Quadratic convergence of Newton's method to the optimal solution of second-order cone optimization

Ali Mohammad-Nezhad Tamás Terlaky

Department of Industrial and Systems Engineering
Lehigh University

October 23, 2017

Outline

Second-order cone optimization

Convergence under strict complementarity

Failure of strict complementarity

Outline

Standard form

We aim to solve

$$(P)\quad \min\{c^T x \;:\; Ax = b,\; x \in \mathcal{L}_+^{\bar{n}}\},$$
$$(D)\quad \max\{b^T y \;:\; A^T y + s = c,\; s \in \mathcal{L}_+^{\bar{n}}\},$$

where

- $\mathcal{L}_+^{\bar{n}} := \mathbb{L}_+^{n_1} \times \ldots \times \mathbb{L}_+^{n_p}$,

$$\mathbb{L}_+^{n_i} := \{x^i := (x_1^i, \ldots, x_{n_i}^i)^T \in \mathbb{R}^{n_i} : x_1^i \geq \|x_{2:n_i}^i\|\}, \quad i = 1, \ldots, p,$$

- $A \in \mathbb{R}^{m \times \bar{n}}$, $c \in \mathbb{R}^{\bar{n}}$, $b \in \mathbb{R}^m$,
- $A := (A_1, \ldots, A_p)$, $x := (x^1; \ldots; x^p)$, $s := (s^1; \ldots; s^p)$, and $c := (c^1; \ldots; c^p)$,
- $\bar{n} := \sum_{i=1}^p n_i$.

Regularity conditions

Assumption

A is assumed to be a full row rank matrix.

Assumption (Interior point condition)

There exists a primal-dual feasible $(x; y; s)$ so that $x, s \in \text{int}(\mathcal{L}_+^{\bar{n}})$.

- As a result, the optimal set is written as the set of solutions of

$$\begin{aligned} Ax = b, &\quad x \in \mathcal{L}_+^{\bar{n}}, \\ A^T y + s = c, &\quad s \in \mathcal{L}_+^{\bar{n}}, \\ x \circ s = 0, & \end{aligned}$$

where $x \circ s = (x^1 \circ s^1; \ldots; x^p \circ s^p)$, and

$$x^i \circ s^i = \begin{pmatrix} (x^i)^T s^i \\ x_1^i s_{2:n_i}^i + s_1^i x_{2:n_i}^i \end{pmatrix}, \quad \forall\ i = 1, \ldots, p.$$

Interior point method

For $\mu > 0$, we solve a system of perturbed optimality conditions:

$$\begin{array}{lll} Ax & = b, & x \in \operatorname{int}(\mathcal{L}_+^{\bar{n}}), \\ A^T y + s & = c, & s \in \operatorname{int}(\mathcal{L}_+^{\bar{n}}), \\ x \circ s & = \mu e. & \end{array}$$

where $e^i = (1; \mathbf{0}) \in \mathbb{R}^{n_i}$, and $e = (e^1; \ldots; e^p)$.

- This system has a unique solution, the so called central solution.
- As $\mu \to 0$, the trajectory converges to a *maximally complementary solution*.

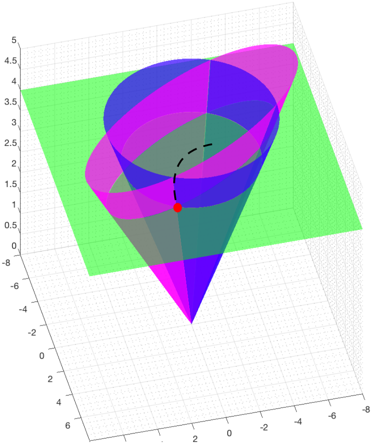
Let $\mathcal{P}^*$ and $\mathcal{D}^*$ be the sets of primal and dual optimal solutions.

