Recent Developments in Nonlinear Optimization Theory

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

July 9-11, 2006

This talk is dedicated to Professor Jiye Han.

To motivate our discussions, let us consider the following simple one-dimensional optimization problem

$$\begin{array}{ll} \min\limits_{x \in \Re} & \frac{1}{2}x^2 \\ \text{s.t.} & x \geq 0\,. \end{array}$$

The corresponding Lagrangian function is

$$L(x,\lambda) := \frac{1}{2}x^2 + \langle \lambda, x \rangle\,, \quad (x,\lambda) \in \Re^2\,.$$

The unique optimal solution and its corresponding Lagrangian multiplier are given by

$$x^* = 0 \quad \& \quad \lambda^* = 0\,,$$

which satisfy the Karush-Kuhn-Tucker (KKT) condition

$$\nabla_x L(x^*, \lambda^*) = x^* + \lambda^* = 0\,, \quad 0 \leq x^* \perp (-\lambda^*) \geq 0\,.$$

The Hessian of L with respect to x^* is:

$$\nabla^2_{xx} L(x^*, \lambda^*) = I \quad (\text{the best one can dream of})\,.$$

Now, let us consider the following equivalent form:

$$\begin{array}{ll} \min\limits_{(t,x)\in\Re^2} & t \\ \text{s.t.} & x \geq 0\,, \\ & \dfrac{1}{2}x^2 \leq t\,. \end{array}$$

$$\Updownarrow$$

(SOC)

$$\min_{(t,x)\in\Re^2} \quad t$$

$$\text{s.t.} \quad x \geq 0\,,$$

$$\|(2x, 2-t)\|_2 \leq 2+t \iff (2+t, 2x, 2-t) \in \mathcal{K}^3\,,$$

where for each $n \geq 1$, $\mathcal{K}^{n+1}$ is the $(n+1)$-dimensional second-order cone

$$\mathcal{K}^{n+1} := \{(t,x) \in \Re \times \Re^n \,:\, t \geq ||x||_2\}\,.$$

The Lagrangian function for (SOC) is

$$L(t, x, \lambda, \mu) := t + \langle \lambda, x \rangle + \langle \mu, (2 + t, 2x, 2 - t) \rangle .$$

The Hessian of L with respective to (t, x) now turns to be

$$\nabla^2_{(t,x)(t,x)} L(t, x, \lambda, \mu) = 0 \quad \text{(too bad???)} .$$

The seemingly harmless transformations have completed changed the Hessian of the corresponding Lagrangian functions (from I to 0).

This change should be related to the non-polyhedral structure of $\mathcal{K}^{n+1}$.

This simple example suggests that when we talk about second-order optimality conditions and perturbation analysis, we need to include the "curvature" of the non-polyderal set involved.

Let's now turn to the general optimization problem

(OP)

$$\min_{x \in X} \quad f(x)$$
$$\text{s.t.} \quad G(x) \in K\,,$$

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[a] each equipped with a scalar product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$, and K is a closed convex set in Y.

[a] A real vector space $\mathcal{H}$ is called a Hilbert space if there is an "inner product" (or a "scalar product") denoted $\langle \cdot, \cdot \rangle$ satisfying i) $\langle x, y \rangle = \langle y, x \rangle \ \forall\, x, y \in \mathcal{H}$; ii) $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle \ \forall\, x, y$, and $z \in \mathcal{H}$; iii) $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle \ \forall\, \alpha \in \Re$ and $x, y \in \mathcal{H}$; iv) $\langle x, x \rangle \geq 0 \ \forall\, x \in \mathcal{H}$; and v) $\langle x, x \rangle = 0$ only if $x = 0$.

Some notation:

For any given $\overline{x} \in X$ and $\varepsilon > 0$, let the open ball be
$\mathbb{B}_\varepsilon(\overline{x}) := \{x \in X \,:\, \|x - \overline{x}\| < \varepsilon\}$.

Suppose that X' and Y' are two finite-dimensional real Hilbert spaces and that $F : X \times X' \mapsto Y'$. If F is Fréchet-differentiable at $(x, x') \in X \times X'$, then we use $\mathcal{J}F(x, x')$ (respectively, $\mathcal{J}_x F(x, x')$) to denote the Fréchet-derivative of F at (x, x') (respectively, the partial Fréchet-derivative of F at (x, x') with respect to x).

