Optimality Conditions for General Constrained Optimization

Yinyu Ye

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

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

Chapter 11.1-8
Figure 1: Global and Local Minimizers of One-Variable Function in Interval $[a \ e]$
A differentiable function \( f \) of one variable defined on an interval \( F = [a, e] \). If an interior-point \( \bar{x} \) is a local/global minimizer, then \( f'(\bar{x}) = 0 \); if the left-end-point \( \bar{x} = a \) is a local minimizer, then \( f'(a) \geq 0 \); if the right-end-point \( \bar{x} = e \) is a local minimizer, then \( f'(e) \leq 0 \). first-order necessary condition (FONC) summarizes the three cases by a unified set of optimality/complementarity conditions:

\[
a \leq x \leq e, \quad f'(x) = y^a + y^e, \quad y^a \geq 0, \quad y^e \leq 0, \quad y^a(x - a) = 0, \quad y^e(x - e) = 0.
\]

If \( f'(\bar{x}) = 0 \), then it is also necessary that \( f(x) \) is locally convex at \( \bar{x} \) for it being a local minimizer.

How to tell the function is locally convex at the solution? It is necessary \( f''(\bar{x}) \geq 0 \), which is called the second-order necessary condition (SONC), which we would explored further.

These conditions are still not, in general, sufficient. It does not distinguish between local minimizers, local maximizers, or saddle points.

If the second-order sufficient condition (SOSC): \( f''(\bar{x}) > 0 \), is satisfied or the function is strictly locally convex, then \( \bar{x} \) is a local minimizer.

Thus, if the function is \textit{convex} everywhere, the first-order necessary condition is already \textit{sufficient}. 
Second-Order Optimality Condition for Unconstrained Optimization

**Theorem 1** *(First-Order Necessary Condition)* Let \( f(\mathbf{x}) \) be a \( C^1 \) function where \( \mathbf{x} \in \mathbb{R}^n \). Then, if \( \mathbf{x} \) is a minimizer, it is necessarily \( \nabla f(\mathbf{x}) = \mathbf{0} \).

**Theorem 2** *(Second-Order Necessary Condition)* Let \( f(\mathbf{x}) \) be a \( C^2 \) function where \( \mathbf{x} \in \mathbb{R}^n \). Then, if \( \mathbf{x} \) is a minimizer, it is necessarily

\[
\nabla f(\mathbf{x}) = \mathbf{0} \quad \text{and} \quad \nabla^2 f(\mathbf{x}) \succeq \mathbf{0}.
\]

Furthermore, if \( \nabla^2 f(\mathbf{x}) \succ \mathbf{0} \), then the condition becomes *sufficient*.

The proofs would be based on 2nd-order Taylor’s expansion at \( \bar{\mathbf{x}} \) such that if these conditions are not satisfied, then one would be find a second-order descent-direction \( \mathbf{d} \) and a small constant \( \bar{\alpha} > 0 \) such that

\[
f(\bar{\mathbf{x}} + \alpha \mathbf{d}) < f(\bar{\mathbf{x}}), \quad \forall 0 < \alpha \leq \bar{\alpha}.
\]

Again, they may still not be sufficient, e.g., \( f(x) = x^3 \).
General Constrained Optimization

\[(GCO) \quad \min \quad f(x) \]
\[\text{s.t.} \quad h(x) = 0 \in \mathbb{R}^m, \]
\[c(x) \geq 0 \in \mathbb{R}^p.\]

We have dealt the cases when the feasible region is convex polyhedron and/or the feasible can be represented by nonlinear convex cones.

We now study the case that the only assumption is that all functions are in $C^1$, and $C^2$ later, either convex or nonconvex.

We again establish optimality conditions to qualify/verify any local optimizers. These conditions give us qualitative structures of (local) optimizers and lead to quantitative algorithms to numerically find a local optimizer or an KKT solution.
Lagrangian Function of Constrained Optimization

It is more convenient to introduce the Lagrangian Function associated with generally constrained optimization:

\[ L(x, y, s) = f(x) - y^T h(x) - s^T c(x), \]

where multipliers \( y \) of the equality constraints are “free” and \( s \geq 0 \) for the “greater or equal to” inequality constraints.

Lagrangian Function can be viewed as a function aggregated the original objective function plus the penalized terms on constraint violations.