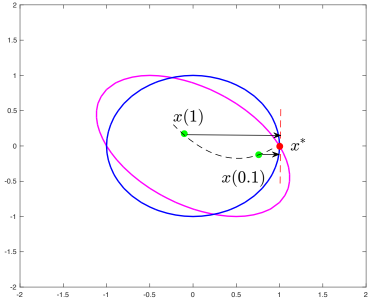
Definition

An optimal solution $(x^*; y^*; s^*) \in \mathcal{P}^* \times \mathcal{D}^*$ is maximally complementary if

$$(x^*; y^*; s^*) \in \operatorname{ri}(\mathcal{P}^* \times \mathcal{D}^*).$$

Illustration of the interior point method





Outline

Second-order cone optimization

Convergence under strict complementarity

Failure of strict complementarity

Quadratic convergence of Newton's method

The optimality conditions can be written as $F((x;y;s)) = 0$ and $x, s \in \mathcal{L}_+^{\bar{n}}$, where

$$F((x;y;s)) := \begin{pmatrix} Ax - b \\ A^T y + s - c \\ x \circ s \end{pmatrix}.$$

The Jacobian of F is given by

$$\nabla F((x;y;s)) := \begin{pmatrix} A & 0 & 0 \\ 0 & A^T & I \\ L(s) & 0 & L(x) \end{pmatrix}.$$

$L(x)$ is a block diagonal matrix:

$$L(x) := \operatorname{diag}(L(x^1), \ldots, L(x^p)),$$
$$L(x^i) := \begin{pmatrix} x_1^i & (x_{2:n_i}^i)^T \\ x_{2:n_i}^i & x_1^i I_{n_i - 1} \end{pmatrix}$$

- ∇F is Lipschitz continuous with global constant $\tau_1 = 2$.

Sufficient conditions for nonsingularity

The Jacobian ∇F is nonsingular (Alizadeh and Goldfarb) at $(x^*; y^*; s^*)$ if

- $(x^*; y^*; s^*)$ is strictly complementary,
- $(x^*; y^*; s^*)$ is primal-dual nondegenerate.

Definition (Strict complementarity)

An optimal solution $(x^*; y^*; s^*) \in \mathcal{P}^* \times \mathcal{D}^*$ is strictly complementary if

$$x^* + s^* \in \operatorname{int}(\mathcal{L}_+^{\bar{n}}).$$

Let $\tan(x^i, \mathbb{L}_+^{n_i})$ be the tangent space to $\mathbb{L}_+^{n_i}$ at x^i.

Definition (Nondegeneracy-Transversality)

A primal-feasible solution x is called nondegenerate if

$$\tan(x^1, \mathbb{L}_+^{n_1}) \times \ldots \times \tan(x^p, \mathbb{L}_+^{n_p}) + \operatorname{Ker}(A) = \mathbb{R}^{\bar{n}}.$$

A dual feasible solution $(y; s)$ is called nondegenerate if

$$\tan(s^1, \mathbb{L}_+^{n_1}) \times \ldots \times \tan(s^p, \mathbb{L}_+^{n_p}) + \mathcal{R}(A^T) = \mathbb{R}^{\bar{n}}.$$

Distance to the optimal set

Let $(\hat{x}; \hat{y}; \hat{s})$ be a primal-dual optimal solution.

The primal and dual optimal sets can be equivalently written as

$$\begin{cases} x \in \hat{x} + \mathrm{Ker}(A), \\ \hat{s}^T x = 0, \\ x \in \mathcal{L}_+^{\bar{n}}, \end{cases} \qquad \begin{cases} s \in \hat{s} + \mathcal{R}(A^T), \\ \hat{x}^T s = 0, \\ s \in \mathcal{L}_+^{\bar{n}}. \end{cases}$$

The distance between $(x(\mu); y(\mu); s(\mu))$ and the affine space in the above system:

- can be bounded by Hoffman error bound.
- θ_1 and θ_2 are Hoffman condition numbers for the primal and dual systems.

