CHAPTER 13 SUFFICIENCY (OUR BEST 560 CONTRIBUTION).

13.1 THE FUNDAMENTAL THEOREM OF NONLINEAR PROGRAMMING

RECALL:

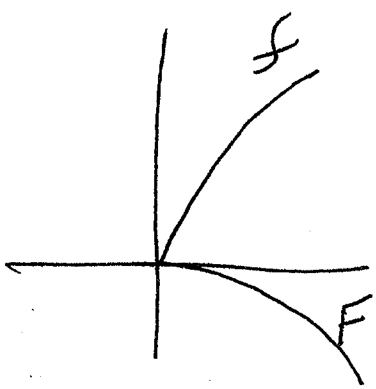
DEF 3.10.3 (LOCAL SUPPORT FUNCTION)

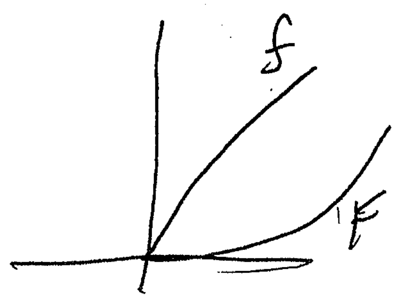
F LOWER SUPPORT FUNCTION FOR f AT x^* IF

(i) $F(x) \leq f(x) \quad \forall x \in S \cap N(x^*)$

(ii) $F(x^*) = f(x^*)$

(iii) $F'(x^*) = 0$





OBSERVATION: x^* MINIMIZES F

$$\Downarrow \quad \not\Uparrow$$

x^* MINIMIZES f

PROPOSITION 13.1.1
~~Let choose the~~ LET (x^*, λ^*, μ^*) BE A KKT OF PROB NLP

THEN $F(x) = l(x, \lambda^*, \mu^*)$ IS A LOCAL SUPPORT

FUNCTION FOR f.

PROOF.

$$l(x, \lambda^*, \mu^*) = f(x) + \Sigma \lambda_i \overset{=0}{h_i(x)} - \Sigma \mu_i \overset{\geq 0}{g_i(x)}$$

$$\leq f(x)$$

$$l(x^*, \lambda^*, \mu^*) = f(x^*)$$

$$l'_x(x^*, \lambda^*, \mu^*) = 0$$

THEOREM 13.1.2 (THE FUNDAMENTAL THEOREM OF NONLINEAR PROG)

LET (x^*, λ^*, μ^*) BE A KKT POINT OF PROB NLP.

IF x^* IS A (LOCAL, GLOBAL, STRICT) MINIMIZER OF $F(x) = \ell(x, \lambda^*, \mu^*)$

THEN x^* IS A (LOCAL, GLOBAL, STRICT) MINIMIZER OF f OVER S.

PROOF $\quad f(x^*) = F(x^*) \leq F(x) \leq f(x)$

4

SECTION 13.2 SUFFICIENCY FOR NONLINEAR PROGRAMMING.
RECALL:

3.12 Fundamental Principles for Second-Order Sufficiency

Theorem 3.12.1 (Fundamental Principles for Second-Order Sufficiency). *Consider Problem* (3.1) *and* $x^* \in S$.
Assume

(a) X *is a normed linear space.*

(b) f *is Fréchet differentiable in a neighborhood of* x^* *and has a second Fréchet derivative at* $x*$.

(c) There exists a local lower support function F *for* f *at* x^*.

Then

(i) The condition that for each $\{x_k \neq x^*\} \subset S$ *converging to* x^* *satisfying*

$$\lim_k f'(x^*)\left(\frac{x_k - x^*}{\|x_k - x^*\|}\right) = 0, \tag{3.62}$$

it follows that

$$\liminf_k F''(x^*)\left(\frac{x_k - x^*}{\|x_k - x^*\|}, \frac{x_k - x^*}{\|x_k - x^*\|}\right) > 0, \tag{3.63}$$

is sufficient for x^* *to be a strict local minimizer of* f *in* S.

Moreover, if $X = \mathbb{R}^n$, *then statement (i) is equivalent to the statement*

(ii) The condition that for each nonzero $z \in T_s(S, x^*)$ *satisfying*

$$f'(x^*)(z) = 0 \tag{3.64}$$

it follows that

$$F''(x^*)(z, z) > 0, \tag{3.65}$$

is sufficient for x^* *to be a strict local minimizer of* f *in* S.

