McMaster University

Advanced Optimization Laboratory

Title:

An Algorithm for Perturbed Second-order Cone Programs

Authors:

Yu Xia

AdvOl-Report No. 2004/17

October 2004, Hamilton, Ontario, Canada

An Algorithm for Perturbed Second-order Cone Programs

Yu Xia*

October 9, 2004

Abstract

The second-order cone programming problem is reformulated into several new systems of nonlinear equations. Assume the perturbation of the data is in a certain neighborhood of zero. Then starting from a solution to the old problem, the semismooth Newton's iterates converge Q-quadratically to a solution of the perturbed problem. The algorithm is globalized. Numerical examples show that the algorithm is good for "warm starting" – for some instances, the solution of a perturbed problem is hit in two iterations.

Key words. Second-order cone, complementarity, semismooth, warm start, Newton's method

1 Introduction and Notations

1.1 Motivation

The aim of this paper is to develop a "warm starting" algorithms for the second-order cone programing (SOCP) problem. A large number of real-world applications are solved by SOCP; see [2, 25, 26, 27, 28, 31, 36, 44], etc. Also, SOCP are used as subroutines for some more general optimization problems; see [18], for instance. For many applications, "warm starting" algorithms may reduce computations (see [17, 21], etc.). For example, in [50], the Euclidean facilities location problem (abbreviated as EFL) and the Steiner minimal tree (abbreviated as SMT) problem under known topology are cast as SOCP problems. When an EFL or an SMT problem needs to be solved under environmental changes, or when a series of similar EFL or SMT problems constitute a bigger project, it is often the case that there is a small change of the cost function in the EFL problem, or a little move of the regular points in the SMT problem. Then, it is reasonable to expect that a new optimum is close to the old solution and to use the old solution to "warm start" the new problems. Although SOCP can be approximated by interior point methods (IPM) (see [1, 3, 6, 34, 35, 36, 38, 45, 48] and references therein), they are reputed to be unsuitable for re-optimization. The reason why an optimal solution of an SOCP problem usually can not "warm start" another slightly perturbed SOCP problem through IPMs is that the starting point of any IPM must be in the interior of the cone, while optima generally are achieved at the boundary.

Several warm starting IPMs have been proposed for a simpler problem – the linear programming (LP), but they all have some restrictive assumptions. The approaches in [21, 33, 51] are not doable without information about the intermediate solutions of the previous problem. To use the shifted barrier function studied in [17], one needs to know a bound on the set of all dual feasible slack vectors of the perturbed problem, calculation of which needs extra work ([46]). One can also perturb the optimal solution of the old problem to get an initial interior point for the new problem, as is

*Computing and Software, McMaster University, 1280 Main street West, Hamilton, Ontario L8S 4K1, Canada (yuxia@cas.mcmaster.ca). This research was supported in part through the U.S. National Science Foundation grant CCR-9901991.

advocated in [29], but the amount of perturbation is heuristic. If the perturbation is too big, the initial point may be far from optima; if the perturbation is too small, the iterations may be stuck at an infeasible vertex. Typically, a warm starting interior point method addresses either the change of the size of the problem, or the perturbation of the parameters, but not both. It is known that the simplex method is good to "warm start" the LP, but no analogous simplex method for the SOCP is known yet. In conclusion, to our knowledge, there is no efficient re-optimization method for SOCP at this moment.

Another drawback of interior point methods is that when strict complementarity conditions are not satisfied at optima, the linear system becomes increasingly ill-conditioned as the iterates approaching the solution, which causes interior point methods numerically unstable (see [3]).

Since Newton's method for systems of nonlinear equations ensures variables converge Q-quadratically locally, we reformulate the SOCP model into systems of nonlinear equations and solve them by semismooth Newton's method of [32, 41]. As this method doesn't require each iterate in the interior of the cone, the old solution is a good starting point for a new problem slightly different in both the size and the parameters. The semismooth Newton's method also doesn't require strict complementarity.

Reformulating the KKT system for general nonlinear programming (NLP) into equations has being studied before, see for instance [15, 37, 41] and references therein, but assumptions of their local analysis typically include second-order sufficiency conditions and linear independence of the gradients of constratints, which are not applicable to the SOCP. Furthermore, none of them are considered in the aspect of re-optimization. Our perturbation analysis can be extended to certain NLPs. Besides, our global algorithm is different from that of other nonsmooth methods. We use perturbation and reformulation, which can be generalized to some NLPs, too. We notice that [13, 19] also deal with second-order cone complementarity problems; however, our approach is different from theirs.

This paper is divided into seven parts. In §2, we convert the second-order cone program into a system of equalities and inequalities. In §3, the complementarity conditions are transformed into equations; hence the system can be solved via semismooth Newton's method which is globally convergent, locally Q-quadratic convergent. In §4, we show that after small perturbations of data, including addition or removal of constraints or variables, the algorithm finds the new solution Q-quadratically from the old solution. Globalization of the algorithm is discussed in §5. Numerical examples are presented in §6. Properties of this method are summarized in §7.

1.2 Notations

The following notations will be used throughout this paper.

Superscripts refer to iteration numbers; subscripts signify block numbers. We use bold lower case letters for column vectors, lower case letters for entries of a vector, and capital letters for matrices. Primal and dual variables are indexed from 0; the jth entry of vector $\mathbf{x}_i$ is denoted as $(\mathbf{x}_i)_j$. Semicolons ";" concatenate column vectors. In this way, $(\mathbf{x};\mathbf{y};\mathbf{z}) = (\mathbf{x}^T, \mathbf{y}^T, \mathbf{z}^T)^T$.

Let $\mathcal{Q}_{n+1}$ indicate a second-order cone (Loréntz cone, ice-cream cone, quadratic cone) in $\mathbb{R}^{n+1}$:

$$\mathcal{Q}_{n+1} \stackrel{\text{def}}{=} \left\{ \mathbf{x} \in \mathbb{R}^{n+1} \colon x_0 \geq \sqrt{\sum_{i=1}^{n} x_i^2} \right\}.$$

It is self dual (see [12, 36]). We write $\mathbf{x} \geq_{\mathcal{Q}_{n+1}} \mathbf{0}$ interchangeably with $\mathbf{x} \in \mathcal{Q}_{n+1}$, since it is a partial order. As convention ([1]), we partition $\mathcal{Q}_{n+1}$ into three disjointed sets: $\{\mathbf{0}\}$, int $\mathcal{Q}$, and bd $\mathcal{Q}$, where bd $\mathcal{Q}$ is the boundary of $\mathcal{Q}$ excluding $\mathbf{0}$.

A second-order cone programming problem is represented as the following:

$$\begin{array}{ll} \min & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \geq_{\mathcal{Q}} \mathbf{0}. \end{array} \tag{1}$$

In the above model, $\mathbf{c} \stackrel{\text{def}}{=} (\mathbf{c}_1; \ldots; \mathbf{c}_n)$, $A \stackrel{\text{def}}{=} [A_1 \ldots A_n]$, $\mathcal{Q} \stackrel{\text{def}}{=} \mathcal{Q}_{N_1} \times \cdots \times \mathcal{Q}_{N_n}$, and $\mathbf{x} \stackrel{\text{def}}{=} (\mathbf{x}_1; \ldots; \mathbf{x}_n)$, where for $i = 1, \ldots, n$, $N_i \in \mathbb{N}$ is the dimension of variable $\mathbf{x}_i$, $A_i \in \mathbb{R}^{m \times N_i}$, $\mathbf{c}_i \in \mathbb{R}^{N_i}$, and $\mathbf{b} \in \mathbb{R}^m$ are data.

We use $\bar{\mathbf{x}}$ to represent the subvector of $\mathbf{x}$ excluding x_0, $\bar{\bar{\mathbf{x}}}$ to represent the subvector of $\mathbf{x}$ excluding x_0 and x_1, i.e., $\mathbf{x} = (x_0, \bar{\mathbf{x}}^T)^T = (x_0, x_1, \bar{\bar{\mathbf{x}}}^T)^T$.
Given a matrix $A = [\mathbf{a}_0, \mathbf{a}_1, \ldots, \mathbf{a}_n]$, $\bar{A}$ stands for the submatrix consisting of the last n columns, i.e., $\bar{A} \stackrel{\text{def}}{=} [\mathbf{a}_1, \ldots, \mathbf{a}_n]$. The symbol $\bar{\bar{A}}$ is used to represent the submatrix of A excluding the first two columns: $\bar{A} = [\mathbf{a}_1, \bar{\bar{A}}]$. We use subscribe $\bar{1}$ to designate the index set without index 1. Under this notation, $A_{\bar{1}} \stackrel{\text{def}}{=} [\mathbf{a}_0, \mathbf{a}_2, \ldots, \mathbf{a}_n]$; $\mathbf{z}_{\bar{1}} \stackrel{\text{def}}{=} [z_0; z_2; \ldots; z_n]$.

The symbol $\delta(\mathbf{x}|X)$ stands for the indicator function of a set X at $\mathbf{x}$.

We use $D[f(\mathbf{x}); \Delta\mathbf{x}]$ to symbolize the directional derivative of f at $\mathbf{x}$ with respect to $\Delta\mathbf{x}$.

We let R_{n+1} correspond to the $(n+1)$ by $(n+1)$ diagonal matrix:

$$R_{n+1} \stackrel{\text{def}}{=} \begin{pmatrix} 1 & & & \\ & -1 & & \\ & & \ddots & \\ & & & -1 \end{pmatrix}.$$

We write $\mathbf{e}_{n(i)}$ to denote the vector in $\mathbb{R}^n$ whose ith entry is 1 and the remaining entries are 0.

For a scalar a, $[a]_+ \stackrel{\text{def}}{=} \max(a, 0)$, $[a]_- \stackrel{\text{def}}{=} \min(a, 0)$.

When the dimension is clear from the context, we will omit subscripts for R_{n+1}, $e_{n(i)}$, etc.

2 The System

In this section we consider a general convex program which includes (1):

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & \mathbf{x} \in H \\ & \mathbf{x} \geq_{\mathcal{Q}} \mathbf{0}. \end{array} \tag{2}$$

Here, $f : \mathbb{R}^{N_1+\cdots+N_n} \mapsto \mathbb{R}$ is a proper convex function, H is a convex set.

Indicating the normal cone to H at $\mathbf{x}$ as $\mathrm{N}_H(\mathbf{x})$, the subdifferential of f at $\mathbf{x}$ as $\partial f(\mathbf{x})$, we transform (2) into the following system:

$$\mathbf{0} \in \partial f(\lambda_1\mathbf{z}_1; \ldots; \lambda_n\mathbf{z}_n) - (\omega_1 R\mathbf{z}_1; \ldots; \omega_n R\mathbf{z}_n) + \mathrm{N}_H(\lambda_1\mathbf{z}_1; \ldots; \lambda_n\mathbf{z}_n) \tag{3a}$$

$$(\lambda_1\mathbf{z}_1; \ldots; \lambda_n\mathbf{z}_n) \in H \tag{3b}$$

$$(\mathbf{z}_i)_0 = 1 \ (i = 1, \ldots, n) \tag{3c}$$

$$\lambda_i\omega_i \left((\mathbf{z}_i)_0^2 - \bar{\mathbf{z}}_i^T\bar{\mathbf{z}}_i\right) = 0 \ (i = 1, \ldots, n) \tag{3d}$$

$$(\mathbf{z}_i)_0^2 - \bar{\mathbf{z}}_i^T\bar{\mathbf{z}}_i \geq 0 \ (i = 1, \ldots, n) \tag{3e}$$

$$\lambda_i \geq 0 \ (i = 1, \ldots, n) \tag{3f}$$

$$\omega_i \geq 0 \ (i = 1, \ldots, n). \tag{3g}$$

An advantage of the above system to others is that it includes the case when $\mathbf{x} = \mathbf{0}$. For briefness, in what follows, we designate $\boldsymbol{\lambda} \stackrel{\text{def}}{=} (\lambda_1, \ldots, \lambda_n)^T$, $\boldsymbol{\omega} \stackrel{\text{def}}{=} (\omega_1, \ldots, \omega_n)^T$. The next Lemma shows the equivalence of (3) and (2).

Lemma 1 *Assume f is a proper convex function, and the objective value of (2) is below bounded on the feasible set. Then:*
(i) If $(\boldsymbol{\omega}^*; \boldsymbol{\lambda}^*; \mathbf{z}^*)$ solves (3), $\mathbf{x}^* = (\lambda_1^*\mathbf{z}_1^*; \ldots; \lambda_n^*\mathbf{z}_n^*)$ solves (2).

(ii) Furthermore, assume $H = \cap_{i=1}^{m} H_i$ *where* H_i *is a polyhedral convex set for* $i = 1, \ldots r$, *and* $\cap_{i=1}^{r} H_i \cap_{i=r+1}^{m} \operatorname{ri} H_i \cap \operatorname{int} \mathcal{Q} \neq \emptyset$. *Then for any* $\mathbf{x}^*$ *satisfying* (2), *there exists* $(\boldsymbol{\omega}^*; \boldsymbol{\lambda}^*; \mathbf{z}^*)$ *solving* (3), *and* $\mathbf{x}^* = (\lambda_1^* \mathbf{z}_1^*; \ldots; \lambda_n^* \mathbf{z}_n^*)$.

To prove the lemma, we first describe the normal cone to $\mathcal{Q}$.

Proposition 1 *For any* $\mathbf{x} \in \operatorname{bd} \mathcal{Q}_{n+1}$, $\mathbf{x} \neq \mathbf{0}$,

$$\mathrm{N}_{\mathcal{Q}_{n+1}}(\mathbf{x}) = \begin{cases} \{\mathbf{0}\} & \mathbf{x} \in \operatorname{int} Q_{n+1}, \\ \{\lambda(-x_0, x_1, \ldots, x_n)^T \colon \lambda \geq 0\} & \mathbf{x} \in \operatorname{bd} \mathcal{Q}_{n+1}, \\ -\mathcal{Q}_{n+1} & \mathbf{x} = \mathbf{0}. \end{cases}$$

Proof: We will omit the subscript "$n+1$", just write "$\mathcal{Q}$" in the proof. By definition,

$$\mathrm{N}_{\mathcal{Q}}(\mathbf{x}) = \{\mathbf{z} \colon \mathbf{z}^T(\mathbf{y} - \mathbf{x}) \leq 0, \forall\, \mathbf{y} \geq_{\mathcal{Q}} \mathbf{0}\}. \tag{4}$$

For $\mathbf{x} = \mathbf{0}$ and $\mathbf{x} \in \operatorname{int} \mathcal{Q}$, the representations are easy to verify. Next, we will give that for $\mathbf{x} \in \operatorname{bd} \mathcal{Q}$ through similar techniques in [1, 12]. First, we will prove $\mathrm{N}_{\mathcal{Q}}(\mathbf{x}) \subseteq \{\lambda(-x_0, x_1, \ldots, x_n)^T \colon \lambda \geq 0\}$.