Lemma (Hölderian error bound)

Let $\big(x(\mu); y(\mu); s(\mu)\big)$ be a central solution with

$$\mu \leq \hat{\mu} := \min\left\{\frac{1}{\theta_1 p}, \frac{1}{\theta_2 p}\right\}.$$

Then there exists $(x; y; s) \in \mathcal{P}^* \times \mathcal{D}^*$, $\gamma > 0$, and $\kappa > 0$ so that

$$\|x(\mu) - x\| \leq \kappa(p\mu)^\gamma, \quad \|y(\mu) - y\| \leq \kappa(p\mu)^\gamma, \quad \|s(\mu) - s\| \leq \kappa(p\mu)^\gamma.$$

Quadratic convergence to a strictly complementary solution

Theorem

Assume that there exists $\beta_1 > 0$ *so that*

$$\|\nabla F((x^*; y^*; s^*))^{-1}\| \leq \beta_1.$$

Let a central solution $(x(\mu); y(\mu); s(\mu))$ *with*

$$\mu < \min\left\{p^{-1}\left(4\sqrt{3}\beta_1\kappa\right)^{-\frac{1}{\gamma}},\ \hat{\mu}\right\}$$

be given.

From $(x(\mu); y(\mu); s(\mu))$ *Newton's method is quadratically convergent to* $(x^*; y^*; s^*)$*.*

- $(x(\mu); y(\mu); s(\mu))$ needs to be in the convergence region of Newton's method.

Outline

Failure of strict complementarity

Without the strict complementarity condition:

- ∇F might be singular at an optimal solution,
- Newton's method is not applicable.
- Convergence to an optimal solution is not better than linear.

We can release the dependence on the strict complementarity condition:

- We need to identify the optimal partition of the problem.

Back to the complementarity condition

The complementarity condition $x^i \circ s^i = 0$ implies:

$x^i(\epsilon)$ $\Cap$	$s^i(\epsilon)$ $\Cap$	Strictly complementary
$\text{int}(\mathbb{L}_+^{n_i})$	$\{0\}$	Yes
$\{0\}$	$\text{int}(\mathbb{L}_+^{n_i})$	Yes
$\text{bd}(\mathbb{L}_+^{n_i}) \setminus \{0\}$	$\text{bd}(\mathbb{L}_+^{n_i}) \setminus \{0\}$	Yes
$\{0\}$	$\{0\}$	No
$\text{bd}(\mathbb{L}_+^{n_i}) \setminus \{0\}$	$\{0\}$	No
$\{0\}$	$\text{bd}(\mathbb{L}_+^{n_i}) \setminus \{0\}$	No

- The complementarity for linear optimization reduces to only three cases.

Optimal partition

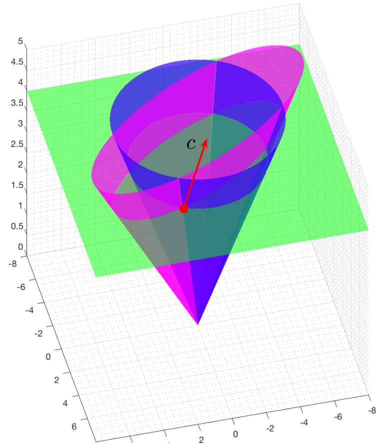
The index set $\{1, \ldots, p\}$ is partitioned into four subsets $\mathcal{B}, \mathcal{N}, \mathcal{R}$, and $\mathcal{T} := (\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3)$:

