CORC Technical Report TR-2004-07

An active set method for single-cone second-order cone programs

E. Erdoğan[*] G. Iyengar[†]

June 25, 2005

Abstract

We develop an active set method for solving second-order cone programs that may have an arbitrary number of linear constraints but are restricted to have only one second-order cone constraint. Problems of this form arise in the context of robust optimization and trust region methods. The proposed active set method exploits the fact that a second-order cone program with only one second-order cone constraint and no inequality constraints can be solved in closed form.

1 Introduction

In this paper we are concerned with the following special case of a second-order cone program (SOCP).

$$\begin{array}{rrcl} \min & \mathbf{f}^T\mathbf{x}, & & \\ \text{subject to} & \mathbf{H}\mathbf{x} & = & \mathbf{g}, \\ & \mathbf{E}\mathbf{x} & \geq & \mathbf{0}, \\ & \mathbf{D}\mathbf{x} & \succeq & \mathbf{0}, \end{array} \tag{1}$$

where $\mathbf{x} \in \mathbf{R}^n$, $\mathbf{f} \in \mathbf{R}^n$, $\mathbf{H} \in \mathbf{R}^{m\times n}$, $\mathbf{g} \in \mathbf{R}^m$, $\mathbf{E} \in \mathbf{R}^{l\times n}$, $\mathbf{D} \in \mathbf{R}^{p\times n}$, and $\succeq$ denotes the partial order with respect to the standard conic quadratic cone $\mathcal{Q} = \{(z_0, \bar{\mathbf{z}})^T \in \mathbf{R}^p : z_0 \geq \sqrt{\bar{\mathbf{z}}^T\bar{\mathbf{z}}}\} \subset \mathbf{R}^p$. We shall call the optimization problem (1) a single-cone SOCP since it is restricted to have only *one* second-order cone constraint.

Our interest in single-cone SOCPs stems from the fact that they arise as the robust counterpart of uncertain linear programs (LPs). Many decision problems in engineering and operations research can be formulated as LPs of the form

$$\begin{array}{rl} \min & \mathbf{c}^T\mathbf{x}, \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq \mathbf{0}. \end{array}$$

Solution techniques for LPs compute a solution assuming that the parameters $(\mathbf{A}, \mathbf{b}, \mathbf{c})$ are known exactly. However, in practice, these parameters are typically the result of some measurement or

[*]IEOR Department, Columbia University, New York, New York 10027. `ee168@columbia.edu`. Research partially supported by NSF grant CCR-00-09972

[†]IEOR Department, Columbia University, New York, New York 10027. `gi10@columbia.edu`. Research partially supported by NSF grants CCR-00-09972, DMS-01-04282 and ONR grant N000140310514.

estimation process and are, therefore, never certain. LPs whose parameters are not known exactly are called *uncertain* LPs. Several strategies have been proposed to address parameter uncertainty in optimization problems. One approach is to solve the LP for a nominal set of parameters $(\mathbf{A}_0, \mathbf{b}_0, \mathbf{c}_0)$ and then analyze the quality of the solution using a post-optimization tool such as sensitivity analysis [5]. This approach is particularly attractive when the uncertainty is "small" in an appropriate sense. In the stochastic programming approach, the uncertainty is assumed to be random with a known distribution, and samples from this known distribution are used to compute good solutions [13]. However, identifying appropriate distributions for the parameters is not straightforward. Also, as the dimension of the problem grows, the complexity of the stochastic program quickly becomes prohibitive. Recently Ben-Tal and Nemirovski [2, 3, 4] proposed *robust* optimization as another approach to address data uncertainty. In this approach, the uncertain parameters $(\mathbf{A}, \mathbf{b}, \mathbf{c})$ are assumed to belong to a bounded uncertainty set $\mathcal{U}$ and the goal of the *robust* counterpart is to compute a minimax optimal solution. The results in [2, 3, 4] establish that, when $\mathcal{U}$ satisfies some regularity properties, the robust counterpart can be reformulated as an SOCP and, therefore, can be solved efficiently both in theory [12] and in practice [15].

The robust counterpart of an uncertain LP where the parameters $(\mathbf{A}, \mathbf{b})$ are completely known and the uncertain cost vector $\mathbf{c}$ belongs to an ellipsoidal uncertainty set can be reformulated as a single-cone SOCP [3] (see also Section 5). In many engineering applications the constraints in the LP are given by design considerations and are, therefore, fixed and certain. For example, in routing problems, arising in the context of road or air traffic control and communication networks, the capacities are determined at the network design stage; therefore, the constraints in the problem, namely the flow balance equations and capacity constraints, are completely known when the routing problem is to be solved. However, the "cost" of an arc is typically a non-linear function of the capacity and flows in the network, and measuring this cost is often difficult and expensive [7]. The "cost" of a feasible flow can often be modeled as an uncertain linear function with an ellipsoidal uncertainty set by using the so-called delta method [11]. Production planning is another natural example where the constraints are fixed and only the costs are uncertain. Here the vector $\mathbf{c}$ denotes the vector of future expected market prices for the various raw materials, and is, typically, estimated from historical prices via linear regression. Since the confidence regions associated with linear regression are ellipsoidal [9, 11], the resulting robust counterpart is a single-cone SOCP.

From the equivalence

$$\|\mathbf{Pu}\| \leq 1 \quad \Leftrightarrow \quad u_0 = 1, \ \begin{bmatrix} u_0 \\ \mathbf{Pu} \end{bmatrix} \succeq \mathbf{0},$$

it follows that the trust region problem is a special case of a single-cone SOCP. This provides another motivation for developing active set methods for single cone SOCPs. Note that formulating the trust region problem as a single-cone SOCP allows one to consider hyperbolic and parabolic trust regions.

Alizadeh and Goldfarb [1] showed that, under appropriate regularity conditions, the optimal solution of a single-cone SOCP with no inequality constraints can be computed in closed form. We use this result to explicitly compute the value of the Lagrangian obtained by dualizing the non-negativity constraints $\mathbf{Ex} \geq \mathbf{0}$. We compute the optimal dual multipliers using an active set method, and then recover an optimal primal solution using the results in [1]. The formulation of the appropriate dual problem is discussed in Section 2, the active set method is detailed in Section 3, and Section 4 details how to recover an optimal solution of (1).

Clearly, any algorithm for solving general SOCPs can be used to solve a single-cone SOCP.

All known codes for solving SOCPs, e.g. SeDuMi [15] and MOSEK®, are based on interior point methods. Our efforts in developing an active set method for the single-cone SOCP were motivated, in part, by the observation that active set methods are known to solve convex quadratic programs efficiently. Our goal was to investigate whether a simple active set algorithm outperforms general purpose SOCP codes at least for certain problem classes. We report the results of our computational experiments in Section 5.

2 Formulation of the Lagrangian dual

In this section we formulate a Lagrangian dual for the single-cone SOCP (1). We assume that $\mathbf{H} \in \mathbf{R}^{m\times n}$ has full row rank and the following constraint qualification holds.

Assumption 1 *There exists $\bar{\mathbf{x}} \in \mathbf{R}^n$ such that $\mathbf{H}\bar{\mathbf{x}} = \mathbf{g}$, $\mathbf{E}\bar{\mathbf{x}} \geq \mathbf{0}$, and $\mathbf{D}\bar{\mathbf{x}} \succ \mathbf{0}$.*

The active set algorithm proposed in this paper exploits the following result from [1].

Lemma 1 *Suppose the pair of primal-dual SOCPs*

$$\begin{array}{rlrl} \min & \mathbf{c}^T\mathbf{x}, & \max & \mathbf{b}^T\mathbf{y}, \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b}, & \text{subject to} & \mathbf{A}^T\mathbf{y} + \mathbf{z} = \mathbf{c}, \\ & \mathbf{x} \succeq \mathbf{0}, & & \mathbf{z} \succeq \mathbf{0}, \end{array} \tag{2}$$

are both *strictly feasible. Then the optimal solution of the primal SOCP is given by*

$$\mathbf{x}^* = \left(\sqrt{\frac{-\mathbf{b}^T(\mathbf{A}\mathbf{R}\mathbf{A}^T)^{-1}\mathbf{b}}{\mathbf{c}^T\mathbf{P}_R\mathbf{c}}}\right)\mathbf{P}_R\mathbf{c} + \mathbf{R}\mathbf{A}^T(\mathbf{A}\mathbf{R}\mathbf{A}^T)^{-1}\mathbf{b}, \tag{3}$$

where

$$\mathbf{R} = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & -\mathbf{I} \end{bmatrix}, \quad \mathbf{P}_R = \mathbf{R} - \mathbf{R}\mathbf{A}^T(\mathbf{A}\mathbf{R}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{R},$$

and $\mathbf{I}$ denotes an identity matrix.

Remark 1 *In Lemma 1 we have implicitly assumed that $\mathbf{A}\mathbf{R}\mathbf{A}^T$ is non-singular. A similar result holds when $\mathbf{A}\mathbf{R}\mathbf{A}^T$ is singular. See [1] for details.*

In order to reformulate (1) into a form similar to the primal SOCP in (2), we dualize the non-negativity constraints to obtain the Lagrangian

$$\begin{array}{rcrl} q(\boldsymbol{\lambda}) & \equiv & \min & (\mathbf{f} - \mathbf{E}^T\boldsymbol{\lambda})^T\mathbf{x}, \\ & & \text{subject to} & \mathbf{H}\mathbf{x} = \mathbf{g}, \\ & & & \mathbf{D}\mathbf{x} \succeq \mathbf{0}, \end{array} \tag{4}$$

where $\boldsymbol{\lambda} \in \mathbf{R}^l_+$ denotes the Lagrange multipliers for the inequality constraints. Note that the result in [1] applies only when the primal and the dual SOCPs are *both* strictly feasible. For SOCPs, feasibility is a subtle issue, e.g. the fact that the primal is bounded does not imply that the dual is feasible [4]; therefore, one has to be careful in applying the results in [1]. Elementary properties of convex duality [6] implies the following claim.

Claim 1 *Let $q(\boldsymbol{\lambda})$ denote the Lagrangian defined in (4). Let $\mathbf{x}^*$ and v^* denote, respectively, any optimal solution and the optimum value of (1). Then*

(a) $v^* = \max\left\{q(\boldsymbol{\lambda}) : \boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\lambda} \in \mathcal{D}_q\right\}$*, where* $\mathcal{D}_q = \left\{\boldsymbol{\lambda} : q(\boldsymbol{\lambda}) > -\infty\right\}$*,*

(b) $\mathbf{x}^* \in \operatorname{argmin}\left\{(\mathbf{f} - \mathbf{E}^T\boldsymbol{\lambda}^*)^T\mathbf{x} : \mathbf{H}\mathbf{x} = \mathbf{g}, \mathbf{D}\mathbf{x} \succeq \mathbf{0}\right\}$*, where* $\boldsymbol{\lambda}^* \in \operatorname{argmax}\left\{q(\boldsymbol{\lambda}) : \boldsymbol{\lambda} \geq 0, \boldsymbol{\lambda} \in \mathcal{D}_q\right\}$*.*

Thus, an optimal solution to (1) can be obtained by first computing an optimal multiplier $\boldsymbol{\lambda}^* \in \operatorname{argmax}\left\{q(\boldsymbol{\lambda}) : \boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\lambda} \in \mathcal{D}_q\right\}$, and then computing an optimal $\mathbf{x}^*$ by solving $q(\boldsymbol{\lambda}^*)$. In Section 2.1 we show how to compute the value of the Lagrange dual function $q(\boldsymbol{\lambda})$ for a fixed value of $\boldsymbol{\lambda} \in \mathcal{D}_q$, in Section 3 we describe an active set algorithm to solve for the optimal dual multipliers $\boldsymbol{\lambda}^*$, and in Section 4 we show how to recover the optimal primal solution $\mathbf{x}^*$.

2.1 Computing the Lagrangian $q(\boldsymbol{\lambda})$

Claim 1 allows us to restrict ourselves to $\boldsymbol{\lambda} \geq \mathbf{0}$ such that $q(\boldsymbol{\lambda}) > -\infty$, i.e. $\boldsymbol{\lambda} \in \mathcal{D}_q \cap \mathbf{R}^l_+$, without any loss in generality. Fix $\mathbf{y} \succeq \mathbf{0}$ and consider the optimization problem in $\mathbf{x}$

$$q(\boldsymbol{\lambda}, \mathbf{y}) \equiv \begin{array}{rl} \min & (\mathbf{f} - \mathbf{E}^T\boldsymbol{\lambda})^T\mathbf{x}, \\ \text{subject to} & \mathbf{H}\mathbf{x} = \mathbf{g}, \\ & \mathbf{D}\mathbf{x} = \mathbf{y}. \end{array} \tag{5}$$

Note that $q(\boldsymbol{\lambda}) > -\infty$ if, and only if, $q(\boldsymbol{\lambda}, \mathbf{y}) > -\infty$ for all $\mathbf{y} \succeq \mathbf{0}$. Since $\mathbf{H}$ has full row rank, $\mathbf{H}\mathbf{x} = \mathbf{g}$ if, and only if, $\mathbf{x} = \mathbf{x}_0 + \mathbf{B}\mathbf{z}$, where $\mathbf{x}_0 = \mathbf{H}^T(\mathbf{H}\mathbf{H}^T)^{-1}\mathbf{g} \in \mathbf{R}^n$, $\mathbf{B} \in \mathbf{R}^{n\times(n-m)}$ is any orthonormal basis for the nullspace $\mathcal{N}(\mathbf{H})$ of $\mathbf{H}$, and $\mathbf{z} \in \mathbf{R}^{n-m}$. Thus, we have that

$$q(\boldsymbol{\lambda}, \mathbf{y}) = (\mathbf{f} - \mathbf{E}^T\boldsymbol{\lambda})^T\mathbf{x}_0 + \begin{array}{rl} \min & (\mathbf{f} - \mathbf{E}^T\boldsymbol{\lambda})^T\mathbf{B}\mathbf{z}, \\ \text{subject to} & \mathbf{D}\mathbf{B}\mathbf{z} = \mathbf{y} - \mathbf{D}\mathbf{x}_0. \end{array} \tag{6}$$

Since $\mathbf{D}\mathbf{B} \in \mathbf{R}^{p\times(n-m)}$ the following three cases exhaust all possibilities.

(i) $\mathbf{rank}(\mathbf{D}\mathbf{B}) = r < \min\{p, n-m\}$: In this case, the singular value decomposition (SVD) of $\mathbf{D}\mathbf{B}$ has the following form

$$\mathbf{D}\mathbf{B} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T = \begin{bmatrix}\mathbf{U}_0 & \mathbf{U}_1\end{bmatrix}\begin{bmatrix}\boldsymbol{\Sigma}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{V}_0^T \\ \mathbf{V}_1^T\end{bmatrix} = \mathbf{U}_0\boldsymbol{\Sigma}_0\mathbf{V}_0^T,$$

where $\mathbf{U}_0 \in \mathbf{R}^{p\times r}$, $\mathbf{U}_1 \in \mathbf{R}^{p\times(p-r)}$, $\mathbf{V}_0 \in \mathbf{R}^{(n-m)\times r}$, $\mathbf{V}_1 \in \mathbf{R}^{(n-m)\times(n-m-r)}$, and $\boldsymbol{\Sigma}_0 \in \mathbf{R}^{r\times r}$ is a diagonal matrix. Consequently, $\mathbf{U}_1^T(\mathbf{y} - \mathbf{D}\mathbf{x}_0) = \mathbf{0}$, and $\mathbf{z} = \mathbf{V}_0\boldsymbol{\Sigma}_0^{-1}\mathbf{U}_0^T(\mathbf{y} - \mathbf{D}\mathbf{x}_0) + \mathbf{V}_1\mathbf{t}$, where $\mathbf{t} \in \mathbf{R}^{n-m-r}$. Thus,

$$q(\boldsymbol{\lambda}, \mathbf{y}) = (\mathbf{f} - \mathbf{E}^T\boldsymbol{\lambda})^T\mathbf{z}_0 + \boldsymbol{\xi}^T\mathbf{y} + \min_{\mathbf{t}}\left\{(\mathbf{f} - \mathbf{E}^T\boldsymbol{\lambda})^T\mathbf{B}\mathbf{V}_1\mathbf{t}\right\}, \tag{7}$$

where $\mathbf{z}_0 = (\mathbf{I} - \mathbf{B}\mathbf{V}_0\boldsymbol{\Sigma}_0^{-1}\mathbf{U}_0^T\mathbf{D})\mathbf{x}_0$, and $\boldsymbol{\xi} = \mathbf{U}_0\boldsymbol{\Sigma}_0^{-1}\mathbf{V}_0^T\mathbf{B}^T(\mathbf{f} - \mathbf{E}^T\boldsymbol{\lambda})$. From (7), we have

$$\mathbf{q}(\boldsymbol{\lambda}, \mathbf{y}) > -\infty \quad \Leftrightarrow \quad \mathbf{V}_1^T\mathbf{B}^T(\mathbf{f} - \mathbf{E}^T\boldsymbol{\lambda}) = \mathbf{0}, \tag{8}$$

and in that case

$$q(\boldsymbol{\lambda}) = (\mathbf{f} - \mathbf{E}^T\boldsymbol{\lambda})^T\mathbf{z}_0 + \bar{q}(\boldsymbol{\xi}), \tag{9}$$

where

$$\begin{aligned} \bar{q}(\boldsymbol{\xi}) \quad = \quad & \min \quad \boldsymbol{\xi}^T \mathbf{y}, \\ & \text{subject to} \quad \mathbf{A}\mathbf{y} = \mathbf{b}, \\ & \qquad \mathbf{y} \succeq \mathbf{0}, \end{aligned} \tag{10}$$

and

$$\begin{aligned} \mathbf{z}_0 &= (\mathbf{I} - \mathbf{B}\mathbf{V}_0\boldsymbol{\Sigma}_0^{-1}\mathbf{U}_0^T\mathbf{D})\mathbf{H}^T(\mathbf{H}\mathbf{H}^T)^{-1}\mathbf{g}, \\ \boldsymbol{\xi} &= \mathbf{U}_0\boldsymbol{\Sigma}_0^{-1}\mathbf{V}_0^T\mathbf{B}^T(\mathbf{f} - \mathbf{E}^T\boldsymbol{\lambda}), \\ \mathbf{b} &= \mathbf{U}_1^T\mathbf{D}\mathbf{H}^T(\mathbf{H}\mathbf{H}^T)^{-1}\mathbf{g}, \\ \mathbf{A} &= \mathbf{U}_1^T. \end{aligned} \tag{11}$$