By convexity of $\mathcal{Q}$, $\forall\, \mathbf{v} \in \mathcal{Q}$, we have $\mathbf{v} + \mathbf{x} \in \mathcal{Q}$. Setting $\mathbf{y} = \mathbf{v} + \mathbf{x}$ in (4), one sees the normal vector $\mathbf{z}$ satisfies:

$$\forall\, \mathbf{v} \geq_{\mathcal{Q}} \mathbf{0},\ \langle \mathbf{z}, \mathbf{v} \rangle \leq \mathbf{0}. \tag{5}$$

Let $\mathbf{v} = (\|\bar{\mathbf{z}}\|_2 \,; \bar{\mathbf{z}})$ in (5); then one can get

$$z_0 \leq 0, \tag{6}$$

$$z_0^2 \geq \sum_{i=1}^{n} z_i^2. \tag{7}$$

Letting $\mathbf{y}$ in (4) be $\mathbf{0}$ and then $2\mathbf{x}$, one can see

$$z_0 x_0 = -\sum_{i=1}^{n} z_i x_i. \tag{8}$$

Since $\mathbf{x} \geq_{\mathcal{Q}} \mathbf{0}$, we have

$$x_0^2 \geq \sum_{i=1}^{n} x_i^2. \tag{9}$$

For an arbitrary scalar α, add (7), 2α multiplying (8) and α^2 multiplying (9) together:

$$(z_0 + \alpha x_0)^2 \geq \sum_{i=1}^{n} (z_i - \alpha x_i)^2. \tag{10}$$

Let $\alpha = -\frac{z_0}{x_0}$ in (10). Notice $\alpha \geq 0$ by (6) and $\mathbf{x} \geq_{\mathcal{Q}} \mathbf{0}$. So (10) is valid iff $z_0 = \frac{z_0}{x_0} x_0$, $z_i = -\frac{z_0}{x_0} x_i$ $(i = 1, \ldots, n)$. This shows

$$\mathrm{N}_{\mathcal{Q}}(\mathbf{x}) \subseteq \{\lambda(-x_0, x_1, \ldots, x_n)^T \colon \lambda \geq 0\}.$$

The other direction

$$\mathrm{N}_{\mathcal{Q}}(\mathbf{x}) \supseteq \{\lambda(-x_0, x_1, \ldots, x_n)^T \colon \lambda \geq 0\}.$$

is obvious since $\mathbf{x} \in \operatorname{bd} \mathcal{Q}$ and $\mathcal{Q}$ is self-dual. ■

Remark 1 *By [43],* $\mathrm{N}_{\mathcal{Q}}(\mathbf{x}) = \partial\delta(\mathbf{x}|\mathcal{Q})$. *Further more, every* $\mathbf{y} \in \partial\delta(\mathbf{x}|\mathcal{Q})$ *is the normal to a nonvertical supporting plane to the graph of* $\delta(\cdot|\mathcal{Q})$ *at* $[\mathbf{x}, \delta(\mathbf{x}|\mathcal{Q})]$. *For* $\mathbf{x} \in \operatorname{bd}\mathcal{Q}$, *the derivative of* $\sum_{i=1}^{n}\mathbf{x}_n^2 - \mathbf{x}_0^2$ *are the normals to the supporting plane to the graph of* $\delta(\mathbf{x}|\mathcal{Q})$.

Now we proceed to prove Lemma 1.

Proof: By [43], a necessary and sufficient condition for $\mathbf{x}^*$ to belong to the minimum set of f is that

$$\mathbf{0} \in \partial\left[f(\mathbf{x}^*) + \delta\left(\mathbf{x}^*|H \cap \mathcal{Q}\right)\right]. \tag{11}$$

Note that $\partial\delta(\mathbf{x}^*|H \cap \mathcal{Q}) = \mathrm{N}_{\mathbf{x}^*}\{H \cap \mathcal{Q}\}$. According to [43, Theorem 23.8] and its corollary,

$$\mathrm{N}_{H\cap\mathcal{Q}} \supseteq \mathrm{N}_{\mathcal{Q}} + \mathrm{N}_H;$$

if in addition, $\cap_{i=1}^{r} H_i \cap_{i=r+1}^{m} \operatorname{ri} H_i \cap \operatorname{int}\mathcal{Q} \neq \emptyset$,

$$\mathrm{N}_{H\cap\mathcal{Q}} = \mathrm{N}_{\mathcal{Q}} + \mathrm{N}_H\,.$$

Proof of **(i)**: Assume $(\boldsymbol{\omega}^*, \boldsymbol{\lambda}^*, \mathbf{z}^*)$ is a solution to (3). Then with regard to (3b), (3f) and (3e), $\mathbf{x}^* = (\lambda_1^*\mathbf{z}_1^*, \ldots, \lambda_n^*\mathbf{z}_n^*)$ is feasible for (2). Moreover, by (3g) and Proposition 1, $-\omega_i^* R\mathbf{z}_i^* \in \mathrm{N}_{\mathcal{Q}_i}(\mathbf{x}_i^*)$. Hence $\mathbf{x}^*$ is optimal by (11).

Under the hypotheses of **(ii)**, assume $\mathbf{x}^*$ is a solution to (2). We let $\lambda_i^* = (x_i^*)_0$. By (11), $\exists\,\mathbf{v}^* \in \partial f(\mathbf{x}^*)$, $\mathbf{u}^* \in \mathrm{N}_{\mathcal{Q}}(\mathbf{x}^*)$, $\mathbf{w}^* \in \mathrm{N}_H(\mathbf{x}^*)$, such that $\mathbf{v}^* + \mathbf{u}^* + \mathbf{w}^* = \mathbf{0}$. When $\mathbf{x}_i^* \in \operatorname{int}\mathcal{Q}_{N_i}$, we set $\omega_i^* = 0$ and $\mathbf{z}_i^* = \frac{1}{\lambda_i^*}\mathbf{x}_i^*$. When $\mathbf{x}_i^* \in \operatorname{bd}\mathcal{Q}_{N_i}$, let $\mathbf{z}_i^* = \frac{1}{\lambda_i^*}\mathbf{x}_i^*$. By lemma 1, $\exists\,\alpha \geq 0$, such that $\mathbf{u}_i^* = -\alpha R\mathbf{x}_i^*$. So we set $\omega_i^* = \alpha^* \cdot \lambda_i^*$. When $\mathbf{x}_i^* = \mathbf{0}$. By Proposition 1, $-\mathbf{u}_i^* \in \mathcal{Q}_{N_i}$. We set $\omega_i^* = (u_i^*)_0$. If $\omega_i^* \neq 0$, we set $\mathbf{z}_i^* = -\frac{1}{\omega_i^*}R\mathbf{u}_i^*$; otherwise, we set $\mathbf{z}_i^* = \mathbf{e}_0$. ∎

Remark 2 *In a previous version, we assume* f *is differentiable and* H *is affine. Let* $\sqrt{\bar{\mathbf{x}}_i^T\bar{\mathbf{x}}_i} - (\mathbf{x}_i)_0 \leq 0\,(i = 1, \ldots, n)$ *represent* $\mathbf{x}_i \in \mathcal{Q}_{N_i}$. *Professor Paul Tseng pointed out that this constraint is convex and applying [43, Theorem 28.2] one can conclude the existence of Lagrangian multipliers under the same assumptions as that in Lemma 1; and using subgradients, the KKT system is equivalent to* (3), *which is valid even if* f *is not differentiable.*

3 Reformulations

we will transform (3) into some systems of equations, which will be solved by semismooth Newton's method. Since linear objective and polyhedral convex domain are good enough for many applications, we assume $f(\mathbf{x}) = \mathbf{c}^T\mathbf{x}$ and $H = \{\mathbf{x}\colon A\mathbf{x} = \mathbf{b}\}$ in the following context, although all the analyses can be extended to nonlinear objectives and general convex constraints. Note that under these hypotheses, $\mathrm{N}(H) = \{A^T\mathbf{y}\}$ when A has full row rank.

3.1 Preliminary

We will use some results in [40, 41, 42] about semismooth Newton-type optimization, because its assumptions are weaker than those of some other papers (see [40, 42]). For completeness, we cite them here.

Let $G : \mathbb{R}^n \mapsto \mathbb{R}^m$ denote a Lipschitz continuous mapping, and let $\mathbf{x}$ be a given point in $\mathbb{R}^n$. Then by Rademacher's theorem, G is differentiable almost everywhere in the sense of Lebesgue measure in any neighborhood of $\mathbf{x}$ in which G is Lipschitz. Let D_G be the set at which G is differentiable. Define

$$\partial_B G(\mathbf{x}) \stackrel{\text{def}}{=} \left\{ \lim_{\substack{\mathbf{x}_i \to \mathbf{x} \\ \mathbf{x}_i \in D_G}} \nabla G(\mathbf{x}_i) \right\},$$

Then $\partial G(\mathbf{x}) = \operatorname{conv}\partial_B G(\mathbf{x})$, where $\partial G(\mathbf{x})$ is the generalized Jacobian of G at $\mathbf{x}$ ([8]).

Definition 1 *([41, 32]) A function G is stated as semismooth at $\mathbf{x}$ if G is locally Lipschitz continuous at $\mathbf{x}$ and*

$$\lim_{\substack{\mathbf{h}'\to\mathbf{h}\\ t\downarrow 0}} \frac{G(\mathbf{x}+t\mathbf{h}')-G(\mathbf{x})}{t} = \lim_{\substack{V\in\partial G(\mathbf{x}+t\mathbf{h}')\\ \mathbf{h}'\to\mathbf{h},\, t\downarrow 0}} \{V\mathbf{h}'\}.$$

Semismooth functions include convex functions and smooth functions, and are directionally differentiable (see [41]).

Definition 2 *([40, 42]) Suppose G is semismooth at $\mathbf{x}$. Then G is said to be strongly semismooth at $\mathbf{x}$ if there exists a constant L and a neighborhood N of $\mathbf{x}$ such that $\forall\, \mathbf{x}+\mathbf{h}\in N$,*

$$\|G'(\mathbf{x}+\mathbf{h};\mathbf{h})-G'(\mathbf{x};\mathbf{h})\|_2 \le L\,\|h\|_2^2\,.$$

It is also argued in [42] that a vector valued function is strongly semismooth iff each component is strongly semismooth; a function with locally Lipschitzian derivative (LC^1 function) is strongly semismooth everywhere; the sum and the min of two LC^1 functions are strongly semismooth.

Definition 3 *([40]) G is said to be strongly BD-regular at $\mathbf{x}$ if $\forall\, V\in\partial_B G(\mathbf{x})$, V is nonsingular.*

Theorem 1 *([40, Theorem 3.1]) Assume $\mathbf{x}^*$ is a solution of G, G is strongly semismooth and strongly BD-regular at $\mathbf{x}^*$. Then the sequence*

$$\mathbf{x}^{k+1} = \mathbf{x}^k - {V^k}^{-1} G(\mathbf{x}^k), \quad \text{where } V^k\in\partial_B G(\mathbf{x}^k), \tag{12}$$

is well defined and converges Q-quadratically to $\mathbf{x}^*$ in a neighborhood of $\mathbf{x}^*$. Besides, if $G(\mathbf{x}^k)\neq\mathbf{0}$ for all k; then $\lim_{k\to\infty}\frac{\|G(\mathbf{x}^{k+1})\|_2}{\|G(\mathbf{x}^k)\|_2}=0$.

In this section, we focus on the non-singularity of each element of ∂G for the sake of perturbation analysis in §4, although Theorem 1 only requires strongly BD-regular at optimum.

For all the nonlinear equation reformulations, we assume the solution satisfies:

$$\begin{gathered}\lambda_i \text{ and } \omega_i \text{ not be zero at the same time } (i=1,\ldots,n);\\ \boldsymbol{\lambda}\neq\mathbf{0}.\end{gathered} \tag{13}$$

The justification of the above assumptions is that if there exists i such that $\lambda_i=\omega_i=0$, the *ith* block can be ignored; Suppose $\boldsymbol{\lambda}=\mathbf{0}$, then $\mathbf{b}$ must also be zero, and the Lagrangian multiplier is not unique. We partition the index set at optimum into five parts:

$$\begin{aligned}
L_1 &\stackrel{\text{def}}{=} \{i\colon \lambda_i>0,\ \omega_i>0,\ \mathbf{z}_i\in \operatorname{bd}\mathcal{Q}_{N_i}\},\\
L_2 &\stackrel{\text{def}}{=} \{i\colon \lambda_i>0,\ \omega_i=0,\ \mathbf{z}_i\in \operatorname{bd}\mathcal{Q}_{N_i}\},\\
L_3 &\stackrel{\text{def}}{=} \{i\colon \lambda_i>0,\ \omega_i=0,\ \mathbf{z}_i\in \operatorname{int}\mathcal{Q}_{N_i}\},\\
L_4 &\stackrel{\text{def}}{=} \{i\colon \lambda_i=0,\ \omega_i>0,\ \mathbf{z}_i\in \operatorname{bd}\mathcal{Q}_{N_i}\},\\
L_5 &\stackrel{\text{def}}{=} \{i\colon \lambda_i=0,\ \omega_i>0,\ \mathbf{z}_i\in \operatorname{int}\mathcal{Q}_{N_i}\}.
\end{aligned}$$

For any $\mathbf{z}\in\mathbb{R}^{n+1}$, $z_0=1$, we define $K_{\mathbf{z}}$ (see [5]) as: when $\|\bar{\mathbf{z}}\|_2\neq 0$ and $z_1\neq -\|\bar{\mathbf{z}}\|_2$,

$$K_{\mathbf{z}} \stackrel{\text{def}}{=} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \mathbf{0}^T\\ \frac{z_1}{2\|\bar{\mathbf{z}}\|_2} & -\frac{z_1}{2\|\bar{\mathbf{z}}\|_2} & -\frac{\bar{\mathbf{z}}^T}{\sqrt{2}\|\bar{\mathbf{z}}\|_2}\\ \frac{\bar{\mathbf{z}}}{2\|\bar{\mathbf{z}}\|_2} & -\frac{\bar{\mathbf{z}}}{2\|\bar{\mathbf{z}}\|_2} & \frac{1}{\sqrt{2}}\left(I-\frac{\bar{\mathbf{z}}\bar{\mathbf{z}}^T}{\|\bar{\mathbf{z}}\|_2(\|\bar{\mathbf{z}}\|_2+z_1)}\right)\end{pmatrix};$$

when $z_1 = -\|\bar{\mathbf{z}}\|_2$ or $\|\bar{\mathbf{z}}\|_2 = 0$,

$$K_{\mathbf{z}} \overset{\text{def}}{=} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & & \\ -\frac{1}{2} & \frac{1}{2} & & \\ & & -\frac{1}{\sqrt{2}} & \\ & & & \frac{1}{\sqrt{2}} I \end{pmatrix}.$$

It is easy to verify that $K_{\mathbf{z}}^{-1} = 2K_{\mathbf{z}}^T$, and

$$\mathbf{z} = K_{\mathbf{z}} \begin{pmatrix} 1 + \|\bar{\mathbf{z}}\|_2 \\ 1 - \|\bar{\mathbf{z}}\|_2 \\ \mathbf{0} \end{pmatrix}, \quad R\mathbf{z} = K_{\mathbf{z}} \begin{pmatrix} 1 - \|\bar{\mathbf{z}}\|_2 \\ 1 + \|\bar{\mathbf{z}}\|_2 \\ \mathbf{0} \end{pmatrix}. \tag{14}$$

3.2 Min Function

We represent the complementarity conditions in (1) by min functions.

$$(\omega_1 R\mathbf{z}_1; \ldots; \omega_n R\mathbf{z}_n) - A^T\mathbf{y} - \mathbf{c} = \mathbf{0} \tag{15a}$$

$$\lambda_1 A_1\mathbf{z}_1 + \cdots + \lambda_n A_n\mathbf{z}_n - \mathbf{b} = \mathbf{0} \tag{15b}$$

$$(\mathbf{z}_i)_0 - 1 = 0 \ (i = 1, \ldots, n) \tag{15c}$$

$$\min(\lambda_i, \omega_i, \mathbf{z}_i^T R\mathbf{z}_i) = 0 \ (i = 1, \ldots, n). \tag{15d}$$

For an element in the generalized Jacobian of (15), we left multiply $\mathrm{Diag}(K_{\mathbf{z}_i}^T)$ to row (15a), right multiply $\mathrm{Diag}(K_{\mathbf{z}_i})$ to columns corresponding to $\mathbf{z}$. Dropping the subscript for blocks, we write each of its block as the following.