In theory, one can adjust the penalty multipliers \( (y, s \geq 0) \) to repeatedly solve the following so-called Lagrangian Relaxation Problem:

\[ (LRP) \min_x L(x, y, s). \]
Consider the (intersection) of Hypersurfaces (vs. Hyperplanes):

\[ \{ \mathbf{x} \in \mathbb{R}^n : \mathbf{h}(\mathbf{x}) = \mathbf{0} \in \mathbb{R}^m, m \leq n \} \]

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

For a point \( \bar{\mathbf{x}} \) on the surface, we call it a regular point if \( \nabla \mathbf{h}(\bar{\mathbf{x}}) \) have rank \( m \) or the rows are linearly independent. For example, \( (0; 0) \) is not a regular point of

\[ \{(x_1; x_2) \in \mathbb{R}^2 : x_1^2 + (x_2 - 1)^2 - 1 = 0, x_1^2 + (x_2 + 1)^2 - 1 = 0 \}. \]

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

\[ \mathbf{h}(\mathbf{x}(t)) = \mathbf{0}, \quad \mathbf{x}(0) = \bar{\mathbf{x}}, \quad \dot{\mathbf{x}}(0) = \mathbf{d}. \]

\( T_{\bar{\mathbf{x}}} \) is called the tangent-space or tangent-plane of the constraints at \( \bar{\mathbf{x}} \).
\[
\begin{align*}
\min \quad & (x_1)^2 + (x_2)^2 \\
\text{s.t.} \quad & (x_1)^2/4 + (x_2)^2 - 1 = 0
\end{align*}
\]

Figure 2: A Nonlinear Equality Constrained Minimization with Constraint Tangents
Lemma 1  Let $\bar{x}$ be a feasible solution and a regular point of the hypersurface of

$$\{x : h(x) = 0, \ c_i(x) = 0, i \in A_{\bar{x}}\}$$

where active-constraint set $A_{\bar{x}} = \{i : c_i(\bar{x}) = 0\}$. If $\bar{x}$ is a (local) minimizer of (GCO), then there must be no $d$ to satisfy linear constraints:

$$\nabla f(\bar{x})d < 0$$
$$\nabla h(\bar{x})d = 0 \in R^m,$$
$$\nabla c_i(\bar{x})d \geq 0, \ \forall i \in A_{\bar{x}}.$$

This lemma was proved when constraints are linear in which case $d$ is a feasible direction, but needs more work otherwise since there is no feasible direction when constraints are nonlinear.
Suppose we have a \( \bar{d} \) satisfies all linear constraints. Then \( \nabla f(\bar{x})\bar{d} < 0 \) so that \( \bar{d} \) is a descent-direction vector. Denote the active-constraint set at \( \bar{d} \) among the linear inequalities by \( A^d_{\bar{x}} (\subset A_{\bar{x}}) \). Then, \( \bar{x} \) remains a regular point of hypersurface of

\[
\{x : h(x) = 0, c_i(x) = 0, i \in A^d_{\bar{x}}\}.
\]

Thus, there is a curve \( x(t) \) such that

\[
h(x(t)) = 0, \quad c_i(x(t)) = 0, i \in A^d_{\bar{x}}, \quad x(0) = \bar{x}, \quad \dot{x}(0) = \bar{d},
\]

for \( t \in [0, a] \) of a sufficiently small positive constant \( a \).

Also, \( \nabla c_i(\bar{x})\bar{d} > 0, \forall i \not\in A^d_{\bar{x}} \) and \( c_i(\bar{x}) > 0, \forall i \not\in A_{\bar{x}} \). From Taylor's theorem, \( c_i(x(t)) > 0 \) for all \( i \not\in A^d_{\bar{x}} \) so that \( x(t) \) is a feasible curve to the original (GCO) problem for \( t \in [0, a] \). Thus, \( \bar{x} \) must be also a local minimizer among all local solutions on the curve \( x(t) \).

