Chapter 2

Inequality Constraints

2.1 Optimality Conditions

Early in multivariate calculus we learn the significance of differentiability in finding minimizers. In this section we begin our study of the interplay between convexity and differentiability in optimality conditions.

For an initial example, consider the problem of minimizing a function $f : C \to \mathbf{R}$ on a set C in $\mathbf{E}$. We say a point $\bar{x}$ in C is a *local minimizer of* f *on* C if $f(x) \geq f(\bar{x})$ for all points x in C close to $\bar{x}$. The *directional derivative* of a function f at $\bar{x}$ in a direction $d \in \mathbf{E}$ is

$$f'(\bar{x}; d) = \lim_{t \downarrow 0} \frac{f(\bar{x} + td) - f(\bar{x})}{t},$$

when this limit exists. When the directional derivative $f'(\bar{x}; d)$ is actually linear in d (that is, $f'(\bar{x}; d) = \langle a, d \rangle$ for some element a of $\mathbf{E}$) then we say f is *(Gâteaux) differentiable* at $\bar{x}$, with *(Gâteaux) derivative* $\nabla f(\bar{x}) = a$. If f is differentiable at every point in C then we simply say f is differentiable (on C). An example we use quite extensively is the function $X \in \mathbf{S}^n_{++} \mapsto \log \det X$. An exercise shows this function is differentiable on $\mathbf{S}^n_{++}$ with derivative X^{-1}.

A convex cone which arises frequently in optimization is the *normal cone* to a convex set C at a point $\bar{x} \in C$, written $N_C(\bar{x})$. This is the convex cone of *normal vectors*, vectors d in $\mathbf{E}$ such that $\langle d, x - \bar{x} \rangle \leq 0$ for all points x in C.

Proposition 2.1.1 (First order necessary condition) *Suppose that C is a convex set in $\mathbf{E}$ and that the point $\bar{x}$ is a local minimizer of the function $f : C \to \mathbf{R}$. Then for any point x in C, the directional derivative, if it exists, satisfies $f'(\bar{x}; x - \bar{x}) \geq 0$. In particular, if f is differentiable at $\bar{x}$, then the condition $-\nabla f(\bar{x}) \in N_C(\bar{x})$ holds.*

Proof. If some point x in C satisfies $f'(\bar{x}; x - \bar{x}) < 0$, then all small real $t > 0$ satisfy $f(\bar{x} + t(x - \bar{x})) < f(\bar{x})$, but this contradicts the local minimality of $\bar{x}$. $\square$

The case of this result where C is an open set is the canonical introduction to the use of calculus in optimization: local minimizers $\bar{x}$ must be *critical points* (that is, $\nabla f(\bar{x}) = 0$). This book is largely devoted to the study of first order necessary optimality conditions for a local minimizer of a function subject to constraints. In that case local minimizers $\bar{x}$ may not lie in the interior of the set C of interest, so the normal cone $N_C(\bar{x})$ is not simply $\{0\}$.

The next result shows that when f is convex the first order condition above is *sufficient* for $\bar{x}$ to be a global minimizer of f on C.

Proposition 2.1.2 (First order sufficient condition) *Suppose that the set $C \subset \mathbf{E}$ is convex and that the function $f : C \to \mathbf{R}$ is convex. Then for any points $\bar{x}$ and x in C, the directional derivative $f'(\bar{x}; x - \bar{x})$ exists in $[-\infty, +\infty)$. If the condition $f'(\bar{x}; x - \bar{x}) \geq 0$ holds for all x in C, or in particular if the condition $-\nabla f(\bar{x}) \in N_C(\bar{x})$ holds, then $\bar{x}$ is a global minimizer of f on C.*

Proof. A straightforward exercise using the convexity of f shows the function

$$t \in (0, 1] \mapsto \frac{f(\bar{x} + t(x - \bar{x})) - f(\bar{x})}{t}$$

is nondecreasing. The result then follows easily (Exercise 7). $\square$

In particular, any critical point of a convex function is a global minimizer.

The following useful result illustrates what the first order conditions become for a more concrete optimization problem. The proof is outlined in Exercise 4.

Corollary 2.1.3 (First order conditions for linear constraints) *For a convex set $C \subset \mathbf{E}$, a function $f : C \to \mathbf{R}$, a linear map $A : \mathbf{E} \to \mathbf{Y}$ (where $\mathbf{Y}$ is a Euclidean space) and a point b in $\mathbf{Y}$, consider the optimization problem*

$$\inf\{f(x) \mid x \in C,\ Ax = b\}. \tag{2.1.4}$$

Suppose the point $\bar{x} \in \operatorname{int} C$ satisfies $A\bar{x} = b$.

(a) *If $\bar{x}$ is a local minimizer for the problem (2.1.4) and f is differentiable at $\bar{x}$ then $\nabla f(\bar{x}) \in A^*\mathbf{Y}$.*

(b) *Conversely, if $\nabla f(\bar{x}) \in A^*\mathbf{Y}$ and f is convex then $\bar{x}$ is a global minimizer for (2.1.4).*

The element $y \in \mathbf{Y}$ satisfying $\nabla f(\bar{x}) = A^* y$ in the above result is called a *Lagrange multiplier*. This kind of construction recurs in many different forms in our development.

In the absence of convexity, we need second order information to tell us more about minimizers. The following elementary result from multivariate calculus is typical.

Theorem 2.1.5 (Second order conditions) *Suppose the twice continuously differentiable function $f : \mathbf{R}^n \to \mathbf{R}$ has a critical point $\bar{x}$. If $\bar{x}$ is a local minimizer then the Hessian $\nabla^2 f(\bar{x})$ is positive semidefinite. Conversely, if the Hessian is positive definite then $\bar{x}$ is a local minimizer.*

(In fact for $\bar{x}$ to be a local minimizer it is sufficient for the Hessian to be positive semidefinite locally; the function $x \in \mathbf{R} \mapsto x^4$ highlights the distinction.)

