Modern Optimization Theory: Optimality Conditions and Perturbation Analysis

> Defeng Sun Department of Mathematics National University of Singapore Republic of Singapore

> > July 5-10, 2006

3 **Perturbation Analysis** Let us consider (OP) $\min_{x \in X} \quad f(x)$ s.t. $G(x) \in K$, where $f: X \to \Re$ and $G: X \to Y$ are \mathcal{C}^2 (twice continuously differentiable), X, Y finite-dimensional real Hilbert vector spaces each equipped with a scalar product $\langle \cdot, \cdot \rangle$ and its induced norm $\|\cdot\|$, and K is a closed convex set in Y.

The Lagrangian function $L: X \times Y \to \Re$ for (OP) is defined by

$$L(x,\mu) := f(x) + \langle \mu, G(x) \rangle, \quad (x,\mu) \in X \times Y.$$

If \bar{x} is a locally optimal solution to (OP) and the following Robinson's CQ holds at \bar{x} :

 $0 \in \inf\{G(\bar{x}) + \mathcal{J}G(\bar{x})X - K\},\$

(or $\mathcal{J}G(\bar{x})X + \mathcal{T}_K(G(\bar{x})) = Y)$,

then there exists a Lagrangian multiplier $\bar{\mu} \in Y$, together with \bar{x} , satisfying the KKT condition:

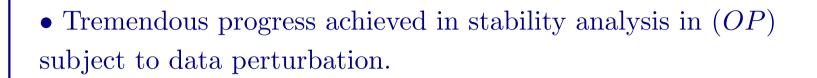
 $\nabla_x L(\bar{x}, \bar{\mu}) = 0$ and $\bar{\mu} \in \mathcal{N}_K(G(\bar{x}))$,

(or $\nabla_x L(\bar{x}, \bar{\mu}) = 0$ and $G(\bar{x}) = \Pi_K (G(\bar{x}) + \bar{\mu})$)

and equivalently if K is a closed convex cone

 $\nabla f(\bar{x}) + \nabla G(\bar{x})\bar{\mu} = 0$ and $K \ni G(\bar{x}) \perp (-\bar{\mu}) \in K^*$.

Let $\mathcal{M}(\bar{x})$ denote the set of Lagrangian multipliers.



• K is a polyhedral set, the theory quite complete. Especially for

 $egin{aligned} \min_{x\in\Re^n} & f(x) \ ext{s.t.} & h(x) = 0 \,, \ & g(x) \leq 0 \,. \end{aligned}$

(NLP)

For (NLP), Robinson's CQ reduces to the Mangasarian-Fromovitz constraint qualification (MFCQ):

 $\begin{cases} \mathcal{J}h_i(\bar{x}), & i = 1, \dots, m, \text{ are linearly independent,} \\ \exists \ d \in X: & \mathcal{J}h_i(\bar{x})d = 0, \ i = 1, \dots, m, \ \mathcal{J}g_j(\bar{x})d < 0, \ j \in \mathcal{I}(\bar{x}), \end{cases}$

where

$$\mathcal{I}(\bar{x}) := \{ j : g_j(\bar{x}) = 0, j = 1, \dots, p \}.$$

A stronger notion than the MFCQ in (NLP) is the linear independence constraint qualification (LICQ):

 $\{\mathcal{J}h_i(\bar{x})\}_{i=1}^m$ and $\{\mathcal{J}g_j(\bar{x})\}_{j\in\mathcal{I}(\bar{x})}$ are linearly independent.

 $\mathcal{M}(\bar{x})$ is nonempty and bounded if and only if the MFCQ holds at \bar{x} while the LICQ implies that $\mathcal{M}(\bar{x})$ is a singleton.

In 1980, Robinson^a introduced the far-reaching concept of <u>strong regularity</u> for generalized equations (KKT system is a special case) and <u>the strong second order sufficient condition (SSOSC)</u> for (NLP) (the later is also developed by Luenberger^b).

Robinson proved for (NLP):

 $SSOSC + LICQ \implies Strong Regularity.$

^aS.M. ROBINSON. Strongly regular generalized equations. Mathematics of Operations Research 5 (1980) 43–62.

