

# Optimality Conditions for Constrained Optimization

## 1. First-Order Conditions

In this section we consider first-order optimality conditions for the constrained problem

$$\mathcal{P} : \begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & x \in \Omega, \end{array}$$

where  $f_0 : \mathbb{R}^n \rightarrow \mathbb{R}$  is continuously differentiable and  $\Omega \subset \mathbb{R}^n$  is closed and non-empty. The first step in the analysis of the problem  $\mathcal{P}$  is to derive conditions that allow us to recognize when a particular vector  $\bar{x}$  is a solution, or local solution, to the problem. For example, when we minimize a function of one variable we first take the derivative and see if it is zero. If it is, then we take the second derivative and check that it is positive. If this is also true, then we know that the point under consideration is a local minimizer of the function. Of course, the presence of constraints complicates this kind of test.

To understand how an optimality test might be derived in the constrained case, let us first suppose that we are at a feasible point  $x$  and we wish to find a better point  $\tilde{x}$ . That is, we wish to find a point  $\tilde{x}$  such that  $\tilde{x} \in \Omega$  and  $f(\tilde{x}) < f(x)$ . As in the unconstrained case, one way to do this is to find a direction  $d$  in which the directional derivative of  $f$  in the direction  $d$  is negative:  $f'(x; d) < 0$ . We know that for such directions we can reduce the value of the function by moving away from the point  $x$  in the direction  $d$ . However, moving in such a direction may violate feasibility. That is, it may happen that  $x + td \notin \Omega$  regardless of how small we take  $t > 0$ . To avoid this problem, we consider the notion of a *feasible direction*.

DEFINITION 1.1. [FEASIBLE DIRECTIONS]

Given a subset  $\Omega$  of  $\mathbb{R}^n$  and a point  $x \in \Omega$ , we say that a direction  $d \in \mathbb{R}^n$  is a *feasible direction* for  $\Omega$  at  $x$  if there is a  $\bar{t} > 0$  such that  $x + td \in \Omega$  for all  $t \in [0, \bar{t}]$ .

THEOREM 1.1. If  $\bar{x}$  is a local solution to the problem  $\mathcal{P}$ , then  $f'(\bar{x}; d) \geq 0$  for all feasible directions  $d$  for  $\Omega$  at  $\bar{x}$  for which  $f'(\bar{x}; d)$  exists.

PROOF. The proof is a straightforward application of the definitions. If the result were false, then there would be a direction of descent for  $f$  at  $\bar{x}$  that is also a feasible direction for  $\Omega$  at  $\bar{x}$ . But then moving a little bit in this direction both keeps us in  $\Omega$  and strictly reduces the value of  $f$ . This contradicts the assumption that  $\bar{x}$  is a local solution. Therefore, the result must be true.  $\square$

Unfortunately, this result is insufficient in many important cases. The insufficiency comes from the dependence on the notion of *feasible direction*. For example, if

$$\Omega = \{(x_1, x_2)^T : x_1^2 + x_2^2 = 1\},$$

then the only feasible direction at any point of  $\Omega$  is the zero direction. Hence, regardless of the objective function  $f$  and the point  $\bar{x} \in \Omega$ , we have that  $f'(\bar{x}; d) \geq 0$  for every feasible direction to  $\Omega$  at  $\bar{x}$ . In this case, Theorem 1.1 has no content.

To overcome this deficiency we introduce a general notion of *tangency* that considers all directions  $d$  pointing into  $\Omega$  at  $x \in \Omega$  in a limiting sense. Define the *tangent cone* to  $\Omega$  at a point  $x \in \Omega$  to be the set of limiting directions obtained from sequences in  $\Omega$  that converge to  $x$ . Specifically, the tangent cone is given by

$$T(x | \Omega) := \{d : \exists \tau_i \searrow 0 \text{ and } \{x_i\} \subset \Omega, \text{ with } x_i \rightarrow x, \text{ such that } \tau_i^{-1}(x_i - x) \rightarrow d\}.$$

EXAMPLE 1.1. (1) If  $\Omega = \{x : Ax = b\}$ , where  $A \in \mathbb{R}^{m \times n}$  and  $b \in \mathbb{R}^m$ , then  $T(x | \Omega) = \text{Nul}(A)$  for every  $x \in \Omega$ .

Reason: Let  $x \in \Omega$ . Note that if  $d \in \text{Nul}(A)$ , then for every  $t \geq 0$  we have  $A(x + td) = Ax + tAd = Ax = b$  so that  $d \in T(x|\Omega)$ . Since  $d \in \text{Nul}(A)$  was chosen arbitrarily, this implies that  $\text{Nul}(A) \subset T(x|\Omega)$ . Hence we only need to establish the reverse inclusion to verify the equivalence of these sets.

Let  $d \in T(x|\Omega)$ . Then, by definition, there are sequences  $t_i \downarrow 0$  and  $\{x^i\} \subset \Omega$  with  $x^i \rightarrow x$  such that  $d^i \rightarrow d$  where  $d^i = t_i^{-1}(x^i - x)$ ,  $i = 1, 2, \dots$ . Note that

$$Ad^i = t_i^{-1}A(x^i - x) = t_i^{-1}[Ax^i - Ax] = t_i^{-1}[b - b] = 0 \quad \forall i = 1, 2, \dots$$

Therefore,  $Ad = \lim_{i \rightarrow \infty} Ad^i = 0$  so that  $d \in \text{Nul}(A)$ . Since  $d$  was chosen arbitrarily from  $T(x|\Omega)$ , we have  $T(x|\Omega) \subset \text{Nul}(A)$  which proves the equivalence.

(2) If  $\Omega = \{(x_1, x_2)^T : x_1^2 + x_2^2 = 1\}$ , then  $T(x|\Omega) = \{(y_1, y_2) : x_1 y_1 + x_2 y_2 = 0\}$ .

(3) A convex set is said to be polyhedral if it can be represented as the solution set of a finite number of linear equality and/or inequality constraints. Thus, for example the constraint region for an LPs is a convex polyhedron. If it is assumed that  $\Omega$  is a convex polyhedron, then

$$T(x|\Omega) = \bigcup_{\lambda \geq 0} \lambda(\Omega - x) = \{\lambda(y - x) \mid \lambda \geq 0, y \in \Omega\}.$$

(4) If  $\Omega$  is a convex subset of  $\mathbb{R}^n$ , then

$$T(x|\Omega) = \overline{\bigcup_{\lambda \geq 0} \lambda(\Omega - x)} = \text{cl} \{\lambda(y - x) \mid \lambda \geq 0, y \in \Omega\}.$$

**THEOREM 1.2. [BASIC CONSTRAINED FIRST-ORDER NECESSARY CONDITIONS]**

Suppose that the function  $f_0 : \mathbb{R}^n \rightarrow \mathbb{R}$  in  $\mathcal{P}$  is continuously differentiable near the point  $\bar{x} \in \Omega$ . If  $\bar{x}$  is a local solution to  $\mathcal{P}$ , then

$$f'_0(\bar{x}; d) \geq 0 \quad \text{for all } d \in T(\bar{x}|\Omega).$$

PROOF. Note that the MVT (Mean Value Theorem) implies that

$$f'_0(\bar{x}; d) = \lim_{\tau \searrow 0} \frac{f_0(\bar{x} + \tau d) - f_0(\bar{x})}{\tau} = \lim_{\substack{s \rightarrow d \\ \tau \searrow 0}} \frac{f_0(\bar{x} + \tau s) - f_0(\bar{x})}{\tau}$$

since  $f_0$  is continuously differentiable.

Suppose  $\bar{x}$  is a local solution to  $\mathcal{P}$  and let  $d \in T(\bar{x}|\Omega)$ . Since  $d \in T(\bar{x}|\Omega)$ , there is a sequence  $\{x_i\} \subset \Omega$  and  $t_i \downarrow 0$  such that  $x_i \rightarrow \bar{x}$  and  $s_i = t_i^{-1}(x_i - \bar{x}) \rightarrow d$ . Note that  $\bar{x} + t_i d \approx \bar{x} + t_i s_i = x_i$ , and so  $f(\bar{x} + t_i s_i) = f(x_i) \geq f(\bar{x})$ . Using the representation of the directional derivative given above, we obtain

$$f'_0(\bar{x}; d) = \lim_{\substack{s \rightarrow d \\ \tau \searrow 0}} \frac{f_0(\bar{x} + \tau s) - f_0(\bar{x})}{\tau} = \lim_{i \rightarrow \infty} \frac{f_0(\bar{x} + t_i s_i) - f_0(\bar{x})}{t_i} = \lim_{i \rightarrow \infty} \frac{f_0(x_i) - f_0(\bar{x})}{t_i} \geq 0.$$

□

This general result is not particularly useful on its own since it refers to the abstract notion of tangent cone. In order to make it useful, we need to be able to compute the tangent cone once a representation for  $\Omega$  is given. We now show how this can be done.