(ii) $\mathbf{rank}(\mathbf{DB}) = n - m < p$: In this case, we have

$$\mathbf{DB} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T = \begin{bmatrix}\mathbf{U}_0 & \mathbf{U}_1\end{bmatrix}\begin{bmatrix}\boldsymbol{\Sigma}_0 \\ \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{V}_0^T\end{bmatrix} = \mathbf{U}_0\boldsymbol{\Sigma}_0\mathbf{V}_0^T,$$

where $\mathbf{U}_0 \in \mathbf{R}^{p\times(n-m)}$, $\mathbf{U}_1 \in \mathbf{R}^{p\times(p-n+m)}$, $\mathbf{V}_0 \in \mathbf{R}^{(n-m)\times(n-m)}$, and $\boldsymbol{\Sigma}_0 \in \mathbf{R}^{(n-m)\times(n-m)}$ is a diagonal matrix. Thus, (6) is feasible if, and only if,

$$\mathbf{U}_1^T(\mathbf{y} - \mathbf{D}\mathbf{x}_0) = \mathbf{0}. \tag{12}$$

Since $\mathbf{V}_0$ has full rank, it follows that when (12) holds we have $\mathbf{z} = \mathbf{V}_0\boldsymbol{\Sigma}_0^{-1}\mathbf{U}_0^T(\mathbf{y} - \mathbf{D}\mathbf{x}_0)$. Consequently,

$$q(\boldsymbol{\lambda}, \mathbf{y}) = (\mathbf{f} - \mathbf{E}^T\boldsymbol{\lambda})^T\mathbf{z}_0 + \boldsymbol{\xi}^T\mathbf{y}, \tag{13}$$

where $\mathbf{z}_0 = (\mathbf{I} - \mathbf{B}\mathbf{V}_0\boldsymbol{\Sigma}_0^{-1}\mathbf{U}_0^T\mathbf{D})\mathbf{x}_0$, and $\boldsymbol{\xi} = \mathbf{U}_0\boldsymbol{\Sigma}_0^{-1}\mathbf{V}_0^T\mathbf{B}^T(\mathbf{f} - \mathbf{E}^T\boldsymbol{\lambda})$. Thus, (9), (10) and (11) remain valid in this case.

(iii) $\mathbf{rank}(\mathbf{DB}) = p < n - m$: The SVD of $\mathbf{DB}$ is given by

$$\mathbf{DB} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T = \begin{bmatrix}\mathbf{U}_0\end{bmatrix}\begin{bmatrix}\boldsymbol{\Sigma}_0 & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{V}_0^T \\ \mathbf{V}_1^T\end{bmatrix} = \mathbf{U}_0\boldsymbol{\Sigma}_0\mathbf{V}_0^T,$$

where $\mathbf{U}_0 \in \mathbf{R}^{p\times p}$, $\mathbf{V}_0 \in \mathbf{R}^{(n-m)\times p}$, $\mathbf{V}_1 \in \mathbf{R}^{(n-m)\times(n-m-p)}$, and $\boldsymbol{\Sigma}_0 \in \mathbf{R}^{p\times p}$ is a diagonal matrix. Since $\mathbf{U}_0$ has full rank, (6) is always feasible. Thus, $\mathbf{z} = \mathbf{V}_0\boldsymbol{\Sigma}_0^{-1}\mathbf{U}_0^T(\mathbf{y} - \mathbf{D}\mathbf{x}_0) + \mathbf{V}_1\mathbf{t}$, where $\mathbf{t} \in \mathbf{R}^{n-m-p}$. Consequently,

$$q(\boldsymbol{\lambda}, \mathbf{y}) = (\mathbf{f} - \mathbf{E}^T\boldsymbol{\lambda})^T\mathbf{z}_0 + \boldsymbol{\xi}^T\mathbf{y} + \min_{\mathbf{t}}\left\{(\mathbf{f} - \mathbf{E}^T\boldsymbol{\lambda})^T\mathbf{B}\mathbf{V}_1\mathbf{t}\right\}, \tag{14}$$

where $\mathbf{z}_0 = (\mathbf{I} - \mathbf{B}\mathbf{V}_0\boldsymbol{\Sigma}_0^{-1}\mathbf{U}_0^T\mathbf{D})\mathbf{x}_0$, and $\boldsymbol{\xi} = \mathbf{U}_0\boldsymbol{\Sigma}_0^{-1}\mathbf{V}_0^T\mathbf{B}^T(\mathbf{f} - \mathbf{E}^T\boldsymbol{\lambda})$. From (14), we have

$$\mathbf{q}(\boldsymbol{\lambda}, \mathbf{y}) > -\infty \quad \Leftrightarrow \quad \mathbf{V}_1^T\mathbf{B}^T(\mathbf{f} - \mathbf{E}^T\boldsymbol{\lambda}) = \mathbf{0}. \tag{15}$$

Thus,

$$q(\boldsymbol{\lambda}) = (\mathbf{f} - \mathbf{E}^T\boldsymbol{\lambda})^T\mathbf{z}_0 + \widehat{q}(\boldsymbol{\xi}), \tag{16}$$

where

$$\begin{aligned} \widehat{q}(\boldsymbol{\xi}) \quad = \quad & \min \quad \boldsymbol{\xi}^T\mathbf{y}, \\ & \text{subject to} \quad \mathbf{y} \succeq \mathbf{0}, \end{aligned} \tag{17}$$

and

$$\begin{aligned} \mathbf{z}_0 &= (\mathbf{I} - \mathbf{B}\mathbf{V}_0\boldsymbol{\Sigma}_0^{-1}\mathbf{U}_0^T\mathbf{D})\mathbf{H}^T(\mathbf{H}\mathbf{H}^T)^{-1}\mathbf{g}, \\ \boldsymbol{\xi} &= \mathbf{U}_0\boldsymbol{\Sigma}_0^{-1}\mathbf{V}_0^T\mathbf{B}^T(\mathbf{f} - \mathbf{E}^T\boldsymbol{\lambda}). \end{aligned} \tag{18}$$

Since the structures of the optimization problems (10) and (17), although similar, are not identical; the corresponding active set methods are also similar, but not identical. In the paper we focus on developing an active set method for optimizing the Lagrangian defined in (9). The active set method for optimizing the Lagrangian defined in (16) is in Appendix B.

Lemma 2 *Let $\bar{q} : \mathbf{R}^p \mapsto \mathbf{R}$ denote the function defined in (10). Then the domain $\mathcal{D}_{\bar{q}} = \{\boldsymbol{\xi} : \ \bar{q}(\boldsymbol{\xi}) > -\infty\}$ is given by*

$$\mathcal{D}_{\bar{q}} = \begin{cases} \mathbf{R}^p, & \gamma < 0, \\ \left\{\boldsymbol{\xi} : \mathbf{e}^T\mathbf{P}\boldsymbol{\xi} \geq 0, (\mathbf{e}^T\mathbf{P}\boldsymbol{\xi})^2 - \gamma\|\mathbf{P}\boldsymbol{\xi}\|^2 \geq 0\right\}, & \gamma \geq 0, \end{cases} \tag{19}$$

where $\mathbf{e} = (1, \mathbf{0}^T)^T$, $\mathbf{P} = \mathbf{I} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}$, $\mathbf{a} = \mathbf{A}\mathbf{e}$, and $\gamma = \frac{1}{2} - \mathbf{a}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{a}$. For all $\boldsymbol{\xi} \in \mathcal{D}_{\bar{q}}$,

$$\bar{q}(\boldsymbol{\xi}) = \mathbf{v}^T\boldsymbol{\xi} + f(\mathbf{P}\boldsymbol{\xi}),$$

where

$$\mathbf{v} = \begin{cases} \mathbf{R}\mathbf{A}^T(\mathbf{A}\mathbf{R}\mathbf{A}^T)^{-1}\mathbf{b}, & \gamma \neq 0, \\ \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b}, & \gamma = 0, \end{cases} \tag{20}$$

$$f(\mathbf{u}) = \begin{cases} \frac{\sqrt{-\gamma(\mathbf{b}^T(\mathbf{A}\mathbf{R}\mathbf{A}^T)^{-1}\mathbf{b})}}{\gamma}\sqrt{(\mathbf{e}^T\mathbf{u})^2 - \gamma\|\mathbf{u}\|^2}, & \gamma \neq 0, \\ \left(\frac{\|\mathbf{y}_0\|^2 - 2(\mathbf{e}^T\mathbf{y}_0)^2}{2\mathbf{e}^T\mathbf{y}_0}\right)\mathbf{e}^T\mathbf{u} - \mathbf{e}^T\mathbf{y}_0\left(\frac{\|\mathbf{u}\|^2 - 2(\mathbf{e}^T\mathbf{u})^2}{2\mathbf{e}^T\mathbf{u}}\right), & \gamma = 0, \end{cases} \tag{21}$$

and $\mathbf{y}_0 = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b}$.

The proof of this result is fairly straightforward, and is, therefore, relegated to Appendix A.

3 Active set algorithm for the Lagrangian dual problem

Note that from (11) we have that $\mathbf{A}\boldsymbol{\xi} = \mathbf{U}_1^T\mathbf{U}_0\boldsymbol{\Sigma}_0^{-1}\mathbf{V}_0^T\mathbf{B}^T(\mathbf{f} - \mathbf{E}^T\boldsymbol{\lambda}) = \mathbf{0}$, i.e. $\boldsymbol{\xi} = \mathbf{P}\boldsymbol{\xi}$. Thus, (8), (9), (10), (11), and Lemma 2 imply that the Lagrangian dual problem is given by

$$\begin{array}{rrcrcl} \max & (\mathbf{f} - \mathbf{E}^T\boldsymbol{\lambda})^T\mathbf{z}_0 + \mathbf{v}^T\boldsymbol{\xi} + f(\boldsymbol{\xi}), \\ \text{subject to} & \mathbf{L}\boldsymbol{\lambda} & + & \boldsymbol{\xi} & = & \mathbf{h}, \\ & \mathbf{M}\boldsymbol{\lambda} & & & = & \mathbf{p}, \\ & \boldsymbol{\lambda} & & & \geq & \mathbf{0}, \\ & & & \boldsymbol{\xi} & \in & \mathcal{K}, \end{array} \tag{22}$$

where

$$\begin{array}{rcl} \mathbf{L} & = & \mathbf{U}_0\boldsymbol{\Sigma}_0^{-1}\mathbf{V}_0^T\mathbf{B}^T\mathbf{E}^T, \\ \mathbf{h} & = & \mathbf{U}_0\boldsymbol{\Sigma}_0^{-1}\mathbf{V}_0^T\mathbf{B}^T\mathbf{f}, \\ \mathbf{M} & = & \mathbf{V}_1^T\mathbf{B}^T\mathbf{E}^T, \\ \mathbf{p} & = & \mathbf{V}_1^T\mathbf{B}^T\mathbf{f}, \end{array}$$

and

$$\mathcal{K} = \begin{cases} \mathbf{R}^p, & \gamma < 0, \\ \{\mathbf{z} : \mathbf{e}^T\mathbf{z} \geq 0, (\mathbf{e}^T\mathbf{z})^2 - \gamma\|\mathbf{z}\|^2 \geq 0\}, & \gamma \geq 0, \end{cases}$$

where $\mathbf{v}$ and $f(\cdot)$ are as defined in (20) and (21) respectively. In the rest of the paper we denote the system of linear equalities in (22) by $\mathcal{A}[\boldsymbol{\lambda}, \boldsymbol{\xi}, \mathbf{h}, \mathbf{p}] = \mathbf{0}$.

When $\gamma \le 0$, the constraints in (22) are linear, hence (22) can be solved using any standard active set method for optimizing a concave function over a polytope. Moreover, γ is strictly positive for all single-cone SOCPs arising in the context of robust optimization. Therefore, in this paper we focus on constructing an active set algorithm for the case when γ is strictly positive. In the rest of this section we prove that the LagrangeDual algorithm displayed in Figure 1 computes an optimal solution of (22). We adopt the convention that a solution algorithm returns the empty set as a solution if, and only if, the problem is infeasible.

Let $\mathcal{C} = \big\{\boldsymbol{\xi} : \exists \boldsymbol{\lambda} \ge \mathbf{0} \text{ s.t. } \mathcal{A}[\boldsymbol{\lambda}, \boldsymbol{\xi}, \mathbf{h}, \mathbf{p}] = \mathbf{0}\big\}$. Then $\mathcal{C} \cap \mathcal{K} = \{\mathbf{h} - \mathbf{M}\boldsymbol{\lambda} : \mathbf{M}\boldsymbol{\lambda} = \mathbf{p}, \mathbf{h} - \mathbf{M}\boldsymbol{\lambda} \in \mathcal{K}, \boldsymbol{\lambda} \ge \mathbf{0}\}$. We construct the active set algorithm by considering the following three mutually exclusive cases: $\mathcal{C} \cap \mathcal{K} = \emptyset$, $\mathcal{C} \cap \mathcal{K} \subset \partial\mathcal{K}$, and $\mathcal{C} \cap \mathbf{int}(\mathcal{K}) \ne \emptyset$. In order to distinguish between these three cases we "homogenize" the set $\mathcal{C} \cap \mathcal{K}$ and solve the following least squares problem

$$
\begin{array}{rrcrcl}
\min & \multicolumn{5}{l}{\|\alpha\mathbf{h} - \mathbf{L}\boldsymbol{\lambda}\|^2,} \\
\text{subject to} & \alpha\mathbf{e}^T\mathbf{h} & - & \mathbf{e}^T\mathbf{L}\boldsymbol{\lambda} & = & 1, \\
 & \alpha\mathbf{p} & - & \mathbf{M}\boldsymbol{\lambda} & = & \mathbf{0}, \\
 & \alpha & & & \ge & 0, \\
 & & & \boldsymbol{\lambda} & \ge & \mathbf{0}.
\end{array}
\tag{23}
$$

Let $(\alpha^{(0)}, \boldsymbol{\mu}^{(0)})$ denote the optimal solution of (23). Then one of the following four mutually exclusive conditions holds.

(i) Either (23) is infeasible or $\|\alpha^{(0)}\mathbf{h} - \mathbf{L}\boldsymbol{\mu}^{(0)}\|^2 > \frac{1}{\gamma}$. Since (23) was constructed by "homogenizing" $\mathcal{C} \cap \mathcal{K}$, it follows that either $\mathcal{C} \cap \mathcal{K} = \emptyset$ or $\mathcal{C} \cap \mathcal{K} = \{\mathbf{0}\}$. The latter can be checked by solving an LP.

(ii) $\|\alpha^{(0)}\mathbf{h} - \mathbf{L}\boldsymbol{\mu}^{(0)}\|^2 < \frac{1}{\gamma}$. We have the following two possibilities:

   (a) $\alpha^{(0)} > 0$: $\boldsymbol{\mu}^{(2)} = \frac{1}{\alpha^{(0)}}\boldsymbol{\mu}^{(0)}$ satisfies $\mathbf{h} - \mathbf{L}\boldsymbol{\mu}^{(2)} \in \mathcal{C} \cap \mathbf{int}(\mathcal{K})$.

   (b) $\alpha^{(0)} = 0$: It is easy to check that $\boldsymbol{\mu}^{(0)}$ is a recession direction of the polytope $\mathcal{P} = \{\boldsymbol{\lambda} : \mathbf{M}\boldsymbol{\lambda} = \mathbf{p}, \boldsymbol{\lambda} \ge \mathbf{0}\}$ and $-\mathbf{e}^T\mathbf{L}\boldsymbol{\mu}^{(0)} = 1$. Let $\widehat{\boldsymbol{\lambda}} \in \mathcal{P}$ (in particular, if $\boldsymbol{\mu}^{(1)}$ is well defined, one can set $\widehat{\boldsymbol{\lambda}} = \boldsymbol{\mu}^{(1)}$). Then, by definition, $\boldsymbol{\lambda}_\omega = \widehat{\boldsymbol{\lambda}} + \omega\boldsymbol{\mu}^{(0)} \in \mathcal{P}$ for all $\omega \ge 0$. Since $\mathbf{e}^T(\mathbf{h} - \mathbf{L}\boldsymbol{\lambda}_\omega) > 0$ for all large enough ω, and $\lim_{\omega\to\infty}\big\{\|\mathbf{h} - \mathbf{L}\boldsymbol{\lambda}_\omega\|/(\mathbf{e}^T(\mathbf{h} - \mathbf{L}\boldsymbol{\lambda}_\omega))\big\} < \frac{1}{\sqrt{\gamma}}$, it follows that there exists $\omega > 0$ such that $\boldsymbol{\mu}^{(2)} = \boldsymbol{\lambda}_\omega$ satisfies $\mathbf{h} - \mathbf{L}\boldsymbol{\mu}^{(2)} \in \mathcal{C} \cap \mathbf{int}(\mathcal{K})$.

   In this case, LagrangeDual completes the optimization by calling the ActiveSet algorithm displayed in Figure 2.

(iii) $\|\alpha^{(0)}\mathbf{h} - \mathbf{L}\boldsymbol{\mu}^{(0)}\|^2 = \frac{1}{\gamma}$. In this case $\mathcal{C} \cap \mathbf{int}(\mathcal{K}) = \emptyset$ and one has to consider the following two possibilities.