^bD.G. LUENBERGER. Introduction to Linear and Nonlinear Programming, Addison-Wesley (London, 1973.) Jongen, Mobert, Rückmann, and Tammer^a; Bonnans and Sulem^b; Dontchev and Rockafellar^c proved:

 $SSOSC + LICQ \iff Strong Regularity.$

^aH.TH. JONGEN, T. MOBERT, J. RÜCKMANN, AND K. TAMMER. On inertia and Schur complement in optimization. *Linear Algebra and its Applications* 95 (1987) 97–109.

^bJ.F. BONNANS AND A. SULEM. Pseudopower expansion of solutions of generalized equations and constrained optimization problems. *Mathematical Pro*gramming 70 (1995) 123–148.

^cA.L. DONTCHEV AND R.T. ROCKAFELLAR. Characterizations of strong regularity for variational inequalities over polyhedral convex sets. *SIAM Journal on Optimization* 6 (1996) 1087–1105. In the above characterizations, K is a polyhedral set. Here we focus on the nonlinear semidefinite programming

(NLSDP)

$$\min_{x \in X} \quad f(x) \\ \text{s.t.} \quad h(x) = 0 , \\ g(x) \in \mathcal{S}^p_+ .$$

Difficulty:

 \mathcal{S}^p_+ is not a polyhedral set.

Let $A \in \mathcal{S}^p$ have the following spectral decomposition

$$A = P\Lambda P^T,$$

where Λ is the diagonal matrix of eigenvalues of A and P is a corresponding orthogonal matrix of orthonormal eigenvectors. Then

$$A_+ := \Pi_{\mathcal{S}^p_+}(A) = P\Lambda_+ P^T.$$

Define

$$\alpha := \{i : \lambda_i > 0\}, \ \beta := \{i : \lambda_i = 0\}, \ \gamma := \{i : \lambda_i < 0\}.$$

Write

$$\Lambda = \begin{bmatrix} \Lambda_{\alpha} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Lambda_{\gamma} \end{bmatrix} \text{ and } P = \begin{bmatrix} P_{\alpha} & P_{\beta} & P_{\gamma} \end{bmatrix}.$$

Define $U \in \mathcal{S}^p$:

$$U_{ij} := \frac{\max\{\lambda_i, 0\} + \max\{\lambda_j, 0\}}{|\lambda_i| + |\lambda_j|}, \quad i, j = 1, \dots, p,$$

where 0/0 is defined to be 1.

The tangent cone of \mathcal{S}^p_+ at $A_+ = \prod_{\mathcal{S}^p_+} (A)$ is ^a:

$$\mathcal{T}_{\mathcal{S}^p_+}(A_+) = \{ B \in \mathcal{S}^p : P^T_{\bar{\alpha}} B P_{\bar{\alpha}} \succeq 0 \}.$$

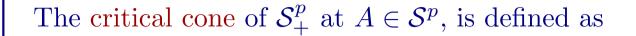
and the lineality space of $\mathcal{T}_{\mathcal{S}^p_+}(A_+)$, i.e., the largest linear space in $\mathcal{T}_{\mathcal{S}^p_+}(A_+)$,

$$\lim \left(\mathcal{T}_{\mathcal{S}^p_+}(A_+) \right) = \{ B \in \mathcal{S}^n : P^T_{\bar{\alpha}} B P_{\bar{\alpha}} = 0 \},$$

where $\bar{\alpha} := \{1, \ldots, p\} \setminus \alpha$ and $P_{\bar{\alpha}} := [P_{\beta} \ P_{\gamma}].$

^aV.I. ARNOLD. Matrices depending on parameters. Russian Mathematical Surveys, 26 (1971) 29–43.

2



$$C(A; \mathcal{S}^p_+) := \mathcal{T}_{\mathcal{S}^p_+}(A_+) \cap (A_+ - A)^{\perp},$$
$$= \left\{ B \in \mathcal{S}^p : P^T_{\beta} B P_{\beta} \succeq 0, \ P^T_{\beta} B P_{\gamma} = 0, \ P^T_{\gamma} B P_{\gamma} = 0 \right\}$$

The affine hull of $C(A; \mathcal{S}^P_+)$, aff $(C(A; \mathcal{S}^P_+))$, can be written as

aff
$$(C(A; \mathcal{S}^p_+)) = \{B \in \mathcal{S}^p : P^T_\beta B P_\gamma = 0, P^T_\gamma B P_\gamma = 0\}.$$

Definition 3.1 For any $B \in S^p$, define the linear-quadratic function $\Upsilon_B : S^p \times S^p \to \Re$ by

$$\Upsilon_B(\Gamma, A) := 2 \left\langle \Gamma, A B^{\dagger} A \right\rangle, \quad (\Gamma, A) \in \mathcal{S}^p \times \mathcal{S}^p,$$

where B^{\dagger} is the Moore-Penrose pseudo-inverse of B.