Example 3.12.2. *We wish to see what our fundamental principle for second-order sufficiency, Theorem 3.12.1, gives with respect to unconstrained minimization in $\mathbb{R}^n$. Consider $f : \mathbb{R}^n \to \mathbb{R}$ and $x^* \in \mathbb{R}^n$. Assume that f has a Fréchet derivative in a neighborhood of x^* and has a second Fréchet derivative at x^*. Our first-order necessary condition is that $\nabla f(x^*) = 0$. Hence, we let f serve as its own support function. This means that one of our assumptions must be $\nabla f(x^*) = 0$. This assumption is not restrictive since it must hold if sufficiency holds. It is clear that $T_\ell(\mathbb{R}^n, x^*) = \mathbb{R}^n$; hence $T_s(\mathbb{R}^n, x^*) = \mathbb{R}^n$, since the latter tangent cone contains the former. We see from part (ii) of Theorem 3.12.1 that x^* will be a strict local minimizer if*

$$f''(x^*)(z,z) = \langle \nabla^2 f(x^*)z, z\rangle > 0 \quad \text{for all } z \neq 0.$$

Hence, sufficiency conditions for x^* to be a strict local minimizer of $f : \mathbb{R}^n \to \mathbb{R}$ are

(i) f is Fréchet differentiable in a neighborhood of x^* and has a second Fréchet derivative at x^*,

(ii) $\nabla f(x^*) = 0$, and

(iii) $\nabla^2 f(x^*)$, the Hessian matrix of f at x^*, is positive definite.

RECALL PROPOSITION 13.1.1 THAT SAYS ~~THE~~ THAT FOR PROBLEM NLP, IF (x^*, λ^*, μ^*) IS A KKT POINT, THEN THE LAGRANGIAN

$$F(x) = \ell(x, \lambda^*, \mu^*)$$

IS A LOWER SUPPORT FUNCTION FOR f.

Sufficiency For Nonlinear Programming Problems in $\mathbb{R}^n$

Theorem 13.2.1 ***Application Theorem One: Sufficiency for Constrained Optimization in $\mathbb{R}^n$.*** *Consider Problem (NLP), where $f, h_i, g_i : \mathbb{R}^n \to \mathbb{R}$ and are twice continuously differentiable. Sufficient conditions for x^* to be a strict local solution of Problem (NLP) are:*

(i) There exists λ^*, μ^* such that (x^*, λ^*, μ^*) is a KKT point.

(ii) Whenever $z \neq 0$ is such that

$$\begin{aligned} z^T \nabla g_i(x^*) &\geq 0 \text{ for } i \in B(x^*) \\ z^T \nabla g_i(x^*) &= 0 \text{ for } i \in B(x^*, \mu^*) = \{i \mid \mu_i^* > 0\} \\ z^T \nabla h_i(x^*) &= 0 \text{ for } i = 1, \ldots, m, \end{aligned}$$

then $z^T \nabla_x^2 \ell(x^*, \lambda^*, \mu^*)\, z > 0$.

PROOF. OUR TASK IS TO SHOW THAT THE IMPLICATION GIVEN IN (ii) OF OUR FUNDAMENTAL PRINCIPLE FOR SECOND-ORDER SUFFICIENCY, THEOREM 3.12.1, HOLDS USING THE LAGRANGIAN AS SUPPORT FUNCTION WHEN ~~TO~~ THE CONDITIONS GIVEN IN THEOREM 13.2.1 HOLD. SO WE ASSUME THE CONDITIONS OF THE LATTER THEOREM AND CONSIDER THE HYPOTHESIS IN (ii) OF THE FORMER THEOREM. SPECIFICALLY CONSIDER $\eta \in T_{0}(S, x^*)$ SATISFYING $\eta^T \nabla f(x^*) = 0$.

WE KNOW THAT

~~NOW KNOW THAT~~ $z = \|z\| \lim_k \dfrac{(x_k - x^*)}{\|x_k - x^*\|}$

for a sequence $\{x_k\}$ satisfying

(a) x_k is feasible

(b) $x_k \neq x^*$

(c) $x_k \to x^*$.

Without loss of generality ~~we~~ assume $\|z\| = 1$.

