Theorem 15 *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$. Consider*

$$g(x) = h(f(x)) \qquad \text{for all } x \in \mathbb{R}^n.$$

The function g is convex if either of the following two conditions is satisfied:

(1) f is convex, h is nondecreasing and convex.

(2) f is concave, h is nonincreasing and convex.

By applying Theorem 15, we can see that:

$$e^{f(x)} \qquad \text{is convex if } f \text{ is convex,}$$

$$\frac{1}{f(x)} \qquad \text{is convex if } f \text{ is concave and } f(x) > 0 \text{ for all } x.$$

2.3 Convex Constrained Optimization Problems

In this section, we consider a generic convex constrained optimization problem. We introduce the basic terminology, and study the existence of solutions and the optimality conditions. We conclude this section with the projection problem and projection theorem. which is important for the subsequent algorithmic development

2.3.1 Constrained Problem

Consider the following constrained optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_1(x) \le 0, \ldots, g_m(x) \le 0 \\ & Bx = d \\ & x \in X, \end{array} \tag{2.5}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is an objective function, $X \subseteq \mathbb{R}^n$ is a given set, $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$ are constraint functions, B is a $p \times n$ matrix, and $d \in \mathbb{R}^p$. Let $g(x) \le 0$ compactly denote the inequalities $g_1(x) \le 0, \ldots, g_m(x) \le 0$. Define

$$C = \{x \in \mathbb{R}^n \mid g(x) \le 0,\ Bx = d,\ x \in X\}. \tag{2.6}$$

We refer to the set C as *constraint set or feasible set.* The problem is *feasible* when C is nonempty. We refer to the value $\inf_{x \in C} f(x)$ as the *optimal value* and denote it by f^*, i.e.,

$$f^* = \inf_{x \in C} f(x),$$

where C is given by Eq. (2.6). A vector x^* is *optimal (solution)* when x^* is *feasible* and attains the optimal value f^*, i.e.,

$$g(x^*) \le 0, \qquad Bx^* = d, \qquad x^* \in X, \qquad f(x^*) = f^*.$$

Before attempting to solve problem (2.5), there are some important questions to be answered, such as:

- Is the problem infeasible, i.e., is C empty?
- Is $f^* = +\infty$?[1].
- Is $f^* = -\infty$?

If the answer is "yes" to any of the preceding questions, then it does not make sense to consider the problem in (2.5) any further. The problem is of interest only when f^* is finite. In this case, the particular instances when the problem has a solution are of interest in many applications.

A feasibility problem is the problem of determining whether the constraint set C in Eq. (2.6) is empty or not. It can be posed as an optimization problem with the objective function $f(x) = 0$ for all $x \in \mathbb{R}^n$. In particular, *a feasibility problem* can be reduced to the following minimization problem:

$$\begin{aligned} &\text{minimize} && 0 && (2.7)\\ &\text{subject to} && g(x) \leq 0,\ Bx = d,\ x \in X. && (2.8)\end{aligned}$$

In many applications, the feasibility problem can be a hard problem on its own. For example, the stability in linear time invariant systems often reduces to the feasibility problem where we want to determine whether there exist positive definite matrices P and Q such that

$$A^T P + PA = -Q.$$

Equivalently, the question is whether the set

$$\{(P,Q) \in S^n_{++} \times S^n_{++} \mid A^T P + PA = -Q\}$$

is nonempty, where S^n_{++} denotes the set of $n \times n$ positive definite matrices.

From now on, we assume that the problem in Eq. (2.5) is feasible, and we focus on the issues of finiteness of the optimal value f^* and the existence of optimal solutions x^*. In particular, for problem (2.5), we use the following assumption.

Assumption 1 *The functions f and $g_i, i = 1,\ldots,m$ are convex over $\mathbb{R}^n$. The set X is closed and convex. The set $C = \{x \in \mathbb{R}^n \mid g(x) \leq 0,\ Bx = d,\ x \in\ X\}$ is nonempty.*

Under Assumption 1, the functions f and $g_i, i = 1,\ldots,m$ are continuous over $\mathbb{R}^n$ (see Theorem 10).