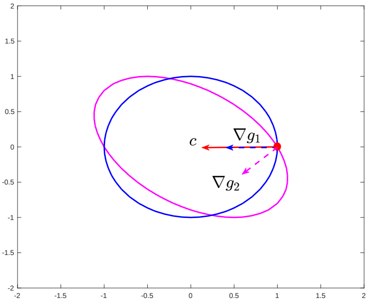
$$
\begin{aligned}
\mathcal{B} &:= \left\{i \mid x_1^i > \|x_{2:n_i}^i\|, \text{ for some } x \in \mathcal{P}^*\right\},\\
\mathcal{N} &:= \left\{i \mid s_1^i > \|s_{2:n_i}^i\|, \text{ for some } s \in \mathcal{D}^*\right\},\\
\mathcal{R} &:= \left\{i \mid x_1^i = \|x_{2:n_i}^i\| > 0, \ s_1^i = \|s_{2:n_i}^i\| > 0, \text{ for some } (x;y;s) \in \mathcal{P}^* \times \mathcal{D}^*\right\},\\
\mathcal{T}_1 &:= \left\{i \mid x^i = s^i = 0, \text{ for all } (x;y;s) \in \mathcal{P}^* \times \mathcal{D}^*\right\},\\
\mathcal{T}_2 &:= \left\{i \mid s^i = 0, \text{ for all } (y;s) \in \mathcal{D}^*, \ x_1^i = \|x_{2:n_i}^i\| > 0, \text{ for some } x \in \mathcal{P}^*\right\},\\
\mathcal{T}_3 &:= \left\{i \mid x^i = 0, \text{ for all } x \in \mathcal{P}^*, \ s_1^i = \|s_{2:n_i}^i\| > 0, \text{ for some } (y;s) \in \mathcal{D}^*\right\}.
\end{aligned}
$$

- Note that $\mathcal{B}, \mathcal{N}, \mathcal{R}$, and $\mathcal{T}$ are mutually disjoint.
- We call $(\mathcal{B}, \mathcal{N}, \mathcal{R}, \mathcal{T})$ the optimal partition.
- A solution $(x^*; y^*; s^*)$ is strictly complementary iff $\mathcal{T} = \emptyset$.

Example $(\mathcal{R}, \mathcal{T}_2 \neq \emptyset)$

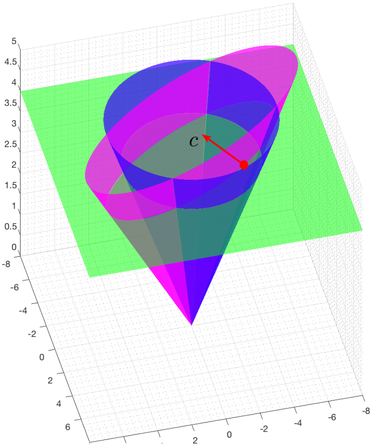
- The cone in pink is *weakly inactive*, i.e., the cone constraint is active with zero Lagrange multiplier.
- This is a *nondegenerate* optimal solution

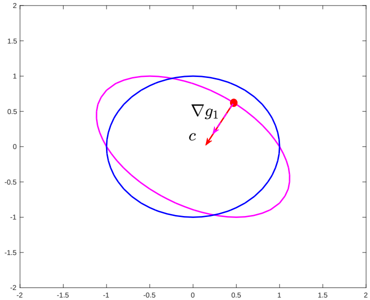




Example ($\mathcal{B}, \mathcal{R} \neq \emptyset$)

- The optimal solution is in the interior of the blue cone.
- The optimal solution is on the boundary of the pink cone.





Identification of the optimal partition

We define the following condition numbers:

$$\sigma_{\mathcal{B}} := \min_{i\in\mathcal{B}} \max_{x\in\mathcal{P}^*} \{x_1^i - \|x_{2:n_i}^i\|\},$$

$$\sigma_{\mathcal{N}} := \min_{i\in\mathcal{N}} \max_{(y,s)\in\mathcal{D}^*} \{s_1^i - \|s_{2:n_i}^i\|\},$$

$$\sigma_1 := \min\{\sigma_{\mathcal{B}}, \sigma_{\mathcal{N}}\},$$

$$\sigma_2 := \min_{i\in\mathcal{R}} \max_{(x;y;s)\in\mathcal{P}^*\times\mathcal{D}^*} \{x_1^i + s_1^i - \|x_{2:n_i}^i + s_{2:n_i}^i\|\},$$

$$\sigma_3 := \max_{(x;y;s)\in\mathcal{P}^*\times\mathcal{D}^*} \left\{\left\|(x;y;s)\right\|\right\}.$$

