7.1

Optimality Conditions for NLP

Katta G. Murty, IOE 611 Lecture slides

HISTORY: 1-dimensional unconstrained min., developed in 17th century as Newton was developing calculus. Soon, extended to multidimensional unconstrained min.

Leibniz (a co-developer of calculus with Newton) was first to distinguish between max and min, and Maclaurin (also did fundamental work on Taylor series) was first to give method to distinguish between local max and local min using 2nd and higher order conditions, in his book published 1742.

Opt. conds. for equality constrained NLP are foundation on NLP theory. Inspiration from problems in mechanics, these conds. developed 18th, 19th centuries. First results by L. Euler, J. L. Lagrange, first published in book by J. L. Lagrange 1788.

Opt. conds. for NLPs involving some inequality constraints studied beginning with J. Fourier 1798; and later by A. Cournot, M. Ostrogradsky, C. F. Gauss, J. Farkas, G. Hamel and several others. Rigorous development completed in W. Karush's M. S.

thesis (U. of Chicago) in 1939, and later in essentially same form by Kuhn, Tucker in 1951.

Basic Principles For Deriving Opt. Conds.

1. If $\bar{x}$ is a local min for an NLP, there cannot be any descent feasible direction at $\bar{x}$.

   Is the converse true?

   EXAMPLE (PEANO, 1884) : Consider the unconstrained min of $\theta(x) = (x_2 - x_1^2)(x_2 - 2x_1^2)$ in R^2 and the point $\bar{x} = 0$.

2. If $\theta(x)$ is the objective function to be minimized in an NLP, $\bar{x}$ is a local min for the NLP, and y a feasible direction at $\bar{x}$ for this NLP, and $\alpha > 0$ is such that $\bar{x} + \lambda y$ is feasible $\forall$ $0 \leq \lambda \leq \alpha$, then for the following one dimensional problem, $\lambda = 0$ must be a local min.

$$\text{minimize} \quad f(\lambda) = \theta(\bar{x} + \lambda y) \quad \text{over } 0 \leq \lambda \leq \alpha$$

3. Let $\bar{x}$ be a local min for an NLP in which $\theta(x)$ is being minimized. If $x = g(\lambda) = (g_1(\lambda), \dots, g_m(\lambda))^T$ defines a differentiable curve satisfying: $g(0) = \bar{x}$, and $g(\lambda)$ lies in the feasible

region for $0 \le \lambda \le \alpha$ for some $\alpha > 0$, then $\lambda = 0$ must be a local min for the following one dimensional problem

$$\text{minimize} \quad f(\lambda) = \theta(g(\lambda)) \quad \text{over } 0 \le \lambda \le \alpha$$

Using these, we can derive necessary opt. conds. for higher dimensional problems using the known conds. for one dimensional problems.

How to represent curves in R^n?

Most popular is the parametric representation. All coordinates of a general point on curve are expressed as functions of a sngle parameter λ, as in $\quad x(\lambda) = (x_1(\lambda), \ldots, x_n(\lambda))^T$.

Curve differentiable if $\frac{dx_j(\lambda)}{d\lambda}$ exists $\forall j$

Curve twicw differentiable if $\frac{d^2x_j(\lambda)}{d\lambda^2}$ exists $\forall j$

If $x(0) = \bar{x}$, tangent line to curve at $\bar{x}$ is the line $\{x = \bar{x} + \delta\frac{dx(0)}{d\lambda} : \delta \text{ real}\}$

Opt. Conds. for Unconstrained 1-Dimensional Problem

$$\min f(\lambda), \qquad \lambda \in R^1$$

The conditions for $\bar{\lambda}$ to be a local min or max are:

1. 1st order nec. cond.: $f'(\bar{\lambda}) = 0$.
2. 2nd order nec. cond: $f'(\bar{\lambda}) = 0$, and if $f''(\bar{\lambda}) \neq 0$, then $f''(\bar{\lambda}) > 0$ (for local min), or $f''(\bar{\lambda}) < 0$ (for local max).
3. Higher order conds: If 1st and some of following derivatives vanish at $\bar{\lambda}$, then $\bar{\lambda}$ is (or is not) a local opt if first nonvanishing derivative at it is of even (odd) order. If it is of even order, $\bar{\lambda}$ is a local min (or local max) if this derivative is positive (negative).