In what follows,*we denote the set of optimal solutions of problem (2.5) by X^*.*

2.3.2 Existence of Solutions

Here, we provide some results on the existence of solutions. Under Assumption 1, these results are consequences of Theorems 3 and 4 of Section 1.2.4.

[1]This happens in general only when $\text{dom} f \cap C = \emptyset$.

Theorem 16 *Let Assumption 1 hold. In addition, let $X \subseteq \mathbb{R}^n$ be bounded. Then, the optimal set X^* of problem (2.5) is nonempty, compact, and convex.*

Proof. At first, we show that the constraint set C is compact. The set C is the intersection of the level sets of continuous functions g_i and hyperplanes (for $j = 1, \ldots, p$, each set $\{x \in \mathbb{R}^n \mid b_j^T x = d_j\}$ is a hyperplane), which are all closed sets. Therefore, C is closed. Since $C \subseteq X$ and X is bounded (because it is compact), the set C is also bounded. Hence, by Lemma 8 of Section 1.2.2, the set C is compact. The function f is continuous (by convexity over $\mathbb{R}^n$). Hence, by Weierstrass Theorem (Theorem 3), the optimal value f^* of problem (2.5) is finite and its optimal set X^* is nonempty.

The set X^* is closed since it can be represented as the intersection of closed sets:

$$X^* = C \cap \{x \in \mathbb{R}^n \mid f(x) \leq f^*\}.$$

Furthermore, X^* is bounded since $X^* \subseteq C$ and C is bounded. Hence, X^* is compact.

We now show that X^* is convex. Note that C is convex as it is given as the intersection of convex sets. Furthermore, the level set $\{x \in \mathbb{R}^n \mid f(x) \leq f^*\}$ is convex by convexity of f. Hence, X^* is the intersection of two convex sets and, thus, X^* is convex. ■

As seen in the proof of Theorem 16, the set C is closed and convex under Assumption 1. Also, under this assumption, the set X^* is closed and convex but possibly empty. The boundedness of X is the key assumption ensuring nonemptiness and boundedness of X^*.

The following theorem is based on Theorem 4. We provide it without a proof. (The proof can be constructed similar to that of Theorem 16. The only new detail is in part (i), where using the coercivity of f, we show that the level sets of f are bounded.)

Theorem 17 *Let Assumption 1 hold. Furthermore, let any of the following conditions be satisfied:*

(i) The function f is coercive over C.

(ii) For some $\gamma \in \mathbb{R}$, the set $\{x \in C \mid f(x) \leq \gamma\}$ is nonempty and compact.

(iii) The set C is compact.

Then, the optimal set X^* of problem (2.5) is nonempty, compact, and convex.

For a quadratic convex objective and a linear constraint set, the existence of solutions is equivalent to finiteness of the optimal value. Furthermore, the issue of existence of solutions can be resolved by checking a "linear condition", as seen in the following theorem.

Theorem 18 *Consider the problem*

$$\begin{array}{ll} \textit{minimize} & f(x) = x^T P x + c^T x \\ \textit{subject to} & Ax \leq b, \end{array}$$

where P is an $n \times n$ positive semidefinite matrix, $c \in \mathbb{R}^n$, A is an $m \times n$ matrix, and $b \in \mathbb{R}^m$. The following statements are equivalent:

(1) The optimal value f^* is finite.

(2) The optimal set X^* is nonempty.

(3) If $Ay \leq 0$ and $Py = 0$ for some $y \in \mathbb{R}^n$, then $c^T y \geq 0$.

The proof of Theorem 18 requires the notion of recession directions of convex closed sets, which is beyond the scope of these notes. The interested reader can find more discussion on this in Bertsekas, Nedić and Ozdaglar [9] (see there Proposition 2.3.5), or in Auslender and Teboulle [2].

As an immediate consequence of Theorem 18, we can derive the conditions for existence of solutions of linear programming problems.

Corollary 1 *Consider a linear programming problem*

$$\begin{array}{ll} \textit{minimize} & f(x) = c^T x \\ \textit{subject to} & Ax \leq b. \end{array}$$

The following conditions are equivalent for the LP problem:

(1) The optimal value f^* is finite.