We begin by assuming that  $\Omega$  has the form

$$(58) \quad \Omega := \{x : f_i(x) \leq 0, i = 1, \dots, s, f_i(x) = 0, i = s + 1, \dots, m\},$$

where each  $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$  is continuously differentiable on  $\mathbb{R}^n$ . Observe that if  $x \in \Omega$  and  $d \in T(x|\Omega)$  then there are sequences  $\{x_k\} \subset \Omega$  and  $\tau_k \searrow 0$  with  $x_k \rightarrow x$  such that  $\tau_k^{-1}(x_k - x) \rightarrow d$ . Setting  $d_k = \tau_k^{-1}(x_k - x)$  for all  $k$  we have that

$$f'_i(x; d) = \lim_{k \rightarrow \infty} \frac{f_i(x + \tau_k d_k) - f_i(x)}{\tau_k}$$

equals 0 for  $i \in \{s + 1, \dots, m\}$  and is less than or equal to 0 for  $i \in I(x)$  where

$$I(x) := \{i : i \in \{1, \dots, s\}, f_i(x) = 0\}.$$

Consequently,

$$T(x|\Omega) \subset \{d : \nabla f_i(x)^T d \leq 0, i \in I(x), \nabla f_i(x)^T d = 0, i = s + 1, \dots, m\}.$$

The set on the right hand side of this inclusion is a computationally tractable. Moreover, in a certain sense, the cases where these two sets do not coincide are exceptional. For this reason we make the following definition.

## DEFINITION 1.2. [REGULARITY]

We say that the set  $\Omega$  is regular at  $x \in \Omega$  if

$$T(x | \Omega) = \{d \in \mathbb{R}^n : f'_i(x; d) \leq 0, i \in I(x), f'_i(x; d) = 0 \text{ } i = s + 1, \dots, m\}.$$

But it is important to note that not every set is regular.

## EXERCISE 1.1. Graph the set

$$\Omega := \{x \in \mathbb{R}^2 \mid -x_1^3 \leq x_2 \leq x_1^3\},$$

and show that it is not regular at the origin. This is done by first showing that

$$T_\Omega(0) = \{(d_1, d_2)^T \mid d_1 \geq 0, d_2 = 0\}.$$

Then set

$$f_1(x_1, x_2) = -x_1^3 - x_2 \quad \text{and} \quad f_2(x_1, x_2) = -x_1^3 + x_2,$$

so that  $\Omega = \{(x_1, x_2)^T \mid f_1(x_1, x_2) \leq 0, f_2(x_1, x_2) \leq 0\}$ . Finally, show that

$$\{d \mid \nabla f_1(0, 0)^T d \leq 0, \nabla f_2(0, 0)^T d \leq 0\} = \{(d_1, d_2)^T \mid d_2 = 0\} \neq T_\Omega(0).$$

Next let us suppose we are at a given point  $x \in \Omega$  and that we wish to obtain a new point  $x_+ = x + td$  for which  $f(x_+) < f(x)$  for some direction  $d \in \mathbb{R}^n$  and steplength  $t > 0$ . A good candidate for a search direction  $d$  is one that minimizes  $f'(x; d)$  over all directions that point into  $\Omega$  up to first-order. That is, we should minimize  $\nabla f(x)^T d$  over the set of tangent directions. Remarkably, this search for a *feasible direction of steepest descent* can be posed as the following linear program (assuming regularity):

$$(59) \quad \begin{array}{ll} \max & (-\nabla f_0(\bar{x}))^T d \\ \text{subject to} & \nabla f_i(\bar{x})^T d \leq 0 \quad i \in I(\bar{x}) \\ & \nabla f_i(\bar{x})^T d = 0 \quad i = s + 1, \dots, m. \end{array}$$

The dual of (59) is the linear program

$$(60) \quad \begin{array}{ll} \min & 0 \\ \text{subject to} & \sum_{i \in I(\bar{x})} u_i \nabla f_i(\bar{x}) + \sum_{i=s+1}^m u_i \nabla f_i(\bar{x}) = -\nabla f_0(\bar{x}) \\ & 0 \leq u_i, \quad i \in I(\bar{x}). \end{array}$$

If we assume that  $\bar{x}$  is a local solution to  $\mathcal{P}$ , Theorem 1.2 tells us that the maximum in (59) is less than or equal to zero. But  $d = 0$  is feasible for (59), hence the maximum value in (59) is zero. Therefore, by the Strong Duality Theorem for Linear Programming, the linear program (60) is feasible, that is, there exist scalars  $u_i$ ,  $i \in I(\bar{x}) \cup \{s + 1, \dots, m\}$  with  $u_i \geq 0$  for  $i \in I(\bar{x})$  such that

$$(61) \quad 0 = \nabla f_0(\bar{x}) + \sum_{i \in I(\bar{x})} u_i \nabla f_i(\bar{x}) + \sum_{i=s+1}^m u_i \nabla f_i(\bar{x}).$$

This observation yields the following result.

## THEOREM 1.3. [CONSTRAINED FIRST-ORDER OPTIMALITY CONDITIONS]

Let  $\bar{x} \in \Omega$  be a local solution to  $\mathcal{P}$  at which  $\Omega$  is regular. Then there exist  $u \in \mathbb{R}^m$  such that

- (1)  $0 = \nabla_x L(\bar{x}, u)$ ,
- (2)  $0 = u_i f_i(\bar{x})$  for  $i = 1, \dots, s$ , and
- (3)  $0 \leq u_i$ ,  $i = 1, \dots, s$ ,

where the mapping  $L : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$  is defined by

$$L(x, u) := f_0(x) + \sum_{i=1}^m u_i f_i(x)$$

and is called the Lagrangian for the problem  $\mathcal{P}$ .

PROOF. For  $i \in I(\bar{x}) \cup \{s + 1, \dots, m\}$  let  $u_i$  be as given in (61) and for  $i \in \{1, \dots, s\} \setminus I(\bar{x})$  set  $u_i = 0$ . Then this choice of  $u \in \mathbb{R}^m$  satisfies (1)–(3) above.  $\square$

## DEFINITION 1.3. [KKT CONDITIONS]

Let  $x \in \mathbb{R}^n$  and  $u \in \mathbb{R}^m$ . We say that  $(x, u)$  is a Karush-Kuhn-Tucker (KKT) pair for  $\mathcal{P}$  if

- (1)  $f_i(x) \leq 0$   $i = 1, \dots, s$ ,  $f_i(x) = 0$   $i = s + 1, \dots, m$  (Primal feasibility),
- (2)  $u_i \geq 0$  for  $i = 1, \dots, s$  (Dual feasibility),
- (3)  $0 = u_i f_i(x)$  for  $i = 1, \dots, s$  (complementarity), and
- (4)  $0 = \nabla_x L(x, u)$  (stationarity of the Lagrangian).

Given  $x \in \mathbb{R}^n$ , if there is a  $u \in \mathbb{R}^m$  such that  $(x, u)$  is a Karush-Kuhn-Tucker pair for  $\mathcal{P}$ , then we say that  $x$  is a KKT point for  $\mathcal{P}$  (we also refer to such an  $x$  as a stationary point for  $\mathcal{P}$ ).  $\square$

## 2. Regularity and Constraint Qualifications

We now briefly discuss conditions that yield the regularity of  $\Omega$  at a point  $x \in \Omega$ . These conditions should be testable in the sense that there is a finitely terminating algorithm that can determine whether they are satisfied or not satisfied. The condition that we will concentrate on is the so called *Mangasarian-Fromovitz constraint qualification* (MFCQ).

DEFINITION 2.1. [MFCQ]

We say that a point  $x \in \Omega$  satisfies the Mangasarian-Fromovitz constraint qualification (or MFCQ) at  $x$  if

- (1) there is a  $d \in \mathbb{R}^n$  such that

$$\begin{aligned} \nabla f_i(x)^T d &< 0 \text{ for } i \in I(x), \\ \nabla f_i(x)^T d &= 0 \text{ for } i = s + 1, \dots, m, \end{aligned}$$

and

- (2) the gradients  $\{\nabla f_i(x) \mid i = s + 1, \dots, m\}$  are linearly independent.