	ω	λ	$\mathbf{z}^T$	$\mathbf{y}^T$
(15a)	$\frac{1}{2}\begin{pmatrix} 1-\|\bar{\mathbf{z}}\|_2 \\ 1+\|\bar{\mathbf{z}}\|_2 \\ \mathbf{0} \end{pmatrix}$		$\frac{\omega}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \\ & & -I \end{pmatrix}$	$-(AK_{\mathbf{z}})^T$
(15b)		$AK_{\mathbf{z}}\begin{pmatrix} 1+\|\bar{\mathbf{z}}\|_2 \\ 1-\|\bar{\mathbf{z}}\|_2 \\ \mathbf{0} \end{pmatrix}$	$\lambda AK_{\mathbf{z}}$	
(15c)			$\left(\frac{1}{2}, \frac{1}{2}, \mathbf{0}^T\right)$	
(15d)	p	q	$l\left(1 - \|\bar{\mathbf{z}}\|_2, 1 + \|\bar{\mathbf{z}}\|_2, \mathbf{0}^T\right)$	

Here, $0 \le \alpha \le 1$, and

$$\begin{cases} p = 1,\ q = l = 0 & 0 = \omega_i < \lambda_i,\ \mathbf{z}_i \in \operatorname{int} \mathcal{Q}; \\ q = 1,\ p = l = 0 & 0 = \lambda_i < \omega_i,\ \mathbf{z}_i \in \operatorname{int} \mathcal{Q}; \\ l = 1,\ p = q = 0 & 0 < \lambda_i,\ 0 < \omega_i,\ \mathbf{z}_i \in \operatorname{bd} \mathcal{Q}; \\ p = 1 - \alpha,\ l = \alpha, q = 0 & 0 = \omega_i < \lambda_i,\ \mathbf{z}_i \in \operatorname{bd} \mathcal{Q}; \\ q = 1 - \alpha,\ l = \alpha,\ p = 0 & 0 = \lambda_i < \omega_i,\ \mathbf{z}_i \in \operatorname{bd} \mathcal{Q}. \end{cases} \tag{16}$$

case 1: $\omega \neq 0$, $q = 0$ **.**
The conditions imply that $p = 0$, $\mathbf{z} \in \operatorname{bd} \mathcal{Q}$, $l = 1$.
Use $l(1 + \|\bar{\mathbf{z}}\|_2)$ to eliminate row (15d) and column z_1; then use $\frac{1}{2}(1 + \|\bar{\mathbf{z}}\|_2)$ to eliminate column ω and the $2nd$ row of (15a); next use $\frac{1}{2}$ to eliminate column z_0 and row (15c).
When $\lambda \neq 0$, case 1 includes solely L_1. This block is reduced to

	λ	$\bar{\mathbf{z}}^T$	$\mathbf{y}^T$
(15b)	$\frac{\omega}{2\lambda}\begin{pmatrix} 0 \\ & -I \end{pmatrix}$		$-(AK_{\mathbf{z}})_{\bar{1}}^T$
	$(AK_{\mathbf{z}})_{\bar{1}}$		

When $\lambda = 0$, case 1 includes exclusively L_4, and this block can be reduced to the following matrix.

	$\mathbf{y}^T$
	$-(AK_{\mathbf{z}})_0^T$
(15b)	$(AK_{\mathbf{z}})_0$

case 2: $\omega \neq 0$, $l = 0$ **.**
Clearly, $p = 0$, $\lambda = 0$, $q = 1$. This case includes just L_4 and L_5.
We first eliminate row (15d) and column λ by q, then subtract column z_0 from column z_1 to eliminate row (15c) and column z_0 by $\frac{1}{2}$. Next, we add the $1st$ row of (15a) to the $2nd$ row of (15a); so we can eliminate the $1st$ row of (15a) and column z_1 by $\frac{\omega}{2}$, the $2nd$ row of (15a) and column ω by 1, and the remaining by $-\frac{\omega}{2}I$. Thus we needn't worry about this block when analyzing the nonsingularity of (15) by assumption (13).
case 3: $\omega \neq 0$, $q \neq 0$, $l \neq 0$ **.**
Assume $l = \alpha$, $q = 1 - \alpha\,(0 < \alpha < 1)$. The conditions imply $\mathbf{z} \in \operatorname{bd} \mathcal{Q}$, $\lambda = 0$; so this case includes only L_4.
We first use $\frac{1}{2}(1+\|\bar{\mathbf{z}}\|_2)$ to delete column ω and the $2nd$ row of (15a), then use $\frac{1}{2}$ to eliminate column z_0 and row (15c), subtract $\frac{\omega}{4\alpha}$ multiplying (15d) from the $1st$ row of (15a). Hence this block can be reduced to the following matrix.

		$\mathbf{y}^T$
	$-\frac{(1-\alpha)\omega}{8\alpha}$	$-(AK_{\mathbf{z}})_0^T$
(15b)	$(AK_{\mathbf{z}})_0$	

case 4: $\omega = 0$, $l = 0$**.**
Under (13), $\lambda \neq 0$; hence $q = 0$, $p = 1$. This block consists of L_2 and L_3.
Assume dual nondegeneracy, then $(A_i K_{\mathbf{z}_i})$ has linearly independent columns (see [3]). Hence the column corresponding to λ is nonzero. We first eliminate column ω and row (15d) by p, then subtract the first two columns of $\mathbf{z}$ multiplying $\frac{1}{\lambda}(1 + \|\bar{\mathbf{z}}\|_2\,, 1 - \|\bar{\mathbf{z}}\|_2)^T$ from column λ. This block is reduced to the following matrix.

	$\mathbf{z}^T$	$\mathbf{y}^T$
		$-(AK_{\mathbf{z}})^T$
(15b)	$AK_{\mathbf{z}}$	

case 5: $\omega = 0$, $l \neq 0$**.**
For this block, $\lambda \neq 0$, $q = 0$ and $\mathbf{z} \in \operatorname{bd} \mathcal{Q}$. This block includes only L_2.
When $p = 0$, l must be 1.
We first use $\frac{1+\|\bar{\mathbf{z}}\|_2}{2}$ to eliminate column ω and the $2nd$ row of (15a), then use $(1 + \|\bar{\mathbf{z}}\|_2)$ to eliminate column z_1 and row (15d), use $\frac{1}{2}$ to get rid of column z_0 and row (15c). This block is reduced to the following matrix.

	$\lambda \quad \bar{\bar{\mathbf{z}}}^T$	$\mathbf{y}^T$
		$-(AK_{\mathbf{z}})_{\bar{1}}^T$
(15b)	$(AK_{\mathbf{z}})_{\bar{1}}$	

When $p = 1 - \alpha$, $l = \alpha\,(0 < \alpha < 1)$.
We first subtract $\frac{2}{\lambda}$ multiplying column z_0 from column λ to eliminate row (15c) and column λ, then subtract $\frac{1}{1-\alpha}$ multiplying row (15d) from the $2nd$ row of (15a). This block is reduced to the following matrix.

	$\mathbf{z}^T$	$\mathbf{y}^T$
	$\begin{pmatrix} 0 & -\frac{2\alpha}{(1-\alpha)\lambda} \\ & 0 \end{pmatrix}$	$-(AK_{\mathbf{z}})^T$
(15b)	$AK_{\mathbf{z}}$	

The nonsingularity result of (15) is the following.

Theorem 2 *Suppose a solution* $\mathbf{w}^* \stackrel{\text{def}}{=} (\boldsymbol{\omega}^*, \boldsymbol{\lambda}^*, \mathbf{z}^*, \mathbf{y}^*)$ *to* (15) *satisfies prima-dual nondegeneracy and also assumption*(13). *Then each element of the generalized Jacobian of* (15) *is nonsingular at* $\mathbf{w}^*$. *Let G denote* (15). *Then the sequence* (12) *converges Q-quadratically to* $\mathbf{w}^*$.

Proof: As in [2, 3], we partition the primal variable $\mathbf{x}_i = \lambda_i \mathbf{z}_i$ into three parts $\mathbf{x} = (\mathbf{x}_B; \mathbf{x}_I; \mathbf{x}_O)$, where $\mathbf{x}_B$ is the collection of all the boundary blocks, $\mathbf{x}_I$ includes all the interior blocks, and $\mathbf{x}_O$ collects all the zero blocks. Rearrange the order of the index set so that $\mathbf{x}_B = (\mathbf{x}_1, \ldots, \mathbf{x}_r)$. It is shown in [2, 3] that primal nondegeneracy means matrix in the following form having linearly independent rows for all $\alpha_1, \ldots, \alpha_r$ and $\boldsymbol{\nu}$ that are not all zeros.

$$\begin{pmatrix} A_1 & \ldots & A_r & A_I & A_O \\ \alpha_1 (R\mathbf{x}_1)^T & \ldots & \alpha_r (R\mathbf{x}_r)^T & \mathbf{0}^T & \boldsymbol{\nu}^T \end{pmatrix} \tag{17}$$

Since right multiplying a nonsingular matrix to a full row rank matrix doesn't change the row rank of the latter, we right multiplying $\mathrm{Diag}(K_{\mathbf{z}_i})$ to (17). Notice $i \in B$ means $\|\bar{\mathbf{z}}\|_2 = 1$, $\lambda_i > 0$. By [3, Lemma 4] ([2, Lemma 19]), primal nondegeneracy is equivalent to:

$$\begin{bmatrix} (A_1 K_{\mathbf{z}_1})_0 & (A_1 K_{\mathbf{z}_1})_{2:N_1} & \ldots & (A_r K_{\mathbf{z}_r})_0 & (A_r K_{\mathbf{z}_r})_{2:N_r} & A_I K_{\mathbf{z}_I} \end{bmatrix}$$

has linearly independent rows.
Using the notation of $L_1, \ldots, L_5$, we write the above matrix as

$$\begin{bmatrix} ((AK_{\mathbf{z}})_{L_1 L_2})_{\bar{1}} & (AK_{\mathbf{z}})_{L_3} \end{bmatrix}.$$

Similarly, we partition the dual variable $\mathbf{s}_i = \omega_i R\mathbf{z}_i$ into $\mathbf{s} = (\mathbf{s}_B; \mathbf{s}_I; \mathbf{s}_O)$, with $\mathbf{s}_B = (\mathbf{s}_1; \ldots; \mathbf{s}_s)$ being the concatenation of boundary blocks of $\mathbf{s}$, $\mathbf{s}_I$ including all the interior blocks of $\mathbf{s}$, and $\mathbf{s}_O$ collecting all the zero blocks of $\mathbf{s}$. And A is partitioned in the same manner: $A = (\tilde{A}_B, \tilde{A}_I, \tilde{A}_O)$. By [3], dual nondegeneracy means

$$\begin{pmatrix} \tilde{A}_1 R\mathbf{s}_1 & \ldots & \tilde{A}_s R\mathbf{s}_s & \tilde{A}_O \end{pmatrix} \tag{18}$$

has linearly independent columns. Since $\tilde{A}_i R\mathbf{s}_i = \tilde{A}_i K_{\mathbf{z}_i} K_{\mathbf{z}_i}^{-1} R\mathbf{s}_i$, by (14), we see that dual nondegeneracy means

$$\begin{bmatrix} \left(\tilde{A}_1 K_{\mathbf{z}_1}\right)_0 & \ldots \left(\tilde{A}_s K_{\mathbf{z}_s}\right)_0 & \tilde{A}_O K_{\mathbf{z}_O} \end{bmatrix}$$

has linearly independent columns.
The above matrix can be written as

$$\begin{bmatrix} ((AK_{\mathbf{z}})_{L_1 L_4})_0 & (AK_{\mathbf{z}})_{L_2 L_3} \end{bmatrix}.$$

Notice adding some columns to a full row rank matrix doesn't change the row rank; and after deleting some columns from a full column rank matrix, the matrix still has full column rank. Hence, primal nondegeneracy implies

$$\begin{bmatrix} \left((AK_{\mathbf{z}})_{L_1}\right)_{\bar{1}} \left((AK_{\mathbf{z}})_{L_{4(l\neq 0)}}\right)_0 & \left((AK_{\mathbf{z}})_{L_{2(p=0)}}\right)_{\bar{1}} & (AK_{\mathbf{z}})_{L_{2(p\neq 0)} L_3} \end{bmatrix}$$

has linearly independent rows.
Dual nondegeneracy implies

$$\begin{bmatrix} \left((AK_{\mathbf{z}})_{L_1}\right)_0 \left((AK_{\mathbf{z}})_{L_{4(l\neq 0)}}\right)_0 & \left((AK_{\mathbf{z}})_{L_{2(p=0)}}\right)_{\bar{1}} & (AK_{\mathbf{z}})_{L_{2(p\neq 0)} L_3} \end{bmatrix}$$

has linearly independent columns.

As in [4], we choose all columns of $\left((AK_{\mathbf{z}})_{L_1 L_{4(l\neq 0)}}\right)_0$, $\left((AK_{\mathbf{z}})_{L_{2(p=0)}}\right)_{\bar{1}}$, and $(AK_{\mathbf{z}})_{L_{2(p\neq 0)}L_3}$, along with some columns from $((AK_{\mathbf{z}})_{L_1})_{2:n}$ to form an m by m nonsingular matrix B_1, and collect the remaining columns of $((AK_{\mathbf{z}})_{L_1})_{2:n}$ into B_2. The nonsingularity of ∂G is reduced to the following matrix being nonsingular:

$$\begin{pmatrix} \tilde{I}_1 & & -B_1^T \\ & \tilde{I}_2 & -B_2^T \\ B_1 & B_2 & \end{pmatrix}. \tag{19}$$

Here, $\tilde{I}_1$ is a diagonal matrix with diagonal elements $\left[0, -\left(\frac{\omega_i}{2\lambda_i}\right)_{i\in L_1}, -\left(\frac{(1-\alpha_i)\omega_i}{8\alpha_i}\right)_{\substack{i\in L_4 \\ (l_i\neq 0, q_i\neq 0)}}, -\left(\frac{2\alpha_i}{(1-\alpha_i)\lambda_i}\right)_{\substack{i\in L_2 \\ (l_i\neq 0, p_i\neq 0)}}\right]$. And $\tilde{I}_2 = -\mathrm{Diag}\left(\frac{\omega_i}{2\lambda_i} I\right)_{i\in L_1}$. By the second condition of (13), B_1 is nonempty.

We first subtract $\tilde{I}_1 B_1^{-1}$ left multiplying the $3rd$ block row of (19) from the $1st$ block row of (19), then add $\tilde{I}_1 B_1^{-1} B_2 \tilde{I}_2^{-1}$ multiplying the $2nd$ block row to the $1st$ block row of (19). Hence (19) is reduced to

$$B_1^T + \tilde{I}_1 B_1^{-1} B_2 \tilde{I}_2^{-1} B_2^T = (I + \tilde{I}_1 B_1^{-1} B_2 \tilde{I}_2^{-1} B_2^T B_1^{-T}) B_1^T,$$

which is nonsingular because it is in the form $(I + N_1 N_2) B_1^T$ with $N_1 = \tilde{I}_1$ and N_2 symmetric negative semidefinite and B_1^T nonsingular. Even if B_2 is empty, (19) is still nonsingular.

It is easy to verify that $\min(\omega_i, \lambda_i, \mathbf{z}_i^T R \mathbf{z}_i)$ is strongly semismooth. In view of the arguments following Definition 2, we conclude that (15) is strongly semismooth. By Theorem 1, we see the Q-quadratic convergence rate of the sequence (12). ■

3.3 $[\cdot]_+$ and $[\cdot]_-$ Functions

We reformulate (1) into the following system,

$$([\omega_1]_+ R\mathbf{z}_1; \ldots; [\omega_n]_+ R\mathbf{z}_n) - A^T\mathbf{y} - \mathbf{c} = \mathbf{0} \tag{20a}$$

$$[\lambda_1]_+ A_1\mathbf{z}_1 + \cdots + [\lambda_n]_+ A_n\mathbf{z}_n - \mathbf{b} = \mathbf{0} \tag{20b}$$

$$(\mathbf{z}_i)_0 - 1 = 0 \ (i = 1, \ldots, n) \tag{20c}$$

$$[\omega_i]_- + [\lambda_i]_- + \mathbf{z}_i^T R \mathbf{z}_i = 0 \ (i = 1, \ldots, n). \tag{20d}$$

Assumption 13 implies one of λ_i and ω_i must be positive $(i = 1, \ldots, n)$.

As in the previous section, we left multiply $\mathrm{Diag}(K_{\mathbf{z}_i}^T)$ to block row (20a), right multiply $\mathrm{Diag}(K_{\mathbf{z}_i})$ to block column $\mathbf{z}$, and omit the subscript for block numbers. Then each block of the generalized Jacobian is in the following form.

	ω	λ	$\mathbf{z}^T$	$\mathbf{y}^T$
(20a)	$\frac{p}{2}\begin{pmatrix} 1-\|\bar{\mathbf{z}}\|_2 \\ 1+\|\bar{\mathbf{z}}\|_2 \\ \mathbf{0} \end{pmatrix}$		$\frac{[\omega]_+}{2}\begin{pmatrix} 0 & 1 & \\ 1 & 0 & \\ & & -I \end{pmatrix}$	$-(AK_{\mathbf{z}})^T$
(20b)		$qAK_{\mathbf{z}}\begin{pmatrix} 1+\|\bar{\mathbf{z}}\|_2 \\ 1-\|\bar{\mathbf{z}}\|_2 \\ \mathbf{0} \end{pmatrix}$	$[\lambda]_+ AK_{\mathbf{z}}$	
(20c)			$(\frac{1}{2}, \frac{1}{2}, \mathbf{0}^T)$	
(20d)	$1-p$	$1-q$	$(1-\|\bar{\mathbf{z}}\|_2, 1+\|\bar{\mathbf{z}}\|_2, \mathbf{0}^T)$	

Hence p and q meet the conditions below:

$$\begin{cases} p = 1 & \omega_i > 0 \\ p \in [0,1] & \omega_i = 0 \\ p = 0 & \omega_i < 0 \end{cases}, \quad \begin{cases} q = 1 & \lambda_i > 0 \\ q \in [0,1] & \lambda_i = 0 \\ q = 0 & \lambda_i < 0 \end{cases}. \tag{21}$$

case 1 : $p = 1, q = 1$.
In this case, $\mathbf{z} \in \operatorname{bd} \mathcal{Q}$.
We first eliminate row (20d) and column z_1 by $(1 + \|\bar{\mathbf{z}}\|_2)$, then eliminate $2nd$ row of (20a) and column ω by $\frac{p}{2}(1 + \|\bar{\mathbf{z}}\|_2)$, eliminate row (20c) and column z_0 by $\frac{1}{2}$.
When $\lambda = 0$, by assumption (13), $\omega > 0$. Therefore, case 1 is made up of L_4. This block is reduced to the following matrix.