Let $\nabla F(x, x') := \mathcal{J}F(x, x')^*$, the adjoint of $\mathcal{J}F(x, x')$ (respectively, $\nabla_x F(x, x') := \mathcal{J}_x F(x, x')^*$, the adjoint of $\mathcal{J}_x F(x, x')$).

If F is twice Fréchet-differentiable at $(x, x') \in X \times X'$, we define

$$\mathcal{J}^2 F(x, x') := \mathcal{J}(\mathcal{J} F)(x, x')$$

$$\mathcal{J}^2_{xx} F(x, x') := \mathcal{J}_x(\mathcal{J}_x F)(x, x'),$$

$$\nabla^2 F(x, x') := \mathcal{J}(\nabla F)(x, x')\,,$$

$$\nabla^2_{xx} F(x, x') := \mathcal{J}_x(\nabla_x F)(x, x')\,.$$

For any closed set $D \subseteq Y$, we write $\mathcal{T}_D^i(y)$ and $\mathcal{T}_D(y)$ for the inner tangent cone and the contingent (Bouligand) cone of D at y, respectively. That is,

$$\mathcal{T}_D^i(y) = \{d \in Y \ : \ \mathrm{dist}(y + td, D) = o(t)\,,\ t \geq 0\}$$

and

$$\mathcal{T}_D(y) = \{d \in Y \ : \ \exists\ t_k \downarrow 0,\ \mathrm{dist}(y + t_k d, D) = o(t_k)\}.$$

When D is a closed convex set, the inner tangent cone and the contingent cone are equal:

$$\mathcal{T}_D(y) = \mathcal{T}_D^i(y) = \{d \in Y \, : \, \mathrm{dist}(y + td, D) = o(t)\,,\, t \geq 0\}, \quad y \in D\,.$$

We use $\mathcal{N}_K(y)$ to denote the normal cone of K at y in the sense of convex analysis

$$\mathcal{N}_K(y) \, = \begin{cases} \{d \in Y \, : \, \langle d, z - y\rangle \leq 0 \;\; \forall\, z \in K\} & \text{if } y \in K, \\ \emptyset & \text{if } y \notin K. \end{cases}$$

The inner and outer second order tangent sets[a] to the set D at the point $y \in D$ and in the direction $d \in Y$ are defined by

$$\mathcal{T}_D^{i,2}(y,d) := \{w \in Y \,:\, \mathrm{dist}(y + td + \frac{1}{2}t^2 w, D) = o(t^2)\,,\, t \geq 0\}$$

and

$$\mathcal{T}_D^{2}(y,d) := \{w \in Y \,:\, \exists\, t_k \downarrow 0 \,\&\, \mathrm{dist}(y + t_k d + \frac{1}{2}t_k^2 w, D) = o(t_k^2)\}.$$

[a] J.F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*, Springer (New York, 2000).

We have $\mathcal{T}_D^{i,2}(z,d) \subseteq \mathcal{T}_D^2(y,d)$ and $\mathcal{T}_D^{i,2}(z,d) = \emptyset$ (respectively, $\mathcal{T}_D^2(z,d) = \emptyset$) if $d \notin \mathcal{T}_D^i(y)$ (respectively, $d \notin \mathcal{T}_D(y)$).

In general, $\mathcal{T}_D^{i,2}(z,d) \neq \mathcal{T}_D^2(z,d)$ even if D is convex. However, when $K := \{0\} \times \mathcal{S}_+^p \subset Y := \Re^m \times \mathcal{S}^p$,

$$\mathcal{T}_K^{i,2}(y,d) = \mathcal{T}_K^2(y,d) \quad \forall\, y, d \in Y.$$

where $\mathcal{S}^p$ is the linear space of all $p \times p$ real symmetric matrices, and $\mathcal{S}_+^p$ is the cone of all $p \times p$ positive semidefinite matrices.

Recall that for any set $D \subseteq Z$, the support function of the set D is defined as

$$\sigma(y, D) := \sup_{z \in D} \langle z, y \rangle, \quad y \in Y\,.$$

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

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

Let $\bar{x}$ be a feasible solution to (OP). Robinson's constraint qualification (CQ) is as follows:

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

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

If $\bar{x}$ is a locally optimal solution to (OP) and Robinson's CQ holds at $\bar{x}$, then there exists a Lagrangian multiplier $\bar{\mu} \in Y$, together with $\bar{x}$, satisfying the KKT condition:

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

and equivalently if K is a closed convex cone

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

where K^* is the dual cone of K given by

$$K^* := \{d \in Y \,:\, \langle d, y\rangle \geq 0 \quad \forall\, y \in K\}\,.$$

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

• Tremendous progress achieved in necessary and sufficient second-order optimality conditions and stability analysis in (OP) subject to data perturbation.