To illustrate the effect of constraints on second order conditions, consider the framework of Corollary 2.1.3 (First order conditions for linear constraints) in the case $\mathbf{E} = \mathbf{R}^n$, and suppose $\nabla f(\bar{x}) \in A^*\mathbf{Y}$ and f is twice continuously differentiable near $\bar{x}$. If $\bar{x}$ is a local minimizer then $y^T \nabla^2 f(\bar{x}) y \geq 0$ for all vectors y in $N(A)$. Conversely, if $y^T \nabla^2 f(\bar{x}) y > 0$ for all nonzero y in $N(A)$ then $\bar{x}$ is a local minimizer.

We are already beginning to see the broad interplay between analytic, geometric and topological ideas in optimization theory. A good illustration is the separation result of Section 1.1, which we now prove.

Theorem 2.1.6 (Basic separation) *Suppose that the set $C \subset \mathbf{E}$ is closed and convex, and that the point y does not lie in C. Then there exist a real b and a nonzero element a of $\mathbf{E}$ such that $\langle a, y \rangle > b \geq \langle a, x \rangle$ for all points x in C.*

Proof. We may assume C is nonempty, and define a function $f : \mathbf{E} \to \mathbf{R}$ by $f(x) = \|x - y\|^2/2$. Now by the Weierstrass proposition (1.1.3) there exists a minimizer $\bar{x}$ for f on C, which by the First order necessary condition (2.1.1) satisfies $-\nabla f(\bar{x}) = y - \bar{x} \in N_C(\bar{x})$. Thus $\langle y - \bar{x}, x - \bar{x} \rangle \leq 0$ holds for all points x in C. Now setting $a = y - \bar{x}$ and $b = \langle y - \bar{x}, \bar{x} \rangle$ gives the result. $\square$

We end this section with a rather less standard result, illustrating another idea which is important later, the use of "variational principles" to treat problems where minimizers may not exist, but which nonetheless have "approximate" critical points. This result is a precursor of a principle due to Ekeland, which we develop in Section 7.1.

Proposition 2.1.7 *If the function $f : \mathbf{E} \to \mathbf{R}$ is differentiable and bounded below then there are points where f has small derivative.*

Proof. Fix any real $\epsilon > 0$. The function $f + \epsilon\|\cdot\|$ has bounded level sets, so has a global minimizer x^ϵ by the Weierstrass proposition (1.1.3). If the vector $d = \nabla f(x^\epsilon)$ satisfies $\|d\| > \epsilon$ then, from the inequality

$$\lim_{t\downarrow 0} \frac{f(x^\epsilon - td) - f(x^\epsilon)}{t} = -\langle \nabla f(x^\epsilon), d\rangle = -\|d\|^2 < -\epsilon\|d\|,$$

we would have for small $t > 0$ the contradiction

$$\begin{aligned} -t\epsilon\|d\| &> f(x^\epsilon - td) - f(x^\epsilon) \\ &= (f(x^\epsilon - td) + \epsilon\|x^\epsilon - td\|) \\ &\quad - (f(x^\epsilon) + \epsilon\|x^\epsilon\|) + \epsilon(\|x^\epsilon\| - \|x^\epsilon - td\|) \\ &\geq -\epsilon t\|d\| \end{aligned}$$

by definition of x^ϵ and the triangle inequality. Hence $\|\nabla f(x^\epsilon)\| \leq \epsilon$. □

Notice that the proof relies on consideration of a *nondifferentiable* function, even though the result concerns derivatives.

Exercises and Commentary

The optimality conditions in this section are very standard (see for example [132]). The simple variational principle (Proposition 2.1.7) was suggested by [95].

1. Prove the normal cone is a closed convex cone.

2. **(Examples of normal cones)** For the following sets $C \subset \mathbf{E}$, check C is convex and compute the normal cone $N_C(\bar{x})$ for points $\bar{x}$ in C:

   (a) C a closed interval in $\mathbf{R}$.

   (b) $C = B$, the unit ball.

   (c) C a subspace.

   (d) C a closed halfspace: $\{x \mid \langle a, x\rangle \leq b\}$ where $0 \neq a \in \mathbf{E}$ and $b \in \mathbf{R}$.

   (e) $C = \{x \in \mathbf{R}^n \mid x_j \geq 0 \text{ for all } j \in J\}$ (for $J \subset \{1, 2, \ldots, n\}$).

3. **(Self-dual cones)** Prove each of the following cones K satisfy the relationship $N_K(0) = -K$.

   (a) $\mathbf{R}^n_+$

   (b) $\mathbf{S}^n_+$

   (c) $\{x \in \mathbf{R}^n \mid x_1 \geq 0,\ x_1^2 \geq x_2^2 + x_3^2 + \cdots + x_n^2\}$

4. **(Normals to affine sets)** Given a linear map $A : \mathbf{E} \to \mathbf{Y}$ (where $\mathbf{Y}$ is a Euclidean space) and a point b in $\mathbf{Y}$, prove the normal cone to the set $\{x \in \mathbf{E} \mid Ax = b\}$ at any point in it is A^*Y. Hence deduce Corollary 2.1.3 (First order conditions for linear constraints).

5. Prove that the differentiable function $x_1^2 + x_2^2(1 - x_1)^3$ has a unique critical point in $\mathbf{R}^2$, which is a local minimizer, but has no global minimizer. Can this happen on $\mathbf{R}$?

6. **(The Rayleigh quotient)**

   (a) Let the function $f : \mathbf{R}^n \setminus \{0\} \to \mathbf{R}$ be continuous, satisfying $f(\lambda x) = f(x)$ for all $\lambda > 0$ in $\mathbf{R}$ and nonzero x in $\mathbf{R}^n$. Prove f has a minimizer.

   (b) Given a matrix A in $\mathbf{S}^n$, define a function $g(x) = x^T Ax/\|x\|^2$ for nonzero x in $\mathbf{R}^n$. Prove g has a minimizer.

   (c) Calculate $\nabla g(x)$ for nonzero x.

   (d) Deduce that minimizers of g must be eigenvectors, and calculate the minimum value.

   (e) Find an alternative proof of part (d) by using a spectral decomposition of A.