	λ	$\mathbf{y}^T$
$(20a)_0$		$-(AK_{\mathbf{z}})_0^T$
	$(AK_{\mathbf{z}})_0$	

When $\lambda \neq 0$, case 1 consists of L_1 and L_2. This block is reduced to the following matrix.

	λ $\bar{\bar{\mathbf{z}}}^T$	$\mathbf{y}^T$
(20a)	$\frac{[\omega]_+}{[\lambda]_+} \begin{pmatrix} 0 & \\ & -I \end{pmatrix}$	$-(AK_{\mathbf{z}})_{\bar{1}}^T$
	$(AK_{\mathbf{z}})_{\bar{1}}$	

case 2: $p = 1$, $q = 0$.
It is constituted by L_4 and L_5. We first eliminate column λ and row (20d) by $(1 - q)$. Then we subtract column z_0 from column z_1, add the $1st$ row of (20a) to the $2nd$ row of (20a), so that we can eliminate column z_1 and the $1st$ row of (20a) by $\frac{\omega}{2}$, remove column ω and the $2nd$ row of (20a) by $\frac{1}{2}(1 + \|\bar{\mathbf{z}}\|_2)$, delete column z_0 and row (20c) by $\frac{1}{2}$. Therefore, to analyze the nonsingularity of generalized Jacobian of (20), we can ignore this block by assumption (13), since this type of blocks are not the only blocks.
case 3: $p = 1$, $0 < q < 1$.
By (20d) and assumption (13), this case includes just L_4. We first eliminate column ω and the $2nd$ row of (20a) by $\frac{1}{2}(1 + \|\bar{\mathbf{z}}\|_2)$, then eliminate column z_0 and row (20c) by $\frac{1}{2}$, and subtract $\frac{1-q}{2}$ multiplying column z_1 from column ω. Hence this block is reduced to the following matrix.

		$\mathbf{y}^T$
$(20a)_0$	$\frac{q-1}{8q}\omega$	$-(AK_{\mathbf{z}})_0^T$
	$(AK_{\mathbf{z}})_0$	

case 4: $p = 0$.
This case consists of L_2 and L_3. It implies $q = 1$. We first eliminate column ω and row (20d) by 1, then subtract $\frac{1}{\lambda}\begin{pmatrix} 1+\|\bar{\mathbf{z}}\|_2 \\ 1-\|\bar{\mathbf{z}}\|_2 \end{pmatrix}$ multiplying first two columns of $\mathbf{z}$ from column λ. This block is reduced to the following matrix.

	$\mathbf{y}^T$
(20a)	$-(AK_{\mathbf{z}})^T$
$AK_{\mathbf{z}}$	

case 5: $0 < p < 1$.
Only the set L_2 is included in this case.
Assume dual nondegeneracy, then $(A_i K_{\mathbf{z}_i})_0 \neq \mathbf{0}$ (see [3, 2]). We first subtract $\frac{2}{\lambda}$ multiplying z_0 from column λ; hence we can eliminate column λ and row (20c); then subtract $\frac{p}{1-p}$ multiplying row (20d) from the $2nd$ row of (20a). This block can be reduced to the following matrix.

		$\mathbf{y}^T$
(20a)	$\begin{pmatrix} 0 & \\ -\frac{2p}{(1-p)\lambda} & 0 \end{pmatrix}$	$-(AK_{\mathbf{z}})^T$
	$AK_{\mathbf{z}}$	

Theorem 3 *Assume a solution* $\mathbf{w}^* \stackrel{\text{def}}{=} (\boldsymbol{\omega}^*, \boldsymbol{\lambda}^*, \mathbf{z}^*, \mathbf{y}^*)$ *to* (20) *satisfies primal-dual nondegeneracy and also assumption*(13). *Then each element of the generalized Jacobian of* (20) *is nonsingular at* $\mathbf{w}^*$. *And the sequence* (12) *converges Q-quadratically to* $\mathbf{w}^*$.

Proof: Similar to the proof of Theorem 2. The nonsingular matrix B_1 consists of $\left((AK_{\mathbf{z}})_{L_1, L_{4(\mathbf{q}\neq 0)}}\right)_0$, $\left((AK_{\mathbf{z}})_{L_{2(p=1)}}\right)_{\bar{1}}$, $(AK_{\mathbf{z}})_{L_{2(p\neq 1)}L_3}$, and some columns from $((AK_{\mathbf{z}})_{L_1})_{2:n}$. It is easy to see that $[a]_+$ and $[a]_-$ are strongly semismooth. ■

3.4 General Complementarity Functions

Note that the complementarity condition of (2) is a vertical complementarity problems. By introducing more variables, it can be cast as a box-constrained variational inequality problem (see [14]). When complementarity involves three variables a, b, and c, another way to convert the complementarity into equations is first to replace one of the variables, say b by its absolute value $|b|$, then to apply some nonlinear complementarity functions: $M(a, |b|c)$. One can also apply M twice, use $M\left[a, M(b, c)\right]$.

In this part, we consider a category of nonlinear complementarity functions $M(a, b)$ whose gradients satisfy:

$$\begin{cases} \nabla M(a,b) = (p, 0)^T & (a = 0, b > 0) \\ \nabla M(a,b) = (0, q)^T & (a > 0, b = 0) \end{cases},$$

where $p \neq 0$, $q \neq 0$.
This category of nonlinear complementarity functions include:

$$-\beta ab + [\min(0, a+b)]^2 = 0 \text{ (for } \beta \in (0,2]) \quad \text{see [47, 11, 24]},$$
$$\min(a, b) = 0,$$
$$\phi(a,b) = \sqrt{a^2 + b^2} - a - b = 0 \quad \text{see [15]},$$
$$\theta(|a-b|) - \theta(a) - \theta(b) = 0 \text{ (see [30])}, \text{ where}$$
$$\theta(t) \text{ is a differentiable strictly increasing function from } \mathbb{R} \text{ to } \mathbb{R} \text{ such that}$$
$$\theta'(0) + \theta'(\zeta) \neq 0 \text{ for all } \zeta > 0.$$

One can reformulate (3) into a system of nonlinear equations by replacing a, b, and c with λ_i, ω_i, and $\mathbf{z}_i^T R \mathbf{z}_i$ for $i = 1, \ldots, n$ in $M(a, |b|c)$ or $M\left[a, M(b, c)\right]$. If one uses the nonlinear complementarity functions listed above, the resulting systems are strongly semismooth according to the arguments below Definition 2. Assume strict complementarity is satisfied. Then the only nondifferentiable points of the resulting systems are at $\omega_i = 0$ for $M(\lambda_i, |\omega_i| \mathbf{z}_i^T R \mathbf{z}_i)$; moreover, it is easy to verify that the structure of the Jacobians of these systems are the same as that of (15). Hence one can conclude the following.

Corollary 1 *Under primal-dual nondegeneracy, strict complementarity assumptions, and* (13), *each element in the generalized Jacobian is nonsingular. The sequence* (12) *converges Q-quadratically to* $\mathbf{w}^*$ *for all the above systems.*

3.5 Jordan Algebra Formulation

This formulation is from [5]. Each $\mathbf{z} \in Q_{n+1}$ with $z_0 = 1$ can be decomposed into

$$\mathbf{z} = \alpha(1; \bar{\mathbf{u}}) + \beta(1; -\bar{\mathbf{u}}), \tag{22}$$

where

$$\alpha = \frac{1+\|\bar{\mathbf{z}}\|_2}{2}, \quad \beta = \frac{1-\|\bar{\mathbf{z}}\|_2}{2}, \quad \bar{\mathbf{u}} = \begin{cases} \binom{1}{\mathbf{0}} & \bar{\mathbf{z}} = \mathbf{0} \\ \frac{\bar{\mathbf{z}}}{\|\bar{\mathbf{z}}\|_2} & \text{otherwise.} \end{cases}$$

The decomposition is unique if $\alpha \geq \beta$. So one can write the primal and dual variables $\lambda\mathbf{z}$, $\omega R\mathbf{z}$ as

$$\lambda\mathbf{z} = \lambda_1(1;\bar{\mathbf{u}}) + \lambda_2(1;-\bar{\mathbf{u}}), \quad \omega R\mathbf{z} = \omega_1(1;\bar{\mathbf{u}}) + \omega_2(1;-\bar{\mathbf{u}}),$$

with $\lambda_1 \geq 0$, $\lambda_2 \geq 0$, $\omega_1 \geq 0$, $\omega_2 \geq 0$. Note that the primal (or dual) variable is in int $\mathcal{Q}$ (or bd $\mathcal{Q}$) iff both (or one of) λ_i (or ω_i) are (is) positive. Thus (3d) is equivalent to $\lambda_1\omega_1 = \lambda_2\omega_2 = 0$. Below, we reformulate (1) by min function. It can be generalized to any complementarity function.

$$(\omega_i)_1(1;\bar{\mathbf{u}}_i) + (\omega_i)_2(1;-\bar{\mathbf{u}}_i) - A_i^T\mathbf{y} - \mathbf{c}_i = \mathbf{0}\ (i=1,\ldots,n) \tag{23a}$$

$$A_1\left((\lambda_1)_1(1;\bar{\mathbf{u}}_1) + (\lambda_1)_2(1;-\bar{\mathbf{u}}_1)\right) + \cdots + A_n\left((\lambda_n)_1(1;\bar{\mathbf{u}}_n) + (\lambda_n)_2(1;-\bar{\mathbf{u}}_2)\right) - \mathbf{b} = \mathbf{0} \tag{23b}$$

$$\min\left((\lambda_i)_1, (\omega_i)_1\right) = 0\ (i=1,\ldots,n) \tag{23c}$$

$$\min\left((\lambda_i)_2, (\omega_i)_2\right) = 0\ (i=1,\ldots,n) \tag{23d}$$

$$\bar{\mathbf{u}}_i^T\bar{\mathbf{u}}_i = 1\ (i=1,\ldots,n) \tag{23e}$$

For any $(1;\bar{\mathbf{u}}) \in \mathbb{R}^{n+1}$ with $\|\bar{\mathbf{u}}\|_2 = 1$, define an n by n orthogonal matrix $L_{\bar{\mathbf{u}}}$:

$$L_{\bar{\mathbf{u}}} \overset{\text{def}}{=} \begin{cases} \begin{pmatrix} u_1 & -\bar{\bar{\mathbf{u}}}^T \\ \bar{\bar{\mathbf{u}}} & I - \frac{\bar{\bar{\mathbf{u}}}\bar{\bar{\mathbf{u}}}^T}{1+u_1} \end{pmatrix} & u_1 \neq -1 \\ \begin{pmatrix} -1 & \\ & I \end{pmatrix} & u_1 = -1 \end{cases}.$$

Denote $\left(\bar{A}L_{\bar{u}}\right)_{\bar{1}}$ as the submatrix of $\left(\bar{A}L_{\bar{u}}\right)$ excluding the first column. We left multiply Diag $\begin{pmatrix} 1 & \\ & L_{\bar{\mathbf{u}}_i}^T \end{pmatrix}$ to block rows (23a) of the generalized Jacobian of (23), right multiply Diag $(L_{\bar{\mathbf{u}}_i})$ to its block columns $\bar{\mathbf{u}}^T$. Dropping the subscript for block numbers, we write each block as the following.

	ω_1	ω_2	λ_1	λ_2	$\bar{\mathbf{u}}^T$	$\mathbf{y}^T$
(23a)	$\begin{pmatrix} 1 \\ 1 \\ \mathbf{0} \end{pmatrix}$	$\begin{pmatrix} 1 \\ -1 \\ \mathbf{0} \end{pmatrix}$			$(\omega_1-\omega_2)\begin{pmatrix} \mathbf{0}^T \\ I \end{pmatrix}$	$\begin{matrix} -\mathbf{a}_0^T \\ -(\bar{A}L_{\bar{\mathbf{u}}})^T \end{matrix}$
(23b)			$\mathbf{a}_0 + (\bar{A}L_{\bar{\mathbf{u}}})_1$	$\mathbf{a}_0 - (\bar{A}L_{\bar{\mathbf{u}}})_1$	$(\lambda_1-\lambda_2)\,\bar{A}L_{\bar{\mathbf{u}}}$	
(23c)	p_1		q_1			
(23d)		p_2		q_2		
(23e)					$2(1,\mathbf{0}^T)$	

Here, for $(j=1,2)$,

$$\begin{cases} p_j = 0,\ q_j = 1 & 0 = (\lambda_i)_j < (\omega_i)_j; \\ p_j = 1,\ q_j = 0 & 0 = (\omega_i)_j < (\lambda_i)_j; \\ p_j = 1-\alpha,\ q_j = \alpha\ (0 \leq \alpha \leq 1) & 0 = (\lambda_i)_j = (\omega_i)_j. \end{cases}$$

We assume

$$\lambda_1 \geq \lambda_2, \quad \omega_1 \leq \omega_2; \tag{24}$$

$$(\omega_1 - \omega_2) \text{ and } (\lambda_1 - \lambda_2) \text{ not be zero at the same time.} \tag{25}$$

case 1: $p_1 = p_2 = 0$**.**
Hence $q_1 = q_2 = 1$. This case is constituted by L_4 and L_5. After adding column ω_1 to column ω_2,

we find that to analyze regularity of the generalized Jacobian of (23), we can forget about this case by assumption (13).

case 2: $q_1 = q_2 = 0$ **.**

Thus $p_1 = p_2 = 1$. This case consists of L_2 and L_3 only. We first add column λ_2 to column λ_1, then subtract $\frac{1}{2}$ multiplying column λ_1 from column λ_2. To analyze regularity of the generalized Jacobian, this block can be transformed to the following matrix:

$$\begin{array}{c|cc} & & \mathbf{y}^T \\ \hline (23b) & & -\left[A\begin{pmatrix} 1 & \\ & L_{\bar{\mathbf{u}}} \end{pmatrix}\begin{pmatrix} 1 & 1 & \\ 1 & -1 & \\ & & I \end{pmatrix}\right]^T \\ & \left[A\begin{pmatrix} 1 & \\ & L_{\bar{\mathbf{u}}} \end{pmatrix}\begin{pmatrix} 1 & 1 & \\ 1 & -1 & \\ & & I \end{pmatrix}\right] & \end{array}$$

case 3: $p_1 \neq 0$, $q_2 \neq 0$, $q_1 = p_2 = 0$**.**

Hence $\omega_1 = \lambda_2 = 0$.

We first eliminate column ω_1 and row (23c) by p_1, column λ_2 and row (23d) by q_2, column u_1 and row (23e) by 2. Then we add the first row of (23a) to the second row of (23a) to eliminate column ω_2 and the $1st$ row of (23a).