Theorem

Let $(x(\mu), y(\mu), s(\mu))$ be a central solution with

$$\mu < \tilde{\mu} := \min\left\{\frac{\sigma_1^2}{2p^2}, \frac{\sigma_1\sigma_2}{4p^2},\ \frac{1}{p}\left(\frac{1}{4\kappa}\min\left\{\frac{\sigma_1}{2p}, \frac{\sigma_2}{2p}\right\}\right)^{\frac{1}{\gamma}},\ \hat{\mu}\right\}.$$

The the optimal partition $(\mathcal{B}, \mathcal{N}, \mathcal{R}, \mathcal{T})$ can be identified from $(x(\mu), y(\mu), s(\mu))$.

Quadratic convergence to the unique optimal solution

We prove quadratic convergence of Newton's method to the unique optimal solution.

- We need the optimal partition $(\mathcal{B}, \mathcal{N}, \mathcal{R}, \mathcal{T})$ to be known.
- We need $(\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3)$ to be correctly identified.

Assumption

It is assumed that $\mu < \tilde{\mu}$ allows for a complete identification of $(\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3)$.

- The optimal partition is used to reformulate the dual problem.

Existence of $\mathcal{R}$

Assume that the primal and dual nondegeneracy conditions hold.

Lemma

Let $(x^*; y^*; s^*)$ be the unique optimal solution. Then $\mathcal{R} = \emptyset$ implies $\mathcal{T} = \emptyset$.

As a consequence, if $\mathcal{R} = \emptyset$, then

- The unique optimal solution can be obtained by solving two linear systems of equations.
- The primal and dual problems are easy to solve.

In the sequel, we assume that $\mathcal{R} \neq \emptyset$.

Problem reduction

Assume that $\mathcal{T}_1, \mathcal{T}_3 \neq \emptyset$.

- If we drop $c^i - A_i^T y \in \mathbb{L}_+^{n_i}$ for $i \in \mathcal{T}_1 \cup \mathcal{T}_3$, then we get

$$(\mathrm{D}'_{\mathrm{SOCO}}) \quad \max\left\{ b^T y : \; A_i^T y + s^i = c^i, \; s^i \in \mathbb{L}_+^{n_i}, \; i \in \{1, \ldots, p\} \setminus \{\mathcal{T}_1 \cup \mathcal{T}_3\} \right\},$$

and its dual is written as

$$(\mathrm{P}'_{\mathrm{SOCO}}) \quad \min\left\{ \sum_{i \in \{1,\ldots,p\} \setminus \{\mathcal{T}_1 \cup \mathcal{T}_3\}} (c^i)^T x^i : \right.$$
$$\left. \sum_{i \in \{1,\ldots,p\} \setminus \{\mathcal{T}_1 \cup \mathcal{T}_3\}} A_i x^i = b, \; x^i \in \mathbb{L}_+^{n_i}, \quad i \in \{1, \ldots, p\} \setminus \{\mathcal{T}_1 \cup \mathcal{T}_3\} \right\}.$$

Let $(\bar{x}; \bar{y}; \bar{s})$ be an optimal solution of $(\mathrm{P}'_{\mathrm{SOCO}})$ and $(\mathrm{D}'_{\mathrm{SOCO}})$.

