INTERIOR–PROXIMAL PRIMAL–DUAL METHODS

Tuomo Valkonen∗

2017-06-23

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.

Due to arXiv’s inability to handle bibtex properly, and refusal to accept PDFs, references are broken in this file. Please get the correctly typeset version from http://tuomov.iki.fi/publications/.

1 INTRODUCTION

Interior point methods exhibit fast convergence on several non-smooth non-strongly-convex problems, including linear problems with symmetric cone constraints [1, 2, 3, 4]. 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

\[ \min_x G(x) + F(Kx) \]

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 [5, 6]. 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 [7, 8, 9, 10] 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?

∗Department of Mathematical Sciences, University of Liverpool, United Kingdom. tuomo.valkonen@iki.fi
The saddle point form of (P) is

\[
\min_x \max_y G(x) + \langle Kx, y \rangle - F^*(y).
\]

A popular algorithm for solving this problem is the primal–dual method of Chambolle and Pock [10]. As discovered in [11], the method can most concisely be written as a \textit{preconditioned proximal point method}, solving on each iteration for \(u^{i+1} = (x^{i+1}, y^{i+1})\) the variational inclusion

\[
(PP_0) \quad 0 \in H(u^{i+1}) + M_{i+1}(u^{i+1} - u^i),
\]

where the monotone operator

\[
H(u) := \left( \frac{\partial G(x) + K^*y}{\partial F^*(y) - Kx} \right)
\]

encodes the optimality condition \(0 \in H(\tilde{u})\) for (S). For the standard proximal point method [12], 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 \textit{preconditioning} or step length operator is given for suitably chosen step length parameters \(\tau_i, \sigma_{i+1}, \theta_i > 0\) by

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

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 \(\|x^i - \tilde{x}\|^2\). If both \(G\) and \(F^*\) are strongly convex, then the method converges linearly.

In our earlier work [13, 14, 15], 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:

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

\textit{for a symmetric cone} \(\mathcal{K}\), \textit{can we replace} \(M_{i+1}\) \textit{in (PP_0) by a non-linear interior point preconditioner that yields tractable sub-problems and a fast, convergent algorithm?}

Generalised proximal point methods motivated by interior point methods have been considered before in [16, 17, 18, 19]. Here the approach has essentially been to replace the squared distance in the proximal point method \(x^{i+1} := \arg \min_{x \in \mathcal{K}} G(x) + \frac{1}{2\tau_i} \|x - x^i\|^2\) for \(\min_{x \in \mathcal{K}} G(x)\) by a suitable Bregman distance supported on \(\text{int} \mathcal{K} \times \text{int} \mathcal{K}\), typically \(D(x, x') := \text{tr}(x \circ \ln x - x \circ \ln x' + x' - x)\). To the best of our knowledge, no convergence rates have been obtained using this approach. 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: in the general case only \(O(1/N)\), but linear convergence in the second-order cone under non-degeneracy and \(A = \langle a, \cdot \rangle\) for \(a \in \text{int} \mathcal{K}\). 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 \( \text{int} \mathcal{K} \), and "compatible" with \( \partial F^* \) 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 \( \partial F^* \) itself is missing.

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 [20], but primal–dual methods generally still require the strong convexity of \( F^* \) as well. For ADMM additional local estimates exist under quadratic [21, 22] or polyhedral assumptions [23]. 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 \( G \) [24, 25].

Our theoretical results depend on the convergence theory for non-linearly preconditioned proximal point methods from [14]. 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

\[
(x, z)_T := (Tx, z), \quad \text{and} \quad \|x\|_T := \sqrt{(x, x)_T}, \quad (x, z \in X).
\]

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 \( \tilde{u} = (\tilde{x}, \tilde{y}) \) satisfying the optimality conditions \( 0 \in H(\tilde{y}) \), in other words

\[
(\text{OC}) \quad -K^*\tilde{y} \in \partial G(\tilde{x}), \quad \text{and} \quad K\tilde{x} \in \partial F^*(\tilde{y}).
\]

For a function \( G \), as above, \( \partial G \) stands the convex subdifferential [26]. For a set \( C \), \( \partial C \) 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 \( \delta_C \) 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 \( F^*(\tilde{y}) = \delta_{B(0, \alpha_i)}(\tilde{y}_i) \) for \( \tilde{y} = (\tilde{y}_1, \ldots, \tilde{y}_n) \in \mathbb{R}^{mn} \). We may lift each \( \tilde{y}_i \) into \( \mathbb{R}^{1+m} \) as
\[ y_i = (y_{i,0}, \bar{y}_i), \text{ and replace } F^* \text{ by} \]
\[
(2.2) \quad \hat{F}^*(y) := \sum_{i=1}^{n} \delta_{C_i}(y_i), \quad \text{where} \quad C_i := \{y_i \in \mathcal{K} \mid Ay = b\},
\]
where, the linear constraint is defined by \( Ay := (y_{0,0}, \ldots, y_{n,0}) \) and \( b := (\alpha_i, \ldots, \alpha_n) \). The cone constraint is given by \( \mathcal{K} = \mathcal{K}_{soc}^n \) for the second-order cone
\[
\mathcal{K}_{soc} := \{y = (y^0, \bar{y}) \in \mathbb{R}^{s+m} \mid y_0 \geq \|\bar{y}\|\}.
\]

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 [27, 28].

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

(i) Powers \( x^m := \sum_{j=1}^{r} \lambda_j^m e_i \) when meaningful,
(ii) The determinant \( \det x := \prod_{j} \lambda_j \), and
(iii) The trace \( \text{tr} x := \sum_{j=1}^{r} \lambda_j \).
(iv) The inner product \( \langle x, y \rangle := \text{tr}(x \circ y) \), and the
(v) 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 \( Q_x := 2L(x)^2 - L(x) \). The invertibility of \( x \) is equivalent to the invertibility of \( Q_x \). Other important properties include \([27, 1]\)

\[
\begin{align*}
&\text{(vi) } Q_{x \circ y} = Q_x Q_y \text{ for } x \in \mathbb{R}, \\
&\text{(vii) } Q_{Q_x y} = Q_x Q_y Q_x \text{ (the fundamental formula of quadratic presentations),} \\
&\text{(viii) } Q_x x^{-1} = x, \\
&\text{(ix) } Q_x e = x^2, \text{ and} \\
&\text{(x) } \det(Q_x y) = \det(x^2) y = \det(x)^2 y.
\end{align*}
\]

Moreover, \( Q_x \) is self-adjoint with respect to the inner product defined above, and the eigenvalues are products \( \lambda_i(x) \lambda_j(x) \) \([28, 27]\), so that

\[
\lambda_{\min}^2(\mathcal{J}) ||y||^2 \leq \langle Q_x y, y \rangle \leq \lambda_{\max}^2(\mathcal{J}) ||y||^2 \quad \text{for all } y \text{ when } x \geq 0.
\]

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

\[
x \circ y = (x^T y, x_0 y_0 + y_0 \tilde{x}),
\]

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

\[
\langle x, y \rangle = 2x^T y.
\]

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

### 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 \([27]\) \( \mathcal{K}^* = -\mathcal{K} \), or the self-scaled cones of \([4]\). Their important properties include \([27, 28]\):

\[
\begin{align*}
&\text{(i) } \text{int } \mathcal{K} = \{ x \in \mathcal{J} \mid x \text{ is positive-definite} \} = \{ x \in \mathcal{J} \mid L(x) \text{ pos. def.} \}, \\
&\text{(ii) } \langle x, y \rangle \geq 0 \text{ for all } y \in \mathcal{K} \text{ iff } x \in \mathcal{K}, \text{ and} \\
&\text{(iii) } \langle x, y \rangle > 0 \text{ for all } y \in \mathcal{K} \setminus \{0\} \text{ iff } x \in \text{int } \mathcal{K}. \\
&\text{(iv) } Q_x \text{ for } x \in \text{int } \mathcal{K} \text{ maps } \mathcal{K} \text{ onto itself.} \\
&\text{(v) } \text{For } x, y \in \text{int } \mathcal{K}, \text{ there exists unique } a \in \mathcal{K}, \text{ such that } x = Q_a y.
\end{align*}
\]
(vi) For any \( x, y \in \mathcal{K} \), \( \langle x, y \rangle = 0 \) iff \( x \circ y = 0 \) \cite{29}.