   (Another approach to this problem is given in Section 7.2, Exercise 6.)

7. Suppose a convex function $g : [0, 1] \to \mathbf{R}$ satisfies $g(0) = 0$. Prove the function $t \in (0, 1] \mapsto g(t)/t$ is nondecreasing. Hence prove that for a convex function $f : C \to \mathbf{R}$ and points $\bar{x}, x \in C \subset \mathbf{E}$, the quotient $(f(\bar{x} + t(x - \bar{x})) - f(\bar{x}))/t$ is nondecreasing as a function of t in $(0, 1]$, and complete the proof of Proposition 2.1.2.

8. * **(Nearest points)**

   (a) Prove that if a function $f : C \to \mathbf{R}$ is strictly convex then it has at most one global minimizer on C.

   (b) Prove the function $f(x) = \|x - y\|^2/2$ is strictly convex on $\mathbf{E}$ for any point y in $\mathbf{E}$.

   (c) Suppose C is a nonempty, closed convex subset of $\mathbf{E}$.

      (i) If y is any point in $\mathbf{E}$, prove there is a unique nearest point (or *best approximation*) $P_C(y)$ to y in C, characterized by

      $$\langle y - P_C(y), x - P_C(y)\rangle \le 0 \quad \text{for all } x \in C.$$

      (ii) For any point $\bar{x}$ in C, deduce that $d \in N_C(\bar{x})$ holds if and only if $\bar{x}$ is the nearest point in C to $\bar{x} + d$.

(iii) Deduce, furthermore, that any points y and z in $\mathbf{E}$ satisfy

$$\|P_C(y) - P_C(z)\| \leq \|y - z\|,$$

so in particular the *projection* $P_C : \mathbf{E} \to C$ is continuous.

(d) Given a nonzero element a of $\mathbf{E}$, calculate the nearest point in the subspace $\{x \in \mathbf{E} \mid \langle a, x \rangle = 0\}$ to the point $y \in \mathbf{E}$.

(e) **(Projection on $\mathbf{R}^n_+$ and $\mathbf{S}^n_+$)** Prove the nearest point in $\mathbf{R}^n_+$ to a vector y in $\mathbf{R}^n$ is y^+, where $y_i^+ = \max\{y_i, 0\}$ for each i. For a matrix U in $\mathbf{O}^n$ and a vector y in $\mathbf{R}^n$, prove that the nearest positive semidefinite matrix to $U^T \operatorname{Diag} y U$ is $U^T \operatorname{Diag} y^+ U$.

9. * **(Coercivity)** Suppose that the function $f : \mathbf{E} \to \mathbf{R}$ is differentiable and satisfies the growth condition $\lim_{\|x\| \to \infty} f(x)/\|x\| = +\infty$. Prove that the gradient map ∇f has range $\mathbf{E}$. (Hint: Minimize the function $f(\cdot) - \langle a, \cdot \rangle$ for elements a of $\mathbf{E}$.)

10. (a) Prove the function $f : \mathbf{S}^n_{++} \to \mathbf{R}$ defined by $f(X) = \operatorname{tr} X^{-1}$ is differentiable on $\mathbf{S}^n_{++}$. (Hint: Expand the expression $(X + tY)^{-1}$ as a power series.)

    (b) Define a function $f : \mathbf{S}^n_{++} \to \mathbf{R}$ by $f(X) = \log \det X$. Prove $\nabla f(I) = I$. Deduce $\nabla f(X) = X^{-1}$ for any X in $\mathbf{S}^n_{++}$.

11. ** **(Kirchhoff's law [9, Chapter 1])** Consider a finite, undirected, connected graph with vertex set V and edge set E. Suppose that α and β in V are distinct vertices and that each edge ij in E has an associated "resistance" $r_{ij} > 0$ in $\mathbf{R}$. We consider the effect of applying a unit "potential difference" between the vertices α and β. Let $V_0 = V \setminus \{\alpha, \beta\}$, and for "potentials" x in $\mathbf{R}^{V_0}$ we define the "power" $p : \mathbf{R}^{V_0} \to \mathbf{R}$ by

$$p(x) = \sum_{ij \in E} \frac{(x_i - x_j)^2}{2r_{ij}},$$

where we set $x_\alpha = 0$ and $x_\beta = 1$.

(a) Prove the power function p has compact level sets.

(b) Deduce the existence of a solution to the following equations (describing "conservation of current"):

$$\begin{aligned} \sum_{j \,:\, ij \in E} \frac{x_i - x_j}{r_{ij}} &= 0 \quad \text{for } i \text{ in } V_0 \\ x_\alpha &= 0 \\ x_\beta &= 1. \end{aligned}$$

(c) Prove the power function p is strictly convex.

(d) Use part (a) of Exercise 8 to show that the conservation of current equations in part (b) have a unique solution.

12. ** **(Matrix completion [86])** For a set $\Delta \subset \{(i,j) \,|\, 1 \le i \le j \le n\}$, suppose the subspace $L \subset \mathbf{S}^n$ of matrices with (i,j)th entry of zero for all (i,j) in Δ satisfies $L \cap \mathbf{S}^n_{++} \neq \emptyset$. By considering the problem (for $C \in \mathbf{S}^n_{++}$)

$$\inf\{\langle C, X\rangle - \log\det X \mid X \in L \cap \mathbf{S}^n_{++}\},$$

use Section 1.2, Exercise 14 and Corollary 2.1.3 (First order conditions for linear constraints) to prove there exists a matrix X in $L \cap \mathbf{S}^n_{++}$ with $C - X^{-1}$ having (i,j)th entry of zero for all (i,j) not in Δ.

13. ** **(BFGS update, cf. [80])** Given a matrix C in $\mathbf{S}^n_{++}$ and vectors s and y in $\mathbf{R}^n$ satisfying $s^T y > 0$, consider the problem

$$\inf\{\langle C, X\rangle - \log\det X \mid Xs = y,\ X \in \mathbf{S}^n_{++}\}.$$

(a) Prove that for the problem above, the point

$$X = \frac{(y - \delta s)(y - \delta s)^T}{s^T(y - \delta s)} + \delta I$$

is feasible for small $\delta > 0$.

