Chapter 7

Karush–Kuhn–Tucker Theory

7.1 An Introduction to Metric Regularity

Our main optimization models so far are inequality-constrained. A little thought shows our techniques are not useful for equality-constrained problems like

$$\inf\{f(x) \mid h(x) = 0\}.$$

In this section we study such problems by linearizing the feasible region $h^{-1}(0)$ using the contingent cone.

Throughout this section we consider an open set $U \subset \mathbf{E}$, a closed set $S \subset U$, a Euclidean space $\mathbf{Y}$, and a continuous map $h : U \to \mathbf{Y}$. The restriction of h to S we denote $h|_S$. The following easy result (Exercise 1) suggests our direction.

Proposition 7.1.1 *If h is Fréchet differentiable at the point $x \in U$ then*

$$K_{h^{-1}(h(x))}(x) \subset N(\nabla h(x)).$$

Our aim in this section is to find conditions guaranteeing equality in this result.

Our key tool is the next result. It states that if a closed function attains a value close to its infimum at some point then a nearby point minimizes a slightly perturbed function.

Theorem 7.1.2 (Ekeland variational principle) *Suppose the function $f : \mathbf{E} \to (\infty, +\infty]$ is closed and the point $x \in \mathbf{E}$ satisfies $f(x) \leq \inf f + \epsilon$ for some real $\epsilon > 0$. Then for any real $\lambda > 0$ there is a point $v \in \mathbf{E}$ satisfying the conditions* (*a*) $\|x - v\| \leq \lambda$,

(*b*) $f(v) \leq f(x)$, *and*

(*c*) v *is the unique minimizer of the function* $f(\cdot) + (\epsilon/\lambda)\|\cdot - v\|$.

Proof. We can assume f is proper, and by assumption it is bounded below. Since the function

$$f(\cdot) + \frac{\epsilon}{\lambda}\|\cdot - x\|$$

therefore has compact level sets, its set of minimizers $M \subset \mathbf{E}$ is nonempty and compact. Choose a minimizer v for f on M. Then for points $z \neq v$ in M we know

$$f(v) \leq f(z) < f(z) + \frac{\epsilon}{\lambda}\|z - v\|,$$

while for z not in M we have

$$f(v) + \frac{\epsilon}{\lambda}\|v - x\| < f(z) + \frac{\epsilon}{\lambda}\|z - x\|.$$

Part (c) follows by the triangle inequality. Since v lies in M we have

$$f(z) + \frac{\epsilon}{\lambda}\|z - x\| \geq f(v) + \frac{\epsilon}{\lambda}\|v - x\| \text{ for all } z \text{ in } \mathbf{E}.$$

Setting $z = x$ shows the inequalities

$$f(v) + \epsilon \geq \inf f + \epsilon \geq f(x) \geq f(v) + \frac{\epsilon}{\lambda}\|v - x\|.$$

Properties (a) and (b) follow. □

As we shall see, precise calculation of the contingent cone $K_{h^{-1}(h(x))}(x)$ requires us first to bound the distance of a point z to the set $h^{-1}(h(x))$ in terms of the function value $h(z)$. This leads us to the notion of "metric regularity". In this section we present a somewhat simplified version of this idea, which suffices for most of our purposes; we defer a more comprehensive treatment to a later section. We say h is *weakly metrically regular* on S at the point x in S if there is a real constant k such that

$$d_{S \cap h^{-1}(h(x))}(z) \leq k\|h(z) - h(x)\| \text{ for all } z \text{ in } S \text{ close to } x.$$

Lemma 7.1.3 *Suppose* $0 \in S$ *and* $h(0) = 0$. *If* h *is not weakly metrically regular on* S *at zero then there is a sequence* $v_r \to 0$ *in* S *such that* $h(v_r) \neq 0$ *for all* r, *and a strictly positive sequence* $\delta_r \downarrow 0$ *such that the function*

$$\|h(\cdot)\| + \delta_r\|\cdot - v_r\|$$

is minimized on S *at* v_r.

Proof. By definition there is a sequence $x_r \to 0$ in S such that

$$d_{S \cap h^{-1}(0)}(x_r) > r\|h(x_r)\| \text{ for all } r. \tag{7.1.4}$$

For each index r we apply the Ekeland principle with

$$f = \|h\| + \delta_S, \quad \epsilon = \|h(x_r)\|, \quad \lambda = \min\{r\epsilon, \sqrt{\epsilon}\}, \quad \text{and } x = x_r$$

to deduce the existence of a point v_r in S such that

(a) $\|x_r - v_r\| \le \min\left\{r\|h(x_r)\|, \sqrt{\|h(x_r)\|}\right\}$ and

(c) v_r minimizes the function

$$\|h(\cdot)\| + \max\left\{r^{-1}, \sqrt{\|h(x_r)\|}\right\} \|\cdot - v_r\|$$

on S.

Property (a) shows $v_r \to 0$, while (c) reveals the minimizing property of v_r. Finally, inequality (7.1.4) and property (a) prove $h(v_r) \neq 0$. □

We can now present a convenient condition for weak metric regularity.

Theorem 7.1.5 (Surjectivity and metric regularity) *If h is strictly differentiable at the point x in S and*

$$\nabla h(x)(T_S(x)) = \mathbf{Y}$$

then h is weakly metrically regular on S at x.

Proof. Notice first h is locally Lipschitz around x (see Theorem 6.2.3). Without loss of generality, suppose $x = 0$ and $h(0) = 0$. If h is not weakly metrically regular on S at zero then by Lemma 7.1.3 there is a sequence $v_r \to 0$ in S such that $h(v_r) \neq 0$ for all r, and a real sequence $\delta_r \downarrow 0$ such that the function

$$\|h(\cdot)\| + \delta_r\|\cdot - v_r\|$$

is minimized on S at v_r. Denoting the local Lipschitz constant by L, we deduce from the sum rule (6.1.6) and the Exact penalization proposition (6.3.2) the condition

$$0 \in \partial_\circ(\|h\|)(v_r) + \delta_r B + L\partial_\circ d_S(v_r).$$

Hence there are elements u_r of $\partial_\circ(\|h\|)(v_r)$ and w_r of $L\partial_\circ d_S(v_r)$ such that $u_r + w_r$ approaches zero.

By choosing a subsequence we can assume

$$\|h(v_r)\|^{-1}h(v_r) \to y \neq 0$$

and an exercise then shows $u_r \to (\nabla h(0))^* y$. Since the Clarke subdifferential is closed at zero (Section 6.2, Exercise 12), we deduce

$$-(\nabla h(0))^* y \in L\partial_\circ d_S(0) \subset N_S(0).$$

However, by assumption there is a nonzero element p of $T_S(0)$ such that $\nabla h(0)p = -y$, so we arrive at the contradiction

$$0 \geq \langle p, -(\nabla h(0))^* y \rangle = \langle \nabla h(0)p, -y \rangle = \|y\|^2 > 0,$$

which completes the proof. □

We can now prove the main result of this section.

