High-Order Optimality Conditions II

Yinyu Ye
Department of Management Science and Engineering
Stanford University
Stanford, CA 94305, U.S.A.

http://www.stanford.edu/~yyye

General Nonlinearly-Constrained Optimization

$$
\begin{array}{lll}
(NCO) & \min & f(\mathbf{x}) \\
 & \text{s.t.} & \mathbf{h}(\mathbf{x}) = \mathbf{0} \in R^m, \\
 & & \mathbf{c}(\mathbf{x}) \le \mathbf{0} \in R^p.
\end{array}
$$

We dealt the case when the feasible region is convex and the objective is convex, then we handled the case when the objective is a general C^1 function.

We now study the case that the only assumption is that all functions are in C^1, either convex or noncovex.

We again establish optimality conditions to qualify local optimizers. These conditions give us qualitative structures of (local) optimizers and lead to quantitative algorithms to find a numerical optimizer.

Hypersurface and Implicit Function Theorem

Consider the hypersurface:

$$\{\mathbf{x} \in R^n : \mathbf{h}(\mathbf{x}) = \mathbf{0} \in R^m,\ m \le n\}$$

When functions $h_i(\mathbf{x})$s are C^1 functions, we say the surfacesmooth.

For a point $\mathbf{x}^*$ on the surface, we call it a regular point if $\nabla \mathbf{h}(\mathbf{x}^*)$ have rank m or the rows are linearly independent. For example, $(0;\ 0)$ is not a regular point of $\{(x_1;\ x_2) \in R^2 :\ x_1^2 + (x_2-1)^2 - 1 = 0,\ x_1^2 + (x_2+1)^2 - 1 = 0\}$.

Based on the Implicit Function Theorem, if $\mathbf{x}^*$ is a regular point and $m < n$, then for every $\mathbf{d} \in \mathcal{T}^* = \{\mathbf{z} :\ \nabla \mathbf{h}(\mathbf{x}^*)\mathbf{z} = \mathbf{0}\}$ there exists a curve $\mathbf{x}(t)$ on the hypersurface, parametrized by a scalar t in a sufficiently small interval $[-a \quad a]$, such that

$$\mathbf{h}(\mathbf{x}(t)) = \mathbf{0}, \quad \mathbf{x}(0) = \mathbf{x}^*, \quad \dot{\mathbf{x}}(0) = \mathbf{d}.$$

$$\min \quad (x_1)^2 + (x_2)^2 \quad \text{s.t.} \quad (x_1)^2/4 + (x_2)^2 - 1 = 0$$