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

(NLP)

$$\begin{array}{ll} \min\limits_{x \in \Re^n} & f(x) \\ \text{s.t.} & h(x) = 0\,, \\ & g(x) \leq 0\,. \end{array}$$

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

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

where

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

A stronger notion than the MFCQ in (NLP) is the LICQ:

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

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

Robinson proved for (NLP):

$$\text{SSOSC} + \text{LICQ} \Longrightarrow \text{Strong Regularity.}$$

[a] S.M. Robinson. Strongly regular generalized equations. *Mathematics of Operations Research* 5 (1980) 43–62.

[b] D.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:

$$\text{SSOSC} + \text{LICQ} \Longleftarrow \text{Strong Regularity.}$$

[a] H.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.

[b] J.F. Bonnans and A. Sulem. Pseudopower expansion of solutions of generalized equations and constrained optimization problems. *Mathematical Programming* 70 (1995) 123–148.

[c] A.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. In this talk, we focus on the nonlinear semidefinite programming

$(NLSDP)$

$$\begin{array}{ll} \min\limits_{x \in X} & f(x) \\ \text{s.t.} & h(x) = 0\,, \\ & g(x) \in \mathcal{S}^p_+\,. \end{array}$$

Difficulty:

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

Note that (NLSDP) can be equivalently written as either semi-infinite programming problem

$$\begin{aligned} \min_{x \in X} \quad & f(x) \\ \text{s.t.} \quad & h(x) = 0\,, \\ & d^T g(x) d \geq 0 \quad \forall \ \|d\|_2 = 1 \end{aligned}$$

or nonsmooth optimization problem

$$\begin{aligned} \min_{x \in X} \quad & f(x) \\ \text{s.t.} \quad & h(x) = 0\,, \\ & \lambda_{\min}(g(x)) \geq 0\,, \end{aligned}$$

where $\lambda_{\min}(g(x))$ is the smallest eigenvalue of $g(x)$.

Indeed, early in seventies and eighties of the last century, researchers working on semi-infinite programming problems and nonsmooth optimization problems realized that in order to get satisfactory second-order necessary and sufficient conditions, an additional term, which represents the curvature of the set K, must be added.

As mentioned earlier, we shall use (NLSDP) as an example to demonstrate this.

Let $\Xi : \mathcal{O} \subseteq Y \rightarrow Z$ be a locally Lipschitz continuous function on the open set $\mathcal{O}$, where Z is another finite-dimensional real Hilbert space.

We denote by $\mathcal{O}_\Xi$ the set of points in $\mathcal{O}$ where Ξ is Fréchet differentiable. Then Clarke's generalized Jacobian[a] of Ξ at y is:

$$\partial \Xi(y) := \text{conv}\{\partial_B \Xi(y)\},$$

where "conv" denotes the convex hull and

$$\partial_B \Xi(y) := \{V \,:\, V = \lim_{k\rightarrow\infty} \mathcal{J}\Xi(y^k)\,,\; y^k \rightarrow y\,,\; y^k \in \mathcal{O}_\Xi\}.$$

[a] F.H. Clarke. *Optimization and Nonsmooth Analysis*, John Wiley and Sons (New York, 1983).

Let D be a closed convex set in Z. Let $\Pi_D : Z \to Z$ denote the metric projector over D:

$$\begin{array}{ll} \min & \frac{1}{2}\langle z-y, z-y\rangle \\ \text{s.t.} & z \in D. \end{array}$$

The operator $\Pi_D(\cdot)$ is F-differentiable almost everywhere in Z and for any $y \in Z$, $\partial\Pi_D(y)$ is well defined.

Lemma.[a] For any $y \in Z$ and $V \in \partial\Pi_D(y)$, (a) V is self-adjoint; (b) $\langle d, Vd\rangle \geq 0 \quad \forall\, d \in Z$; and (c) $\langle Vd, d - Vd\rangle \geq 0 \quad \forall\, d \in Z$.