Theorem 7.1.6 (Liusternik) *If h is strictly differentiable at the point x and $\nabla h(x)$ is surjective then the set $h^{-1}(h(x))$ is tangentially regular at x and*

$$K_{h^{-1}(h(x))}(x) = N(\nabla h(x)).$$

Proof. Assume without loss of generality that $x = 0$ and $h(0) = 0$. In light of Proposition 7.1.1, it suffices to prove

$$N(\nabla h(0)) \subset T_{h^{-1}(0)}(0).$$

Fix any element p of $N(\nabla h(0))$ and consider a sequence $x^r \to 0$ in $h^{-1}(0)$ and $t_r \downarrow 0$ in $\mathbf{R}_{++}$. The previous result shows h is weakly metrically regular at zero, so there is a constant k such that

$$d_{h^{-1}(0)}(x^r + t_r p) \leq k\|h(x^r + t_r p)\|$$

holds for all large r, and hence there are points z^r in $h^{-1}(0)$ satisfying

$$\|x^r + t_r p - z^r\| \leq k\|h(x^r + t_r p)\|.$$

If we define directions $p^r = t_r^{-1}(z^r - x^r)$ then clearly the points $x^r + t_r p^r$ lie in $h^{-1}(0)$ for large r, and since

$$\begin{aligned}
\|p - p^r\| &= \frac{\|x^r + t_r p - z^r\|}{t_r} \\
&\leq \frac{k\|h(x^r + t_r p) - h(x^r)\|}{t_r} \\
&\to k\|(\nabla h(0))p\| \\
&= 0,
\end{aligned}$$

we deduce $p \in T_{h^{-1}(0)}(0)$. □

Exercises and Commentary

Liusternik's original study of tangent spaces appeared in [130]. Closely related ideas were pursued by Graves [85] (see [65] for a good survey). The Ekeland principle first appeared in [69], motivated by the study of infinite-dimensional problems where techniques based on compactness might be unavailable. As we see in this section, it is a powerful idea even in finite dimensions; the simplified version we present here was observed in [94]. See also Exercise 14 in Section 9.2. The inversion technique we use (Lemma 7.1.3) is based on the approach in [101]. The recognition of "metric" regularity (a term perhaps best suited to nonsmooth analysis) as a central idea began largely with Robinson; see [162, 163] for example. Many equivalences are discussed in [5, 168].

1. Suppose h is Fréchet differentiable at the point $x \in S$.

   (a) Prove for any set $D \supset h(S)$ the inclusion
   $$\nabla h(x) K_S(x) \subset K_D(h(x)).$$

   (b) If h is constant on S, deduce
   $$K_S(x) \subset N(\nabla h(x)).$$

   (c) If h is a real function and x is a local minimizer of h on S, prove
   $$-\nabla h(x) \in (K_S(x))^-.$$

2. **(Lipschitz extension)** Suppose the real function f has Lipschitz constant k on the set $C \subset \mathbf{E}$. By considering the infimal convolution of the functions $f + \delta_C$ and $k\|\cdot\|$, prove there is a function $\tilde{f} : \mathbf{E} \to \mathbf{R}$ with Lipschitz constant k that agrees with f on C. Prove furthermore that if f and C are convex then $\tilde{f}$ can be assumed convex.

3. * **(Closure and the Ekeland principle)** Given a subset S of $\mathbf{E}$, suppose the conclusion of Ekeland's principle holds for all functions of the form $g + \delta_S$ where the function g is continuous on S. Deduce S is closed. (Hint: For any point x in $\operatorname{cl} S$, let $g = \|\cdot - x\|$.)

4. ** Suppose h is strictly differentiable at zero and satisfies
   $$h(0) = 0,\ v_r \to 0,\ \|h(v_r)\|^{-1} h(v_r) \to y, \text{ and } u_r \in \partial_\circ(\|h\|)(v_r).$$
   Prove $u_r \to (\nabla h(0))^* y$. Write out a shorter proof when h is continuously differentiable at zero.

5. ** Interpret Exercise 27 (Conical open mapping) in Section 4.2 in terms of metric regularity.

6. ** **(Transversality)** Suppose the set $V \subset \mathbf{Y}$ is open and the set $R \subset V$ is closed. Suppose furthermore h is strictly differentiable at the point x in S with $h(x)$ in R and

$$\nabla h(x)(T_S(x)) - T_R(h(x)) = \mathbf{Y}. \tag{7.1.7}$$

(a) Define the function $g : U \times V \to \mathbf{Y}$ by $g(z, y) = h(z) - y$. Prove g is weakly metrically regular on $S \times R$ at the point $(x, h(x))$.

(b) Deduce the existence of a constant k' such that the inequality

$$d_{(S\times R)\cap g^{-1}(g(x,h(x)))}(z, y) \le k' \|h(z) - y\|$$

holds for all points (z, y) in $S \times R$ close to $(x, h(x))$.

(c) Apply Proposition 6.3.2 (Exact penalization) to deduce the existence of a constant k such that the inequality

$$d_{(S\times R)\cap g^{-1}(g(x,h(x)))}(z, y) \le k(\|h(z) - y\| + d_S(z) + d_R(y))$$

holds for all points (z, y) in $U \times V$ close to $(x, h(x))$.

(d) Deduce the inequality

$$d_{S\cap h^{-1}(R)}(z) \le k(d_S(z) + d_R(h(z)))$$

holds for all points z in U close to x.

(e) Imitate the proof of Liusternik's theorem (7.1.6) to deduce the inclusions

$$T_{S\cap h^{-1}(R)}(x) \supset T_S(x) \cap (\nabla h(x))^{-1} T_R(h(x))$$

and

$$K_{S\cap h^{-1}(R)}(x) \supset K_S(x) \cap (\nabla h(x))^{-1} T_R(h(x)).$$

(f) Suppose h is the identity map, so

$$T_S(x) - T_R(x) = \mathbf{E}.$$

If either R or S is tangentially regular at x, prove

$$K_{R\cap S}(x) = K_R(x) \cap K_S(x).$$

(g) **(Guignard)** By taking polars and applying the Krein–Rutman polar cone calculus (3.3.13) and condition (7.1.7) again, deduce

$$N_{S\cap h^{-1}(R)}(x) \subset N_S(x) + (\nabla h(x))^* N_R(h(x)).$$

(h) If C and D are convex subsets of $\mathbf{E}$ satisfying $0 \in \text{core}\,(C - D)$ (or $\text{ri}\,C \cap \text{ri}\,D \neq \emptyset$), and the point x lies in $C \cap D$, use part (e) to prove

$$T_{C \cap D}(x) = T_C(x) \cap T_D(x).$$

7. ** **(Liusternik via inverse functions)** We first fix $\mathbf{E} = \mathbf{R}^n$. The classical inverse function theorem states that if the map $g : U \to \mathbf{R}^n$ is continuously differentiable then at any point x in U at which $\nabla g(x)$ is invertible, x has an open neighbourhood V whose image $g(V)$ is open, and the restricted map $g|_V$ has a continuously differentiable inverse satisfying the condition

$$\nabla \left(g|_V\right)^{-1}(g(x)) = (\nabla g(x))^{-1}.$$

Consider now a continuously differentiable map $h : U \to \mathbf{R}^m$, and a point x in U with $\nabla h(x)$ surjective, and fix a direction d in the null space $N(\nabla h(x))$. Choose any $(n \times (n - m))$ matrix D making the matrix $A = (\nabla h(x), D)$ invertible, define a function $g : U \to \mathbf{R}^n$ by $g(z) = (h(z), Dz)$, and for a small real $\delta > 0$ define a function $p : (-\delta, \delta) \to \mathbf{R}^n$ by

$$p(t) = g^{-1}(g(x) + tAd).$$

(a) Prove p is well-defined providing δ is small.

