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 11.1-11.8
We have dealt the cases when the feasible region is a convex polyhedron and/or the feasible can be represented by nonlinear convex cones intersect linear equality constraints.

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.

The main proof idea is that if $\bar{x}$ is a local minimier of (GCO), then it must be a local minimizer of the problem where the constraints are linearized using the First-Order Taylor expansion.
Hypersurface and Implicit Function Theorem

Consider the (intersection) of Hypersurfaces (vs. Hyperplanes):

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

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

For a point \( \bar{x} \) on the surface, we call it a regular point if \( \nabla h(\bar{x}) \) have rank \( m \) or the rows, or the gradient vector of each \( h_i(\cdot) \) at \( \bar{x} \), 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 (Appendix A of the Text), if \( \bar{x} \) is a regular point and \( m < n \), then for every \( d \in T_{\bar{x}} = \{ z : \nabla h(\bar{x}) z = 0 \} \) there exists a curve \( \mathbf{x}(t) \) on the hypersurface, parametrized by a scalar \( t \) in a sufficiently small interval \([ -a, a ]\), such that

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

\( T_{\bar{x}} \) is called the tangent-space or tangent-plane of the constraints at \( \bar{x} \).
Figure 1: Tangent Plane on a Hypersurface at Point $\mathbf{x}^*$
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:

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

(1)

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.

$\bar{x}$ being a regular point is often referred as a Constraint Qualification condition.
Proof

Suppose we have a $\mathbf{d}$ satisfies all linear constraints. Then $\nabla f(\mathbf{x})\mathbf{d} < 0$ so that $\mathbf{d}$ is a descent-direction vector. Denote the active-constraint set at $\mathbf{d}$ among the linear inequalities in (1) by $A^d_{\mathbf{x}} (\subset A_{\mathbf{x}})$. Then, $\mathbf{x}$ remains a regular point of hypersurface of

$$\{ \mathbf{x} : h(\mathbf{x}) = 0, \ c_i(\mathbf{x}) = 0, i \in A^d_{\mathbf{x}} \}.$$

Thus, there is a curve $\mathbf{x}(t)$ such that

$$h(\mathbf{x}(t)) = 0, \ c_i(\mathbf{x}(t)) = 0, i \in A^d_{\mathbf{x}}, \ \mathbf{x}(0) = \mathbf{x}, \ \dot{\mathbf{x}}(0) = \mathbf{d},$$

for $t \in [0 \ a]$ of a sufficiently small positive constant $a$.

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

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

$$0 \leq \phi'(0) = \nabla f(\mathbf{x}(0))\dot{\mathbf{x}}(0) = \nabla f(\mathbf{x})\mathbf{d} < 0, \ \Rightarrow \ a \ contradiction.$$
Theorem 1 (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 $(\bar{y}, \bar{s} \geq 0)$

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

(2)

and (complementarity slackness)

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

The proof is again based on the Alternative System Theory or Farkas Lemma. The complementarity slackness 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 $\bar{s}_i = 0$.

A solution who satisfies these conditions is called an KKT point or solution of (GCO) – any local minimizer $\bar{x}$, if it is also a regular point, must be an KKT solution; but the reverse may not be true.
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, so that the KKT condition (2) can be written as

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

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:

$$(LRP) \quad \min_x L(x, y, s).$$
Constraint Qualification and the KKT Theorem

One condition for a local minimizer \( \bar{x} \) that must always be an KKT solution is the constraint qualification: \( \bar{x} \) is a regular point of the constraints. Otherwise, a local minimizer may not be an KKT solution: Consider \( \bar{x} = (0; 0) \) of a convex nonlinearly-constrained problem

\[
\min x_1, \quad \text{s.t.} \quad x_1^2 + (x_2 - 1)^2 - 1 \leq 0, \quad x_1^2 + (x_2 + 1)^2 - 1 \leq 0.
\]

On the other hand, even the regular point condition does not hold, the KKT theorem may still true:

\[
\min x_2, \quad \text{s.t.} \quad x_1^2 + (x_2 - 1)^2 - 1 \leq 0, \quad x_1^2 + (x_2 + 1)^2 - 1 \leq 0,
\]

that is, \( \bar{x} = (0; 0) \) is an KKT solution of the latter problem.