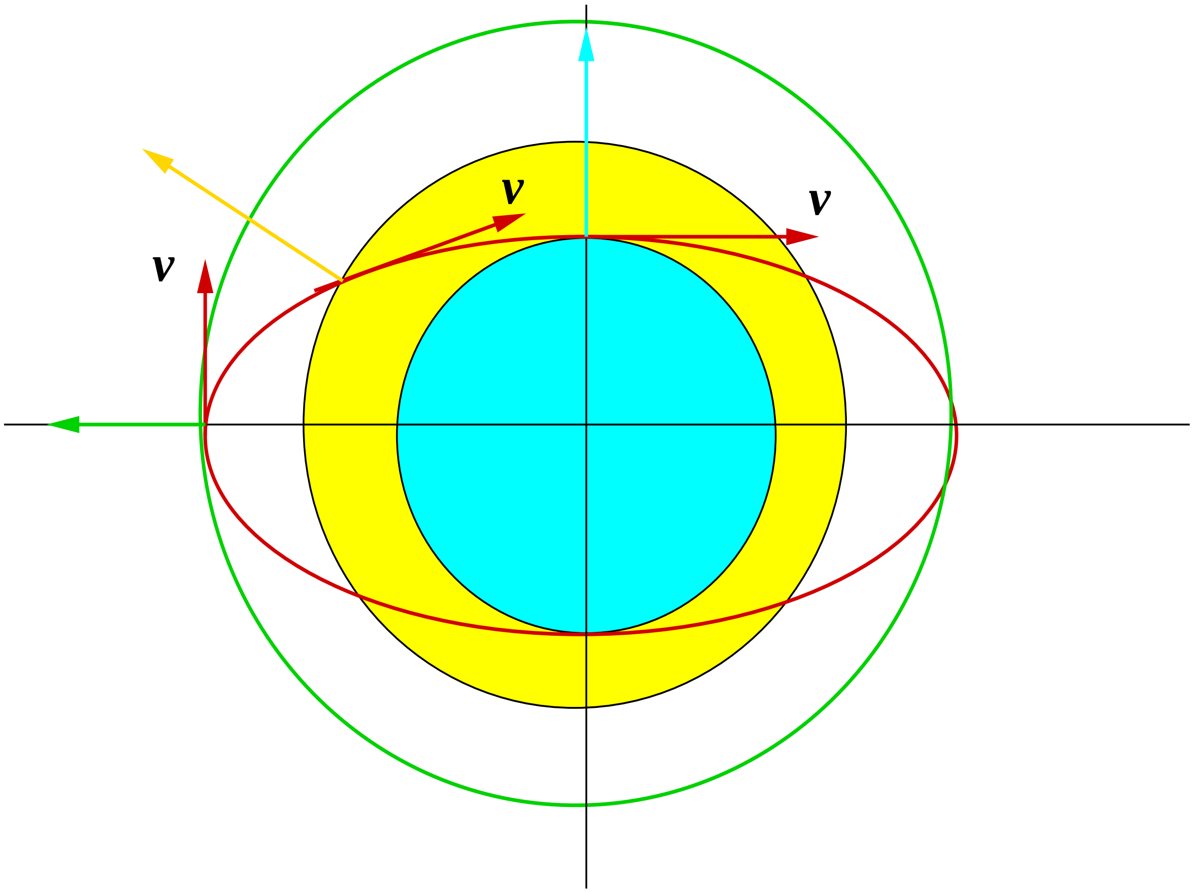

Figure 1: FONC for Nonlinear Equality Constrained Minimization

First-Order Condition for Nonlinearly-Constrained Optimization I

Lemma 1 *Let* $\mathbf{x}^*$ *be a regular point of the hypersurface of*

$$\{\mathbf{x} :\ \mathbf{h}(\mathbf{x}) = \mathbf{0},\ c_i(\mathbf{x}) = 0, i \in \mathcal{C}^*\}$$

where active-constraint set $\mathcal{C}^* = \{i :\ c_i(\mathbf{x}^*) = 0\}$*. If* $\mathbf{x}^*$ *is a (local) minimizer of (NCO), then* $\mathbf{d} = \mathbf{0}$ *must be a minimizer of the following linear program:*

$$\begin{array}{ll} \min & \nabla f(\mathbf{x}^*)\mathbf{d} \\ \textit{s.t.} & \nabla \mathbf{h}(\mathbf{x}^*)\mathbf{d} = \mathbf{0} \in R^m, \\ & \nabla c_i(\mathbf{x}^*)\mathbf{d} \le 0,\ i \in \mathcal{C}^*. \end{array}$$

Proof

First, we notice that $\mathbf{d} = \mathbf{0}$ is a feasible solution for the linear program. Thus, if $\mathbf{0}$ is not a minimizer, we must have a $\bar{\mathbf{d}}$ such that $\nabla f(\mathbf{x}^*)\bar{\mathbf{d}} < 0$, that is, $\bar{\mathbf{d}}$ is a descent-direction vector for the objective function.

Denote the active-constraint set at $\bar{\mathbf{d}}$ of the linear program by $\mathcal{C}'$ $(\subset \mathcal{C}^*)$. Then, $\mathbf{x}^*$ remains a regular point of hypersurface of

$$\{\mathbf{x} : \ \mathbf{h}(\mathbf{x}) = \mathbf{0}, \ c_i(\mathbf{x}) = 0, i \in \mathcal{C}'\}.$$

Thus, as it was showed earlier, there is a curve $\mathbf{x}(t)$ such that

$$\mathbf{h}(\mathbf{x}(t)) = \mathbf{0}, \quad c_i(\mathbf{x}(t)) = 0, \ i \in \mathcal{C}', \quad \mathbf{x}(0) = \mathbf{x}^*, \quad \dot{\mathbf{x}}(0) = \bar{\mathbf{d}},$$

for $t \in [-a \quad a]$ of a sufficiently small positive constant a.