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

(vii) The barrier function \( B(x) := -\log(\det x) \) tends to infinity as \( x \) goes to \( \text{bd} \mathcal{K} \).

(viii) \( \nabla B(x) = -x^{-1} \) and \( \nabla^2 B(x) = Q_x^{-1} \) (differentiated wrt. the norm in \( \mathcal{J} \)).

(ix) The normal cone \( N_K(x) = \{-y \in \mathcal{K} \mid \langle y, x \rangle = 0 \} \) for \( x \in \mathcal{K} \) \cite{30, 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,

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

If \( 0 \neq x \in \text{bd} \mathcal{K} \), we have \( x^2 = 2x_0x \). 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 \( N_K(x) = \{-aRx \mid a \geq 0 \} \).

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 \( (y_\mu, d_\mu, z_\mu) \in \text{int} \mathcal{K} \times \text{int} \mathcal{K} \times \mathbb{R}^k \) to the system

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

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

(SCLP) \quad 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

(2.5) \quad \min_{y \in \mathcal{K}, Ay=b} \langle c, y \rangle - \mu \log \det(y). \]
The set of solutions to $(\text{SCLP}_\mu)$ for varying $\mu > 0$ is called the central path. From [2, Theorem 2.2] we know that if there exists a primal–dual interior feasible point, i.e., some $(y^*, d^*, z^*) \in \text{int} \mathcal{K} \times \text{int} \mathcal{K} \times \mathbb{R}^k$ such that $Ay^* = b$ and $A^t z^* + c = d^*$, then there exists a solution $(y_\mu, d_\mu, z_\mu)$ to $(\text{SCLP}_\mu)$ 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 2.1.** Suppose there exists a primal interior feasible point $y^* \in \text{int} \mathcal{K} \cap \{ y \in \mathcal{K} \mid Ay = b \}$. Then there exists a solution $(y_\mu, d_\mu, z_\mu) \in \text{int} \mathcal{K} \times \text{int} \mathcal{K} \times \mathbb{R}^k$ to $(\text{SCLP}_\mu)$ for all $\mu > 0$.

**Proof.** The article [2] 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 level-bounded, proper, and lower semicontinuous, the problem (2.5) has a solution $y$. This $y$ has to satisfy $(\text{SCLP}_\mu)$ for some $d$ and $z$. Now [2, Theorem 2.2] applies. \qed

Practical methods [4, 1] for solving (SCLP) by closely following the central path are based on scaling the iterates $(y^t, d^t)$ by $Q_p$ for a suitable $p \in \text{int} \mathcal{K}$. We will need this scaling for different purposes, and therefore recall the following basic properties.

**Lemma 2.2.** Let $p \in \text{int} \mathcal{K}$, and $y, d \in \mathcal{K}$. Define $\tilde{y} := Q_p^{1/2} y$, and $\tilde{d} := Q_p^{-1/2} d$. Then

(i) $y \circ d = 0$ if and only if $\tilde{y} \circ \tilde{d} = 0$.

(ii) If $y, d \in \text{int} \mathcal{K}$ and $\mu > 0$, then $y \circ d = \mu e$ if and only if $\tilde{y} \circ \tilde{d} = \mu e$.

(iii) $(\text{SCLP})$ (resp. $(\text{SCLP}_\mu)$) is satisfied for $y$ and $d$ if and only if it is satisfied for $\tilde{y}$ and $\tilde{d}$ with $A$ and $c$ replaced by $\tilde{A} := AQ_p^{-1/2}$ and $\tilde{c} := Q_p^{-1/2} c$.

**Proof.** The claim (i) is a consequence of the properties Section 2.2(iv) and (vi). The claim (iii) is the content of [1, Lemma 28]. Finally, to establish (iii), the remaining linear equations in $(\text{SCLP})$ and $(\text{SCLP}_\mu)$ are obvious. \qed

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 $y + d \in \text{int} \mathcal{K}$. We say that $y$ is primal non-degenerate if

\begin{equation}
\label{2.6}
v = A^t z \quad \text{and} \quad y \circ v = 0 \implies v = 0.
\end{equation}

Likewise $d$ is dual non-degenerate if

\begin{equation}
\label{2.7}
A v = 0 \quad \text{and} \quad d \circ v = 0 \implies v = 0.
\end{equation}

**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 [31], 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 [32, 33, 34, 35, 36]. After all, in typical interior point methods, one is not interested in solving (SCLP) exactly; rather, one is interested in staying close to the central path while decreasing $\mu$ fast. So here we provide the result necessary for our work.
Lemma 2.3. Let $\tilde{y}, \tilde{d} \in \mathcal{K}$ and $\tilde{e} \in \mathbb{R}^k$ solve (SCLP). Also let $y_\mu, d_\mu, z_\mu \in \text{int} \mathcal{K}$ and $z_\mu \in \mathbb{R}^k$ solve (SCLP$_\mu$) for some $\mu > 0$. If $\tilde{y}$ and $\tilde{d}$ are strictly complementary, and both primal and dual non-degenerate, then

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

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 $(y_\mu, d_\mu, z_\mu)$ solves (SCLP$_\mu$) if and only if $y_\mu = \tilde{y} + \Delta y$ and $d_\mu = \tilde{d} + \Delta d$ with

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

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

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

This simply says that

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

From [2, Corollary 4.9] we know that the operator $M_{\tilde{y}, \tilde{d}}$ is invertible when the solution $(\tilde{y}, \tilde{d})$ is strictly complementary and both primal and dual non-degenerate. Moreover, for $(y_\mu, d_\mu)$ satisfying (SCLP$_\mu$), we know from [2, Corollary 4.6] that $M_{y_\mu, d_\mu}$ is invertible. In fact, both $M_{y_\mu, d_\mu}$ and $M_{\tilde{y}, \tilde{d}}$ are positive definite: in both cases, $(y, d) = (\tilde{y}, \tilde{d})$, and $(y, d) = (y_\mu, d_\mu)$, the map $m(\xi) := \langle \xi, M_{y,d}\xi \rangle$ 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 $(y, d) = (\tilde{y}, \tilde{d})$ the positivity follows from the assumed primal and dual non-degeneracy, as the operators $L(\tilde{y})$ and $L(\tilde{d})$ are positive semi-definite. For $(y, d) = (y_\mu, d_\mu) \in \text{int} \mathcal{K} \times \text{int} \mathcal{K}$, 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 [7, cf.|proof of Lemma 32]as-2003.

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

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

where the first inequality holds by the orthogonality of $\Delta y$ and $\Delta d$. The claim follows. \hfill \Box
2.5 STRONG MONOTONICITY OF THE BARRIER

If the barrier $B(y) = -\log(\det y)$ 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 2.4.** Let $y, y' \in \text{int} \mathcal{K}$, and denote $d := y^{-1}$, and $d' := (y')^{-1}$. Then

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

**Proof.** There exists a unique $w \in \text{int} \mathcal{K}$ s.t. $d' = Q_w^{-1}y$ and $d = Q_w^{-1}y'$; see, e.g., [4, Corollary 3.1]. We thus see (2.9) to hold if

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

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

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

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

$$
  Q_w^{-1} = Q_{w^{-1}} = Q_{y^{-1/2}}(Q_{y^{-1/2}}d')^{1/2} = Q_{y^{-1/2}}Q_{y^{-1/2}}d_y^{-1/2}.
$$

Following [37, p.42], from this we quickly compute

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

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

Continuing from (2.12), we also have

$$
  Q_w^{-1} = Q_{y^{-1/2}}[Q_{y^{-1/2}}d_y^{-1/2}]^{1/2}Q_{y^{-1/2}}.
$$

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

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

This proves (2.10) and consequently (2.9).\qed

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

**Lemma 2.5.** Let $y, d \in \text{int} \mathcal{K}$ 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

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

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

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

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

9
Proof. Let $Q_w$ be as in the proof of Lemma 2.4.

\[ -\langle d - \tilde{d}, y - \tilde{y} \rangle \stackrel{(2.15)}{=} -\langle d - d', y - \tilde{y} \rangle = \langle y - y', y - \tilde{y} \rangle_{Q_w^1} \]

\[ = \langle y - \tilde{y}, y - \tilde{y} \rangle_{Q^2} + \langle \tilde{y} - y', y - \tilde{y} \rangle_{Q^2} \]

\[ \geq (1 - \alpha)\|y - \tilde{y}\|_{Q^2}^2 - \frac{1}{16\alpha}\|y' - \tilde{y}\|_{Q_w^1}^2. \]