---

[a] F. Meng, D. Sun, and G. Zhao. Semismoothness of solutions to generalized equations and the Moreau-Yosida regularization. *Mathematical Programming* 104 (2005) 561–581.

For A and B in $\mathcal{S}^p$,

$$\langle A, B \rangle := \mathrm{Tr}\left(A^T B\right) = \mathrm{Tr}\left(AB\right) ,$$

where "Tr" denotes the trace of a square matrix (i.e., the sum of all diagonal elements of the symmetric matrix). 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, one can check without difficulty that (see Higham[a] and Tseng[b]):

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

[a] N.J. Higham. Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra and Applications* 103 (1988) 103–118.

[b] P. Tseng. Merit functions for semi-definite complementarity problems. *Mathematical Programming* 83 (1998) 159–185.

Define

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

Write

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

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

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

where $0/0$ is defined to be 1.

The operator $\Pi_{\mathcal{S}^p_+}(\cdot)$ is directionally differentiable.[a] Sun and Sun[b] showed that $\Pi_{\mathcal{S}^p_+}(\cdot)$ is strongly semismooth at A, i.e., in addition to the directional differentiability of $\Pi_{\mathcal{S}^p_+}(\cdot)$ at A, for any $H \in \mathcal{S}^p$ and $V \in \partial\Pi_{\mathcal{S}^p_+}(A+H)$ we have

$$\Pi_{\mathcal{S}^p_+}(A+H) - \Pi_{\mathcal{S}^p_+}(A) - V(H) = O(\|H\|^2)$$

and $\Pi'_{\mathcal{S}^p_+}(A;H)$ is given by

[a] J.F. Bonnans, R. Cominetti, and A. Shapiro. Sensitivity analysis of optimization problems under second order regularity constraints. *Mathematics of Operations Research* 23 (1998) 803–832 and Second order optimality conditions based on parabolic second order tangent sets. *SIAM Journal on Optimization* 9 (1999) 466–493.

[b] D. Sun and J. Sun. Semismooth matrix valued functions. *Mathematics of Operations Research* 27 (2002) 150–169.

$$\Pi'_{\mathcal{S}^p_+}(A;H) = P \begin{bmatrix} P_\alpha^T H P_\alpha & P_\alpha^T H P_\beta & U_{\alpha\gamma} \circ P_\alpha^T H P_\gamma \\ P_\beta^T H P_\alpha & \Pi_{\mathcal{S}^{|\beta|}_+}(P_\beta^T H P_\beta) & 0 \\ P_\gamma^T H P_\alpha \circ U_{\alpha\gamma}^T & 0 & 0 \end{bmatrix} P^T,$$

where $\circ$ denotes the Hadamard product. Note that $\Pi'_{\mathcal{S}^p_+}(A;H)$ does not depend on any particularly chosen P.

When $|\beta| = 0$, $\Pi_{\mathcal{S}^n_+}(\cdot)$ is continuously differentiable around A and the above formula reduces to the classical result of Löwner[a].

[a]K. Löwner. *Über monotone matrixfunctionen.* Mathematische Zeitschrift 38 (1934) 177–216.

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

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

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

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

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

[a] V.I. Arnold. *Matrices depending on parameters.* Russian Mathematical Surveys, 26 (1971) 29–43.

One may use the following relations to get $\mathcal{T}_{\mathcal{S}^p_+}(A_+)$ directly:

$$
\begin{aligned}
& \mathcal{T}_{\mathcal{S}^p_+}(A_+) \\
= & \{B \in \mathcal{S}^p \ : \ \mathrm{dist}(A_+ + tB, \mathcal{S}^p_+) = o(t), \ t \geq 0\} \\
= & \{B \in \mathcal{S}^p \ : \ \|A_+ + tB - \Pi_{\mathcal{S}^p_+}(A_+ + tB)\| = o(t), \ t \geq 0\} \\
= & \{B \in \mathcal{S}^p \ : \ \|A_+ + tB - [A_+ + t\Pi'_{\mathcal{S}^p_+}(A_+; B) + o(t)]\| = o(t), t \geq 0\} \\
= & \{B \in \mathcal{S}^p \ : B = \Pi'_{\mathcal{S}^p_+}(A_+; B)\} \,.
\end{aligned}
$$