(b) Prove the following properties:

  (i) p is continuously differentiable.

  (ii) $p(0) = x$.

  (iii) $p'(0) = d$.

  (iv) $h(p(t)) = h(x)$ for all small t.

(c) Deduce that a direction d lies in $N(\nabla h(x))$ if and only if there is a function $p : (-\delta, \delta) \to \mathbf{R}^n$ for some $\delta > 0$ in $\mathbf{R}$ satisfying the four conditions in part (b).

(d) Deduce $K_{h^{-1}(h(x))}(x) = N(\nabla h(x))$.

7.2 The Karush–Kuhn–Tucker Theorem

The central result of optimization theory describes first order necessary optimality conditions for the general nonlinear problem

$$\inf\{f(x) \mid x \in S\}, \tag{7.2.1}$$

where, given an open set $U \subset \mathbf{E}$, the objective function is $f : U \to \mathbf{R}$ and the feasible region S is described by equality and inequality constraints:

$$S = \{x \in U \mid g_i(x) \leq 0 \text{ for } i = 1, 2, \ldots, m,\ h(x) = 0\}. \tag{7.2.2}$$

The equality constraint map $h : U \to \mathbf{Y}$ (where $\mathbf{Y}$ is a Euclidean space) and the inequality constraint functions $g_i : U \to \mathbf{R}$ (for $i = 1, 2, \ldots, m$) are all continuous. In this section we derive necessary conditions for the point $\bar{x}$ in S to be a local minimizer for the problem (7.2.1).

In outline, the approach takes three steps. We first extend Liusternik's theorem (7.1.6) to describe the contingent cone $K_S(\bar{x})$. Next we calculate this cone's polar cone using the Farkas lemma (2.2.7). Finally, we apply the Contingent necessary condition (6.3.10) to derive the result.

As in our development for the inequality-constrained problem in Section 2.3, we need a regularity condition. Once again, we denote the set of indices of the active inequality constraints by $I(\bar{x}) = \{i \mid g_i(\bar{x}) = 0\}$.

Assumption 7.2.3 (The Mangasarian–Fromovitz constraint qualification) *The active constraint functions g_i (for i in $I(\bar{x})$) are Fréchet differentiable at the point $\bar{x}$, the equality constraint map h is strictly differentiable, with a surjective gradient, at $\bar{x}$, and the set*

$$\{p \in N(\nabla h(\bar{x})) \mid \langle \nabla g_i(\bar{x}), p \rangle < 0 \textit{ for } i \textit{ in } I(\bar{x})\} \tag{7.2.4}$$

is nonempty.

Notice in particular that the set (7.2.4) is nonempty in the case where the map $h : U \to \mathbf{R}^q$ has components $h_1, h_2, \ldots, h_q$ and the set of gradients

$$\{\nabla h_j(\bar{x}) \mid j = 1, 2, \ldots, q\} \cup \{\nabla g_i(\bar{x}) \mid i \in I(\bar{x})\} \tag{7.2.5}$$

is linearly independent (Exercise 1).

Theorem 7.2.6 *Suppose the Mangasarian–Fromovitz constraint qualification (7.2.3) holds. Then the contingent cone to the feasible region S defined by equation (7.2.2) is given by*

$$K_S(\bar{x}) = \{p \in N(\nabla h(\bar{x})) \mid \langle \nabla g_i(\bar{x}), p \rangle \leq 0 \textit{ for } i \textit{ in } I(\bar{x})\}. \tag{7.2.7}$$

Proof. Denote the set (7.2.4) by $\widetilde{K}$ and the right hand side of formula (7.2.7) by K. The inclusion

$$K_S(\bar{x}) \subset K$$

is a straightforward exercise. Furthermore, since $\widetilde{K}$ is nonempty, it is easy to see $K = \operatorname{cl} \widetilde{K}$. If we can show $\widetilde{K} \subset K_S(\bar{x})$ then the result will follow since the contingent cone is always closed.

To see $\widetilde{K} \subset K_S(\bar{x})$, fix an element p of $\widetilde{K}$. Since p lies in $N(\nabla h(\bar{x}))$, Liusternik's theorem (7.1.6) shows $p \in K_{h^{-1}(0)}(\bar{x})$. Hence there are sequences $t_r \downarrow 0$ in $\mathbf{R}_{++}$ and $p^r \to p$ in $\mathbf{E}$ satisfying $h(\bar{x} + t_r p^r) = 0$ for all r. Clearly $\bar{x} + t_r p^r \in U$ for all large r, and we claim $g_i(\bar{x} + t_r p^r) < 0$. For indices i not in $I(\bar{x})$ this follows by continuity, so we suppose $i \in I(\bar{x})$ and $g_i(\bar{x} + t_r p^r) \geq 0$ for all r in some subsequence R of $\mathbf{N}$. We then obtain the contradiction

$$\begin{aligned} 0 &= \lim_{r \to \infty \text{ in } R} \frac{g_i(\bar{x} + t_r p^r) - g_i(\bar{x}) - \langle \nabla g_i(\bar{x}), t_r p^r \rangle}{t_r \|p^r\|} \\ &\geq -\frac{\langle \nabla g_i(\bar{x}), p \rangle}{\|p\|} \\ &> 0. \end{aligned}$$

The result now follows. □

Lemma 7.2.8 *Any linear maps $A : \mathbf{E} \to \mathbf{R}^q$ and $G : \mathbf{E} \to \mathbf{Y}$ satisfy*

$$\{x \in N(G) \mid Ax \leq 0\}^- = A^* \mathbf{R}^q_+ + G^* \mathbf{Y}.$$

Proof. This is an immediate application of Section 5.1, Exercise 9 (Polyhedral cones). □

Theorem 7.2.9 (Karush–Kuhn–Tucker conditions) *Suppose $\bar{x}$ is a local minimizer for problem (7.2.1) and the objective function f is Fréchet differentiable at $\bar{x}$. If the Mangasarian–Fromovitz constraint qualification (7.2.3) holds then there exist multipliers λ_i in $\mathbf{R}_+$ (for i in $I(\bar{x})$) and μ in $\mathbf{Y}$ satisfying*

$$\nabla f(\bar{x}) + \sum_{i \in I(\bar{x})} \lambda_i \nabla g_i(\bar{x}) + \nabla h(\bar{x})^* \mu = 0. \tag{7.2.10}$$

Proof. The Contingent necessary condition (6.3.10) shows

$$\begin{aligned} -\nabla f(\bar{x}) &\in K_S(\bar{x})^- \\ &= \{p \in N(\nabla h(\bar{x})) \mid \langle \nabla g_i(\bar{x}), p \rangle \leq 0 \text{ for } i \text{ in } I(\bar{x})\}^- \\ &= \sum_{i \in I(\bar{x})} \mathbf{R}_+ \nabla g_i(\bar{x}) + \nabla h(\bar{x})^* \mathbf{Y} \end{aligned}$$

using Theorem 7.2.6 and Lemma 7.2.8. □

Exercises and Commentary

A survey of the history of these results may be found in [158]. The Mangasarian–Fromovitz condition originated with [133], while the Karush–Kuhn–Tucker conditions first appeared in [111] and [117]. The idea of penalty functions (see Exercise 11 (Quadratic penalties)) is a common technique in optimization. The related notion of a barrier penalty is crucial for interior point methods; examples include the penalized linear and semidefinite programs we considered in Section 4.3, Exercise 4 (Examples of duals).

