Chapter 10
Optimality Conditions for Linearly Constrained Problems

In the previous chapter we discussed the notion of *stationarity*, which is a necessary optimality condition for problems with differentiable objective functions and closed convex feasible sets. One of the main drawbacks of this concept is that for most feasible sets, it is rather difficult to validate whether this condition is satisfied or not, and it is even more difficult to use it in order to actually solve the underlying optimization problem. Our main objective in this chapter is to derive an equivalent optimality condition that is much easier to handle. We will establish the so-called KKT conditions for the special case of linearly constrained problems.

10.1 ▪ Separation and Alternative Theorems

We begin with a very simple yet powerful result on convex sets, namely the separation theorem between a point and a closed convex set. This result will be the basis for all the optimality conditions that will be discussed later on. Given a set $S \subseteq \mathbb{R}^n$, a hyperplane $H = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^T\mathbf{x} = b\}$ ($\mathbf{a} \in \mathbb{R}^n \backslash \{0\}, b \in \mathbb{R}$) is said to *strictly separate* a point $\mathbf{y} \notin S$ from S if

$$\mathbf{a}^T\mathbf{y} > b$$

and

$$\mathbf{a}^T\mathbf{x} \leq b \text{ for all } \mathbf{x} \in S.$$

An illustration of a separation between a point and a closed and convex set can be seen in Figure 10.1. Our next result shows that a point can always be strictly separated from a closed convex set, as long as it does not belong to it.

Theorem 10.1 (strict separation theorem). *Let $C \subseteq \mathbb{R}^n$ be a nonempty closed and convex set, and let $\mathbf{y} \notin C$. Then there exist $\mathbf{p} \in \mathbb{R}^n \backslash \{0\}$ and $\alpha \in \mathbb{R}$ such that*

$$\mathbf{p}^T\mathbf{y} > \alpha$$

and

$$\mathbf{p}^T\mathbf{x} \leq \alpha \textit{ for all } \mathbf{x} \in C.$$

Proof. By the second projection theorem (Theorem 9.8), the vector $\bar{\mathbf{x}} = P_C(\mathbf{y}) \in C$ satisfies

$$(\mathbf{y} - \bar{\mathbf{x}})^T(\mathbf{x} - \bar{\mathbf{x}}) \leq 0 \text{ for all } \mathbf{x} \in C,$$

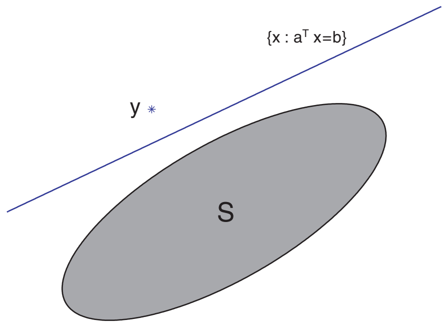

Figure 10.1. *Strict separation of point from a closed and convex set.*

which is the same as

$$(\mathbf{y}-\bar{\mathbf{x}})^T\mathbf{x} \le (\mathbf{y}-\bar{\mathbf{x}})^T\bar{\mathbf{x}} \text{ for all } \mathbf{x} \in C.$$

Denote $\mathbf{p} = \mathbf{y}-\bar{\mathbf{x}} \neq \mathbf{0}$ (since $\mathbf{y} \notin C$) and $\alpha = (\mathbf{y}-\bar{\mathbf{x}})^T\bar{\mathbf{x}}$. Then we have that $\mathbf{p}^T\mathbf{x} \le \alpha$ for all $\mathbf{x} \in C$. On the other hand,

$$\mathbf{p}^T\mathbf{y} = (\mathbf{y}-\bar{\mathbf{x}})^T\mathbf{y} = (\mathbf{y}-\bar{\mathbf{x}})^T(\mathbf{y}-\bar{\mathbf{x}}) + (\mathbf{y}-\bar{\mathbf{x}})^T\bar{\mathbf{x}} = \|\mathbf{y}-\bar{\mathbf{x}}\|^2 + \alpha > \alpha,$$

and the result is established. $\square$

As was already mentioned, the latter separation theorem is extremely important since it is the basis for many optimality conditions. We begin by using it in order to prove an *alternative theorem*, which is known in the literature as *Farkas' lemma*. We refer to it as an alternative theorem since it essentially states that exactly one of two systems ("alternatives") is feasible.