(2) The optimal set X^* is nonempty.

(3) If $Ay \leq 0$ for some $y \in \mathbb{R}^n$, then $c^T y \geq 0$.

A linear programming (LP) problem that has a solution, it always has a solution of a specific structure. This specific solution is due to the geometry of the polyhedral constraint set. We describe this specific solution for an LP problem in a *standard form*:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0, \end{array} \tag{2.9}$$

where A is an $m \times n$ matrix and $b \in \mathbb{R}^m$. The feasible set for the preceding LP problem is the polyhedral set $\{x \in \mathbb{R}^n \mid Ax = b,\ x \geq 0\}$.

Definition 5 *We say that x is a basic feasible solution for the LP in Eq. (2.9), when x is feasible and there are n linearly independent constraints among the constraints that x satisfies as equalities.*

The preceding definition actually applies to an LP problem in any form and not necessarily in the standard form. Furthermore, a basic solution of an LP is a vertex (or extreme point) of the (polyhedral) constraint set of the given LP, which are out of the scope of these lecture notes. The interested readers may find more on this, for example, in the textbook on linear optimization by Bertsimas and Tsitsiklis [11].

Note that for a given polyhedral set, there can be only finitely many basic solutions. However, the number of such solutions may be very large. For example, the cube $\{x \in \mathbb{R}^n \mid 0 \le x_i \le 1,\ i = 1, \ldots, n\}$ is given by $2n$ inequalities, and it has 2^n basic solutions.

We say that a vector x is *a basic solution* if it satisfies Definition 5 apart from being feasible. Specifically, x is a basic solution for (2.9) if there are n linearly independent constraints among the constraints that x satisfies as equalities. A basic solution x is *degenerate* if more than n constraints are satisfied as equalities at x (active at x). Otherwise, it is *nondegenerate*.

Example 8 *Consider the polyhedral set given by*

$$\begin{array}{ll} \textit{minimize} & x_1 + x_2 + x_3 \le 2 \\ \textit{subject to} & x_2 + 2x_3 \le 2 \\ & x_1 \le 1 \\ & x_3 \le 1 \\ & x_1 \ge 0,\ x_2 \ge 0,\ x_3 \ge 0. \end{array}$$

The vector $\tilde{x} = (1, 1, 0)$ is a nondegenerate basic feasible solution since there are exactly three linearly independent constraints that are active at $\tilde{x}$, specifically,

$$x_1 + x_2 + x_3 \le 2, \qquad x_1 \le 1, \qquad x_3 \ge 0.$$

The vector $\hat{x} = (1, 0, 1)$ is a degenerate feasible solution since there are five constraints active at $\hat{x}$, namely

$$x_1 + x_2 + x_3 \le 2, \qquad x_2 + 2x_3 \le 2, \qquad x_1 \le 1, \qquad x_3 \le 1, \qquad x_2 \ge 0.$$

Out of these, for example, the last three are linearly independent.

We are now ready to state a fundamental result for linear programming solutions.

Theorem 19 *Consider an LP problem. Assume that its constraint set has at least one basic feasible solution and that the LP has an optimal solution. Then, there exists an optimal solution which is also a basic feasible solution.*

2.3.3 Optimality Conditions

In this section, we deal with a differentiable convex function. We have the following.

Theorem 20 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable convex function, and let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set. Consider the problem*

$$\begin{array}{ll} \textit{minimize} & f(x) \\ \textit{subject to} & x \in C. \end{array}$$

A vector x^* is optimal for this problem if and only if $x^* \in C$ and

$$\nabla f(x^*)^T (z - x^*) \ge 0 \qquad \textit{for all } z \in C.$$

Proof. For the sake of simplicity, we prove the result assuming that f is continuously differentiable.