Proposition 3.1 Suppose that $B \in \mathcal{S}^p_+$ and $\Gamma \in \mathcal{N}_{\mathcal{S}^p_+}(B)$, *i.e.*,

 $B = \Pi_{\mathcal{S}^p_+}(B + \Gamma) \,.$

Then for any $V \in \partial \Pi_{\mathcal{S}^p_+}(B+\Gamma)$ and $\Delta B, \Delta \Gamma \in \mathcal{S}^p$ such that $\Delta B = V(\Delta B + \Delta \Gamma)$, it holds that

 $\langle \Delta B, \Delta \Gamma \rangle \ge -\Upsilon_B(\Gamma, \Delta B).$

Let \bar{x} be a stationary point of (NLSDP). Let $(\bar{\zeta}, \overline{\Gamma}) \in \mathcal{M}(\bar{x})$ such that

$$\nabla_x L(\bar{x}, \bar{\zeta}, \overline{\Gamma}) = 0, \quad -h(\bar{x}) = 0, \quad \text{and} \quad \overline{\Gamma} \in \mathcal{N}_{\mathcal{S}^p_+}(g(\bar{x})).$$

Let $A := g(\bar{x}) + \bar{\Gamma}$ and^a

$$g(\bar{x}) = P \begin{bmatrix} \Lambda_{\alpha} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} P^{T}, \text{ and } \overline{\Gamma} = P \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Lambda_{\gamma} \end{bmatrix} P^{T}.$$

^aSince $g(\bar{x})$ and $\overline{\Gamma}$ commute, we can simultaneously diagonalize them.

The critical cone $C(\bar{x})$ of (NLSDP) at \bar{x} is

$$C(\bar{x}) = \left\{ d : \mathcal{J}h(\bar{x})d = 0, \mathcal{J}g(\bar{x})d \in \mathcal{T}_{\mathcal{S}^{p}_{+}}(g(\bar{x})), \mathcal{J}f(\bar{x})d = 0 \right\}$$
$$= \left\{ d : \mathcal{J}h(\bar{x})d = 0, \quad P^{T}_{\beta}(\mathcal{J}g(\bar{x})d)P_{\beta} \succeq 0, \right.$$
$$P^{T}_{\beta}(\mathcal{J}g(\bar{x})d)P_{\gamma} = 0, \quad P^{T}_{\gamma}(\mathcal{J}g(\bar{x})d)P_{\gamma} = 0 \right\}.$$

The difficulty is that the affine hull of $C(\bar{x})$, $\operatorname{aff}(C(\bar{x}))$, has no explicit formula.

Define the following outer approximation set to $\operatorname{aff}(C(\bar{x}))$ with respect to $(\bar{\zeta}, \overline{\Gamma})$ by

$$\operatorname{app}(\bar{\zeta},\overline{\Gamma}) := \left\{ d : \mathcal{J}h(\bar{x})d = 0, \quad \mathcal{J}g(\bar{x})d \in \operatorname{aff}\left(C(A;\mathcal{S}^p_+)\right) \right\}.$$

It holds that

$$\operatorname{app}(\bar{\zeta}, \overline{\Gamma}) = \left\{ d : \mathcal{J}h(\bar{x})d = 0, \ P_{\beta}^{T}(\mathcal{J}g(\bar{x})d)P_{\gamma} = 0, \\ P_{\gamma}^{T}(\mathcal{J}g(\bar{x})d)P_{\gamma} = 0 \right\}.$$

Then by the definition of $\operatorname{aff}(C(\bar{x}))$, we have for any $(\bar{\zeta}, \overline{\Gamma}) \in \mathcal{M}(\bar{x})$ that

 $\operatorname{aff}(C(\bar{x})) \subseteq \operatorname{app}(\bar{\zeta},\overline{\Gamma}).$

The two sets $\operatorname{aff}(C(\overline{x}))$ and $\operatorname{app}(\overline{\zeta},\overline{\Gamma})$ coincide if the strict complementary condition holds at $(\overline{x},\overline{\zeta},\overline{\Gamma})$:

 $\operatorname{rank}(g(\bar{x})) + \operatorname{rank}(\overline{\Gamma}) = p,$

where "rank" denotes the rank of a square matrix.