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

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

\[ Q_w^{-1} = Q_a^{1/2} Q_{a/2, w}^{-1} Q_a^{1/2} = Q_a^{1/2} Q_w^{-1} Q_a^{1/2}. \]

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

\[ Q_w^{-1} = Q_{d/2}^{-1} Q_{d^{-1/2}, Q_d} Q_{d^{-1/2}} \]

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

\[ \square \]

Lemma 2.6. Let $y, d \in \text{int} K$ and $\tilde{y}, \tilde{d} \in K$ with $u \circ d = \mu e$ for some $\mu > 0$, and $\tilde{y} \circ \tilde{d} = 0$. Suppose there exist $y', d' \in K$ such that $\langle d - d', y - \tilde{y} \rangle = 0$ and $y' \circ d' = \mu e$. Then for any $\alpha \in (0, 1)$ holds

\[ -\langle d - \tilde{d}, y - \tilde{y} \rangle \geq \frac{(1 - \alpha)\mu}{\lambda_{\max}(\tilde{y})\lambda_{\max}(y')}\|y - y'\|_{Q_a}^2 - \frac{\lambda_{\max}(d)\lambda_{\max}(d')}{4\alpha\mu}\|y' - \tilde{y}\|_{Q_a}^2. \]

Proof. We apply Lemma 2.5 with $\tilde{d}, d,$ and $d'$ replaced by $\tilde{d}/\mu, d/\mu,$ and $d'/\mu$. This causes the right-hand-side of the estimate (2.16) to be multiplied by $\mu$, along with both $\lambda_{\max}(d)$ and $\lambda_{\max}(d')$ to be divided by $\mu$.

\[ \square \]

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

Proposition 2.7. Suppose $Ay = b$ implies $(a, y) = b_0$ for some $a \in \text{int} K$ and $b_0 > 0$. Fix $\mu > 0$, and let $(y, d, z) \in \text{int} K \times \text{int} K \times \mathbb{R}^k$ solve (SCLP)$_\mu$. Likewise, suppose $(y_\mu, d_\mu, z_\mu) \in \text{int} K \times \text{int} K \times \mathbb{R}^k$ solves (SCLP)$_\mu$ for $c = \hat{c}$, where $(\tilde{y}, \tilde{d}, \tilde{z})$ solves (SCLP) for $c = \hat{c}$. If $\tilde{y}$ and $\tilde{d}$ are strictly complementary, $\tilde{d}$ dual non-degenerate, and $\tilde{y}$ primal non-degenerate, then for any $\alpha \in (0, 1)$ holds

\[ -\langle d - \tilde{d}, y - \tilde{y} \rangle \geq \frac{(1 - \alpha)\mu}{b_0^2}\|y - \tilde{y}\|_{Q_a}^2 - \frac{C_{c, \mu}C_{\hat{c}, \mu}r}{\alpha\lambda_{\min}(M_{\tilde{y}, d})^2\mu}, \]

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

\[ C_{c, \mu} := \frac{\mu r + 2b_0\|c\|_{Q_a^*}}{\lambda_{\min}(y^*)}. \]
Proof: We begin by applying Lemma 2.6 with \((y', d')\) set to the \(\mu\)-approximation \((y_\mu, d_\mu)\) to \((\tilde{y}, \tilde{d})\) provided by Lemma 2.3. Inserting (2.8) into (2.18), we therefore obtain
\[
\langle d - \tilde{d}, y - \tilde{y}\rangle \geq \frac{(1 - \alpha)\mu}{\lambda_{\max}(\tilde{y})\lambda_{\max}(\tilde{y}_\mu)}\|y - y_\mu\|_Q^2 - \frac{\mu\lambda_{\max}(d)\lambda_{\min}(d_\mu)\|\alpha\{M_{y, d}\}^2}{\lambda_{\min}(M_{y, d})^2}. \tag{2.21}
\]

It remains to estimate the eigenvalues in this expression.

First of all, we easily derive the necessary bounds on \(\lambda_{\max}(\tilde{y})\) and \(\lambda_{\max}(y')\) as
\[
\lambda_{\max}(\tilde{y}) \leq \text{tr}(\tilde{y}) = \langle e, \tilde{y}\rangle = \langle a, y\rangle = b_0. \tag{2.22}
\]

Secondly, regarding the estimate on \(\lambda_{\max}(d)\), we fix some \(y^* \in \text{int} \mathcal{K}\) satisfying \(Ay^* = b\). Such a point exist by our assumption of there existing solutions to (SCLP\(_\mu\)); see also Lemma 2.1. Since \(d = A'z + c\) for some \(z \in \mathbb{R}^k\), and \(d \circ y = \mu e\), we then derive
\[
\lambda_{\min}(y^*)\lambda_{\max}(d) \leq \lambda_{\min}(y^*)(\epsilon, d) \leq \langle y^*, d \rangle = \langle \tilde{y}^*, d \rangle
\]
\[
= \langle \tilde{y}^*, d \rangle + \langle \tilde{y}^* - \tilde{y}, d \rangle = \mu r + \langle \tilde{y}^* - \tilde{y}, e \rangle \leq \mu r + \|e\|\|\lambda_{\min}(\tilde{y}) + \lambda_{\max}(y^*)\| \leq \mu r + 2b_0\|e\|.
\]

In the last inequality we have used (2.22) for both \(\tilde{y}\) and \(\tilde{y}^*\). Since \(y^* \in \mathcal{K}\), so that \(\lambda_{\min}(y^*) > 0\), and \(\|e\| = \|c\|_{Q; J}^\dagger\), this gives the claimed bounds on \(\lambda_{\max}(d)\) and \(\lambda_{\max}(d')\). \(\square\)

2.6 STRONG MONOTONICITY OF THE BARRIER IN THE SECOND-ORDER CONE

In the second-order cone \(\mathcal{K} = \mathcal{K}_{\text{soc}} \subset \mathbb{E}_{1+m}\), under suitable constraints \(Ay = b\), we have a stronger result.

Lemma 2.8. Suppose \(y, y', d, d' \in \text{int} \mathcal{K}_{\text{soc}}\) with \(y \circ d = y' \circ d' = \mu e\) for given \(\mu > 0\). Then
\[
-\langle d - d', y - y'\rangle_J \geq \frac{\text{det}(d) + \text{det}(d')}{\mu}\|y - y'\|_{\mathbb{R}}^2,
\tag{2.23}
\]
where \(\|y - y'\|_{\mathbb{R}}^2 := \|\tilde{y} - \tilde{y}'\|_{\mathbb{R}^m}^2 = (y_0 - y_0')\|y - y'\|^2 = -\det(y - y')\).

Proof: We have \(d = \mu Ry/\det(y) = \mu^{-1}(d)Ry\). Likewise \(d' = \mu^{-1}(d')Ry\). We write for brevity \(\beta := \mu^{-1}\text{det}(d)\) and \(\beta' := \mu^{-1}\text{det}(d')\). Then
\[
-\langle d - d', y - y'\rangle_J = -\langle \beta' y - \beta' y', y - y'\rangle_J = 2\langle \beta y - \beta y', y - y'\rangle_{-R},
\]
where the second “inner product” is \(\langle x, y\rangle_{-R} := -\langle Rx, y\rangle_{\mathbb{E}_{1+m}}\). We can thus write
\[
-\langle d - d', y - y'\rangle_J = 2\beta\|y - y'\|_{-R}^2 + 2\langle \beta - \beta'\rangle\langle y, y - y'\rangle_{-R}
\]
as well as
\[
-\langle d - d', y - y'\rangle_J = 2\beta'\|y - y'\|_{-R}^2 + 2\langle \beta - \beta'\rangle\langle y, y - y'\rangle_{-R}.
\]
Summing these two expressions we deduce
\[
-\langle d - d', y - y'\rangle_J = (\beta + \beta')\|y - y'\|_{-R}^2 + (\beta - \beta')(\|y\|_{-R}^2 - \|y'\|_{-R}^2).
\tag{2.24}
\]
Now observe that
\[ \|y\|^2 - \|\tilde{y}\|^2 = -\det(y) = -\mu^2/\det(d). \]
Thus
\[ (\beta - \beta')(\|y\|^2 - \|\tilde{y}\|^2) = \mu(\det(d) - \det(d'))(\det(d')^{-1} - \det(d)^{-1}) \]
\[ = \mu(\det(d') - \det(d))^2 / (\det(d) \det(d')) > 0. \]

This and (2.24) immediately prove the claim. \( \square \)

For solutions of (SCLP\( \mu \)) 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
\[ D_F(w, d) := \|Q_w^{1/2} d - \mu_w a e\| \quad \text{for} \quad \mu_w, d = \langle w, d \rangle / r, \quad (w, d \in \mathcal{K}). \]

This distance is typically used to define the so-called short-step neighbourhood of the central path [?, see, e.g.,] as-2003.

