0.01s	30	0.09s	27	0.08s	4	0.13s	34	1.42s	13	0.52s
PEDI	16	0.04s	270	0.73s	280	0.75s	86	3.78s	–	–	400	17.76s
Dual FB	12	0.03s	6	0.02s	9	0.02s	14	0.62s	21	0.96s	12	0.53s

TABLE 2. H^1 denoising performance: CPU time and number of iterations (at a resolution of 10) to reach given duality gap, distance to target, or primal objective value.

	low resolution						high resolution					
	gap ≤ -150dB		tgt ≤ -100dB		val ≤ -100dB		gap ≤ -150dB		tgt ≤ -100dB		val ≤ -100dB	
Method	iter	time	iter	time	iter	time	iter	time	iter	time	iter	time
PDHGM	360	0.91s	–	–	180	0.46s	380	11.48s	–	–	120	3.60s
PEDI	120	0.31s	87	0.22s	54	0.14s	51	1.69s	39	1.28s	24	0.78s
Dual FB	44	0.11s	43	0.11s	22	0.05s	17	0.74s	18	0.78s	8	0.32s

5.3. **Performance analysis and concluding remarks.** As expected, the performance of PEDI on TV denoising is not particularly good, reflecting the $O(1/N)$ rates from Theorem 4.3. For H^1 denoising we observe significantly improved convergence, reflecting the linear rates from Theorem 4.5, and of dual forward–backward splitting. While PEDI eventually has better gap behaviour than dual forward–backward splitting, overall, however, the method appears no match for the latter in our sample problems. The results for the high resolution and low resolution problem are comparable. Since the low-resolution problem has size of order 10^4, and the high resolution problem has size of the relatively large order 10^6, this suggests good scalability of the algorithm. Further research is required to see whether there are problems for which the overall Primal Euclidean(Proximal)–Dual Interior or similar approaches provide competitive algorithms.

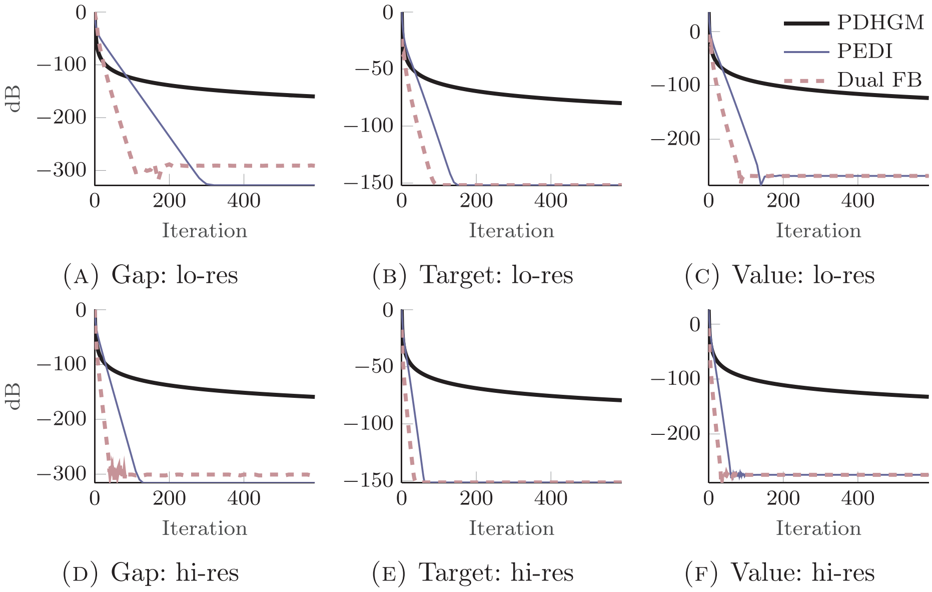
Irrespective of the limited practicality of PEDI, our theoretical analysis helps to bridge the gap in performance between direct primal or dual methods, and primal–dual methods. After all, we have obtained linear rates without the strong convexity of both G and F^* in the saddle point problem (S). As a next step to take from here, it will be interesting to see if convergence rates can be derived in our overall setup for the "distance-like" preconditioners from [9, 23, 27, 43]. Moreover, we are puzzled by what, if anything, makes the second-order cone special? Finally, numerically we have only considered problems of the form given in Remark 4.2, where the interior point sub-problem can be solved exactly. This is sufficient for most image processing and similar applications. However, it would be interesting to know whether we can combine a path-following interior point algorithm for its solution into the overall proximal point method. Such an approach may yield a primal–dual version of the work in [37].