(b) Prove the problem has an optimal solution using Section 1.2, Exercise 14.

(c) Use Corollary 2.1.3 (First order conditions for linear constraints) to find the solution. (The solution is called the *BFGS update* of C^{-1} under the *secant condition* $Xs = y$.)

(See also [61, p. 205].)

14. ** Suppose intervals $I_1, I_2, \ldots, I_n \subset \mathbf{R}$ are nonempty and closed and the function $f : I_1 \times I_2 \times \ldots \times I_n \to \mathbf{R}$ is differentiable and bounded below. Use the idea of the proof of Proposition 2.1.7 to prove that for any $\epsilon > 0$ there exists a point $x^\epsilon \in I_1 \times I_2 \times \ldots \times I_n$ satisfying

$$(-\nabla f(x^\epsilon))_j \in N_{I_j}(x^\epsilon_j) + [-\epsilon, \epsilon] \quad (j = 1, 2, \ldots, n).$$

15. * **(Nearest polynomial with a given root)** Consider the Euclidean space of complex polynomials of degree no more than n, with inner product

$$\Big\langle \sum_{j=0}^n x_j z^j\,,\ \sum_{j=0}^n y_j z^j \Big\rangle = \sum_{j=0}^n \overline{x_j} y_j.$$

Given a polynomial p in this space, calculate the nearest polynomial with a given complex root α, and prove the distance to this polynomial is $(\sum_{j=0}^{n} |\alpha|^{2j})^{(-1/2)}|p(\alpha)|$.

2.2 Theorems of the Alternative

One well-trodden route to the study of first order conditions uses a class of results called "theorems of the alternative", and, in particular, the Farkas lemma (which we derive at the end of this section). Our first approach, however, relies on a different theorem of the alternative.

Theorem 2.2.1 (Gordan) *For any elements $a^0, a^1, \ldots, a^m$ of $\mathbf{E}$, exactly one of the following systems has a solution:*

$$\sum_{i=0}^{m} \lambda_i a^i = 0, \quad \sum_{i=0}^{m} \lambda_i = 1, \quad 0 \le \lambda_0, \lambda_1, \ldots, \lambda_m \in \mathbf{R} \tag{2.2.2}$$

$$\langle a^i, x \rangle < 0 \quad \text{for } i = 0, 1, \ldots, m, \quad x \in \mathbf{E}. \tag{2.2.3}$$

Geometrically, Gordan's theorem says that the origin does not lie in the convex hull of the set $\{a^0, a^1, \ldots, a^m\}$ if and only if there is an open halfspace $\{y \mid \langle y, x \rangle < 0\}$ containing $\{a^0, a^1, \ldots, a^m\}$ (and hence its convex hull). This is another illustration of the idea of separation (in this case we separate the origin and the convex hull).

Theorems of the alternative like Gordan's theorem may be proved in a variety of ways, including separation and algorithmic approaches. We employ a less standard technique using our earlier analytic ideas and leading to a rather unified treatment. It relies on the relationship between the optimization problem

$$\inf\{f(x) \mid x \in \mathbf{E}\}, \tag{2.2.4}$$

where the function f is defined by

$$f(x) = \log\Big(\sum_{i=0}^{m} \exp\langle a^i, x \rangle\Big), \tag{2.2.5}$$

and the two systems (2.2.2) and (2.2.3). We return to the surprising function (2.2.5) when we discuss conjugacy in Section 3.3.

Theorem 2.2.6 *The following statements are equivalent:*

(*i*) *The function defined by (2.2.5) is bounded below.*

(*ii*) *System (2.2.2) is solvable.*

(*iii*) *System (2.2.3) is unsolvable.*

Proof. The implications (ii) $\Rightarrow$ (iii) $\Rightarrow$ (i) are easy exercises, so it remains to show (i) $\Rightarrow$ (ii). To see this we apply Proposition 2.1.7. We deduce that for each $k = 1, 2, \ldots,$ there is a point x^k in $\mathbf{E}$ satisfying

$$\|\nabla f(x^k)\| = \Big\| \sum_{i=0}^{m} \lambda_i^k a^i \Big\| < \frac{1}{k},$$

where the scalars

$$\lambda_i^k = \frac{\exp\langle a^i, x^k\rangle}{\sum_{r=0}^m \exp\langle a^r, x^k\rangle} > 0$$

satisfy $\sum_{i=0}^m \lambda_i^k = 1$. Now the limit λ of any convergent subsequence of the bounded sequence (λ^k) solves system (2.2.2). $\square$

The equivalence of (ii) and (iii) gives Gordan's theorem.

We now proceed by using Gordan's theorem to derive the Farkas lemma, one of the cornerstones of many approaches to optimality conditions. The proof uses the idea of the *projection* onto a linear subspace $\mathbf{Y}$ of $\mathbf{E}$. Notice first that $\mathbf{Y}$ becomes a Euclidean space by equipping it with the same inner product. The projection of a point x in $\mathbf{E}$ onto $\mathbf{Y}$, written $P_{\mathbf{Y}}x$, is simply the nearest point to x in $\mathbf{Y}$. This is well-defined (see Exercise 8 in Section 2.1), and is characterized by the fact that $x - P_{\mathbf{Y}}x$ is orthogonal to $\mathbf{Y}$. A standard exercise shows $P_{\mathbf{Y}}$ is a linear map.

Lemma 2.2.7 (Farkas) *For any points $a^1, a^2, \ldots, a^m$ and c in $\mathbf{E}$, exactly one of the following systems has a solution:*

$$\sum_{i=1}^m \mu_i a^i = c, \quad 0 \le \mu_1, \mu_2, \ldots, \mu_m \in \mathbf{R} \tag{2.2.8}$$

$$\langle a^i, x\rangle \le 0 \quad \textit{for} \quad i = 1, 2, \ldots, m, \quad \langle c, x\rangle > 0, \quad x \in \mathbf{E}. \tag{2.2.9}$$

Proof. Again, it is immediate that if system (2.2.8) has a solution then system (2.2.9) has no solution. Conversely, we assume (2.2.9) has no solution and deduce that (2.2.8) has a solution by using induction on the number of elements m. The result is clear for $m = 0$.