   (a) $\alpha^{(0)} > 0$: $\boldsymbol{\mu}^{(2)} = \frac{1}{\alpha^{(0)}}\boldsymbol{\mu}^{(0)}$ satisfies $\mathbf{h} - \mathbf{L}\boldsymbol{\mu}^{(2)} \in \mathcal{C} \cap \partial\mathcal{K}$. Since the optimal value of (23) is $1/\gamma$ and the Euclidean norm is a strictly convex function, it follows that $\mathcal{C} \cap \mathcal{K} = \big\{\omega(\mathbf{h} - \mathbf{L}\boldsymbol{\mu}^{(2)}) : \mathbf{h} - \mathbf{L}\boldsymbol{\lambda} = \omega(\mathbf{h} - \mathbf{L}\boldsymbol{\mu}^{(2)}), \mathbf{M}\boldsymbol{\lambda} = \mathbf{p}, \boldsymbol{\lambda} \ge \mathbf{0}, \omega \ge 0\big\}$. Since $f(\xi) = 0$ for all $\xi \in \mathcal{C} \cap \mathcal{K}$ (see (21)), it follows that the optimization problem (22) reduces to the LP

$$
\begin{array}{rrcrcl}
\max & \multicolumn{5}{l}{(\mathbf{f} - \mathbf{E}^T\boldsymbol{\lambda})^T\mathbf{z}_0 + \mathbf{v}^T(\mathbf{h} - \mathbf{L}\boldsymbol{\mu}^{(2)})\omega,} \\
\text{subject to} & \mathbf{L}\boldsymbol{\lambda} & + & (\mathbf{h} - \mathbf{L}\boldsymbol{\mu}^{(2)})\omega & = & \mathbf{h}, \\
 & \mathbf{M}\boldsymbol{\lambda} & & & = & \mathbf{p}, \\
 & \boldsymbol{\lambda} & & & \ge & \mathbf{0}, \\
 & & & \omega & \ge & 0.
\end{array}
\tag{24}
$$

The LagrangeDual **Algorithm**:
Input: Optimization problem (22).
Output: Optimal solution of (22).

set $\boldsymbol{\mu}^{(1)} \leftarrow \text{argmax}\left\{-\mathbf{z}_0^T \mathbf{E}^T \boldsymbol{\lambda} : \mathcal{A}[\boldsymbol{\lambda}, \mathbf{0}, \mathbf{h}, \mathbf{p}] = \mathbf{0}, \boldsymbol{\lambda} \geq \mathbf{0}\right\}$

set $(\alpha^{(0)}, \boldsymbol{\mu}^{(0)}) \leftarrow \text{argmin}\{(23)\}$.

if $(\alpha^{(0)}, \boldsymbol{\mu}^{(0)}) = \emptyset$ or $\left(\|\alpha^{(0)}\mathbf{h} - \mathbf{L}\boldsymbol{\mu}^{(0)}\|^2 > 1/\gamma\right)$

    set $\boldsymbol{\mu} \leftarrow \boldsymbol{\mu}^{(1)}$

else if $\left(\|\alpha^{(0)}\mathbf{h} - \mathbf{L}\boldsymbol{\mu}^{(0)}\|^2 < 1/\gamma\right)$

    if $(\alpha^{(0)} > 0)$ **set** $\boldsymbol{\mu}^{(2)} \leftarrow \frac{1}{\alpha^{(0)}}\boldsymbol{\mu}^{(0)}$
    else
        if $\boldsymbol{\mu}^{(1)} \neq \emptyset$ **set** $\boldsymbol{\mu} \leftarrow \boldsymbol{\mu}^{(1)}$; **else choose** $\boldsymbol{\mu} \in \{\boldsymbol{\lambda} : \mathbf{M}\boldsymbol{\lambda} = \mathbf{p}, \boldsymbol{\lambda} \geq \mathbf{0}\}$
        choose $\widehat{\omega}$ s.t. $\mathbf{h} - \mathbf{L}(\boldsymbol{\mu} + \widehat{\omega}\boldsymbol{\mu}^{(0)}) \in \textbf{int}(\mathcal{K})$
        set $\boldsymbol{\mu}^{(2)} \leftarrow \boldsymbol{\mu} + \widehat{\omega}\boldsymbol{\mu}^{(0)}$
    end
    set $\boldsymbol{\mu} \leftarrow$ ActiveSet$(\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)})$

else

    if $(\alpha^{(0)} > 0)$

        set $\boldsymbol{\mu}^{(2)} \leftarrow \frac{1}{\alpha^{(0)}}\boldsymbol{\mu}^{(0)}$
        set $(\omega, \boldsymbol{\mu}) \leftarrow \text{argmin}\{(24)\}$
    else

        set $(\omega, \boldsymbol{\mu}) \leftarrow \text{argmin}\{(25)\}$
    end

end

return $\boldsymbol{\mu}$

Figure 1: Lagrangian Dual Algorithm

(b) $\alpha^{(0)} = 0$: The recession direction $\boldsymbol{\mu}^{(0)}$ satisfies $-\mathbf{L}\boldsymbol{\mu}^{(0)} \in \partial\mathcal{K}$. An argument similar to the one in part (a) implies that (22) reduces to the LP

$$\begin{array}{rrcccl} \max & \multicolumn{5}{l}{(\mathbf{f} - \mathbf{E}^T\boldsymbol{\lambda})^T\mathbf{z}_0 + \mathbf{v}^T(\mathbf{h} - \mathbf{L}\boldsymbol{\mu}^{(2)})\omega,} \\ \text{subject to} & \mathbf{L}\boldsymbol{\lambda} & + & (\mathbf{h} - \mathbf{L}\boldsymbol{\mu}^{(2)})\omega & = & \mathbf{h}, \\ & \mathbf{M}\boldsymbol{\lambda} & & & = & \mathbf{p}, \\ & \boldsymbol{\lambda} & & & \geq & \mathbf{0}, \\ & & & \omega & \geq & 0. \end{array} \tag{25}$$

Next, we establish the correctness of the procedure ActiveSet displayed in Figure 2. We begin by showing that for any optimal solution $(\boldsymbol{\xi}^*, \boldsymbol{\lambda}^*)$ of (22) either $\boldsymbol{\xi}^* = \mathbf{0}$ or $\boldsymbol{\xi}^* \in \mathcal{C} \cap \mathbf{int}(\mathcal{K})$, i.e. $\boldsymbol{\xi} \notin \mathcal{C} \cap \big(\partial\mathcal{K} \backslash \{\mathbf{0}\}\big)$.

Lemma 3 *Suppose $\mathcal{C} \cap \mathbf{int}(\mathcal{K}) \neq \emptyset$ and let $(\boldsymbol{\xi}^*, \boldsymbol{\lambda}^*)$ denote any optimal solution of (22). Then $\boldsymbol{\xi}^* \notin \mathcal{C} \cap \big(\partial\mathcal{K} \backslash \{\mathbf{0}\}\big)$.*

Proof: Assume otherwise, i.e. $\boldsymbol{\xi}^* \in \mathcal{C} \cap \big(\partial\mathcal{K} \backslash \{\mathbf{0}\}\big)$ for some optimal solution $(\boldsymbol{\xi}^*, \boldsymbol{\lambda}^*)$. Let $(\boldsymbol{\xi}_0, \boldsymbol{\lambda}_0)$ denote any feasible solution of (22) with $\boldsymbol{\xi}_0 \in \mathcal{C} \cap \mathbf{int}(\mathcal{K})$. For $\beta \in [0, 1]$, let $(\boldsymbol{\xi}_\beta, \boldsymbol{\lambda}_\beta)$ denote the convex combination $(\boldsymbol{\xi}_\beta, \boldsymbol{\lambda}_\beta) = \beta(\boldsymbol{\xi}_0, \boldsymbol{\lambda}_0) + (1 - \beta)(\boldsymbol{\xi}^*, \boldsymbol{\lambda}^*)$ and let $r(\beta)$ denote the objective value of (22) evaluated at $(\boldsymbol{\xi}_\beta, \boldsymbol{\lambda}_\beta)$. Then

$$\begin{aligned} r(\beta) &= \mathbf{v}^T\boldsymbol{\xi}_\beta + \mathbf{f}^T\mathbf{z}_0 - \mathbf{z}_0^T\mathbf{E}^T\boldsymbol{\lambda}_\beta + f(\boldsymbol{\xi}_\beta), \\ &= \mathbf{v}^T\boldsymbol{\xi}_\beta + \mathbf{f}^T\mathbf{z}_0 - \mathbf{z}_0^T\mathbf{E}^T\boldsymbol{\lambda}_\beta + \theta\sqrt{(\mathbf{e}^T\boldsymbol{\xi}_\beta)^2 - \gamma\|\boldsymbol{\xi}_\beta\|^2}, \\ &\geq \underbrace{\big(\mathbf{v}^T\boldsymbol{\xi}^* + \mathbf{f}^T\mathbf{z}_0 - \mathbf{z}_0^T\mathbf{E}^T\boldsymbol{\lambda}^*\big)}_{=r(0)} + \beta\underbrace{\big(\mathbf{v}^T(\boldsymbol{\xi}_0 - \boldsymbol{\xi}^*) - \mathbf{z}_0^T\mathbf{E}^T(\boldsymbol{\lambda}_0 - \boldsymbol{\lambda}^*)\big)}_{\triangleq\delta} \\ &\quad + \theta\sqrt{\beta^2\left((\mathbf{e}^T\boldsymbol{\xi}_0)^2 - \gamma\|\boldsymbol{\xi}_0\|^2\right) + 2\beta(1-\beta)\left((\mathbf{e}^T\boldsymbol{\xi}_0)(\mathbf{e}^T\boldsymbol{\xi}^*) - \gamma\|\boldsymbol{\xi}_0\|\|\boldsymbol{\xi}^*\|\right)}, \end{aligned} \tag{26}$$

where $\theta = \frac{\sqrt{-\gamma\mathbf{b}^T(\mathbf{ARA}^T)^{-1}\mathbf{b}}}{\gamma}$ and the last inequality follows from the fact that $(\mathbf{e}^T\boldsymbol{\xi}^*)^2 - \gamma\|\boldsymbol{\xi}^*\|^2 = 0$. Since $\boldsymbol{\xi}_0 \in \mathcal{C} \cap \mathbf{int}(\mathcal{K})$ we have $\epsilon = \min\big\{(\mathbf{e}^T\boldsymbol{\xi}_0)^2 - \gamma\|\boldsymbol{\xi}_0\|^2, (\mathbf{e}^T\boldsymbol{\xi}_0)(\mathbf{e}^T\boldsymbol{\xi}^*) - \gamma\|\boldsymbol{\xi}_0\|\|\boldsymbol{\xi}^*\|\big\} > 0$. From (26) we have that $r(\beta) - r(0) \geq \theta\sqrt{\epsilon}\sqrt{2\beta - \beta^2} + \beta\delta$. Choose β_0 as follows.

$$\beta_0 = \begin{cases} 1 & \delta \geq 0, \\ 1 + \frac{\delta}{\sqrt{\theta^2\epsilon + \delta^2}} & \delta < 0. \end{cases}$$

Then it follows that $\beta_0 > 0$ and $r(\beta_0) - r(0) > 0$. A contradiction. ■

The ActiveSet algorithm receives as input

(i) $\boldsymbol{\mu}^{(1)} = \mathrm{argmax}\{-\mathbf{z}_0^T\mathbf{E}^T\boldsymbol{\lambda} : \mathcal{A}[\boldsymbol{\lambda}, \mathbf{0}, \mathbf{h}, \mathbf{p}], \boldsymbol{\lambda} \geq \mathbf{0}\}$, and

(ii) a vector $\boldsymbol{\mu}^{(2)}$ such that $\mathbf{h} - \mathbf{L}\boldsymbol{\mu}^{(2)} \in \mathcal{C} \cap \mathbf{int}(\mathcal{K})$.

When $\boldsymbol{\mu}^{(1)} \neq \emptyset$, the algorithm calls the procedure FindDirection that returns an ascent direction at $(\boldsymbol{\xi}, \boldsymbol{\lambda}) = (\mathbf{0}, \boldsymbol{\mu}^{(1)})$, if it exists; otherwise it returns $(\mathbf{0}, \mathbf{0})$. If FindDirection returns $(\mathbf{0}, \mathbf{0})$, it follows that $(\mathbf{0}, \boldsymbol{\mu}^{(1)})$ is optimal and the algorithm terminates; otherwise ActiveSet calls the procedure FindStep$((\boldsymbol{\xi}, \boldsymbol{\lambda}), (\mathbf{d}_\xi, \mathbf{d}_\lambda), \alpha_q)$ that computes the iterate $(\boldsymbol{\xi}^{(0)}, \boldsymbol{\lambda}^{(0)})$ as follows.

$$(\boldsymbol{\xi}^{(0)}, \boldsymbol{\lambda}^{(0)}) = (\boldsymbol{\xi}, \boldsymbol{\lambda}) + \alpha_{\min}(\mathbf{d}_\xi, \mathbf{d}_\lambda), \quad \alpha_{\min} = \min\{\max\{\alpha : \boldsymbol{\lambda} + \alpha\mathbf{d}_\lambda \geq \mathbf{0}\}, \alpha_q\}.$$

The ACTIVESET **Algorithm**:
Input: Optimization problem (22), $\boldsymbol{\mu}^{(1)}$, and $\boldsymbol{\mu}^{(2)}$.
Output: Optimal solution of (22).

```
quit ← 0    k ← 0
if (μ^(1) ≠ ∅)
        (d_ξ, d_λ) ← FindDirection(∅)
        if (v^T d_ξ − z_0^T E^T d_λ + f(d_ξ) ≤ 0) return (0, μ^(1))
        else (ξ^(0), λ^(0)) ← FindStep((0, μ^(1)), (d_ξ, d_λ), ∞)
else
        λ^(0) ← μ^(2)
end if
W^(k) ← Σ_{i: λ_i^(k) = 0} e_i e_i^T
/* e_i denotes the i-th column of an identity matrix */
while (∼quit)
        (d_ξ, d_λ, α_q) ← FindOpt(ξ^(k), λ^(k), W^(k))
        if ((d_ξ, d_λ) ≠ 0)
            (ξ^(k+1), λ^(k+1)) ← FindStep((ξ^(k), λ^(k)), (d_ξ, d_λ), α_q)
            k ← k + 1
            W^(k) ← Σ_{i: λ_i^(k) = 0} e_i e_i^T
        else
            ρ ← FindMultipliers(W^(k), ξ^(k))
            if (max_i{ρ_i} ≤ 0)
                quit ← 1
            else
                choose j: ρ_j > 0
                W^(k+1) ← W^(k) − e_j e_j^T
            end
        end
end
return λ^(k)
```

Figure 2: Active Set Algorithm

Since $\alpha \mathbf{d}_\xi \in \mathcal{K}$ for all $\alpha \geq 0$, $\alpha_{\min}$ is only limited by the non-negativity constraints on $\boldsymbol{\lambda}$. Note that the iterate $(\boldsymbol{\xi}^{(0)}, \boldsymbol{\lambda}^{(0)})$ satisfies $\boldsymbol{\xi}^{(0)} \in \mathbf{int}(\mathcal{K})$; therefore, the optimum solution $(\boldsymbol{\xi}^*, \boldsymbol{\lambda}^*)$ also satisfies $\boldsymbol{\xi}^* \in \mathbf{int}(\mathcal{K})$ by Lemma 3.

Next, we show that the procedure FindDirection can be implemented efficiently. The pair $(\mathbf{d}_\xi, \mathbf{d}_\lambda)$ is an ascent direction at $(\mathbf{0}, \boldsymbol{\mu}^{(1)})$ if, and only if, $(\mathbf{d}_\xi, \mathbf{d}_\lambda)$ is a recession direction for the set

$$\begin{array}{rcl} -\mathbf{z}_0^T \mathbf{E}^T \mathbf{d}_\lambda + \mathbf{v}^T \mathbf{d}_\xi + \theta\sqrt{(\mathbf{e}^T \mathbf{d}_\xi)^2 - \gamma\|\mathbf{d}_\xi\|^2} > 0, && \\ \mathcal{A}[\mathbf{d}_\lambda, \mathbf{d}_\xi, \mathbf{0}, \mathbf{0}] &=& \mathbf{0}, \\ \mathbf{d}_\xi &\in& \mathcal{K}, \end{array} \tag{27}$$

Lemma 4 *Let $\mathcal{A}_{\mathbf{W}}[\mathbf{u}, \mathbf{v}, \boldsymbol{\gamma}, \boldsymbol{\nu}] = \mathbf{0}$ denote the system of linear equalities*

$$\begin{bmatrix} \mathbf{L} & \mathbf{I} \\ \mathbf{M} & \mathbf{0} \\ \mathbf{W} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\nu} \\ \mathbf{0} \end{bmatrix},$$

where $(\mathbf{u}, \mathbf{v})$ are variables, and $(\boldsymbol{\gamma}, \boldsymbol{\nu}, \mathbf{W})$ are parameters. Then a recession direction $(\mathbf{d}_\xi, \mathbf{d}_\lambda)$ for the set

$$\begin{array}{rcl} -\mathbf{z}_0^T \mathbf{E}^T \mathbf{d}_\lambda + \mathbf{v}^T \mathbf{d}_\xi + \theta\sqrt{(\mathbf{e}^T \mathbf{d}_\xi)^2 - \gamma\|\mathbf{d}_\xi\|^2} > 0, && \\ \mathcal{A}_{\mathbf{W}}[\mathbf{d}_\lambda, \mathbf{d}_\xi, \mathbf{0}, \mathbf{0}] &=& \mathbf{0}, \\ \mathbf{d}_\xi &\in& \mathcal{K}, \end{array} \tag{28}$$

if it exists, can be computed by solving two systems of linear equalities.

