Second Order Optimality Conditions for Constrained Nonlinear Programming

Lecture 10, Continuous Optimisation

Oxford University Computing Laboratory, HT 2006

Notes by Dr Raphael Hauser (hauser@comlab.ox.ac.uk)

Additions by Dr Nick Gould (n.gould@rl.ac.uk)

We again consider the general nonlinear optimisation problem

$$
\begin{aligned}
\text{(NLP)} \quad & \min_{x\in\mathbb{R}^n} f(x) \\
& \text{s.t. } g_i(x) = 0 \quad (i \in \mathcal{E}), \\
& \phantom{\text{s.t. }} g_i(x) \geq 0 \quad (i \in \mathcal{I}).
\end{aligned}
$$

We will now derive second order optimality conditions for (NLP).

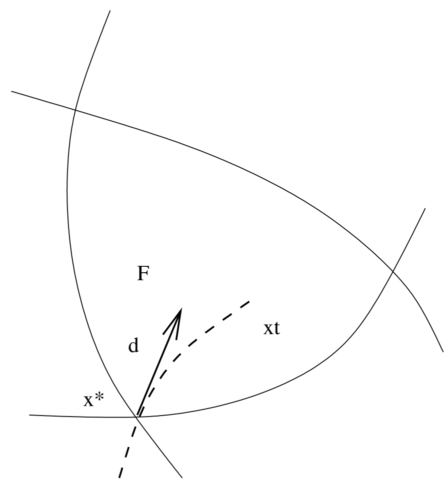
For that purpouse, we assume that f and the g_i $(i \in \mathcal{E} \cup \mathcal{I})$ are *twice* continuously differentiable functions.

Definition 1: Let $x^* \in \mathbb{R}^n$ be a feasible point for (NLP) and let $x \in C^2\Big((-\epsilon,\epsilon),\mathbb{R}^n\Big)$ be a path such that

$$
\begin{aligned}
x(0) &= x^*, \\
d &:= \frac{d}{dt}x(0) \neq 0, \\
g_i(x(t)) &= 0 \qquad (i \in \mathcal{E}, t \in (-\epsilon,\epsilon)), \\
g_i(x(t)) &\geq 0 \qquad (i \in \mathcal{I}, t \in [0,\epsilon)).
\end{aligned}
\tag{1}
$$

Thus, we can imagine that $x(t)$ is a smooth piece of trajectory of a point particle that passes through x^* at time $t = 0$ with nonzero speed d and moves into the feasible domain.

We call $x(t)$ a *feasible exit path* from x^* and the tangent vector $d = \frac{d}{dt}x(0)$ a *feasible exit direction* from x^*.



The second order optimality analysis is based on the following observation:

If x^* is a local minimiser of (NLP) and $x(t)$ is a feasible exit path from x^* then x^* must also be a local minimiser for the univariate constrained optimisation problem

$$\begin{aligned} &\min f(x(t)) \\ &\text{s.t. } t \geq 0 \end{aligned}$$

Before we start looking at such problems more closely, we develop an alternative characterisation of feasible exit directions from x^*.

On the other hand, if the LICQ holds at x^* then Lemma 1 of Lecture 9 shows that (2) implies the existence of a feasible exit path from x^* such that

$$\frac{d}{dt}x(0) = d, \tag{3}$$
$$g_i(x(t)) = td^{\mathsf{T}}\nabla g_i(x^*) \quad (i \in \mathcal{E} \cup \mathcal{A}(x^*). \tag{4}$$

Thus, when the LICQ holds then (2) is also a *sufficient* condition and hence an exact characterisation for d to be a feasible exit path from x^*.