1. **(Linear independence implies Mangasarian–Fromovitz)** If the set of gradients (7.2.5) is linearly independent, then by considering the equations

$$\begin{aligned} \langle \nabla g_i(\bar{x}), p \rangle &= -1 \quad \text{for } i \text{ in } I(\bar{x}) \\ \langle \nabla h_j(\bar{x}), p \rangle &= 0 \quad \text{for } j = 1, 2, \ldots, q, \end{aligned}$$

   prove the set (7.2.4) is nonempty.

2. Consider the proof of Theorem 7.2.6.

   (a) Prove $K_S(\bar{x}) \subset K$.

   (b) If $\widetilde{K}$ is nonempty, prove $K = \mathrm{cl}\, \widetilde{K}$.

3. **(Linear constraints)** If the functions g_i (for i in $I(\bar{x})$) and h are affine, prove the contingent cone formula (7.2.7) holds.

4. **(Bounded multipliers)** In Theorem 7.2.9 (Karush–Kuhn–Tucker conditions), prove the set of multiplier vectors (λ, μ) satisfying equation (7.2.10) is compact.

5. **(Slater condition)** Suppose the set U is convex, the functions

$$g_1, g_2, \ldots, g_m : U \to \mathbf{R}$$

   are convex and Fréchet differentiable, and the function $h : \mathbf{E} \to \mathbf{Y}$ is affine and surjective. Suppose further there is a point $\hat{x}$ in $h^{-1}(0)$ satisfying $g_i(\hat{x}) < 0$ for $i = 1, 2, \ldots, m$. For any feasible point $\bar{x}$ for problem (7.2.1), prove the Mangasarian–Fromovitz constraint qualification holds.

6. **(Largest eigenvalue)** For a matrix A in $\mathbf{S}^n$, use the Karush–Kuhn–Tucker theorem to calculate

$$\sup\{x^T A x \mid \|x\| = 1,\ x \in \mathbf{R}^n\}.$$

7. * **(Largest singular value [100, p. 135])** Given any $m \times n$ matrix A, consider the optimization problem

$$\alpha = \sup\{x^T Ay \mid \|x\|^2 = 1,\ \|y\|^2 = 1\} \tag{7.2.11}$$

and the matrix

$$\widetilde{A} = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}.$$

(a) If μ is an eigenvalue of $\widetilde{A}$, prove $-\mu$ is also.

(b) If μ is a nonzero eigenvalue of $\widetilde{A}$, use a corresponding eigenvector to construct a feasible solution to problem (7.2.11) with objective value μ.

(c) Deduce $\alpha \geq \lambda_1(\widetilde{A})$.

(d) Prove problem (7.2.11) has an optimal solution.

(e) Use the Karush–Kuhn–Tucker theorem to prove any optimal solution of problem (7.2.11) corresponds to an eigenvector of $\widetilde{A}$.

(f) **(Jordan [108])** Deduce $\alpha = \lambda_1(\widetilde{A})$. (This number is called the *largest singular value of* A.)

8. ** **(Hadamard's inequality [88])** The matrix with columns $x^1, x^2, \ldots, x^n$ in $\mathbf{R}^n$ we denote by $(x^1, x^2, \ldots, x^n)$. Prove $(\bar{x}^1, \bar{x}^2, \ldots, \bar{x}^n)$ solves the problem

$$\begin{array}{lrcl} \inf & -\det(x^1, x^2, \ldots, x^n) & & \\ \text{subject to} & \|x^i\|^2 & = & 1 \ \text{ for } i = 1, 2, \ldots, n \\ & x^1, x^2, \ldots, x^n & \in & \mathbf{R}^n \end{array}$$

if and only if the matrix $(\bar{x}^1, \bar{x}^2, \ldots, \bar{x}^n)$ has determinant equal to one and has columns forming an orthonormal basis, and deduce the inequality

$$\det(x^1, x^2, \ldots, x^n) \leq \prod_{i=1}^{n} \|x^i\|.$$

9. **(Nonexistence of multipliers [77])** Define a function $\text{sgn} : \mathbf{R} \to \mathbf{R}$ by

$$\text{sgn}(v) = \begin{cases} 1 & \text{if } v > 0 \\ 0 & \text{if } v = 0 \\ -1 & \text{if } v < 0 \end{cases}$$

and a function $h : \mathbf{R}^2 \to \mathbf{R}$ by

$$h(u, v) = v - \text{sgn}(v)(u^+)^2.$$

(a) Prove h is Fréchet differentiable at $(0, 0)$ with derivative $(0, 1)$.

(b) Prove h is not continuous on any neighbourhood of $(0,0)$, and deduce it is not strictly differentiable at $(0,0)$.

(c) Prove $(0,0)$ is optimal for the problem

$$\inf\{f(u,v) \mid h(u,v) = 0\},$$

where $f(u,v) = u$, and yet there is no real λ satisfying

$$\nabla f(0,0) + \lambda \nabla h(0,0) = (0,0).$$

(Exercise 14 in Section 8.1 gives an approach to weakening the conditions required in this section.)

10. * **(Guignard optimality conditions [87])** Suppose the point $\bar{x}$ is a local minimizer for the optimization problem

$$\inf\{f(x) \mid h(x) \in R,\ x \in S\}$$

where $R \subset \mathbf{Y}$. If the functions f and h are strictly differentiable at $\bar{x}$ and the transversality condition

$$\nabla h(\bar{x}) T_S(\bar{x}) - T_R(h(\bar{x})) = \mathbf{Y}$$

holds, use Section 7.1, Exercise 6 (Transversality) to prove the optimality condition

$$0 \in \nabla f(\bar{x}) + \nabla h(\bar{x})^* N_R(h(\bar{x})) + N_S(\bar{x}).$$

11. ** **(Quadratic penalties [136])** Take the nonlinear program (7.2.1) in the case $\mathbf{Y} = \mathbf{R}^q$ and now let us assume all the functions

$$f, g_1, g_2, \ldots, g_m, h_1, h_2, \ldots, h_q : U \to \mathbf{R}$$

are continuously differentiable on the set U. For positive integers k we define a function $p_k : U \to \mathbf{R}$ by

$$p_k(x) = f(x) + k\Big(\sum_{i=1}^{m} (g_i^+(x))^2 + \sum_{j=1}^{q} (h_j(x))^2\Big).$$

Suppose the point $\bar{x}$ is a local minimizer for the problem (7.2.1). Then for some compact neighbourhood W of $\bar{x}$ in U we know $f(x) \geq f(\bar{x})$ for all feasible points x in W. Now define a function $r_k : W \to \mathbf{R}$ by

$$r_k(x) = p_k(x) + \|x - \bar{x}\|^2,$$

and for each $k = 1, 2, \ldots$ choose a point x^k minimizing r_k on W.