Suppose then that the result holds in any Euclidean space and for any set of $m - 1$ elements and any element c. Define $a^0 = -c$. Applying Gordan's theorem (2.2.1) to the unsolvability of (2.2.9) shows there are scalars $\lambda_0, \lambda_1, \ldots, \lambda_m \ge 0$ in $\mathbf{R}$, not all zero, satisfying $\lambda_0 c = \sum_1^m \lambda_i a^i$. If $\lambda_0 > 0$ the proof is complete, so suppose $\lambda_0 = 0$ and without loss of generality $\lambda_m > 0$.

Define a subspace of $\mathbf{E}$ by $\mathbf{Y} = \{y \mid \langle a^m, y\rangle = 0\}$, so by assumption the system

$$\langle a^i, y\rangle \le 0 \ \text{ for } i = 1, 2, \ldots, m-1, \quad \langle c, y\rangle > 0, \quad y \in \mathbf{Y},$$

or equivalently

$$\langle P_{\mathbf{Y}} a^i, y\rangle \le 0 \ \text{ for } i = 1, 2, \ldots, m-1, \quad \langle P_{\mathbf{Y}} c, y\rangle > 0, \quad y \in \mathbf{Y},$$

has no solution.

By the induction hypothesis applied to the subspace $\mathbf{Y}$, there are nonnegative reals $\mu_1, \mu_2, \ldots, \mu_{m-1}$ satisfying $\sum_{i=1}^{m-1} \mu_i P_{\mathbf{Y}} a^i = P_{\mathbf{Y}} c$, so the vector $c - \sum_1^{m-1} \mu_i a^i$ is orthogonal to the subspace $\mathbf{Y} = (\mathrm{span}\,(a^m))^\perp$. Thus some real μ_m satisfies

$$\mu_m a^m = c - \sum_1^{m-1} \mu_i a^i. \tag{2.2.10}$$

If μ_m is nonnegative we immediately obtain a solution of (2.2.8), and if not then we can substitute $a^m = -\lambda_m^{-1} \sum_1^{m-1} \lambda_i a^i$ in equation (2.2.10) to obtain a solution. $\Box$

Just like Gordan's theorem, the Farkas lemma has an important geometric interpretation which gives an alternative approach to its proof (Exercise 6): any point c not lying in the *finitely generated cone*

$$C = \Big\{ \sum_1^m \mu_i a^i \,\Big|\, 0 \le \mu_1, \mu_2, \ldots, \mu_m \in \mathbf{R} \Big\} \tag{2.2.11}$$

can be separated from C by a hyperplane. If x solves system (2.2.9) then C is contained in the closed halfspace $\{a \mid \langle a, x\rangle \le 0\}$, whereas c is contained in the complementary open halfspace. In particular, it follows that any finitely generated cone is closed.

Exercises and Commentary

Gordan's theorem appeared in [84], and the Farkas lemma appeared in [75]. The standard modern approach to theorems of the alternative (Exercises 7 and 8, for example) is via linear programming duality (see, for example, [53]). The approach we take to Gordan's theorem was suggested by Hiriart–Urruty [95]. Schur-convexity (Exercise 9) is discussed extensively in [134].

1. Prove the implications (ii) $\Rightarrow$ (iii) $\Rightarrow$ (i) in Theorem 2.2.6.
2. (a) Prove the orthogonal projection $P_{\mathbf{Y}} : \mathbf{E} \to \mathbf{Y}$ is a linear map.

   (b) Give a direct proof of the Farkas lemma for the case $m = 1$.
3. Use the Basic separation theorem (2.1.6) to give another proof of Gordan's theorem.
4. * Deduce Gordan's theorem from the Farkas lemma. (Hint: Consider the elements $(a^i, 1)$ of the space $\mathbf{E} \times \mathbf{R}$.)
5. * **(Carathéodory's theorem [52])** Suppose $\{a^i \mid i \in I\}$ is a finite set of points in $\mathbf{E}$. For any subset J of I, define the cone

$$C_J = \Big\{ \sum_{i \in J} \mu_i a^i \,\Big|\, 0 \le \mu_i \in \mathbf{R},\ (i \in J) \Big\}.$$

(a) Prove the cone C_I is the union of those cones C_J for which the set $\{a^i \mid i \in J\}$ is linearly independent. Furthermore, prove directly that any such cone C_J is closed.

(b) Deduce that any finitely generated cone is closed.

(c) If the point x lies in $\operatorname{conv}\{a^i \mid i \in I\}$, prove that in fact there is a subset $J \subset I$ of size at most $1 + \dim \mathbf{E}$ such that x lies in $\operatorname{conv}\{a^i \mid i \in J\}$. (Hint: Apply part (a) to the vectors $(a^i, 1)$ in $\mathbf{E} \times \mathbf{R}$.)

(d) Use part (c) to prove that if a subset of $\mathbf{E}$ is compact then so is its convex hull.

6. * Give another proof of the Farkas lemma by applying the Basic separation theorem (2.1.6) to the set defined by (2.2.11) and using the fact that any finitely generated cone is closed.

7. ** **(Ville's theorem)** With the function f defined by (2.2.5) (with $\mathbf{E} = \mathbf{R}^n$), consider the optimization problem

$$\inf\{f(x) \mid x \geq 0\} \tag{2.2.12}$$

and its relationship with the two systems

$$\sum_{i=0}^{m} \lambda_i a^i \geq 0, \quad \sum_{i=0}^{m} \lambda_i = 1, \\ 0 \leq \lambda_0, \lambda_1, \ldots, \lambda_m \in \mathbf{R} \tag{2.2.13}$$

and

$$\langle a^i, x \rangle < 0 \quad \text{for } i = 0, 1, \ldots, m, \quad x \in \mathbf{R}^n_+. \tag{2.2.14}$$

Imitate the proof of Gordan's theorem (using Section 2.1, Exercise 14) to prove the following are equivalent:

(i) Problem (2.2.12) is bounded below.