Let x^* be optimal. Suppose that for some $\hat{z} \in C$ we have

$$\nabla f(x^*)^T(\hat{z} - x^*) < 0.$$

Since f is continuously differentiable, by the first-order Taylor expansion [Theorem 6(a)], we have for all sufficiently small $\alpha > 0$,

$$f(x^* + \alpha(\hat{z} - x^*)) = f(x^*) + \alpha \nabla f(x^*)^T(\hat{z} - x^*) + o(\alpha) < f(x^*),$$

with $x^* \in C$ and $\hat{z} \in C$. By the convexity of C, we have $x^* + \alpha(\hat{z} - x^*) \in C$. Thus, this vector is feasible and has a smaller objective value than the optimal point x^*, which is a contradiction. Hence, we must have $\nabla f(x^*)^T(z - x^*) \geq 0$ for all $z \in C$.

Suppose now that $x^* \in C$ and

$$\nabla f(x^*)^T(z - x^*) \geq 0 \qquad \text{for all } z \in C. \tag{2.10}$$

By convexity of f [see Theorem 12], we have

$$f(x^*) + \nabla f(x^*)^T(z - x^*) \leq f(z) \qquad \text{for all } z \in C,$$

implying that

$$\nabla f(x^*)^T(z - x^*) \leq f(z) - f(x^*).$$

This and Eq. (2.10) further imply that

$$0 \leq f(z) - f(x^*) \qquad \text{for all } z \in C.$$

Since $x^* \in C$, it follows that x^* is optimal. ∎

We next discuss several implications of Theorem 20, by considering some special choices for the set C. Let C be the entire space, i.e., $C = \mathbb{R}^n$. The condition

$$\nabla f(x^*)^T(z - x^*) \geq 0 \qquad \text{for all } z \in C$$

reduces to

$$\nabla f(x^*)^T d \geq 0 \qquad \text{for all } d \in \mathbb{R}^n. \tag{2.11}$$

In turn, this is equivalent to

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

Thus, by Theorem 20, *a vector x^* is a minimum of f over $\mathbb{R}^n$ if and only if $\nabla f(x^*) = 0$.*

Let the set C be affine, i.e., the problem of interest is

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b, \end{array} \tag{2.12}$$

where A is an $m \times n$ matrix and $b \in \mathbb{R}^m$. In this case, the condition of Theorem 20 reduces to

$$\nabla f(x^*)^T y \geq 0 \qquad \text{for all } y \in N_A,$$

where N_A is the null space of the matrix A. Thus, the gradient $\nabla f(x^*)$ is orthogonal to the null space N_A. Since the range of A^T is orthogonal to N_A [see Eq. (1.2)], it follows that $\nabla f(x^*)$ belongs to the range of A^T, implying that

$$\nabla f(x^*) + A^T\lambda^* = 0 \qquad \text{for some } \lambda^* \in \mathbb{R}^m.$$

Hence, by Theorem 20, *a vector x^* solves problem (2.12) if and only if $Ax^* = b$ and there exists $\lambda^* \in \mathbb{R}^m$ such that $\nabla f(x^*) + A^T\lambda^* = 0$.* The relation $\nabla f(x^*) + A^T\lambda^* = 0$ is known as *primal optimality condition*. It is related to Lagrangian duality, which is the focus of Section 2.5.

Let C be the nonnegative orthant in $\mathbb{R}^n$, i.e., the problem of interest is

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \geq 0. \end{array} \tag{2.13}$$

For this problem, the condition of Theorem 20 is equivalent to

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

Therefore, *a vector x^* solves problem (2.13) if and only if $x^* \geq 0$ and $\nabla f(x^*)^T x^* = 0$.* The relation $\nabla f(x^*)^T x^* = 0$ is known as *complementarity condition*, and the terminology comes again from the Lagrangian duality theory.

Let C be a simplex in $\mathbb{R}^n$, i.e., the problem of interest is

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \geq 0, \ \displaystyle\sum_{i=1}^n x_i = a, \end{array} \tag{2.14}$$

where $a > 0$ is a scalar. By Theorem 20, x^* is optimal if and only if

$$\sum_{i=1}^n \frac{\partial f(x^*)}{x_i}(x_i - x_i^*) \geq 0 \qquad \text{for all } x_i \geq 0 \text{ with } \textstyle\sum_{i=1}^n x_i = a.$$