When $\lambda_1 \neq 0$, $\omega_2 = 0$, this case includes solely L_2. When $\lambda_1 \neq 0$, $\omega_2 \neq 0$, this case includes just L_1. This block can be reduced to the following matrix.

$$\begin{array}{c|cc} & \lambda_1 \quad \bar{\bar{\mathbf{u}}}^T & \mathbf{y}^T \\ \hline (23b) & \begin{pmatrix} 0 & \\ -\frac{\omega_2}{\lambda_1} & I \end{pmatrix} & \begin{matrix} -\mathbf{a}_0^T - (\bar{A}L_{\bar{\mathbf{u}}})_1^T \\ -(\bar{A}L_{\bar{\mathbf{u}}})_{\bar{1}}^T \end{matrix} \\ & \begin{matrix} \mathbf{a}_0 + (\bar{A}L_{\bar{\mathbf{u}}})_1 & (\bar{A}L_{\bar{\mathbf{u}}})_{\bar{1}} \end{matrix} & \end{array}$$

When $\lambda_1 = 0$, hence $\omega_2 \neq 0$. Case 2 includes only L_4. This block can be reduced to the following matrix.

$$\begin{array}{c|cc} & \lambda_1 & \mathbf{y}^T \\ \hline (23b)_0 & & -\mathbf{a}_0^T - (\bar{A}L_{\bar{\mathbf{u}}})_1^T \\ & \mathbf{a}_0 + (\bar{A}L_{\bar{\mathbf{u}}})_1 & \end{array}$$

case 4: $p_1 \neq 0$, $q_1 \neq 0$**.**

Hence, $q_2 = 1$, $p_2 = 0$. This case includes merely L_4. Assume $p_1 = 1 - \alpha$, $q_1 = \alpha$ $(0 < \alpha < 1)$. We first remove column λ_2 and row (23d) by q_2, column $\mathbf{u}_1$ and row (23e) by 2, then add the first row of (23a) to the second row of (23a) to get rid of column ω_2 and the first row of (15a) by 1. Then we subtract $\frac{\alpha}{1-\alpha}$ multiplying column ω_1 from column λ_1. This block can be reduced to the following matrix.

$$\begin{array}{c|cc} & & \mathbf{y}^T \\ \hline (23b)_0 & -\frac{2\alpha}{1-\alpha} & -\mathbf{a}_0^T - (\bar{A}L_{\bar{\mathbf{u}}})_1^T \\ & \mathbf{a}_0 + (\bar{A}L_{\bar{\mathbf{u}}})_1 & \end{array}$$

case 5: $p_2 \neq 0$, $q_2 \neq 0$ **.**

Hence, $q_1 = 0$, $p_1 = 1$. This case consists of L_2. Assume $p_2 = 1 - \alpha$, $q_2 = \alpha$ $(0 < \alpha < 1)$. We first eliminate column ω_1 and row (23c) by 1, column u_1 and row (23e) by 2, then add the $1st$ row of (23a) to the $2nd$ row of (23a). Next, we subtract $\frac{\alpha}{1-\alpha}$ multiplying column ω_2 from column λ_2 to eliminate column ω_2 and row (23d). Then subtract $\frac{1}{2}$ multiplying the $2nd$ row of (23a) from the $1st$ row of (23a). Hence, this block is reduced to the following matrix.

$$\begin{array}{c|cc} & & \mathbf{y} \\ \hline (23b) & \begin{pmatrix} 0 & \\ -\frac{2\alpha}{1-\alpha} & \\ & 0 \end{pmatrix} & -\left[A\begin{pmatrix} 1 & \\ & L_{\bar{\mathbf{u}}} \end{pmatrix}\begin{pmatrix} 1 & 1 & \\ 1 & -1 & \\ & & I \end{pmatrix}\right]^T \\ & \left[A\begin{pmatrix} 1 & \\ & L_{\bar{\mathbf{u}}} \end{pmatrix}\begin{pmatrix} 1 & 1 & \\ 1 & -1 & \\ & & I \end{pmatrix}\right] & \end{array}$$

Theorem 4 *Suppose a solution* $\mathbf{w}^* \stackrel{\text{def}}{=} (\boldsymbol{\omega}^*, \boldsymbol{\lambda}^*, \mathbf{u}^*, \mathbf{y}^*)$ *to* (23) *satisfies prima-dual nondegeneracy and also assumptions* (24) *and* (25). *Then every element in the generalized Jacobian of* (23) *is nonsingular at* $\mathbf{w}^*$. *Therefore, the sequence* (12) *converges Q-quadratically to* $\mathbf{w}^*$.

Proof: We right multiply $\mathrm{Diag}\left[\begin{pmatrix} 1 & \\ & L_{\bar{\mathbf{u}}_i} \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}_I\right]$ to (17). Denote $B \stackrel{\text{def}}{=} A\mathrm{Diag}\left[\begin{pmatrix} 1 & \\ & L_{\bar{\mathbf{u}}_i} \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}_I\right]$. After adding some columns to (17) and deleting some columns from (18), one finds that primal nondegeneracy implies $((B_{L_1 L_{2(p_2=0)}})_{\bar{1}}\ B_{L_{2(p_2\neq 0)}L_3}\ (B_{L_4(p_1\neq 0)})_0)$ having linearly independent rows. Dual nondegeneracy implies $((B_{L_1 L_{4(p_1\neq 0)}})_0\ (B_{L_2(p_2=0)})_{\bar{1}}\ B_{L_{2(p_2\neq 0)}L_3})$ having linearly independent columns.
The proof of Theorem 2 can be carried over here with $\tilde{I}_2 = \mathrm{Diag}\left(-\frac{(\omega_i)_2}{(\lambda_i)_1} I\right)_{L_1}$ and diagonal elements of $\tilde{I}_1$ being

$$\left[0,\ \left(-\frac{2\alpha_i}{1-\alpha_i}\right)_{\substack{L_4 \\ \left(\substack{p_1\neq 0, \\ q_1\neq 0}\right)}},\ \left(\begin{matrix} 0 & -\frac{2\alpha_i}{1-\alpha_i} \\ 0 & \end{matrix}\right)_{\substack{L_2 \\ \left(\substack{p_2\neq 0, \\ q_2\neq 0}\right)}},\ \left(-\frac{(\omega_i)_2}{(\lambda_i)_1} I\right)_{L_1}\right].$$

■

4 Perturbation

The main results of this section is the following.

Theorem 5 *Let* G^{old} *denote any of the systems* (15), (20), (23), *or the systems in § 3.4. Let* $\mathbf{w}^{old}$ *be a solution to* G^{old}. *Suppose every element in* $\partial G^{old}(\mathbf{w}^{old})$ *is nonsingular. Then there exist positive scalars* ν *and* υ, *such that if the perturbations* $(\Delta A, \Delta \mathbf{b}, \Delta \mathbf{c})$ *satisfy*

$$\|\Delta A\|_2 \leq \nu, \quad \left\|(\Delta \mathbf{c} + \Delta A^T \mathbf{y}^{old}; \Delta \mathbf{b} - \Delta A \mathbf{x}^{old})\right\|_\infty < \upsilon,$$

then the perturbed problem is solvable, and starting from $\mathbf{w}^{old}$, *the iterates* (12) *converge Q-quadratically to a solution of the new problem* $\mathbf{w}^{new}$.

The perturbation of A also includes addition or deletion of variables or constraints, because one only needs to view the added variables as 0 and the deleted variables as constants at the original problem. To prove the theorem, we first give a lemma.

Lemma 2 *In a neighborhood* $\mathcal{N}(\mathbf{w}^{old})$ *of* $\mathbf{w}^{old}$, *there exists a constant* $\rho \geq 0$ *independent of* $\mathbf{w} \in \mathcal{N}(\mathbf{w}^{old})$, *such that for any* $\mathbf{w} + \Delta \mathbf{w} \in \mathcal{N}(\mathbf{w}^{old})$, $V \in \partial G(\mathbf{w} + \Delta \mathbf{w})$, *we have*

$$\begin{aligned} \|V \Delta \mathbf{w} - G'(\mathbf{w}; \Delta \mathbf{w})\|_2 &\leq \rho \|\Delta \mathbf{w}\|_2^2, \\ \|G(\mathbf{w} + \Delta \mathbf{w}) - G(\mathbf{w}) - G'(\mathbf{w}; \Delta \mathbf{w})\|_2 &\leq \rho \|\Delta \mathbf{w}\|_2^2. \end{aligned} \tag{26}$$

Proof: We only need to show that every component of G satisfies the lemma, and the intersection of these neighborhoods is nonempty.
1. $\min(a, b, c)$
Represent $(\lambda_i, \omega_i, \mathbf{z}_i^T R \mathbf{z}_i)$ as (a, b, c) in any order. Assume when $a < b \leq c$, then $a + \Delta a < b + \Delta b$ and $a + \Delta a < c + \Delta c$; when $a = b < c$, then $b + \Delta b < c$ and $a + \Delta a < c$; when $a = b = c$, then $(\Delta a, \Delta b, \Delta c)$ can be arbitrary. Let $\mathbf{w} \stackrel{\text{def}}{=} (a, b, c)^T$. Then the left-hand-sides of (26) are zero; so $\rho = 0$. When a is replaced by $\mathbf{z}^T R \mathbf{z}$, $\rho = 1$.
2. $[\cdot]_+$ and $[\cdot]_-$
Assume if $a < 0$, then $a + \Delta a < 0$; if $a > 0$, then $a + \Delta a > 0$; there is no restriction on Δa when $a = 0$. Then the left-hand-sides of (26) are zero; so $\rho = 0$.

3. The Fischer-Burmeister function $\phi(a,b)$

When $\lambda_i^2+\omega_i^2>0$, we require each element in $\mathcal{N}(\mathbf{w}^{\rm old})$ satisfy $\sqrt{\lambda_i^2+\omega_i^2}\geq\frac{1}{2}\sqrt{{\lambda_i}^{\rm old}{}^2+{\omega_i}^{\rm old}{}^2}$. It is easily verified that $\rho=\frac{4}{\sqrt{{a^{\rm old}}^2+{b^{\rm old}}^2}}$. No restriction on $\mathcal{N}$ if ${\lambda_i}^{\rm old}={\omega_i}^{\rm old}=0$.

For $\phi\left(\omega_i,\left(\frac{\mathbf{z}^TR\mathbf{z}}{2}\right)\right)$, we require each element in $\mathcal{N}(\mathbf{w}^{\rm old})\setminus\{\mathbf{w}^{\rm old}\}$ be differentiable, and satisfy $\sqrt{\omega_i^2+\left(\frac{\mathbf{z}_i^TR\mathbf{z}_i}{2}\right)^2}\geq\frac{\sqrt{{\omega_i}^{\rm old}{}^2+\left(\frac{\mathbf{z}_i^{\rm old}{}^TR\mathbf{z}_i^{\rm old}}{2}\right)^2}}{2}$. We also assume $\|(a;\mathbf{z})\|_2\leq 2$, $\|(\Delta\, a;\Delta\,\mathbf{z})\|_2\leq 2$. Hence $\rho=1+\frac{40}{\sqrt{{a^{\rm old}}^2+\left(\frac{\mathbf{z}^{\rm old}{}^TR\mathbf{z}^{\rm old}}{2}\right)^2}}$.

4. Others

All the other maps employed by G is either linear or in the form $\lambda A\mathbf{x}$, where λ is a scalar, $\mathbf{x}$ is a vector and A is a matrix. So the neighborhood for any map other than the above employed by (3) satisfying lemma 2 is the whole space and $\rho=\frac{1}{2}\|A\|_2$. ■

Now we proceed to prove the theorem.

Proof: We will first consider perturbations of $\mathbf{b}$ and $\mathbf{c}$.

Denote the perturbed system as $G=G^{\rm old}-(\Delta\,\mathbf{c};\Delta\,\mathbf{b};\mathbf{0})$. Obviously, for any $\mathbf{w}$, $\partial G(\mathbf{w})=\partial G^{\rm old}(\mathbf{w})$. Hence G is Lipschitzian near $\mathbf{w}^{\rm old}$, and each element in $\partial G(\mathbf{w}^{\rm old})$ is nonsingular. Let B indicate the open Euclidean unit ball, and $\bar{B}$ its closure. By [8, Lemma 1, Lemma 2 in Chapter 7], there exist positive δ and r, such that for any $\mathbf{w}\in\mathbf{w}^{\rm old}+rB$ and $V\in\partial G(\mathbf{w})$, $\left\|V^{-1}\right\|_2\leq\frac{1}{\delta}$. And if $\mathbf{w}_1$ and $\mathbf{w}_2$ lie in $\mathbf{w}^{\rm old}+r\bar{B}$, $\|G(\mathbf{w}_1)-G(\mathbf{w}_2)\|_2\geq\delta\|\mathbf{w}_1-\mathbf{w}_2\|_2$. Replacing r by lr with $0\leq l\leq 1$ in the proof of [8, Lemma 3 in Chapter 7.1], one can verify that $G(\mathbf{w}^{\rm old}+lrB)$ contains $G(\mathbf{w}^{\rm old})+(\frac{1}{2}lr\delta)B$.

We can always find $0<l^*\leq 1/2$, such that $\mathbf{w}^{\rm old}+2l^*rB\subseteq\mathcal{N}(\mathbf{w}^{\rm old})$ and $l^*\leq\min(\frac{\delta}{2\rho r},\frac{1}{r})$, where $\mathcal{N}(\mathbf{w}^{\rm old})$ and ρ are defined in the previous lemma. Suppose $\|(\Delta\,\mathbf{c};\Delta\,\mathbf{b})\|_2<\frac{1}{2}l^*r\delta$. Then the new problem has a solution, designated as $\mathbf{w}^{\rm new}$, contained in $\mathbf{w}^{\rm old}+l^*rB$. We will use induction to prove the Q-quadratic convergence of the sequence (12) to $\mathbf{w}^{\rm new}$ from $\mathbf{w}^{\rm old}$.

Apparently, $\left\|\mathbf{w}^{\rm old}-\mathbf{w}^{\rm new}\right\|_2<l^*r$. Assume $\left\|\mathbf{w}^k-\mathbf{w}^{\rm new}\right\|_2<l^*r$. Then

$$\left\|\mathbf{w}^k-\mathbf{w}^{\rm old}\right\|_2\leq\left\|\mathbf{w}^k-\mathbf{w}^{\rm new}\right\|_2+\left\|\mathbf{w}^k-\mathbf{w}^{\rm old}\right\|_2<2l^*r.$$

Thus, $\mathbf{w}^k\in\mathcal{N}(\mathbf{w}^{\rm old})$. Similar to the proof of Theorem 3.2 in [41], we see

$$\begin{aligned}\left\|\mathbf{w}^{k+1}-\mathbf{w}^{\rm new}\right\|_2&=\left\|\mathbf{w}^k-\mathbf{w}^{\rm new}-{V^k}^{-1}G(\mathbf{w}^k)\right\|_2\\&\leq\left\|{V^k}^{-1}\left[G(\mathbf{w}^k)-G(\mathbf{w}^{\rm new})-G'(\mathbf{w}^{\rm new};\mathbf{w}^k-\mathbf{w}^{\rm new})\right]\right\|_2\\&+\left\|{V^k}^{-1}\left[V^k(\mathbf{w}^k-\mathbf{w}^{\rm new})-G'(\mathbf{w}^{\rm new};\mathbf{w}^k-\mathbf{w}^{\rm new})\right]\right\|_2\\&\leq 2\frac{\rho}{\delta}\left\|\mathbf{w}^k-\mathbf{w}^{\rm new}\right\|_2^2.\end{aligned}$$

The last inequality is due to lemma 2. By induction,

$$\left\|\mathbf{w}^{k+1}-\mathbf{w}^{\rm new}\right\|_2\leq 2\frac{\rho}{\delta}\left\|\mathbf{w}^k-\mathbf{w}^{\rm new}\right\|_2^2<2\frac{\rho}{\delta}(l^*r)^2\leq l^*r.$$

Now we add perturbation of A.

Since $A\mathbf{x}=(A+\Delta\,A)\mathbf{x}-\Delta\,A\mathbf{x}$,

$$G(\mathbf{w}^{\rm old})-G^{\rm old}(\mathbf{w}^{\rm old})=\left(-\Delta\,\mathbf{c}-\Delta\,A^T\mathbf{y}^{\rm old};\Delta\,A\mathbf{x}^{\rm old}-\Delta\,\mathbf{b};\mathbf{0}\right).$$

Note that perturbations may only modify ρ in lemma 2, not $\mathcal{N}(\mathbf{w}^{\rm old})$. Also observe that only $\Delta\,A$ may change ρ, and ρ depends linearly on A by 4 of the proof of lemma 2. So there exists $\nu_1>0$, such that when $\|\Delta\,A\|_2\leq\nu_1$, we have $\rho^{\rm new}\leq 2\rho^{\rm old}$.

Because ∂G is uppersemicontinuous (see [8]), according to perturbation lemma, there exists a positive number ν_2, so that when $\|\Delta A\|_2 \leq \nu_2$, for any $\mathbf{w} \in \mathbf{w}^{\text{old}} + \frac{r}{2}B$, and any $V \in \partial G(\mathbf{w})$, we have $\left\|V^{-1}\right\|_2 \leq \frac{2}{\delta}$.
Therefore, $G(\mathbf{w}^{\text{old}} + \frac{1}{2}lrB)$ contains $G(\mathbf{w}^{\text{old}}) + \frac{1}{8}lr\delta B$.