We have the following key result which we shall not prove.

THEOREM 2.1. [MFCQ  $\rightarrow$  REGULARITY] Let  $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $i = 1, 2, \dots, m$  be  $C^1$  near  $\bar{x} \in \Omega$ . If the MFCQ holds at  $\bar{x}$ , then  $\Omega$  is regular at  $\bar{x}$ .

The MFCQ is algorithmically verifiable. This is seen by considering the LP

$$(62) \quad \begin{aligned} \min \quad & 0 \\ \text{subject to} \quad & \nabla f_i(\bar{x})^T d \leq -1 \quad i \in I(\bar{x}) \\ & \nabla f_i(\bar{x})^T d = 0 \quad i = s + 1, \dots, m. \end{aligned}$$

Clearly, the MFCQ is satisfied at  $\bar{x}$  if and only if the above LP is feasible and the gradients  $\{\nabla f_i(\bar{x}) \mid i = s + 1, \dots, m\}$  are linearly independent. This observation also leads to a *dual* characterization of the MFCQ by considering the dual of the LP (62).

LEMMA 2.1. [DUAL MFCQ]

The MFCQ is satisfied at a point  $\bar{x} \in \Omega$  if and only if the only solution to the system

$$\begin{aligned} \sum_{i=1}^m u_i \nabla f_i(\bar{x}) &= 0, \\ u_i f_i(\bar{x}) &= 0 \quad i = 1, 2, \dots, s, \text{ and} \\ u_i &\geq 0 \quad i = 1, 2, \dots, s, \end{aligned}$$

is  $u_i = 0$ ,  $i = 1, 2, \dots, m$ .

PROOF. The dual the LP (62) is the LP

$$(63) \quad \begin{aligned} \min \quad & \sum_{i \in I(\bar{x})} u_i \\ \text{subject to} \quad & \sum_{i \in I(\bar{x})} u_i \nabla f_i(\bar{x}) + \sum_{i=s+1}^m u_i \nabla f_i(\bar{x}) = 0 \\ & 0 \leq u_i, \quad i \in I(\bar{x}). \end{aligned}$$

This LP is always feasible, simply take all  $u_i$ 's equal to zero. Hence, by the Strong Duality Theorem of Linear Programming, the LP (62) is feasible if and only if the LP (63) is finite valued in which case the optimal value in both is zero. That is, the MFCQ holds at  $\bar{x}$  if and only if the optimal value in (63) is zero and the gradients

$\{\nabla f_i(\bar{x}) \mid i = s+1, \dots, m\}$  are linearly independent. The latter statement is equivalent to the statement that the only solution to the system

$$\begin{aligned} \sum_{i=1}^m u_i \nabla f_i(\bar{x}) &= 0, \\ u_i f_i(\bar{x}) &= 0 \quad i = 1, 2, \dots, s, \text{ and} \\ u_i &\geq 0 \quad i = 1, 2, \dots, s, \end{aligned}$$

is  $u_i = 0$ ,  $i = 1, 2, \dots, m$ . □

Techniques similar to these show that the MFCQ is a local property. That is, if it is satisfied at a point then it must be satisfied on a neighborhood of that point. The MFCQ is a powerful tool in the analysis of constraint systems as it implies many useful properties. One such property is established in the following result.

**THEOREM 2.2.** [MFCQ  $\rightarrow$  COMPACT MULTIPLIER SET]

Let  $\bar{x} \in \Omega$  be a local solution to  $\mathcal{P}$  at which the set of Karush-Kuhn-Tucker multipliers

$$(64) \quad KKT(\bar{x}) := \left\{ u \in \mathbb{R}^m \left| \begin{array}{l} \nabla_x L(\bar{x}, u) = 0 \\ u_i f_i(\bar{x}) = 0, \quad i = 1, 2, \dots, s, \\ 0 \leq u_i, \quad i = 1, 2, \dots, s \end{array} \right. \right\}$$

is non-empty. Then  $KKT(\bar{x})$  is a compact set if and only if the MFCQ is satisfied at  $\bar{x}$ .

**PROOF.** ( $\Rightarrow$ ) If MFCQ is not satisfied at  $\bar{x}$ , then from the Strong Duality Theorem for linear programming, Lemma 2.1, and the LP (63) guarantees the existence of a non-zero vector  $\bar{u} \in \mathbb{R}^m$  satisfying

$$\sum_{i=1}^m u_i \nabla f_i(\bar{x}) = 0 \text{ and } 0 \leq u_i \text{ with } 0 = u_i f_i(\bar{x}) \text{ for } i = 1, 2, \dots, s.$$

Then for each  $u \in KKT(\bar{x})$  we have that  $u + t\bar{u} \in KKT(\bar{x})$  for all  $t > 0$ . Consequently,  $KKT(\bar{x})$  cannot be compact.

( $\Leftarrow$ ) If  $KKT(\bar{x})$  is not compact, there is a sequence  $\{u^j\} \subset KKT(\bar{x})$  with  $\|u^j\| \uparrow +\infty$ . With no loss of generality, we may assume that

$$\frac{u^j}{\|u^j\|} \rightarrow u.$$

But then

$$\begin{aligned} u_i &\geq 0, \quad i = 1, 2, \dots, s, \\ u_i f_i(\bar{x}) &= \lim_{i \rightarrow \infty} \frac{u^j}{\|u^j\|} f_i(\bar{x}) = 0, \quad i = 1, 2, \dots, s, \text{ and} \\ \sum_{i=1}^m u_i f_i(\bar{x}) &= \lim_{i \rightarrow \infty} \frac{\nabla_x L(\bar{x}, u^j)}{\|u^j\|} = 0. \end{aligned}$$

Hence, by Lemma 2.1, the MFCQ cannot be satisfied at  $\bar{x}$ . □

Before closing this section we introduce one more constraint qualification. This is the so called *LI* condition and is associated with the uniqueness of the multipliers..

**DEFINITION 2.2 (LINEAR INDEPENDENCE CONDITION).** The *LI* condition is said to be satisfied at the point  $x \in \Omega$  if the constraint gradients

$$\{\nabla f_i(x) \mid i \in I(x) \cup \{s+1, \dots, m\}\}$$

are linearly independent.

Clearly, the LI condition implies the MFCQ. However, it is a much stronger condition in the presence of inequality constraints. In particular, the LI condition implies the uniqueness of the multipliers at a local solution to  $\mathcal{P}$ .

### 3. Second-Order Conditions

Second-order conditions are introduced by way of the Lagrangian. As is illustrated in the following result, the multipliers provide a natural way to incorporate the curvature of the constraints.

**THEOREM 3.1.** [CONSTRAINED SECOND-ORDER SUFFICIENCY]

Let  $\Omega$  have representation (58) and suppose that each of the functions  $f_i$ ,  $i = 0, 1, 2, \dots, m$  are  $C^2$ . Let  $\bar{x} \in \Omega$ . If  $(\bar{x}, \bar{u}) \in \mathbb{R}^n \times \mathbb{R}^m$  is a Karush-Kuhn-Tucker pair for  $\mathcal{P}$  such that

$$d^T \nabla_x^2 L(\bar{x}, \bar{u}) d > 0$$

for all  $d \in T_\Omega(\bar{x})$ ,  $d \neq 0$ , with  $\nabla f_0(\bar{x})^T d = 0$ , then there is an  $\epsilon > 0$  and  $\nu > 0$  such that

$$f_0(x) \geq f_0(\bar{x}) + \nu \|x - \bar{x}\|^2$$

for every  $x \in \Omega$  with  $\|x - \bar{x}\| \leq \epsilon$ , in particular  $\bar{x}$  is a strict local solution to  $\mathcal{P}$ .