(a) Prove $r_k(x^k) \leq f(\bar{x})$ for each $k = 1, 2, \ldots$.

(b) Deduce
$$\lim_{k \to \infty} g_i^+(x^k) = 0 \quad \text{for } i = 1, 2, \ldots, m$$
and
$$\lim_{k \to \infty} h_j(x^k) = 0 \quad \text{for } j = 1, 2, \ldots, q.$$

(c) Hence show $x^k \to \bar{x}$ as $k \to \infty$.

(d) Calculate $\nabla r_k(x)$.

(e) Deduce
$$-2(x^k - \bar{x}) = \nabla f(x^k) + \sum_{i=1}^{m} \lambda_i^k \nabla g_i(x^k) + \sum_{j=1}^{q} \mu_j^k \nabla h_j(x^k)$$
for some suitable choice of vectors λ^k in $\mathbf{R}_+^m$ and μ^k in $\mathbf{R}^q$.

(f) By taking a convergent subsequence of the vectors
$$\|(1, \lambda^k, \mu^k)\|^{-1}(1, \lambda^k, \mu^k) \in \mathbf{R} \times \mathbf{R}_+^m \times \mathbf{R}^q,$$
show from parts (c) and (e) the existence of a nonzero vector $(\lambda_0, \lambda, \mu)$ in $\mathbf{R} \times \mathbf{R}_+^m \times \mathbf{R}^q$ satisfying the *Fritz John conditions*:

  (i) $\lambda_i g_i(\bar{x}) = 0$ for $i = 1, 2, \ldots, m$.

  (ii) $\lambda_0 \nabla f(\bar{x}) + \sum_{i=1}^m \lambda_i \nabla g_i(\bar{x}) + \sum_{j=1}^q \mu_j \nabla h_j(\bar{x}) = 0$.

(g) Under the assumption of the Mangasarian–Fromovitz constraint qualification (7.2.3), show that the Fritz John conditions in part (f) imply the Karush–Kuhn–Tucker conditions.

7.3 Metric Regularity and the Limiting Subdifferential

In Section 7.1 we presented a convenient test for the weak metric regularity of a function at a point in terms of the surjectivity of its strict derivative there (Theorem 7.1.5). This test, while adequate for most of our purposes, can be richly refined using the limiting subdifferential.

As before, we consider an open set $U \subset \mathbf{E}$, a Euclidean space $\mathbf{Y}$, a closed set $S \subset U$, and a function $h : U \to \mathbf{Y}$ which we assume throughout this section is locally Lipschitz. We begin with the full definition of metric regularity, strengthening the notion of Section 7.1. We say h is *metrically regular on S at* the point x in S if there is a real constant k such that the estimate

$$d_{S\cap h^{-1}(y)}(z) \leq k\|h(z) - y\|$$

holds for all points z in S close to x and all vectors y in $\mathbf{Y}$ close to $h(x)$. (Before we only required this to be true when $y = h(x)$.)

Lemma 7.3.1 *If h is not metrically regular on S at x then there are sequences (v_r) in S converging to x, (y_r) in $\mathbf{Y}$ converging to $h(x)$, and (ϵ_r) in $\mathbf{R}_{++}$ decreasing to zero such that, for each index r, we have $h(v_r) \neq y_r$ and the function*

$$\|h(\cdot) - y_r\| + \epsilon_r\| \cdot - v_r\|$$

is minimized on S at v_r.

Proof. The proof is completely analogous to that of Lemma 7.1.3: we leave it as an exercise. $\square$

We also need the following chain-rule-type result; we leave the proof as an exercise.

Lemma 7.3.2 *At any point x in $\mathbf{E}$ where $h(x) \neq 0$ we have*

$$\partial_a\|h(\cdot)\|(x) = \partial_a\langle\|h(x)\|^{-1}h(x), h(\cdot)\rangle(x).$$

Using this result and a very similar proof to Theorem 7.1.5, we can now extend the surjectivity and metric regularity result.

Theorem 7.3.3 (Limiting subdifferential and regularity) *If a point x lies in S and no nonzero element w of $\mathbf{Y}$ satisfies the condition*

$$0 \in \partial_a\langle w, h(\cdot)\rangle(x) + N_S^a(x)$$

then h is metrically regular on S at x.

Proof. If h is not metrically regular, we can apply Lemma 7.3.1, so with that notation the function

$$\|h(\cdot) - y_r\| + \epsilon_r\| \cdot - v_r\|$$

is minimized on S at v_r. By Proposition 6.3.2 (Exact penalization) we deduce for large enough real L

$$\begin{aligned} 0 &\in \partial_a(\|h(\cdot) - y_r\| + \epsilon_r\| \cdot - v_r\| + L d_S(\cdot))(v_r) \\ &\subset \partial_a\|h(\cdot) - y_r\|(v_r) + \epsilon_r B + L\partial_a d_S(v_r) \end{aligned}$$

for all r, using the Limiting subdifferential sum rule (6.4.4). If we write $w_r = \|h(v_r) - y_r\|^{-1}(h(v_r) - y_r)$, we obtain by Lemma 7.3.2

$$0 \in \partial_a\langle w_r, h(\cdot)\rangle(v_r) + \epsilon_r B + L\partial_a d_S(v_r),$$

so there are elements u_r in $\partial_a\langle w_r, h(\cdot)\rangle(v_r)$ and z_r in $L\partial_a d_S(v_r)$ such that $\|u_r + z_r\| \le \epsilon_r$. The sequences (w_r), (u_r), and (z_r) are all bounded, so by taking subsequences we can assume w_r approaches some nonzero vector w, z_r approaches some vector z, and u_r approaches $-z$.

Now, using the sum rule again we observe

$$u_r \in \partial_a\langle w, h(\cdot)\rangle(v_r) + \partial_a\langle w_r - w, h(\cdot)\rangle(v_r)$$

for each r. The local Lipschitz constant of the function $\langle w_r - w, h(\cdot)\rangle$ tends to zero, so since $\partial_a\langle w, h(\cdot)\rangle$ is a closed multifunction at x (by Section 6.4, Exercise 5) we deduce

$$-z \in \partial_a\langle w, h(\cdot)\rangle(x).$$

Similarly, since $\partial_a d_S(\cdot)$ is closed at x, we see

$$z \in L\partial_a d_S(x) \subset N_S^a(x)$$

by Exercise 4, and this contradicts the assumption of the theorem. □

This result strengthens and generalizes the elegant test of Theorem 7.1.5, as the next result shows.

Corollary 7.3.4 (Surjectivity and metric regularity) *If h is strictly differentiable at the point x in S and*

$$(\nabla h(x)^*)^{-1}(N_S^a(x)) = \{0\}$$

or, in particular,

$$\nabla h(x)(T_S(x)) = \mathbf{Y}$$

then h is metrically regular on S at x.