Let $\nu = \min(\nu_1, \nu_2)$. Assume $\|\Delta A\|_2 \leq \nu$. Then as the proof above, let $l^* \leq \min\left(\frac{\delta}{4\rho^{\text{old}}r}, \frac{1}{r}, \frac{1}{2}\right)$, and $\upsilon = \frac{1}{8}l^*r\delta$, we can get Q-quadratic convergence for the sequence (12) to $\mathbf{w}^{\text{new}}$, too. ■

5 Globalization

Let G indicate one of the nonlinear equation reformulations of (1) given before. Denote $\mathbf{w} \stackrel{\text{def}}{=} (\boldsymbol{\omega}; \boldsymbol{\lambda}; \mathbf{z}; \mathbf{y})$. Set $\Psi(\mathbf{w}) \stackrel{\text{def}}{=} \frac{1}{2}G(\mathbf{w})^T G(\mathbf{w})$; then a root of G is the same thing as a global minimum of Ψ. Therefore, one may want to apply any gradient-type method for systems of nonlinear equations (see [10]) to globalize the algorithm. For our system, however, a difficulty involved in this strategy is that Ψ may not be differentiable everywhere. It is known that ϕ^2 (see [20]) and $\theta(t) = t|t|$ (see [30]) are differentiable, but not are the squares of other nonlinear complementarity functions. At a nondifferentiable point of Ψ, there're more than one element in $\partial(\Psi)$, which may result in ambiguity in line-search rules and stopping criteria. To overcome this problem, we propose two approaches: perturbation and reformulation.

5.1 Perturbation

This part is devoted to the globalized algorithm with perturbed Armijo-type stepsize rule.

The nondifferentiable points are isolated. Therefore, if the intended starting point is nondifferentiable, one can always find an initial differentiable point that is arbitrarily close to it. Choose constants $s > 0$, $\sigma > 0$, $\beta \in (0, 1), \gamma \in (\beta, 1)$. For each $k \geq 0$, $\Psi(\mathbf{w}^k)$ is differentiable.

1. Set $\alpha^{k,0} = s$, i = 0.

2. Find the largest nonnegative integer l for which

$$\Psi(\mathbf{w}^k) - \Psi\left(\mathbf{w}^k + \beta^l\alpha^{k,i}\,\Delta\,\mathbf{w}^k\right) \geq -\sigma\beta^l\alpha^{k,i}\nabla\Psi(\mathbf{w}^k)^T\,\Delta\,\mathbf{w}^k.$$

3. If $\Psi\left(\mathbf{w}^k + \beta^l\alpha^{k,i}\,\Delta\,\mathbf{w}^k\right)$ is nondifferentiable, find $t \in [\gamma, 1)$ so that $\Psi\left(\mathbf{w}^k + t\beta^l\alpha^{k,i}\,\Delta\,\mathbf{w}^k\right)$ is differentiable, set $\alpha^{k,i+1} = t\beta^l\alpha^{k,i}$, $i + 1 \to i$, go to step 2.
Otherwise, set $\alpha^k = \beta^l\alpha^{k,i}$, $\mathbf{w}^{k+1} = \mathbf{w}^k + \alpha^k\,\Delta\,\mathbf{w}$, $k + 1 \to k$.

It is not difficult to find t in step 3. Since $\Psi(\mathbf{w}^k)$ is differentiable, the stepsizes only should not be the following.

- $\sum_{i=1}^n |\lambda_i|A_i\mathbf{z}_i - \mathbf{b}$
  Nondifferentiable at $\lambda_i = 0$ for $i = 1, \ldots, n$. Therefore, one only needs to circumvent stepsizes $-\lambda_i^k/\,\Delta\,\lambda_i^k$ when $\Delta\,\lambda_i^k \neq 0$.

- $\min(\lambda_i, \omega_i, \mathbf{z}_i^T R\mathbf{z}_i)$ for $i = 1, \ldots, n$
  Nondifferentiable at $\lambda_i = \omega_i$, $\lambda_i = \mathbf{z}_i^T R\mathbf{z}_i$, $\omega_i = \mathbf{z}_i^T R\mathbf{z}_i$. So, stepsizes shouldn't be $\frac{\lambda_i^k - \omega_i^k}{\Delta\,\omega_i^k - \Delta\,\lambda_i^k}$ for $\Delta\,\lambda_i^k \neq \Delta\,\omega_i^k$;
  $\frac{\Delta\,\lambda_i^k - 2\mathbf{z}_i^{k\,T}R\,\Delta\,\mathbf{z}_i^k \pm \sqrt{\left(2\mathbf{z}_i^{k\,T}R\,\Delta\,\mathbf{z}_i^k - \Delta\,\lambda_i^k\right)^2 - 4\left(\Delta\,\mathbf{z}_i^{k\,T}R\,\Delta\,\mathbf{z}_i^k\right)\left(\mathbf{z}_i^{k\,T}R\mathbf{z}_i^k - \lambda_i^k\right)}}{2\,\Delta\,\mathbf{z}_i^{k\,T}R\,\Delta\,\mathbf{z}_i^k}$,
  $\frac{\Delta\,\omega_i^k - 2\mathbf{z}_i^{k\,T}R\,\Delta\,\mathbf{z}_i^k \pm \sqrt{\left(2\mathbf{z}_i^{k\,T}R\,\Delta\,\mathbf{z}_i^k - \Delta\,\omega_i^k\right)^2 - 4\left(\Delta\,\mathbf{z}_i^{k\,T}R\,\Delta\,\mathbf{z}_i^k\right)\left(\mathbf{z}_i^{k\,T}R\mathbf{z}_i^k - \omega_i^k\right)}}{2\,\Delta\,\mathbf{z}_i^{k\,T}R\,\Delta\,\mathbf{z}_i^k}$,
  if they're real and $\Delta\,\mathbf{z}_i^{k\,T}R\,\Delta\,\mathbf{z}_i^k \neq 0$.

- $[\cdot]_+$ and $[\cdot]_-$
  For $i = 1, \ldots, n$, nondifferentiable at $\lambda_i = 0$ and $\omega_i = 0$. So not to choose stepsizes $-\lambda_i^k / \Delta\, \lambda_i^k$, $-\omega_i^k / \Delta\, \omega_i^k$.

All the other nondifferentiable functions are modifications of the above functions; Therefore, the number of nondifferentiable points on the line segment $(\mathbf{w}^k, \mathbf{w}^k + p\, \Delta\, \mathbf{w}^k]$, where $p \in (0, 1]$, is finite. To get around them, one can calculate these stepsizes; alternatively, one may just randomly choose a point $t' \in [\gamma, 1)$. Since the set of nondifferentiable points has measure zero, the probability of $\mathbf{w}^k + t'\beta^l\alpha^{k,i}\, \Delta\, \mathbf{w}^k$ being differentiable is 1. If $\mathbf{w}^k + t'\beta^l\alpha^{k,i}\, \Delta\, \mathbf{w}^k$ is differentiable, let $t = t'$. Otherwise, keep on selecting points t' in $[\gamma, 1] \setminus T$, where $T \subset [\gamma, 1)$ is the set of already tried points, until $\mathbf{w}^k + t'\beta^l\alpha^{k,i}\, \Delta\, \mathbf{w}^k$ is differentiable.

Next, we modify the convergence analysis of gradient method with Armijo rule (see [7, p. 38, Prop. 1.2.1]), give the following results.

Proposition 2 *Suppose the sequences $\{\mathbf{w}^k\}_{k=1}^{\infty}$ and $\{\Delta\, \mathbf{w}^k\}_{k=1}^{\infty}$ are bounded. Then $\{\mathbf{w}^k\}_{k=1}^{\infty}$ has limit points. Furthermore, assume for each sub-sequence $\{\mathbf{w}^{k_i}\}_{i=1}^{\infty}$ converging to a nonstationary point of Ψ,*

$$\limsup_{i \to \infty} \nabla \Psi(\mathbf{w}^{k_i})^T\, \Delta\, \mathbf{w}^{k_i} < 0. \tag{27}$$

Then each limit point of $\{\mathbf{w}^k\}_{k=1}^{\infty}$ is a stationary point of Ψ.

Proof: Since $\{\mathbf{w}^k\}_{k=1}^{\infty}$ is bounded, it has limit points. Let $\tilde{\mathbf{w}}$ be a limit point of $\{\mathbf{w}^k\}$. If $\tilde{\mathbf{w}}$ is not a stationary point, let $\{\mathbf{w}^{k_i}\}_{i=1}^{\infty}$ be a subsequence converging to $\tilde{\mathbf{w}}$. Because Ψ is continuous, $\Psi(\mathbf{w}^{k_i})$ converges to $\Psi(\tilde{\mathbf{w}})$. From the definition of Armijo rule,

$$\Psi(\mathbf{w}^k) - \Psi(\mathbf{w}^{k+1}) \geq -\sigma\alpha^k \nabla \Psi(\mathbf{w}^k)^T\, \Delta\, \mathbf{w}^k, \text{ for } k = 1, 2, \ldots$$

Therefore,

$$\alpha^{k_i} \nabla \Psi(\mathbf{w}^{k_i})^T\, \Delta\, \mathbf{w}^{k_i} \to 0. \tag{28}$$

From (27) and (28), we have

$$\alpha^{k_i} \to 0.$$

That means $\exists\, p > 0$, $\forall\, i \geq p$, the stepsize is reduced at least once. By the definition of perturbed Armijo rule, $\forall\, i \geq p$, for some $\varrho^{k_i} \in [\frac{1}{\beta}, \frac{1}{\gamma\beta})$:

$$\Psi(\mathbf{w}^{k_i}) - \Psi\left(\mathbf{w}^{k_i} + \alpha^{k_i}\varrho^{k_i}\, \Delta\, \mathbf{w}^{k_i}\right) < -\sigma\alpha^{k_i}\varrho^{k_i} \nabla \Psi(\mathbf{w}^{k_i})^T\, \Delta\, \mathbf{w}^{k_i}, \tag{29}$$

and $\Psi\left(\mathbf{w}^{k_i} + \alpha^{k_i}\varrho^{k_i}\, \Delta\, \mathbf{w}^{k_i}\right)$ is differentiable.

From (27), without loss of generality, we assume $\Delta\, \mathbf{w}^{k_i} \neq \mathbf{0}$. Denote

$$\mathbf{d}^{k_i} \stackrel{\text{def}}{=} \frac{\Delta\, \mathbf{w}^{k_i}}{\|\Delta\, \mathbf{w}^{k_i}\|_2}, \qquad \bar{\alpha}^{k_i} \stackrel{\text{def}}{=} \alpha^{k_i}\varrho^{k_i} \left\|\Delta\, \mathbf{w}^{k_i}\right\|_2.$$

Since $\left\|\Delta\, \mathbf{w}^{k_i}\right\|_2$ is bounded, $\bar{\alpha}^{k_i} \to 0$. Therefore, (29) can be written as: $\forall i \geq p$:

$$\frac{\Psi(\mathbf{w}^{k_i}) - \Psi\left(\mathbf{w}^{k_i} + \bar{\alpha}^{k_i}\mathbf{d}^{k_i}\right)}{\bar{\alpha}^{k_i}} < -\sigma \nabla \Psi(\mathbf{w}^{k_i})^T \mathbf{d}^{k_i}. \tag{30}$$

By chain rule, similar to the proof of Lemma 2, one can verify that $\exists\, h > 0, \rho > 0$, so that $\forall\ \|\mathbf{w}' - \tilde{\mathbf{w}}\|_2 \leq h$ and $\|\mathbf{w}'' - \tilde{\mathbf{w}}\|_2 \leq h$:

$$\|\Psi(\mathbf{w}'') - \Psi(\mathbf{w}') - \Psi'[\mathbf{w}'; (\mathbf{w}'' - \mathbf{w}')]\|_2 \leq \rho\, \|\mathbf{w}'' - \mathbf{w}'\|_2^2.$$

Since $\mathbf{w}^{k_i} \to \tilde{\mathbf{w}}$, $\bar{\alpha}^{k_i} \to 0$, $\left\|\mathbf{d}^{k_i}\right\|_2 = 1$, we have: $\exists\, q \geq p$, so that $\forall\, i > q$:

$$\left\|\mathbf{w}^{k_i} - \tilde{\mathbf{w}}\right\|_2 \leq \frac{h}{2}, \qquad \bar{\alpha}^{k_i} \leq \frac{h}{2}.$$

Hence

$$\left\|\mathbf{w}^{k_i} + \bar{\alpha}^{k_i}\mathbf{d}^{k_i} - \tilde{\mathbf{w}}\right\|_2 \leq h.$$

Therefore,

$$\frac{\Psi(\mathbf{w}^{k_i}) - \Psi\left(\mathbf{w}^{k_i} + \bar{\alpha}^{k_i}\mathbf{d}^{k_i}\right)}{\bar{\alpha}^{k_i}} \geq -\nabla\Psi(\mathbf{w}^{k_i})\mathbf{d}^{k_i} - \bar{\alpha}^{k_i}\rho\left\|\mathbf{d}^{k_i}\right\|_2^2 \tag{31}$$

Combine (30) and (31), we have

$$-\frac{\bar{\alpha}^{k_i}\rho\left\|\mathbf{d}^{k_i}\right\|_2^2\left\|\Delta\,\mathbf{w}^{k_i}\right\|_2}{1-\sigma} < \nabla\Psi(\mathbf{w}^{k_i})^T\,\Delta\,\mathbf{w}^{k_i}. \tag{32}$$

Since $\left\|\mathbf{d}^{k_i}\right\|_2 = 1$, $\Delta\,\mathbf{w}^{k_i}$ is bounded, and $\bar{\alpha}^{k_i} \to 0$, we have the left hand side of (32) converges to zeor, which contradicts to (27). ■

Corollary 2 *If the Jacobian of G is nonsingular for each k, let*

$$\Delta\,\mathbf{w}^k = -\nabla G^{-1}(\mathbf{w}^k)G(\mathbf{w}^k).$$

Otherwise, let

$$\Delta\,\mathbf{w}^k = -\left(\nabla G(\mathbf{w}^k)^T\nabla G(\mathbf{w}^k) + c^k I\right)^{-1}\nabla G(\mathbf{w}^k)^T G(\mathbf{w}^k).$$

Assume $\{\mathbf{w}^k\}_{i=1}^{\infty}$ and $\{\Delta\,\mathbf{w}^k\}_{i=1}^{\infty}$ are bounded. Suppose $c^k \to 0$. Then the sequence $\{\mathbf{w}^k\}_{i=1}^{\infty}$ generated by the gradient method with perturbed Armijo rule has limit points which are stationary points of Ψ.

Remark 3 *The perturbation scheme can be extended to some other Armijo-type line search techniques, such as the nonmonotone line search rule ([23]).*

5.2 Reformulation

In some algorithms for directional differentiable nonsmooth equations, such as [37, 40], the *kth* search direction $\Delta\,\mathbf{w}^k$ is a solution to the equation:

$$G(\mathbf{w}^k) + G'(\mathbf{w}^k; \Delta\,\mathbf{w}^k) = 0. \tag{33}$$

The Armijo-type stepsize rule is then changed to:
For given constants $s > 0$, $\sigma > 0$, $\beta \in (0,1)$, find the largest nonnegative integer l for which

$$\Psi(\mathbf{w}^k) - \Psi\left(\mathbf{w}^k + \beta^l s\,\Delta\,\mathbf{w}^k\right) \geq -\sigma\beta^l s\Psi'(\mathbf{w}^k; \Delta\,\mathbf{w}^k).$$

The equation (33) is not easy to solve in general. In this part, we will show how to solve it for (3). Since the set of nondifferentiable points of all our systems has measure zero, the probability of hitting these points is zero; consequently, the probability of applying the reformulations below during the iteration is zero. Therefore, the expected total computation doesn't include that below. For simplicity, we drop off the iteration number k in this part.

5.2.1 Min Function

We reformulate (33) with respect to (15) in this part.