**PROOF.** Suppose to the contrary that no such  $\epsilon > 0$  and  $\nu > 0$  exist, then there exist sequences  $\{x_k\} \subset \Omega$ ,  $\{\nu_k\} \subset \mathbb{R}_+$  such that  $x_k \rightarrow \bar{x}$ ,  $\nu_k \downarrow 0$ , and

$$f_0(x_k) \leq f_0(\bar{x}) + \nu_k \|x_k - \bar{x}\|^2$$

for all  $k = 1, 2, \dots$ . For every  $x \in \Omega$  we know that  $\bar{u}^T f(x) \leq 0$  and  $0 = \bar{u}^T f(\bar{x})$  where the  $i$ th component of  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is  $f_i$ . Hence

$$\begin{aligned} L(x_k, \bar{u}) \leq f_0(x_k) &\leq f_0(\bar{x}) + \nu_k \|x_k - \bar{x}\|^2 \\ &= L(\bar{x}, \bar{u}) + \nu_k \|x_k - \bar{x}\|^2. \end{aligned}$$

Therefore,

$$(65) \quad f_0(\bar{x}) + \nabla f_0(\bar{x})^T (x_k - \bar{x}) + o(\|x_k - \bar{x}\|) \leq f_0(\bar{x}) + \nu_k \|x_k - \bar{x}\|^2$$

and

$$(66) \quad \begin{aligned} L(\bar{x}, \bar{u}) &+ \nabla_x L(\bar{x}, \bar{u})^T (x_k - \bar{x}) \\ &+ \frac{1}{2} (x_k - \bar{x})^T \nabla_x^2 L(\bar{x}, \bar{u}) (x_k - \bar{x}) + o(\|x_k - \bar{x}\|^2) \\ &\leq L(\bar{x}, \bar{u}) + \nu_k \|x_k - \bar{x}\|^2. \end{aligned}$$

With no loss of generality, we can assume that

$$d_k := \frac{x_k - \bar{x}}{\|x_k - \bar{x}\|} \rightarrow \bar{d} \in T_\Omega(\bar{x}).$$

Dividing (65) through by  $\|x_k - \bar{x}\|$  and taking the limit we find that  $\nabla f_0(\bar{x})^T \bar{d} \leq 0$ . Since

$$T_\Omega(\bar{x}) \subset \{d : \nabla f_i(\bar{x})^T d \leq 0, i \in I(\bar{x}), \nabla f_i(\bar{x})^T d = 0, i = s+1, \dots, m\},$$

we have  $\nabla f_i(\bar{x})^T \bar{d} \leq 0$ ,  $i \in I(\bar{x}) \cup \{0\}$  and  $\nabla f_i(\bar{x})^T \bar{d} = 0$  for  $i = s+1, \dots, m$ . On the other hand,  $(\bar{x}, \bar{u})$  is a Karush-Kuhn-Tucker point so

$$\nabla f_0(\bar{x})^T \bar{d} = - \sum_{i \in I(\bar{x})} \bar{u}_i \nabla f_i(\bar{x})^T \bar{d} \geq 0.$$

Hence  $\nabla f_0(\bar{x})^T \bar{d} = 0$ , so that

$$\bar{d}^T \nabla_x^2 L(\bar{x}, \bar{u}) \bar{d} > 0.$$

But if we divide (66) by  $\|x_k - \bar{x}\|^2$  and take the limit, we arrive at the contradiction

$$\frac{1}{2} \bar{d}^T \nabla_x^2 L(\bar{x}, \bar{u}) \bar{d} \leq 0,$$

whereby the result is established.  $\square$

The assumptions required to establish Theorem 3.1 are somewhat strong but they do lead to a very practical and, in many cases, satisfactory second-order sufficiency result. In order to improve on this result one requires a much more sophisticated mathematical machinery. We do not take the time to develop this machinery. Instead we simply state a very general result. The statement of this result employs the entire set of Karush-Kuhn-Tucker multipliers  $KKT(\bar{x})$ .

**THEOREM 3.2 (GENERAL CONSTRAINED SECOND-ORDER NECESSITY AND SUFFICIENCY).** Let  $\bar{x} \in \Omega$  be a point at which  $\Omega$  is regular.

(1) If  $\bar{x}$  is a local solution to  $\mathcal{P}$ , then  $KKT(\bar{x}) \neq \emptyset$ , and for every  $d \in T(\bar{x} | \Omega)$  there is a  $u \in KKT(\bar{x})$  such that

$$d^T \nabla_x^2 L(\bar{x}, u) d \geq 0.$$

(2) If  $KKT(\bar{x}) \neq \emptyset$ , and for every  $d \in T(\bar{x} | \Omega)$ ,  $d \neq 0$ , for which  $\nabla f_0(\bar{x})^T d = 0$  there is a  $u \in KKT(\bar{x})$  such that

$$d^T \nabla_x^2 L(\bar{x}, u) d > 0,$$

then there is an  $\epsilon > 0$  and  $\nu > 0$  such that

$$f_0(x) \geq f_0(\bar{x}) + \nu \|x - \bar{x}\|^2$$

for every  $x \in \Omega$  with  $\|x - \bar{x}\| \leq \epsilon$ , in particular  $\bar{x}$  is a strict local solution to  $\mathcal{P}$ .

#### 4. Optimality Conditions in the Presence of Convexity

As we saw in the unconstrained case, convexity can have profound implications for optimality and optimality conditions. To begin with, we have the following very powerful result whose proof is identical to the proof in the unconstrained case.

**THEOREM 4.1.** [CONVEXITY+LOCAL OPTIMALITY→GLOBAL OPTIMALITY]

Suppose that  $f_0 : \mathbb{R}^n \rightarrow \mathbb{R}$  is convex and that  $\Omega \subset \mathbb{R}^n$  is a convex set. If  $\bar{x} \in \mathbb{R}^n$  is a local solution to  $\mathcal{P}$ , then  $\bar{x}$  is a global solution to  $\mathcal{P}$ .

**PROOF.** Suppose there is a  $\hat{x} \in \Omega$  with  $f_0(\hat{x}) < f_0(\bar{x})$ . Let  $\epsilon > 0$  be such that

$$f_0(\bar{x}) \leq f_0(x) \quad \text{whenever} \quad \|x - \bar{x}\| \leq \epsilon \text{ and } x \in \Omega,$$

and

$$\epsilon < 2\|\bar{x} - \hat{x}\|.$$

Set  $\lambda := \epsilon(2\|\bar{x} - \hat{x}\|)^{-1} < 1$  and  $x_\lambda := \bar{x} + \lambda(\hat{x} - \bar{x}) \in \Omega$ . Then  $\|x_\lambda - \bar{x}\| \leq \epsilon/2$  and  $f_0(x_\lambda) \leq (1 - \lambda)f_0(\bar{x}) + \lambda f_0(\hat{x}) < f_0(\bar{x})$ . This contradicts the choice of  $\epsilon$  and so no such  $\hat{x}$  exists.  $\square$

We also have the following first-order necessary conditions for optimality. The proof of this result again follows that for the unconstrained case.

**THEOREM 4.2.** [1<sup>ST</sup>-ORDER NECESSITY AND SUFFICIENCY]

Suppose that  $f_0 : \mathbb{R}^n \rightarrow \mathbb{R}$  is convex and that  $\Omega \subset \mathbb{R}^n$  is a convex set, and let  $\bar{x} \in \Omega$ . Then the following statements are equivalent.

- (i)  $\bar{x}$  is a local solution to  $\mathcal{P}$ .
- (ii)  $f'_0(\bar{x}; y - \bar{x}) \geq 0$  for all  $y \in \Omega$ .
- (iii)  $\bar{x}$  is a global solution to  $\mathcal{P}$ .

**PROOF.** The implication (i)⇒(ii) follows from Theorem 1.1 since each of the directions  $d = y - \bar{x}$ ,  $y \in \Omega$  is a feasible direction for  $\Omega$  at  $\bar{x}$  due to the convexity of  $\Omega$ . To see the implication (ii)⇒(iii), we again resort to the subdifferential inequality. Let  $y$  be any other point in  $\Omega$ . Then  $d = y - \bar{x} \in T_\Omega(\bar{x})$  and so by the subdifferential inequality we have

$$f_0(y) \geq f_0(\bar{x}) + f'_0(\bar{x}; y - \bar{x}) \geq f_0(\bar{x}).$$

Since  $y \in \Omega$  was arbitrary the implication (ii)⇒(iii) follows. The implication (iii)⇒(i) is trivial.  $\square$

The utility of this result again depends on our ability to represent the tangent cone  $T_\Omega(\bar{x})$  in a computationally tractable manner. Following the general case, we assume that the set  $\Omega$  has the representation (58):

$$(67) \quad \Omega := \{x : f_i(x) \leq 0, i = 1, \dots, s, f_i(x) = 0, i = s + 1, \dots, m\}.$$

The first issue we must address is to determine reasonable conditions on the functions  $f_i$  that guarantee that the set  $\Omega$  is convex. We begin with the following elementary facts about convex functions and convex sets whose proofs we leave to the reader.