Let \( \phi(t) = f(x(t)) \). Then, \( t = 0 \) must be a local minimizer of \( \phi(t) \) for \( 0 \leq t \leq a \) so that

\[
0 = \phi'(0) = \nabla f(x(0))\dot{x}(0) = \nabla f(\bar{x})\bar{d} < 0, \quad \Rightarrow \text{ a contradiction}.
\]
Theorem 3 (First-Order or KKT Optimality Condition) Let $\bar{x}$ be a (local) minimizer of (GCO) and it is a regular point of $\{x : h(x) = 0, c_i(x) = 0, i \in A_{\bar{x}}\}$. Then, for some multipliers $(\bar{y}, \bar{s} \geq 0)$

$$\nabla f(\bar{x}) = \bar{y}^T \nabla h(\bar{x}) + \bar{s}^T \nabla c(\bar{x})$$

that is,

$$\nabla_x L(\bar{x}, \bar{y}, \bar{s}) = 0;$$

and (complementarity)

$$\bar{s}_i c_i(\bar{x}) = 0, \forall i.$$ 

The proof is again based on the Alternative System Theory. The complementarity condition is from that $c_i(\bar{x}) = 0$ for all $i \in A_{\bar{x}}$, and for $i \notin A_{\bar{x}}$, we simply set $\bar{s}_i = 0$. 

11
Now in addition we assume all functions are in $C^2$, that is, twice continuously differentiable. Recall the tangent linear sub-space at $\bar{x}$:

$$T_{\bar{x}} := \{ z : \nabla h(\bar{x})z = 0, \nabla c_i(\bar{x})z = 0 \forall i \in A_{\bar{x}} \}.$$ 

**Theorem 4** Let $\bar{x}$ be a (local) minimizer of (GCO) and a regular point of hypersurface $$\{ x : h(x) = 0, c_i(x) = 0, i \in A_{\bar{x}} \},$$ and let $\bar{y}, \bar{s}$ denote Lagrange multipliers such that $(\bar{x}, \bar{y}, \bar{s})$ satisfies the (first-order) KKT conditions of (GCO). Then, it is necessary to have

$$d^T \nabla^2_x L(\bar{x}, \bar{y}, \bar{s})d \geq 0 \quad \forall \ d \in T_{\bar{x}}.$$

The Hessian of the Lagrangian function need to be positive semidefinite on the tangent-space.
Proof

The proof reduces to one-dimensional case by considering the objective function \( \phi(t) = f(x(t)) \) on the feasible curve \( x(t) \). Since 0 is a (local) minimizer of \( \phi(t) \),

\[
0 \leq \phi''(t)|_{t=0} = \dot{x}(0)^T \nabla^2 f(\bar{x}) \ddot{x}(0) + \nabla f(\bar{x}) \dot{x}(0) = d^T \nabla^2 f(\bar{x}) d + \nabla f(\bar{x}) \dot{x}(0).
\]

Let all active constraints (including the equality ones) be \( h(x) = 0 \) and differentiating equations \( \bar{y}^T h(x(t)) = \sum_i \bar{y}_i h_i(x(t)) = 0 \) twice, we obtain

\[
0 = \dot{x}(0)^T \left[ \sum_i \bar{y}_i \nabla^2 h_i(\bar{x}) \right] \ddot{x}(0) + \bar{y}^T \nabla h(\bar{x}) \dot{x}(0) = d^T \left[ \sum_i \bar{y}_i \nabla^2 h_i(\bar{x}) \right] d + \bar{y}^T \nabla h(\bar{x}) \dot{x}(0).
\]

Let the second expression subtracted from the first one on both sides and use the FONC:

\[
0 \leq d^T \nabla^2 f(\bar{x}) d - d^T \left[ \sum_i \bar{y}_i \nabla^2 h_i(\bar{x}) \right] d + \nabla f(\bar{x}) \dot{x}(0) - \bar{y}^T \nabla h(\bar{x}) \dot{x}(0)
\]

\[
= d^T \nabla^2 f(\bar{x}) d - d^T \left[ \sum_i \bar{y}_i \nabla^2 h_i(\bar{x}) \right] d
\]

\[
= d^T \nabla^2 x L(\bar{x}, \bar{y}, \bar{s}) d.
\]

Note that this inequality holds for every \( d \in T_{\bar{x}} \).
Theorem 5  Let $\bar{x}$ be a regular point of (GCO) and let $\bar{y}, \bar{s}$ be the Lagrange multipliers such that $(\bar{x}, \bar{y}, \bar{s})$ satisfies the (first-order) KKT conditions of (GCO). Then, if in addition

$$d^T \nabla^2_x L(\bar{x}, \bar{y}, \bar{s}) d > 0 \quad \forall \, 0 \neq d \in T_{\bar{x}},$$

then $\bar{x}$ is a local minimizer of (GCO).

