# Modern Optimization Theory:

## Optimality Conditions and Perturbation Analysis

Defeng Sun

Department of Mathematics

National University of Singapore

Republic of Singapore

July 5-10, 2006

# 1 Robinson's Constraint Qualification and Optimality Conditions

Let us first consider the following simple one-dimensional optimization problem

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

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

$$\nabla_x L(x^*, \lambda^*) = x^* + \lambda^* = 0\,, \quad 0 \le x^* \perp \lambda^* \ge 0\,.$$

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 forms:

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

$$\Updownarrow$$

$(SOC)$

$$\begin{array}{ll} \min\limits_{(t,x)\in\Re^2} & t \\ \text{s.t.} & x \leq 0\,, \\ & \|(2x, 2-t)\|_2 \leq 2+t \iff (2+t, 2x, 2-t) \in \mathcal{K}^3\,, \end{array}$$

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

$(OP)$

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

where $f : X \to \Re$ and $G : X \to Y$ are $\mathcal{C}^1$ (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 $x = 0$.

The problem (OP) is very general and includes

**(i)** Linear programming (LP): when $f$ is a linear functional, $g$ is affine, and $K$ is a polyhedral convex cone.

**(ii)** Nonlinear programming (NLP): when $f$ or $g$ is nonlinear and $K$ is a polyhedral convex cone.

**(iii)** Linear conic programming: when $f$ is a linear functional, $g$ is affine, and $K$ is a closed (non-polyhedral) convex cone.

**(iv)** Nonlinear conic programming: when $f$ or $g$ is nonlinear and $K$ is a closed (non-polyhedral) convex cone.

In particular, it includes the nonlinear semidefinite programming

$(NLSDP)$

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

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.

Difficulty:

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

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

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

or nonsmooth optimization problem

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

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.

## Some notation:

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[a] 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 $x$).

[a] A function $\Psi : X \rightarrow Y$ is said to be Fréchet-differentiable at $x \in X$ if there exists a linear operator, denoted by $\mathcal{J}\Psi(x)$, such that

$$\Psi(x + \Delta x) - \Psi(x) - \mathcal{J}\Psi(x)(\Delta x) = o(\|\Delta x\|).$$

For example, if $\Psi(x) = Ax + xA^T$, where $x \in \mathcal{S}^p$ and $A \in \Re^{p\times p}$, then

$$\mathcal{J}\Psi(x)(\Delta x) = A\Delta x + \Delta x A^T \quad \forall\, \Delta x \in \mathcal{S}^p.$$

Another example is $\Psi(x) = x^2$, $x \in \mathcal{S}^p$. By the definition, one can check directly

$$\mathcal{J}\Psi(x)(\Delta x) = x(\Delta x) + (\Delta x)x \quad \forall\, \Delta x \in \mathcal{S}^{\cdot}$$

Let

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

be the adjoint[a] of $\mathcal{J}F(x, x')$ (respectively, $\nabla_x F(x, x') := \mathcal{J}_x F(x, x')^*$, the adjoint of $\mathcal{J}_x F(x, x')$).

[a] For a linear operator $A : X \to Y$, its adjoint is the unique linear operator mapping $Y$ into $X$, denoted by $A^*$, satisfies

$$\langle y, Ax \rangle_Y = \langle A^* y, x \rangle_X \quad \forall x \in X \text{ and } y \in Y.$$

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

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

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

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

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

## Some definitions.

**Definition 1.1** *The following two sets are called the upper and lower limits of a parameterized family $A_t$, of subsets of $Y$:*

$$\limsup_{t\to t_0} A_t := \{y\in Y : \exists\, t_n\to t_0 \text{ such that } y_n\to y \text{ for some } y_n\in A_{t_n}\}$$

*and*

$$\liminf_{t\to t_0} A_t := \{y\in Y : \text{for every } t_n\to t_0\ \exists\ y_n\in A_{t_n} \text{ such that } y_n\to y\}.$$

**Definition 1.2** *For any closed set $D \subseteq Y$ and a point $y \in D$, we define the radial cone*

$$\mathcal{R}_D(y) := \{d \in Y \,:\, \exists\, t^* > 0 \text{ such that } y + td \in D \;\forall\, t \in [0, t^*]\};$$

*the inner tangent cone*

$$\mathcal{T}_D^i(y) := \liminf_{t \downarrow 0} \frac{D - y}{t};$$

*the contingent (Bouligand) cone*

$$\mathcal{T}_D(y) := \limsup_{t \downarrow 0} \frac{D - y}{t};$$

*and the Clarke tangent cone*

$$\mathcal{T}_D^c(y) := \liminf_{\substack{t \downarrow 0 \\ D \ni y' \to y}} \frac{D - y'}{t}.$$

Obviously, we have

$$\mathcal{R}_D(y) \subset \mathcal{T}_D^i(y) \subset \mathcal{T}_D(y)\,.$$

The contingent, inner, and Clarke tangent cones are closed while the radial cone is not closed.