**LEMMA 4.1.** If  $C_i \subset \mathbb{R}^n$ ,  $i = 1, 2, \dots, N$ , are convex sets, then so is the set  $C = \bigcap_{i=1}^N C_i$ .

**LEMMA 4.2.** If  $h : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$  is a convex function, then for every  $\alpha \in \mathbb{R}$  the set

$$\text{lev}_h(\alpha) = \{x \mid h(x) \leq \alpha\}$$

is a convex set.

These facts combine to give the following result.

LEMMA 4.3. *If the functions  $f_i$ ,  $i = 1, 2, \dots, s$  are convex and the functions  $f_i$ ,  $i = s + 1, \dots, m$  are linear, then the set  $\Omega$  given by (67) is a convex set.*

REMARK 4.1. *Recall that a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is said to be linear if there exists  $c \in \mathbb{R}^n$  and  $\alpha \in \mathbb{R}$  such that  $f(x) = c^T x + \alpha$ .*

PROOF. Note that

$$\Omega = \left( \bigcap_{i=1}^m \text{lev}_{f_i}(0) \right) \cap \left( \bigcap_{i=s+1}^m \text{lev}_{-f_i}(0) \right),$$

where each of the functions  $f_i, i = 1, \dots, m$  and  $-f_i, i = s + 1, \dots, m$  is convex. Therefore, the convexity of  $\Omega$  follows from Lemmas 4.2 and 4.1.  $\square$

In order to make the link to the KKT condition in the presence of convexity, we still require the regularity of the set  $\Omega$  at the point of interest  $\bar{x}$ . If the set  $\Omega$  is a polyhedral convex set, i.e.

$$\Omega = \{x \mid Ax \leq a, Bx = b\}$$

for some  $A \in \mathbb{R}^{s \times n}, a \in \mathbb{R}^s, B \in \mathbb{R}^{(m-s) \times n}$ , and  $b \in \mathbb{R}^{(m-s)}$ , then the set  $\Omega$  is everywhere regular (Why?). In the general convex case this may not be true. However, convexity can be used to derive a much simpler test for the regularity of non-polyhedral convex sets.

DEFINITION 4.1 (THE SLATER CONSTRAINT QUALIFICATION). *Let  $\Omega \subset \mathbb{R}^n$  be as given in (67) with  $f_i$ ,  $i = 1, \dots, s$  convex and  $f_i$ ,  $i = s + 1, \dots, m$  linear. We say that  $\Omega$  satisfies the Slater constraint qualification if there exists  $\tilde{x} \in \Omega$  such that  $f_i(\tilde{x}) < 0$  for  $i = 1, \dots, s$ .*

THEOREM 4.3 (CONVEXITY AND REGULARITY). *Suppose  $\Omega \subset \mathbb{R}^n$  is as given in (67) with  $f_i$ ,  $i = 1, \dots, s$  convex and  $f_i$ ,  $i = s + 1, \dots, m$  linear. If either  $\Omega$  is polyhedral convex or satisfies the Slater constraint qualification, then  $\Omega$  is regular at every point  $\bar{x} \in \Omega$  at which the function  $f_i$ ,  $i = 1, \dots, s$  are differentiable.*

We do not present the proof of this result as it takes us too far afield of our study. Nonetheless, we make use of this fact in the following result of the KKT conditions.

THEOREM 4.4 (CONVEXITY+REGULARITY  $\rightarrow$  (OPTIMALITY  $\Leftrightarrow$  KKT CONDITIONS)). *Let  $f_0 : \mathbb{R}^n \rightarrow \mathbb{R}$  be a differentiable convex function and let  $\Omega$  be as given in Lemma 4.3 where each of the function  $f_i$ ,  $i = 1, \dots, s$  is differentiable.*

- (i) *If  $\bar{x} \in \Omega$  is a KKT point for  $\mathcal{P}$ , then  $\bar{x}$  is a global solution to  $\mathcal{P}$ .*
- (ii) *Suppose the functions  $f_i$ ,  $i = 0, 1, \dots, s$  are continuously differentiable. If  $\bar{x}$  is a solution to  $\mathcal{P}$  at which  $\Omega$  is regular, then  $\bar{x}$  is a KKT point for  $\mathcal{P}$ .*

PROOF. Part (ii) of this theorem is just a restatement of Theorem 1.3 and so we need only prove Part (i).

Since  $\bar{x}$  is a KKT point there exists  $\bar{y} \in \mathbb{R}^m$  such that  $(\bar{x}, \bar{y})$  is a KKT pair for  $\mathcal{P}$ . Consider the function  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  given by

$$h(x) = L(x, \bar{y}) = f_0(x) + \sum_{i=1}^m \bar{y}_i f_i(x).$$

By construction, the function  $h$  is convex with  $0 = \nabla h(\bar{x}) = \nabla_x L(\bar{x}, \bar{y})$ . Therefore,  $\bar{x}$  is a global solution to the problem  $\min_{x \in \mathbb{R}^n} h(x)$ . Also note that for every  $x \in \Omega$  we have

$$\sum_{i=1}^m \bar{y}_i f_i(x) \leq 0,$$

since  $\bar{y}_i f_i(x) \leq 0$   $i = 1, \dots, s$  and  $\bar{y}_i f_i(x) = 0$   $i = s + 1, \dots, m$ . Consequently,

$$\begin{aligned} f_0(\bar{x}) &= h(\bar{x}) \leq h(x) = L(x, \bar{y}) \\ &= f_0(x) + \sum_{i=1}^m \bar{y}_i f_i(x) \\ &\leq f_0(x) \end{aligned}$$

for all  $x \in \Omega$ . This establishes Part (i).  $\square$



If all of the functions  $f_i$   $i = 0, 1, \dots, m$  are twice continuously differentiable, then the second-order sufficiency conditions stated in Theorem 3.1 apply. However, in the presence of convexity another kind of second-order condition is possible that does not directly incorporate curvature information about the functions  $f_i$   $i = 1, \dots, m$ . These second-order conditions are most appropriate when  $\Omega$  is polyhedral convex.

**THEOREM 4.5. [2<sup>ND</sup>-ORDER OPTIMALITY CONDITIONS FOR POLYHEDRAL CONSTRAINTS]**

Let  $f_0 : \mathbb{R}^n \rightarrow \mathbb{R}$  be  $C^2$  and  $\bar{x}$  be an element of the convex set  $\Omega$ .

- (1) (necessity) If  $\bar{x} \in \mathbb{R}^n$  is a local solution to  $\mathcal{P}$  with  $\Omega$  a polyhedral convex set, then  $\nabla f_0(\bar{x})^T d \geq 0$  for all  $d \in T_\Omega(\bar{x})$  and

$$d^T \nabla^2 f_0(\bar{x}) d \geq 0$$

for all  $d \in T_\Omega(\bar{x})$  with  $\nabla f_0(\bar{x})^T d = 0$ .

- (2) (sufficiency) If  $\bar{x} \in \mathbb{R}^n$  is such that  $\nabla f_0(\bar{x})^T (y - \bar{x}) \geq 0$  for all  $d \in T_\Omega(\bar{x})$  and

$$d^T \nabla^2 f_0(\bar{x}) d > 0$$

for all  $d \in T_\Omega(\bar{x}) \setminus \{0\}$  with  $\nabla f_0(\bar{x})^T d = 0$ , then there exist  $\epsilon, \nu > 0$  such that

$$f_0(x) \geq f_0(\bar{x}) + \nu \|x - \bar{x}\|^2$$

for all  $x \in \Omega$  with  $\|x - \bar{x}\| \leq \epsilon$ .