Proof. Since it is easy to check for any element w of $\mathbf{Y}$ the function $\langle w, h(\cdot)\rangle$ is strictly differentiable at x with derivative $\nabla h(x)^* w$, the first condition implies the result by Theorem 7.3.3. On the other hand, the second condition implies the first, since for any element w of $(\nabla h(x)^*)^{-1}(N_S^a(x))$ there is an element z of $T_S(x)$ satisfying $\nabla h(x) z = w$, and now we deduce

$$\|w\|^2 = \langle w, w\rangle = \langle w, \nabla h(x) z\rangle = \langle \nabla h(x)^* w, z\rangle \leq 0$$

using Exercise 4, so $w = 0$. □

As a final extension to the idea of metric regularity, consider now a closed set $D \subset \mathbf{Y}$ containing $h(x)$. We say h is *metrically regular on S at x with respect to D* if there is a real constant k such that

$$d_{S \cap h^{-1}(y+D)}(z) \leq k d_D(h(z) - y)$$

for all points z in S close to x and vectors y close to 0. Our previous definition was the case $D = \{h(x)\}$. This condition estimates how far a point $z \in S$ is from feasibility for the system

$$h(z) \in y + D, \quad z \in S,$$

in terms of the constraint error $d_D(h(z) - y)$.

Corollary 7.3.5 *If the point x lies in the closed set $S \subset \mathbf{E}$ with $h(x)$ in the closed set $D \subset \mathbf{Y}$, and no nonzero element w of $N_D^a(h(x))$ satisfies the condition*

$$0 \in \partial_a \langle w, h(\cdot)\rangle(x) + N_S^a(x),$$

then h is metrically regular on S at x with respect to D.

Proof. Define a function $\widetilde{h} : U \times \mathbf{Y} \to \mathbf{Y}$ by $\widetilde{h}(z, y) = h(z) - y$, a set $\widetilde{S} = S \times D$, and a point $\widetilde{x} = (x, h(x))$. Since by Exercise 5 we have

$$N_{\widetilde{S}}^a(\widetilde{x}) = N_S^a(x) \times N_D^a(h(x))$$

and

$$\partial_a \langle w, \widetilde{h}(\cdot)\rangle(\widetilde{x}) = \partial_a \langle w, h(\cdot)\rangle(x) \times \{-w\}$$

for any element w of $\mathbf{Y}$, there is no nonzero w satisfying the condition

$$0 \in \partial_a \langle w, \widetilde{h}(\cdot)\rangle(\widetilde{x}) + N_{\widetilde{S}}^a(\widetilde{x}),$$

so $\widetilde{h}$ is metrically regular on $\widetilde{S}$ at $\widetilde{x}$ by Theorem 7.3.3 (Limiting subdifferential and regularity). Some straightforward manipulation now shows h is metrically regular on S at x with respect to D. □

The case $D = \{h(x)\}$ recaptures Theorem 7.3.3.

A nice application of this last result estimates the distance to a level set under a Slater-type assumption, a typical illustration of the power of metric regularity.

Corollary 7.3.6 (Distance to level sets) *If the function $g : U \to \mathbf{R}$ is locally Lipschitz around a point x in U satisfying*

$$g(x) = 0 \quad \text{and} \quad 0 \notin \partial_a g(x)$$

then there is a real constant $k > 0$ such that the estimate

$$d_{g^{-1}(-\mathbf{R}_+)}(z) \leq kg(z)^+$$

holds for all points z in $\mathbf{E}$ close to x.

Proof. Let $S \subset U$ be any closed neighbourhood of x and apply Corollary 7.3.5 with $h = g$ and $D = -\mathbf{R}_+$. □

Exercises and Commentary

In many circumstances, metric regularity is in fact equivalent to weak metric regularity (see [25]). The power of the limiting subdifferential as a tool in recognizing metric regularity was first observed by Mordukhovich [144]; there is a comprehensive discussion in [145, 168].

1. * Prove Lemma 7.3.1.

2. * Assume $h(x) \neq 0$.

   (a) Prove
   $$\partial_- \|h(\cdot)\|(x) = \partial_- \langle \|h(x)\|^{-1} h(x), h(\cdot)\rangle(x).$$

   (b) Prove the analogous result for the limiting subdifferential. (You may use the Limiting subdifferential sum rule (6.4.4).)

3. **(Metric regularity and openness)** If h is metrically regular on S at x, prove h is *open* on S at x; that is, for any neighbourhood U of x we have $h(x) \in \operatorname{int} h(U \cap S)$.

4. ** **(Limiting normals and distance functions)** Given a point z in $\mathbf{E}$, suppose y is a nearest point to z in S.

   (a) If $0 \leq \alpha < 1$, prove the unique nearest point to $\alpha z + (1 - \alpha)y$ in S is y.

   (b) For z not in S, deduce every element of $\partial_- d_S(z)$ has norm one.

   (c) For any element w of $\mathbf{E}$, prove
   $$d_S(z + w) \leq d_S(z) + d_S(y + w).$$

   (d) Deduce $\partial_- d_S(z) \subset \partial_- d_S(y)$.

Now consider a point x in S.

(e) Prove ϕ is an element of $\partial_a d_S(x)$ if and only if there are sequences (x^r) in S approaching x, and (ϕ^r) in $\mathbf{E}$ approaching ϕ satisfying $\phi^r \in \partial_- d_S(x^r)$ for all r.

(f) Deduce $\mathbf{R}_+ \partial_a d_S(x) \subset N_S^a(x)$.

(g) Suppose ϕ is an element of $\partial_- \delta_S(x)$. For any real $\epsilon > 0$, apply Section 6.4, Exercise 3 (Local minimizers) and the Limiting subdifferential sum rule to prove

$$\phi \in (\|\phi\| + \epsilon)\partial_a d_S(x) + \epsilon B.$$

(h) By taking limits, deduce

$$N_S^a(x) = \mathbf{R}_+ \partial_a d_S(x).$$

(i) Deduce

$$N_S(x) = \mathrm{cl}\,(\mathrm{conv}\, N_S^a(x)),$$

and hence

$$T_S(x) = N_S^a(x)^-.$$

(Hint: Use Section 6.4, Exercise 7 (Limiting and Clarke subdifferentials).)

(j) Hence prove the following properties are equivalent:

(i) $T_S(x) = \mathbf{E}$.

(ii) $N_S^a(x) = \{0\}$.

(iii) $x \in \mathrm{int}\, S$.

5. **(Normals to products)** For closed sets $S \subset \mathbf{E}$ and $D \subset \mathbf{Y}$ and points x in S and y in D, prove

$$N_{S \times D}^a(x, y) = N_S^a(x) \times N_D^a(y).$$

6. * Complete the remaining details of the proof of Corollary 7.3.5.

7. Prove Corollary 7.3.6 (Distance to level sets).

8. **(Limiting versus Clarke conditions)** Define a set

$$S = \{(u, v) \in \mathbf{R}^2 \mid u \leq 0 \text{ or } v \leq 0\}$$

and a function $h : \mathbf{R}^2 \to \mathbf{R}$ by $h(u, v) = u + v$. In Corollary 7.3.4 (Surjectivity and metric regularity), prove the limiting normal cone condition holds at the point $x = 0$, and yet the Clarke tangent cone condition fails.

9. ** **(Normals to level sets)** Under the hypotheses of Corollary 7.3.6 (Distance to level sets), prove

$$N^a_{g^{-1}(-\mathbf{R}_+)}(x) = \mathbf{R}_+ \partial_a g(x).$$

(Hint: Use Exercise 4 and the Max rule (Section 6.4, Exercise 10(g).)