From Definitions 1.1 and 1.2 we have the following equivalent forms for $\mathcal{T}_D^i(y)$ and $\mathcal{T}_D(y)$:

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

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

where for each $w \in Y$, $\mathrm{dist}(w, D) := \inf\{\|w - d\| \,:\, d \in D\}$.

**Proposition 1.1** *If $D$ is a closed convex set and $y \in D$, then*

$$\mathcal{R}_D(y) = \bigcup_{t>0} \{t^{-1}(D - y)\}$$

*and*

$$\mathcal{T}_D(y) = \mathcal{T}_D^i(y) = \mathcal{T}_D^c(y) = \mathrm{cl}[\mathcal{R}_D(y)]\,,$$

*where "cl" denotes the topological closure.*

Therefore, 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

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

Let $Z$ be another Hilbert space and $D$ 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 globally Lipschitz continuous with modulus 1.

Next, we demonstrate how to compute $\mathcal{T}_{\mathcal{S}^p_+}(\cdot)$.

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

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

where "Tr" denotes the trace of a square matrix (i.e., the sum of all diagonal elements of the symmetric matrix). Let $A$ have the following spectral decomposition

$$\boxed{A = P\Lambda P^T,}$$

where $\Lambda$ 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 (more about this part in the next talk):

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

where $\Pi_{\mathcal{S}^p_+}(A)$ is the metric projector of $A$ onto $\mathcal{S}^p_+$ under the above trace inner product.

Note that computing $A_+$ is equivalent to computing the full eigen-decomposition of $A$, which in turn needs $9n^3$ flops. For a typical Pentium IV type desktop PC, it needs about 10 seconds for $n = 1,000$ and less than 90 seconds for $n = 2,000$.

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

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

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

$\Pi_{\mathcal{S}^p_+}$ is directionally differentiable, i.e., there exists a positive homogeneous function denoted by $\Pi'_{\mathcal{S}^p_+}(A;\cdot)$, such that for any $H \in \mathcal{S}^p$,

$$\Pi_{\mathcal{S}^p_+}(A+tH) - \Pi_{\mathcal{S}^p_+}(A) - t\Pi'_{\mathcal{S}^p_+}(A;H) = o(t) \quad \forall\, t \downarrow 0\,,$$

with

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

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].

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

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

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

---

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

[b] 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:

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

and

$$\Pi'_{\mathcal{S}^p_+}(A_+; B) = P \begin{bmatrix} P_\alpha^T B P_\alpha & P_\alpha^T H P_{\bar{\alpha}} \\ P_{\bar{\alpha}}^T H P_\alpha & \Pi_{\mathcal{S}^{|\bar{\alpha}|}_+}(P_{\bar{\alpha}}^T H P_{\bar{\alpha}}) \end{bmatrix} P^T.$$

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

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

Now, let us define Robinson's constraint qualification (CQ).

**Definition 1.3** *Let $\bar{x}$ be a feasible solution to (OP). Robinson's constraint qualification is as follows:*

$$0 \in \operatorname{int}\{G(\bar{x}) + \mathcal{J}G(\bar{x})X - K\}, \tag{1}$$

*where "int" denotes the topological interior.*

**Proposition 1.2** *Suppose that $G(\bar{x}) \in K$. Then Robinson's CQ (1) is equivalent to*

$$\mathcal{J}G(\bar{x})X + \mathcal{T}_K(G(\bar{x})) = Y\,. \tag{2}$$

**Proposition 1.3** *If $Y$ is the Cartesian product of $Y_1$ and $Y_2$, and $K = K_1 \times K_2 \subset Y_1 \times Y_2$, where $K_1$ and $K_2$ are closed convex subsets of $Y_1$ and $Y_2$, respectively. Let $G(x) = (G_1(x), G_2(x)) \in Y_1 \times Y_2$. Assume that $G(\bar{x}) \in K$. Suppose that $\mathcal{J}G_1(\bar{x})$ is onto and that $K_2$ has a nonempty interior. Then Robinson's CQ (1) is equivalent to the existence of $d \in X$ such that*

$$\left\{ \begin{array}{l} G_1(\bar{x}) + \mathcal{J}G_1(\bar{x})d \in K_1 \, . \\ \\ G_2(\bar{x}) + \mathcal{J}G_2(\bar{x})d \in \mathrm{int}(K_2) \, . \end{array} \right. \qquad (3)$$

[If $K_1 = \{0\}$, the first relation in (3) becomes an equation.]