**PROOF.** (1) Since  $\Omega$  is polyhedral convex, we have  $T_\Omega(\bar{x}) = \bigcup_{\lambda \geq 0} (\Omega - \bar{x})$ . Therefore, the fact that  $\nabla f_0(\bar{x})^T d \geq 0$  for all  $d \in T_\Omega(\bar{x})$  follows from Theorem 4.2. Next let  $d \in T_\Omega(\bar{x}) = \bigcup_{\lambda \geq 0} (\Omega - \bar{x})$  be such that  $\nabla f_0(\bar{x})^T d = 0$ . Then there is a  $y \in \Omega$ ,  $y \neq \bar{x}$ , and a  $\lambda_0 > 0$  such that  $d = \lambda_0(y - \bar{x})$ . Let  $\epsilon > 0$  be such that  $f_0(\bar{x}) \leq f_0(x)$  for all  $x \in \Omega$  with  $\|x - \bar{x}\| \leq \epsilon$ . Set  $\bar{\lambda} = \min\{\lambda_0, \epsilon(\lambda_0\|y - \bar{x}\|)^{-1}\} > 0$  so that  $\bar{x} + \lambda d \in \Omega$  and  $\|\bar{x} - (\bar{x} + \lambda d)\| \leq \epsilon$  for all  $\lambda \in [0, \bar{\lambda}]$ . By hypothesis, we now have

$$\begin{aligned} f_0(\bar{x}) &\leq f_0(\bar{x} + \lambda d) \\ &= f_0(\bar{x}) + \lambda \nabla f_0(\bar{x})^T (y - \bar{x}) + \frac{\lambda^2}{2} d^T \nabla^2 f_0(\bar{x}) d + o(\lambda^2) \\ &= f_0(\bar{x}) + \frac{\lambda^2}{2} d^T \nabla^2 f_0(\bar{x}) d + o(\lambda^2), \end{aligned}$$

where the second equality follows from the choice of  $d$  ( $\nabla f_0(\bar{x})^T d = 0$ ). Therefore  $d^T \nabla^2 f_0(\bar{x}) d \geq 0$ .

(2) We show that  $f_0(\bar{x}) \leq f_0(x) - \nu \|x - \bar{x}\|^2$  for some  $\nu > 0$  for all  $x \in \Omega$  near  $\bar{x}$ . Indeed, if this were not the case there would exist sequences  $\{x_k\} \subset \Omega$ ,  $\{\nu_k\} \subset \mathbb{R}_+$  with  $x_k \rightarrow \bar{x}$ ,  $\nu_k \downarrow 0$ , and

$$f_0(x_k) < f_0(\bar{x}) + \nu_k \|x_k - \bar{x}\|^2$$

for all  $k = 1, 2, \dots$  where, with no loss of generality,  $\frac{x_k - \bar{x}}{\|x_k - \bar{x}\|} \rightarrow d$ . Clearly,  $d \in T_\Omega(\bar{x})$ . Moreover,

$$\begin{aligned} f_0(\bar{x}) + \nabla f_0(\bar{x})^T (x_k - \bar{x}) &+ o(\|x_k - \bar{x}\|) \\ &= f_0(x_k) \\ &\leq f_0(\bar{x}) + \nu_k \|x_k - \bar{x}\|^2 \end{aligned}$$

so that  $\nabla f_0(\bar{x})^T d = 0$ .

Now, since  $\nabla f_0(\bar{x})^T (x_k - \bar{x}) \geq 0$  for all  $k = 1, 2, \dots$ ,

$$\begin{aligned} f_0(\bar{x}) + \frac{1}{2} (x_k - \bar{x})^T \nabla^2 f_0(\bar{x}) (x_k - \bar{x}) &+ o(\|x_k - \bar{x}\|^2) \\ &\leq f_0(\bar{x}) + \nabla f_0(\bar{x})^T (x_k - \bar{x}) + \frac{1}{2} (x_k - \bar{x})^T \nabla^2 f_0(\bar{x}) (x_k - \bar{x}) \\ &\quad + o(\|x_k - \bar{x}\|^2) \\ &= f_0(x_k) \\ &< f_0(\bar{x}) + \nu_k \|x_k - \bar{x}\|^2. \end{aligned}$$

Hence,

$$\left( \frac{x_k - \bar{x}}{\|x_k - \bar{x}\|} \right)^T \nabla^2 f_0(\bar{x}) \left( \frac{x_k - \bar{x}}{\|x_k - \bar{x}\|} \right) \leq \nu_k + \frac{o(\|x_k - \bar{x}\|^2)}{\|x_k - \bar{x}\|^2}$$

Taking the limit in  $k$  we obtain the contradiction

$$0 < d^T \nabla^2 f_0(\bar{x}) d \leq 0,$$

whereby the result is established.  $\square$

Although it is possible to weaken the assumption of polyhedrality in Part 1, such weakenings are somewhat artificial as they essentially imply that  $T_\Omega(x) = \bigcup_{\lambda \geq 0} (\Omega - x)$ . The following example illustrates what can go wrong when the assumption of polyhedrality is dropped.

EXAMPLE 4.1. *Consider the problem*

$$\begin{array}{ll} \min & \frac{1}{2}(x_2 - x_1^2) \\ \text{subject to} & 0 \leq x_2, \quad x_1^3 \leq x_2^2. \end{array}$$

Observe that the constraint region in this problem can be written as  $\Omega := \{(x_1, x_2)^T : |x_1|^{\frac{3}{2}} \leq x_2\}$ , therefore

$$\begin{aligned} f_0(x) &= \frac{1}{2}(x_2 - x_1^2) \\ &\geq \frac{1}{2}(|x_1|^{\frac{3}{2}} - |x_1|^2) \\ &= \frac{1}{2}|x_1|^{\frac{3}{2}}(1 - |x_1|^{\frac{1}{2}}) > 0 \end{aligned}$$

whenever  $0 < |x_1| \leq 1$ . Consequently, the origin is a strict local solution for this problem. Nonetheless,

$$T_\Omega(0) \cap [\nabla f_0(0)]^\perp = \{(\delta, 0)^T : \delta \in \mathbb{R}\},$$

while

$$\nabla^2 f_0(0) = \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix}.$$

That is, even though the origin is a strict local solution, the Hessian of  $f_0$  is not positive semidefinite on  $T_\Omega(0)$ .

When using the second-order conditions given above, one needs to be careful about the relationship between the Hessian of  $f_0$  and the set  $K := T_\Omega(x) \cap [\nabla f_0(x)]^\perp$ . In particular, the positive definiteness (or semidefiniteness) of the Hessian of  $f_0$  on the cone  $K$  does not necessarily imply the positive definiteness (or semidefiniteness) of the Hessian of  $f_0$  on the subspace spanned by  $K$ . This is illustrated by the following example.

EXAMPLE 4.2. *Consider the problem*

$$\begin{array}{ll} \min & (x_1^2 - \frac{1}{2}x_2^2) \\ \text{subject to} & -x_1 \leq x_2 \leq x_1. \end{array}$$

Clearly, the origin is the unique global solution for this problem. Moreover, the constraint region for this problem,  $\Omega$ , satisfies

$$T_\Omega(0) \cap [\nabla f(0)]^\perp = T_\Omega(0) = \Omega,$$

with the span of  $\Omega$  being all of  $\mathbb{R}^2$ . Now, while the Hessian of  $f_0$  is positive definite on  $\Omega$ , it is not positive definite on all of  $\mathbb{R}^2$ .

In the polyhedral case it is easy to see that the sufficiency result in Theorem 4.5 is equivalent to the sufficiency result of Theorem 3.1. However, in the nonpolyhedral case, these results are not comparable. It is easy to see that Theorem 4.5 can handle situations where Theorem 3.1 does not apply even if  $\Omega$  is given in the form (58). Just let one of the active constraint functions be nondifferentiable at the solution. Similarly, Theorem 3.1 can provide information when Theorem 4.5 does not. This is illustrated by the following example.

EXAMPLE 4.3. *Consider the problem*

$$\begin{array}{ll} \min & x_2 \\ \text{subject to} & x_1^2 \leq x_2. \end{array}$$

Clearly,  $\bar{x} = 0$  is the unique global solution to this convex program. Moreover,

$$\begin{aligned} f_0(\bar{x}) + \frac{1}{2} \|x - \bar{x}\|^2 &= \frac{1}{2}(x_1^2 + x_2^2) \\ &\leq \frac{1}{2}(x_2 + x_2^2) \\ &\leq x_2 = f_0(x) \end{aligned}$$

for all  $x$  in the constraint region  $\Omega$  with  $\|x - \bar{x}\| \leq 1$ . It is easily verified that this growth property is predicted by Theorem 4.5.

### 5. Convex Optimization, Saddle Point Theory, and Lagrangian Duality

In this section we extend the duality theory for linear programming to general problems of convex optimization. This is accomplished using the saddle point properties of the Lagrangian in convex optimization. Again, consider the problem

$$\begin{aligned} \mathcal{P} \quad & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i \leq 0, \quad i = 1, 2, \dots, s \\ & && f_i(x) = 0, \quad i = s + 1, \dots, m, \end{aligned}$$

where it is assumed that the functions  $f_0, f_1, \dots, f_s$  are convex functions mapping  $\mathbb{R}^n$  to  $\overline{\mathbb{R}}$ , and  $f_{s+1}, \dots, f_m$  are affine mappings from  $\mathbb{R}^n$  to  $\mathbb{R}$ . We denote the constraint region for  $\mathcal{P}$  by  $\Omega$ .