Divide the index set $L = \{1, 2, \ldots, n\}$ into the following five subsets.

$$L_0 \overset{\text{def}}{=} \{i\colon \min(\lambda_i, \omega_i, \mathbf{z}_i^T R\mathbf{z}_i) \text{ is differentiable}\},$$
$$L_{\lambda\omega} \overset{\text{def}}{=} \{i\colon \lambda_i = \omega_i < \mathbf{z}_i^T R\mathbf{z}_i\}, \qquad L_{\lambda\mathbf{z}} \overset{\text{def}}{=} \{i\colon \lambda_i = \mathbf{z}_i^T R\mathbf{z}_i < \omega_i\},$$
$$L_{\omega\mathbf{z}} \overset{\text{def}}{=} \{i\colon \omega_i = \mathbf{z}_i^T R\mathbf{z}_i < \lambda_i\}, \qquad L_{\lambda\omega\mathbf{z}} \overset{\text{def}}{=} \{i\colon \lambda_i = \omega_i = \mathbf{z}_i^T R\mathbf{z}_i\}.$$

It is easy to see that the directional derivatives are

$$D\left[\min(\lambda_i, \omega_i, \mathbf{z}_i^T R\mathbf{z}_i); \begin{pmatrix}\Delta\lambda_i\\ \Delta\omega_i\\ \Delta\mathbf{z}_i\end{pmatrix}\right] = \begin{cases}\min(\Delta\lambda_i, \Delta\omega_i) & i \in L_{\lambda\omega},\\ \min\left(\Delta\lambda_i, 2\mathbf{z}_i^T R\,\Delta\mathbf{z}_i\right) & i \in L_{\lambda\mathbf{z}},\\ \min\left(\Delta\omega_i, 2\mathbf{z}_i^T R\,\Delta\mathbf{z}_i\right) & i \in L_{\omega\mathbf{z}},\\ \min\left(\Delta\lambda_i, \Delta\omega_i, 2\mathbf{z}_i^T R\,\Delta\mathbf{z}_i\right) & i \in L_{\lambda\omega\mathbf{z}}.\end{cases}$$

Let p_j, q_j, l_j be determined by (16). We transform (33) into:

$$\begin{array}{ll}
R\mathbf{z}_j\,\Delta\omega_j + \omega_j R\,\Delta\mathbf{z}_j - A_j^T\,\Delta\mathbf{y} = A_j^T\mathbf{y} + \mathbf{c}_j - \omega_j R\mathbf{z}_j & (j \in L)\\
\sum_{i\in L}(\Delta\lambda_i A_i\mathbf{z}_i + \lambda_i A_i\,\Delta\mathbf{z}_i) = \mathbf{b} - \sum_{i\in L}\lambda_i A_i\mathbf{z}_i & \\
\Delta(\mathbf{z}_j)_0 = 1 - (\mathbf{z}_j)_0 & (j \in L)\\
p_j\,\Delta\omega_i + q_j\,\Delta\lambda_j + 2l_j\mathbf{z}_j^T R\,\Delta\mathbf{z}_j = -\min\left(\lambda_j, \omega_j, \mathbf{z}_j^T R\mathbf{z}_j\right) & (j \in L_0)\\
\begin{cases}u_j = \Delta\lambda_j + \min(\lambda_j, \omega_j, \mathbf{z}_j^T R\mathbf{z}_j)\\ v_j = \Delta\omega_j + \min(\lambda_j, \omega_j, \mathbf{z}_j^T R\mathbf{z}_j)\\ u_j \ge 0, \quad v_j \ge 0, \quad u_j v_j = 0\end{cases} & (j \in L_{\lambda\omega})\\
\begin{cases}u_j = \Delta\lambda_j + \min(\lambda_j, \omega_j, \mathbf{z}_j^T R\mathbf{z}_j)\\ w_j = 2\mathbf{z}_j^T R\,\Delta\mathbf{z}_j + \min(\lambda_j, \omega_j, \mathbf{z}_j^T R\mathbf{z}_j)\\ u_j \ge 0, \quad w_j \ge 0, \quad u_j w_j = 0\end{cases} & (j \in L_{\lambda\mathbf{z}})\\
\begin{cases}v_j = \Delta\omega_j + \min(\lambda_j, \omega_j, \mathbf{z}_j^T R\mathbf{z}_j)\\ w_j = 2\mathbf{z}_j^T R\,\Delta\mathbf{z}_j + \min(\lambda_j, \omega_j, \mathbf{z}_j^T R\mathbf{z}_j)\\ v_j \ge 0, \quad w_j \ge 0, \quad v_j w_j = 0\end{cases} & (j \in L_{\omega\mathbf{z}})\\
\begin{cases}u_j = \Delta\lambda_j + \min(\lambda_j, \omega_j, \mathbf{z}_j^T R\mathbf{z}_j)\\ v_j = \Delta\omega_j + \min(\lambda_j, \omega_j, \mathbf{z}_j^T R\mathbf{z}_j)\\ w_j = 2\mathbf{z}_j^T R\,\Delta\mathbf{z}_j + \min(\lambda_j, \omega_j, \mathbf{z}_j^T R\mathbf{z}_j)\\ u_j \ge 0, \quad v_j \ge 0, \quad w_j \ge 0, \quad u_j v_j w_j = 0\end{cases} & (j \in L_{\lambda\omega\mathbf{z}})
\end{array}$$

5.2.2 $[\cdot]_+$ and $[\cdot]_-$ Functions

Define the index sets:

$$L_\lambda \overset{\text{def}}{=} \{i \in L\colon \lambda_i = 0\}, \qquad L_\omega \overset{\text{def}}{=} \{i \in L\colon \omega_i = 0\}.$$

Note that

$$D([0]_+; \Delta a) = [\Delta a]_+, \qquad D([0]_-; \Delta a) = [\Delta a]_-.$$

Let

$$u_j^s = [\Delta s]_+, \qquad v_j^s = -[\Delta s]_- \qquad (j \in L_s,\ s \in \{\lambda, \omega\}).$$

Then

$$\Delta\lambda_j = u_j^\lambda - v_j^\lambda\ (j \in L_\lambda), \qquad \Delta\omega_j = u_j^\omega - v_j^\omega\ (j \in L_\omega).$$

And

$$u_j^\omega v_j^\omega = 0 \qquad u_j^\omega \geq 0 \qquad v_j^\omega \geq 0 \qquad (j \in L_\omega),$$
$$u_j^\lambda v_j^\lambda = 0 \qquad u_j^\lambda \geq 0 \qquad v_j^\lambda \geq 0 \qquad (j \in L_\lambda).$$

Let p_j and q_j $(j \in L)$ be as defined in (21). Then (33) with respect to (20) can be reformulated as the following.

$$u_j^\omega R\mathbf{z}_j - A_j^T \Delta \mathbf{y} = \mathbf{c}_j + A_j^T \mathbf{y} \qquad (j \in L_\omega)$$
$$\Delta \omega_j p_j R\mathbf{z}_j + [\omega_j]_+ R \Delta \mathbf{z}_j - A_j^T \Delta \mathbf{y} = \mathbf{c}_j + A_j^T \mathbf{y} - [\omega_j]_+ R\mathbf{z}_j \ (j \in L \setminus L_\omega)$$
$$\sum_{i \in L_\lambda} u_i^\lambda A_i \mathbf{z}_i + \sum_{i \in L \setminus L_\lambda} (\Delta \lambda_i q_i A_i \mathbf{z}_i + [\lambda_i]_+ A_i \Delta \mathbf{z}_i) = \mathbf{b} - \sum_{i \in L} [\lambda_i]_+ A_i \mathbf{z}_i$$
$$\Delta (\mathbf{z}_j)_0 = 1 - (\mathbf{z}_j)_0 \qquad (j \in L)$$
$$(1 - p_j) \Delta \omega_j + (1 - q_j) \Delta \lambda_j + 2\mathbf{z}_j^T R \Delta \mathbf{z}_j = -\mathbf{z}_j^T R\mathbf{z}_j - [\omega_j]_- - [\lambda_j]_-$$
$$(j \in L \setminus (L_\lambda \cup L_\omega))$$
$$(1 - p_j) \Delta \omega_j - v_j^\lambda + 2\mathbf{z}_j^T R \Delta \mathbf{z}_j = -\mathbf{z}_j^T R\mathbf{z}_j - [\omega_j]_- - [\lambda_j]_- \ (j \in L_\lambda \setminus L_\omega)$$
$$-v_j^\omega + (1 - q_j) \Delta \lambda_j + 2\mathbf{z}_j^T R \Delta \mathbf{z}_j = -\mathbf{z}_j^T R\mathbf{z}_j - [\omega_j]_- - [\lambda_j]_- \ (j \in L_\omega \setminus L_\lambda)$$
$$-v_j^\omega - v_j^\lambda + 2\mathbf{z}_j^T R \Delta \mathbf{z}_j = -\mathbf{z}_j^T R\mathbf{z}_j - [\omega_j]_- - [\lambda_j]_- \qquad (j \in L_\omega \cap L_\lambda)$$
$$u_j^\omega v_j^\omega = 0 \quad u_j^\omega \geq 0 \quad v_j^\omega \geq 0 \qquad (j \in L_\omega)$$
$$u_j^\lambda v_j^\lambda = 0 \quad u_j^\lambda \geq 0 \quad v_j^\lambda \geq 0 \qquad (j \in L_\lambda)$$

5.2.3 General Complementarity Functions and Jordan Algebra Reformulations

When $\lambda_j = 0$, $D[|\lambda_j|; \Delta \lambda_j] = |\Delta \lambda_j|$. If $\Delta \lambda_j$ only appears in (3b) of (33), we assume $\Delta \lambda_j \geq 0$ when $\lambda_j = 0$, and the result will not be affected. Therefore, we replace $D[|\lambda_j|; \Delta \lambda_j]$ by $\Delta \lambda_j$, and add $\Delta \lambda_j \geq 0$.

Complementarity functions composed by absolute value, min, $[\cdot]_+$ and $[\cdot]_-$ in (33) at nondifferentiable points can be treated like (15), (20). For Fischer-Burmeister function,

$$D[\phi(0,0); (\Delta a, \Delta b)] = \sqrt{\Delta a^2 + \Delta b^2} - \Delta a - \Delta b.$$

6 Numerical Examples

We have implemented (3) with perturbed Armijo line search, and the perturbed nonmonotone line search in MATLAB. Our preliminary computational results show that generally the numbers of function evaluation and iteration are fewer for nonmonotone line search if the initial point is far from the optimum, but the numbers are almost the same for the two line search strategies when starting close to the optimum. We present some test results with perturbed nonmonotone line search below.

Given positive numbers ϵ, *steptol*, *itlimit*, *conlimit*, the algorithm terminates when it finds an ϵ-solution of G, or when the stepsize is less than *steptol*, or when the iteration number exceeds *itlimit*.

Do while $\|G\|_\infty \geq \epsilon$, $\left\|\mathbf{w}^{k+1} - \mathbf{w}^k\right\|_\infty \geq$ *steptol*, and $k \leq$ *itlimit*.

1. Calculate $\nabla G(\mathbf{w}^k)$, and estimate it's condition number.

2. If the estimated condition number of $G(\mathbf{w}^k)$ is smaller than *conlimit*,

$$\Delta\mathbf{w} = -\nabla G(\mathbf{w}^k)^{-1} G(\mathbf{w}^k);$$

otherwise,

$$\Delta\mathbf{w} = -\left(\nabla G(\mathbf{w}^k)^T \nabla G(\mathbf{w}^k) + c^k I\right)^{-1} \nabla G(\mathbf{w}^k)^T G(\mathbf{w}^k).$$

3. Do line search to determine the step size α.
4. $\mathbf{w}^{k+1} = \mathbf{w}^k + \alpha\Delta\mathbf{w}$; $k+1 \to k$.

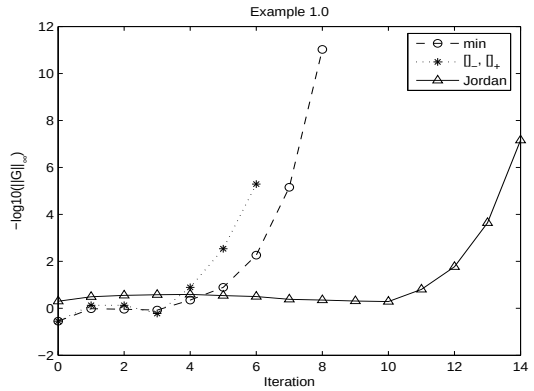
We use the suggested values in [10] to terminate our program. On the PC running the program, the machine accuracy is $\tau = 2.220446e-016$; so we set $\epsilon = \tau^{1/3} = 6.055454e-006$, $steptol = \tau^{2/3} = 3.666853e-011$, $itlimit = 100$.

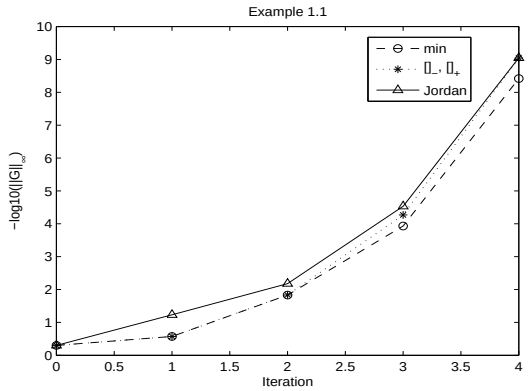
6.1 Example 1

Our first computational example is the SMT problem from [50, example 1]. Our initial points were calculated from the approximated values below. We've used the min, $[]_-$ and $[]_+$, and Jordan algebra reformulations.

Initial coordinates of the 8 Steiner points in example 1.0

index	x-coordinate	y-coordinate	index	x-coordinate	y-coordinate
1	0.6	6.5	5	7.2	1.8
2	0.8	3.5	6	5.2	2.1
3	1.7	1.2	7	2.5	7.5
4	4.1	0.8	8	3.9	7.0





The output is summarized in figure 'Example 1.0'. x-axis is the iteration numbers, y-axis is $-\log_{10}$ of $\|G\|_\infty$. It shows local Q-quadratic convergence. Our initial network-cost is larger than that of iteration 7 in [50], but each of our network-cost at the $7th$ iteration of (15), the $5th$ of (20), and the $13th$ of (23) is better than that of the $23th$–the last iteration in [50].

To test reoptimization, we perturbed each coordinate of the regular points by a scalar in $(-0.5, 0.5)$. The resulting coordiates are in the table below. Starting from the solutions to example 1.0, the Newton's method for each of our formula found an optimum in 4 iterations; see figure 'Example 1.1'.

The coordinates of the 10 regular points in example 1.1

index	x	y	index	x	y
9	2.06225265	9.06259293	14	7.55387796	0.97892289
10	0.82034497	6.63177002	15	8.92332597	3.05143468
11	1.24810704	3.85186112	16	5.04443039	3.90964814
12	1.65588987	1.36153760	17	3.42613689	6.64003516
13	3.66904285	0.86330140	18	7.43136476	7.22161716

In the next example, we set point 9 to $(2.5, 9.0)$. We get the solution in 2 iterations for each of the formula from the old solution.

6.2 Example 2

Our example 2 are randomly generated second-order cone programs. The results are summarized in the table below.

Output of example 2

tp	syst	succ	initial gap	final gap	iteration	fun eval
1	(15)	100	0.708798	-14.091132	1.060000	2.060000
	(20)	100	0.708798	-14.023286	1.000000	2.000000
	(23)	100	0.708798	-14.114863	1.000000	2.000000
2	(15)	100	-0.468689	-7.392421	3.600000	4.630000
	(20)	100	-0.468689	-8.334704	4.070000	5.150000
	(23)	100	-0.468689	-7.735485	6.750000	16.160000
3	(15)	97	-0.719331	-7.772418	8.412371	20.917526
	(20)	98	-0.719363	-7.578901	10.316327	33.663265
	(23)	98	-0.719363	-7.638747	15.438776	77.887755
4	(15)	100	-1.675727	-7.835776	3.660000	6.300000
	(20)	100	-1.676413	-7.879695	4.110000	8.280000
	(23)	100	-1.676288	-7.314917	7.770000	26.850000
5	(15)	100	-0.443995	-7.691475	6.110000	10.610000
	(20)	100	-0.443996	-7.316729	7.160000	17.410000
	(23)	100	-0.443996	-7.401255	11.040000	45.700000
6	(15)	99	0.040911	-7.873853	4.494949	9.191919
	(20)	100	0.045377	-7.943580	4.530000	6.230000
	(23)	100	0.045377	-7.659177	6.700000	15.610000
7	(15)	56	-0.647448	-7.792780	8.964286	22.303571
	(20)	56	-0.647448	-7.858531	10.535714	33.517857
	(23)	56	-0.647448	-7.774833	15.267857	76.232143
8	(15)	22	0.272251	-7.448693	17.272727	82.863636
	(20)	8	0.341327	-13.353607	40.750000	853.000000
	(23)	28	0.299279	-7.369137	31.750000	449.678571
9	(15)	99	0.255684	-7.401473	7.454545	20.313131
	(20)	99	0.253370	-7.834493	7.272727	12.424242
	(23)	99	0.254746	-7.694833	9.616162	25.111111

We first generated the solutions, each of which has 10 10-dimensional blocks. Type of primal solution is [b, o, i, b, b, i, o, o, b, b]; type of dual solution is [b, i, o, b, b, o, i, b, b, b]; where 'b' means the block is in bd $\mathcal{Q}$, 'o' means the block is zero, 'i' means the block is in int $\mathcal{Q}$. Number of linear constraints $m = 33$. According to [10], the line search Newton direction is independent of the scaling of the dependent or the independent variables. So we set each element of A and y in our randomly generated problem in $(-1, 1)$, the first element of each nonzero block of primal and dual solution 1, the remaining elements random numbers. We randomly generated 100 instances of the problem, and solved them by (15), (20), (23). Because strict complementarity is not satisfied at the $8th$ block, iterates of interior point method will not converge to this solution. For starting point $\mathbf{y} = \mathbf{0}$, $\mathbf{x}$ and $\mathbf{s}$ the optimal solution, the output is summarized at block row 'type 1'. All the instances converge to the optimal solutions. The block rows 'type 2-10' show the summaries of various perturbed problems as below. The starting point is the solution to the old problem. Since some blocks are in the boundary of the second-order cone, interior point method can not start from it.