Proof: Use Taylor series expansion.

Opt. Conds. for Unconstrained min in R^n

Consider $\min \theta(x), \quad x \in R^n$.

Here are the conditions for $\bar{x}$ to be a local min to this problem.

1. 1st Order Nec. Cond.: $\nabla\theta(\bar{x}) = 0$
2. 2nd Order Nec. Conds.: $\nabla\theta(\bar{x}) = 0$, and $\nabla^2_{xx}\theta(\bar{x})$ is PSD.
3. Suff. Conds.: $\nabla\theta(\bar{x}) = 0$, and $\nabla^2_{xx}\theta(\bar{x})$ is PD.
4. Nec. & Suff. cond. for global min when $\theta(x)$ is convex: $\nabla\theta(\bar{x}) = 0$.

Examples: 1. $\theta(x) = 2x_1^2 + x_2^2 + x_3^2 + x_1x_2 + x_1x_3 + x_2x_3 - 9x_1 - 9x_2 - 8x_3$.

2. $\theta(x) = 2x_1^2 + x_3^2 + 2x_1x_2 + 2x_1x_3 + 4x_2x_3 + 4x_1 - 8x_2 + 2x_3$.

3. $\theta(x) = -2x_1^2 - x_2^2 + x_1x_2 - 10x_1 + 6x_2$.

Can check These conds. efficiently. Notice gap between the 2nd order nec. and suff. cond. when $\theta(x)$ is nonconvex. When in gap, not able to conclude that either $\bar{x}$ is, or is not, a local min .

Opt. Conds. In Terms of Feasible Directions

Consider $\min \theta(x)$ s. to constraints. Let Γ be set of feasible solutions.

Opt. conds. for $\bar{x} \in \Gamma$ to be local minimum are:

1. 1st order noc. conds.: $\nabla\theta(\bar{x})y \geq 0 \quad \forall$ feasible directions y at $\bar{x}$.

2. 2nd order noc. conds.: $\nabla\theta(\bar{x})y \geq 0 \quad \forall$ feasible directions y at $\bar{x}$; and $y^T\nabla^2_{xx}\theta(\bar{x})y \geq 0 \quad \forall$ feasible directions y at $\bar{x}$ satisfying $\nabla\theta(\bar{x})y = 0$.

When Γ convex, set of feasible directions at $\bar{x}$ is $\{x - \bar{x} : x \in \Gamma, x \neq \bar{x}\}$.

These 1st order nec. conds. when Γ convex, lead to **Variational Inequality Problem**: VI(K, f): INPUT: A vector function $f(x) = (f_1(x), \ldots, f_n(x))^T$ defined over R^n, and a subset $K \subset R^n$.

OUTPUT DESIRED: Find an $x^* \in K$ satisfying $(x - x^*)^T f(x^*) \geq 0 \quad \forall x \in K$.

Tangent Planes

Let Γ be the feasible region to an NLP, and $\bar{x} \in \Gamma$. The **Tangent Plane** at $\bar{x}$ to Γ is defined to be the set of all directions of tangent lines at $\bar{x}$ to differentiable curves through $\bar{x}$ lying in Γ, i.e.,

$\{[\frac{dx(\lambda)}{d\lambda}]_{\lambda=0} : x(\lambda)$ is a differentiable curve lying in Γ for a positive length around $\lambda = 0, x(0) = \bar{x}\}$.

Tangent planes play major role in deriving opt. conds. For general sets determining tangent planes hard. However for set of feasible solutions of following system,

$$h(x) = (h_1(x), \ldots, h_m(x))^T = 0$$

tangent plane at a feasible solution $\bar{x}$ has a simple characterization if $\bar{x}$ satisfies a condition called a **CQ (constraint qualification)**.