(ii) System (2.2.13) is solvable.

(iii) System (2.2.14) is unsolvable.

Generalize by considering the problem $\inf\{f(x) \mid x_j \geq 0 \ (j \in J)\}$.

8. ** **(Stiemke's theorem)** Consider the optimization problem (2.2.4) and its relationship with the two systems

$$\sum_{i=0}^{m} \lambda_i a^i = 0, \quad 0 < \lambda_0, \lambda_1, \ldots, \lambda_m \in \mathbf{R} \tag{2.2.15}$$

and

$$\langle a^i, x \rangle \leq 0 \quad \text{for } i = 0, 1, \ldots, m, \quad \text{not all } 0, \quad x \in \mathbf{E}. \tag{2.2.16}$$

Prove the following are equivalent:

(i) Problem (2.2.4) has an optimal solution.
(ii) System (2.2.15) is solvable.
(iii) System (2.2.16) is unsolvable.

Hint: Complete the following steps.

(a) Prove (i) implies (ii) by Proposition 2.1.1.
(b) Prove (ii) implies (iii).
(c) If problem (2.2.4) has no optimal solution, prove that neither does the problem

$$\inf\Big\{\sum_{i=0}^{m}\exp y_i \;\Big|\; y \in K\Big\}, \tag{2.2.17}$$

where K is the subspace $\{(\langle a^i, x\rangle)_{i=0}^{m} \mid x \in \mathbf{E}\} \subset \mathbf{R}^{m+1}$. Hence, by considering a minimizing sequence for (2.2.17), deduce system (2.2.16) is solvable.

Generalize by considering the problem $\inf\{f(x) \mid x_j \geq 0 \ (j \in J)\}$.

9. ** **(Schur-convexity)** The *dual cone* of the cone $\mathbf{R}^n_{\geq}$ is defined by

$$\big(\mathbf{R}^n_{\geq}\big)^+ = \{y \in \mathbf{R}^n \mid \langle x, y\rangle \geq 0 \text{ for all } x \text{ in } \mathbf{R}^n_{\geq}\}.$$

(a) Prove a vector y lies in $(\mathbf{R}^n_{\geq})^+$ if and only if

$$\sum_1^j y_i \geq 0 \;\text{ for } j = 1, 2, \ldots, n-1, \quad \sum_1^n y_i = 0.$$

(b) By writing $\sum_1^j [x]_i = \max_k \langle a^k, x\rangle$ for some suitable set of vectors a^k, prove that the function $x \mapsto \sum_1^j [x]_i$ is convex. (Hint: Use Section 1.1, Exercise 7.)
(c) Deduce that the function $x \mapsto [x]$ is $(\mathbf{R}^n_{\geq})^+$-*convex*, that is:

$$\lambda[x] + (1-\lambda)[y] - [\lambda x + (1-\lambda)y] \in \big(\mathbf{R}^n_{\geq}\big)^+ \;\text{ for } 0 \leq \lambda \leq 1.$$

(d) Use Gordan's theorem and Proposition 1.2.4 to deduce that for any x and y in $\mathbf{R}^n_{\geq}$, if $y - x$ lies in $(\mathbf{R}^n_{\geq})^+$ then x lies in $\operatorname{conv}(\mathbf{P}^n y)$.
(e) A function $f : \mathbf{R}^n_{\geq} \to \mathbf{R}$ is *Schur-convex* if

$$x, y \in \mathbf{R}^n_{\geq}, \; y - x \in \big(\mathbf{R}^n_{\geq}\big)^+ \quad \Rightarrow \quad f(x) \leq f(y).$$

Prove that if f is convex, then it is Schur-convex if and only if it is the restriction to $\mathbf{R}^n_{\geq}$ of a *symmetric* convex function $g : \mathbf{R}^n \to \mathbf{R}$ (where by symmetric we mean $g(x) = g(\Pi x)$ for any x in $\mathbf{R}^n$ and any permutation matrix Π).

2.3 Max-functions

This section is an elementary exposition of the first order necessary conditions for a local minimizer of a differentiable function subject to differentiable inequality constraints. Throughout this section we use the term "differentiable" in the Gâteaux sense, defined in Section 2.1. Our approach, which relies on considering the local minimizers of a *max-function*

$$g(x) = \max_{i=0,1,\ldots,m} \{g_i(x)\}, \tag{2.3.1}$$

illustrates a pervasive analytic idea in optimization: *nonsmoothness*. Even if the functions $g_0, g_1, \ldots, g_m$ are smooth, g may not be, and hence the gradient may no longer be a useful notion.

Proposition 2.3.2 (Directional derivatives of max-functions) *Let $\bar{x}$ be a point in the interior of a set $C \subset \mathbf{E}$. Suppose that continuous functions $g_0, g_1, \ldots, g_m : C \to \mathbf{R}$ are differentiable at $\bar{x}$, that g is the max-function (2.3.1), and define the index set $K = \{i \mid g_i(\bar{x}) = g(\bar{x})\}$. Then for all directions d in $\mathbf{E}$, the directional derivative of g is given by*

$$g'(\bar{x}; d) = \max_{i \in K} \{\langle \nabla g_i(\bar{x}), d \rangle\}. \tag{2.3.3}$$

Proof. By continuity we can assume, without loss of generality, $K = \{0, 1, \ldots, m\}$; those g_i not attaining the maximum in (2.3.1) will not affect $g'(\bar{x}; d)$. Now for each i, we have the inequality

$$\liminf_{t \downarrow 0} \frac{g(\bar{x} + td) - g(\bar{x})}{t} \geq \lim_{t \downarrow 0} \frac{g_i(\bar{x} + td) - g_i(\bar{x})}{t} = \langle \nabla g_i(\bar{x}), d \rangle.$$

Suppose

$$\limsup_{t \downarrow 0} \frac{g(\bar{x} + td) - g(\bar{x})}{t} > \max_i \{\langle \nabla g_i(\bar{x}), d \rangle\}.$$