Lemma 10.2 (Farkas' lemma). *Let* $\mathbf{c} \in \mathbb{R}^n$ *and* $\mathbf{A} \in \mathbb{R}^{m\times n}$. *Then* **exactly** *one of the following systems has a solution:*

I. $\mathbf{A}\mathbf{x} \le \mathbf{0}, \mathbf{c}^T\mathbf{x} > 0.$

II. $\mathbf{A}^T\mathbf{y} = \mathbf{c}, \mathbf{y} \ge \mathbf{0}.$

Before proceeding to the proof of the lemma, let us begin with an illustration. For that, consider the following example:

$$\mathbf{A} = \begin{pmatrix} 1 & 5 \\ -1 & 2 \end{pmatrix}, \qquad \mathbf{c} = \begin{pmatrix} -1 \\ 9 \end{pmatrix},$$

Stating that system I is *infeasible* means that the system $\mathbf{A}\mathbf{x} \le \mathbf{0}$ implies the inequality $\mathbf{c}^T\mathbf{x} \le 0$. Thus, the relevant question is whether the inequality $-x_1 + 9x_2 \le 0$ holds whenever the two inequalities

$$\begin{aligned} x_1 + 5x_2 &\le 0, \\ -x_1 + 2x_2 &\le 0 \end{aligned}$$

are satisfied. The answer to this question is affirmative. Indeed, we can see the implication by noting that adding twice the second inequality to the first inequality yields the desired inequality $-x_1 + 9x_2 \le 0$. Thus, the argument for showing the implication is that the row vector $\mathbf{c}^T$ can be written as a conic combination of the rows of $\mathbf{A}$, or in other words, that $\mathbf{c}$ is a conic combination of the columns of $\mathbf{A}^T$:

$$\begin{pmatrix} 1 \\ 5 \end{pmatrix} + 2\begin{pmatrix} -1 \\ 2 \end{pmatrix} = \begin{pmatrix} -1 \\ 9 \end{pmatrix}$$

or

$$\underbrace{\begin{pmatrix} 1 & -1 \\ 5 & 2 \end{pmatrix}}_{\mathbf{A}^T} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \underbrace{\begin{pmatrix} -1 \\ 9 \end{pmatrix}}_{\mathbf{c}}.$$

The interesting question is whether it is always correct that a system of linear inequalities ("base system") implies another linear inequality ("new inequality") if and only if the new inequality can be written as a conic combination of the inequalities in the base system. The answer to this question, according to Farkas' lemma, is yes! We will actually prefer to state Farkas' lemma in the spirit of this discussion. This can be done since an alternative theorem stating that exactly one of two statements A and B is true is equivalent to a result stating that B is equivalent to the denial of A.

Lemma 10.3 (Farkas' lemma, second formulation). *Let* $\mathbf{c} \in \mathbb{R}^n$ *and* $\mathbf{A} \in \mathbb{R}^{m \times n}$. *Then the following two claims are equivalent:*

A. *The implication* $\mathbf{A}\mathbf{x} \leq \mathbf{0} \Rightarrow \mathbf{c}^T\mathbf{x} \leq 0$ *holds true.*

B. *There exists* $\mathbf{y} \in \mathbb{R}^m_+$ *such that* $\mathbf{A}^T\mathbf{y} = \mathbf{c}$.

Proof. Suppose that system B is feasible, meaning that there exists $\mathbf{y} \in \mathbb{R}^m_+$ such that $\mathbf{A}^T\mathbf{y} = \mathbf{c}$. To see that the implication A holds, suppose that $\mathbf{A}\mathbf{x} \leq \mathbf{0}$ for some $\mathbf{x} \in \mathbb{R}^n$. Then multiplying this inequality from the left by $\mathbf{y}^T$ (a valid operation since $\mathbf{y} \geq \mathbf{0}$) yields

$$\mathbf{y}^T\mathbf{A}\mathbf{x} \leq 0.$$

Finally, using the fact that $\mathbf{c}^T = \mathbf{y}^T\mathbf{A}$, we obtain the desired inequality

$$\mathbf{c}^T\mathbf{x} \leq 0.$$

The reverse direction is much less obvious. Suppose that the implication A is satisfied, and let us show that system B is feasible. Suppose in contradiction that system B is infeasible, and consider the following closed and convex set:

$$S = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \mathbf{A}^T\mathbf{y} \text{ for some } \mathbf{y} \in \mathbb{R}^m_+\}.$$

The closedness of the above set follows from Lemma 6.32. The infeasibility of B means that $\mathbf{c} \notin S$. By Theorem 10.1, it follows that there exists a vector $\mathbf{p} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$ and $\alpha \in \mathbb{R}$ such that $\mathbf{p}^T\mathbf{c} > \alpha$ and

$$\mathbf{p}^T\mathbf{x} \leq \alpha \text{ for all } \mathbf{x} \in S. \tag{10.1}$$