Consider an index i with $x_i^* > 0$. Let $j \neq i$ and consider a feasible vector x with $x_i = 0$, $x_j = x_j^* + x_i^*$ and all the other coordinates the same as those of x^*. By using this vector in the preceding relation, we obtain

$$\left(\frac{\partial f(x^*)}{x_j} - \frac{\partial f(x^*)}{x_i}\right) x_i^* \geq 0 \qquad \text{for all } i \text{ such that } x_i^* > 0,$$

or equivalently

$$\frac{\partial f(x^*)}{x_i} \leq \frac{\partial f(x^*)}{x_j} \qquad \text{for all } i \text{ such that } x_i^* > 0. \tag{2.15}$$

Hence, x^* is an optimal solution to problem (2.14) if and only if x^* satisfies relation (2.15).

Let us illustrate the optimality conditions for a simplical constraint set on the problem of optimal routing in a communication network (see [5] and [17]).

Example 9 *(Optimal Routing) Consider a directed graph modeling a data communication network. Let $\mathcal{S}$ be a set of origin-destination pairs, i.e., each $s \in \mathcal{S}$ is an ordered pair (i_s, j_s) of nodes i_s and j_s in the network, with i_s being the origin and j_s being the destination of s. Let y_s be the traffic flow of s (data units/second) i.e., the arrival rate of traffic entering the network at the origin of s and exiting the network at the destination of s. The traffic flow of s is routed through different paths in the network. There is a cost associated with using the links $\mathcal{L}$ of the network, namely, the cost of sending a flow z_{ij} on the link $(i,j) \in \mathcal{L}$ is $f_{ij}(z_{ij})$, where f_{ij} is convex and continuously differentiable. The problem is to decide on paths along which the flow y_s should be routed, so as to minimize the total cost.*

To formalize the problem, we introduce the following notation:

- *$\mathcal{P}_s$ is the set of all directed paths from the origin of s to the destination of s.*

- *x_s is the part of the flow y_s routed through the path p with $p \in \mathcal{P}_s$.*

Let x denote a vector of path-flow variables, i.e.,

$$x = \{x_p \mid p \in \mathcal{P}_s,\ s \in \mathcal{S}\}.$$

Then, the routing problem can be casted as the following convex minimization:

$$\begin{aligned}
&\textit{minimize} && f(x) = \sum_{(i,j)\in\mathcal{L}} f_{ij}\left(\sum_{\{p \mid (i,j)\in p\}} x_p\right) \\
&\textit{subject to} && \sum_{p\in\mathcal{P}_s} x_p = y_s \qquad \textit{for all } s \in \mathcal{S} \\
& && x_p \ge 0 \qquad \textit{for all } p \in \mathcal{P}_s \textit{ and all } s \in \mathcal{S}.
\end{aligned}$$

The cost on link (i,j) depends on the total flow through that link, i.e., the sum of all flows x_p along paths p that contain the link (i,j). The problem is convex in variable x, with differentiable objective function and a constraint set given by a Cartesian product of simplices (one simplex per s).

We now consider the optimality conditions for the routing problem. Note that

$$\frac{\partial f(x)}{x_p} = \sum_{(i,j)\in p} f'_{ij}(z_{ij}),$$

with z_{ij} being the total flow on the link (i,j). When $f'_{ij}(z_{ij})$ is viewed as the length of the link (i,j) evaluated at z_{ij}, the partial derivative $\frac{\partial f(x)}{x_p}$ is the length of the path p. By the necessary and sufficient conditions for a simplex [cf. Eq. 2.14], for all $s \in \mathcal{S}$, we have $x_p^* > 0$ when

$$\frac{\partial f(x^*)}{x_p} \le \frac{\partial f(x^*)}{x_{\tilde{p}}} \qquad \textit{for all } \tilde{p} \in \mathcal{P}_s.$$

This relation means that a set of path-flows is optimal if and only if the flow is positive only on the shortest paths (where link length is measured by the first derivative). It also means that at an optimum x^*, for an $s \in \mathcal{S}$, all the paths $p \in \mathcal{P}_s$ carrying a positive flow $x_p^* > 0$ have the same length (i.e., the traffic y_s is routed through the paths of equal length).