(A) Gap: lo-res (B) Target: lo-res (C) Value: lo-res

(D) Gap: hi-res (E) Target: hi-res (F) Value: hi-res

FIGURE 1. TV denoising convergence behaviour: high and low resolution images; gap, distance to target solution, and primal objective value in decibels.



(A) Gap: lo-res (B) Target: lo-res (C) Value: lo-res

(D) Gap: hi-res (E) Target: hi-res (F) Value: hi-res

FIGURE 2. H^1 denoising convergence behaviour: high and low resolution images; gap, distance to target solution, and primal objective value in decibels.

Acknowledgements

The final stages of this research have been performed with the support of the EPSRC First Grant EP/P021298/1, "PARTIAL Analysis of Relations in Tasks of Inversion for Algorithmic Leverage".

A data statement for the EPSRC

The source code and data used to produce the numerical results of this publication have been deposited at doi: 10.5281/zenodo.1402031.

References

[1] F. Alizadeh and D. Goldfarb, *Second-order cone programming*, Math. Program. **95** (2003), 3–51

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

[3] D. Boley, *Local linear convergence of the alternating direction method of multipliers on quadratic or linear programs*, SIAM J. Optim. **23** (2013), 2183–2207.

[4] K. Bredies and D. A. Lorenz, *Linear convergence of iterative soft-thresholding*, J. Fourier Anal. Appl. **14** (2008), 813–837.

[5] R. S. Burachik, L. M. G. Drummond and S. Scheimberg, *On some properties and an application of the logarithmic barrier method*, Math. Program. **111** (2008), 95–112.

[6] A. Chambolle, *An algorithm for mean curvature motion*, Interfaces Free Bound. **6** (2004), 195.

[7] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision **40** (2011), 120–145.

[8] G. H.-G. Chen and R. T. Rockafellar, *Convergence rates in forward–backward splitting*, SIAM J. Optim. **7** (1997), 421–444.

[9] J.-S. Chen and S. Pan, *An entropy-like proximal algorithm and the exponential multiplier method for convex symmetric cone programming*, Computational Optimization and Applications **47** (2010), 477–499.

[10] J. X. da Cruz Neto, O. P. Ferreira, P. R. Oliveira and R. C. M. Silva, *Central paths in semidefinite programming, generalized proximal-point method and cauchy trajectories in riemannian manifolds*, J. Optim. Theory Appl. **139** (2008), 227.

[11] J. C. de Los Reyes, C.-B. Schönlieb and T. Valkonen, *Bilevel parameter learning for higher-order total variation regularisation models*, J. Math. Imaging Vision **57** (2017), 1–25.

[12] E. Esser, X. Zhang and T. F. Chan, *A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science*, SIAM J. Imaging Sci. **3** (2010), 1015–1046.

[13] J. Faraut and A. Korányi, *Analysis on Symmetric Cones*, Oxford University Press, 1994.

[14] L. Faybusovich, *Euclidean jordan algebras and interior-point algorithms*, Positivity **1** (1997), 331–357.

[15] L. Faybusovich, *Linear systems in Jordan algebras and primal-dual interior-point algorithms*, J. Comput. Appl. Math. **86** (1997), 149–175.

[16] D. Gabay, *Applications of the method of multipliers to variational inequalities*, in: *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, volume 15, Edited by M. Fortin and R. Glowinski, North-Holland, 1983, pp. 299–331.

[17] M. Grant and S. Boyd, *Graph implementations for nonsmooth convex programs*, in: *Recent Advances in Learning and Control*, Edited by V. Blondel, S. Boyd and H. Kimura, Lecture Notes in Control and Information Sciences, Springer-Verlag Limited, 2008, pp. 95–110.

[18] M. Grant and S. Boyd, *CVX: Matlab software for disciplined convex programming, version 2.1*, `http://cvxr.com/cvx` (2014).

[19] M. Halick, E. de Klerk and C. Roos, *Limiting behavior of the central path in semidefinite optimization*, Optimization Methods and Software **20** (2005), 99–113.