Regularity Condition: The feasible sol. $\bar{x}$ of above system satisfies this CQ (and hence called a **regular feasible solution** or **regular point**) if $\{\nabla h_i(\bar{x}) : i = 1 \text{ to } m\}$ is linearly independent, i.e., if Jacobian $\nabla h(\bar{x})$ has full row rank.

THEOREM: If $\bar{x}$ is a regular feasible sol. of above system, the tangent plane at $\bar{x}$ to the set of feasible solutions is $\{y : \nabla h(\bar{x})y = 0, y \neq 0\}$.

IMPLICIT FUNCTION THEOREM: Let $\bar{x} \in R^n$ be a feasible sol. of system of eqs. $f_i(x) = 0$, $i = 1$ to m s. th. $(\frac{\partial f_i(\bar{x})}{\partial x_j} : i = 1 \text{ to } m, j = 1 \text{ to } m)$ is nonsingular. Then there exists an open nbhd. $\mathcal{D}$ of $\bar{\chi} = (\bar{x}_{m+1}, \ldots, \bar{x}_n)^T$ in R^{n-m} in which the variables $x_1, \ldots, x_m$ can be expressed as differentiable functions of $\chi = (x_{m+1}, \ldots, x_n)^T$ on the set of feasible sols. of the original system of eqs. Further, the partial derivatives at $\bar{x}$ of these functions, $\frac{\partial x_i(\bar{\chi})}{\partial x_j}$ for $i = 1$ to m, $j = m + 1$ to n, are obtained by solving the system of eqs.

$$\sum_{r=1}^{m} \frac{\partial f_i(\bar{x})}{\partial x_r}\frac{\partial x_r(\bar{\chi})}{\partial x_j} + \frac{\partial f_i(\bar{x})}{\partial x_j} = 0, \quad j = m + 1 \text{ to } n, i = 1 \text{ to } m$$

Example for Implicit Func. Theorem: $n = 5$, $m = 2$, system is

$$
\begin{aligned}
f_1(x) &= x_1 + x_2 + x_3 + x_4 - x_5 - 12 = 0 \\
f_2(x) &= -x_1 + x_2 - 2x_3 - x_4 + 4x_5 - 2 = 0
\end{aligned}
$$

and solution is $\bar{x} = (5, 7, 0, 0, 0)^T$.

Examples for T. Planes:

(i) $x = (x_1, x_2)^T, \quad h_1(x) = x_1 = 0$

(ii) $x = (x_1, x_2)^T, \quad h_1(x) = x_1^3 = 0$

Opt. Conds. for Equality Constrained NLP

Consider min $\theta(x)$ s. to $h_i(x) = 0$, $i = 1$ to m.

The **Lagrangian** for this problem is defined to be $L(x, \mu) = \theta(x) - \Sigma_{i=1}^{m} \mu_i h_i(x)$ where $\mu = (\mu_1, \ldots, \mu_m)$ is known as the vector of **Lagrange Multipliers**.

The opt. conds. for this problem are:

1. 1st order nec. opt. conds.: If $\bar{x}$ is a a local min. and either all constraints are linear, or the regularity cond. holds at $\bar{x}$, there must exist Lagrange multiplier vector $\bar{\mu} = (\bar{\mu}_1, \ldots, \bar{\mu}_m)$ s. th.

$$\nabla\theta(\bar{x}) - \sum_{i=1}^{m} \bar{\mu}_i \nabla h_i(\bar{x}) = 0$$

   i.e., $\nabla_x L(\bar{x}, \bar{\mu}) = 0$.