See the proof in Chapter 11.8 of LY.
\[
\min \ (x_1)^2 + (x_2)^2 \quad \text{s.t.} \quad -(x_1)^2/4 - (x_2)^2 + 1 \leq 0
\]

Figure 3: FONC and SONC for Constrained Minimization
\[ L(x_1, x_2, y) = (x_1)^2 + (x_2)^2 - y\left(-\frac{(x_1)^2}{4} - (x_2)^2 + 1\right), \]

\[ \nabla_x L(x_1, x_2, y) = (2x_1(1 + y/4), 2x_2(1 + y)), \]

\[ \nabla^2_x L(x_1, x_2, y) = \begin{pmatrix} 2(1 + y/4) & 0 \\ 0 & 2(1 + y) \end{pmatrix} \]

\[ T_x := \{ (z_1, z_2) : (x_1/4)z_1 + x_2z_2 = 0 \}. \]

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:

\[ \nabla^2_x L(2, 0, -4) = \begin{pmatrix} 0 & 0 \\ 0 & -6 \end{pmatrix}, \quad T_x = \{(z_1, z_2) : z_1 = 0\} \]

Then the Hessian is not positive semidefinite on \( T_x \) since

\[ d^T \nabla^2_x L(2, 0, -4) d = -6d_2^2 \leq 0. \]

Consider the third KKT point:

\[ \nabla^2_x L(0, 1, -1) = \begin{pmatrix} 3/2 & 0 \\ 0 & 0 \end{pmatrix}, \quad T_x = \{(z_1, z_2) : z_2 = 0\} \]

Then the Hessian is positive definite on \( T_x \) since

\[ d^T \nabla^2_x L(0, 0, -1) d = (3/2)d_1^2 > 0, \quad \forall \mathbf{0} \neq d \in T_x. \]
Summary Theorem of KKT Conditions for GCO

We now consider optimality conditions for problems having three types of inequalities:

\[
\begin{align*}
\text{(GCO)} & \quad \min f(x) \\
\text{s.t.} & \quad c_i(x) \quad (\leq, =, \geq) \quad 0, \quad i = 1, \ldots, m,
\end{align*}
\]

For any feasible point \( x \) of (GCO) define the active constraint set by \( A_x = \{ i : c_i(x) = 0 \} \).

Let \( \bar{x} \) be a local minimizer for (GCO) and \( \bar{x} \) is a regular point on the hypersurface of the active constraints. Then there exist multipliers \( \bar{y} \) such that

\[
\begin{align*}
\nabla f(\bar{x}) & = \bar{y}^T \nabla c(\bar{x}) \\
\bar{y}_i & (\leq, \text{free}, \geq) \quad 0, \quad i = 1, \ldots, m, \\
\bar{y}_i c_i(\bar{x}) & = 0.
\end{align*}
\]
In the second-order test, we typically like to know whether or not
\[ d^T Q d \geq 0, \quad \forall d, \quad \text{s.t. } A d = 0 \]
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 \( d \) is in the null space of matrix \( A \) if and only if
\[ d = (I - A^T (AA^T)^{-1} A) u = P_A u \]
for some vector \( u \in \mathbb{R}^n \), where \( P_A \) is called the projection matrix of \( A \). Thus, the test becomes whether or not
\[ u^T P_A Q P_A u \geq 0, \quad \forall u \in \mathbb{R}^n, \]
that is, we just need to test positive semidefiniteness of \( P_A Q P_A \) as usual.
Another way is to apply SDP relaxation:

\[
(SDP) \quad \min Q \cdot D \\
\text{s.t.} \quad A_i^T A_i \cdot D = 0; \forall i \\
D \succeq 0,
\]

where \( A_i \) is the \( i \)th row vector of \( A \). The objective value is bounded below by 0 if the dual has a feasible solution:

\[
(SDD) \quad \min 0^T y \\
\text{s.t.} \quad Q - \sum_i y_i A_i^T A_i \succeq 0.
\]

Why?