Remark 2 *Although* FindDirection *computes an ascent direction of the set (28) for the special case $\mathbf{W} = \mathbf{0}$, we prove the result for general $\mathbf{W}$ since we need such a result at a later stage.*

Proof: The set in (28) has a recession direction if, and only if, the optimization problem

$$\begin{array}{rrcl} \max & -\mathbf{z}_0^T \mathbf{E}^T \mathbf{d}_\lambda + \mathbf{v}^T \mathbf{d}_\xi + \theta\sqrt{(\mathbf{e}^T \mathbf{d}_\xi)^2 - \gamma\|\mathbf{d}_\xi\|^2}, && \\ \text{subject to} & \mathcal{A}_{\mathbf{W}}[\mathbf{d}_\lambda, \mathbf{d}_\xi, \mathbf{0}, \mathbf{0}] &=& \mathbf{0}, \\ & \mathbf{d}_\xi &\in& \mathcal{K}, \end{array} \tag{29}$$

is unbounded.

An argument similar to the one employed in the proof of Lemma 3 establishes that one can restrict attention to $(\mathbf{d}_\xi, \mathbf{d}_\lambda)$ satisfying $\mathbf{d}_\xi \in \mathbf{int}(\mathcal{K}) \cup \{\mathbf{0}\}$. The direction $(\mathbf{d}_\xi, \mathbf{d}_\lambda)$ can be computed by considering the following three cases:

(a) First consider positive recession directions of the form $(\mathbf{0}, \mathbf{d}_\lambda)$. It is easy to see that the all such directions are solutions of the following set of linear equalities.

$$\begin{array}{rcl} -\mathbf{z}_0^T \mathbf{E}^T \mathbf{d}_\lambda &=& 1, \\ \mathcal{A}_{\mathbf{W}}[\mathbf{d}_\lambda, \mathbf{0}, \mathbf{0}, \mathbf{0}] &=& \mathbf{0}. \end{array} \tag{30}$$

(b) Next, suppose (30) is infeasible; however, there still exists a positive recession direction for (29). Set $\mathbf{e}^T \mathbf{d}_\xi = 1$ in (29) to obtain

$$\begin{array}{rrcl} \max & -\mathbf{z}_0^T \mathbf{E}^T \mathbf{d}_\lambda + \mathbf{v}^T \mathbf{d}_\xi + \theta\sqrt{1 - \gamma\|\mathbf{d}_\xi\|^2}, && \\ \text{subject to} & \mathcal{A}_{\mathbf{W}}[\mathbf{d}_\lambda, \mathbf{d}_\xi, \mathbf{0}, \mathbf{0}] &=& \mathbf{0}, \\ & \mathbf{e}^T \mathbf{d}_\xi &=& 1, \\ & \gamma\|\mathbf{d}_\xi\|^2 &\leq& 1. \end{array} \tag{31}$$

Since (30) is assumed to be infeasible, (31) is bounded. Setting $\mathbf{d}_\xi = -\mathbf{L}\mathbf{d}_\lambda$, we get

$$\begin{array}{rrcl} \max & -\left(\mathbf{E}\mathbf{z}_0 + \mathbf{L}^T\mathbf{v}\right)^T\mathbf{d}_\lambda + \theta\sqrt{1-\gamma\|\mathbf{L}\mathbf{d}_\lambda\|^2}, & & \\ \text{subject to} & \mathcal{A}_{\mathbf{W}}[\mathbf{d}_\lambda, -\mathbf{L}\mathbf{d}_\lambda, \mathbf{0}, \mathbf{0}] & = & \mathbf{0}, \\ & \mathbf{e}^T\mathbf{L}\mathbf{d}_\lambda & = & -1, \\ & \gamma\|\mathbf{L}\mathbf{d}_\lambda\|^2 & \leq & 1. \end{array} \tag{32}$$

Since the optimal $\mathbf{d}_\xi^* \in \mathbf{int}(\mathcal{K})$, we have $\gamma\|\mathbf{L}\mathbf{d}_\lambda^*\|^2 < 1$, and therefore, the optimal Lagrange multiplier corresponding to this constraint is zero. Thus, the Lagrangian $\mathcal{L}$ of (32) reduces to

$$\mathcal{L} = -(\mathbf{E}\mathbf{z}_0 + \mathbf{L}^T\mathbf{v})^T\mathbf{d}_\lambda + \theta\sqrt{1-\gamma\|\mathbf{L}\mathbf{d}_\lambda\|^2} - \boldsymbol{\tau}^T\mathbf{M}\mathbf{d}_\lambda - \boldsymbol{\rho}^T\mathbf{W}\mathbf{d}_\lambda - \eta(\mathbf{e}^T\mathbf{L}\mathbf{d}_\lambda + 1)$$

and the first-order optimality conditions are given by

$$\begin{array}{lllllcl} \frac{\theta\gamma}{\beta}\mathbf{L}^T\mathbf{L}\mathbf{d}_\lambda & +\mathbf{M}^T\boldsymbol{\tau} & +\mathbf{W}^T\boldsymbol{\rho} & +\mathbf{L}^T\mathbf{e}\eta & = & -(\mathbf{E}\mathbf{z}_0 + \mathbf{L}^T\mathbf{v}), \\ \mathbf{M}\mathbf{d}_\lambda & & & & = & \mathbf{0}, \\ \mathbf{W}\mathbf{d}_\lambda & & & & = & \mathbf{0}, \\ \mathbf{e}^T\mathbf{L}\mathbf{d}_\lambda & & & & = & -1, \end{array} \tag{33}$$

where $\beta = \sqrt{1-\gamma\|\mathbf{L}\mathbf{d}_\lambda\|^2}$. Since we are looking for solutions $\mathbf{d}_\xi = -\mathbf{L}\mathbf{d}_\lambda \in \mathbf{int}(\mathcal{K})$, we are only interested in the solutions to (33) that satisfy $\beta > 0$.

By setting $\bar{\boldsymbol{\rho}} = \beta\boldsymbol{\rho}$, $\bar{\boldsymbol{\tau}} = \beta\boldsymbol{\tau}$, and $\bar{\eta} = \beta\eta$, we see that (33) is equivalent to

$$\underbrace{\begin{bmatrix} \theta\gamma\mathbf{L}^T\mathbf{L} & \mathbf{M}^T & \mathbf{W}^T & \mathbf{L}^T\mathbf{e} \\ \mathbf{M} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{W} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{e}^T\mathbf{L} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}}_{\triangleq\mathbf{K}} \begin{bmatrix} \mathbf{d}_\lambda \\ \bar{\boldsymbol{\tau}} \\ \bar{\boldsymbol{\rho}} \\ \bar{\eta} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ -1 \end{bmatrix} - \beta \begin{bmatrix} \mathbf{E}\mathbf{z}_0 + \mathbf{L}^T\mathbf{v} \\ \mathbf{0} \\ \mathbf{0} \\ 0 \end{bmatrix}. \tag{34}$$

Suppose $\mathbf{K}$ is non-singular. Let $\mathbf{w} = (\bar{\boldsymbol{\tau}}^T, \bar{\boldsymbol{\rho}}^T, \bar{\eta})^T$, $\mathbf{b}_1 = (\mathbf{0}^T, \mathbf{0}^T, -1)^T$ and $\mathbf{b}_2 = \mathbf{E}\mathbf{z}_0 + \mathbf{L}^T\mathbf{v}$. Partition $\mathbf{K}^{-1}$ into submatrices

$$\mathbf{K}^{-1} = \begin{bmatrix} \mathbf{K}_{11}^{-1} & \mathbf{K}_{12}^{-1} \\ \mathbf{K}_{12}^{-T} & \mathbf{K}_{22}^{-1} \end{bmatrix}$$

such that

$$\begin{bmatrix} \mathbf{d}_\lambda \\ \mathbf{w} \end{bmatrix} = \mathbf{K}^{-1}\begin{bmatrix} \mathbf{0} \\ \mathbf{b}_1 \end{bmatrix} - \beta\mathbf{K}^{-1}\begin{bmatrix} \mathbf{b}_2 \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{K}_{12}^{-1}\mathbf{b}_1 - \beta\mathbf{K}_{11}^{-1}\mathbf{b}_2 \\ \mathbf{K}_{22}^{-1}\mathbf{b}_1 - \beta\mathbf{K}_{12}^{-T}\mathbf{b}_2 \end{bmatrix}$$

This partition implies that $\mathbf{K}_{12}^{-T}\mathbf{L}^T\mathbf{L}\mathbf{K}_{11}^{-1} = \mathbf{0}$. Therefore,

$$\begin{array}{rcl} \beta^2 + \gamma\|\mathbf{L}\mathbf{d}_\lambda\|^2 - 1 & = & \beta^2(1 + \gamma\|\mathbf{L}\mathbf{K}_{11}^{-1}\mathbf{b}_2\|^2) - 2\beta\gamma(\mathbf{L}\mathbf{K}_{12}^{-1}\mathbf{b}_1)^T\mathbf{L}\mathbf{K}_{11}^{-1}\mathbf{b}_2 + \gamma\|\mathbf{L}\mathbf{K}_{12}^{-1}\mathbf{b}_1\|^2 - 1, \\ & = & \beta^2(1 + \gamma\|\mathbf{L}\mathbf{K}_{11}^{-1}\mathbf{b}_2\|^2) + \gamma\|\mathbf{L}\mathbf{K}_{12}^{-1}\mathbf{b}_1\|^2 - 1. \end{array}$$

Consequently, the unique positive solution of the quadratic equation $\beta^2 = 1 - \gamma\|\mathbf{L}\mathbf{d}_\lambda\|^2$ is

$$\beta = \sqrt{\frac{1 - \gamma\|\mathbf{L}\mathbf{K}_{12}^{-1}\mathbf{b}_1\|^2}{1 + \gamma\|\mathbf{L}\mathbf{K}_{11}^{-1}\mathbf{b}_2\|^2}}.$$

Thus, (33) has a solution if, and only if, $1 - \gamma\|\mathbf{L}\mathbf{K}_{12}^{-1}\mathbf{b}_1\|^2 > 0$.

The case where $\mathbf{K}$ is singular can be handled by taking the singular value decomposition of $\mathbf{K}$ and working in the appropriate range spaces.

(c) In case one is not able to produce a solution in either (a) or (b), it follows that the optimal solution of (29) is 0, and $(\mathbf{d}_\xi, \mathbf{d}_\lambda) = (\mathbf{0}, \mathbf{0})$ achieves this value.

■

When ActiveSet algorithm enters the `while` loop, we are guaranteed that $\boldsymbol{\xi}^* \in \mathbf{int}(\mathcal{K})$. Within the loop, one has to compute the optimal value of

$$\begin{array}{rrcl} \max & -\mathbf{z}_0^T \mathbf{E}^T \boldsymbol{\lambda} + \mathbf{v}^T \boldsymbol{\xi} + f(\boldsymbol{\xi}), & & \\ \text{subject to} & \mathcal{A}_{\mathbf{W}}[\boldsymbol{\lambda}, \boldsymbol{\xi}, \mathbf{h}, \mathbf{p}] & = & \mathbf{0}, \\ & \boldsymbol{\xi} & \in & \mathcal{K}\backslash\{\mathbf{0}\}, \end{array} \tag{35}$$

where $\mathbf{W}$ denotes the current inactive set, i.e. $\mathbf{W} = \sum_{i:\lambda_i = 0} \mathbf{e}_i \mathbf{e}_i^T$. At this stage we have already determined that $\boldsymbol{\xi} = \mathbf{0}$ is *not* optimal for (22); therefore, by Lemma 3 it follows that we can restrict ourselves to $\boldsymbol{\xi} \in \mathbf{int}(\mathcal{K})$. The procedure $(\mathbf{d}_\xi, \mathbf{d}_\lambda, \alpha_q) = \text{FindOpt}(\boldsymbol{\xi}, \boldsymbol{\lambda}, \mathbf{W})$ takes as input the current iterate and the current $\mathbf{W}$; and returns an output $(\mathbf{d}_\xi, \mathbf{d}_\lambda, \alpha_q)$ that satisfies the following.

(i) When (35) is bounded, $(\mathbf{d}_\xi, \mathbf{d}_\lambda) = (\boldsymbol{\xi}^*, \boldsymbol{\lambda}^*) - (\boldsymbol{\xi}, \boldsymbol{\lambda})$, where $(\boldsymbol{\xi}^*, \boldsymbol{\lambda}^*)$ is the optimal solution of (35) and $\alpha_q = 1$;

(ii) When (35) is unbounded, $(\mathbf{d}_\xi, \mathbf{d}_\lambda)$ is any recession direction of the feasible set of (35) satisfying $-\mathbf{z}_0^T \mathbf{E}^T \mathbf{d}_\lambda + \mathbf{v}^T \mathbf{d}_\xi + f(\mathbf{d}_\xi) > 0$ and $\alpha_q = \infty$.

When $(\mathbf{d}_\xi, \mathbf{d}_\lambda) = (\mathbf{0}, \mathbf{0})$, the ActiveSet algorithm checks the Lagrange multipliers $\boldsymbol{\rho}$ corresponding to the constraints $\mathbf{W}\boldsymbol{\lambda} = \mathbf{0}$ by calling the procedure FindMultipliers that computes the solution of

$$\begin{bmatrix} \mathbf{W}^T & \mathbf{M}^T \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho} \\ \boldsymbol{\tau} \end{bmatrix} = - \left[\mathbf{E}\mathbf{z}_0 + \mathbf{L}^T \mathbf{v} + \mathbf{L}^T \nabla f(\boldsymbol{\xi}^*) \right]. \tag{36}$$

If the signs of all the Lagrange multipliers are consistent with the KKT conditions, i.e. $\max_i\{\rho_i\} \le 0$, the algorithm terminates; otherwise, it drops one of the constraints with the incorrect sign. Lemma 6 establishes that ActiveSet terminates finitely. Thus, all that remains to be shown is that FindOpt can be implemented efficiently.

Lemma 5 *Suppose there exists a feasible $(\bar{\boldsymbol{\xi}}, \bar{\boldsymbol{\lambda}})$ for (35) such that $\bar{\boldsymbol{\xi}} \in \mathbf{int}(\mathcal{K})$. Then (35) can be solved in closed form by solving at most three systems of linear equations.*

Proof: Let $\mathbf{d}_\xi = \boldsymbol{\xi} - \bar{\boldsymbol{\xi}}$ and $\mathbf{d}_\lambda = \boldsymbol{\lambda} - \bar{\boldsymbol{\lambda}}$. Then (35) is equivalent to

$$\begin{array}{rrcl} \max & -\mathbf{z}_0^T \mathbf{E}^T \mathbf{d}_\lambda + \mathbf{v}^T \mathbf{d}_\xi + \theta\sqrt{(\mathbf{e}^T(\bar{\boldsymbol{\xi}} + \mathbf{d}_\xi))^2 - \gamma\|\bar{\boldsymbol{\xi}} + \mathbf{d}_\xi\|^2}, & & \\ \text{subject to} & \mathcal{A}_{\mathbf{W}}[\mathbf{d}_\lambda, \mathbf{d}_\xi, \mathbf{0}, \mathbf{0}] & = & \mathbf{0}, \\ & \bar{\boldsymbol{\xi}} + \mathbf{d}_\xi & \in & \mathcal{K}\backslash\{\mathbf{0}\}. \end{array} \tag{37}$$

First, suppose (37) is unbounded, i.e. there exists $(\mathbf{d}_\xi, \mathbf{d}_\lambda)$ such that

$$\begin{aligned} &\lim_{t\to\infty} \Big\{ - t\mathbf{z}_0^T \mathbf{E}^T \mathbf{d}_\lambda + t\mathbf{v}^T \mathbf{d}_\xi + \theta\sqrt{(\mathbf{e}^T(\bar{\boldsymbol{\xi}} + t\mathbf{d}_\xi))^2 - \gamma\|\bar{\boldsymbol{\xi}} + t\mathbf{d}_\xi\|^2} \Big\} \\ &\quad = \lim_{t\to\infty} \Big\{ - t\mathbf{z}_0^T \mathbf{E}^T \mathbf{d}_\lambda + t\mathbf{v}^T \mathbf{d}_\xi + t\theta\sqrt{(\mathbf{e}^T(\bar{\boldsymbol{\xi}}/t + \mathbf{d}_\xi))^2 - \gamma\|\bar{\boldsymbol{\xi}}/t + \mathbf{d}_\xi\|^2} \Big\} \quad = \quad +\infty. \end{aligned}$$

Since $\bar{\boldsymbol{\xi}}/t \to 0$, it follows that (37) is unbounded if, and only if, $(\mathbf{d}_\xi, \mathbf{d}_\lambda)$ is a recession direction for

$$\begin{array}{rcl} -\mathbf{z}_0^T\mathbf{E}^T\mathbf{d}_\lambda + \mathbf{v}^T\mathbf{d}_\xi + \theta\sqrt{(\mathbf{e}^T\mathbf{d}_\xi)^2 - \gamma\|\mathbf{d}_\xi\|^2} > 0, & & \\ \mathcal{A}_\mathbf{W}[\mathbf{d}_\lambda, \mathbf{d}_\xi, \mathbf{0}, \mathbf{0}] & = & \mathbf{0}, \\ \mathbf{d}_\xi & \in & \mathcal{K}. \end{array} \tag{38}$$

Since (38) is the same as (28), it follows that a positive recession direction for (37), if it exists, can be computed by solving at most two systems of linear equations.