In general, these two sets may be different even if $\mathcal{M}(\bar{x})$ is a singleton as in the case for (NLP).

Proposition 3.2 Suppose that $(\overline{\zeta}, \overline{\Gamma})$ satisfies the following strict constraint qualification:

$$\begin{pmatrix} \mathcal{J}h(\bar{x}) \\ \mathcal{J}g(\bar{x}) \end{pmatrix} X + \begin{pmatrix} 0 \\ \mathcal{T}_{\mathcal{S}^p_+}(g(\bar{x})) \cap \overline{\Gamma}^{\perp} \end{pmatrix} = \begin{pmatrix} \Re^m \\ \mathcal{S}^p \end{pmatrix}.$$

Then $\mathcal{M}(\bar{x})$ is a singleton, i.e., $\mathcal{M}(\bar{x}) = \{(\bar{\zeta}, \overline{\Gamma})\}, and$ aff $(C(\bar{x})) = \operatorname{app}(\bar{\zeta}, \overline{\Gamma}).$ Recall that the "no-gap" second order necessary condition and the second order sufficient condition for (NLSDP) can be stated as follows:

Theorem 3.1 Let $K = \{0\} \times S^p_+ \subset \Re^m \times S^p$. Suppose that \bar{x} is a locally optimal solution to (NLSDP) and Robinson's CQ holds at \bar{x} . Then

 $\sup_{\mu \in \mathcal{M}(\bar{x})} \left\{ \left\langle d, \nabla_{xx}^2 L(\bar{x}, \mu) d \right\rangle - \sigma \left(\mu, \mathcal{T}_K^2(G(\bar{x}), \mathcal{J}G(\bar{x}) d) \right) \right\} \ge 0$

for all $d \in C(\bar{x})$.

(continued)

Conversely, let \bar{x} be a feasible solution to (NLSDP) such that $\mathcal{M}(\bar{x})$ is nonempty. Suppose that Robinson's CQ holds at \bar{x} . Then the following condition

$$\sup_{\mu \in \mathcal{M}(\bar{x})} \left\{ \left\langle d, \nabla_{xx}^2 L(\bar{x}, \mu) d \right\rangle - \sigma \left(\mu, \mathcal{T}_K^2(G(\bar{x}), \mathcal{J}G(\bar{x})d) \right) \right\} > 0$$

for all $d \in C(\bar{x}) \setminus \{0\}$ is necessary and sufficient for the quadratic growth condition at the point \bar{x} :

 $f(x) \ge f(\bar{x}) + c \|x - \bar{x}\|^2 \quad \forall x \in \widehat{N} \text{ such that } G(x) \in K$

for some constant c > 0 and a neighborhood \widehat{N} of \overline{x} in X.

Proposition 3.3 Let \bar{x} be a feasible solution to (NLSDP) such that $\mathcal{M}(\bar{x})$ is nonempty. Then for any $(\zeta, \Gamma) \in \mathcal{M}(\bar{x})$ with $\zeta \in \Re^m$ and $\Gamma \in S^p$, one has

$$\Upsilon_{g(\bar{x})}(\Gamma, \mathcal{J}g(\bar{x})d) = \sigma\left(\Gamma, \mathcal{T}^2_{\mathcal{S}^p_+}(g(\bar{x}), \mathcal{J}g(\bar{x})d)\right) \quad \forall d \in C(\bar{x}),$$

where

$$\Upsilon_B(\Gamma, A) = 2 \langle \Gamma, AB^{\dagger}A \rangle, \quad (\Gamma, A) \in \mathcal{S}^p \times \mathcal{S}^p.$$