Let $z_k = \dfrac{(x_k - x^*)}{\|x_k - x^*\|}$ and let $\delta_k = \|x_k - x^*\|$, so that

$$\delta_k z_k = x_k - x^*.$$

NOW

$$\frac{g_i(x^* + \delta_k z_k) - g_i(x^*)}{\delta_k} \;\geq\; 0,$$

therefore $\nabla g_i(x^*)^T z \geq 0$ for $i \in B(x^*)$. Also,

$$\frac{h_i(x^* + \delta_k z_k) - h_i(x^*)}{\delta_k} \;=\; 0,$$

so $\nabla h_i(x^*)^T z = 0$ for $i = 1, \ldots, m$. It follows that

$$0 \;=\; \nabla f(x^*)^T z \;=\; \sum_{i \in B(x^*)} \mu_i^* \, \nabla g_i(x^*)^T z.$$

Thus, $\nabla g_i(x^*)^T z = 0$ for $i \in$ ~~$B(x^*)$~~ $B(x^*, \mu^*)$ and ~~$\eta = \bar{\eta}$ and~~ all of the ~~assumptions~~ conditions in (ii) of Theorem 13.2.1 are satisfied. Hence, we have

$$z^T \nabla_x^2 L(x^*, \lambda^*, \mu^*) z \;>\; 0$$

and

Our general principle for sufficiency holds. □

√

§ 9

13.3 Sufficiency For Nonlinear Programming Problems in Hilbert Space

In this section, we study sufficiency results in infinite dimensions. First, we consider the case of an unconstrained optimization problem.

Theorem 13.3.1 ***Application Theorem Two: Sufficiency for Unconstrained Optimization in Hilbert Space.*** *Consider $f : H \to \mathbb{R}$, where H is a Hilbert space. Assume that $f \in C^2$. Sufficient conditions for x^* to be a strict local minimizer of f are*

(i) $\nabla f(x^*) = 0$.

(ii) $\eta^T \nabla^2 f(x^*) \eta \geq c\|\eta\|^2$ *for all* $\eta \neq 0$ *in* H *and some* $c > 0$.

Proof: Clearly (i) and (ii) of Theorem 13.2.1 are satisfied. □

Remark 13.3.2 *If $H = \mathbb{R}^n$, then (ii) of Theorem 13.3.1 is implied by $\eta^T \nabla^2 f(x^*) \eta > 0$ for all $\eta \neq 0$. This fact follows directly from the fact that the Rayleigh quotient $(\eta^T \nabla^2 f(x^*) \eta)/(\eta^T \eta)$ is minimized by the smallest eigenvalue of $\nabla^2 f(x^*)$, which is necessarily positive.*

Consider Problem (~~11.1~~ NLP), where f, h_i, and g_i all have their domains in H. Problem (~~11.1~~ NLP) is now an infinite-dimensional constrained nonlinear program. For the remainder of this section we use the notation

$$
\begin{aligned}
B^* &= \{i : g_i(x^*) = 0\} \\
D^* &= \{i : g_i(x^*) = 0,\ \mu_i^* > 0\} \quad = B(x^*, \mu^*) \\
T^* &= \{d \in H : \nabla h_i(x^*)^T d = 0,\ i = 1, \ldots, m; \\
&\qquad \nabla g_i(x^*)^T d = 0,\ i \in D^*; \\
&\qquad \nabla g_i(x^*)^T d \geq 0,\ i \in B^*\}. \quad = \mathcal{T}\ell(x^*, \mu^*)
\end{aligned}
$$

To the best of my knowledge, the following result is new and not in previous literature.

Theorem 13.3.2 ***Application Theorem Three: Sufficiency for Constrained Optimization in Hilbert Space.*** *Assume Problem (~~11.1~~ NLP) $\in C^2$. Sufficient conditions for x^* to be a strict local solution of Problem (~~11.1~~ NLP) are*

(i) There exist λ^*, μ^* such that ~~$\nabla \ell(x^*, \lambda^*, \mu^*) = 0$ and $\mu_i^* g_i(x^*) = 0$.~~ (x^*, λ^*, μ^*) is a KKT point.

(ii) $d^T \nabla_x^2 \ell(x^*, \lambda^*, \mu^*)\, d \ \geq\ \alpha \|d\|^2$ *for all* $d \in T^*$.