Next, suppose (37) is bounded. By introducing a scaling parameter α, (37) can be reformulated as

$$\begin{array}{rl} \max & -(\mathbf{E}\mathbf{z}_0 + \mathbf{L}^T\mathbf{v})^T\mathbf{d}_\lambda + \theta\sqrt{1 - \gamma\|\alpha\bar{\boldsymbol{\xi}} - \mathbf{L}\mathbf{d}_\lambda\|^2}, \\ \text{subject to} & \begin{array}{rcl} \mathcal{A}_\mathbf{W}[\mathbf{d}_\lambda, -\mathbf{L}\mathbf{d}_\lambda, \mathbf{0}, \mathbf{0}] & = & \mathbf{0}, \\ \alpha & \geq & 0, \\ -\mathbf{e}^T\mathbf{L}\mathbf{d}_\lambda + \mathbf{e}^T\bar{\boldsymbol{\xi}}\alpha & = & 1, \\ \gamma\|\alpha\bar{\boldsymbol{\xi}} - \mathbf{L}\mathbf{d}_\lambda\|^2 & \leq & 1. \end{array} \end{array}$$

Since (37) is bounded, i.e. it does not have any positive recession direction, we have that $\alpha^* > 0$. Also, by Lemma 3 it follows that $\alpha^*\bar{\boldsymbol{\xi}} + \mathbf{d}_\xi^* \in \mathbf{int}(K)$, i.e. $\gamma\|\alpha^*\bar{\boldsymbol{\xi}} - \mathbf{L}\mathbf{d}_\lambda^*\|^2 < 1$, therefore the optimal Lagrange multiplier corresponding to this constraint is zero. Consequently, the Lagrangian $\mathcal{L}$ reduces to

$$\mathcal{L} = -(\mathbf{E}\mathbf{z}_0 + \mathbf{L}^T\mathbf{v})^T\mathbf{d}_\lambda + \theta\sqrt{1 - \gamma\|\alpha\bar{\boldsymbol{\xi}} - \mathbf{L}\mathbf{d}_\lambda\|^2} - \boldsymbol{\tau}^T\mathbf{M}\mathbf{d}_\lambda - \boldsymbol{\rho}^T\mathbf{W}\mathbf{d}_\lambda - \eta(\mathbf{e}^T\alpha\bar{\boldsymbol{\xi}} - \mathbf{e}^T\mathbf{L}\mathbf{d}_\lambda - 1).$$

The first-order optimality conditions are given by

$$\begin{array}{rcrcrcrcrcl} \frac{\theta\gamma}{\beta}\mathbf{L}^T\mathbf{L}\mathbf{d}_\lambda & - & \frac{\theta\gamma}{\beta}\mathbf{L}^T\bar{\boldsymbol{\xi}}\alpha & - & \mathbf{L}^T\mathbf{e}\eta & + & \mathbf{M}^T\boldsymbol{\tau} & + & \mathbf{W}^T\boldsymbol{\rho} & = & -(\mathbf{E}\mathbf{z}_0 + \mathbf{L}^T\mathbf{v}), \\ -\frac{\theta\gamma}{\beta}\bar{\boldsymbol{\xi}}^T\mathbf{L}\mathbf{d}_\lambda & + & \frac{\theta\gamma}{\beta}\|\bar{\boldsymbol{\xi}}\|^2\alpha & + & \mathbf{e}^T\bar{\boldsymbol{\xi}}\eta & & & & & = & 0, \\ -\mathbf{e}^T\mathbf{L}\mathbf{d}_\lambda & + & \mathbf{e}^T\bar{\boldsymbol{\xi}}\alpha & & & & & & & = & 1, \\ \mathbf{M}\mathbf{d}_\lambda & & & & & & & & & = & \mathbf{0}, \\ \mathbf{W}\mathbf{d}_\lambda & & & & & & & & & = & \mathbf{0}, \end{array} \tag{39}$$

where $\beta = \sqrt{1 - \gamma\|\alpha\bar{\boldsymbol{\xi}} - \mathbf{L}\mathbf{d}_\lambda\|^2}$. Set $\bar{\boldsymbol{\rho}} = \beta\boldsymbol{\rho}$, $\bar{\boldsymbol{\tau}} = \beta\boldsymbol{\tau}$, and $\bar{\eta} = \beta\eta$. Then (39) is equivalent to

$$\underbrace{\begin{bmatrix} \theta\gamma\mathbf{L}^T\mathbf{L} & -\theta\gamma\mathbf{L}^T\bar{\boldsymbol{\xi}} & -\mathbf{L}^T\mathbf{e} & \mathbf{M}^T & \mathbf{W}^T \\ -\theta\gamma\bar{\boldsymbol{\xi}}^T\mathbf{L} & \theta\gamma\|\bar{\boldsymbol{\xi}}\|^2 & \mathbf{e}^T\bar{\boldsymbol{\xi}} & \mathbf{0} & \mathbf{0} \\ -\mathbf{e}^T\mathbf{L} & \mathbf{e}^T\bar{\boldsymbol{\xi}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{M} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{W} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}}_{\triangleq\mathbf{K}} \begin{bmatrix} \mathbf{d}_\lambda \\ \alpha \\ \bar{\eta} \\ \bar{\boldsymbol{\tau}} \\ \bar{\boldsymbol{\rho}} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ 0 \\ 1 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} - \beta \begin{bmatrix} \mathbf{E}\mathbf{z}_0 + \mathbf{L}^T\mathbf{v} \\ 0 \\ 0 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}. \tag{40}$$

Suppose $\mathbf{K}$ is non-singular. Let $\widehat{\mathbf{d}} = (\mathbf{d}_\lambda^T, \alpha)^T$, $\mathbf{w} = (\bar{\eta}, \bar{\boldsymbol{\tau}}^T, \bar{\boldsymbol{\rho}}^T)^T$, $\mathbf{b}_1 = (1, \mathbf{0}^T, \mathbf{0}^T)^T$, and $\mathbf{b}_2 = ((\mathbf{E}\mathbf{z}_0 + \mathbf{L}^T\mathbf{v})^T, \mathbf{0}^T)^T$. Partition $\mathbf{K}^{-1}$ such that

$$\begin{bmatrix} \widehat{\mathbf{d}} \\ \mathbf{w} \end{bmatrix} = \mathbf{K}^{-1}\begin{bmatrix} \mathbf{0} \\ \mathbf{b}_1 \end{bmatrix} - \beta\mathbf{K}^{-1}\begin{bmatrix} \mathbf{b}_2 \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{K}_{12}^{-1}\mathbf{b}_1 - \beta\mathbf{K}_{11}^{-1}\mathbf{b}_2 \\ \mathbf{K}_{22}^{-1}\mathbf{b}_1 - \beta\mathbf{K}_{12}^{-T}\mathbf{b}_2 \end{bmatrix}$$

This partition implies that $\mathbf{K}_{12}^{-T}[-\mathbf{L},\bar{\boldsymbol{\xi}}]^T[-\mathbf{L},\bar{\boldsymbol{\xi}}]\mathbf{K}_{11}^{-1} = \mathbf{0}$. Therefore,

$$\begin{aligned}\beta^2 + \gamma\|[-\mathbf{L},\bar{\boldsymbol{\xi}}]\widehat{\mathbf{d}}\|^2 - 1 &= \beta^2(1+\gamma\|[-\mathbf{L},\bar{\boldsymbol{\xi}}]\mathbf{K}_{11}^{-1}\mathbf{b}_2\|^2) - 2\beta\gamma([-\mathbf{L},\bar{\boldsymbol{\xi}}]\mathbf{K}_{12}^{-1}\mathbf{b}_1)^T[-\mathbf{L},\bar{\boldsymbol{\xi}}]\mathbf{K}_{11}^{-1}\mathbf{b}_2 \\ &\quad + \gamma\|[-\mathbf{L},\bar{\boldsymbol{\xi}}]\mathbf{K}_{12}^{-1}\mathbf{b}_1\|^2 - 1, \\ &= \beta^2(1+\gamma\|\mathbf{L}\mathbf{K}_{11}^{-1}\mathbf{b}_2\|^2) + \gamma\|\mathbf{L}\mathbf{K}_{12}^{-1}\mathbf{b}_1\|^2 - 1.\end{aligned}$$

Consequently, the unique positive solution of the quadratic equation $\beta^2 = 1 - \gamma\|[-\mathbf{L},\bar{\boldsymbol{\xi}}]\widehat{\mathbf{d}}\|^2$ is

$$\beta = \sqrt{\frac{1-\gamma\|[-\mathbf{L},\bar{\boldsymbol{\xi}}]\mathbf{K}_{12}^{-1}\mathbf{b}_1\|^2}{1+\gamma\|[-\mathbf{L},\bar{\boldsymbol{\xi}}]\mathbf{K}_{11}^{-1}\mathbf{b}_2\|^2}}.$$

The case where $\mathbf{K}$ is singular can be handled by taking the SVD of $\mathbf{K}$ and working in the appropriate range spaces. ■

In our numerical experiments we found that solving (40) as a least squares problem was much faster than computing the inverse or the SVD of $\mathbf{K}$.

Lemma 5 implies that at each iteration of the ActiveSet algorithm, we have to solve at most three systems of linear equations, namely the equations (30), (34), and (40). Next we show that the special structures of these systems of linear equalities can be leveraged to solve them more efficiently. We will demonstrate our technique on the linear system (34). Extensions to (30) and (40) are straightforward.

The matrix $\mathbf{K}$ in (34) is a $(l+n-m-r+1+w)$-dimensional square matrix, where $r = \mathbf{rank}\,(\mathbf{DB})$ and w is the cardinality of the current inactive set, i.e. number of rows of $\mathbf{W}$. Only the matrix $\mathbf{W}$ changes from one iteration to the next – all the other elements of $\mathbf{K}$ remain fixed. This fact can be leveraged as follows.

1. The equality $\mathbf{W}\mathbf{d}_\lambda = \mathbf{0}$ sets the components of $\mathbf{d}_\lambda$ corresponding to the current inactive set to zero. Removing these variables and dropping the corresponding rows of $\mathbf{K}$ reduces the dimension of $\mathbf{K}$ to $l+n-m-r+1$. Thus, this simple operation ensures that the size of the linear equations remains independent of the cardinality of the inactive set.

2. Let
$$\tilde{\mathbf{d}} = \begin{bmatrix}\mathbf{d}_\lambda \\ \bar{\boldsymbol{\tau}} \\ \bar{\eta}\end{bmatrix}, \quad \tilde{\mathbf{K}} = \begin{bmatrix}\theta\gamma\mathbf{L}^T\mathbf{L} & \mathbf{M}^T & \mathbf{L}^T\mathbf{e} \\ \mathbf{M} & \mathbf{0} & \mathbf{0} \\ \mathbf{e}^T\mathbf{L} & \mathbf{0} & \mathbf{0}\end{bmatrix} \in \mathbf{R}^{(l+n-m-r+1)\times(l+n-m-r+1)}, \tag{41}$$
$\mathbf{B}_1$ be any orthonormal basis for row space of $\tilde{\mathbf{K}}$, and $\mathbf{B}_2$ be any orthonormal basis for the nullspace $\mathcal{N}(\tilde{\mathbf{K}})$. Then, $\tilde{\mathbf{d}} = \mathbf{B}_1\boldsymbol{\mu} + \mathbf{B}_2\boldsymbol{\zeta}$, where $\boldsymbol{\mu} \in \mathbf{R}^{r_K}$, $\boldsymbol{\zeta} \in \mathbf{R}^{l+n-m-r+1-r_K}$, and $r_K = \mathbf{rank}(\tilde{\mathbf{K}})$. An SVD-based argument similar to the one in Section 2.1 (detailed in Appendix C) shows that the dimension of $\mathbf{K}$ can be reduced to $l+n-m-r+1-r_K+w$.

These observations suggest that one can speed up FindOpt as follows: If $(l+n-m-r+1) < (l+n-m-r+1-r_K+w)$, i.e. if $w > r_K$, solve (34) using the first dimension reduction technique; otherwise, use the second dimension reduction.

In each iteration either new rows are added to $\mathbf{W}$ or some of the rows of $\mathbf{W}$ are dropped. Since every row of $\mathbf{W}$ is a row of an identity matrix, one can suitably adapt the revised simplex method [5] to efficiently update the iterates. For example, adding a new row to $\mathbf{W}$ forces an entry of $\mathbf{d}_\lambda$ to be equal to zero, i.e. a variable leaves the basis, and introduces a new variable through $\bar{\boldsymbol{\rho}}$, i.e. a variable enters the basis. This process, although requiring a careful bookkeeping of variables and bases, is fairly straightforward.

We conclude this section with the following finite convergence result.

Lemma 6 *The* ActiveSet *algorithm terminates after a finite number of iterations.*

Proof: Let $\mathcal{A}_j$, $j \geq 1$, denote the active set on the j-th call to the procedure FindMultipliers. Since every iteration of ActiveSet strictly improves the objective value of (22), it follows that $\mathcal{A}_{j_1} \neq \mathcal{A}_{j_2}$ for all $j_1 \neq j_2$. Since the size of the active set can only increase between successive calls to FindMultipliers, it follows that ActiveSet terminates after, at most, $l2^l$ iterations, where l is number of inequality constraints in the single-cone SOCP (1). ■

4 Recovering an optimal solution

Let $\boldsymbol{\lambda}^*$ denote the solution returned by LagrangeDual, i.e. $\boldsymbol{\lambda}^*$ is optimal for (22). Set $\boldsymbol{\xi}^* = \mathbf{U}_0\boldsymbol{\Sigma}_0^{-1}\mathbf{V}_0^T\mathbf{B}^T(\mathbf{f} - \mathbf{E}^T\boldsymbol{\lambda}^*)$, and using Lemma (2) obtain the closed form optimal solution $\mathbf{y}^*$ to $\bar{q}(\xi^*)$ defined in (10). Then all $\mathbf{x}^*$ satisfying

$$\mathbf{x}^* = \mathbf{x}_0 + \mathbf{B}\mathbf{z}^* = \mathbf{x}_0 + \mathbf{B}\mathbf{V}_0\boldsymbol{\Sigma}_0^{-1}\mathbf{U}_0^T(\mathbf{y}^* - \mathbf{D}\mathbf{x}_0) + \mathbf{B}\mathbf{V}_1\mathbf{t},$$

where $\mathbf{x}_0 = \mathbf{H}^T(\mathbf{H}\mathbf{H}^T)^{-1}\mathbf{g}$, $\mathbf{t} \in \mathbf{R}^{n-m-r}$, and $r = \mathbf{rank}(\mathbf{DB})$, are optimal for (1). Thus, if $\mathbf{V}_1 \neq \emptyset$, i.e. $\mathbf{rank}(\mathbf{DB}) \neq n-m$, the optimal solution is not unique; in fact, an entire affine space is optimal.

5 Computational experiments

In this section we discuss the computational performance of the LagrangeDual algorithm on special classes of single-cone SOCPs that arise in the context of robust optimization.

Consider the following LP

$$\begin{array}{rl} \min & \mathbf{c}^T\mathbf{z}, \\ \text{subject to} & \mathbf{A}\mathbf{z} = \mathbf{b}, \\ & \mathbf{z} \geq \mathbf{0}, \end{array} \tag{42}$$

where $\mathbf{c}, \mathbf{z} \in \mathbf{R}^{\bar{n}}$, $\mathbf{A} \in \mathbf{R}^{\bar{m}\times\bar{n}}$, and $\mathbf{b} \in \mathbf{R}^{\bar{m}}$. Suppose the constraint matrix $(\mathbf{A}, \mathbf{b})$ is known exactly; however, the cost vector $\mathbf{c}$ is uncertain and is only known to lie within an ellipsoidal uncertainty set $\mathcal{S}$ given by

$$\mathcal{S} = \{\mathbf{c} = \mathbf{c}_0 + \mathbf{P}^T\boldsymbol{\alpha} : \boldsymbol{\alpha} \in \mathbf{R}^s, \boldsymbol{\alpha}^T\boldsymbol{\alpha} \leq 1\}.$$

We will call (42) an LP with uncertain cost. Such an LP is a special case of a more general class of uncertain LPs where the constraints are also uncertain [2, 3].

Let $f(\mathbf{z}) = \max_{\mathbf{c}\in\mathcal{S}}\{\mathbf{c}^T\mathbf{z}\}$ denote the worst case cost of the decision $\mathbf{z}$. Then we have that

$$f(\mathbf{z}) = \mathbf{c}_0^T\mathbf{z} + \max_{\{\boldsymbol{\alpha}:\boldsymbol{\alpha}^T\boldsymbol{\alpha}\leq 1\}} \left\{\boldsymbol{\alpha}^T\mathbf{P}\mathbf{z}\right\} = \mathbf{c}_0^T\mathbf{z} + \|\mathbf{P}\mathbf{z}\|.$$

The robust counterpart of the uncertain LP is defined as follows [2, 3].

$$\begin{array}{rl} \min & \mathbf{c}_0^T\mathbf{z} + \|\mathbf{P}\mathbf{z}\|, \\ \text{subject to} & \mathbf{A}\mathbf{z} = \mathbf{b}, \\ & \mathbf{z} \geq \mathbf{0}. \end{array} \tag{43}$$

By defining,

$$\mathbf{x} = \begin{bmatrix} \mathbf{z} \\ y_0 \\ \mathbf{y} \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} \mathbf{A} & 0 & \mathbf{0} \\ \mathbf{P} & 0 & -\mathbf{I} \end{bmatrix}, \quad \mathbf{E} = \begin{bmatrix} \mathbf{I} & 0 & \mathbf{0} \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} \mathbf{0} & 1 & \mathbf{0} \\ \mathbf{0} & 0 & \mathbf{I} \end{bmatrix}, \quad \mathbf{g} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} \mathbf{c}_0 \\ 1 \\ \mathbf{0} \end{bmatrix},$$

it is easy to see that (43) can be reformulated as a single-cone SOCP. The constant γ for problems of the form (43) is given by $\gamma = 0.5$. Thus, we are in a position to use the LagrangeDual algorithm.

All the systems of linear equations encountered during the course of the LagrangeDual algorithm were solved using the MATLAB® function `mldivide` and all the computations were carried out using MATLAB R13 on a PC with a Pentium M (1.50GHz) and 512 MB of RAM. For moderate values of $(\bar{n}, \bar{m})$ the LP that defines $\boldsymbol{\mu}^{(1)}$ (see Figure 1) was solved using SeDuMi. For large $(\bar{n}, \bar{m})$ $\boldsymbol{\mu}^{(1)}$ was computed using the simplex algorithm.

In the first set of experiments, the LP instances were randomly generated. In particular, the entries of matrix $\mathbf{A}$ and the cost vector $\mathbf{c}_0$ were drawn independently at random according to the uniform distribution on the unit $[0, 1]$ interval. To ensure feasibility of (42), the vector $\mathbf{b}$ was set to $\mathbf{b} = \mathbf{A}\mathbf{w}$, where each component of the vector $\mathbf{w}$ was generated independently at random from the uniform distribution on $[0, 1]$. The matrix $\mathbf{P}$ defining the uncertainty set $\mathcal{S}$ was set equal to the $\bar{n}$-dimensional identity matrix and for each $(\bar{n}, \bar{m})$ pair, we generated 50 random instances.