Therefore, finding an KKT solution is a plausible way to find a local minimizer.
We now consider optimality conditions for problems having three types of inequalities:

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

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

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

\[
\begin{align*}
\nabla f(\bar{x}) & = \bar{y}^T \nabla c(\bar{x}) \quad \text{(Lagrangian Derivative Conditions (LDC))} \\
\bar{y}_i (\leq, '\text{free}', \geq) & = 0, \ i = 1, \ldots, m, \quad \text{(Multiplier Sign Constraints (MSC))} \\
\bar{y}_i c_i(\bar{x}) & = 0, \quad \text{(Complementarity Slackness Conditions (CSC))}.
\end{align*}
\]

The complete First-Order KKT Conditions consist of these four parts!
Recall SOCP Relaxation of Sensor Network Localization

Given $a_k \in \mathbb{R}^2$ and Euclidean distances $d_k$, $k = 1, 2, 3$, find $x \in \mathbb{R}^2$ such that

$$\begin{align*}
\min_x & \quad 0^T x, \\
\text{s.t.} & \quad \|x - a_k\|^2 - d_k^2 \leq 0, \quad k = 1, 2, 3,
\end{align*}$$

$$L(x, y) = 0^T x - \sum_{k=1}^{3} y_k (\|x - a_k\|^2 - d_k^2),$$

$$\begin{align*}
0 &= \sum_{k=1}^{3} y_k (x - a_k) \quad \text{(LDC)} \\
y_k &\leq 0, \quad k = 1, 2, 3, \quad \text{(MSC)} \\
y_k (\|x - a_k\|^2 - d_k^2) &= 0. \quad \text{(CSC)}
\end{align*}$$
Arrow-Debreu’s Exchange Market with Linear Economy

Each trader $i$, equipped with a good bundle vector $w_i$, trade with others to maximize its individual utility function. The equilibrium price is an assignment of prices to goods so as when every producer sells his/her own good bundle and buys a maximal bundle of goods then the market clears. Thus, trader $i$’s optimization problem, for given prices $p_j, j \in G$, is

$$\text{maximize} \quad u_i^T x_i := \sum_{j \in P} u_{ij} x_{ij}$$

subject to

$$p^T x_i := \sum_{j \in P} p_j x_{ij} \leq p^T w_i,$$

$$x_{ij} \geq 0, \quad \forall j,$$

Then, the equilibrium price vector is the one such that there are maximizers $x(p)_i$s

$$\sum_i x(p)_{ij} = \sum_i w_{ij}, \quad \forall j.$$
Example of Arrow-Debreu’s Model

Traders 1, 2 have good bundle

\[ w_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad w_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \]

Their optimization problems for given prices \( p_x, p_y \) are:

\[
\begin{align*}
\text{max} \quad & 2x_1 + y_1 \\
\text{s.t.} \quad & p_x \cdot x_1 + p_y \cdot y_1 \leq p_x, \\
& x_1, y_1 \geq 0
\end{align*}
\]

\[
\begin{align*}
\text{max} \quad & 3x_2 + y_2 \\
\text{s.t.} \quad & p_x \cdot x_2 + p_y \cdot y_2 \leq p_y, \\
& x_2, y_2 \geq 0
\end{align*}
\]

One can normalize the prices \( p \) such that one of them equals 1. This would be one of the problems in HW2.
Equilibrium conditions of the Arrow-Debreu market

Similarly, the necessary and sufficient equilibrium conditions of the Arrow-Debreu market are

\[ p_j \geq u_{ij} \cdot \frac{p^T w_i}{x_i}, \quad \forall i, j, \]

\[ \sum_i x_{ij} = \sum_i w_{ij} \quad \forall j, \]

\[ p_j > 0, \ x_i \geq 0, \quad \forall i, j; \]

where the budget for trader \( i \) is replaced by \( p^T w_i \). Again, the nonlinear inequality can be rewritten as

\[ \log(u^T_i x_i) + \log(p_j) - \log(p^T w_i) \geq \log(u_{ij}), \ \forall i, j, u_{ij} > 0. \]

Let \( y_j = \log(p_j) \) or \( p_j = e^{y_j} \) for all \( j \). Then, these inequalities become

\[ \log(u^T_i x_i) + y_j - \log(\sum_j w_{ij} e^{y_j}) \geq \log(u_{ij}), \ \forall i, j, u_{ij} > 0. \]

Note that the function on the left is concave in \( x_i \) and \( y_j \).