Lemma

$(\bar{x}; \bar{y}; \bar{s})$ *is primal-dual nondegenerate.*

Problem reduction

It follows from the optimality conditions that

$$\begin{aligned} (x^*)^i &= \bar{x}^i, && i \in \mathcal{B} \cup \mathcal{R} \cup \mathcal{T}_2, \\ (x^*)^i &= 0, && i \in \mathcal{N} \cup \mathcal{T}_1 \cup \mathcal{T}_3, \\ y^* &= \bar{y}, && \\ (s^*)^i &= c^i - A_i^T \bar{y}, && i \in \mathcal{N} \cup \mathcal{R} \cup \mathcal{T}_3, \\ (s^*)^i &= 0, && i \in \mathcal{T}_1 \cup \mathcal{T}_2. \end{aligned}$$

Thus, if we remove the columns of $\mathcal{T}_1$ and $\mathcal{T}_3$

- we can recover the unique optimal solutions of (P) and (D).

Primal reformulation

Let $\nu^i \in \mathbb{R}^{n_i}$ for $i \in \mathcal{B} \cup \mathcal{R} \cup \mathcal{T}_2$.

The unique optimal solution $\bar{x}$ can be obtained by solving

$$
\begin{aligned}
(\mathrm{P}_{\mathrm{NLO}}) \quad \min \quad & \sum_{i \in \mathcal{B}\cup\mathcal{R}\cup\mathcal{T}_2} (c_i)^T \nu^i \\
\text{s.t.} \quad & \sum_{i \in \mathcal{B}\cup\mathcal{R}\cup\mathcal{T}_2} A_i \nu^i = b, \\
& (\nu^i)^T R_i \nu^i = 0, \qquad i \in \mathcal{R} \cup \mathcal{T}_2, \\
& \nu \in \mathcal{V},
\end{aligned}
$$

where

$$
\mathcal{V} := \left\{ \nu \mid \nu_1^i > 0,\ i \in \mathcal{R} \cup \mathcal{T}_2,\ \ \nu^i \in \operatorname{int}(\mathbb{L}_+^{n_i}),\ i \in \mathcal{B} \right\}.
$$

- $(\mathrm{P}_{\mathrm{NLO}})$ has a unique globally optimal solution.

Dual reformulation

Let $w \in \mathbb{R}^m$, $z^i \in \mathbb{R}^{n_i}$ for $i \in \mathcal{R} \cup \mathcal{N}$.

The unique optimal solution $(\bar{y}; \bar{s})$ is the globally optimal solution of

$$\begin{array}{llll} (\mathrm{D}_{\mathrm{NLO}}) & \min & -b^T w & \\ & \text{s.t.} & A_i^T w = c^i, & i \in \mathcal{B} \cup \mathcal{T}_2, \\ & & A_i^T w + z^i = c^i, & i \in \mathcal{R} \cup \mathcal{N}, \\ & & (z^i)^T R_i z^i = 0, & i \in \mathcal{R}, \\ & & z \in \mathcal{W}, & \end{array}$$

where

$$\mathcal{W} := \left\{ z \mid z_1^i > 0,\ i \in \mathcal{R},\ \ z^i \in \operatorname{int}(\mathbb{L}_+^{n_i}),\ i \in \mathcal{N} \right\}.$$

- $(\mathrm{D}_{\mathrm{NLO}})$ has a unique globally optimal solution.

First-order optimality conditions

Let $u^i \in \mathbb{R}^{n_i}$ for $i \in \mathcal{B} \cup \mathcal{T}_2 \cup \mathcal{R} \cup \mathcal{N}$ and $v \in \mathbb{R}^{|\mathcal{R}|}$.

The first-order optimality conditions for $(\mathrm{D}_{\mathrm{NLO}})$ are given by

$$
\begin{cases}
-\sum_{i \in \mathcal{B} \cup \mathcal{T}_2 \cup \mathcal{R} \cup \mathcal{N}} A_i u^i & = b, \\
-u^i - 2v_i R_i z^i & = 0, \quad i \in \mathcal{R}, \\
-u^i & = 0, \quad i \in \mathcal{N}, \\
A_i^T w & = c^i, \quad i \in \mathcal{B} \cup \mathcal{T}_2, \\
A_i^T w + z^i & = c^i, \quad i \in \mathcal{R} \cup \mathcal{N}, \\
(z^i)^T R_i z^i & = 0, \quad i \in \mathcal{R}, \\
z \in \mathcal{W}. &
\end{cases}
$$

- It bears a striking resemblance to the optimality conditions of second-cone program.