Then some real sequence $t_k \downarrow 0$ and real $\epsilon > 0$ satisfy

$$\frac{g(\bar{x} + t_k d) - g(\bar{x})}{t_k} \geq \max_i \{\langle \nabla g_i(\bar{x}), d \rangle\} + \epsilon \quad \text{for all } k \in \mathbf{N}$$

(where $\mathbf{N}$ denotes the sequence of natural numbers). We can now choose a subsequence R of $\mathbf{N}$ and a fixed index j so that all integers k in R satisfy $g(\bar{x} + t_k d) = g_j(\bar{x} + t_k d)$. In the limit we obtain the contradiction

$$\langle \nabla g_j(\bar{x}), d \rangle \geq \max_i \{\langle \nabla g_i(\bar{x}), d \rangle\} + \epsilon.$$

Hence

$$\limsup_{t \downarrow 0} \frac{g(\bar{x} + td) - g(\bar{x})}{t} \leq \max_i \{\langle \nabla g_i(\bar{x}), d \rangle\},$$

and the result follows. □

For most of this book we consider optimization problems of the form

$$\left.\begin{array}{lrcl} \inf & f(x) & & \\ \text{subject to} & g_i(x) & \leq & 0 \ \text{ for } i \in I \\ & h_j(x) & = & 0 \ \text{ for } j \in J \\ & x & \in & C, \end{array}\right\} \tag{2.3.4}$$

where C is a subset of $\mathbf{E}$, I and J are finite index sets, and the *objective function* f and *inequality* and *equality constraint functions* g_i $(i \in I)$ and h_j $(j \in J)$, respectively, are continuous from C to $\mathbf{R}$. A point x in C is *feasible* if it satisfies the constraints, and the set of all feasible x is called the *feasible region*. If the problem has no feasible points, we call it *inconsistent*. We say a feasible point $\bar{x}$ is a *local minimizer* if $f(x) \geq f(\bar{x})$ for all feasible x close to $\bar{x}$. We aim to derive first order necessary conditions for local minimizers.

We begin in this section with the differentiable inequality constrained problem

$$\left.\begin{array}{lrcl} \inf & f(x) & & \\ \text{subject to} & g_i(x) & \leq & 0 \ \text{ for } i = 1, 2, \ldots, m \\ & x & \in & C. \end{array}\right\} \tag{2.3.5}$$

For a feasible point $\bar{x}$ we define the *active set* $I(\bar{x}) = \{i \mid g_i(\bar{x}) = 0\}$. For this problem, assuming $\bar{x} \in \operatorname{int} C$, we call a vector $\lambda \in \mathbf{R}^m_+$ a *Lagrange multiplier vector* for $\bar{x}$ if $\bar{x}$ is a critical point of the *Lagrangian*

$$L(x; \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$

(in other words, $\nabla f(\bar{x}) + \sum \lambda_i \nabla g_i(\bar{x}) = 0$), and *complementary slackness* holds: $\lambda_i = 0$ for indices i not in $I(\bar{x})$.

Theorem 2.3.6 (Fritz John conditions) *Suppose problem (2.3.5) has a local minimizer $\bar{x} \in \operatorname{int} C$. If the functions f, g_i $(i \in I(\bar{x}))$ are differentiable at $\bar{x}$ then there exist $\lambda_0, \lambda_i \in \mathbf{R}_+$ $(i \in I(\bar{x}))$, not all zero, satisfying*

$$\lambda_0 \nabla f(\bar{x}) + \sum_{i \in I(\bar{x})} \lambda_i \nabla g_i(\bar{x}) = 0.$$

Proof. Consider the function

$$g(x) = \max\{f(x) - f(\bar{x}),\ g_i(x) \mid i \in I(\bar{x})\}.$$

Since $\bar{x}$ is a local minimizer for the problem (2.3.5), it is a local minimizer of the function g, so all directions $d \in \mathbf{E}$ satisfy the inequality

$$g'(\bar{x}; d) = \max\{\langle \nabla f(\bar{x}), d\rangle, \langle \nabla g_i(\bar{x}), d\rangle \mid i \in I(\bar{x})\} \geq 0,$$

by the First order necessary condition (2.1.1) and Proposition 2.3.2 (Directional derivatives of max-functions). Thus the system

$$\langle \nabla f(\bar{x}), d \rangle < 0, \quad \langle \nabla g_i(\bar{x}), d \rangle < 0 \text{ for } i \in I(\bar{x})$$

has no solution, and the result follows by Gordan's theorem (2.2.1). □

One obvious disadvantage remains with the Fritz John first order conditions above: if $\lambda_0 = 0$ then the conditions are independent of the objective function f. To rule out this possibility we need to impose a regularity condition or "constraint qualification", an approach which is another recurring theme. The easiest such condition in this context is simply the linear independence of the gradients of the active constraints $\{\nabla g_i(\bar{x}) \mid i \in I(\bar{x})\}$. The culminating result of this section uses the following weaker condition.

Assumption 2.3.7 (The Mangasarian–Fromovitz constraint qualification) *There is a direction d in $\mathbf{E}$ satisfying $\langle \nabla g_i(\bar{x}), d \rangle < 0$ for all indices i in the active set $I(\bar{x})$.*

Theorem 2.3.8 (Karush–Kuhn–Tucker conditions) *Suppose problem (2.3.5) has a local minimizer $\bar{x}$ in $\mathrm{int}\, C$. If the functions f, g_i (for $i \in I(\bar{x})$) are differentiable at $\bar{x}$, and if the Mangasarian–Fromovitz constraint qualification (2.3.7) holds, then there is a Lagrange multiplier vector for $\bar{x}$.*

Proof. By the trivial implication in Gordan's theorem (2.2.1), the constraint qualification ensures $\lambda_0 \neq 0$ in the Fritz John conditions (2.3.6). □

Exercises and Commentary

The approach to first order conditions of this section is due to [95]. The Fritz John conditions appeared in [107]. The Karush–Kuhn–Tucker conditions were first published (under a different regularity condition) in [117], although the conditions appear earlier in an unpublished master's thesis [111]. The Mangasarian–Fromovitz constraint qualification appeared in [133]. A nice collection of optimization problems involving the determinant, similar to Exercise 8 (Minimum volume ellipsoid), appears in [47] (see also [183]). The classic reference for inequalities is [91].