Since $\mathbf{0} \in S$, we can conclude that $\alpha \geq 0$, and hence also that $\mathbf{p}^T\mathbf{c} > 0$. In addition, (10.1) is equivalent to

$$\mathbf{p}^T\mathbf{A}^T\mathbf{y} \leq \alpha \text{ for all } \mathbf{y} \geq \mathbf{0}$$

or to

$$(\mathbf{Ap})^T \mathbf{y} \le \alpha \text{ for all } \mathbf{y} \ge 0, \tag{10.2}$$

which implies that $\mathbf{Ap} \le 0$. Indeed, if there was an index $i \in \{1,2,\dots,m\}$ such that $[\mathbf{Ap}]_i > 0$, then for $\mathbf{y} = \beta \mathbf{e}_i$, we would have $(\mathbf{Ap})^T \mathbf{y} = \beta[\mathbf{Ap}]_i$, which is an expression that goes to ∞ as $\beta \to \infty$. Taking a large enough β will contradict (10.2). We have thus arrived at a contradiction to the assumption that the implication A holds (using the vector $\mathbf{p}$), and consequently B is satisfied. $\square$

The next alternative theorem, called "Gordan's theorem," is heavily based on Farkas' lemma.

Theorem 10.4 (Gordan's alternative theorem). *Let $\mathbf{A} \in \mathbb{R}^{m\times n}$. Then exactly one of the following two systems has a solution:*

A. $\mathbf{Ax} < 0$.

B. $\mathbf{p} \ne 0, \mathbf{A}^T\mathbf{p} = 0, \mathbf{p} \ge 0$.

Proof. Suppose that system A has a solution. We will prove that system B is infeasible. Assume in contradiction that B is feasible, meaning that there exists $\mathbf{p} \ne 0$ satisfying $\mathbf{A}^T\mathbf{p} = 0, \mathbf{p} \ge 0$. Multiplying the equality $\mathbf{A}^T\mathbf{p} = 0$ from the left by $\mathbf{x}^T$ yields

$$(\mathbf{Ax})^T\mathbf{p} = 0,$$

which is impossible since $\mathbf{Ax} < 0$ and $0 \ne \mathbf{p} \ge 0$.

Now suppose that system A does not have a solution. Note that system A is equivalent to (s is a scalar)

$$\begin{aligned} \mathbf{Ax} + s\mathbf{e} &\le 0, \\ s &> 0. \end{aligned}$$

The latter system can be rewritten as

$$\tilde{\mathbf{A}}\begin{pmatrix}\mathbf{x}\\ s\end{pmatrix} \le 0, \qquad \mathbf{c}^T\begin{pmatrix}\mathbf{x}\\ s\end{pmatrix} > 0,$$

where $\tilde{\mathbf{A}} = \begin{pmatrix}\mathbf{A} & \mathbf{e}\end{pmatrix}$ and $\mathbf{c} = \mathbf{e}_{n+1}$. The infeasibility of A is thus equivalent to the infeasibility of the system

$$\tilde{\mathbf{A}}\mathbf{w} \le 0, \quad \mathbf{c}^T\mathbf{w} > 0, \quad \mathbf{w} \in \mathbb{R}^{n+1}.$$

By Farkas' lemma, there exists $\mathbf{z} \in \mathbb{R}^m_+$ such that

$$\begin{pmatrix}\mathbf{A}^T\\ \mathbf{e}^T\end{pmatrix}\mathbf{z} = \mathbf{c};$$

that is, there exists $\mathbf{z} \in \mathbb{R}^m_+$ such that

$$\mathbf{A}^T\mathbf{z} = 0, \qquad \mathbf{e}^T\mathbf{z} = 1.$$

Since $\mathbf{e}^T\mathbf{z} = 1$, it follows in particular that $\mathbf{z} \ne 0$, and we have thus shown the existence of $0 \ne \mathbf{z} \in \mathbb{R}^m_+$ such that $\mathbf{A}^T\mathbf{z} = 0$; that is, that system B is feasible. $\square$

10.2 ▪ The KKT conditions

We will now show how Gordan's alternative theorem can be used to establish a very useful optimality criterion that is in fact a special case of the so-called Karush–Kuhn–Tucker (abbreviated KKT) conditions, which will be discussed later on in Chapter 11. The new optimality condition follows from the stationarity condition already discussed in Chapter 9.