Also, $\nabla c_i(\mathbf{x}^*)\bar{\mathbf{d}} < 0, \ \forall i \notin \mathcal{C}'$ and $c_i(\mathbf{x}^*) < 0, \ \forall i \notin \mathcal{C}^*$. From Taylor's theorem, $c_i(\mathbf{x}(t)) < 0$ for all $i \notin \mathcal{C}'$ so that $\mathbf{x}(t)$ is a feasible curve to the original (NCO) problem for $t \in [-a \quad a]$. Thus, $\mathbf{x}^*$ must be a local minimizer of the curve $\mathbf{x}(t)$.

Let $\phi(t) = f(\mathbf{x}(t))$. Then, $t = 0$ must be a local minimizer of $\phi(t)$ for $-a \le t \le a$ so that

$$0 = \phi'(0) = \nabla f(\mathbf{x}(0))\dot{\mathbf{x}}(0) = \nabla f(\mathbf{x}^*)\bar{\mathbf{d}}$$

which results in a contradiction.

First-Order Condition for Nonlinearly-Constrained Optimization II

Therefore, the first-order or duality condition applies to the linear program, which leads to

Theorem 1 *(First-Order or KKT Optimality Condition) Let* $\mathbf{x}^*$ *be a (local) minimizer of (NCO) and it is a regular point of* $\{\mathbf{x} : \ \mathbf{h}(\mathbf{x}) = \mathbf{0}, \ c_i(\mathbf{x}) = 0, i \in \mathcal{C}^*\}$*. Then,*

$$\nabla_x L(\mathbf{x}^*, \mathbf{y}^*, \mathbf{s}^*) = \mathbf{0}$$

and

$$s_i^* c_i(\mathbf{x}^*) = 0, \ \forall i$$

for some multipliers $(\mathbf{y}^*, \mathbf{s}^* \geq \mathbf{0})$*.*

The first equality condition and $\mathbf{s}^* \geq \mathbf{0}$ is based on the linear programming duality where the pair of $(\mathbf{y}^*, \mathbf{s}^*)$ is a dual optimizer pair. The complementarity condition is from that $c_i(\mathbf{x}^*) = 0$ for all $i \in \mathcal{A}^*$, and for $i \notin \mathcal{A}^*$, we simply set $s_i^* = 0$.

Second-Order Necessary Conditions for NCO

Now in addition we assume all functions are twice continuously differentiable.

Again let

$$T^* := \{\mathbf{z} : \nabla \mathbf{h}(\mathbf{x}^*)\mathbf{z} = \mathbf{0},\ \nabla c_i(\mathbf{x}^*)\mathbf{z} = 0 \forall i \in \mathcal{C}^*\}.$$

T^* is sometimes called the tangent linear space of the active constraints at $\mathbf{x}^*$.

Theorem 2 *Let $\mathbf{x}^*$ be a (local) minimizer of (NCO) and a regular point of $\{\mathbf{x} :\ \mathbf{h}(\mathbf{x}) = \mathbf{0},\ c_i(\mathbf{x}) = 0, i \in \mathcal{C}^*\}$, and let $\mathbf{y}^*, \mathbf{s}^*$ denote Lagrange multipliers such that $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{s}^*)$ satisfies the (first-order) KKT conditions of (NCO). Then, it is necessary to have*

$$\mathbf{z}^T \nabla^2_{\mathbf{x}} L(\mathbf{x}^*, \mathbf{y}^*, \mathbf{s}^*)\mathbf{z} \geq 0 \qquad \forall \mathbf{z} \in T^*.$$