The critical cone of $\mathcal{S}_+^p$ at $A \in \mathcal{S}^p$, is defined as

$$C(A;\mathcal{S}_+^p) := \mathcal{T}_{\mathcal{S}_+^p}(A_+) \cap (A_+ - A)^\perp,$$

$$= \left\{ B \in \mathcal{S}^p \; : \; P_\beta^T B P_\beta \succeq 0, \; P_\beta^T B P_\gamma = 0, \; P_\gamma^T B P_\gamma = 0 \right\}.$$

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

$$\mathrm{aff}\left(C(A;\mathcal{S}_+^p)\right) = \left\{ B \in \mathcal{S}^p \; : \; P_\beta^T B P_\gamma = 0, \; P_\gamma^T B P_\gamma = 0 \right\}.$$

Lemma. Let $\Psi : X \to Y$ be $\mathcal{C}^1$ on an open neighborhood $\widehat{N}$ of $\bar{x}$ and $\Xi : \mathcal{O} \subseteq Y \to Z$ be a locally Lipschitz continuous function on an open set $\mathcal{O}$ containing $\bar{y} := \Psi(\bar{x})$.

Suppose that Ξ is directionally differentiable at every point in $\mathcal{O}$ and that $J\Psi(\bar{x}) : X \to Y$ is onto. Then it holds that

$$\partial_B \Phi(\bar{x}) = \partial_B \Xi(\bar{y}) \mathcal{J}\Psi(\bar{x}), \quad \Phi(x) := \Xi(\Psi(x)), \quad x \in \widehat{N}.$$

By using the above lemma and

$$\partial_B \Pi_{\mathcal{S}^p_+}(A) = \partial_B \Theta(0)\,, \qquad \Theta(\cdot) := \Pi'_{\mathcal{S}^p_+}(A;\cdot)\,,$$

we obtain

Proposition. For any $V \in \partial_B \Pi_{\mathcal{S}^p_+}(A)$ (respectively, $\partial \Pi_{\mathcal{S}^p_+}(A)$), there exists a $W \in \partial_B \Pi_{\mathcal{S}^{|\beta|}_+}(0)$ (respectively, $\partial \Pi_{\mathcal{S}^{|\beta|}_+}(0)$) such that

$$V(H) = P \begin{bmatrix} \widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & U_{\alpha\gamma} \circ \widetilde{H}_{\alpha\gamma} \\ \widetilde{H}^T_{\alpha\beta} & W(\widetilde{H}_{\beta\beta}) & 0 \\ \widetilde{H}^T_{\alpha\gamma} \circ U^T_{\alpha\gamma} & 0 & 0 \end{bmatrix} P^T \quad \forall\, H \in \mathcal{S}^p, \tag{1}$$

where $\widetilde{H} := P^T H P$.

Conversely, for any $W \in \partial_B \Pi_{\mathcal{S}^{|\beta|}_+}(0)$ (respectively, $\partial \Pi_{\mathcal{S}^{|\beta|}_+}(0)$) , there exists a $V \in \partial_B \Pi_{\mathcal{S}^p_+}(A)$ (respectively, $\partial \Pi_{\mathcal{S}^p_+}(A)$) such that (1) holds.

Definition. For any given $B \in \mathcal{S}^p$, define the linear-quadratic function $\Upsilon_B : S^p \times \mathcal{S}^p \to \Re$ by

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

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

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

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

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]

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

[a] Since $g(\bar{x})$ and $\overline{\Gamma}$ commute, we can simultaneously diagonalize them.

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

$$
\begin{aligned}
C(\bar{x}) &= \left\{ d : \mathcal{J}h(\bar{x})d = 0, \mathcal{J}g(\bar{x})d \in \mathcal{T}_{\mathcal{S}^p_+}(g(\bar{x})), \mathcal{J}f(\bar{x})d = 0 \right\} \\
&= \Big\{ d : \mathcal{J}h(\bar{x})d = 0, \quad P_\beta^T(\mathcal{J}g(\bar{x})d)P_\beta \succeq 0, \\
&\qquad P_\beta^T(\mathcal{J}g(\bar{x})d)P_\gamma = 0, \quad P_\gamma^T(\mathcal{J}g(\bar{x})d)P_\gamma = 0 \Big\}.
\end{aligned}
$$

The difficulty is that the affine hull of $C(\bar{x})$, $\mathrm{aff}(C(\bar{x}))$, has no explicit formula. Define the following outer approximation set to $\mathrm{aff}(C(\bar{x}))$ with respect to $(\bar{\zeta}, \overline{\Gamma})$ by

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

It holds that

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

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

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

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. Suppose that $(\bar{\zeta}, \overline{\Gamma})$ 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 $\mathrm{aff}(C(\bar{x})) = \mathrm{app}(\bar{\zeta}, \overline{\Gamma})$.