2. 2nd order nec. opt. conds.: If $\bar{x}$ is a a local min. and either all constraints are linear, or the regularity cond. holds at $\bar{x}$, there must exist Lagrange multiplier vector $\bar{\mu} = (\bar{\mu}_1, \ldots, \bar{\mu}_m)$ s. th.

$$\begin{aligned} \nabla_x L(\bar{x}, \bar{\mu}) &= \nabla\theta(\bar{x}) - \sum_{i=1}^{m} \bar{\mu}_i \nabla h_i(\bar{x}) = 0 \\ y^T \nabla^2_{xx}(L(\bar{x}, \bar{\mu}))y &\geq 0 \ \ \forall y \in T = \{y : \nabla h(\bar{x})y = 0\} \end{aligned}$$

3. Suff. Cond. for a strong local min: If $\bar{x}$ is feasible, and there exists a Lagrange Multiplier vector $\bar{\mu}$ s. th.

$$\begin{aligned} \nabla_x L(\bar{x}, \bar{\mu}) &= \nabla\theta(\bar{x}) - \sum_{i=1}^{m} \bar{\mu}_i \nabla h_i(\bar{x}) = 0 \\ y^T \nabla^2_{xx}(L(\bar{x}, \bar{\mu}))y &> 0 \ \ \forall y \neq 0 \in T = \{y : \nabla h(\bar{x})y = 0\} \end{aligned}$$

then $\bar{x}$ is a strong local min for the NLP.

Given $\bar{x}$ all of the above conditions can be checked efficiently. Notice that here also, there is a small gap between the nec. and suff. conds. for a local min.

Examples

1. min $\theta(x) = x_1 x_2$ s. to $x_1 + x_2 = 2$.

2. min $-x_1 - x_2$ s. to $x_1^2 + x_2^2 - 8 = 0$

3. min $\theta(x) = 2x_1^3 + \frac{1}{2}x_2^2 + x_1x_2 + \frac{1}{24}x_1$ s. to $x_1 + x_2 = 2$.

4. [A solution of NLP may not be an unconstrained minimizer of the Lagrangian]: min x_1^3 s. to $x_1 + 1 = 0$.

The Lagrange Multiplier Technique: Technique for solving equality constrained NLP by solving the 1st order nec. conds. as a system of nonlinear eqs.

Opt. Conds. for Linearly Constrained NLP

When there are inequality constraints, the 1st order nec. conds. are called KKT Conditions.

For linearlyconstrained problems, no CQ are required to derive the necessary optimality conds.

Consider: $\min \theta(x)$ s. to $A_{i.}x \begin{cases} = b_i, & i = 1 \text{ to } m \\ \geq b_i, & i = m+1 \text{ to } m+p \end{cases}$

Let $\bar{x}$ be a feasible sol. Let $P(\bar{x}) = \{i : m+1 \leq i \leq m+p$ and $A_{i.}\bar{x} = b_i\}$, i.e., the index set of active inequality constraints at $\bar{x}$.

Then T. plane at $\bar{x}$ is $T(\bar{x}) = \{y : A_{i.}y \begin{cases} = 0, & i = 1 \text{ to } m \\ \geq 0, & i \in P(\bar{x}) \end{cases} \}$

The Lagrangian is $L(x, \pi) = \theta(x) - \Sigma_{i=1}^{m+p} \pi_i(A_{i.}x - b_i)$ where $\pi = (\pi_1, \ldots, \pi_{m+p})$ is the Lagrange Multiplier vector. The opt. conds. for the feasible sol. $\bar{x}$ to be a local min are:

1. 1st order nec. opt. conds.: There must exist a Lagrange multiplier vector $\bar{\pi}$ s. th.

$$
\begin{aligned}
\nabla_x L(\bar{x}, \bar{\pi}) &= \nabla\theta(\bar{x}) - \bar{\pi}A = 0 \\
\pi_i &\geq 0, \quad \forall i \in \{m+1, \ldots, m+p\} \\
\bar{\pi}_i(A_{i.}\bar{x} - b_i) &= 0, \quad \forall i \in \{m+1, \ldots, m+p\}
\end{aligned}
$$

The 3rd condition in the above is known as the **Complementary Slackness Condition**.