Sketch of Proof

The proof is based on that fact that $\mathbf{x}^*$ is a local minimizer of (NCO) and a regular point, so that it is a local minimizer of any twice differentiable curve $\mathbf{x}(t)$ in the feasible region passing through $\mathbf{x}^* = \mathbf{x}(0)$ with $\dot{\mathbf{x}}(0) = \mathbf{z},\ \mathbf{z} \in T^*$. Thus, the Hessian of the Lagrangian function need to be positive semidefinite on the tangent space, since the second derivative of $f(\mathbf{x}(t))$ equals $\mathbf{z}^T \nabla^2_{\mathbf{x}} L(\mathbf{x}^*, \mathbf{y}^*, \mathbf{s}^*)\mathbf{z}$ for $\mathbf{z} \in T^*$.

Second-Order Sufficient Conditions for NCO

Theorem 3 *Let* $\mathbf{x}^*$ *be a regular point of (NCO) and let* $\mathbf{y}^*, \mathbf{s}^*$ *be the Lagrange multipliers such that* $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{s}^*)$ *satisfies the (first-order) KKT conditions of (NCO). Then, if in addition*

$$\mathbf{z}^T \nabla^2_{\mathbf{x}} L(\mathbf{x}^*, \mathbf{y}^*, \mathbf{s}^*)\mathbf{z} > 0 \qquad \forall \mathbf{0} \neq \mathbf{z} \in T^*,$$

then $\mathbf{x}^*$ *is a local minimizer of (NCO).*

The proof can be found in Chapter 11.8 of LY.

$$\min \quad (x_1)^2 + (x_2)^2 \quad \text{s.t.} \quad -(x_1)^2/4 - (x_2)^2 + 1 \leq 0$$