In the absence of convexity, the point $x^* \in C$ satisfying the condition

$$f(x^*)^T(z - x^*) \geq 0 \qquad \text{for all } z \in C$$

is referred to as *a stationary point*. Such a point may be a local or global minimum of f over C. *A global minimum of f over C* is any solution to the problem of minimizing f over C. *A local minimum of f over C* is a point $\tilde{x} \in C$ for which there exists a ball $B(\tilde{x}, r)$ such that there is no "better" point among the points that belong to the ball $B(\tilde{x}, r)$ and the set C, i.e., a ball $B(\tilde{x}, r)$ such that

$$f(x) \geq f(\tilde{x}) \qquad \text{for all } x \in C \cap B(\tilde{x}, r).$$

For convex problems (i.e., convex f and C), there is no distinction between local and global minima: *every local minimum is also global in convex problems.* This makes solving convex minimization problems "easier" than solving a more general "nonconvex" problems.

Let us note that for a strictly convex function, the optimal solution to the problem of minimizing f over C is unique (of course, when a solution exists). We state this result in the following theorem, whose proof follows from the definition of strict convexity.

Theorem 21 *Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set and let f be a strictly convex function over C. If the problem of minimizing f over C has a solution, then the solution is unique.*

Proof. To arrive at a contradiction, assume that the optimal set X^* has more than one point. Let x_1^* and x_2^* be two distinct solutions, i.e., $f(x_1^*) = f(x_2^*) = f^*$ and $x_1^* \neq x_2^*$. Also, let $\alpha \in (0, 1)$. Since f is convex, the set X^* is also convex implying that $\alpha x_1^* + (1 - \alpha)x_2^* \in X^*$. Hence,

$$f(\alpha x_1^* + (1 - \alpha)x_2^*) = f^*. \tag{2.16}$$

At the same time, by strict convexity of f over C and the relation $X^* \subseteq C$, we have that f is strictly convex over X^*, so that

$$f(\alpha x_1^* + (1 - \alpha)x_2^*) < \alpha f(x_1^*) + (1 - \alpha)f(x_2^*) = f^*,$$

which contradicts relation (2.16). Therefore, the solution must be unique. ■

2.3.4 Projection Theorem

One special consequence of Theorems 4 and 20 is the Projection Theorem. The theorem guarantees the existence and uniqueness of the projection of a vector on a closed convex set. This result has a wide range of applications.

For a given nonempty set $C \subseteq \mathbb{R}^n$ and a vector $\hat{x}$, the projection problem is the problem of determining the point $x^* \in C$ that is the closest to $\hat{x}$ among all $x \in C$ (with respect to the Euclidean distance). Formally, the problem is given by

$$\begin{array}{ll} \text{minimize} & \|x - \hat{x}\|^2 \\ \text{subject to} & x \in C. \end{array} \tag{2.17}$$

In general, such a problem may not have an optimal solution and the solution need not be unique (when it exists). However, when the set C is closed and convex set, the solution exists and it is unique, as seen in the following theorem.

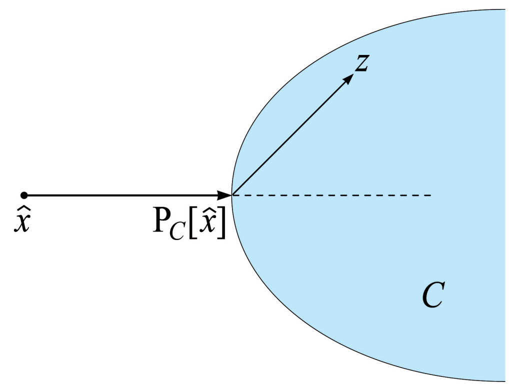

Figure 2.9: The projection of a vector $\hat{x}$ on the closed convex set C is the vector $P_C[\hat{x}] \in C$ that is the closest to $\hat{x}$ among all $x \in C$, with respect to the Euclidean distance.

Theorem 22 *(Projection Theorem) Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set and $\hat{x} \in \mathbb{R}^n$ be a given arbitrary vector.*

(a) The projection problem in Eq. (2.17) has a unique solution.