Theorem 10.5 (KKT conditions for linearly constrained problems; necessary optimality conditions). *Consider the minimization problem*

$$\text{(P)} \quad \begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & \mathbf{a}_i^T\mathbf{x} \le b_i, \quad i=1,2,\dots,m, \end{array}$$

where f is continuously differentiable over $\mathbb{R}^n$, $\mathbf{a}_1,\mathbf{a}_2,\dots,\mathbf{a}_m \in \mathbb{R}^n$, $b_1,b_2,\dots,b_m \in \mathbb{R}$, and let $\mathbf{x}^*$ be a local minimum point of (P). Then there exist $\lambda_1,\lambda_2,\dots,\lambda_m \ge 0$ such that

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i \mathbf{a}_i = \mathbf{0} \tag{10.3}$$

and

$$\lambda_i(\mathbf{a}_i^T\mathbf{x}^* - b_i) = 0, \quad i=1,2,\dots,m. \tag{10.4}$$

Proof. Since $\mathbf{x}^*$ is a local minimum point of (P), it follows by Theorem 9.2 that $\mathbf{x}^*$ is a stationary point, meaning that $\nabla f(\mathbf{x}^*)^T(\mathbf{x}-\mathbf{x}^*) \ge 0$ for every $\mathbf{x} \in \mathbb{R}^n$ satisfying $\mathbf{a}_i^T\mathbf{x} \le b_i$ for any $i=1,2,\dots,m$. Let us denote the set of *active constraints* by

$$I(\mathbf{x}^*) = \{i : \mathbf{a}_i^T\mathbf{x}^* = b_i\}.$$

Making the change of variables $\mathbf{y} = \mathbf{x}-\mathbf{x}^*$, we obtain that $\nabla f(\mathbf{x}^*)^T\mathbf{y} \ge 0$ for any $\mathbf{y} \in \mathbb{R}^n$ satisfying $\mathbf{a}_i^T(\mathbf{y}+\mathbf{x}^*) \le b_i$ for any $i=1,2,\dots,m$, that is, for any $\mathbf{y} \in \mathbb{R}^n$ satisfying

$$\begin{array}{ll} \mathbf{a}_i^T\mathbf{y} \le 0, & i \in I(\mathbf{x}^*), \\ \mathbf{a}_i^T\mathbf{y} \le b_i - \mathbf{a}_i^T\mathbf{x}^*, & i \notin I(\mathbf{x}^*). \end{array}$$

We will show that in fact the second set of inequalities in the latter system can be removed, that is, that the following implication is valid:

$$\mathbf{a}_i^T\mathbf{y} \le 0 \text{ for all } i \in I(\mathbf{x}^*) \Rightarrow \nabla f(\mathbf{x}^*)^T\mathbf{y} \ge 0.$$

Suppose then that $\mathbf{y}$ satisfies $\mathbf{a}_i^T\mathbf{y} \le 0$ for all $i \in I(\mathbf{x}^*)$. Since $b_i - \mathbf{a}_i^T\mathbf{x}^* > 0$ for all $i \notin I(\mathbf{x}^*)$, it follows that there exists a small enough $\alpha > 0$ for which[4] $\mathbf{a}_i^T(\alpha\mathbf{y}) \le b_i - \mathbf{a}_i^T\mathbf{x}^*$. Thus, since in addition $\mathbf{a}_i^T(\alpha\mathbf{y}) \le 0$ for any $i \in I(\mathbf{x}^*)$ it follows by the stationarity condition that $\nabla f(\mathbf{x}^*)^T(\alpha\mathbf{y}) \ge 0$, and hence that $\nabla f(\mathbf{x}^*)^T\mathbf{y} \ge 0$. We have thus shown that

$$\mathbf{a}_i^T\mathbf{y} \le 0 \text{ for all } i \in I(\mathbf{x}^*) \Rightarrow \nabla f(\mathbf{x}^*)^T\mathbf{y} \ge 0.$$

Thus, by Farkas' lemma it follows that there exist $\lambda_i \ge 0, i \in I(\mathbf{x}^*)$, such that

$$-\nabla f(\mathbf{x}^*) = \sum_{i \in I(\mathbf{x}^*)} \lambda_i \mathbf{a}_i.$$

[4] Indeed, if $\mathbf{a}_i^T\mathbf{y} \le 0$ for all $i \notin I(\mathbf{x}^*)$, then we can take $\alpha = 1$. Otherwise, denote $J = \{i \notin I(\mathbf{x}^*) : \mathbf{a}_i^T\mathbf{y} > 0\}$ and take $\alpha = \min_{i \in J} \frac{b_i - \mathbf{a}_i^T\mathbf{x}^*}{\mathbf{a}_i^T\mathbf{y}}$.

