SECTION C: CONTINUOUS OPTIMISATION
LECTURE 10: SECOND ORDER OPTIMALITY CONDITIONS FOR CONSTRAINED NONLINEAR PROGRAMMING

HONOUR SCHOOL OF MATHEMATICS, OXFORD UNIVERSITY
HILARY TERM 2005, DR RAPHAEL HAUSER

1. Second Order Conditions. We will now derive second order optimality conditions for the smooth constrained optimisation problem (NLP), and we will therefore assume that f and the g_i $(i = 1, \ldots, m)$ are twice continuously differentiable functions.

1.1. Feasible Exit Paths. Our analysis will be based on the notion of feasible exit directions and exit paths:

DEFINITION 1.1. *Let $x^* \in \mathbb{R}^n$ be a feasible point and let $x \in C^2\big((-\epsilon, \epsilon), \mathbb{R}^n\big)$ be a path such that*

$$\begin{aligned} x(0) &= x^*, \\ d = \frac{d}{dt}x(0) &\neq 0, \\ g_i(x(t)) &= 0 \qquad (i \in \mathcal{E}, t \in (-\epsilon, \epsilon)), \\ g_i(x(t)) &\geq 0 \qquad (i \in \mathcal{I}, t \in [0, \epsilon)). \end{aligned} \tag{1.1}$$

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, see Figure 1.1. 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^*.*

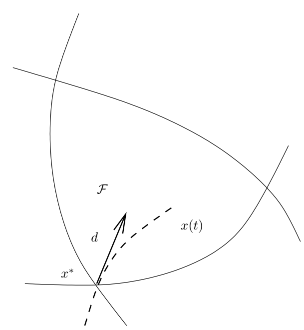

FIG. 1.1. *A feasible exit path from x^*.*

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

$$\begin{aligned} &\min f(x(t)) \\ &\text{s.t. } \; t \geq 0 \end{aligned} \tag{1.2}$$

Before we start looking at such problems more closely, let us characterise the set of feasible exit directions from x^*. Note that (1.1) implies

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

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

$$\begin{aligned} d &\neq 0, \\ d^{\mathrm{T}}\nabla g_i(x^*) &= 0 \quad (i \in \mathcal{E}), \\ d^{\mathrm{T}}\nabla g_j(x^*) &\geq 0 \quad (j \in \mathcal{A}(x^*)). \end{aligned} \tag{1.3}$$

On the other hand, if the LICQ holds at x^* then Lemma 2.3 of Lecture 9 shows that (1.3) implies the existence of a feasible exit path from x^* such that

$$\frac{d}{dt}x(0) = d, \tag{1.4}$$

$$g_i(x(t)) = td^{\mathrm{T}}\nabla g_i(x^*) \quad (i \in \mathcal{E} \cup \mathcal{A}(x^*). \tag{1.5}$$

Thus, when the LICQ holds then (1.3) is also a sufficient condition for d to be a feasible exit path from x^*.

1.2. Necessary Second Order Optimality Conditions. Let x^* be a local minimiser of (NLP) where the LICQ holds, let $x(t)$ be a feasible exit path from x^* with exit direction d, and let us consider the restricted problem (1.2).

The first order necessary optimality conditions of Lecture 9 say that there exists a vector λ^* of Lagrange multipliers such that (x^*, λ^*) satisfies the KKT conditions

$$\begin{aligned} D_x\mathcal{L}(x^*, \lambda^*) &= 0, \\ \lambda_j^* &\geq 0 \quad (j \in \mathcal{I}), \\ \lambda_i^* g_i(x^*) &= 0 \quad (i \in \mathcal{E} \cup \mathcal{I}), \\ g_j(x^*) &\geq 0 \quad (j \in \mathcal{I}), \\ g_i(x^*) &= 0 \quad (i \in \mathcal{E}), \end{aligned} \tag{1.6}$$

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

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

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

In the first case there exists an index $j \in \mathcal{A}(x^*)$ such that $d^{\mathrm{T}}\nabla g_j(x^*) > 0$. Taylor's theorem shows

$$\begin{aligned} f(x(t)) &= f(x^*) + td^{\mathrm{T}}\nabla f(x^*) + O(t^2) \\ &\overset{\mathrm{KKT}}{=} f(x^*) + t\sum_{i=1}^{m} \lambda_i^* d^{\mathrm{T}}\nabla g_i(x^*) + O(t^2) \\ &\geq f(x^*) + t\lambda_j^* d^{\mathrm{T}}\nabla g_j(x^*) + O(t^2) \\ &> f(x^*) \qquad \forall\, 0 < t \ll 1. \end{aligned}$$

Thus, in this case f strictly increases along the path $x(t)$ for small positive t.

In the second case we have

$$\lambda_i^* d^{\mathrm{T}} \nabla g_i(x^*) = 0 \quad (i \in \mathcal{I} \cup \mathcal{E}) \tag{1.7}$$

and the above argument fails to show 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 (1.2). But this might very well be a local maximiser of the restricted problem, see Figure 1.2. Just like in unconstrained optimisation, second order derivatives yield more information in this case.