**Lemma 2.9.** Suppose \( y, d \in \text{int} \mathcal{K}_{soc} \) with \( y \circ d = \mu e \) and \( \langle a, y \rangle = b_0 \) for some \( \mu, b_0 > 0 \) and \( a \in \text{int} \mathcal{K}_{soc} \). Then
\[ 2\mu^2 + \sqrt{2} b_0 D_F(a^{-1}, d) \mu \leq \det(d) \leq \frac{4\mu^2 + \sqrt{2} b_0 D_F(a^{-1}, d) \mu}{b_0^2 \det(a)}. \]

**Proof.** We define \( \tilde{y} := Q_w^{1/2} y \) and \( \tilde{d} := Q_w^{-1/2} d \). Then \( \langle e, \tilde{y} \rangle = \langle a, y \rangle = b_0 \), and by [1, Lemma 28], \( \tilde{y} \circ \tilde{d} = \mu e \). These conditions expand to \( \tilde{y}_0 d_0 + \tilde{y}_0^T \tilde{d} = \mu, \tilde{y}_0 \tilde{d} + d_0 \tilde{y} = 0 \), and \( 2\tilde{y}_0 = b_0 \). (In the latter, recall that the \( E_{1+m} \)-inner product satisfies \( \langle e, \tilde{y} \rangle = 2e^T \tilde{y} \).) We reduce this system to \( d_0^2 - \|d\|^2 - 2d_0 \mu / b_0 = 0 \), from where we solve
\[ d_0 = \frac{\mu + \sqrt{\mu^2 + b_0^2 \|d\|^2}}{b_0}. \]

Thus
\[ \det(d) = d_0^2 - \|d\|^2 = \frac{2\mu^2 + 2\mu \sqrt{\mu^2 + b_0^2 \|d\|^2}}{b_0^2}, \]
from which we easily estimate
\[ \frac{2\mu^2 + 2\mu b_0 \|d\|}{b_0^2} \leq \det(d) \leq \frac{4\mu^2 + 2\mu b_0 \|d\|}{b_0^2}. \]

To finish deriving (2.25), from Section 2.1(x) we recall that \( \det(d) = \det(a) \det(d) \). We also have \( r d_0 = \langle d, e \rangle = \langle d, a^{-1} \rangle \) for the rank \( r = 2 \), so
\[ \sqrt{2} \|d\|_{\mathcal{R}^n} = \|d - d_0 e\|_{\mathcal{J}} = \|Q_w^{1/2} d - \mu_w a^{-1} a e\|_{\mathcal{J}} = D_F(a^{-1}, d), \]
where we emphasise the standard Euclidean norm on \( d \in \mathcal{R}^n \) versus the \( \sqrt{2} \)-scaled standard norm on \( \mathcal{J} \). With this, (2.27) gives (2.25). \( \square \)
If $D_F(a^{-1}, \tilde{d}) > 0$, or alternatively $\det(\tilde{y}) > 0$, then the next proposition shows local strong monotonicity of the barrier for $d$ close to $\tilde{d}$ and $\mu > 0$ small. Moreover, if $D_F(a^{-1}, \tilde{d}) > 0$, the factor of strong monotonicity does not vanish as $\mu \searrow 0$.

**Proposition 2.10.** Let $\mathcal{K} = \mathcal{K}_{\text{soc}}$, and suppose $Ay = b$ implies $\langle a, y \rangle = b_0$ for some $a \in \text{int} \mathcal{K}$ and $b_0 > 0$. Let $(y, d, z) \in \text{int} \mathcal{K} \times \text{int} \mathcal{K} \times \mathbb{R}^k$ solve (SCLP), and likewise that $(\tilde{y}, \tilde{d}, \tilde{z}) \in \mathcal{K} \times \mathcal{K} \times \mathbb{R}^k$ solve (SCLP) for $c = \tilde{c}$. Then

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

**Proof.** We have

$$
0 = \tilde{y} \circ \tilde{d} = (\tilde{y}_0 \tilde{d}_0 + \tilde{y}^T \tilde{d}, \tilde{y}_0 \tilde{d} + \tilde{d}_0 \tilde{y}).
$$

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

$$
\tilde{\beta} = \frac{\tilde{y}}{\tilde{y}_0^2} = \frac{\tilde{d}_0}{\tilde{y}_0} = \frac{\|\tilde{d}\|_{Q_m}}{\tilde{y}_0} = \frac{\sqrt{Z}D_F(a^{-1}, \tilde{d})}{b_0} \geq 0.
$$

In the final step we have reasoned as in (2.28). We may therefore repeat the steps of Lemma 2.8 until (2.24) to obtain

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

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

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

If $a = e$, we have $y_0 = \tilde{y}_0 = b_0/2$, so that $2\|y - \tilde{y}\|_{R}^2 = \|y - \tilde{y}\|_{R}^2$. The final equality from (2.31) also gives $\tilde{\beta} = \sqrt{2}D_F(e, \tilde{d})/b_0$. With the help of Lemma 2.9, (2.33) thus yields

$$
\langle d - \tilde{d}, y - \tilde{y} \rangle \geq \frac{2\mu + \sqrt{2}b_0[D_F(e, d) + D_F(e, \tilde{d})]}{b_0^2} \|y - \tilde{y}\|_{R}^2 - \mu + \frac{\sqrt{2}b_0 D_F(e, \tilde{d})}{4\mu + \sqrt{2}b_0 D_F(e, d)} \frac{2\mu + \sqrt{2}b_0 D_F(e, d)}{b_0^2} \det(\tilde{y}).
$$

Since $\lambda_{\min}(e) = \det(e) = 1$, the estimate (2.29) is immediate in the case $a = e$. 

13
If \( a \neq e \), we define \( \hat{y} := Q_a^{1/2}y \), and \( \hat{d} := Q_a^{-1/2}d \) as in Lemma 2.9. Then \((\hat{y}, \hat{d}, z)\) continues to satisfy (SCLP), with \( A \) replaced by \( \hat{A} := AQ_a^{-1/2} \) and \( \xi := Q_a^{-1/2}c \). The same holds with (SCLP) for \( \hat{y} := Q_a^{1/2}y \) and \( \hat{d} := Q_a^{-1/2}d \). Therefore, (2.34) holds for these transformed variables. Since \( D_F(e, d) = D_F(a^{-1}, d) \), as well as \( \|y - \tilde{y}\|^2 = \|y - \hat{y}\|_{Q_a}^2 \), and \(-\langle d - d', y - y'\rangle = -\langle d - \hat{d}, \hat{y} - \tilde{y}\rangle \), we obtain the claim.

**Corollary 2.11.** Let \( \mathcal{K} = \mathcal{K}_{soc} \), and suppose \( A = (a, \cdot) \) for some \( a \in \text{int} \mathcal{K} \). Suppose moreover that \( \langle a^{-1}, c \rangle = \langle a^{-1}, \hat{c} \rangle = 0 \). Let \((y, z) \in \text{int} \mathcal{K} \times \text{int} \mathcal{K} \times \mathbb{R}^k \) solve (SCLP), and likewise that \((\hat{y}, \hat{d}, z) \in \mathcal{K} \times \mathcal{K} \times \mathbb{R}^k \) solve (SCLP) for \( \hat{c} = \hat{c} \). If \( \hat{c} \neq 0 \), then

\[
\langle d - \hat{d}, y - \hat{y} \rangle \geq \frac{\mu + 2^{-1/2} b_0 \|c\|_{\mathcal{Q}_a} + \|\hat{c}\|_{\mathcal{Q}_a}}{b_0^2} \|y - \hat{y}\|_{\mathcal{Q}_a}^2 - \mu.
\]

Otherwise, if \( \hat{c} = 0 \) with \( \hat{y} = ba^{-1}/2 \), then

\[
\langle d - \hat{d}, y - \hat{y} \rangle \geq \frac{\mu + 2^{-1/2} b_0 \|c\|_{\mathcal{Q}_a}}{b_0^2} \|y - \hat{y}\|_{\mathcal{Q}_a}^2.
\]

We say that (2.35) is strong monotonicity of the barrier “with a penalty”, \( \mu \).