Definition 3.2 Let \bar{x} be a stationary point of (NLSDP). We say that the strong second order sufficient condition (SSOSC) holds at \bar{x} if

$$\sup_{(\zeta,\Gamma)\in\mathcal{M}(\bar{x})}\left\{\left\langle d,\nabla^2_{xx}L(\bar{x},\zeta,\Gamma)d\right\rangle-\Upsilon_{g(\bar{x})}(\Gamma,\mathcal{J}g(\bar{x})d)\right\}>0$$

for all $d \in \widehat{C}(\bar{x}) \setminus \{0\}$, where for any $(\zeta, \Gamma) \in \mathcal{M}(\bar{x})$, $(\zeta, \Gamma) \in \Re^m \times S^p$ and

$$\widehat{C}(\bar{x}) := \bigcap_{(\zeta, \Gamma) \in \mathcal{M}(\bar{x})} \operatorname{app}(\zeta, \Gamma).$$

Next, we define a <u>nondegeneracy condition</u> for (NLSDP), which is an analogue of the <u>LICQ</u> for (NLP). The concept of nondegeneracy originally appeared in Robinson^a for (OP).

Definition 3.3 We say that a feasible point \bar{x} to (OP) is constraint nondegenerate if

 $\mathcal{J}G(\bar{x})X + \ln(\mathcal{T}_K(\bar{y})) = Y,$

where $\bar{y} := G(\bar{x})$.

^aS.M. ROBINSON. Local structure of feasible sets in nonlinear programming, Part II: Nondegeneracy. *Mathematical Programming Study* 22 (1984) 217–230.

Write down the KKT condition as $F(x,\zeta,\Gamma): = \begin{bmatrix} \nabla L(x,\zeta,\Gamma) \\ -h(x) \\ -g(x) + \Pi_{\mathcal{S}^{p}_{+}}(g(x) + \Gamma) \end{bmatrix} = \begin{bmatrix} \nabla_{x}L(x,\zeta,\Gamma) \\ -h(x) \\ \Gamma - \Pi_{\mathcal{S}^{p}_{-}}(\Gamma + g(x)) \end{bmatrix} = 0,$ which is equivalent to the following generalized equation:

 $0 \in \phi(z) + \mathcal{N}_D(z),$

where ϕ is \mathcal{C}^1 and D is a closed convex set in Z.

Definition 3.4 [Robinson'80] Let \overline{z} be a solution of the generalized equation. We say that \overline{z} is a <u>strongly regular</u> solution if there exist neighborhoods \mathcal{B} of the origin $0 \in \mathbb{Z}$ and \mathcal{V} of \overline{z} such that for every $\delta \in \mathcal{B}$, the following linearized generalized equation

 $\delta \in \phi(\bar{z}) + \mathcal{J}\phi(\bar{z})(z - \bar{z}) + \mathcal{N}_D(z)$

has a unique solution in \mathcal{V} , denoted by $z_{\mathcal{V}}(\delta)$, and the mapping $z_{\mathcal{V}}: \mathcal{B} \to \mathcal{V}$ is Lipschitz continuous.

Let U be a Banach space and $f: X \times U \to \Re$ and $G: X \times U \to Y$. We say that (f(x, u), G(x, u)), with $u \in U$, is a C^2 -smooth parameterization of (OP) if $f(\cdot, \cdot)$ and $g(\cdot, \cdot)$ are C^2 and there exists a $\bar{u} \in U$ such that $f(\cdot, \bar{u}) = f(\cdot)$ and $G(\cdot, \bar{u}) = G(\cdot)$. The corresponding parameterized problem takes the form:

 (OP_u)

$$\min_{x \in X} \quad f(x, u)$$

s.t. $G(x, u) \in K$.

We say that a parameterization is canonical if $U := X \times Y$, $\bar{u} = (0,0) \in X \times Y$, and

 $(f(x, u), G(x, u)) := (f(x) - \langle u_1, x \rangle, G(x) + u_2), \quad x \in X.$

Definition 3.5 [Bonnans and Shapiro'00] Let \bar{x} be a stationary point of (OP). We say that the <u>uniform second order (quadratic) growth condition</u> holds at \bar{x} with respect to a C^2 -smooth parameterization (f(x, u), G(x, u)) if there exist c > 0 and neighborhoods \mathcal{V}_X of \bar{x} and \mathcal{V}_U of \bar{u} such that for any $u \in \mathcal{V}_U$ and any stationary point $x(u) \in \mathcal{V}_X$ of (OP_u), the following holds:

$$f(x,u) \ge f(x(u),u) + c \|x - x(u)\|^2 \quad \forall x \in \mathcal{V}_X \text{ such that } G(x,u) \in K.$$