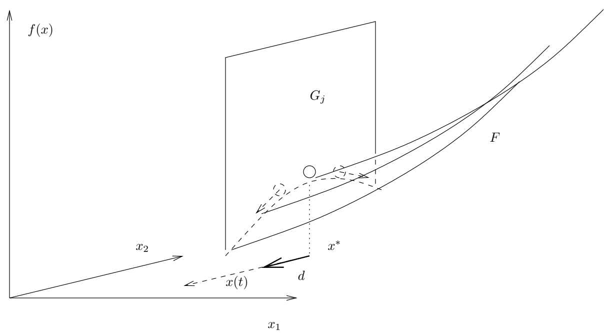

Fig. 1.2. *Mechanistic interpretation of second order conditions: the dotted line shows the graph of f restricted to a feasible exit path from x^* (position of the solid ball) satisfying* (1.7)*. The ball is not at a local minimiser for* (1.2)*. On the other hand, for all feasible exit paths pointing strictly away from the constraint surface G_j the ball sits at a strict local minimiser of the restricted problem* (1.2).

Theorem 1.2 (Second Order Necessary 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^*) satisfies the KKT conditions. Then we have*

$$d^{\mathrm{T}} D_{xx} \mathcal{L}(x^*, \lambda^*) d \geq 0 \tag{1.8}$$

for all feasible exit directions d from x^* that satisfy (1.7).

Proof. Let $d \neq 0$ satisfy (1.3) and (1.7), and let $x \in C^2\big((-\epsilon, \epsilon), \mathbb{R}^n\big)$ be a feasible exit path from x^* corresponding to d. Then

$$\mathcal{L}\big(x(t), \lambda^*\big) \overset{(1.5)}{=} f(x(t)) - \sum_{i=1}^{m} \lambda_i^* t d^{\mathrm{T}} \nabla g_i(x^*) \overset{(1.7)}{=} f(x(t)).$$

Therefore, Taylor's theorem implies

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

If it were the case that $d^{\mathrm{T}} D_{xx} \mathcal{L}(x^*, \lambda^*) d < 0$ 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 $d^{\mathrm{T}} D_{xx} \mathcal{L}(x^*, \lambda^*) d \geq 0$. ☐

1.3. 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 (1.8) to a strict inequality? Our next result shows that this is indeed the case.

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 (1.2). 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 (1.2) for all possible feasible exit paths from x^*.

THEOREM 1.3 (Sufficient Optimality Conditions). *Let* $(x^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^m$ *be such that the KKT conditions* (1.6) *hold, the LICQ holds, and*

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

for all feasible exit directions $d \in \mathbb{R}^n$ *from* x^* *that satisfy* (1.7). *Then* x^* *is a strict local minimiser.*

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

$$f(x_k) \leq f(x^*) \quad \forall\, k \in \mathbb{N}. \tag{1.9}$$

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 $k_i \equiv i$. One can check that d must satisfy the conditions (1.3) the same way as for feasible exit directions. By Taylor's theorem,

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

Therefore,

$$\nabla f(x^*)^{\mathrm{T}} d = \lim_{k\to\infty} \nabla f(x^*)^{\mathrm{T}} d_k \leq 0. \tag{1.10}$$

On the other hand, the KKT conditions imply

$$d^{\mathrm{T}} \nabla f(x^*) = \sum_{i=1}^{m} \lambda_i^* d^{\mathrm{T}} \nabla g_i(x^*) \geq 0. \tag{1.11}$$

But (1.10) and (1.11) can be jointly true only if (1.7) holds. The assumption of the theorem therefore implies that

$$d^{\mathrm{T}} D_{xx} \mathcal{L}(x^*, \lambda^*) d > 0. \tag{1.12}$$

On the other hand,

$$\begin{aligned}
f(x^*) &\geq f(x_k) \\
&\overset{\text{KKT}}{\geq} f(x_k) - \sum_{i=1}^{m} \lambda_i^* g_i(x_k) \\
&= \mathcal{L}(x_k, \lambda^*) \\
&= \mathcal{L}(x^*, \lambda^*) + \|x_k - x^*\| D_x \mathcal{L}(x^*, \lambda^*) d_k^{\mathrm{T}} + \frac{\|x_k - x^*\|^2}{2} d_k^{\mathrm{T}} D_{xx} \mathcal{L}(x^*, \lambda^*) d_k \\
&\qquad + O(\|x_k - x^*\|^3) \\
&\overset{\text{KKT}}{=} f(x^*) + \frac{\|x_k - x^*\|^2}{2} d_k^{\mathrm{T}} D_{xx} \mathcal{L}(x^*, \lambda^*) d_k + O(\|x_k - x^*\|^3),
\end{aligned}$$

or

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

Taking limits, we obtain

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

Since this contradicts (1.12), our assumption about the existence of the sequence $(x_k)_{\mathbb{N}}$ must have been wrong. □