**Proof.** We do not until the very end of the proof use the assumption \( A = (a, \cdot) \). For now, we use the weaker assumption that \( Ay = b \) implies \( \langle a, y \rangle = b_0 \). We apply Proposition 2.10. This gives

\[
\langle d - \hat{d}, y - \hat{y} \rangle \geq \frac{\mu + 2^{-1/2} b_0 \|D_F(a^{-1}, d) + D_F(a^{-1}, \hat{d})\|}{b_0^2} \|y - \hat{y}\|_{\mathcal{Q}_a}^2 - \mu
\]

\[
+ \frac{2^{-1/2} b_0 \mu}{2 \mu + 2^{-1/2} b_0 D_F(a^{-1}, d)} \mu + \frac{2^{-1/2} b_0 D_F(a^{-1}, d)}{b_0^2/2} \det(Q_a^{1/2} \hat{y}).
\]

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

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

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

\[
\langle d - \hat{d}, y - \hat{y} \rangle \geq \frac{\mu + 2^{-1/2} b_0 D_F(a^{-1}, d)}{b_0^2} \|y - \hat{y}\|_{\mathcal{Q}_a}^2.
\]

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

\[
\langle d - \hat{d}, y - \hat{y} \rangle \geq \frac{\mu + 2^{-1/2} b_0 \|D_F(a^{-1}, d) + D_F(a^{-1}, \hat{d})\|}{b_0^2} \|y - \hat{y}\|_{\mathcal{Q}_a}^2 - \mu.
\]
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

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

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

\[ \square \]

3. AN ABSTRACT PRECONDITIONED PROXIMAL POINT ITERATION

In this section, we recall some of the core results from [14]. We start by setting

\[
(3.1) \quad H(u) := \left( \frac{\partial G(x) + K^* y}{\partial F^*(y) - Kx} \right),
\]

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

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

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

\[
(PP) \quad 0 \in W_{i+1} H(u^{i+1}) + V_{i+1}(u^{i+1})
\]

for the unknown next iterate $u^{i+1}$. To obtain convergence rates for the resulting method, the idea from [14, 13] 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

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

which is an operator measure of strong monotonicity of $H$.

The next lemma, which is relatively trivial to prove [14], forms the basis from which our work proceeds.

**Theorem 3.1.** Let us be given $K \in \mathcal{L}(X; Y)$, $G \in C(X)$, and $F^* \in C(Y)$ 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

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

15
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

\[
\frac{1}{2}||u^{i+1} - u^i||_Z^2 + \frac{1}{2}||u^{i+1} - \hat{u}||_{Z_{i+1}(\Xi_{i+1}(\lambda)+M_{i+1})-Z_{i+2}M_{i+2}}^2 \\
+ (\partial F^*(y^{i+1}) - \partial F^*(\hat{y}), y^{i+1} - \hat{y})_{\Psi_{i+1}\Sigma_{i+1}} + \langle Z_{i+1}V_{i+1}(u^{i+1}), u^{i+1} - \hat{u} \rangle \geq -\Delta_{i+1}
\]

holds, then

\[
\frac{1}{2}||u^N - \hat{u}||_Z^2 \leq \frac{1}{2}||u^0 - \hat{u}||_Z^2 + \sum_{i=0}^{N-1} \Delta_{i+1}, \quad (N \geq 1).
\]

**Proof.** This is [14, 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 \( \Gamma = \gamma I \).

It is possible to extend this theorem to provide an estimate on an ergodic duality gap [? see] Theorem 4.6 of the proxtest. 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 \( y \in \text{int} \mathcal{K} \) with \( Ay = b \). Then the subdifferential sum formula (see, e.g., [26]) applies, so that

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

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

\[
N_\mathcal{K}(y) = \{ -d \mid d \in \mathcal{K}, \langle p, d \rangle = 0 \} \quad (y \in \mathcal{K}).
\]

Thus \( 0 \in H(\hat{u}) \) may also be written as the existence of \( (\hat{x}, \hat{y}, \hat{d}, \hat{z}) \in X \times \mathcal{K} \times \mathcal{K} \times Z \) with

\[
-\mathcal{K}^*\hat{y} \in \partial G(\hat{x}), \quad A\hat{y} = b, \quad A^*\hat{z} - K\hat{x} = \hat{d}, \quad \hat{y} \circ \hat{d} = 0.
\]

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 [14] could be used to extend the analysis to the gap. Moreover, following the approach of [15], 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

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

where $d^{i+1} \in \text{int} \mathcal{K}$ 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 (FP) may be written as the system

\begin{align}
(4.2a) & \quad 0 \in \tau_i \partial G(x^{i+1}) + \tau_i K^* y^{i+1} + (x^{i+1} - x^i), \\
(4.2b) & \quad 0 \in \sigma_{i+1}[A^* z^{i+1} - K x^i - d^{i+1}], \quad \text{as well as} \\
(4.2c) & \quad y^{i+1} \circ d^{i+1} = \mu_{i+1} e \quad \text{and} \quad Ay^{i+1} = b \quad \text{with} \quad y^{i+1}, d^{i+1} \in \text{int} \mathcal{K}.
\end{align}

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 (4.2). Suppose for some $\omega_{i+1}, \delta_{i+1} \in \mathbb{R}$ for all $i \in \mathbb{N}$ that

\begin{align}
(4.4a) & \quad -\langle d^{i+1} - \bar{d}, y^{i+1} - \bar{y} \rangle \geq \omega_{i+1} ||y^{i+1} - \bar{y}||^2 - \delta_{i+1}, \\
(4.4b) & \quad \psi_{i+1} \sigma_{i+1} = \phi_i \tau_i, \\
(4.4c) & \quad 2\omega_{i+1} \geq \tau_i ||K||^2, \quad \text{and} \\
(4.4d) & \quad \phi_{i+1} \leq \phi_i (1 + 2 \tau_i \bar{y}).
\end{align}

Then (Co-T) 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

$$
(4.5) \quad Z_{i+1} M_{i+1} = \begin{pmatrix}
\phi_i I & 0 \\
0 & 0
\end{pmatrix} \succeq 0.
$$

**Proof.** The condition (Co-T) now reads

\begin{equation}
(4.6) \quad \frac{1}{2} ||u^{i+1} - u^i||^2_{Z_{i+1} M_{i+1}} + \frac{1}{2} ||u^{i+1} - \bar{u}||^2_{Z_{i+1}} + \psi_{i+1} \sigma_{i+1} (K(x^{i+1} - x^i), y^{i+1} - \bar{y}) \quad \text{step in local norm} \\
\text{lin. precondition. upd. d.} \quad \text{de-coupling term from $V'$} \\
+ \psi_{i+1} \sigma_{i+1} (A^* (z^{i+1} - \bar{z}), y^{i+1} - \bar{y}) - \psi_{i+1} \sigma_{i+1} (d^{i+1} - \bar{d}, y^{i+1} - \bar{y}) \geq -\Delta_{i+1} \quad F^* \text{ term from (Co-T) as well as } d^{i+1} \text{ from } V'
\end{equation}

with the linear preconditioner update discrepancy

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

The expansion and estimate (4.5) are trivially verified along with the self-adjointness of $Z_{i+1} M_{i+1}$. This expansion allows us to write

$$
D_{i+2} = \begin{pmatrix}
\phi_i (1 + 2 \tau_i \bar{y}) I - \phi_{i+1} I \\
-2 \psi_{i+1} \sigma_{i+1} K
\end{pmatrix}.
$$
We use (4.4b) to cancel the off-diagonals of $D_{t+2}$ in (4.6). Then we use the fact that $A(y^{i+1} - \tilde{y}) = 0$ to cancel the first term on the second line of (4.6). Finally, we use $\Delta_{t+1} = \psi_{t+1}\sigma_{t+1}d_{t+1}$ and (4.4a) to estimate the second term on the second line of (4.6). This gives the condition

$$\phi_i \frac{1}{2} \| x^{i+1} - x^i \|^2 + \frac{\psi_{t+1}\sigma_{t+1}d_{t+1}}{2} \| y^{i+1} - \tilde{y} \|^2 + \frac{\phi_i(1 + 2y\tau_i) - \phi_{i+1}}{2} \| x^{i+1} - \tilde{x} \|^2 + \psi_{t+1}\sigma_{t+1}(K(x^{i+1} - x^i), y^{i+1} - \tilde{y}) \geq 0.\tag{4.7}$$

Application of (4.4d), as well as Cauchy’s inequality to the inner product term, shows that (4.7) and consequently (Co-I) is satisfied if

$$\psi_{t+1}\sigma_{t+1}d_{t+1} \geq \frac{1}{2} \langle e \rangle.$$

This follows from (4.4b) and (4.4c).