**Theorem 2**  The equilibrium set of the Arrow-Debreu Market is convex in allocations and the logarithmic of prices.
Cobb-Douglas Utility:

\[ u_i(x_i) = \prod_{j \in G} x_{ij}^{u_{ij}}, \ x_{ij} \geq 0. \]

Leontief Utility:

\[ u_i(x_i) = \min_{j \in G} \left\{ \frac{x_{ij}}{u_{ij}}, \ x_{ij} \geq 0 \right\}. \]

Again, the equilibrium price vector is the one such that there are maximizers to clear the market.
Example of Geometric Optimization

Consider the Geometric Optimization Problem

\[
\begin{align*}
\min_x & \quad \sum_{i=1}^m \left( a_i \prod_{j=1}^n x_{ij}^{u_{ij}} \right) \\
\text{s.t.} & \quad \prod_{j=1}^n x_{kj}^{c_{kj}} = b_k, \quad k = 1, \ldots, K \\
& \quad x_j > 0, \quad \forall j,
\end{align*}
\]

where the coefficients \( a_i \geq 0 \ \forall i \) and \( b_k > 0 \ \forall k \).

\[
\begin{align*}
\min_{x,y,z} & \quad xy + yz + zx \\
\text{s.t.} & \quad xyz = 1 \\
& \quad (x, y, x) > 0.
\end{align*}
\]
Let \( y_j = \log(x_j) \) so that \( x_j = e^{y_j} \). Then the problem becomes

\[
\min_x \quad \sum_{i=1}^{m} \left( a_i e^{\sum_{j=1}^{n} u_{ij} y_j} \right)
\]

s.t. \[
\sum_{j=1}^{n} c_{kj} y_j = \log(b_k), \quad k = 1, \ldots, K
\]

\( y_j \text{ free } \forall j. \)

This is a convex objective function with linear constraints!

\[
\min_{u,v,w} \quad e^{u+v} + e^{v+w} + e^{w+u}
\]

s.t. \( u + v + w = 0 \)

\( (u, v, w) \text{ free.} \)

Now the KKT solution suffices!
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 3** 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.
The proof reduces to one-dimensional case by considering the objective function \( \phi(t) = f(x(t)) \) for the feasible curve \( x(t) \) on the surface of ALL active constraints. Since 0 is a (local) minimizer of \( \phi(t) \) in an interval \([-a a]\) for a sufficiently small \( a > 0 \), we must have \( \phi'(0) = 0 \) so that

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

Let all active constraints (including the equality ones) be \( h(x) = 0 \) and differentiating equations

\[
\tilde{y}^T h(x(t)) = \sum_i \tilde{y}_i h_i(x(t)) = 0
\]

twice, we obtain

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

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

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

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

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

Note that this inequality holds for every \( d \in T_{\bar{x}} \).
Second-Order Sufficient Conditions for GCO

**Theorem 4** Let $\bar{x}$ be a regular point of (GCO) with **equality constraints only** and let $\bar{y}$ be the Lagrange multipliers such that $(\bar{x}, \bar{y})$ satisfies the (first-order) KKT conditions of (GCO). Then, if in addition

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

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

See the proof in Chapter 11.5 of LY.

The SOSC for general (GCO) is proved in Chapter 11.8 of LY.
\[
\text{min } (x_1)^2 + (x_2)^2 \quad \text{s.t. } (x_1)^2/4 + (x_2)^2 - 1 = 0
\]

Figure 2: FONC and SONC for Constrained Minimization
\[ L(x_1, x_2, y) = (x_1)^2 + (x_2)^2 - y(- (x_1)^2/4 - (x_2)^2 + 1), \]

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

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

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

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:

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

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

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

Consider the third KKT point:

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

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

\[
d^T \nabla_x^2 L(0, 0, -1) d = (3/2)d_1^2 > 0, \ \forall 0 \neq d \in T_x.
\]

This would be sufficient for the third KKT solution to be a local minimizer.
In the second-order test, we typically like to know whether or not

\[ d^T Q d \geq 0, \quad \forall d, \text{ s.t. } A d = 0 \]

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

\[ d = (I - A^T (A A^T)^{-1} A) u = P_A u \]

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.
Spherical Constrained Nonconvex Quadratic Optimization

\[
\text{(SCQP)} \quad \min \quad x^T Q x + c^T x \\
\text{s.t.} \quad \|x\|^2 (\leq, =) 1.
\]

**Theorem 5** The FONC and SONC, that is, the following conditions on \(x\), together with the multiplier \(y\),

\[
\|x\|^2 \ (\leq, =) \ 1, \ (OPC') \\
2Qx + c - 2yx = 0, \ (LDC') \\
y \ (\leq, '\text{free}') \ 0, \ (MSC') \\
y(1 - \|x\|^2) = 1, \ (CSC') \\
(Q - yI) \succeq 0, \ (SOC').
\]

are necessary and sufficient for finding the global minimizer of (SCQP).