7.4 Second Order Conditions

Optimality conditions can be refined using second order information; we saw an early example in Theorem 2.1.5 (Second order conditions). Because of the importance of curvature information for Newton-type methods in numerical optimization, second order conditions are widely useful.

In this section we present prototypical second order conditions for constrained optimization. Our approach is a simple and elegant blend of convex analysis and metric regularity.

Consider an open set $U \subset \mathbf{E}$, a Euclidean space $\mathbf{Y}$. Given any function $h : U \to \mathbf{Y}$ that is Fréchet differentiable on U, the gradient map ∇h is a function from U to the vector space $L(\mathbf{E}, \mathbf{Y})$ of all linear maps from $\mathbf{E}$ to $\mathbf{Y}$ with the operator norm

$$\|A\| = \max_{x \in B_{\mathbf{E}}} \|Ax\| \quad (A \in L(\mathbf{E}, \mathbf{Y})).$$

If this map ∇h is itself Fréchet differentiable at the point $\bar{x}$ in U then we say h is *twice Fréchet differentiable* at $\bar{x}$: the gradient $\nabla^2 h(\bar{x})$ is a linear map from $\mathbf{E}$ to $L(\mathbf{E}, \mathbf{Y})$, and for any element v of $\mathbf{E}$ we write

$$(\nabla^2 h(\bar{x})v)(v) = \nabla^2 h(\bar{x})(v, v).$$

In this case h has the following *quadratic approximation* at $\bar{x}$:

$$h(\bar{x} + v) = h(\bar{x}) + \nabla h(\bar{x})v + \frac{1}{2}\nabla^2 h(\bar{x})(v, v) + o(\|v\|^2) \;\text{ for small } v.$$

We suppose throughout this section that the functions $f : U \to \mathbf{R}$ and h are twice Fréchet differentiable at $\bar{x}$, and that the closed convex set S contains $\bar{x}$. We consider the nonlinear optimization problem

$$\inf\{f(x) \mid h(x) = 0, \; x \in S\}, \tag{7.4.1}$$

and we define the *narrow critical cone at* $\bar{x}$ by

$$C(\bar{x}) = \{d \in \mathbf{R}_+(S - \bar{x}) \mid \nabla f(\bar{x})d \le 0, \; \nabla h(\bar{x})d = 0\}.$$

Theorem 7.4.2 (Second order necessary conditions) *Suppose that the point $\bar{x}$ is a local minimum for the problem (7.4.1), that the direction d lies in the narrow critical cone $C(\bar{x})$, and that the condition*

$$0 \in \operatorname{core}(\nabla h(\bar{x})(S - \bar{x})) \tag{7.4.3}$$

holds. Then there exists a multiplier λ in $\mathbf{Y}$ such that the Lagrangian

$$L(\cdot) = f(\cdot) + \langle \lambda, h(\cdot) \rangle \tag{7.4.4}$$

satisfies the conditions

$$\nabla L(\bar{x}) \in -N_S(\bar{x}) \tag{7.4.5}$$

and

$$\nabla^2 L(\bar{x})(d,d) \geq 0. \tag{7.4.6}$$

Proof. Consider first the convex program

$$\inf\{\nabla f(\bar{x})z \mid \nabla h(\bar{x})z = -\nabla^2 h(\bar{x})(d,d),\ z \in \mathbf{R}_+(S-\bar{x})\}. \tag{7.4.7}$$

Suppose the point z is feasible for problem (7.4.7). It is easy to check for small real $t \geq 0$ the path

$$x(t) = \bar{x} + td + \frac{t^2}{2}z$$

lies in S. Furthermore, the quadratic approximation shows this path almost satisfies the original constraint for small t:

$$\begin{aligned} h(x(t)) &= h(\bar{x}) + t\nabla h(\bar{x})d + \frac{t^2}{2}(\nabla h(\bar{x})z + \nabla^2 h(\bar{x})(d,d)) + o(t^2) \\ &= o(t^2). \end{aligned}$$

But condition (7.4.3) implies in particular that $\nabla h(\bar{x})T_S(\bar{x}) = \mathbf{Y}$; in fact these conditions are equivalent, since the only convex set whose closure is $\mathbf{Y}$ is $\mathbf{Y}$ itself (see Section 4.1, Exercise 20(a) (Properties of the relative interior)). So, by Theorem 7.1.5 (Surjectivity and metric regularity), h is (weakly) metrically regular on S at $\bar{x}$. Hence the path above is close to feasible for the original problem: there is a real constant k such that, for small $t \geq 0$, we have

$$d_{S\cap h^{-1}(0)}(x(t)) \leq k\|h(x(t))\| = o(t^2).$$

Thus we can perturb the path slightly to obtain a set of points

$$\{\tilde{x}(t) \mid t \geq 0\} \subset S \cap h^{-1}(0)$$

satisfying $\|\tilde{x}(t) - x(t)\| = o(t^2)$.

Since $\bar{x}$ is a local minimizer for the original problem (7.4.1), we know

$$f(\bar{x}) \leq f(\tilde{x}(t)) = f(\bar{x}) + t\nabla f(\bar{x})d + \frac{t^2}{2}(\nabla f(\bar{x})z + \nabla^2 f(\bar{x})(d,d)) + o(t^2)$$

using the quadratic approximation again. Hence $\nabla f(\bar{x})d \geq 0$, so in fact $\nabla f(\bar{x})d = 0$, since d lies in $C(\bar{x})$. We deduce

$$\nabla f(\bar{x})z + \nabla^2 f(\bar{x})(d,d) \geq 0.$$

We have therefore shown the optimal value of the convex program (7.4.7) is at least $-\nabla^2 f(\bar{x})(d,d)$.

For the final step in the proof, we rewrite problem (7.4.7) in Fenchel form:

$$\inf_{z \in \mathbf{E}} \left\{ \left(\langle \nabla f(\bar{x}), z \rangle + \delta_{\mathbf{R}_+(S-\bar{x})}(z) \right) + \delta_{\{-\nabla^2 h(\bar{x})(d,d)\}}(\nabla h(\bar{x}) z) \right\}.$$

Since condition (7.4.3) holds, we can apply Fenchel duality (3.3.5) to deduce there exists $\lambda \in \mathbf{Y}$ satisfying

$$\begin{aligned} -\nabla^2 f(\bar{x})(d,d) &\leq -\delta^*_{\mathbf{R}_+(S-\bar{x})}(-\nabla h(\bar{x})^* \lambda - \nabla f(\bar{x})) - \delta^*_{\{-\nabla^2 h(\bar{x})(d,d)\}}(\lambda) \\ &= -\delta_{N_S(\bar{x})}(-\nabla h(\bar{x})^* \lambda - \nabla f(\bar{x})) + \langle \lambda, \nabla^2 h(\bar{x})(d,d) \rangle, \end{aligned}$$

whence the result. □

Under some further conditions we can guarantee that for *any* multiplier λ satisfying the first order condition (7.4.5), the second order condition (7.4.6) holds for *all* directions d in the narrow critical cone (see Exercises 2 and 3).