Definition 1 implies

$$d^{\mathsf{T}}\nabla g_i(x^*) = \frac{d}{dt}g_i(x(t))|_{t=0} = \begin{cases} \frac{d}{dt}0 = 0 & (i \in \mathcal{E}), \\ \lim_{t\to 0+} \frac{g_i(x(t))-0}{t} \geq 0 & (i \in \mathcal{A}(x^*)). \end{cases}$$

Therefore, the following are *necessary* conditions for $d \in \mathbb{R}^n$ to be a feasible exit direction from x^*:

$$\begin{aligned} d &\neq 0, \\ d^{\mathsf{T}}\nabla g_i(x^*) &= 0 \quad (i \in \mathcal{E}), \\ d^{\mathsf{T}}\nabla g_j(x^*) &\geq 0 \quad (j \in \mathcal{A}(x^*)). \end{aligned} \tag{2}$$

Second Order Necessary Optimality Conditions

Let x^* be a local minimiser of (NLP) where the LICQ holds. The KKT conditions say that there exists a vector λ^* of Lagrange multipliers such that

$$\begin{aligned} D_x\mathcal{L}(x^*, \lambda^*) &= 0, \\ \lambda_j^* &\geq 0 \quad (j \in \mathcal{I}), \\ \lambda_i^* g_i(x^*) &= 0 \quad (i \in \mathcal{E} \cup \mathcal{I}), \\ g_j(x^*) &\geq 0 \quad (j \in \mathcal{I}), \\ g_i(x^*) &= 0 \quad (i \in \mathcal{E}), \end{aligned} \tag{5}$$

where $\mathcal{L}(x, \lambda) = f(x) - \sum_i \lambda_i g_i$ is the Lagrangian associated with (NLP).

Now let $x(t)$ be a feasible exit path from x^* with exit direction d, and let us consider the restricted problem

$$\begin{aligned} &\min f(x(t)) \\ &\text{s.t. } t \geq 0 \end{aligned} \tag{6}$$

Since x^* is a local minimiser of (NLP), $t = 0$ must be a local minimiser of (6).

By Taylor's theorem and the KKT conditions,

$$\begin{aligned} f(x(t)) &= f(x^*) + td^\mathsf{T} \nabla f(x^*) + O(t^2) \\ &= f(x^*) + t \sum_{i=1}^{m} \lambda_i^* d^\mathsf{T} \nabla g_i(x^*) + O(t^2). \end{aligned}$$

We thus wish to show that for small $t \geq 0$,

$$t \sum_{i=1}^{m} \lambda_i^* d^\mathsf{T} \nabla g_i(x^*) + O(t^2) \geq 0. \tag{7}$$

Note that

$$\lambda_i^* d^\mathsf{T} \nabla g_i(x^*) = 0 \quad (i \in \mathcal{E} \cup \mathcal{I} \setminus \mathcal{A}(x^*)),$$

so that these terms can be omitted from (7).

But what about indices $j \in \mathcal{A}(x^*)$? We have to distinguish two different cases:

Case 1: there exists an index $j \in \mathcal{A}(x^*)$ such that $d^\mathsf{T} \nabla g_j(x^*) > 0$.

Then for all $0 < t \ll 1$,

$$\begin{aligned} f(x(t)) &= f(x^*) + t \sum_{i=1}^{m} \lambda_i^* d^\mathsf{T} \nabla g_i(x^*) + O(t^2) \\ &\geq f(x^*) + t\lambda_j^* d^\mathsf{T} \nabla g_j(x^*) + O(t^2) \\ &> f(x^*). \end{aligned}$$

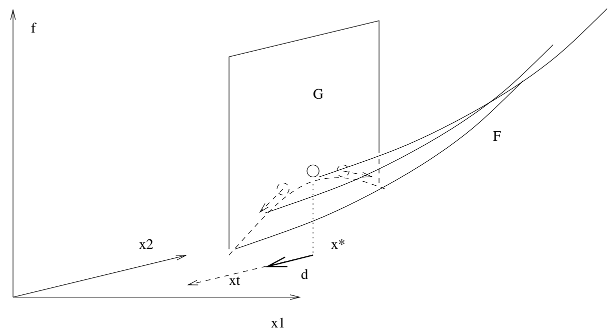
Thus, in this case f strictly increases along the path $x(t)$ for small positive t even if $\frac{d^2}{dt^2} f(x(0))$ was negative. Because of the constraint g_j, nothing can be said about the $D^2_{xx} f(x^*) d$.

Case 2:

$$\lambda_i^* d^\mathsf{T} \nabla g_i(x^*) = 0 \quad (i \in \mathcal{I} \cup \mathcal{E}). \tag{8}$$

In this case the above argument fails to guarantee that f locally increases along path $x(t)$. We only know that $d/dt\, f(x(0)) = 0$, that is, x^* is a stationary point of (6).