1. Prove by induction that if the functions $g_0, g_1, \ldots, g_m : \mathbf{E} \to \mathbf{R}$ are all continuous at the point $\bar{x}$ then so is the max-function $g(x) = \max_i \{g_i(x)\}$.

2. **(Failure of Karush–Kuhn–Tucker)** Consider the following problem:

$$\begin{array}{lrcl} \inf & (x_1+1)^2 + x_2^2 & & \\ \text{subject to} & -x_1^3 + x_2^2 & \leq & 0 \\ & x & \in & \mathbf{R}^2. \end{array}$$

(a) Sketch the feasible region and hence solve the problem.

(b) Find multipliers λ_0 and λ satisfying the Fritz John conditions (2.3.6).

(c) Prove there exists no Lagrange multiplier vector for the optimal solution. Explain why not.

3. **(Linear independence implies Mangasarian–Fromovitz)** If the set of vectors $\{a^1, a^2, \ldots, a^m\}$ in $\mathbf{E}$ is linearly independent, prove directly there exists a direction d in $\mathbf{E}$ satisfying $\langle a^i, d\rangle < 0$ for $i = 1, 2, \ldots, m$.

4. For each of the following problems, explain why there must exist an optimal solution, and find it by using the Karush–Kuhn–Tucker conditions.

(a)
$$\begin{array}{lrcl} \inf & x_1^2 + x_2^2 & & \\ \text{subject to} & -2x_1 - x_2 + 10 & \leq & 0 \\ & -x_1 & \leq & 0. \end{array}$$

(b)
$$\begin{array}{lrcl} \inf & 5x_1^2 + 6x_2^2 & & \\ \text{subject to} & x_1 - 4 & \leq & 0 \\ & 25 - x_1^2 - x_2^2 & \leq & 0. \end{array}$$

5. **(Cauchy–Schwarz and steepest descent)** For a nonzero vector y in $\mathbf{E}$, use the Karush–Kuhn–Tucker conditions to solve the problem

$$\inf\{\langle y, x\rangle \mid \|x\|^2 \leq 1\}.$$

Deduce the Cauchy–Schwarz inequality.

6. * **(Hölder's inequality)** For real $p > 1$, define q by $p^{-1} + q^{-1} = 1$, and for x in $\mathbf{R}^n$ define

$$\|x\|_p = \Big(\sum_1^n |x_i|^p\Big)^{1/p}.$$

For a nonzero vector y in $\mathbf{R}^n$, consider the optimization problem

$$\inf\{\langle y, x\rangle \mid \|x\|_p^p \leq 1\}. \tag{2.3.9}$$

(a) Prove $\frac{d}{du}|u|^p/p = u|u|^{p-2}$ for all real u.

(b) Prove reals u and v satisfy $v = u|u|^{p-2}$ if and only if $u = v|v|^{q-2}$.

(c) Prove problem (2.3.9) has a nonzero optimal solution.

(d) Use the Karush–Kuhn–Tucker conditions to find the unique optimal solution.

(e) Deduce that any vectors x and y in $\mathbf{R}^n$ satisfy $\langle y, x\rangle \le \|y\|_q \|x\|_p$.

(We develop another approach to this theory in Section 4.1, Exercise 11.)

7. * Consider a matrix A in $\mathbf{S}^n_{++}$ and a real $b > 0$.

(a) Assuming the problem

$$\inf\{-\log\det X \mid \operatorname{tr} AX \le b,\ X \in \mathbf{S}^n_{++}\}$$

has a solution, find it.

(b) Repeat using the objective function $\operatorname{tr} X^{-1}$.

(c) Prove the problems in parts (a) and (b) have optimal solutions. (Hint: Section 1.2, Exercise 14.)

8. ** **(Minimum volume ellipsoid)**

(a) For a point y in $\mathbf{R}^n$ and the function $g : \mathbf{S}^n \to \mathbf{R}$ defined by $g(X) = \|Xy\|^2$, prove $\nabla g(X) = Xyy^T + yy^T X$ for all matrices X in $\mathbf{S}^n$.

(b) Consider a set $\{y^1, y^2, \ldots, y^m\} \subset \mathbf{R}^n$. Prove this set spans $\mathbf{R}^n$ if and only if the matrix $\sum_i y^i (y^i)^T$ is positive definite.

Now suppose the vectors $y^1, y^2, \ldots, y^m$ span $\mathbf{R}^n$.

(c) Prove the problem

$$\begin{array}{lrcl} \inf & -\log\det X & & \\ \text{subject to} & \|Xy^i\|^2 - 1 & \le & 0 \quad \text{for } i = 1, 2, \ldots, m \\ & X & \in & \mathbf{S}^n_{++} \end{array}$$

has an optimal solution. (Hint: Use part (b) and Section 1.2, Exercise 14.)

Now suppose $\bar{X}$ is an optimal solution for the problem in part (c). (In this case the set $\{y \in \mathbf{R}^n \mid \|\bar{X}y\| \le 1\}$ is a minimum volume ellipsoid (centered at the origin) containing the vectors $y^1, y^2, \ldots, y^m$.)

(d) Show the Mangasarian–Fromovitz constraint qualification holds at $\bar{X}$ by considering the direction $d = -\bar{X}$.

(e) Write down the Karush–Kuhn–Tucker conditions that $\bar{X}$ must satisfy.

(f) When $\{y^1, y^2, \ldots, y^m\}$ is the standard basis of $\mathbf{R}^n$, the optimal solution of the problem in part (c) is $\bar{X} = I$. Find the corresponding Lagrange multiplier vector.

http://www.springer.com/978-0-387-29570-1

Convex Analysis and Nonlinear Optimization
Theory and Examples
Borwein, J.; Lewis, A.S.
2006, XII, 310 p., Hardcover
ISBN: 978-0-387-29570-1