We contrast the necessary condition above with a rather elementary second order *sufficient* condition. For this we use the *broad critical cone at* $\bar{x}$:

$$\overline{C}(\bar{x}) = \{ d \in K_S(\bar{x}) \mid \nabla f(\bar{x}) d \leq 0, \ \nabla h(\bar{x}) d = 0 \}.$$

Theorem 7.4.8 (Second order sufficient condition) *Suppose for each nonzero direction d in the broad critical cone $\overline{C}(\bar{x})$ there exist multipliers μ in $\mathbf{R}_+$ and λ in $\mathbf{Y}$ such that the Lagrangian*

$$\overline{L}(\cdot) = \mu f(\cdot) + \langle \lambda, h(\cdot) \rangle$$

satisfies the conditions

$$\nabla \overline{L}(\bar{x}) \in -N_S(\bar{x}) \quad \textit{and} \quad \nabla^2 \overline{L}(\bar{x})(d,d) > 0.$$

Then for all small real $\delta > 0$ the point $\bar{x}$ is a strict local minimizer for the perturbed problem

$$\inf\{ f(x) - \delta \|x - \bar{x}\|^2 \mid h(x) = 0, \ x \in S \}. \tag{7.4.9}$$

Proof. Suppose there is no such δ, so there is a sequence of feasible solutions (x_r) for problem (7.4.9) converging to $\bar{x}$ and satisfying

$$\limsup_{r \to \infty} \frac{f(x_r) - f(\bar{x})}{\|x_r - \bar{x}\|^2} \leq 0. \tag{7.4.10}$$

By taking a subsequence, we can assume

$$\lim_{r \to \infty} \frac{x_r - \bar{x}}{\|x_r - \bar{x}\|} = d,$$

and it is easy to check the nonzero direction d lies in $\overline{C}(\bar{x})$. Hence by assumption there exist the required multipliers μ and λ.

From the first order condition we know

$$\nabla\overline{L}(\bar{x})(x_r - \bar{x}) \geq 0,$$

so by the quadratic approximation we deduce as $r \to \infty$

$$\begin{aligned}\mu(f(x_r) - f(\bar{x})) &= \overline{L}(x_r) - \overline{L}(\bar{x}) \\ &\geq \frac{1}{2}\nabla^2\overline{L}(\bar{x})(x_r - \bar{x}, x_r - \bar{x}) + o(\|x_r - \bar{x}\|^2).\end{aligned}$$

Dividing by $\|x_r - \bar{x}\|^2$ and taking limits shows

$$\mu \liminf_{r\to\infty} \frac{f(x_r) - f(\bar{x})}{\|x_r - \bar{x}\|^2} \geq \frac{1}{2}\nabla^2\overline{L}(\bar{x})(d, d) > 0,$$

which contradicts inequality (7.4.10). □

Notice this result is of Fritz John type (like Theorem 2.3.6): we do not assume the multiplier μ is nonzero. Furthermore, we can easily weaken the assumption that the set S is convex to the condition

$$(S - \bar{x}) \cap \epsilon B \subset K_S(\bar{x}) \text{ for some } \epsilon > 0.$$

Clearly the narrow critical cone may be smaller than the broad critical cone, even when S is convex. They are equal if S is *quasipolyhedral* at $\bar{x}$:

$$K_S(\bar{x}) = \mathbf{R}_+(S - \bar{x})$$

(as happens in particular when S is polyhedral). However, even for unconstrained problems there is an intrinsic gap between the second order necessary conditions and the sufficient conditions.

Exercises and Commentary

Our approach here is from [25] (see also [12]). There are higher order analogues [11]. Problems of the form (7.4.11) where all the functions involved are quadratic are called *quadratic programs.* Such problems are particularly well-behaved: the optimal value is attained when finite, and in this case the second order necessary conditions developed in Exercise 3 are also *sufficient* (see [21]). For a straightforward exposition of the standard second order conditions, see [132], for example.

1. **(Higher order conditions)** By considering the function

$$\operatorname{sgn}(x) \exp\left(-\frac{1}{x^2}\right)$$

on $\mathbf{R}$, explain why there is no necessary and sufficient nth order optimality condition.

2. * **(Uniform multipliers)** With the assumptions of Theorem 7.4.2 (Second order necessary conditions), suppose in addition that for all directions d in the narrow critical cone $C(\bar{x})$ there exists a solution z in $\mathbf{E}$ to the system

$$\nabla h(\bar{x})z = -\nabla^2 h(\bar{x})(d,d) \;\text{ and }\; z \in \operatorname{span}(S - \bar{x}).$$

By considering problem (7.4.7), prove that if the multiplier λ satisfies the first order condition (7.4.5) then the second order condition (7.4.6) holds for all d in $C(\bar{x})$. Observe this holds in particular if $S = \mathbf{E}$ and $\nabla h(\bar{x})$ is surjective.

3. ** **(Standard second order necessary conditions)** Consider the problem

$$\left.\begin{array}{lrcl} \inf & f(x) & & \\ \text{subject to} & g_i(x) & \leq & 0 \text{ for } i = 1, 2, \ldots, m \\ & h_j(x) & = & 0 \text{ for } j = 1, 2, \ldots, q \\ & x & \in & \mathbf{R}^n, \end{array}\right\} \qquad (7.4.11)$$

where all the functions are twice Fréchet differentiable at the local minimizer $\bar{x}$ and the set of gradients

$$A = \{\nabla g_i(\bar{x}) \mid i \in I(\bar{x})\} \cup \{\nabla h_j(\bar{x}) \mid j = 1, 2, \ldots, q\}$$

is linearly independent (where we denote the set of indices of the active inequality constraints by $I(\bar{x}) = \{i \mid g_i(\bar{x}) = 0\}$, as usual). By writing this problem in the form (7.4.1) and applying Exercise 2, prove there exist multipliers μ_i in $\mathbf{R}_+$ (for i in $I(\bar{x})$) and $\lambda_1, \lambda_2, \ldots, \lambda_q$ in $\mathbf{R}$ such that the Lagrangian

$$L(\cdot) = f(\cdot) + \sum_{i \in I(\bar{x})} \mu_i g_i + \sum_{j=1}^{q} \lambda_j h_j$$

satisfies the conditions

$$\nabla L(\bar{x}) = 0 \;\text{ and }\; \nabla^2 L(\bar{x})(d,d) \geq 0 \;\text{ for all } d \text{ in } A^{\perp}.$$

4. **(Narrow and broad critical cones are needed)** By considering the set

$$S = \{x \in \mathbf{R}^2 \mid x_2 \geq x_1^2\}$$

and the problem

$$\inf\{x_2 - \alpha x_1^2 \mid x \in S\}$$

for various values of the real parameter α, explain why the narrow and broad critical cones cannot be interchanged in either the Second order necessary conditions (7.4.2) or the sufficient conditions (7.4.8).

5. **(Standard second order sufficient conditions)** Write down the second order sufficient optimality conditions for the general nonlinear program in Exercise 3.

6. * **(Guignard-type conditions)** Consider the problem of Section 7.2, Exercise 10,

$$\inf\{f(x) \mid h(x) \in R,\ x \in S\},$$

where the set $R \subset \mathbf{Y}$ is closed and convex. By rewriting this problem in the form (7.4.1), derive second order optimality conditions.