Constraint qualification

Let $(\bar{w}; \bar{z})$ be the unique globally optimal solution of $(\mathrm{D}_{\mathrm{NLO}})$.

Lemma

Under the dual nondegeneracy condition, the Jacobian of equality constraints at $(\bar{w}; \bar{z})$ has full row rank.

- There exist unique Lagrange multipliers.

Nonsingularity of the Jacobian

The first-order optimality conditions can be written as $G((w; z; u; v)) = 0$ and $z \in \mathcal{W}$, where

$$G((w;z;u;v)) := \begin{pmatrix} -\sum_{i \in \mathcal{B} \cup \mathcal{T}_2 \cup \mathcal{R} \cup \mathcal{N}} A_i u^i - b \\ -u^i - 2 v_i R_i z^i \\ -u^i \\ A_i^T w - c^i \\ A_i^T w + z^i - c^i \\ (z^i)^T R_i z^i \end{pmatrix}.$$

Let $(\bar{u}; \bar{v})$ be the unique Lagrange multipliers associated with $(\bar{w}; \bar{z})$.

Lemma

Under the primal and dual nondegeneracy conditions, the Jacobian ∇G is nonsingular at $(\bar{w}; \bar{z}; \bar{u}; \bar{v})$.

- The primal nondegeneracy leads to a second-order condition at the globally optimal solution.

Quadratic convergence

Let ϵ be the convergence region of Newton's method.

Theorem

Assume that the primal and dual nondegeneracy conditions hold. Let

$$\mu < \min\left\{p^{-1}\left(4\sqrt{2}\beta_2\kappa\left(\sqrt{3}+\frac{2p\sqrt{|\mathcal{R}|}}{\sigma_2}(1+\frac{2\sigma_3}{\sigma_2})\right)\right)^{-\frac{1}{\gamma}},\ \tilde{\mu}\right\},$$

in which β_2 denotes an upper bound for $\left\|\nabla G((\bar{w};\bar{z};\bar{u};\bar{v}))^{-1}\right\|$.

Then Newton's method converges to $(\bar{x};\bar{y};\bar{s})$ with quadratic rate.

Discussion

To establish quadratic convergence:

- If strict complementarity holds,

$$\mu < \min \left\{ p^{-1} (4\sqrt{3}\beta_1 \kappa)^{-\frac{1}{\gamma}},\ \hat{\mu} \right\}.$$

- If strict complementarity fails,

$$\mu < \min \left\{ p^{-1} \left(4\sqrt{2}\beta_2 \kappa \left(\sqrt{3} + \frac{2p\sqrt{|\mathcal{R}|}}{\sigma_2} (1 + \frac{2\sigma_3}{\sigma_2}) \right) \right)^{-\frac{1}{\gamma}},\ \tilde{\mu} \right\}.$$

- Quadratic convergence is harder to achieve when strict complementarity fails.
- μ has to be small enough so that the optimal partition can be identified.

Conclusions:

Under the primal and dual nondegeneracy conditions:

- We proved quadratic convergence of Newton's method to the strict complementarity solution.
- We proved quadratic convergence of Newton's method to the maximally complementary solution.

Future directions:

- Strong second-order conditions to release the assumption on the identification of $(\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3)$.
- Quadratic convergence to the unique optimal solution of semidefinite optimization using the optimal partition.

If you would like to see more about parametric second-order cone optimization, come to the other session tomorrow.

Thank you for your attention

Any questions?