The Lagrangian for  $\mathcal{P}$  is the function

$$L(x, y) = f_0(x) + y_1 f_1(x) + y_2 f_2(x) + \dots + y_m f_m(x),$$

where it is always assumed that  $0 \leq y_i, i = 1, 2, \dots, s$ . Set  $K = \mathbb{R}_+^s \times \mathbb{R}^{m-s} \subset \mathbb{R}^m$ . A pair  $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times K$  is said to be a saddle point for  $L$  if

$$L(\bar{x}, y) \leq L(\bar{x}, \bar{y}) \leq L(x, \bar{y}) \quad \forall (x, y) \in \mathbb{R}^n \times K.$$

We have the following basic saddle point theorem for  $L$ .

**THEOREM 5.1 (SADDLE POINT THEOREM).** *Let  $\bar{x} \in \mathbb{R}^n$ . If there exists  $\bar{y} \in K$  such that  $(\bar{x}, \bar{y})$  is a saddle point for the Lagrangian  $L$ , then  $\bar{x}$  solves  $\mathcal{P}$ . Conversely, if  $\bar{x}$  is a solution to  $\mathcal{P}$  at which the Slater C.Q. is satisfied, then there is a  $\bar{y} \in K$  such that  $(\bar{x}, \bar{y})$  is a saddle point for  $L$ .*

**PROOF.** If  $(\bar{x}, \bar{y}) \in \mathbb{R}^n \times K$  is a saddle point for  $\mathcal{P}$  then

$$\sup_{y \in K} L(\bar{x}, y) = \sup_{y \in K} f_0(\bar{x}) + y_1 f_1(\bar{x}) + y_2 f_2(\bar{x}) + \dots + y_m f_m(\bar{x}) \leq L(\bar{x}, \bar{y}).$$

If for some  $i \in \{1, \dots, s\}$  such that  $f_i(\bar{x}) > 0$ , then we could send  $y_i \uparrow +\infty$  to find that the supremum on the left is  $+\infty$  which is a contradiction, so we must have  $f_i(\bar{x}) \leq 0, i = 1, \dots, s$ . Moreover, if  $f_i(\bar{x}) \neq 0$  for some  $i \in \{s + 1, \dots, m\}$ , then we could send  $y_i \uparrow -\text{sign}(f_i(\bar{x}))\infty$  to again find that the supremum on the left is  $+\infty$  again a contradiction, so we must have  $f_i(\bar{x}) = 0, i = s + 1, \dots, m$ . That is, we must have  $\bar{x} \in \Omega$ . Since  $L(\bar{x}, \bar{y}) = \sup_{y \in K} L(\bar{x}, y)$ , we must have  $\sum_{i=1}^m \bar{y}_i f_i(\bar{x}) = 0$ . Therefore the right half of the saddle point condition implies that

$$f_0(\bar{x}) = L(\bar{x}, \bar{y}) \leq \inf_x L(x, \bar{y}) \leq \inf_{x \in \Omega} L(x, \bar{y}) \leq \inf_{x \in \Omega} f_0(x) \leq f_0(\bar{x}),$$

and so  $\bar{x}$  solves  $\mathcal{P}$ .

Conversely, if  $\bar{x}$  is a solution to  $\mathcal{P}$  at which the Slater C.Q. is satisfied, then there is a vector  $\bar{y}$  such that  $(\bar{x}, \bar{y})$  is a KKT pair for  $\mathcal{P}$ . Primal feasibility ( $\bar{x} \in \Omega$ ), dual feasibility ( $\bar{y} \in K$ ), and complementarity ( $\bar{y}_i f_i(\bar{x}), i = 1, \dots, s$ ) imply that

$$L(\bar{x}, y) \leq f_0(\bar{x}) = L(\bar{x}, \bar{y}) \quad \forall y \in K.$$

On the other hand, dual feasibility and convexity imply the convexity of the function  $L(x, \bar{y})$  in  $x$ . Hence the condition  $0 = \nabla_x L(\bar{x}, \bar{y})$  implies that  $\bar{x}$  is a global minimizer for the function  $x \rightarrow L(x, \bar{y})$ , that is

$$L(\bar{x}, \bar{y}) \leq L(x, \bar{y}) \quad \forall x \in \mathbb{R}^n.$$

Therefore,  $(\bar{x}, \bar{y})$  is a saddle point for  $L$ . □

Note that it is always the case that

$$\sup_{y \in K} \inf_{x \in \mathbb{R}^n} L(x, y) \leq \inf_{x \in \mathbb{R}^n} \sup_{y \in K} L(x, y)$$

since the largest minimum is always smaller than the smallest maximum. On the other hand, if  $(\bar{x}, \bar{y})$  is a saddle point for  $L$ , then

$$\inf_{x \in \mathbb{R}^n} \sup_{y \in K} L(x, y) \leq \sup_{y \in K} L(\bar{x}, y) \leq L(\bar{x}, \bar{y}) \leq \inf_{x \in \mathbb{R}^n} L(x, \bar{y}) \leq \sup_{y \in K} \inf_{x \in \mathbb{R}^n} L(x, y).$$

Hence, if a saddle point for  $L$  exists on  $\mathbb{R}^n \times K$ , then

$$\sup_{y \in K} \inf_{x \in \mathbb{R}^n} L(x, y) = \inf_{x \in \mathbb{R}^n} \sup_{y \in K} L(x, y).$$

Such a result is called a mini-max theorem and provides conditions under which one can exchange and inf-sup for a sup-inf. This mini-max result can be used as a basis for convex duality theory.

Observe that we have already shown that

$$\sup_{y \in K} L(x, y) = \begin{cases} +\infty & \text{if } x \notin \Omega, \\ f_0(x) & \text{if } x \in \Omega. \end{cases}$$

Therefore,

$$\inf_{x \in \mathbb{R}^n} \sup_{y \in K} L(x, y) = \inf_{x \in \Omega} f_0(x).$$

We will call this the *primal* problem. This is the inf-sup side of the saddle point problem. The other side, the sup-inf problem, we will call the *dual* problem with dual objective function

$$g(y) = \inf_{x \in \mathbb{R}^n} L(x, y).$$

The Saddle Point Theorem says that if  $(\bar{x}, \bar{y})$  is a saddle point for  $L$ , then  $\bar{x}$  solves the primal problem,  $\bar{y}$  solves the dual problem, and the optimal values in the primal and dual problems coincide. This is a *Weak Duality Theorem*. The Strong Duality Theorem follows from the second half of the Saddle Point Theorem and requires the use of the Slater Constraint Qualification.

**5.1. Linear Programming Duality.** We now show how the Lagrangian Duality Theory described above gives linear programming duality as a special case. Consider the following LP:

$$\begin{aligned} \mathcal{P} \quad & \text{minimize} && b^T x \\ & \text{subject to} && A^T x \geq c, \quad 0 \leq x. \end{aligned}$$

The Lagrangian is

$$L(x, y, v) = b^T x + y^T (c - A^T x) - v^T x, \quad \text{where } 0 \leq y, \quad 0 \leq v.$$

The dual objective function is

$$g(y, v) = \min_{x \in \mathbb{R}^n} L(x, y, v) = \min_{x \in \mathbb{R}^n} b^T x + y^T (c - A^T x) - v^T x.$$

Our first goal is to obtain a closed form expression for  $g(y, v)$ . This is accomplished by using the optimality conditions for minimizing  $L(x, y, v)$  to eliminate  $x$  from the definition of  $L$ . Since  $L(x, y, v)$  is a convex function in  $x$ , the global solution to  $\min_{x \in \mathbb{R}^n} L(x, y, v)$  is obtained by solving the equation  $0 = \nabla_x L(x, y, v) = b - Ay - v$  with  $0 \leq y, \quad 0 \leq v$ . Using this condition in the definition of  $L$  we get

$$L(x, y, v) = b^T x + y^T (c - A^T x) - v^T x = (b - Ay - v)^T x + c^T y = c^T y,$$

subject to  $b - A^T y = v$  and  $0 \leq y, \quad 0 \leq v$ . Hence the Lagrangian dual problem

$$\begin{aligned} & \text{maximize} && g(y, v) \\ & \text{subject to} && 0 \leq y, \quad 0 \leq v \end{aligned}$$

can be written as

$$\begin{aligned} \mathcal{D} \quad & \text{maximize} && c^T y \\ & \text{subject to} && b - Ay = v, \quad 0 \leq y, \quad 0 \leq v. \end{aligned}$$

Note that we can treat the variable  $v$  as a slack variable in this LP and write

$$\begin{aligned} \mathcal{D} \quad & \text{maximize} && c^T y \\ & \text{subject to} && Ay \leq b, \quad 0 \leq y. \end{aligned}$$

The linear program  $\mathcal{D}$  is the dual to the linear program  $\mathcal{P}$ .