[20] D. Han and X. Yuan, *Local linear convergence of the alternating direction method of multipliers for quadratic programs*, SIAM J. Numer. Anal. **51** (2013), 3446–3457.
[21] B. He and X. Yuan, *Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective*, SIAM J. Imaging Sci. **5** (2012), 119–149.
[22] M. Hong and Z.-Q. Luo, *On the linear convergence of the alternating direction method of multipliers*, Math. Program. **162** (2017), 165–199.
[23] A. Kaplan and R. Tichatschke, *Interior proximal method for variational inequalities on non-polyhedral sets*, Discussiones Mathematicae, Differential Inclusions, Control and Optimization **27** (2007), 71–93.
[24] M. Koecher, *The Minnesota notes on Jordan algebras and their applications*, *Lecture Notes in Mathematics*, volume 1710, Springer-Verlag, Berlin, 1999.
[25] J. Liang, J. Fadili and G. Peyré, *Local linear convergence of forward–backward under partial smoothness*, Adv. Neural Inf. Process. Syst. **27** (2014), 1970–1978.
[26] I. Loris and C. Verhoeven, *On a generalization of the iterative soft-thresholding algorithm for the case of non-separable penalty*, Inverse Problems **27** (2011), 125007.
[27] J. Lpez and E. A. P. Quiroz, *Construction of proximal distances over symmetric cones*, Optimization (2017), doi: 10.1080/02331934.2016.1277998, published online.
[28] R. Monteiro, *Polynomial Convergence of Primal-Dual Algorithms for Semidefinite Programming Based on the Monteiro and Zhang Family of Directions*, SIAM J. Optim. **8** (1998), 797–812.
[29] R. D. Monteiro and F. Zou, *On the existence and convergence of the central path for convex programming and some duality results*, Computational Optimization and Applications **10** (1998), 51–77.
[30] Y. E. Nesterov and M. J. Todd, *Self-scaled barriers and interior-point methods for convex programming*, Math. Oper. Res. **22** (1997), 1–42.
[31] H. Ramírez and D. Sossa, *On the central paths in symmetric cone programming*, J. Optim. Theory Appl. (2016), 1–20.
[32] W. Ring, *Structural properties of solutions to total variation regularization problems*, ESAIM: Math. Model. Numer. Anal. **34** (2000), 799–810.
[33] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1972.
[34] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Optim. **14** (1976), 877–898.
[35] L. Rudin, S. Osher and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D **60** (1992), 259–268.
[36] S. H. Schmieta and F. Alizadeh, *Extension of primal-dual interior point algorithms to symmetric cones*, Math. Program. **96** (2003), 409–438.
[37] Q. Tran-Dinh, A. Kyrillidis and V. Cevher, *An inexact proximal path-following algorithm for constrained convex minimization*, SIAM J. Optim. **24** (2014), 1718–1745.
[38] T. Valkonen, *Diff-convex combinations of Euclidean distances: a search for optima*, number 99 in Jyväskylä Studies in Computing, University of Jyväskylä, 2008, `http://tuomov.iki.fi/m/thesis.pdf`, Ph.D Thesis.
[39] T. Valkonen, *Block-proximal methods with spatially adapted acceleration* (2017), `http://tuomov.iki.fi/m/blockcp.pdf`, submitted, arXiv: 1609.07373.
[40] T. Valkonen, *Testing and non-linear preconditioning of the proximal point method* (2017), `http://tuomov.iki.fi/m/proxtest.pdf`, submitted, arXiv: 1703.05705.
[41] T. Valkonen and T. Pock, *Acceleration of the PDHGM on partially strongly convex functions*, J. Math. Imaging Vision **59** (2017), 394–414.
[42] S. J. Wright and D. Orban, *Properties of the log-barrier function on degenerate nonlinear programs*, Math. Oper. Res. **27** (2002), 585–613.
[43] Z. Yu, Y. Zhu and Q. Cao, *On the convergence of central path and generalized proximal point method for symmetric cone linear programming*, Appl. Math. Inf. Sci. **7** (2013), 2327–2333.

Manuscript received April 16 2018
revised August 1 2018

T. Valkonen
ModeMat, Escuela Politécnica Nacional, Quito, Ecuador; *previously* Department of Mathematical Sciences, University of Liverpool, United Kingdom
E-mail address: tuomo.valkonen@iki.fi