PROOF. OUR TASK IS TO SHOW THAT THE IMPLICATION GIVEN IN (c) OF OUR FUNDAMENTAL PRINCIPLE FOR SECOND-ORDER SUFFICIENCY, THEOREM 3.12.1, HOLDS USING THE LAGRANGIAN AS SUPPORT FUNCTION WHEN THE CONDITIONS GIVEN IN THEOREM 13.2.2 HOLD.

WE BEGIN BY ASSUMING that there exist λ^*, μ^* such THAT (x^*, λ^*, μ^*) IS KKT POINT. ~~that $\nabla_x \ell(x^*, \lambda^*, \mu^*) = 0$.~~ NOW WE ~~ASSUME~~ CONSIDER THE FRONT-END OF THE IMPLICATION (c) IN THEOREM 3.12.1.

Let $\{x_k\}$ be a feasible sequence such that $x_k \neq x^*$, $x_k \to x^*$, and $\lim_k \nabla f(x^*)^T d_k = 0$, where $d_k = (x_k - x^*)/\|x_k - x^*\|$. Let $\sigma_k = \|x_k - x^*\|$.

OUR OBJECTIVE IS TO ESTABLISH (~~3.63~~) THAT (3.63)

$$\liminf_{k} d_k^T \nabla_x^2 \ell(x^*, \lambda^*, \mu^*) d_k > 0.$$

MUST HOLD FOR THIS d_k SEQUENCE. WE ACCOMPLISH THIS BY CONTRADICTION. WE SUPPOSE THAT FOR SOME SUBSEQUENCE OF $\{d_k\}$, ALSO CALLED $\{d_k\}$, WE HAVE

$$\lim_{k\to\infty} d_k^T \nabla_x^2 \ell(x_k, \lambda^*, \mu^*) d_k \le 0. \qquad (13.1)$$

OUR TASK NOW BECOMES TO SHOW THAT THE SUPPOSITION (13.1) CONTRADICTS (ii) OF THEOREM 13.3.2.

12

(I) We know

$$\frac{\overset{=0}{h_i(x^* + \sigma_k d_k)} \; - \; \overset{=0}{h_i(x^*)}}{\sigma_k} \;=\; 0.$$

BY FRÉCHET DIFFERENTION

Therefore,

$$\nabla h_i(x^*)(d_k) \to 0.$$

(II) We also have

$$\frac{g_i(x^* + \sigma_k d_k) \; - \; g_i(x^*)}{\sigma_k} \;\geq\; 0.$$

Thus,

$$\liminf \nabla g_i(x^*) \; d_k \;\geq\; 0 \quad \text{for } i \in B^*.$$

(III) For $i \in D^*$ (i.e. $\mu_i^* > 0$), we have by definition

$$\begin{aligned}\nabla_x \ell(x^*, \lambda^*, \mu^*)^T \, d_k \;&=\; \nabla f(x^*)^T d_k + \sum_{i=1}^{m} \lambda_i^* \, \nabla h_i(x^*)^T \; d_k \\ &\quad - \sum_{i \in D^*} \mu_i^* \, \nabla g_i(x^*)^T \, d_k.\end{aligned}$$

However, since

$$\nabla_x \ell(x^*, \lambda^*, \mu^*) \;=\; 0, \qquad \nabla f(x^*)^T d_k \; \cancel{=} \to \; 0,$$

$$\text{and} \quad \sum_{i=1}^{m} \lambda_i^* \, \nabla h_i(x^*)^T d_k \; \cancel{=} \to \; 0,$$

we know that

$$\sum_{i \in D^*} \mu_i^* \, \nabla g_i(x^*)^T d_k \;\;\to\;\; 0.$$

Thus, $\nabla g_i(x^*)^T d_k \to 0$ for all $i \in D^*$.

(IV) By choosing a subsequence, we have

$$\nabla g_i(x_k)^T d_k \;\to\; \beta_i \;\geq\; 0 \qquad \text{for all } \; i \in B^* - D^*.$$

Let

$$\begin{aligned}\Psi = \{\nabla h_i(x^*) : i = 1, \ldots, m\} \;&\cup\; \{\nabla g_i(x^*) : i \in D^*\} \\ &\cup\; \{\nabla g_i(x^*) : i \in B^* - D^* \text{ and } \beta_i = 0\},\end{aligned}$$

and let $\psi = \text{span}(\Psi)$.