2. 2nd order nec. opt. conds.: There must exist a Lagrange multiplier vector $\bar{\pi}$ s. th.

$$
\begin{aligned}
\nabla_x L(\bar{x}, \bar{\pi}) &= \nabla\theta(\bar{x}) - \bar{\pi}A = 0 \\
\pi_i &\geq 0, \quad \forall i \in \{m+1, \ldots, m+p\} \\
\bar{\pi}_i(A_{i.}\bar{x} - b_i) &= 0, \quad \forall i \in \{m+1, \ldots, m+p\} \\
\text{and} \quad y^T \nabla^2_{xx}(L(\bar{x}, \bar{\pi}))y &\geq 0, \quad \forall y \in T(\bar{x})
\end{aligned}
$$

Given $\bar{x}$, the 1st order conds. can be checked efficiently. However, if $\theta(x)$ is nonconvex, checking the last cond. among the 2nd order conds. is hard (see Murty, Kabadi [1987]).

Opt. Conds. for General Constrained NLP

Consider: $\min \theta(x)$ s. to $g_i(x) \begin{cases} = 0, & i = 1 \text{ to } m \\ \geq 0 & i = m+1 \text{ to } m+p \end{cases}$.

The Lagrangian is $L(x, \pi) = \theta(x) - \Sigma_{i=1}^{m+p} \pi_i g_i(x)$ where $\pi = (\pi_1, \ldots, \pi_{m+p})$ is the Lagrange Multiplier vector.

Let $\bar{x}$ be a feasible sol. Let $P(\bar{x}) = \{i : m+1 \leq i \leq m+p$ and $g_i(\bar{x}) = 0\}$, i.e., the index set of active inequality constraints at $\bar{x}$.

Nec. conds. derived under a CQ. There are several CQ, some weaker than the others. The principal ones are:

Regularity condition: The feasible solution $\bar{x}$ satisfies this (and hence called a **regular point**) if $\{\nabla_x g_i(\bar{x}) : i \in \{1, \ldots, m\} \cup P(\bar{x})\}$ is linearly independent. It is possible to check whether this condition holds at $\bar{x}$ efficiently.

First order CQ: The feasible solution $\bar{x}$ satisfies this CQ if for each $y \in \{y : \nabla_x g_i(\bar{x}) y = 0, \quad i = 1 \text{ to } m; \nabla_x g_i(\bar{x}) y \geq 0, \quad i \in P(\bar{x})\}$, y is the tangent direction to a differentiable curve emanating from $\bar{x}$ and lying in the feasible region. This condition is hard to check.

Second order CQ: The feasible solution $\bar{x}$ satisfies this CQ if for each $y \in \{y : \nabla_x g_i(\bar{x})y = 0, i \in \{1, \ldots, m\} \cup P(\bar{x})\}$, there exists a twice differentiable curve emanating from $\bar{x}$ and lying in the region $\{x : g_i(x) = 0, i \in \{1, \ldots, m\} \cup P(\bar{x})\}$, for which y is the tangent direction at $\bar{x}$. This condition is hard to check.

Mangasarian-Fromovitz CQ: The feasible solution $\bar{x}$ satisfies this CQ if the set $\{d : \nabla_x g_i(\bar{x})d = 0, \quad i = 1 \text{ to } m\} \cap \{d : \nabla_x g_i(\bar{x})d > 0, \quad i \in P(\bar{x})\} \neq \emptyset$, and $\{\nabla_x g_i(\bar{x}) : i = 1 \text{ to } m\}$ is linearly independent. This condition can be checked efficiently.