We define $\tau_i$ through (4.4c) for a lower bound $\omega_{n,i+1}$ of $\omega_{i+1}$. Likewise, we take (4.4d) as an equality as the definition of $\phi_{i+1}$. We observe that $\sigma_{i+1}$ and $\psi_{i+1}$ are irrelevant to the algorithm in (4.3), as will be the specific choice of $\phi_0 > 0$ to the satisfaction of (4.4). Taking $\phi_0 = 1$, we obtain Algorithm 4.1 from (4.3).

**Algorithm 4.1 (Barrier-preconditioned primal–dual method).**

**Require:** Linear operator $K \in L(X; J)$, strongly convex $G \in C(X)$, and $F^* \in C(J)$ of the form (4.1). Factor $\gamma > 0$ of the strong convexity of $G$. Rules for $\mu_i, \omega_{n,i} > 0$.

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 $\mu_i, \omega_{n,i}$, and step length
      $$\tau_i := 2\omega_{n,i+1}/\| K \|^2.$$
   5. Update testing parameter
      $$\phi_{i+1} := \phi_i(1 + 2\gamma\tau_i).$$
   6. Perform dual update by solving for $(y^{i+1}, d^{i+1}, z^{i+1}) \in \text{int } K \times \text{int } K \times Z$ 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. Perform primal update
      $$x^{i+1} := (I + \tau_i \partial G)^{-1}(x^i - \tau_i K^*y^{i+1}).$$
4. **until** a stopping criterion is satisfied.
Remark 4.2 (Solution of Line 6 of Algorithm 4.1). The system on Line 6 is a standard (SCLP). In the second-order cone with $A = \langle e, \cdot \rangle$ and $(e, \mathcal{R}(K)) = \{0\}$, it is easy to solve. Indeed, \( (0, \bar{d}^{i+1}) = -Kx^i \) while \( d_0^{i+1} \) is given by the expression in (2.26). Finally
\[
y^{i+1} = \mu_{i+1}(d_0^{i+1})^{-1} = \frac{\mu_{i+1}Rd^{i+1}}{\det(d_0^{i+1})} = \frac{\mu_{i+1}Rd^{i+1}}{(d_0^{i+1})^2 - ||d^{i+1}\||^2)}.
\]
More general cases \( A = \langle a, \cdot \rangle \) and \( \langle a^{-1}, \mathcal{R}(K) \rangle = \{0\} \) follow by scaling.

4.2 CONVERGENCE RATES IN GENERAL SYMMETRIC CONES

We still need to specify \( \mu_{i+1} \), verify (4.4a), 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 4.1 be satisfied. Assuming that \( Ay = b \) implies \( \langle a, y \rangle = b_0 \) for some \( a \in \text{int} \mathcal{K} \) and \( b_0 > 0 \), suppose there exists a solution \( (\bar{x}, \bar{y}, d, \bar{z}) \in X \times \mathcal{K} \times \mathcal{K} \times Z \) to (IOC) with \( \bar{y} \) and \( d \) strictly complementary, \( d \) dual non-degenerate, and \( \bar{y} \) primal non-degenerate. Suppose further that \( \text{dom} \mathcal{C} \) is bounded, or that the primal iterates \( (x^i)_{i \in \mathbb{N}} \) of Algorithm 4.1 stay bounded through other means. For some constant \( \theta > 0 \) and \( \xi \in (0, b_0^2) \), take
\[
(4.8) \quad \mu_{i+1} := \theta \phi_i^{-1/2}, \quad \text{and} \quad \omega_{x, i+1} := \zeta \lambda_{\min}(a) \mu_{i+1}.
\]
Then \( \|x^N - \bar{x}\|^2 = O(1/N) \).

Remark 4.4. The assumption \( Z = \mathbb{R}^k \) is merely for the simplicity of application of Proposition 2.7 and later Corollary 2.11. There would be nothing stopping us from applying the results on uncountable products of symmetric cones, for example.

Proof. We use Proposition 2.7, which verifies (4.4a) with
\[
\delta_i^{i+1} \leq \hat{C}C_{-Kx^i, \mu_{i+1}}C_{-K\bar{x}, \mu_{i+1}} \mu_{i+1} \quad \text{and} \quad \omega_{x, i+1} = \omega_{x, i+1} = \zeta \lambda_{\min}(a) \mu_{i+1}
\]
for \( C_{-Kx^i, \mu_{i+1}}, C_{-K\bar{x}, \mu_{i+1}} \) 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 \( \mu_{i+1} \) in (4.8) and Line 5 of the algorithm. Therefore \( \delta_{i+1} \leq C \mu_{i+1} \) for some constant \( C > 0 \). From (4.4b) and (4.8) it now follows
\[
(4.9) \quad \Delta_{i+1} := \bar{\psi}_{i+1} \sigma_{i+1} \delta_{i+1} \leq C r_i \phi_i \mu_{i+1} = C \theta r_i \phi_i^{1/2}.
\]

Next we use Theorem 3.1 and Lemma 4.1. For \( C_0 := \frac{1}{2} \|u^0 - \bar{u}\|^2 \), (3.4), (4.5), and (4.9) give the combined estimate
\[
(4.10) \quad \frac{\phi N}{2} \|x^N - \bar{x}\|^2 \leq C_0 + C \theta \sum_{i=0}^{N-1} r_i \phi_i^{1/2}, \quad (N \geq 1).
\]
Inserting \( \omega_{x, i+1} \) and \( \mu_{i+1} \) from (4.8), Lines 4 and 5 of the algorithm say
\[
\phi_{i+1} = \phi_i + \gamma \nu \phi_i^{1/2} \quad \text{and} \quad r_i = \phi_i^{-1/2} \nu / \|K\|^2 \quad \text{for} \quad \nu := 2 \zeta \lambda_{\min}(a) \theta.
\]
It follows [?, see ]tuomov-cpaccel that \( \phi_N = \Theta(N^2) \), while \( \sum_{i=0}^{N-1} r_i \phi_i^{1/2} = N \nu / \|K\|^2 \). Inserting these estimates into (4.10), we verify the \( O(1/N) \) rate.

\[\square\]
4.3 Convergence Rates in the Second-Order Cone

In the second-order cone, we obtain linear convergence under dual non-degeneracy:

**Theorem 4.5.** For $\mathcal{K} = \mathcal{K}_{\text{soc}}$ the second-order cone, $Z = \mathbb{R}^k$, and $A = \langle a, \cdot \rangle$ for some $a \in \text{int} \mathcal{K}$ with $(a^{-1}, \mathcal{R}(K)) = \{0\}$, let the requirements of Algorithm 4.1 be satisfied. Suppose there exists a solution $(\bar{x}, \bar{y}, \bar{d}, \bar{z}) \in X \times \mathcal{K} \times \mathcal{K} \times Z$ to (IOC). If $K\bar{x} = 0$, take $\bar{y} = ba^{-1}/2$ and $\bar{d} = 0$ for some $\theta > 0$ and $\zeta \in (0, b_0^{-2})$, take

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

Suppose further that $\text{dom } G$ is bounded, or that the primal iterates $\{x^i\}_{i \in \mathbb{N}}$ of Algorithm 4.1 stay bounded through other means. Then for some $C, \varepsilon > 0$ holds

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

**Proof.** From Line 4 of the algorithm and (4.11), we expand

$$\tau_i := 2(\xi \phi_i^{-1/2} + \bar{\ell}_{i+1})\lambda_{\text{min}}(a)/\|K\|^2 \quad \text{for} \quad \bar{\ell}_{i+1} := 2^{-1/2}b_0^{-1}\|Kx^i\|_{Q^i}.\tag{4.12}$$

From (4.12) and Line 5, we estimate

$$\phi_i \geq \phi_0 + 2\gamma \theta \sum_{i=0}^{N-1} \phi_i^{1/2}.\tag{4.13}$$

It follows from (4.12) that $\sup_i \tau_i \leq C_\tau$ for some constant $C_\tau > 0$. From (4.11), we also obtain $\mu_{i+1} \searrow 0$.