(V) Define $\nu : \psi \to \mathbb{R}^{|\Psi|}$ so that the components of the vector space $\nu(d)$ are

$$\nabla h_i(x^*)^T d \quad for \quad i = 1, \ldots, m,$$

$$\nabla g_i(x^*)^T d \quad for \quad i \in D^*, \text{ and}$$

$$\nabla g_i(x^*)^T d \quad for \quad i \in B^* - D^* \text{ and } \beta_i = 0.$$

Notice that ν is one-to-one, since

$$\nu d = 0 \implies \nu^T d = 0 \text{ for all } \nu \in \psi \implies d^T d = 0 \implies d = 0.$$

Therefore, $\dim(\psi) = \dim(\mathbb{R}(\nu))$, and ν is a map between the finite-dimensional spaces ψ and $\mathbb{R}(\nu)$. Then, there exists $\alpha > 0$ such that $\|\nu d\| \geq \alpha \|d\|$ for all $d \in \psi$. (Note that $\|\nu(\cdot)\|$ achieves its minimum on the unit sphere.)

(VI) Consider the decomposition $H = \psi \oplus \psi^{\perp}$. We can write $d_k = d_k^1 + d_k^2$ uniquely, where $d_k^1 \in \psi$ and $d_k^2 \in \psi^{\perp}$. Now,

$$\nu(d_k) = \nu(d_k^1) + \nu(d_k^2).$$

However, $\nu(d_k^2) = 0$. From (I), (II), (III), and (IV), we have $\nu(d_k) \to 0$, and so $\nu(d_k^1) \to 0$. Also, since $d_k^1 \in \psi$, $\|\nu(d_k^1)\| \geq \alpha \|d_k^1\|$. Then, $d_k^1 \to 0$.

(VII) Now, for $i \in B^* - D^*$ and $\beta_i \neq 0$, we have

$$\nabla g_i(x^*) d_k \to \beta_i \implies \nabla g_i(x^*) d_k^2 \to \beta_i > 0.$$

(Recall $\nabla g_i(x^*)^T d_k^1 \to 0$.) Thus, for k large, $\nabla g_i(x^*)^T d_k^2 > 0$ and $d_k^2 \in T^*$ for k sufficiently large.

(VIII) We have

$$(d_k^2)^T \nabla_x^2 L(x_*, \lambda^*, \mu^*)\, d_k^2 \geq \alpha \|d_k^2\|^2$$

for k large. In the following equations, we use the shortened notation $\nabla_x^2 L(x_*, \lambda^*, \mu^*) = \nabla_x^2 L^*$. Then,

$$\begin{aligned} d_k^T \left(\nabla_x^2 L^*\right) d_k &= (d_k^1)^T \left(\nabla_x^2 L^*\right) d_k^1 + 2 (d_k^1)^T \left(\nabla_x^2 L^*\right) d_k^2 + (d_k^2)^T \left(\nabla_x^2 L^*\right) d_k^2 \\ &\geq \alpha \|d_k^2\|^2 - 2 \|\nabla_x^2 L^*\| \, \|d_k^1\| \, \|d_k^2\| - \|\nabla_x^2 L^*\| \, \|d_k^1\|^2 \\ &= \alpha \|d_k^2\|^2 - \|\nabla_x^2 L^*\| \left\{ 2\|d_k^1\| \, \|d_k^2\| + \|d_k^1\|^2 \right\} \\ &\geq \frac{\alpha}{2} \text{ for } k \text{ large.} \end{aligned}$$

To obtain this result we used the fact that $d_k^1 \to 0$.

Our conclusion $d_k^T \nabla_x^2 \ell(x^*, \lambda^*, \mu^*)\, d_k \geq \alpha/2$ for k large contradicts the supposition

$$\lim_{k \to \infty} d_k^T \nabla_x^2 \ell(x_k, \lambda^*, \mu^*) d_k \leq 0.$$

Thus, our theorem holds. □

REMARK 1.
THIS BEAUTIFUL PROOF IS DUE TO MY FORMER STUDENT

Richard Byrd

REMARK 2. STATEMENTS IN THE LITERATURE SAY THAT THIS THEOREM DOES NOT HOLD WITHOUT FURTHER ASSUMPTIONS.