(b) A vector $x^* \in C$ is the solution to the projection problem if and only if

$$(x^* - \hat{x})^T(x - x^*) \geq 0 \qquad \text{for all } x \in C.$$

Proof. (a) The function $f(x) = \|x - \hat{x}\|$ is coercive over $\mathbb{R}^n$, and therefore coercive over C (i.e., $\lim_{\|x\|\to\infty,\, x\in C} f(x) = \infty$). The set C is closed, and therefore by Theorem 4, the optimal set X^* for projection problem (2.17) is nonempty.

Furthermore, the Hessian of f is given by $\nabla^2 f(x) = 2I$, which is positively definite everywhere. Therefore, by Theorem 11(b), the function f is strictly convex and the optimal solution is unique [cf. Theorem 21].

(b) By the first-order optimality condition of Theorem 20), we have $x^* \in C$ is a solution to the projection problem if and only if

$$\nabla f(x^*)^T(x - x^*) \geq 0 \qquad \text{for all } x \in C.$$

Since $\nabla f(x) = 2(x - \hat{x})$, the result follows. ■

The projection of a vector $\hat{x}$ to a closed convex set C is illustrated in Figure 2.9. The unique solution x^* to the projection problem is referred to as *the projection of $\hat{x}$ on C*, and it is denoted by $P_C[\hat{x}]$. The projection $P_C[\hat{x}]$ has some special properties, as given in the following theorem.

Theorem 23 *Let $C \subseteq \mathbb{R}^n$ be a nonempty closed convex set.*

(a) The projection mapping $P_C : \mathbb{R}^n \to C$ is nonexpansive, i.e.,

$$\|P_C[x] - P_C[y]\| \leq \|x - y\| \qquad \text{for all } x, y \in \mathbb{R}^n.$$

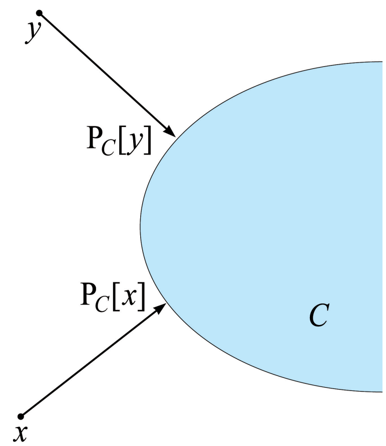

Figure 2.10: The projection mapping $x \mapsto P_C[x]$ is nonexpansive, $\|P_C[x]-P_C[y]\| \leq \|x-y\|$ for all x, y.

(b) The set distance function $d : \mathbb{R}^n \to \mathbb{R}$ given by

$$dist(x, C) = \|P_C[x] - x\|$$

is convex.

Proof. (a) The relation evidently holds for any x and y with $P_C[x] = P_C[y]$. Consider now arbitrary $x, y \in \mathbb{R}^n$ with $P_C[x] \neq P_C[y]$. By Projection Theorem (b), we have

$$(P_C[x] - x)^T(z - P_C[x]) \geq 0 \qquad \text{for all } z \in C, \tag{2.18}$$

$$(P_C[y] - y)^T(z - P_C[y]) \geq 0 \qquad \text{for all } z \in C. \tag{2.19}$$

Using $z = P_C[y]$ in Eq. (2.18) and $z = P_C[x]$ in Eq. (2.19), and by summing the resulting inequalities, we obtain,

$$(P_C[y] - y + x - P_C[x])^T(P_C[x] - P_C[y]) \geq 0.$$

Consequently,

$$(x - y)^T(P_C[x] - P_C[y]) \geq \|P_C[x] - P_C[y]\|^2.$$

Since $P_C[x] \neq P_C[y]$, it follows that $\|y - x\| \geq \|P_C[x] - P_C[y]\|$.

(b) The distance function is equivalently given by

$$dist(x, C) = \min_{z \in C} \|x - z\| \qquad \text{for all } x \in \mathbb{R}^n.$$

The function $h(x, z) = \|x - z\|$ is convex in (x, z) over $\mathbb{R}^n \times \mathbb{R}^n$, and the set C is convex. Hence, by Theorem 13(d), the function $dist(x, C)$ is convex. ■