In particular, for conventional nonlinear programming

$(NLP)$

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

Robinson's CQ reduces to the well-known Mangasarian-Fromovitz constraint qualification (MFCQ):

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

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

**Proposition 1.4** *Let $\bar{x}$ be a feasible solution to (OP) and* $\underline{\Phi := \{x \in X \,:\, G(x) \in K\} = G^{-1}(K)}$. *Then we have*

**(i)** *The point $d = 0$ is an optimal solution to*

$$\begin{array}{ll} \min\limits_{x \in X} & \mathcal{J}f(\bar{x})d \\ \text{s.t.} & d \in \mathcal{T}_{\Phi}(\bar{x}) \,. \end{array} \tag{4}$$

**(ii)** *If Robinson's CQ (1) holds, then $d = 0$ is an optimal solution to the linearized problem*

$$\begin{array}{ll} \min\limits_{x \in X} & \mathcal{J}f(\bar{x})d \\ \text{s.t.} & \mathcal{J}G(\bar{x})d \in \mathcal{T}_{K}(G(\bar{x})) \,. \end{array} \tag{5}$$

**Proof.** (i) Let $d \in \mathcal{T}_\Phi(\bar{x})$. Then there exist sequences $t_n \downarrow 0$ and $x_n = \bar{x} + t_n d + o(t_n)$ such that $x_n \in \Phi$. Since $\bar{x}$ is a local solution to (OP), we obtain

$$0 \leq \lim_{n\to\infty} \frac{f(x_n) - f(\bar{x})}{t_n} = \mathcal{J} f(\bar{x})d\,.$$

Thus, $d = 0$ is an optimal solution to (4).

(ii) Since Robinson's CQ holds, the inner and outer tangent sets to $\Phi$ at $\bar{x}$ coincide and are the same as the feasible solution set of (5). Then (ii) follows from (i). $\square$

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

$$L(x,\mu) := f(x) + \langle \mu, G(x) \rangle\,, \quad (x,\mu) \in X \times Y. \tag{6}$$

We say that $\bar{\mu} \in Y$ is a Lagrangian multiplier of (OP) at $\bar{x}$ if it, together with $\bar{x}$, satisfies the Karush-Kuhn-Tucker (KKT) condition:

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

If $K$ is a closed convex cone, then the KKT system is equivalent to

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

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

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

For example, if $K = \{0\}^m \times \mathcal{S}^p_+$, then

$$K^* = \Re^m \times \mathcal{S}^p_+\,.$$

Let $\mathcal{M}(\bar{x})$, possibly empty set, denote the set of Lagrangian multipliers of (OP) at $\bar{x}$. We call $\bar{x}$ a stationary point of (OP) if $\mathcal{M}(\bar{x}) \neq \emptyset$.

**Theorem 1.1** *Let $\bar{x}$ be a locally optimal solution to (OP). Suppose that Robinson's CQ (1) holds. Then $\mathcal{M}(\bar{x})$ is a nonempty and bounded convex set.*

**Proof.** The proof is based on Proposition 1.4 and some duality theory for the linearized problem (5). We omit the details here. $\square$

Note that the converse part of Theorem 1.1 is also true, i.e., if $\mathcal{M}(\bar{x})$ is nonempty and bounded, then Robinson's CQ (1) holds.

Let $\bar{x}$ be a feasible solution to (OP). The critical cone of (OP) at $\bar{x}$ is defined as

$$\boxed{C(\bar{x}) := \{d \in X \,:\, \mathcal{J}G(\bar{x})d \in \mathcal{T}_K(G(\bar{x})),\ \mathcal{J}f(\bar{x})d \leq 0\}.}$$

The cone $C(\bar{x})$ consists of directions for which the linearized problem (5) does not provide any information about the optimality of $\bar{x}$, and will be useful in the study of second-order optimality conditions.