Table 1 compares the running time of LagrangeDual to that of SeDuMi on the randomly generated instances. Column 3 lists the average of the ratio of running time t_{sed} of SeDuMi to running time t_{alg} of LagrangeDual and Column 4 lists the average of the ratio of t_{sed} to the running time t_{act} of ActiveSet. Note that the running time of ActiveSet is equal to the difference between the running time of LagrangeDual and the time t_{init} required to compute the initial Lagrange multipliers $(\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)})$. The time t_{init} is listed in Column 6. Columns 5 and 7 list, respectively, the average running time t_{alg} of LagrangeDual and the average number of iterations of the `while` loop in ActiveSet.

From the results displayed in Table 1, it is clear that the performance of the LagrangeDual algorithm (including the time spent to obtain the initial Lagrange multipliers) is superior to the SeDuMi when

(i) either the number of variables $\bar{n}$ is small,

(ii) and/or the ratio of the number of constraints to the number of variables $\bar{m}/\bar{n} \leq 0.1$ or $\bar{m}/\bar{n} \geq 0.5$.

The data in Column 4 of Table 1 implies that the performance of LagrangeDual algorithm is superior to the SeDuMi when the time spent to obtain the initial Lagrange multipliers is excluded. This observation suggests that the performance of LagrangeDual is likely to improve if it is initialized using a more efficient LP-solver.

Since network flow problems are a natural class of linear programs where the number of variables is large but the number of constraints is reasonably small, next we tested LagrangeDual on random instances of the uncertain min-cost flow problems. The random networks were generated using the network generator developed by Goldberg [8]. Results are averaged over 10 runs for each pair $(\bar{n}, \bar{m})$.

Table 2 displays the results for the randomly generated network matrices. In order to be consistent with the previous set of results, we continue to denote the number of variables by $\bar{n}$ and

$\bar{n}$	$\bar{m}$	t_{sed}/t_{alg}	t_{sed}/t_{act}	t_{alg}	t_{init}	iterations
100	20	2.5880	14.0623	0.2225	0.1892	6.0400
100	40	2.1039	10.7260	0.2767	0.2230	7.7000
100	60	2.1201	6.9539	0.3435	0.2247	6.6200
100	80	2.8624	9.4862	0.3019	0.2048	3.4600
200	20	2.0144	9.8526	0.8085	0.6560	10.4400
200	50	1.0479	2.0670	1.4371	0.7446	20.3200
200	80	1.4784	2.7523	1.5546	0.7540	22.2800
200	100	1.5296	2.4418	1.7406	0.6909	32.4500
200	125	1.6418	2.2056	1.9302	0.6854	37.5400
200	150	2.6301	4.1392	1.5328	0.6821	20.5200
200	175	2.9156	4.9983	1.3913	0.6639	12.4600
300	30	1.3012	7.8696	2.4029	2.0261	13.1200
300	60	0.7355	1.5558	4.3684	2.1425	26.5800
300	90	0.8926	1.8139	4.7719	2.1576	32.7200
300	120	1.0898	2.3107	5.1750	2.1817	40.3400
300	150	1.0190	1.8005	8.3897	2.0995	75.5400
300	180	1.4281	2.9468	6.4953	2.1301	54.7800
300	210	1.3864	2.3917	5.6592	2.0499	47.4200
300	240	1.8482	3.8423	6.5621	2.1846	46.0600
300	270	2.4261	6.3807	6.6542	2.3099	36.1400
500	50	1.2789	9.5464	10.6301	9.8295	16.2400
500	100	0.6193	1.3513	18.0874	8.9554	34.3800
500	200	0.8238	1.5742	23.3625	8.9827	56.9000
500	300	1.0865	1.9030	26.2689	8.8647	74.1200
500	400	1.4921	2.6069	29.8203	9.0223	76.0400
1000	100	1.1382	9.3120	88.6851	74.3340	40.2400
1000	250	0.6622	1.1354	199.0423	76.2742	84.2800
1000	500	1.0170	1.5283	249.0415	75.6579	121.1200
1000	750	1.5285	2.5135	214.9937	75.5053	92.3800
1500	150	1.1910	14.0390	369.9276	345.9236	46.1400
1500	500	0.7092	1.0651	867.7425	283.1938	138.8200
1500	1000	0.9616	1.2590	1259.9096	280.9712	230.9400
1500	1250	1.5668	2.1655	1024.6152	269.9120	158.5800
2000	200	1.1245	10.2186	604.3456	513.5234	55.2200
2000	500	∞	∞	3456.4591	2183.3089	208.3400
5000	500	∞	∞	4067.1300	2967.4054	405.1200

Table 1: Running time of SeDuMi and the LagrangeDual algorithm

$\bar{n}$	$\bar{m}$	$t_{\text{sed}}/t_{\text{alg}}$	t_{alg}	iterations
1000	100	4.2342	20.5434	53.5000
1000	150	1.7765	37.9068	65.4000
1500	150	3.6572	55.1678	74.8000
1500	250	3.0398	64.8549	88.3000
2000	330	2.4105	817.8575	94.1000

Table 2: Running time of SeDuMi and the LagrangeDual algorithm on networks

the number of constraints by $\bar{m}$. Thus, $\bar{n}$ and $\bar{m}$ denote, respectively, the number of *arcs* and the number of *nodes* in the network. As before, Column 3 lists the average of the ratio of running time t_{sed} of SeDuMi to running time t_{alg} of LagrangeDual. Column 4 and 5 list respectively the average running time of LagrangeDual and the average number of iterations of the `while` loop in ActiveSet. Since the version of LagrangeDual that we implemented did not take advantage of sparsity, in this set of experiments we did not allow SeDuMi to leverage sparsity. From the results of our computational experiments it appears that LagrangeDual is faster than SeDuMi on relatively dense networks. Also, for large networks $\bar{n} \approx 5000$ the SeDuMi failed to solve the problem but LagrangeDual did not have any trouble converging.

We also compared the performance of LagrangeDual with that of SeDuMi on some of the small problems from the NETLIB LP [14] library. All the LP instances were converted to canonical form LPs (42). To define the uncertainty set $\mathcal{S}$, we took the nominal cost vector $\mathbf{c}_0$ as given by the NETLIB LP library, assumed that only the non-zero elements of $\mathbf{c}_0$ are uncertain, and then defined the matrix $\mathbf{P}$ accordingly. In these experiments the performance of SeDuMi was superior to that of LagrangeDual. This is not surprising given that for most of these small problems the ratio $\bar{m}/\bar{n}$ (after the problem was converted to the canonical form) was between 0.1 and 0.6.

Before concluding, we would like to mention that these experiments are biased in favor of SeDuMi. As mentioned in [15] (the version updated for SeDuMi 1.05) SeDuMi "takes full advantage of sparsity" which increases its speed considerably and it uses a dense column factorization proposed in [10]. In addition, most of the subroutines of the SeDuMi are written in C code. On the other hand, the LagrangeDual algorithm was implemented using only MATLAB functions, without any special treatment of sparsity and dense columns.

References

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

[2] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Math. Oper. Res.*, 23(4):769–805, 1998.

[3] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Oper. Res. Lett.*, 25(1):1–13, 1999.

[4] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization*. SIAM, Philadelphia, PA, 2001.

[5] D. Bertsimas and J. N. Tsitsiklis. *Introduction to Linear Optimization*. Athena Scientific, 1997.

[6] S. P. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

[7] N. Gartner, C.J. Messer, and A.K. Rathi. Traffic flow theory: A state of the art report. Technical report, Turner Fairbank Highway Research Center (TFHRC), 1997. Available at http://www.tfhrc.gov/its/tft/tft.htm.

[8] A. Goldberg. C-code for a random network generator. Available via anonymous ftp from ftp://dimacs.rutgers.edu/pub/netflow/generators/network/grid-on-torus/goto.c.

[9] D. Goldfarb and G. Iyengar. Robust portfolio selection problems. *Math. Oper. Res.*, 28(1):1–37, 2003.

[10] D. Goldfarb and K. Scheinberg. A product-form cholesky factorization method for handling dense columns in interior point methods for linear programming. *Mathematical Programming*, 99:1–34, 2004.

[11] W. H. Greene. *Econometric analysis.* Macmillan Pub. Co., 1990.

[12] Y. Nesterov and A. Nemirovski. *Interior-point polynomial algorithms in convex programming.* SIAM, Philadelphia, 1993.

[13] A. Ruszczynski and A. Shapiro, editors. *Stochastic Programming.* Handbook in Operations Research and Management Science. Elsevier, 2003.

[14] The NETLIB LP Test Problem Set. Available at http://cuter.rl.ac.uk/cuter-www/Problems/netlib.shtml.

[15] J. Sturm. Using SeDuMi 1.02, a matlab toolbox for optimization over symmetric cones. *Optim. Methods and Software*, 11-12:625–653, 1999.

A Proofs

A.1 Structural results for single-cone SOCPs

Lemma 7 *Suppose* $\mathbf{A} \in \mathbf{R}^{m\times n}$ *has full row rank and there exists* $\mathbf{d} \succeq \mathbf{0}$, $\mathbf{d} \neq \mathbf{0}$ *such that* $\mathbf{A}\mathbf{d} = \mathbf{0}$. *Let* $\mathbf{a}$ *denote the first column of the matrix* $\mathbf{A}$. *Then*

(a) $\gamma = \frac{1}{2} - \mathbf{a}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{a} \in [0, 0.5]$.

(b) for all $\mathbf{d}$ *such that* $\mathbf{A}\mathbf{d} = \mathbf{0}$, *we have* $\|\mathbf{d}\|^2 \geq \frac{2}{1+2\gamma}(\mathbf{e}^T\mathbf{d})^2$.

(c) $\gamma > 0 \Leftrightarrow \exists \mathbf{d} \succ \mathbf{0} : \mathbf{A}\mathbf{d} = \mathbf{0}$.

(d) $\gamma = 0 \Leftrightarrow \{\mathbf{d} : \mathbf{A}\mathbf{d} = \mathbf{0}, \mathbf{d} \succeq \mathbf{0}\} = \big\{\beta\big(\mathbf{e} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{a}\big) : \beta \geq 0\big\}$.

Proof: Partition the matrix $\mathbf{A}$ as $\mathbf{A} = [\mathbf{a}, \bar{\mathbf{A}}]$. By scaling, we can assume that $\mathbf{d} = (1; \bar{\mathbf{d}}) \succeq \mathbf{0}$. Since $\mathbf{d} \succeq \mathbf{0}$, $\|\bar{\mathbf{d}}\| \leq 1$. Then,

$$\begin{aligned}
\mathbf{A}\mathbf{A}^T - 2\mathbf{a}\mathbf{a}^T &= \mathbf{a}\mathbf{a}^T + \bar{\mathbf{A}}\bar{\mathbf{A}}^T - 2\mathbf{a}\mathbf{a}^T, \\
&= \bar{\mathbf{A}}\bar{\mathbf{A}}^T - \mathbf{a}\mathbf{a}^T, \\
&= \bar{\mathbf{A}}\bar{\mathbf{A}}^T - \bar{\mathbf{A}}\bar{\mathbf{d}}\bar{\mathbf{d}}^T\bar{\mathbf{A}}^T, &(44)\\
&= \bar{\mathbf{A}}(\mathbf{I} - \bar{\mathbf{d}}\bar{\mathbf{d}}^T)\bar{\mathbf{A}}^T \succeq \mathbf{0}, &(45)
\end{aligned}$$

where (44) follows from the fact that $\mathbf{Ad} = \mathbf{a} + \bar{\mathbf{A}}\bar{\mathbf{d}} = \mathbf{0}$, and (45) follows from the fact that $\|\bar{\mathbf{d}}\| \leq 1$. Define

$$\mathbf{M} = \begin{bmatrix} \frac{1}{2} & \mathbf{a}^T \\ \mathbf{a} & \mathbf{AA}^T \end{bmatrix}.$$

Since $\frac{1}{2} > 0$ and the Schur complement of $\frac{1}{2}$ in $\mathbf{M}$ is $\mathbf{AA}^T - 2\mathbf{aa}^T \succeq \mathbf{0}$, it follows that $\mathbf{M} \succeq \mathbf{0}$. Since $\mathbf{A}$ has full row rank, it follows that $\mathbf{AA}^T \succ \mathbf{0}$, and the matrix $\mathbf{M} \succeq \mathbf{0}$ if, and only if, the Schur complement of $\mathbf{AA}^T$

$$\gamma = \frac{1}{2} - \mathbf{a}^T(\mathbf{AA}^T)^{-1}\mathbf{a} \geq 0.$$

Since $(\mathbf{AA}^T)^{-1} \succ \mathbf{0}$, it follows that $\gamma = \frac{1}{2} - \mathbf{a}^T(\mathbf{AA}^T)^{-1}\mathbf{a} \leq \frac{1}{2}$. This establishes part (a).

To establish the other results, consider the following minimum norm problem

$$\begin{array}{rl} \min & \|\mathbf{d}\|^2, \\ \text{subject to} & \mathbf{Ad} = \mathbf{0}, \\ & \mathbf{e}^T\mathbf{d} = 1. \end{array} \tag{46}$$

The optimal solution $\mathbf{d}^*$ and the optimal value v^* of (46) can be obtained easily via the Lagrange multipliers technique, and is given by

$$\mathbf{d}^* = \frac{2}{1+2\gamma}\Big(\mathbf{e} - \mathbf{A}^T(\mathbf{AA}^T)^{-1}\mathbf{a}\Big), \quad v^* = \frac{2}{1+2\gamma}.$$

Thus, it follows that for all $\mathbf{d}$ such that $\mathbf{Ad} = \mathbf{0}$, we have $\|\mathbf{d}\|^2 \geq \frac{2}{1+2\gamma}(\mathbf{e}^T\mathbf{d})^2$.

Since there exists $\mathbf{d} = (1; \bar{\mathbf{d}}) \succeq \mathbf{0}$ with $\mathbf{Ad} = \mathbf{0}$, there exists a $\mathbf{d} \succ \mathbf{0}$ with $\mathbf{Ad} = \mathbf{0}$ if, and only if, $v^* < 2$, i.e. $\gamma > 0$. Moreover when $\gamma = 0$, $\big\{\mathbf{d} : \mathbf{Ad} = \mathbf{0}, \mathbf{d} \succeq \mathbf{0}\big\} = \{\beta\mathbf{d}^* : \beta \geq 0\}$. ■

Lemma 8 *Suppose $\mathbf{A} \in \mathbf{R}^{m\times n}$ has full row rank and consider the following SOCP*

$$\begin{array}{rl} \min & \boldsymbol{\xi}^T\mathbf{y}, \\ \text{subject to} & \mathbf{Ay} = \mathbf{b}, \\ & \mathbf{y} \succeq \mathbf{0}, \end{array} \tag{47}$$

Let $\mathbf{a}$ denote the first column of the matrix $\mathbf{A}$, and let $\gamma = \frac{1}{2} - \mathbf{a}^T(\mathbf{AA}^T)^{-1}\mathbf{a} \neq 0$. Then we have the following.

(i) The dual of (47) is strictly feasible for all $\gamma < 0$.

(ii) When $\gamma > 0$, the dual of (47) is strictly feasible if, and only if, $\mathbf{e}^T\mathbf{P}\boldsymbol{\xi} > 0$ and $(\mathbf{e}^T\mathbf{P}\boldsymbol{\xi})^2 - \gamma\|\mathbf{P}\boldsymbol{\xi}\|^2 > 0$, where $\mathbf{P} = \mathbf{I} - \mathbf{A}^T(\mathbf{AA}^T)^{-1}\mathbf{A}$ denotes the orthogonal projector operator onto $\mathcal{N}(\mathbf{A})$.

Proof: The dual of (47) is given by

$$\begin{array}{rl} \max & \mathbf{b}^T\boldsymbol{\mu}, \\ \text{subject to} & \boldsymbol{\xi} - \mathbf{A}^T\boldsymbol{\mu} \succeq \mathbf{0}. \end{array}$$

Since $\mathbf{A}$ has full row rank, $\boldsymbol{\xi}$ can be written as $\boldsymbol{\xi} = \mathbf{P}\boldsymbol{\xi} + \mathbf{A}^T\mathbf{w}$ for some $\mathbf{w} \in \mathbf{R}^m$. Thus, it follows that there exists a $\boldsymbol{\mu}$ such that $\boldsymbol{\xi} - \mathbf{A}^T\boldsymbol{\mu} \succ \mathbf{0}$ if, and only if, there exists a $\boldsymbol{\mu}$ such that $\mathbf{P}\boldsymbol{\xi} + \mathbf{A}^T\boldsymbol{\mu} \succ \mathbf{0}$.

From the definition of the Lorentz cone, it follows that there exists a $\boldsymbol{\mu}$ such that $\mathbf{P}\boldsymbol{\xi}+\mathbf{A}^T\boldsymbol{\mu} \succ \mathbf{0}$ if, and only if, the optimal value of

$$\begin{array}{rl} \min & \|\alpha\mathbf{P}\boldsymbol{\xi} + \mathbf{A}^T\boldsymbol{\mu}\|^2, \\ \text{subject to} & \alpha\mathbf{e}^T\mathbf{P}\boldsymbol{\xi} + \mathbf{a}^T\boldsymbol{\mu} = 1, \\ & \alpha \geq 0, \end{array} \tag{48}$$

is less than 2.

First consider the case $\gamma < 0$. Note that the solution $\alpha = 0$, $\boldsymbol{\mu} = \frac{(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{a}}{\mathbf{a}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{a}}$ is feasible to (48) with the objective function value $\frac{1}{\mathbf{a}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{a}} = \frac{2}{1-2\gamma} < 2$.