By combining Theorem 3.45 and Proposition 3.136 with Theorem 3.137 in [Bonnans and Shapiro'00], we can state the "no-gap" second order necessary condition and the second order sufficient condition for $(NLSDP)$.

Theorem 1. (Second-Order Necessary and Sufficient Conditions.)

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

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

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

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

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

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

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

Since

$$\mathcal{T}^2_K(G(\bar{x}), \mathcal{J}G(\bar{x})d) \subset \mathcal{T}_{\mathcal{T}_K(G(\bar{x}))}(\mathcal{J}G(\bar{x})d)$$

and

$$\mathcal{T}_{\mathcal{T}_K(G(\bar{x}))}(\mathcal{J}G(\bar{x})d) = \mathrm{cl}\left\{\mathcal{T}_K(G(\bar{x})) + \mathrm{span}(\mathcal{J}G(\bar{x})d)\right\} ,$$

we have for any $\mu \in \mathcal{M}(\bar{x})$ and $d \in C(\bar{x})$,

$$\sigma\left(\mu, \mathcal{T}^2_K(G(\bar{x}), \mathcal{J}G(\bar{x})d)\right) \leq \sigma\left(\mu, \mathcal{T}_{\mathcal{T}_K(G(\bar{x}))}(\mathcal{J}G(\bar{x})d)\right) = 0\,.$$

Thus, unless $0 \in \mathcal{T}_K^2(G(\bar{x}), \mathcal{J}G(\bar{x})d)$ for all $h \in C(\bar{x})$ as in the case when K is a polyhedral convex set, the additional "sigma term" in the necessary and sufficient second-order conditions will not disappear.

Lemma. 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 $\Gamma \in \mathcal{S}^p$, 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

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

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

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

Next, we define a nondegeneracy condition for $(NLSDP)$, which is an analogue of the LICQ for (NLP). The concept of nondegeneracy originally appeared in Robinson[a] for (OP).

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

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

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

[a] S.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

$$F(x,\zeta,\Gamma):$$
$$= \begin{bmatrix} \nabla_x L(x,\zeta,\Gamma) \\ -h(x) \\ -g(x) + \Pi_{\mathcal{S}^p_+}(g(x)+\Gamma) \end{bmatrix} = \begin{bmatrix} \nabla_x L(x,\zeta,\Gamma) \\ -h(x) \\ \Gamma - \Pi_{\mathcal{S}^p_-}(\Gamma + g(x)) \end{bmatrix} = 0,$$

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. [Robinson'80] Let $\bar{z}$ be a solution of the generalized equation. We say that $\bar{z}$ is a strongly regular solution if there exist neighborhoods $\mathcal{B}$ of the origin $0 \in Z$ and $\mathcal{V}$ of $\bar{z}$ 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 *$\mathcal{C}^2$-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)

$$\begin{array}{ll} \min\limits_{x \in X} & f(x,u) \\ \text{s.t.} & G(x,u) \in K \,. \end{array}$$

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. [Bonnans and Shapiro'00] Let $\bar{x}$ be a stationary point of (OP). We say that the uniform second order (quadratic) growth condition holds at $\bar{x}$ with respect to a $\mathcal{C}^2$-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:

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

We say that the uniform second order growth condition holds at $\bar{x}$ if the above inequality holds for every $\mathcal{C}^2$-smooth parameterization of (OP).

Definition. [Kojima[a] and Bonnans and Shapiro'00]

Let $\bar{x}$ be a stationary point of (OP). We say that $\bar{x}$ is strongly stable with respect to a $\mathcal{C}^2$-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 $\mathcal{C}^2$-smooth parameterization, we say that $\bar{x}$ is strongly stable.

[a] M. 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 $\mathrm{ind}(\phi, \bar{z})$ denote the index of a continuous function $\phi : Z \to Z$ at an isolated zero $\bar{z} \in Z$ 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

[a] M.S. Gowda. Inverse and implicit function theorems for H-differentiable and semismooth functions. *Optimization Methods and Software* 19 (2004) 443-461.