We then use Corollary 2.11, which verifies (4.4a) with

$$\omega_{\zeta,i+1} := (\mu_{i+1}\zeta + \ell_{i+1})\lambda_{\text{min}}(a), \quad \begin{cases} \ell_{i+1} := \frac{\|Kx^i\|_{Q^i}}{\sqrt{2}b_0}, \quad \text{and} \quad \delta_{i+1} := 0, & \text{if } K\bar{x} = 0, \\ \ell_{i+1} = \frac{\|K\bar{x}\|_{Q^i} + \|Kx^i\|_{Q^i}}{\sqrt{2}b_0}, \quad \text{and} \quad \delta_{i+1} = \mu_{i+1}, & \text{if } K\bar{x} \neq 0, \end{cases}$$

Setting $\ell := \|K\bar{x}\|_{Q^i}/(\sqrt{2}b_0) > 0$, we have $\ell_{i+1} = \bar{\ell}_{i+1} + \ell$.

Next we use Theorem 3.1 and Lemma 4.1. Recalling (4.4b) and that $\Delta_{i+1} = \psi_{i+1}\sigma_{i+1}\delta_{i+1}$ in Lemma 4.1, setting $C_0 := \frac{1}{8}\|u^0 - \bar{u}\|^2_{Z,M}$, (3.4) and (4.5) yield

$$\frac{\phi_i}{2}\|x^N - \bar{x}\|^2 \leq C_0 + D_N \quad \text{for} \quad D_N := \sum_{i=0}^{N-1} \tau_i\phi_i\delta_{i+1} \quad (N \geq 1).\tag{4.14}$$

In the case $K\bar{x} = 0$, we have $\delta_{i+1} = 0$. As in the proof of Theorem 4.3, by a standard analysis [15, 13], it follows from (4.13) that $\phi_N \geq CN^2$ for some $C > 0$. We therefore get from (4.14) the claimed $O(1/N^2)$ rate.
Consider then the case $K\tilde{x} \neq 0$. We estimate
\begin{equation}
D_N = \sum_{i=0}^{N-1} \tau_i \phi_i \mu_{i+1} \leq C_1 \sum_{i=0}^{N-1} \phi_i \mu_{i+1}
\end{equation}

By Lines 4 and 5 of the algorithm, $\phi_N \geq \phi_0 + 2\gamma \|K\|^{-2} \sum_{i=0}^{N-1} \phi_i \mu_{i+1}$. Using these estimates in (4.14), it follows that $\|x^N - \tilde{x}\|$ is bounded. If $\tilde{\ell}_{i+1} \searrow 0$, (4.12) and (4.13) shows that also $\tau_i \searrow 0$. Restarting our analysis from a later iteration, we can therefore make $C_1 > 0$ arbitrarily small. Consequently, for any $\epsilon > 0$, for large enough $N$ holds $\|x^N - \tilde{x}\| \leq \epsilon$. Since $\ell > 0$, this is in contradiction to $\tilde{\ell}_{i+1} \searrow 0$. We may therefore assume that $\tilde{\ell}_{i+1} \geq \epsilon$ for some $\epsilon > 0$, at least for large $i$. Since our claims are asymptotical, we may without loss of generality assume this for all $i$.

From (4.12), we now estimate $\tau_i \geq \tilde{\epsilon} \lambda_{\text{min}}(a)/\|K\|^2 =: \tau_* > 0$. From Line 5 consequently
\begin{equation}
\phi_{i+1} \geq \phi_i (1 + 2\gamma \tau_*).
\end{equation}

This shows that $\phi_N \geq \Theta((1 + \gamma \tau_*)^N)$ grows exponentially, predicting (4.14) to yield linear rates from if we can control the penalty $D_N$.

Continuing from (4.15), by Hölder’s inequality, since the conjugate exponent of $1/(1-p)$ is $1/p$, for any $p \in (0, 1)$ holds
\[
D_N \leq C_1 \theta \sum_{i=0}^{N-1} \phi_i^{1-p} \phi_i^{p-1/2} \leq C_1 \theta \left( \sum_{i=0}^{N-1} \phi_i \right)^{1-p} \left( \sum_{i=0}^{N-1} \phi_i^{1/(2p)} \right)^{p}.
\]

By (4.16), 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
\[
\phi_N - \phi_0 = 2\gamma \sum_{i=0}^{N-1} \phi_i \tau_i \geq 2\gamma \tau_* \sum_{i=0}^{N-1} \phi_i.
\]

For some constant $C' > 0$ we therefore get
\[
D_N \leq C' (\phi_N - \phi_0)^{1-p} \leq C' \phi_N^{1-p}.
\]

Minding (4.14) and (4.16), this shows the claimed linear rate. \hfill \Box

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
\begin{equation}
\min_{x \in \mathbb{R}^{n_2}} \frac{1}{2} \|z - x\|^2 + \alpha R(x),
\end{equation}

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 $R(x) = \|Dx\|_{L^1}$, 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 \( \|g\|_{2,1} := \sum_{i=1}^{n_1n_2} \sqrt{g_{i,1} + g_{i,2}} \) for \( g = (g_{i,1}, g_{i,2}) \in \mathbb{R}^{2n_1n_2} \).

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

\[
\min_{x \in \mathbb{R}^{n_1n_2}} \max_{y \in \mathcal{J}} \frac{1}{2} \|z - u\|^2_2 + \langle Kx, y \rangle - \delta_{\mathcal{K}^\perp} A^{-1} b(y),
\]

where for \( H^1 \) denoising

\[
\mathcal{J} = \mathbb{R}_{+} \times [2n_1n_2], \quad Kx = (0, Dx), \quad Ay = y_0, \quad b = \alpha,
\]

and for TV denoising, for \( i = 1, \ldots, n_1n_2, \)

\[
\mathcal{J} = (\mathbb{R}_{+} \times [2n_1n_2]), \quad [Kx]_i = (0, [Dx]_{i,1}, [Dx]_{i,2}), \quad Ay = ((y_1)_0, \ldots, (y_{n_1n_2})_0), \quad b = (\alpha, \ldots, \alpha).
\]

In the latter case, Line 6 of Algorithm 4.1 splits into \( n_1n_2 \) 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 [38] 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, \( \mathcal{K} = \mathcal{K}^{n_1n_2}_{\text{soc}} \) 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 (4.4b), a straightforward approach would yield linear convergence when \( \min_i \| [K^\top]_i \| > 0 \). From the structure of the TV denoising problem, it is however easy to see that it can often happen that \( [K^\top]_i = 0 \). 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 [39]. 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. 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 [40].

\(^1\)At the time of writing online at http://roku.sgraphics/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.

| Method   | low resolution | high resolution |
|----------|----------------|-----------------|
|          | gap \leq -50dB | Iter | Time | gap \leq -50dB | Iter | Time | gap \leq -50dB | Iter | Time |
| PDHGM    | 4              | 0.018          | 30   | 0.098 | 27   | 0.085 | 4              | 0.138          | 34   | 1.428 | 13   | 0.528 |
| PEDI     | 16             | 0.043          | 270  | 0.738 | 280  | 0.758 | 86             | 3.785          | -    | -    | 400  | 17.765 |
| Dual FB  | 12             | 0.038          | 9    | 0.025 | 9    | 0.028 | 14             | 0.625          | 21   | 0.968 | 12   | 0.538 |

We compared our algorithm (denoted PEDI, Primal Euclidean–Dual Interior) to the accelerated Chambolle–Pock method (PDHGM, Primal–Dual Hybrid Gradient method, Modified [41]) 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 $L := \sqrt{F/\gamma} \geq \|K\|$ [42]. For the PDHGM, we took $r_0 = 0.52/L$ and $\sigma_0 = 1.9/L$, using the strong convexity parameter $\gamma = 0.9 < 1$ for acceleration. For our method, we took $\zeta = 0.9/h_0^2$ and $\theta = 1/\zeta$, keeping $r_0$ and $\gamma$ 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 $\hat{x}$ 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 $\hat{x}$ in decibels $10 \log_{10} \|x^i - \hat{x}\|^2/\|\hat{x}\|^2$, the primal objective value $\text{val}(x) := G(x) + F(Kx)$ relative to the target $10 \log_{10} \left(\text{val}(\hat{x})/\text{val}(x)\right)^2/\|\hat{x}\|^2$, as well as the duality gap $10 \log_{10} \left(\text{gap}_x/\text{gap}_y^2\right)$, 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 $K^*y^i = \nabla G(x^i) = x^i - z$.