If $\gamma > 0$, then the first part of Lemma 2 shows that $\boldsymbol{\xi}$ has to satisfy $\mathbf{e}^T\mathbf{P}\boldsymbol{\xi} \geq 0$ and $(\mathbf{e}^T\mathbf{P}\boldsymbol{\xi})^2 - \gamma\|\mathbf{P}\boldsymbol{\xi}\|^2 \geq 0$. Otherwise, (47) becomes unbounded and therefore, by the Weak Duality Lemma for SOCPs [1], its dual is infeasible.

The rest of the analysis is very similar to the one used in the proof of Lemma 7 and is left to the reader. ■

A.2 Proof of Lemma 2

By definition, $\boldsymbol{\xi} \in \mathcal{D}_{\bar{q}}$ if, and only if, (10) is bounded, or equivalently the optimal value of the homogeneous problem

$$\begin{array}{rl} \min & \boldsymbol{\xi}^T\mathbf{d}, \\ \text{subject to} & \mathbf{A}\mathbf{d} = \mathbf{0}, \\ & \mathbf{d} \succeq \mathbf{0}, \end{array} \tag{49}$$

is non-negative. Without loss of generality, we assume that $\boldsymbol{\xi} \in \mathcal{N}(\mathbf{A})$. Otherwise, $\boldsymbol{\xi}$ can be decomposed as $\boldsymbol{\xi} = \mathbf{P}\boldsymbol{\xi} + \boldsymbol{\xi}_1$ where $\mathbf{P}\boldsymbol{\xi} \in \mathcal{N}(\mathbf{A})$ and $\boldsymbol{\xi}_1$ belongs to the row space of $\mathbf{A}$ (the space orthogonal to $\mathcal{N}(\mathbf{A})$). Since $\mathbf{A}\mathbf{d} = \mathbf{0}$ implies $\boldsymbol{\xi}_1^T\mathbf{d} = 0$, we can drop $\boldsymbol{\xi}_1$ from the objective.

Lemma 7 part (b) in Appendix A.1 establishes that $\|\mathbf{d}\|^2 \geq \frac{2}{1+2\gamma}(\mathbf{e}^T\mathbf{d})^2$ for all $\mathbf{d}$ such that $\mathbf{A}\mathbf{d} = \mathbf{0}$. Since $\mathbf{d} \succeq \mathbf{0}$ implies that $2(\mathbf{e}^T\mathbf{d})^2 \geq \|\mathbf{d}\|^2$, it follows that $\mathbf{d} = \mathbf{0}$ is the only feasible solution to (49) when $\gamma < 0$. Hence, $\mathcal{D}_{\bar{q}} = \mathbf{R}^p$.

Next, suppose $\gamma \geq 0$. Then (49) is bounded if, and only if,

$$\begin{array}{rl} \min & \boldsymbol{\xi}^T\mathbf{d}, \\ \text{subject to} & \mathbf{A}\mathbf{d} = \mathbf{0}, \\ & \mathbf{e}^T\mathbf{d} = 1, \\ & \mathbf{d}^T\mathbf{d} \leq 2, \end{array} \tag{50}$$

has a non-negative optimal value.

The Lagrangian of (50) is given by

$$\mathcal{L} = \boldsymbol{\xi}^T\mathbf{d} - \widehat{\boldsymbol{\tau}}^T\mathbf{A}\mathbf{d} - \widehat{\delta}(\mathbf{e}^T\mathbf{d} - 1) + \widehat{\beta}(\mathbf{d}^T\mathbf{d} - 2),$$

where $\widehat{\beta} \geq 0$. Setting the derivative $\nabla\mathcal{L} = 0$ we get

$$\mathbf{d} = -\beta\boldsymbol{\xi} + \mathbf{A}^T\boldsymbol{\tau} + \delta\mathbf{e},$$

where β, $\boldsymbol{\tau}$, and δ are rescaled values of $\widehat{\beta}$, $\widehat{\boldsymbol{\tau}}$, and $\widehat{\delta}$; however, $\beta \geq 0$ still holds. Since $\mathbf{A}\boldsymbol{\xi} = \mathbf{0}$, the constraint $\mathbf{A}\mathbf{d} = \mathbf{0}$ yields

$$\boldsymbol{\tau} = -\delta(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{e} = -\delta(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{a}.$$

Next, the constraint $\mathbf{e}^T\mathbf{d} = 1$ implies that

$$1 = -\beta\mathbf{e}^T\boldsymbol{\xi} + \delta\mathbf{e}^T(\mathbf{e} - \mathbf{A}^T(\mathbf{A}\mathbf{A})^{-1}\mathbf{a}) = -\beta\mathbf{e}^T\boldsymbol{\xi} + \delta(1 - \mathbf{a}^T(\mathbf{A}\mathbf{A})^{-1}\mathbf{a}) = -\beta\mathbf{e}^T\boldsymbol{\xi} + \delta\left(\frac{1+2\gamma}{2}\right).$$

Thus,

$$\mathbf{d} = -\beta\boldsymbol{\xi} + \left(\frac{2}{1+2\gamma}\right)(1+\beta\mathbf{e}^T\boldsymbol{\xi})\Big(\mathbf{e} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{a}\Big).$$

From Lemma 7 part (a) we have $\gamma \in [0, 0.5]$, therefore $\mathbf{d}$ is well-defined.

Since (50) has a linear objective and its feasible set is the intersection of an affine set with a Euclidean ball, there exists an optimal solution to (50) that satisfies $\mathbf{d}^T\mathbf{d} = 2$. It is easy to see this when the matrix $[\mathbf{A}; \mathbf{e}^T]$ does not have full column rank. When $[\mathbf{A}; \mathbf{e}^T]$ has full column rank, the system $\mathbf{A}\mathbf{d} = \mathbf{0}, \mathbf{e}^T\mathbf{d} = 1$ admits the unique solution $\tilde{\mathbf{d}} = \frac{2}{1+2\gamma}(\mathbf{e} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{a})$ which implies $\{\mathbf{d} : \mathbf{A}\mathbf{d} = 0, \mathbf{d} \succeq \mathbf{0}\} = \big\{t\big(\mathbf{e} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{a}\big) : t \geq 0\big\}$. Then, by Lemma 7 part (d) in Appendix A.1, we have $\gamma = 0$. Therefore, $\tilde{\mathbf{d}}^T\tilde{\mathbf{d}} = \frac{2}{1+2\gamma} = 2$. Simplifying the constraint $\mathbf{d}^T\mathbf{d} = 2$, we get

$$\beta^2\left(\|\boldsymbol{\xi}\|^2 - \frac{2}{1+2\gamma}(\mathbf{e}^T\boldsymbol{\xi})^2\right) = \frac{4\gamma}{1+2\gamma}.$$

Since $\mathbf{A}\boldsymbol{\xi} = \mathbf{0}$, Lemma 7 part (b) implies that $\|\boldsymbol{\xi}\|^2 \geq \frac{2}{1+2\gamma}(\mathbf{e}^T\boldsymbol{\xi})^2$. Therefore, we only have to consider the following two cases.

(i) $(\mathbf{e}^T\boldsymbol{\xi})^2 = \frac{1+2\gamma}{2}\|\boldsymbol{\xi}\|^2$.

Suppose $\mathbf{e}^T\boldsymbol{\xi} = \sqrt{\frac{1+2\gamma}{2}}\|\boldsymbol{\xi}\|$. Then $\boldsymbol{\xi} \succeq \mathbf{0}$, and the optimal value of (50) is non-negative, or equivalently (49) is bounded. Next, suppose $\mathbf{e}^T\boldsymbol{\xi} = -\sqrt{\frac{1+2\gamma}{2}}\|\boldsymbol{\xi}\|$. Then $\mathbf{d} = -\boldsymbol{\xi} \succeq \mathbf{0}$, and $\mathbf{d}^T\boldsymbol{\xi} = -\|\boldsymbol{\xi}\|^2 < 0$. Therefore, (49) is unbounded.

(ii) $(\mathbf{e}^T\boldsymbol{\xi})^2 < \frac{1+2\gamma}{2}\|\boldsymbol{\xi}\|^2$.

In this case $\beta = \sqrt{\frac{4\gamma}{(1+2\gamma)\|\boldsymbol{\xi}\|^2 - 2(\mathbf{e}^T\boldsymbol{\xi})^2}}$. And (50) has a non-negative optimal value if, and only if,

$$\begin{aligned} 0 &\leq \boldsymbol{\xi}^T\mathbf{d}, \\ &= -\beta\|\boldsymbol{\xi}\|^2 + \left(\frac{2}{1+2\gamma}\right)(1+\beta\mathbf{e}^T\boldsymbol{\xi})(\mathbf{e}^T\boldsymbol{\xi}), \\ &= -\frac{\beta}{1+2\gamma}\Big((1+2\gamma)\|\boldsymbol{\xi}\|^2 - 2(\mathbf{e}^T\boldsymbol{\xi})^2\Big) + \frac{2(\mathbf{e}^T\boldsymbol{\xi})}{1+2\gamma}. \end{aligned}$$

Substituting the value of β and simplifying we get

$$\mathbf{e}^T\boldsymbol{\xi} \geq 0, \quad (\mathbf{e}^T\boldsymbol{\xi})^2 \geq \gamma\|\boldsymbol{\xi}\|^2.$$

Since, as we discussed above, assuming $\boldsymbol{\xi} \in \mathcal{N}(\mathbf{A})$ is equivalent to replacing $\boldsymbol{\xi}$ by $\mathbf{P}\boldsymbol{\xi}$, the result follows.

For the second part of Lemma 2, first consider the case $\gamma \neq 0$, or equivalently $\mathbf{A}\mathbf{R}\mathbf{A}^T$ is non-singular [1]. Using the results of the first part of this Lemma, one can be prove that (see Lemma 8 in Appendix A.1) if $\gamma < 0$, then the dual of (49) is strictly feasible for any $\boldsymbol{\xi} \in \mathbf{R}^p$ and when $\gamma > 0$ the dual of (49) is strictly feasible if, and only if, $\mathbf{e}^T\mathbf{P}\boldsymbol{\xi} \geq 0$ and $(\mathbf{e}^T\mathbf{P}\boldsymbol{\xi})^2 - \gamma\|\mathbf{P}\boldsymbol{\xi}\|^2 > 0$; and from [1] Section 5 it follows that when the dual is strictly feasible

$$\bar{q}(\boldsymbol{\xi}) = \frac{\sqrt{-\gamma(\mathbf{b}^T(\mathbf{ARA}^T)^{-1}\mathbf{b})}}{\gamma}\sqrt{(\mathbf{e}^T\mathbf{P}\boldsymbol{\xi})^2 - \gamma\|\mathbf{P}\boldsymbol{\xi}\|^2} + \boldsymbol{\xi}^T\mathbf{RA}^T(\mathbf{ARA}^T)^{-1}\mathbf{b}.$$

When the dual is not strictly feasible, i.e. $(\mathbf{e}^T\mathbf{P}\boldsymbol{\xi})^2 = \gamma\|\mathbf{P}\boldsymbol{\xi}\|^2$, choose $\widehat{\boldsymbol{\xi}} \in \mathcal{D}_{\bar{q}}$ such that the dual corresponding to $\widehat{\boldsymbol{\xi}}$ is strictly feasible. For $0 < \epsilon \leq 1$, let $\boldsymbol{\xi}_\epsilon = (1-\epsilon)\boldsymbol{\xi} + \epsilon\widehat{\boldsymbol{\xi}}$. Then we have two cases:

(i) $\mathbf{P}\boldsymbol{\xi} = \mathbf{0}$. In this case, $(\mathbf{e}^T\mathbf{P}\boldsymbol{\xi}_\epsilon)^2 - \gamma\|\mathbf{P}\boldsymbol{\xi}_\epsilon\|^2 = \epsilon^2\big((\mathbf{e}^T\mathbf{P}\widehat{\boldsymbol{\xi}})^2 - \gamma\|\mathbf{P}\widehat{\boldsymbol{\xi}}\|^2\big) > 0$.

(ii) $\mathbf{P}\boldsymbol{\xi} \neq \mathbf{0}$. In this case, $\gamma > 0$. Since $\mathbf{e}^T\mathbf{P}\boldsymbol{\xi} - \sqrt{\gamma}\|\mathbf{P}\boldsymbol{\xi}\|$ is a concave function of $\boldsymbol{\xi}$, it follows that

$$\mathbf{e}^T\mathbf{P}\boldsymbol{\xi}_\epsilon - \sqrt{\gamma}\|\mathbf{P}\boldsymbol{\xi}_\epsilon\| \geq \epsilon\big(\mathbf{e}^T\mathbf{P}\widehat{\boldsymbol{\xi}} - \sqrt{\gamma}\|\mathbf{P}\widehat{\boldsymbol{\xi}}\|\big) + (1-\epsilon)\big(\mathbf{e}^T\mathbf{P}\boldsymbol{\xi} - \sqrt{\gamma}\|\mathbf{P}\boldsymbol{\xi}\|\big) = \epsilon\big(\mathbf{e}^T\mathbf{P}\widehat{\boldsymbol{\xi}} - \sqrt{\gamma}\|\mathbf{P}\widehat{\boldsymbol{\xi}}\|\big) > 0.$$

Thus, the dual corresponding to $\boldsymbol{\xi}_\epsilon$ is always strictly feasible and

$$\bar{q}(\boldsymbol{\xi}_\epsilon) = f(\mathbf{P}\boldsymbol{\xi}_\epsilon) + \mathbf{v}^T\boldsymbol{\xi}_\epsilon.$$

Taking the limit as $\epsilon \downarrow 0$ establishes the result.

Next, consider the case $\gamma = 0$, or equivalently $\mathbf{ARA}^T$ is singular. Note that

$$\bar{q}(\boldsymbol{\xi}) = \boldsymbol{\xi}^T\mathbf{y}_0 + \widehat{q}(\mathbf{P}\boldsymbol{\xi}),$$

where $\mathbf{y}_0 = \mathbf{A}^T(\mathbf{AA}^T)^{-1}\mathbf{b}$, and

$$\widehat{q}(\mathbf{P}\boldsymbol{\xi}) = \begin{array}{rl} \min & (\mathbf{P}\boldsymbol{\xi})^T\mathbf{w}, \\ \text{subject to} & \mathbf{Aw} = \mathbf{0}, \\ & \mathbf{y}_0 + \mathbf{w} \succeq \mathbf{0}. \end{array} \tag{51}$$

The following are easy to check linear algebra facts:

(i) $\gamma = 0 \Rightarrow 2(\mathbf{e}^T\mathbf{y}_0)^2 \leq \|\mathbf{y}_0\|^2$.

(ii) $\gamma = 0 \Rightarrow 2(\mathbf{e}^T\mathbf{w})^2 \leq \|\mathbf{w}\|^2$, for all $\mathbf{w} \in \mathcal{N}(\mathbf{A})$. In particular, $\|\mathbf{P}\boldsymbol{\xi}\|^2 \geq 2(\mathbf{e}^T\mathbf{P}\boldsymbol{\xi})^2$.

We solve (51) by first scaling it to reduce it to a minimum norm problem and then optimizing over the scaling factor. Let $\mathbf{w}^*$ denote the optimal solution of (51) and let $\alpha^* = \mathbf{e}^T(\mathbf{y}_0 + \mathbf{w}^*)$. Then

$$\|\mathbf{w}^* + \mathbf{y}_0\|^2 = \|\mathbf{w}^*\|^2 + \|\mathbf{y}_0\|^2 \leq 2(\alpha^*)^2,$$

where the equality follows from the fact that $\mathbf{y}_0^T\mathbf{w}^* = \mathbf{b}^T(\mathbf{AA}^T)^{-1}\mathbf{Aw}^* = 0$. It follows that $\mathbf{w}^*$ is the optimal solution of

$$\begin{array}{rl} \min & (\mathbf{P}\boldsymbol{\xi})^T\mathbf{w}, \\ \text{subject to} & \mathbf{Aw} = 0, \\ & \mathbf{e}^T\mathbf{w} = \alpha - \mathbf{e}^T\mathbf{y}_0, \\ & \|\mathbf{w}\|^2 \leq 2\alpha^2 - \|\mathbf{y}_0\|^2, \end{array} \tag{52}$$

with α set equal to α^*. Using Lagrange multipliers, the optimal value of (52) is given by

$$(\mathbf{P}\boldsymbol{\xi})^T\mathbf{w}_\alpha = -\sqrt{\|\mathbf{P}\boldsymbol{\xi}\|^2 - 2(\mathbf{e}^T\mathbf{P}\boldsymbol{\xi})^2}\sqrt{4\alpha(\mathbf{e}^T\mathbf{y}_0) - \|\mathbf{y}_0\|^2 - 2(\mathbf{e}^T\mathbf{y}_0)^2} + 2(\mathbf{e}^T\mathbf{P}\boldsymbol{\xi})(\alpha - \mathbf{e}^T\mathbf{y}_0). \quad (53)$$

Differentiating this expression with respect to α we get

$$4\alpha^*(\mathbf{e}^T\mathbf{y}_0) - \|\mathbf{y}_0\|^2 - 2(\mathbf{e}^T\mathbf{y}_0)^2 = \left(\frac{\mathbf{e}^T\mathbf{y}_0}{\mathbf{e}^T\mathbf{P}\boldsymbol{\xi}}\right)^2 \left(\|\mathbf{P}\boldsymbol{\xi}\|^2 - 2(\mathbf{e}^T\mathbf{P}\boldsymbol{\xi})^2\right).$$

It is easy to check that $2(\alpha^*)^2 \geq \|\mathbf{y}_0\|^2$. Substituting α^* into (53) and simplifying we get

$$\begin{aligned}\bar{q}(\boldsymbol{\xi}) &= \boldsymbol{\xi}^T\mathbf{y}_0 + (\mathbf{P}\boldsymbol{\xi})^T\mathbf{w}_{\alpha^*}, \\ &= \boldsymbol{\xi}^T\mathbf{y}_0 + \left(\frac{\|\mathbf{y}_0\|^2 - 2(\mathbf{e}^T\mathbf{y}_0)^2}{2\mathbf{e}^T\mathbf{y}_0}\right)\mathbf{e}^T\mathbf{P}\boldsymbol{\xi} - \mathbf{e}^T\mathbf{y}_0\left(\frac{\|\mathbf{P}\boldsymbol{\xi}\|^2 - 2(\mathbf{e}^T\mathbf{P}\boldsymbol{\xi})^2}{2\mathbf{e}^T\mathbf{P}\boldsymbol{\xi}}\right).\end{aligned}$$

B Analysis for the case rank(DB) $= p < n - m$

Note that in this case $\mathbf{A} = \mathbf{U}_1^T = \emptyset$, so $\gamma = \frac{1}{2}$ and $\mathbf{P} = \mathbf{I}$. The following Lemma is easy to prove.