**Proposition 1.5** *Let $\bar{x}$ be a feasible solution to (OP). If $\mathcal{M}(\bar{x}) \neq \emptyset$, then $d = 0$ is an optimal solution to the linearized problem (5) and*

$$C(\bar{x}) := \{d \in X \, : \, \mathcal{J}G(\bar{x})d \in \mathcal{T}_K(G(\bar{x})), \ \mathcal{J}f(\bar{x})d = 0\}.$$

*Moreover, for any $\mu \in \mathcal{M}(\bar{x})$,*

$$C(\bar{x}) := \{d \in X \, : \, \mathcal{J}G(\bar{x})d \in \mathcal{T}_K(G(\bar{x}), \ \langle \mu, \mathcal{J}G(\bar{x})d \rangle = 0\}.$$

Note that for any $\mu \in \mathcal{M}(\bar{x})$,

$$0 = \langle \nabla_x L(\bar{x}, \mu), d \rangle = \mathcal{J}f(\bar{x})d + \langle \mu, \mathcal{J}G(\bar{x})d \rangle \, .$$

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). This is also our major reference book on this part.

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 $K := \{0\} \times \mathcal{S}_+^p \subset Y := \Re^m \times \mathcal{S}^p$,

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

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

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

**Theorem 1.2** *(Second-Order Necessary Condition.)*

*Suppose that $f$ and $G$ are twice continuously differentiable. Let $\bar{x}$ be a local optimal solution to (OP). Suppose that Robinson's CQ (1) holds. Then for every $d \in C(\bar{x})$ and any convex set $T(d) \subset \mathcal{T}^2_K(G(\bar{x}, \mathcal{J}G(\bar{x})d)$, the following inequality holds*

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

Before we state the second order sufficient conditions for (OP), we need below the concept of $\mathcal{C}^2$-cone reducibility.

**Definition 1.4** *A closed (not necessarily convex) set $D \subseteq Y$ is called $\mathcal{C}^2$-cone reducible at a point $\bar{y} \in D$ if there exist a neighborhood $\mathcal{V} \subseteq Y$ of $\bar{y}$, a pointed closed convex cone $Q$ (a cone is said to be pointed if and only its lineality space is the origin) in a finite dimensional space $Z$ and a twice continuously differentiable mapping $\Xi : \mathcal{V} \to Z$ such that:*

*(i) $\Xi(\bar{y}) = 0 \in Z$,*

*(ii) the derivative mapping $\mathcal{J}\Xi(\bar{y}) : Y \to Z$ is onto, and*

*(iii) $\underline{D \cap \mathcal{V} = \{y \in \mathcal{V} \mid \Xi(y) \in Q\}}$. We say that $D$ is $\mathcal{C}^2$-cone reducible if $D$ is $\mathcal{C}^2$-cone reducible at every point $\bar{y} \in Y$ (possibly to a different pointed cone $Q$).*

Many interesting sets such as

the polyhedral convex set,

the second-order cone, and

the cone $S^p_+$

are all $\mathcal{C}^2$-cone reducible, and

the Cartesian product of $\mathcal{C}^2$-cone reducible sets is again $\mathcal{C}^2$-cone reducible

In particular, $K = \{0\} \times \mathcal{S}^p_+$ is $\mathcal{C}^2$-cone reducible.

**Theorem 1.3** (*Second-Order Sufficient Condition.*)

*Suppose that $f$ and $G$ are twice continuously differentiable. Let $\bar{x}$ be a stationary point to (OP). Suppose that Robinson's CQ (1) holds and that the set $K$ is $\mathcal{C}^2$-cone reducible at $\bar{y} := G(\bar{x})$. Then the following condition*

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

*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 $X$.*

By combining Theorems 1.2-1.3 and the $\mathcal{C}^2$-cone reducibility of polyhedral convex sets and $\mathcal{S}_+^p$, we can now state the "no-gap" second order necessary condition and the second order sufficient condition for $(NLSDP)$.

**Theorem 1.4** *(Second-Order Necessary and Sufficient Conditions for (NLSDP).)*

*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*

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

*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

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

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 $d \in C(\bar{x})$,

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

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.

**Example**. 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

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

Exercises.

1. Let $\Psi(x) = Ax^2A^T$, where $x \in \mathcal{S}^p$ and $A \in \Re^{p\times p}$. Compute $\mathcal{J}\Psi(x)$ and $\mathcal{J}^2\Psi(x)$.

2. Prove the converse part of Theorem 1.1.

3. Show that all polyhedral convex sets and second-order-cones are $\mathcal{C}^2$-cone reducible.

For details on topics discussed here, see the following excellent monograph

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