The opt. conds. for the feasible sol. $\bar{x}$ to be a local min are:

1st order nec. conds. : If $\bar{x}$ is a local minimum , and either all the constraints are linear constraints, or $\bar{x}$ satisfies the regularity or the 1st order or the Mangasarian-Fromovitz CQs, then there exists a Lagrange multiplier vector $\bar{\pi} = (\bar{\pi}_1, \ldots, \bar{\pi}_{m+p})$ such that

$$\nabla_x \theta(\bar{x}) = \sum_{i=1}^{m+p} \bar{\pi}_i \nabla_x g_i(\bar{x})$$

$$\bar{\pi}_i \geq 0 \text{ for } i \in \{m+1, \ldots, m+p\}$$

$$\bar{\pi} g_i(\bar{x}) = 0 \text{ for } i \in \{m+1, \ldots, m+p\}$$

These conds. are known as the **KKT (Karush-Kuhn-Tucker) conditions**. Given $\bar{x}$, checking these conds. , can be posed as an LP.

2nd order conds. : If $\bar{x}$ is a local min, and either all the constraints are linear constraints, or $\bar{x}$ satisfies the regularity or the 2nd order or the Mangasarian-Fromovitz CQs, then there exists a Lagrange multiplier vector $\bar{\pi} = (\bar{\pi}_1, \ldots, \bar{\pi}_{m+p})$ such that

$$\nabla_x \theta(\bar{x}) = \sum_{i=1}^{m+p} \bar{\pi}_i \nabla_x g_i(\bar{x})$$

$$\bar{\pi}_i \geq 0 \text{ for } i \in \{m+1, \ldots, m+p\}$$

$$\bar{\pi} g_i(\bar{x}) = 0 \text{ for } i \in \{m+1, \ldots, m+p\}$$

$$\text{and } y^T(\nabla^2_{xx} L(\bar{x}, \bar{\pi}))y \geq 0 \text{ for all } y \in T_1$$

where $T_1 = \{y : \nabla_x g_i(\bar{x})y = 0 \; i \in \{1, \ldots, m\} \cup P(\bar{x})\}$.

Given $\bar{x}$, these conds. can be checked efficiently.

Suff. conds. for strict local min: If the feasible solution $\bar{x}$ is such that there exists a Lagrange multiplier vector $\bar{\pi} = (\bar{\pi}_1, \ldots, \bar{\pi}_{m+p})$ which together with $\bar{x}$ satisfies

$$\nabla_x \theta(\bar{x}) = \sum_{i=1}^{m+p} \bar{\pi}_i \nabla_x g_i(\bar{x})$$

$$\bar{\pi}_i \geq 0 \text{ for } i \in \{m+1, \ldots, m+p\}$$

$$\bar{\pi} g_i(\bar{x}) = 0 \text{ for } i \in \{m+1, \ldots, m+p\}$$

$$\text{and } y^T(\nabla^2_{xx} L(\bar{x}, \bar{\pi}))y > 0 \text{ for all } y \in T_2$$

where $T_2 = \{y : \nabla_x g_i(\bar{x})y = 0$ for all $i \in \{1, \ldots, m\} \cup (\{i : \bar{\pi}_i > 0\} \cap P(\bar{x})\}$; and $\nabla_x g_i(\bar{x})y \geq 0$ for all $i \in \{i : \bar{\pi}_i = 0\} \cap P(\bar{x})\}$.

If (3) is a nonconvex program, verifying whether the last condition among the sufficient optimality conditions holds is hard (Murty and Kabadi [1987]).

A weaker sufficient condition for $\bar{x}$ to be a strict local minimum for (3) is obtained by replacing T_2 in the above condition by the set $T_3 = \{y : \nabla_x g_i(\bar{x})y = 0$ for all $i \in \{1, \ldots, m\} \cup (\{i : \bar{\pi}_i > 0\} \cap P(\bar{x})\}$. This weaker sufficient condition can be checked efficiently.

Nec. and suff. conds. for $\bar{x}$ to be a global minimum if problem is a convex program: The 1st order nec. conds. are nec. and suff. for a global min.

Example: Determining Electricity Flows In above electrical network, current flows in direction of arrow on each arc. Flow variables entered on arc need to be determined. No. on arc j is its resistence r_j. Current flows occur to min. power loss $= \Sigma\, r_j x_j^2$. Solve using an **active set strategy** and 1st order conds.