[b] J.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 $(\bar{\zeta}, \overline{\Gamma}) \in \Re^m \times \mathcal{S}^p$ 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.

[a] D. 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 $V \in \partial_B F(\bar{x}, \bar{\zeta}, \overline{\Gamma})$, $\operatorname{sgn} \det V = \operatorname{ind}(F, (\bar{x}, \bar{\zeta}, \overline{\Gamma})) = \pm 1$.

(h) Φ is a globally Lipschitz homeomorphism.

(i) For every $V \in \partial_B \Phi(0)$, $\operatorname{sgn} \det V = \operatorname{ind}(\Phi, 0) = \pm 1$.

(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).

For an application of Theorem 2, let us look at the augmented Lagrangian method for solving (NLSDP):

For each $c > 0$, the augmented Lagrangian function for (NLSDP) is:

$$\begin{aligned} L_c(x, \zeta, \Xi) : \; &= \; f(x) + \langle \zeta, h(x) \rangle + \frac{c}{2} \|h(x)\|^2 \\ & \quad + \frac{1}{2c} \left[\|\Pi_{\mathcal{S}^p_+}(\Xi - c g(x))\|^2 - \|\Xi\|^2 \right], \end{aligned}$$

where $(x, \zeta, \Xi) \in X \times \Re^m \times Y$.

Let $c_0 > 0$ be given. Let $(\zeta^0, \Xi^0) \in \Re^m \times \mathcal{S}^p_+$ be the initial estimated Lagrange multiplier. At the kth iteration, determine x^k by minimizing $L_{c_k}(x, \zeta^k, \Xi^k)$, compute (ζ^{k+1}, Ξ^{k+1}) by

$$\begin{cases} \zeta^{k+1} := \zeta^k + c_k h(x^k)\,, \\ \Xi^{k+1} := \Pi_{\mathcal{S}^p_+}(\Xi^k - c_k g(x^k))\,, \end{cases}$$

and update c_{k+1} by

$$c_{k+1} := c_k \quad \text{or} \quad c_{k+1} := \kappa c_k$$

according to certain rules, where $\kappa > 1$ is a preselected positive number.

Theorem 3[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 $(NLSDP)$. Suppose that one of (a)-(j) in Theorem 2 holds.

Then we can find positive numbers $\overline{c}$, ϱ_1, and ϱ_2 such that for any $c \geq \overline{c}$, there exist two positive numbers ε and δ (may depend on c) such that for any $(\zeta, \Xi) \in \mathbb{B}_\delta(\overline{\zeta}, \overline{\Xi})$, the problem

$$\min \ L_c(x, \zeta, \Xi) \quad \text{s.t. } x \in \mathbb{B}_\varepsilon(\overline{x})$$

[a]D. Sun, J. Sun, and Liwei Zhang. The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming. Manuscript, Department of Mathematics, National University of Singapore, January 2006.

(continued)

has a unique solution denoted $x_c(\zeta, \Xi)$. The function $x_c(\cdot, \cdot)$ is locally Lipschitz continuous on $\mathbb{B}_\delta(\overline{\zeta}, \overline{\Xi})$ and is semismooth at any point in $\mathbb{B}_\delta(\overline{\zeta}, \overline{\Xi})$, and for any $(\zeta, \Xi) \in \mathbb{B}_\delta(\overline{\zeta}, \overline{\Xi})$, we have

$$\|x_c(\zeta, \Xi) - \overline{x}\| \leq \varrho_1 \|(\zeta, \Xi) - (\overline{\zeta}, \overline{\Xi})\|/c$$

and

$$\|(\zeta_c(\zeta, \Xi), \Xi_c(\zeta, \Xi)) - (\overline{\zeta}, \overline{\Xi})\| \leq \varrho_2 \|(\zeta, \Xi) - (\overline{\zeta}, \overline{\Xi})\|/c\,,$$

where $\zeta_c(\zeta, \Xi)$ and $\xi_c(\zeta, \Xi))$ are defined as

$$\zeta_c(\zeta, \Xi) := \zeta + ch(x_c(\zeta, \Xi)) \quad \text{and} \quad \Xi_c(\zeta, \Xi) := \Pi_{\mathcal{S}^p_+}(\xi - cg(x_c(\zeta, \Xi)))\,.$$

Note that Theorem 3 solved the local convergence and rate of convergence of the augmented Lagrangian function method for (NLSDP).

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?