Type 2: each element of $\Delta\,\mathbf{b}$ is in $\left(-\frac{\|\mathbf{b}\|_2}{m}, \frac{\|\mathbf{b}\|_2}{m}\right)$.
Type 3: each element of $\Delta\,\mathbf{c}$ is in $\left(-\frac{\|\mathbf{c}\|_2}{N}, \frac{\|\mathbf{c}\|_2}{N}\right)$, where $N = \sum_{i=1}^{n} N_i$.
Type 4: each element of $\Delta\,A$ is in $\left(-\frac{\|A\|_{\mathrm{F}}}{mN}, \frac{\|A\|_{\mathrm{F}}}{mN}\right)$.
Type 5: Each element of $\Delta\,A$ is in $0.8\left(-\frac{\|A\|_{\mathrm{F}}}{mN}, \frac{\|A\|_{\mathrm{F}}}{mN}\right)$, each element of $\Delta\,\mathbf{b}$ is in $\left(-\frac{\|\mathbf{b}\|_2}{m}, \frac{\|\mathbf{b}\|_2}{m}\right)$, each element of $\Delta\,\mathbf{c}$ is in $0.5\left(-\frac{\|\mathbf{c}\|_2}{N}, \frac{\|\mathbf{c}\|_2}{N}\right)$.
Type 6: we add a constraint.
Type 7: we delete the last constraint.
Type 8: a 3-dimensional block is added.
Type 9: we delete the last block.

The second column shows the output in that row is obtained from which reformulation of the systems of equations. The column 'succ' shows the number out of the 100 instances whose $\|G\|_\infty$ was reduced under $\tau^{1/3}$. The $4th$ and $5th$ columns list the average value of initial and final $\log_{10}\|G\|_\infty$ of a solved instance. The column 'iteration' shows the average number of iterations per solved instance. The column 'fun eval' shows the average number of function evaluations of a solved instance.

Confirms to the analysis in § 5, of all the line searches for all the instances of the problems, perturbation is only used six times.

7 Properties

We summarize the properties of the algorithm below.

The total number of variables and equations of (3) is about half of that of other systems, which means that our method requires less memory. Without regularization, the Jacobian of each iteration has the same sparsity pattern; so one can use some techniques, such as that in [22], to solve large scale sparse problems. Besides, due to the special structure of nonlinear complementarity reformulation, only a reduced system of (3) needs to be solved (see [9]). To further lessen the work of each iteration but keep the desired convergence rate, one may use some modified Newton's methods, like periodical Newton, quasi-Newton.

The primal-dual variables of each iterate of an interior point method must be in the interior of the second-order cone; while those of sequence (12) can be neither in the second-order cone, nor feasible. Therefore, the optimal solution to an old problem can be used as an initial point for a new one via Newton-type methods. This means a Newton-type method gives a better starting point for the new problem than an IPM does, since optimum is usually achieved in the boundary. Furthermore, one can use the decomposition of the Jacobian for the old solution, if it is available, to solve the linear system with Newton-type methods.

Most asymptotic convergence results about IPMs are the Q-suplinear or Q-quadratic convergence rate of duality gap, which don't imply the same fast convergence rate of iteration sequences. The log-barrier IPM employs some barrier functions $p(\mathbf{w};\mu)$, where μ is the barrier parameter, to transform constrained optimization problems into unconstrained ones. As $\mu \to 0$, the minimizer of $p(\mathbf{w};\mu)$, denoted as $\mu(\mathbf{w})$, goes to a minimizer of the original constrained optimization problem $\mathbf{w}^*$. By the results in [49], for each fixed μ, the iterates of the pure Newton's method on $p(\mathbf{w};\mu)$ converge Q-quadratically to $\mu(\mathbf{w})$ within an $O(\mu^\sigma)$ neighborhood, for some $\sigma \in (1,2]$, of $\mu(\mathbf{w})$, but not to $\mathbf{w}^*$. And the size of Euclidean ball within which the quadratic convergence rate can be achieved decreases to zero as $\mu \to 0$. In [39], it is shown that the iterates of a certain class of primal-dual IPMs with at least two backsolves per iteration converge Q-superlinearly. In summary, the local convergence rate of the iterates of an IPM to $\mathbf{w}^*$ can not be faster than that of a Newton-type method due to the parameter μ.

For local convergence analysis, strict complementarity is not needed for (3). Under primal-dual nondegeneracy and some other conditions, its Jacobian is regular at optimum; hence the solution is

numerically stable and accurate.

Because Newton-type methods don't employ penalty or barrier functions, they have the advantages of penalty free algorithms described in [16].

Newton-type methods are better for perturbed problem, but starting from a point far from optimum, their iteration sequences may be trapped at a local minimum or saddle point of the merit function; while IPMs are not good for re-optimization, in practice, they can converge to an optimum even far from it. So to "cold start" a problem, we suggest using a hybrid algorithm: start with an IPM until the primal, dual infeasibility and duality gap are small, then switch to a Newton-type method. Another approach to "cold start" a problem may be to use some parameterized complementarity functions such as ϕ_μ, the Chen-Mangasarian family of complementarity functions.

We have extended the above results about the normal cone, perturbation analysis, etc. to semidefinite programming, symmetric cone programming, and P-cone programming. Preliminary numerical results show that as for SOCP, Newton-type methods are good for "warm starting" these models.

Acknowledgements

I would like to express my gratitude to my Ph.D. advisor Professor Farid Alizadeh. This work is supported in part through his grants from the U.S. National Science Foundation. Discussions with him motivated the paper and encouraged me to improve its quality. I also want to thank Professor Paul Tseng. Especially, he gave a short proof of the Lagrangian multipliers. His comments and suggestions about the ambiguousness of my previous line search led to § 5, and helped improved the appearance of the paper. He also brought up reference [20] to my attention. As well, I want to thank Professor Jonathan Eckstein for pointing out the ambiguousness of my perturbed line search scheme in the previous draft.

References

[1] I. Adler and F. Alizadeh. Primal-dual interior point algorithms for convex quadratically constrained and semidefinite optimization problems. Technical Report RRR 46-95, RUTCOR, Rutgers University, 1995.

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

[3] F. Alizadeh and S. H. Schmieta. Optimization with semidefinite, quadratic and linear constraints. Technical Report RRR 23-97, RUTCOR, Rutgers Univeristy, 1997.

[4] Farid Alizadeh, Jean-Pierre A. Haeberly, and Michael L. Overton. Primal-dual interior-point methods for semidefinite programming: convergence rates, stability and numerical results. *SIAM J. Optim.*, 8(3):746–768 (electronic), 1998.

[5] Farid Alizadeh and Yu Xia. The Q method for second-order cone programming. Technical Report AdvOl-Report No. 2004/15, McMaster Univeristy, 2004.

[6] Hande Y. Benson and Robert J. Vanderbei. Solving problems with semidefinite and related constraints using interior-point methods for nonlinear programming. *Math. Program.*, 95(2, Ser. B):279–302, 2003.

[7] Dimitri P. Bertsekas. *Nonlinear programming: 2nd Edition.* Athena Scientific, 1999.

[8] Frank H. Clarke. *Optimization and nonsmooth analysis.* John Wiley & Sons Inc., New York, 1983. A Wiley-Interscience Publication.

[9] Tecla De Luca, Francisco Facchinei, and Christian Kanzow. A theoretical and numerical comparison of some semismooth algorithms for complementarity problems. *Comput. Optim. Appl.*, 16(2):173–205, 2000.

[10] John E. Dennis, Jr. and Robert B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*. Prentice-Hall Inc., Englewood Cliffs, N.J., 1983.

[11] Yu. G. Evtushenko and V. A. Purtov. Sufficient conditions for a minimum for nonlinear programming problems. *Dokl. Akad. Nauk SSSR*, 278(1):24–27, 1984.

[12] Jacques Faraut and Adam Korányi. *Analysis on symmetric cones*. The Clarendon Press Oxford University Press, New York, 1994. Oxford Science Publications.

[13] Leonid Faybusovich. Euclidean Jordan algebras and interior-point algorithms. *Positivity*, 1(4):331–357, 1997.

[14] M. C. Ferris and J. S. Pang. Engineering and economic applications of complementarity problems. *SIAM Rev.*, 39(4):669–713, 1997.

[15] A. Fischer. A special Newton-type optimization method. *Optimization*, 24(3-4):269–284, 1992.

[16] Roger Fletcher and Sven Leyffer. Nonlinear programming without a penalty function. *Math. Programming*, 91(2, Ser. A):239–269, 2002.

[17] Robert M. Freund. A potential-function reduction algorithm for solving a linear program directly from an infeasible "warm start". *Math. Programming*, 52(3, Ser. B):441–466 (1992), 1991. Interior point methods for linear programming: theory and practice (Scheveningen, 1990).

[18] M. Fukushima, Luo Z.-Q., and P. Tseng. A sequential quadratically constrained quadratic programming method for differentiable convex minimization. *SIAM J. Optim.*, 13(4):1098–1119 (electronic), 2003.

[19] Masao Fukushima, Zhi-Quan Luo, and Paul Tseng. Smoothing functions for second-order-cone complementarity problems. *SIAM J. Optim.*, 12(2):436–460 (electronic), 2001.

[20] Carl Geiger and Christian Kanzow. On the resolution of monotone complementarity problems. *Comput. Optim. Appl.*, 5(2):155–173, 1996.

[21] Jacek Gondzio. Warm start of the primal-dual method applied in the cutting-plane scheme. *Math. Programming*, 83(1, Ser. A):125–143, 1998.

[22] Jacek Gondzio and Robert Sarkissian. Parallel interior-point solver for structured linear programs. *Math. Programming*, 96(3, Ser. A):561–584, 2003.

[23] L. Grippo, F. Lampariello, and S. Lucidi. A nonmonotone line search technique for Newton's method. *SIAM J. Numer. Anal.*, 23(4):707–716, 1986.

[24] Christian Kanzow. Global convergence properties of some iterative methods for linear complementarity problems. *SIAM J. Optim.*, 6(2):326–341, 1996.

[25] Sunyoung Kim and Masakazu Kojima. Second order cone programming relaxation of nonconvex quadratic optimization problems. *Optim. Methods Softw.*, 15(3-4):201–224, 2001.

[26] Sunyoung Kim and Masakazu Kojima. Exact solutions of some nonconvex quadratic optimization problems via SDP and SOCP relaxations. *Comput. Optim. Appl.*, 26(2):143–154, 2003.

[27] Miguel Sousa Lobo, Lieven Vandenberghe, Stephen Boyd, and Hervé Lebret. Applications of second-order cone programming. *Linear Algebra Appl.*, 284(1-3):193–228, 1998. ILAS Symposium on Fast Algorithms for Control, Signals and Image Processing (Winnipeg, MB, 1997).

[28] Zhi-Quan Luo. Applications of convex optimization in signal processing and digital communication. *Math. Program.*, 97(1-2, Ser. B):177–207, 2003.

[29] Irvin J. Lustig, Roy E. Marsten, and David F. Shanno. Computational experience with a globally convergent primal-dual predictor-corrector algorithm for linear programming. *Math. Programming*, 66(1, Ser. A):123–135, 1994.

[30] O. L. Mangasarian. Equivalence of the complementarity problem to a system of nonlinear equations. *SIAM J. Appl. Math.*, 31(1):89–92, 1976.

[31] J.S. Marron and M. Todd. Distance weighted discrimination. Technical Report 1339, ORIE, Cornell University, July 2002.

[32] Robert Mifflin. Semismooth and semiconvex functions in constrained optimization. *SIAM J. Control Optimization*, 15(6):959–972, 1977.

[33] John E. Mitchell and Michael J. Todd. Solving combinatorial optimization problems using Karmarkar's algorithm. *Math. Programming*, 56(3, Ser. A):245–284, 1992.

[34] Renato D. C. Monteiro and Takashi Tsuchiya. Polynomial convergence of primal-dual algorithms for the second-order cone program based on the MZ-family of directions. *Math. Program.*, 88(1, Ser. A):61–83, 2000.

[35] Yu. E. Nesterov and M. J. Todd. Primal-dual interior-point methods for self-scaled cones. *SIAM J. Optim.*, 8(2):324–364 (electronic), 1998.

[36] Yurii Nesterov and Arkadii Nemirovskii. *Interior-point polynomial algorithms in convex programming*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994.

[37] Jong-Shi Pang. Newton's method for B-differentiable equations. *Math. Oper. Res.*, 15(2):311–341, 1990.

[38] Jiming Peng, Cornelis Roos, and Tamás Terlaky. Primal-dual interior-point methods for second-order conic optimization based on self-regular proximities. *SIAM J. Optim.*, 13(1):179–203 (electronic), 2002.

[39] Florian A. Potra. Q-superlinear convergence of the iterates in primal-dual interior-point methods. *Math. Program.*, 91(1, Ser. A):99–115, 2001.

[40] Li Qun Qi. Convergence analysis of some algorithms for solving nonsmooth equations. *Math. Oper. Res.*, 18(1):227–244, 1993.

[41] Li Qun Qi and Jie Sun. A nonsmooth version of Newton's method. *Math. Programming*, 58(3, Ser. A):353–367, 1993.

[42] Liqun Qi and Houyuan Jiang. Semismooth Karush-Kuhn-Tucker equations and convergence analysis of Newton and quasi-Newton methods for solving these equations. *Math. Oper. Res.*, 22(2):301–325, 1997.

[43] R. Tyrrell Rockafellar. *Convex analysis*. Princeton University Press, Princeton, N.J., 1970. Princeton Mathematical Series, No. 28.

[44] Takashi Sasakawa and Takashi Tsuchiya. Optimal magnetic shield design with second-order cone programming. *SIAM J. Sci. Comput.*, 24(6):1930–1950 (electronic), 2003.

[45] S. H. Schmieta and F. Alizadeh. Associative and Jordan algebras, and polynomial time interior-point algorithms for symmetric cones. *Math. Oper. Res.*, 26(3):543–564, 2001.

[46] Michael J. Todd. On Anstreicher's combined phase i–phase ii projective algorithm for linear programming. *Math. Programming*, 55(1, Ser. A):1–15, 1992.

[47] P. Tseng. Growth behavior of a class of merit functions for the nonlinear complementarity problem. *J. Optim. Theory Appl.*, 89(1):17–37, 1996.

[48] Takashi Tsuchiya. A convergence analysis of the scaling-invariant primal-dual path-following algorithms for second-order cone programming. *Optim. Methods Softw.*, 11/12(1-4):141–182, 1999. Interior point methods.

[49] Stephen J. Wright. On the convergence of the Newton/log-barrier method. *Math. Program.*, 90(1, Ser. A):71–100, 2001.

[50] Guoliang Xue and Yinyu Ye. An efficient algorithm for minimizing a sum of Euclidean norms with applications. *SIAM J. Optim.*, 7(4):1017–1036, 1997.

[51] E. Alper Yildirim and Stephen J. Wright. Warm-start strategies in interior-point methods for linear programming". *SIAM J. Optim.*, 12(3):782–810, 2002.