Lemma 9 *Let $\widehat{q} : \mathbf{R}^p \mapsto \mathbf{R}$ denote the function defined in (17). Then the domain $\mathcal{D}_{\widehat{q}} = \{\boldsymbol{\xi} : \widehat{q}(\boldsymbol{\xi}) > -\infty\}$ is given by*

$$\mathcal{D}_{\widehat{q}} = \left\{\boldsymbol{\xi} : \mathbf{e}^T\boldsymbol{\xi} \geq 0, 2(\mathbf{e}^T\boldsymbol{\xi})^2 - \|\boldsymbol{\xi}\|^2 \geq 0\right\}. \quad (54)$$

where $\mathbf{e} = (1, \mathbf{0}^T)^T$. For all $\boldsymbol{\xi} \in \mathcal{D}_{\widehat{q}}$, we have $\widehat{q}(\boldsymbol{\xi}) = 0$.

Then (15), (16), (17), (18), and Lemma 9 imply that the Lagrangian dual problem is given by

$$\begin{array}{rrcl} \max & (\mathbf{f} - \mathbf{E}^T\boldsymbol{\lambda})^T\mathbf{z}_0, & & \\ \text{subject to} & \mathcal{A}[\boldsymbol{\lambda}, \boldsymbol{\xi}, \mathbf{h}, \mathbf{p}] & = & \mathbf{0}, \\ & \boldsymbol{\lambda} & \geq & \mathbf{0}, \\ & \boldsymbol{\xi} & \in & \mathcal{K}, \end{array} \quad (55)$$

where $\mathcal{A}[\boldsymbol{\lambda}, \boldsymbol{\xi}, \mathbf{h}, \mathbf{p}]$ denotes the set of linear equalities in (22) and $\mathcal{K} = \{\mathbf{z} : \mathbf{e}^T\mathbf{z} \geq 0, 2(\mathbf{e}^T\mathbf{z})^2 - \|\mathbf{z}\|^2 \geq 0\}$.

As in the case discussed in the paper, first set $\boldsymbol{\xi} = \mathbf{0}$ and solve (55). Let $\boldsymbol{\mu}^{(1)}$ be its optimal solution. A direction $(\mathbf{d}_\xi, \mathbf{d}_\lambda)$ is an ascent direction at $(\mathbf{0}, \boldsymbol{\mu}^{(1)})$ if, and only if, $(\mathbf{d}_\xi, \mathbf{d}_\lambda)$ is a recession direction of the set

$$\begin{array}{rcl} -\mathbf{z}_0^T\mathbf{E}^T\mathbf{d}_\lambda & > & 0, \\ \mathcal{A}_{\mathbf{W}}[\mathbf{d}_\lambda, \mathbf{d}_\xi, \mathbf{0}, \mathbf{0}] & = & \mathbf{0}, \\ \mathbf{d}_\xi & \in & \mathcal{K}, \end{array} \quad (56)$$

with the matrix $\mathbf{W} = \mathbf{0}$.

Lemma 10 *A recession direction $(\mathbf{d}_\xi, \mathbf{d}_\lambda)$ of (56), if it exists, can be computed by solving two systems of linear equations.*

Proof: We will find a recession direction of (56) by solving the following problem.

$$\begin{array}{rrcl} \max & -\mathbf{z}_0^T\mathbf{E}^T\mathbf{d}_\lambda, & & \\ \text{subject to} & \mathcal{A}_{\mathbf{W}}[\mathbf{d}_\lambda, \mathbf{d}_\xi, \mathbf{0}, \mathbf{0}] & = & \mathbf{0}, \\ & \mathbf{d}_\xi & \in & \mathcal{K}. \end{array} \quad (57)$$

If the optimal value of this problem is positive, then (56) has a recession direction. The direction $(\mathbf{d}_\xi, \mathbf{d}_\lambda)$ can be computed by considering the following three cases:

(a) Suppose $\mathbf{d}_\xi = \mathbf{0}$. Then $(\mathbf{0}, \mathbf{d}_\lambda)$ is a recession direction for (56) if, and only if, $\mathbf{d}_\lambda$ solves

$$\begin{aligned} -\mathbf{z}_0^T \mathbf{E}^T \mathbf{d}_\lambda &= 1, \\ \mathcal{A}_{\mathbf{W}}[\mathbf{d}_\lambda, \mathbf{0}, \mathbf{0}, \mathbf{0}] &= \mathbf{0}. \end{aligned} \tag{58}$$

(b) Next, suppose (58) is infeasible; however, there exists a positive recession direction for (57). Set $\mathbf{e}^T \mathbf{d}_\xi = 1$ in (57) to obtain

$$\begin{array}{rrcl} \max & -\mathbf{z}_0^T \mathbf{E}^T \mathbf{d}_\lambda, & & \\ \text{subject to} & \mathcal{A}_{\mathbf{W}}[\mathbf{d}_\lambda, \mathbf{d}_\xi, \mathbf{0}, \mathbf{0}] & = & \mathbf{0}, \\ & \mathbf{e}^T \mathbf{d}_\xi & = & 1, \\ & \|\mathbf{d}_\xi\|^2 & \leq & 2. \end{array} \tag{59}$$

Since (58) is assumed to be infeasible, (59) is bounded. Setting $\mathbf{d}_\xi = -\mathbf{L}\mathbf{d}_\lambda$, we get

$$\begin{array}{rrcr} \max & -\big(\mathbf{E}\mathbf{z}_0\big)^T \mathbf{d}_\lambda, & & \\ \text{subject to} & \mathcal{A}_{\mathbf{W}}[\mathbf{d}_\lambda, -\mathbf{L}\mathbf{d}_\lambda, \mathbf{0}, \mathbf{0}] & = & \mathbf{0}, \\ & \mathbf{e}^T \mathbf{L}\mathbf{d}_\lambda & = & -1, \\ & \|\mathbf{L}\mathbf{d}_\lambda\|^2 & \leq & 2. \end{array}$$

Since the objective function of this problem is linear, the optimal $\mathbf{d}_\lambda^*$ satisfies $\|\mathbf{L}\mathbf{d}_\lambda^*\|^2 = 2$ and the Lagrangian function $\mathcal{L}$ is given by

$$\mathcal{L} = -(\mathbf{E}\mathbf{z}_0)^T \mathbf{d}_\lambda - \boldsymbol{\tau}^T \mathbf{M}\mathbf{d}_\lambda - \boldsymbol{\rho}^T \mathbf{W}\mathbf{d}_\lambda - \eta(\mathbf{e}^T \mathbf{L}\mathbf{d}_\lambda + 1) - \beta(\|\mathbf{L}\mathbf{d}_\lambda\|^2 - 2)$$

where $\beta \geq 0$ and the first-order optimality conditions are given by

$$\begin{array}{rcccrr} 2\beta \mathbf{L}^T \mathbf{L}\mathbf{d}_\lambda & +\mathbf{M}^T \boldsymbol{\tau} & +\mathbf{W}^T \boldsymbol{\rho} & +\mathbf{L}^T \mathbf{e}\eta & = & -\mathbf{E}\mathbf{z}_0, \\ \mathbf{M}\mathbf{d}_\lambda & & & & = & \mathbf{0}, \\ \mathbf{W}\mathbf{d}_\lambda & & & & = & \mathbf{0}, \\ \mathbf{e}^T \mathbf{L}\mathbf{d}_\lambda & & & & = & -1, \end{array} \tag{60}$$

and $\beta(\|\mathbf{L}\mathbf{d}_\lambda\|^2 - 2) = 0$. If $\beta = 0$, then (60) can be solved easily. Suppose $\beta > 0$. Then by setting $\bar{\boldsymbol{\rho}} = \frac{1}{\beta}\boldsymbol{\rho}$, $\bar{\boldsymbol{\tau}} = \frac{1}{\beta}\boldsymbol{\tau}$, and $\bar{\eta} = \frac{1}{\beta}\eta$, we see that (60) is equivalent to

$$\underbrace{\begin{bmatrix} 2\mathbf{L}^T \mathbf{L} & \mathbf{M}^T & \mathbf{W}^T & \mathbf{L}^T \mathbf{e} \\ \mathbf{M} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{W} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{e}^T \mathbf{L} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}}_{\triangleq \mathbf{K}} \begin{bmatrix} \mathbf{d}_\lambda \\ \bar{\boldsymbol{\tau}} \\ \bar{\boldsymbol{\rho}} \\ \bar{\eta} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ -1 \end{bmatrix} - \frac{1}{\beta} \begin{bmatrix} \mathbf{E}\mathbf{z}_0 \\ \mathbf{0} \\ \mathbf{0} \\ 0 \end{bmatrix}.$$

Suppose $\mathbf{K}$ is non-singular. Let $\mathbf{w} = (\bar{\boldsymbol{\tau}}^T, \bar{\boldsymbol{\rho}}^T, \bar{\eta})^T$, $\mathbf{b}_1 = (\mathbf{0}^T, \mathbf{0}^T, -1)^T$, and $\mathbf{b}_2 = \mathbf{E}\mathbf{z}_0$. Partition $\mathbf{K}^{-1}$ into submatrices

$$\mathbf{K}^{-1} = \begin{bmatrix} \mathbf{K}_{11}^{-1} & \mathbf{K}_{12}^{-1} \\ \mathbf{K}_{12}^{-T} & \mathbf{K}_{22}^{-1} \end{bmatrix}$$

such that

$$\begin{bmatrix} \mathbf{d}_\lambda \\ \mathbf{w} \end{bmatrix} = \mathbf{K}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{b}_1 \end{bmatrix} - \beta \mathbf{K}^{-1} \begin{bmatrix} \mathbf{b}_2 \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{K}_{12}^{-1} \mathbf{b}_1 - \beta \mathbf{K}_{11}^{-1} \mathbf{b}_2 \\ \mathbf{K}_{22}^{-1} \mathbf{b}_1 - \beta \mathbf{K}_{12}^{-T} \mathbf{b}_2 \end{bmatrix}$$

This partition implies that $\mathbf{K}_{12}^{-T}\mathbf{L}^T\mathbf{L}\mathbf{K}_{11}^{-1}$. Therefore, β is the unique positive root of

$$\begin{aligned} 2 &= \|\mathbf{L}\mathbf{d}_\lambda\|^2, \\ &= \|\mathbf{L}\mathbf{K}_{12}^{-1}\mathbf{b}_1\|^2 - 2\frac{1}{\beta}(\mathbf{L}\mathbf{K}_{12}^{-1}\mathbf{b}_1)^T\mathbf{L}\mathbf{K}_{11}^{-1}\mathbf{b}_2 + \frac{1}{\beta^2}\|\mathbf{L}\mathbf{K}_{11}^{-1}\mathbf{b}_2\|^2, \\ &= \|\mathbf{L}\mathbf{K}_{12}^{-1}\mathbf{b}_1\|^2 - \frac{1}{\beta^2}\|\mathbf{L}\mathbf{K}_{11}^{-1}\mathbf{b}_2\|^2. \end{aligned}$$

Consequently,

$$\beta = \sqrt{\frac{\|\mathbf{L}\mathbf{K}_{11}^{-1}\mathbf{b}_2\|^2}{2 - \|\mathbf{L}\mathbf{K}_{12}^{-1}\mathbf{b}_1\|^2}}.$$

Thus, (60) has a solution if, and only if, $2 - \|\mathbf{L}\mathbf{K}_{12}^{-1}\mathbf{b}_1\|^2 > 0$.

The case where $\mathbf{K}$ is singular can be handled by taking the SVD of $\mathbf{K}$ and working in the appropriate range spaces.

(c) In case one is not able to produce a solution in either (a) or (b), it follows that the optimal solution of (29) is 0, and $(\mathbf{d}_\xi, \mathbf{d}_\lambda) = (\mathbf{0}, \mathbf{0})$ achieves this value.

■

In a typical step of the ActiveSet when $\mathbf{rank}(\mathbf{DB}) = p$, we have to solve the following problem.

$$\begin{array}{rrcl} \max & -\mathbf{z}_0^T\mathbf{E}^T\boldsymbol{\lambda}, & & \\ \text{subject to} & \mathcal{A}_{\mathbf{W}}[\boldsymbol{\lambda}, \boldsymbol{\xi}, \mathbf{h}, \mathbf{p}] & = & \mathbf{0}, \\ & \boldsymbol{\xi} & \in & \mathcal{K}, \end{array} \tag{61}$$

where $\mathbf{W}$ denotes the current inactive set, i.e. $\mathbf{W} = \sum_{i:\lambda_i=0} \mathbf{e}_i\mathbf{e}_i^T$.

Lemma 11 *Suppose there exists a feasible $(\bar{\boldsymbol{\xi}}, \bar{\boldsymbol{\lambda}})$ for (61) such that $\bar{\boldsymbol{\xi}} \in \mathbf{int}(\mathcal{K})$. Then (61) can be solved in closed form by solving at most three systems of linear equations.*

This result can be established using a combination of the techniques used to establish Lemma 10 and Lemma 5.

C Decreasing the Dimension of K

Consider the following system of linear equalities

$$\begin{array}{rcrcl} \tilde{\mathbf{K}}\tilde{\mathbf{d}} & + & \tilde{\mathbf{W}}^T\bar{\boldsymbol{\rho}} & = & \mathbf{b}, \\ \tilde{\mathbf{W}}\tilde{\mathbf{d}} & & & = & \mathbf{0}, \end{array} \tag{62}$$

where $\tilde{\mathbf{K}}$ and $\tilde{\mathbf{d}}$ are given in (41) and $\tilde{\mathbf{W}} = [\mathbf{W}, \mathbf{0}, \mathbf{0}]$. Let SVD of $\tilde{\mathbf{K}}$ be given by

$$\tilde{\mathbf{K}} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T = \mathbf{U}\begin{bmatrix}\boldsymbol{\Sigma}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{V}_0^T \\ \mathbf{V}_1^T\end{bmatrix},$$

where $\mathbf{\Sigma}_0 \in \mathbf{R}^{r_K \times r_K}$ is a diagonal matrix and $r_K = \mathbf{rank}(\tilde{\mathbf{K}})$. Decompose $\tilde{\mathbf{d}} = \mathbf{V}_0\boldsymbol{\mu} + \mathbf{V}_1\boldsymbol{\zeta}$. Then (62) is equivalent to

$$\begin{array}{rcrcl} \mathbf{U}\begin{bmatrix}\mathbf{\Sigma}_0\boldsymbol{\mu}\\ \mathbf{0}\end{bmatrix} & + & \tilde{\mathbf{W}}^T\bar{\boldsymbol{\rho}} & = & \mathbf{b}, \\ \tilde{\mathbf{W}}(\mathbf{V}_0\boldsymbol{\mu} + \mathbf{V}_1\boldsymbol{\zeta}) & & & = & \mathbf{0}, \end{array}$$

which is equivalent to

$$\begin{array}{rcrcl} \begin{bmatrix}\mathbf{\Sigma}_0\boldsymbol{\mu}\\ \mathbf{0}\end{bmatrix} & + & \mathbf{U}^T\tilde{\mathbf{W}}^T\bar{\boldsymbol{\rho}} & = & \mathbf{U}^T\mathbf{b}, \\ \tilde{\mathbf{W}}(\mathbf{V}_0\boldsymbol{\mu} + \mathbf{V}_1\boldsymbol{\zeta}) & & & = & \mathbf{0}. \end{array} \tag{63}$$

Let

$$\mathbf{U}^T = \begin{bmatrix}\mathbf{U}_0^T\\ \mathbf{U}_1^T\end{bmatrix}.$$

Then (63) is equivalent to

$$\begin{array}{rcrcrcl} \mathbf{\Sigma}_0\boldsymbol{\mu} & & & + & \mathbf{U}_0^T\tilde{\mathbf{W}}^T\bar{\boldsymbol{\rho}} & = & \mathbf{U}_0^T\mathbf{b}, \\ & & & & \mathbf{U}_1^T\tilde{\mathbf{W}}^T\bar{\boldsymbol{\rho}} & = & \mathbf{U}_1^T\mathbf{b}, \\ \tilde{\mathbf{W}}\mathbf{V}_0\boldsymbol{\mu} & + & \tilde{\mathbf{W}}\mathbf{V}_1\boldsymbol{\zeta} & & & = & \mathbf{0}. \end{array}$$

Setting $\boldsymbol{\mu} = \mathbf{\Sigma}_0^{-1}(\mathbf{U}_0^T\mathbf{b} - \mathbf{U}_0^T\tilde{\mathbf{W}}^T\bar{\boldsymbol{\rho}})$, we obtain the following system which has a smaller number of variables.

$$\begin{array}{rcrcl} & & \mathbf{U}_1^T\tilde{\mathbf{W}}^T\bar{\boldsymbol{\rho}} & = & \mathbf{U}_1^T\mathbf{b}, \\ \tilde{\mathbf{W}}\mathbf{V}_1\boldsymbol{\zeta} & - & \tilde{\mathbf{W}}\mathbf{V}_0\mathbf{\Sigma}_0^{-1}\mathbf{U}_0^T\tilde{\mathbf{W}}^T\bar{\boldsymbol{\rho}} & = & -\tilde{\mathbf{W}}\mathbf{V}_0\mathbf{\Sigma}_0^{-1}\mathbf{U}_0^T\mathbf{b}. \end{array}$$