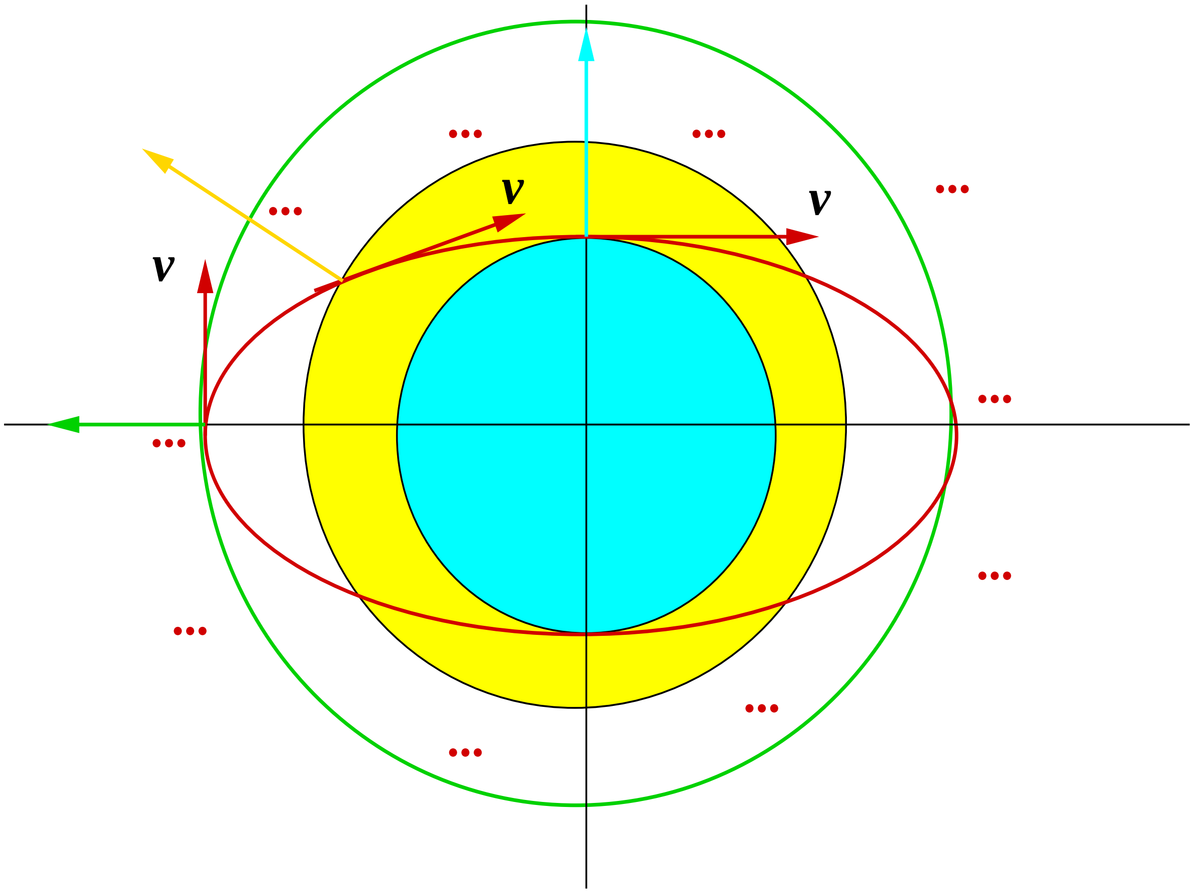

Figure 2: FONC for Nonlinear Inequality Constrained Minimization

$$L(x_1, x_2, y) = (x_1)^2 + (x_2)^2 + y(-(x_1)^2/4 - (x_2)^2 + 1),$$

$$\nabla_x L(x_1, x_2, y) = (2x_1(1 - y/4),\ 2x_2(1 - y)),$$

$$\nabla_x^2 L(x_1, x_2, y) = \begin{pmatrix} 2(1 - y/4) & 0 \\ 0 & 2(1 - y) \end{pmatrix}$$

$$T(\mathbf{x}) := \{(z_1,\ z_2) :\ (x_1/4)z_1 + x_2 z_2 = 0\}.$$

We see that there are two possible values for y: either 4 or 1, which lead to total four KKT points:

$$\begin{pmatrix} x_1 \\ x_2 \\ y \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 4 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \\ 4 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \text{ and } \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}.$$

Consider the first KKT point:

$$\nabla_x^2 L(2,0,4) = \begin{pmatrix} 0 & 0 \\ 0 & -6 \end{pmatrix}, \; T^* = \{(z_1,\ z_2) :\ z_1 = 0\}$$

Then the Hessian is not positive semidefinite on T^* since

$$\mathbf{z}^T \nabla_x^2 L(2,0,4)\mathbf{z} = -6z_2^2 \le 0.$$

Consider the third KKT point:

$$\nabla_x^2 L(0,1,1) = \begin{pmatrix} 3/2 & 0 \\ 0 & 0 \end{pmatrix}, \; T^* = \{(z_1,\ z_2) :\ z_2 = 0\}$$

Then the Hessian is positive definite on T^* since

$$\mathbf{z}^T \nabla_x^2 L(0,0,1)\mathbf{z} = (3/2)z_1^2 > 0,\ \forall \mathbf{0} \ne \mathbf{z} \in T^*.$$

Test positive semidefiniteness in a subspace

In the second-order test, we typically like to know whether or not

$$\mathbf{z}^T Q \mathbf{z} \geq 0, \ \forall \mathbf{z}, \ \text{s.t. } A\mathbf{z} = \mathbf{0}$$

for a given symmetric matrix Q and a rectangle matrix A. (In this case, the subspace is the null space of matrix A.) This test itself might be a nonconvex optimization problem.

But it is known that $\mathbf{z}$ is in the null space of matrix A if and only if $\mathbf{z} = (I - A^T(AA^T)^{-1}A)\mathbf{u} = P_A\mathbf{u}$ for some vector $\mathbf{u} \in R^n$, where P_A is called the projection matrix of A. Thus, the test becomes whether or not

$$\mathbf{u}^T P_A Q P_A \mathbf{u} \geq 0, \ \forall \mathbf{u} \in R^n,$$

that is, we just need to test positive semidefiniteness of $P_A Q P_a$ as usual.