We say that the uniform second order growth condition holds at \bar{x} if the above inequality holds for every C^2 -smooth parameterization of (OP). **Definition 3.6** [Kojima^a and Bonnans and Shapiro'00]

Let \bar{x} be a stationary point of (OP). We say that \bar{x} is <u>strongly stable</u> with respect to a C^2 -smooth parameterization (f(x, u), G(x, u)) if there exist neighborhoods \mathcal{V}_X of \bar{x} and \mathcal{V}_U of \bar{u} such that for any $u \in \mathcal{V}_U$, (OP_u) has a unique stationary point $x(u) \in \mathcal{V}_X$ and $x(\cdot)$ is continuous on \mathcal{V}_U .

If this holds for any C^2 -smooth parameterization, we say that \bar{x} is strongly stable.

^aM. KOJIMA. Strongly stable stationary solutions in nonlinear programs. In: S.M. Robinson, editor, *Analysis and Computation of Fixed Points*, Academic Press (New York, 1980), pp. 93-138.

Let

$$\Phi(\delta) := F'(\bar{x}, \bar{\zeta}, \overline{\Gamma}; \delta).$$

Let $\operatorname{ind}(\phi, \overline{z})$ denote the index of a continuous function $\phi: Z \to Z$ at an isolated zero $\overline{z} \in Z$ used in degree theory.

Based on several recent results of Bonnans and Shapiro'00; Gowda^a; Pang, Sun and Sun^b; Sun and Sun'02, we get

^aM.S. GOWDA. Inverse and implicit function theorems for H-differentiable and semismooth functions. *Optimization Methods and Software* 19 (2004) 443-461.

^bJ.S. PANG, D. SUN, AND J. SUN. Semismooth homeomorphisms and strong stability of semidefinite and Lorentz complementarity problems. *Mathematics of Operations Research* 28 (2003) 39–63.

Theorem 2^a. Let \bar{x} be a locally optimal solution to (NLSDP). Suppose that Robinson's CQ holds at \bar{x} so that \bar{x} is necessarily a stationary point of (NLSDP). Let $(\bar{\zeta}, \overline{\Gamma}) \in \Re^m \times S^p$ be such that $(\bar{x}, \bar{\zeta}, \overline{\Gamma})$ is a KKT point of (NLSDP). Then the following TEN statements are equivalent:

- (a) The SSOSC holds at \bar{x} and \bar{x} is constraint nondegenerate.
- (b) Any element in $\partial F(\bar{x}, \bar{\zeta}, \overline{\Gamma})$ is nonsingular.
- (c) The KKT point $(\bar{x}, \bar{\zeta}, \overline{\Gamma})$ is strongly regular.
- (d) The uniform second order growth condition holds at \bar{x} and \bar{x} is constraint nondegenerate.
- (e) The point \bar{x} is strongly stable and \bar{x} is constraint nondegenerate.

^aD. SUN. The strong second order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications. *Mathematics of Operations Research* 31 (2006).

(continued)

- (f) F is a locally Lipschitz homeomorphism near $(\bar{x}, \bar{\zeta}, \overline{\Gamma})$.
- (g) For every $V \in \partial_B F(\bar{x}, \bar{\zeta}, \overline{\Gamma})$, sgn det $V = \operatorname{ind}(F, (\bar{x}, \bar{\zeta}, \overline{\Gamma})) = \pm 1$.
- (h) Φ is a globally Lipschitz homeomorphism.
- (i) For every $V \in \partial_B \Phi(0)$, sgn det $V = ind(\Phi, 0) = \pm 1$.
- (j) Any element in $\partial \Phi(0)$ is nonsingular.

Note that many more equivalent statements can be added by looking at statements (b) and (g).

Some unsolved problems:

- (Q1) How far can we go beyond the SDP cone? Symmetric cone (SOC is fine)? Homogeneous cone? Hyperbolic cone?
- (Q2) What can we say about the equivalent conditions in Theorem 2 if \bar{x} is assumed to be a stationary point only? Or more generally
- (Q3) How can we characterize the strong regularity for the conic complementarity problems?