### 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. 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 [16, 17, 18, 19]. Moreover, we are puzzled by what, if anything, makes the second-order cone special?
Figure 1: TV denoising convergence behaviour: high and low resolution images; gap, distance to target solution, and primal objective value in decibels.

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

| Method  | iter time | gap ≤ −150dB | iter time | gtt ≤ −100dB | iter time | val ≤ −100dB | iter time |
|--------|-----------|---------------|-----------|--------------|-----------|--------------|-----------|
| PDHGM  | 360 0.918 | −             | −         | 180 0.468    | −         | 380 11.485   | −         |
| PEDI   | 120 0.318 | 87 0.228      | 54 0.148  | 51 1.698     | 39 1.288   | 24 0.788     |           |
| Dual FB| 44 0.118  | 43 0.118      | 22 0.058  | 17 0.748     | 18 0.788   | 8 0.328      |           |

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

All data and source codes will be deposited for readers to access when the final accepted version of the manuscript is submitted.

REFERENCES

[1] S. H. Schmieta and F. Alizadeh, Extension of primal-dual interior point algorithms to symmetric cones, Mathematical Programming 96 (2003), 409–438, doi:10.1007/s10107-003-0380-2.

[2] L. Faybusovich, Euclidean jordan algebras and interior-point algorithms, Positivity 1 (1997), 331–357, doi:10.1023/A:1009701824047.

[3] R. Monteiro, Polynomial Convergence of Primal-Dual Algorithms for Semidefinite Programming Based on the Monteiro and Zhang Family of Directions, SIAM Journal on Optimization 8 (1998), 797–812, doi:10.1137/S1052623496308618.

[4] Y. E. Nesterov and M. J. Todd, Self-scaled barriers and interior-point methods for convex programming, Mathematics of Operations Research 22 (1997), 1–42, doi:10.1137/S1052623495290209.

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

[6] 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 Limited2008, 95–110.

[7] A. Beck and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM Journal on Imaging Sciences 2 (2009), 183–202, doi:10.1137/080716542.
[8] 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, doi:10.1088/0266-5611/27/12/125007.

[9] 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-Holland1983, 299–331.

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

[11] B. He and X. Yuan, *Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective*, SIAM Journal on Imaging Sciences 5 (2012), 119–149, doi:10.1137/100814494.

[12] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM Journal on Optimization 14 (1976), 877–898, doi:10.1137/0314056.

[13] T. Valkonen and T. Pock, *Acceleration of the PDHGM on partially strongly convex functions*, Journal of Mathematical Imaging and Vision (2016), doi:10.1007/s10851-016-0692-2, to appear, arXiv:1511.06566. URL http://iki.fi/tuomov/mathematics/cpaccel.pdf

[14] T. Valkonen, *Testing and non-linear preconditioning of the proximal point method* (2017), submitted, arXiv:1703.05705. URL http://iki.fi/tuomov/mathematics/proxtest.pdf

[15] T. Valkonen, *Block-proximal methods with spatially adapted acceleration* (2016), submitted, arXiv:1609.07373. URL http://iki.fi/tuomov/mathematics/blockcp.pdf

[16] 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, doi:10.12785/amis/070624.

[17] 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, doi:10.1007/s10589-008-9227-0.

[18] J. López and E. A. P. Quiroz, *Construction of proximal distances over symmetric cones*, Optimization (2017), doi:10.1080/02331934.2016.1277998, published online.

[19] 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, doi:10.7151/dmdico1077.

[20] G. H.-G. Chen and R. T. Rockafellar, *Convergence rates in forward–backward splitting*, SIAM Journal on Optimization 7 (1997), 421–444, doi:10.1137/S0895907795290179.
[21] D. Boley, *Local linear convergence of the alternating direction method of multipliers on quadratic or linear programs*, SIAM Journal on Optimization **23** (2013), 2183–2207, doi:10.1137/120878951.

[22] D. Han and X. Yuan, *Local linear convergence of the alternating direction method of multipliers for quadratic programs*, SIAM Journal on Numerical Analysis **51** (2013), 3446–3457, doi:10.1137/120886753.

[23] M. Hong and Z.-Q. Luo, *On the linear convergence of the alternating direction method of multipliers*, Mathematical Programming **162** (2017), 165–199, doi:10.1007/s10107-016-1034-2.

[24] J. Liang, J. Fadili and G. Peyré, *Local linear convergence of forward–backward under partial smoothness*, Advances in Neural Information Processing Systems **27** (2014), 1970–1978. URL http://papers.nips.cc/paper/5260-local-linear-convergence-of-forward-backward-under-partial-smoothness.

[25] K. Bredies and D. A. Lorenz, *Linear convergence of iterative soft-thresholding*, Journal of Fourier Analysis and Applications **14** (2008), 813–837, doi:10.1007/s00041-008-9041-1.

[26] R. T. Rockafellar, *Convex Analysis*, Princeton University Press 1972.

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

[28] M. Koecher, *The Minnesota notes on Jordan algebras and their applications*, Lecture Notes in Mathematics, volume 1710, Springer-Verlag, Berlin 1999.

[29] L. Faybusovich, *Linear systems in Jordan algebras and primal-dual interior-point algorithms*, Journal of Computational and Applied Mathematics **86** (1997), 149–175, doi:10.1016/S0377-0427(97)00153-2.

[30] T. Valkonen, *Diff-convex combinations of Euclidean distances: a search for optima*, number 99 in Jyväskylä Studies in Computing, University of Jyväskylä 2008, Ph.D Thesis. URL http://iki.fi/tuomov/mathematics/thesis.pdf

[31] S. J. Wright and D. Orban, *Properties of the log-barrier function on degenerate nonlinear programs*, Mathematics of Operations Research **27** (2002), 585–613, doi:10.1287/moor.27.3.585.312.

[32] H. Ramirez and D. Sossa, *On the central paths in symmetric cone programming*, Journal of Optimization Theory and Applications (2016), 1–20, doi:10.1007/s10957-016-0989-8.

[33] 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*, Journal of Optimization Theory and Applications **139** (2008), 227, doi:10.1007/s10957-008-9422-2.

[34] 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, doi:10.1023/A:1018339901042.
[35] R. S. Burachik, L. M. G. Drummond and S. Scheimberg, On some properties and an application of the logarithmic barrier method, Mathematical Programming 111 (2008), 95–112, doi:10.1007/s10107-006-0075-3.

[36] M. Halická, E. de Klerk and C. Roos, Limiting behavior of the central path in semidefinite optimization, Optimization Methods and Software 20 (2005), 99–113, doi:10.1080/1055678041000172718.

[37] F. Alizadeh and D. Goldfarb, Second-order cone programming, Mathematical Programming 95 (2003), 3–51, doi:10.1007/s10107-002-0339-5.

[38] L. Rudin, S. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D 60 (1992), 259–268.

[39] W. Ring, Structural properties of solutions to total variation regularization problems, ESAIM: Mathematical Modelling and Numerical Analysis 34 (2000), 799–810, doi:10.1051/m2an:2000104.

[40] J. C. de Los Reyes, C.-B. Schönlieb and T. Valkonen, Bilevel parameter learning for higher-order total variation regularisation models, Journal of Mathematical Imaging and Vision 57 (2017), 1–25, doi:10.1007/s10851-016-0662-8, arXiv:1508.07243. URL http://iki.fi/tuomov/mathematics/tgv_learn.pdf

[41] 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 Journal on Imaging Sciences 3 (2010), 1015–1046, doi:10.1137/09076934X.

[42] A. Chambolle, An algorithm for mean curvature motion, Interfaces and Free Boundaries 6 (2004), 195.