**5.2. Convex Quadratic Programming Duality.** One can also apply the Lagrangian Duality Theory in the context of Convex Quadratic Programming. To see how this is done let  $Q \in \mathbb{R}^{n \times n}$  be symmetric and positive definite, and let  $c \in \mathbb{R}^n$ . Consider the convex quadratic program

$$\begin{aligned} \mathcal{D} \quad & \text{minimize} && \frac{1}{2}x^T Qx + c^T x \\ & \text{subject to} && Ax \leq b, \quad 0 \leq x. \end{aligned}$$

The Lagrangian is given by

$$L(x, y, v) = \frac{1}{2}x^T Qx + c^T x + y^T (A^T x - b) - v^T x \quad \text{where } 0 \leq y, \quad 0 \leq v.$$

The dual objective function is

$$g(y, v) = \min_{x \in \mathbb{R}^n} L(x, y, v).$$

The goal is to obtain a closed form expression for  $g$  with the variable  $x$  removed by using the first-order optimality condition  $0 = \nabla_x L(x, y, v)$ . This optimality condition completely identifies the solution since  $L$  is convex in  $x$ . We have

$$0 = \nabla_x L(x, y, v) = Qx + c + A^T y - v.$$

Since  $Q$  is invertible, we have

$$x = Q^{-1}(v - A^T y - c).$$

Plugging this expression for  $x$  into  $L(x, y, v)$  gives

$$\begin{aligned} g(y, v) &= L(Q^{-1}(v - A^T y - c), y, v) \\ &= \frac{1}{2}(v - A^T y - c)^T Q^{-1}(v - A^T y - c) \\ &\quad + c^T Q^{-1}(v - A^T y - c) + y^T (AQ^{-1}(v - A^T y - c) - b) - v^T Q^{-1}(v - A^T y - c) \\ &= \frac{1}{2}(v - A^T y - c)^T Q^{-1}(v - A^T y - c) - (v - A^T y - c)^T Q^{-1}(v - A^T y - c) - b^T y \\ &= -\frac{1}{2}(v - A^T y - c)^T Q^{-1}(v - A^T y - c) - b^T y. \end{aligned}$$

Hence the dual problem is

$$\begin{aligned} & \text{maximize} && -\frac{1}{2}(v - A^T y - c)^T Q^{-1}(v - A^T y - c) - b^T y \\ & \text{subject to} && 0 \leq y, \quad 0 \leq v. \end{aligned}$$

Moreover,  $(\bar{y}, \bar{v})$  solve the dual problem if and only if  $\bar{x} = Q^{-1}(\bar{v} - A^T \bar{y} - c)$  solves the primal problem with the primal and dual optimal values coinciding.

### Exercises

- (1) Locate all of the KKT points for the following problems. Can you show that these points are local solutions? Global solutions?

(a)

$$\begin{aligned} & \text{minimize} && e^{(x_1-x_2)} \\ & \text{subject to} && e^{x_1} + e^{x_2} \leq 20 \\ & && 0 \leq x_1 \end{aligned}$$

(b)

$$\begin{aligned} & \text{minimize} && e^{(-x_1+x_2)} \\ & \text{subject to} && e^{x_1} + e^{x_2} \leq 20 \\ & && 0 \leq x_1 \end{aligned}$$

(c)

$$\begin{aligned} & \text{minimize} && x_1^2 + x_2^2 - 4x_1 - 4x_2 \\ & \text{subject to} && x_1^2 \leq x_2 \\ & && x_1 + x_2 \leq 2 \end{aligned}$$

(d)

$$\begin{aligned} & \text{minimize} && \frac{1}{2} \|x\|^2 \\ & \text{subject to} && Ax = b \end{aligned}$$

where  $b \in \mathbb{R}^m$  and  $A \in \mathbb{R}^{m \times n}$  satisfies  $\text{Nul}(A^T) = \{0\}$ .

- (2) Show that the set

$$\Omega := \{x \in \mathbb{R}^2 \mid -x_1^3 \leq x_2 \leq x_1^3\}$$

is not regular at the origin. Graph the set  $\Omega$ .

- (3) Construct an example of a constraint region of the form (58) at which the MFCQ is satisfied, but the LI condition is not satisfied.

- (4) Suppose  $\Omega = \{x; Ax \leq b, Ex = h\}$  where  $A \in \mathbb{R}^{m \times n}$ ,  $E \in \mathbb{R}^{k \times n}$ ,  $b \in \mathbb{R}^m$ , and  $h \in \mathbb{R}^k$ .

(a) Given  $x \in \Omega$ , show that

$$T(x|\Omega) = \{d : A_i d \leq 0 \text{ for } i \in I(x), Ed = 0\},$$

where  $A_i$  denotes the  $i$ th row of the matrix  $A$  and  $I(x) = \{i \mid A_i x = b_i\}$ .

(b) Given  $x \in \Omega$ , show that every  $d \in T(x|\Omega)$  is a feasible direction for  $\Omega$  at  $x$ .

(c) Note that parts (a) and (b) above show that

$$T(x|\Omega) = \bigcup_{\lambda > 0} \lambda(\Omega - x)$$

whenever  $\Omega$  is a convex polyhedral set. Why?

- (5) Let  $C \subset \mathbb{R}^n$  be non-empty, closed and convex. For any  $x \in \mathbb{R}^n$  consider the problem of finding the closest point in  $C$  to  $x$  using the 2-norm:

$$\mathcal{D} \quad \begin{aligned} & \text{minimize} && \frac{1}{2} \|x - z\|_2^2 \\ & \text{subject to} && z \in C \end{aligned}$$

Show that  $\bar{z} \in C$  solves this problem if and only if

$$\langle x - \bar{z}, z - \bar{z} \rangle \leq 0 \quad \text{for all } z \in C.$$

- (6) Let  $\Omega$  be a non-empty closed convex subset of  $\mathbb{R}^n$ . The geometric object *dual* to the tangent cone is called the *normal cone*:

$$N(x|\Omega) = \{z; \langle z, d \rangle \leq 0, \text{ for all } d \in T(x|\Omega)\}.$$

(a) Show that if  $\bar{x}$  solves the problem  $\min\{f(x) : x \in \Omega\}$  then

$$-\nabla f(\bar{x}) \in N(\bar{x}|\Omega).$$

(b) Show that

$$N(\bar{x}|\Omega) = \{z : \langle z, x - \bar{x} \rangle \leq 0, \text{ for all } x \in \Omega\}.$$

(c) Let  $\bar{x} \in \Omega$ . Show that  $\bar{x}$  solves the problem  $\min\{\frac{1}{2} \|x - y\|_2^2 : x \in \Omega\}$  for every  $y \in \bar{x} + N(\bar{x}|\Omega)$ .

(7) Consider the functions

$$f(x) = \frac{1}{2}x^T Qx - c^T x$$

and

$$f_t(x) = \frac{1}{2}x^T Qx - c^T x + t\phi(x),$$

where  $t > 0$ ,  $Q \in \mathbb{R}^{n \times n}$  is positive semi-definite,  $c \in \mathbb{R}^n$ , and  $\phi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  is given by

$$\phi(x) = \begin{cases} -\sum_{i=1}^n \ln x_i & , \text{ if } x_i > 0, i = 1, 2, \dots, n, \\ +\infty & , \text{ otherwise.} \end{cases}$$

- (a) Show that  $\phi$  is a convex function.
- (b) Show that both  $f$  and  $f_t$  are convex functions.
- (c) Show that the solution to the problem  $\min f_t(x)$  always exists and is unique.
- (d) Let  $\{t_i\}$  be a decreasing sequence of positive real scalars with  $t_i \downarrow 0$ , and let  $x^i$  be the solution to the problem  $\min f_{t_i}(x)$ . Show that if the sequence  $\{x^i\}$  has a cluster point  $\bar{x}$ , then  $\bar{x}$  must be a solution to the problem  $\min\{f(x) : 0 \leq x\}$ .

*Hint:* Use the KKT conditions for the QP  $\min\{f(x) : 0 \leq x\}$ .