But this might very well be a local maximiser of the restricted problem. Second order derivatives $\frac{d^2}{dt^2} f(x(0))$ now decide whether $t = 0$ is a local minimiser of the restricted problem (6), yielding additional necessary information in this case!



Theorem 1: 2nd Order Necessary Optimality Conditions. Let x^* be a local minimiser of (NLP) where the LICQ holds. Let $\lambda^* \in \mathbb{R}^m$ be a Lagrange multiplier vector such that (x^*, λ^*) satisfy the KKT conditions. Then we have

$$d^{\mathsf{T}} D_{xx}\mathcal{L}(x^*, \lambda^*) d \geq 0 \tag{9}$$

for all feasible exit directions d from x^* that satisfy (8).

Proof:

- Let $d \neq 0$ satisfy (2) and (8), and let $x \in C^2\Big((-\epsilon, \epsilon), \mathbb{R}^n\Big)$ be a feasible exit path from x^* corresponding to d.

- Then

$$\mathcal{L}\Big(x(t), \lambda^*\Big) \overset{(4)}{=} f(x(t)) - \sum_{i=1}^{m} \lambda_i^* t d^{\mathsf{T}} \nabla g_i(x^*) \overset{(8)}{=} f(x(t)).$$

- Therefore, Taylor's theorem implies

$$\begin{aligned} f(x(t)) &= \mathcal{L}(x^*, \lambda^*) + t D_x\mathcal{L}(x^*, \lambda^*) d \\ &\quad + \frac{t^2}{2}\Big(d^{\mathsf{T}} D_{xx}\mathcal{L}(x^*, \lambda^*) d + D_x\mathcal{L}(x^*, \lambda^*) \frac{d^2}{dt^2} x(0)\Big) + O(t^3) \\ &\overset{\text{KKT}}{=} f(x^*) + \frac{t^2}{2} d^{\mathsf{T}} D_{xx}\mathcal{L}(x^*, \lambda^*) d + O(t^3). \end{aligned}$$

- If it were the case that $d^{\mathsf{T}} D_{xx}\mathcal{L}(x^*, \lambda^*) d < 0$ then $f(x(t)) < f(x^*)$ for all t sufficiently small, contradicting the assumption that x^* is a local minimiser. Therefore, it must be the case that $d^{\mathsf{T}} D_{xx}\mathcal{L}(x^*, \lambda^*) d \geq 0$. $\square$

Sufficient Optimality Conditions:

In unconstrained minimisation we found that strengthening the second order condition $D^2 f(x) \succeq 0$ to $D^2 f(x) \succ 0$ led to sufficient optimality conditions.

Does the same happen when we change the inequality in (9) to a strict inequality? Our next result shows that this is indeed the case.

Theorem: Sufficient Optimality Conditions.
Let $(x^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^m$ be such that the KKT conditions (5) hold, the LICQ holds, and

$$d^{\mathsf{T}} D_{xx} \mathcal{L}(x^*, \lambda^*) d > 0$$

for all feasible exit directions $d \in \mathbb{R}^n$ from x^* that satisfy

$$\lambda_i^* d^{\mathsf{T}} \nabla g_i(x^*) = 0 \quad (i \in \mathcal{I} \cup \mathcal{E}).$$

Then x^* is a strict local minimiser.

There are two issues that need to be addressed in the proof:

- The first is that x^* is a strict local minimiser for the restricted problem (6). This is easy to prove using Taylor expansions.

- The second, more delicate issue is to show that it suffices to look at the univariate problems (6) for all possible feasible exit paths from x^*.

Proof:

- Let us assume to the contrary of our claim that x^* is not a local minimiser.

- Then there exists a sequence of feasible points $(x_k)_{\mathbb{N}}$ such that $\lim_{k\to\infty} x_k = x^*$ and
$$f(x_k) \leq f(x^*) \quad \forall k \in \mathbb{N}. \tag{10}$$

- The sequence $\frac{x_k - x^*}{\|x_k - x^*\|}$ lies on the unit sphere which is a compact set. The Bolzano–Weierstrass theorem therefore implies that we can extract a subsequence $(x_{k_i})_{i\in\mathbb{N}}$, $k_i < k_j$

$(i < j)$, such that the limiting direction $d := \lim_{k\to\infty} d_{k_i}$ exists, where

$$d_{k_i} = \frac{x_{k_i} - x^*}{\|x_{k_i} - x^*\|}.$$

- Since d lies on the unit sphere we have $d \neq 0$. Replacing the old sequence by the new one we may assume without loss of generality that $k_i \equiv i$.

- Let us check that d satisfies the conditions

$$\begin{aligned} d &\neq 0, \\ d^\mathsf{T} \nabla g_i(x^*) &= 0 \quad (i \in \mathcal{E}), \\ d^\mathsf{T} \nabla g_j(x^*) &\geq 0 \quad (j \in \mathcal{A}(x^*)). \end{aligned} \tag{11}$$

and hence is a feasible exit direction:

$$\begin{aligned} d^\mathsf{T} \nabla g_j(x^*) &= \lim_{i\to\infty} \frac{g_j(x_i) - g_j(x^*)}{\|x_i - x^*\|} \\ &= \begin{cases} \lim_{i\to\infty} 0 = 0 & (j \in \mathcal{E}), \\ \lim_{i\to\infty} \frac{g_j(x_i) - 0}{\|x_i - x^*\|} \geq 0 & (j \in \mathcal{A}(x^*)). \end{cases} \end{aligned}$$

- By Taylor's theorem,

$$f(x^*) \geq f(x_k) = f(x^*) + \|x_k - x^*\| \nabla f(x^*)^\mathsf{T} d_k + O(\|x_k - x^*\|^2).$$

Therefore,

$$\nabla f(x^*)^\mathsf{T} d = \lim_{k\to\infty} \nabla f(x^*)^\mathsf{T} d_k \leq 0. \tag{12}$$

- On the other hand, the KKT conditions and (11) imply

$$d^\mathsf{T} \nabla f(x^*) = \sum_{i=1}^m \lambda_i^* d^\mathsf{T} \nabla g_i(x^*) \geq 0. \tag{13}$$

- But (12) and (13) can be jointly true only if

$$\lambda_i^* d^\mathsf{T} \nabla g_i(x^*) = 0 \quad (i \in \mathcal{I} \cup \mathcal{E}).$$

- The assumption of the theorem therefore implies that

$$d^\mathsf{T} D_{xx} \mathcal{L}(x^*, \lambda^*) d > 0. \tag{14}$$

- On the other hand,

$$\begin{aligned} f(x^*) &\geq f(x_k) \\ &\overset{\text{KKT}}{\geq} f(x_k) - \sum_{i=1}^m \lambda_i^* g_i(x_k) \quad (\text{since } \lambda_i^* \geq 0 \text{ for } i \in \mathcal{I} \\ &\qquad\qquad\qquad\qquad\qquad\quad \text{and } x_k \text{ is feasible}) \\ &= \mathcal{L}(x_k, \lambda^*) \\ &= \mathcal{L}(x^*, \lambda^*) + \|x_k - x^*\| D_x \mathcal{L}(x^*, \lambda^*) d_k^\mathsf{T} \\ &\quad + \frac{\|x_k - x^*\|^2}{2} d_k^\mathsf{T} D_{xx} \mathcal{L}(x^*, \lambda^*) d_k + O(\|x_k - x^*\|^3) \\ &\overset{\text{KKT}}{=} f(x^*) + \frac{\|x_k - x^*\|^2}{2} d_k^\mathsf{T} D_{xx} \mathcal{L}(x^*, \lambda^*) d_k + O(\|x_k - x^*\|^3), \end{aligned}$$

or

$$d_k^\mathsf{T} D_{xx} \mathcal{L}(x^*, \lambda^*) d_k \leq |O(\|x_k - x^*\|)|.$$

- Taking limits, we obtain

$$d^\top D_{xx}\mathcal{L}(x^*,\lambda^*)d = \lim_{k\to\infty} d_k^\top D_{xx}\mathcal{L}(x^*,\lambda^*)d_k \leq 0.$$

- Since this contradicts (14), our assumption about the existence of the sequence $(x_k)_{\mathbb{N}}$ must have been wrong. $\square$

Reading Assignment: Lecture-Note 10.