Defining $\lambda_i = 0$ for all $i \notin I(\mathbf{x}^*)$, we get that $\lambda_i(\mathbf{a}_i^T\mathbf{x}^* - b_i) = 0$ for all $i = 1,2,\dots,m$ and that

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i \mathbf{a}_i = \mathbf{0}$$

as required. $\square$

The KKT conditions are necessary optimality conditions, but when the objective function is convex, they are both necessary and sufficient global optimality conditions.

Theorem 10.6 (KKT conditions for convex linearly constrained problems; necessary and sufficient optimality conditions). *Consider the minimization problem*

$$\text{(P)} \quad \begin{array}{ll} \min & f(\mathbf{x}) \\ s.t. & \mathbf{a}_i^T\mathbf{x} \le b_i,\ i = 1,2,\dots,m, \end{array}$$

where f is a convex continuously differentiable function over $\mathbb{R}^n$, $\mathbf{a}_1,\mathbf{a}_2,\dots,\mathbf{a}_m \in \mathbb{R}^n$, $b_1,b_2,\dots,b_m \in \mathbb{R}$, and let $\mathbf{x}^*$ be a feasible solution of (P). *Then $\mathbf{x}^*$ is an optimal solution of (P) if and only if there exist $\lambda_1,\lambda_2,\dots,\lambda_m \ge 0$ such that*

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i \mathbf{a}_i = \mathbf{0} \tag{10.5}$$

and

$$\lambda_i(\mathbf{a}_i^T\mathbf{x}^* - b_i) = 0, \quad i = 1,2,\dots,m. \tag{10.6}$$

Proof. If $\mathbf{x}^*$ is an optimal solution of (P), then by Theorem 10.5 there exist $\lambda_1,\lambda_2,\dots,\lambda_m \ge 0$ such that (10.5) and (10.6) are satisfied. To prove the sufficiency, suppose that $\mathbf{x}^*$ is a feasible solution of (P) satisfying (10.5) and (10.6). Let $\mathbf{x}$ be any feasible solution of (P). Define the function

$$h(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^m \lambda_i(\mathbf{a}_i^T\mathbf{x} - b_i).$$

Then by (10.5) it follows that $\nabla h(\mathbf{x}^*) = \mathbf{0}$, and since h is convex, it follows by Proposition 7.8 that $\mathbf{x}^*$ is a minimizer of h over $\mathbb{R}^n$, which combined with (10.6) implies that

$$f(\mathbf{x}^*) = f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i(\mathbf{a}_i^T\mathbf{x}^* - b_i) = h(\mathbf{x}^*) \le h(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^m \lambda_i(\mathbf{a}_i^T\mathbf{x} - b_i) \le f(\mathbf{x}),$$

where the last inequality follows from the fact that $\lambda_i \ge 0$ and $\mathbf{a}_i^T\mathbf{x} - b_i \le 0$ for $i = 1,2,\dots,m$. We have thus proven that $\mathbf{x}^*$ is a global optimal solution of (P). $\square$

The scalars $\lambda_1,\dots,\lambda_m$ that appear in the KKT conditions are also called *Lagrange multipliers*, and each of the multipliers is associated with a corresponding constraint: λ_i is the multiplier associated with the ith constraint $\mathbf{a}_i^T\mathbf{x} \le b_i$. Note that the multipliers associated with inequality constraints are nonnegative. The conditions (10.6) are known in the literature as the *complementary slackness conditions*. We can also generalize Theorems 10.5 and 10.6 to the case where linear *equality* constraints are also present. The main difference is that the multipliers associated with equality constraints are not restricted to be nonnegative. The proof of the variant that also incorporates equality constraints is based on the simple observation that a linear equality constraint $\mathbf{a}^T\mathbf{x} = b$ can be written as two inequality constraints, $\mathbf{a}^T\mathbf{x} \le b$ and $-\mathbf{a}^T\mathbf{x} \le -b$.

Theorem 10.7 (KKT conditions for linearly constrained problems). *Consider the minimization problem*

$$
\text{(Q)} \quad \begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & \mathbf{a}_i^T \mathbf{x} \le b_i, \quad i = 1,2,\dots,m, \\ & \mathbf{c}_j^T \mathbf{x} = d_j, \quad j = 1,2,\dots,p, \end{array}
$$

where f is a continuously differentiable function over $\mathbb{R}^n$, $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m, \mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_p \in \mathbb{R}^n$, $b_1, b_2, \dots, b_m, d_1, d_2, \dots, d_p \in \mathbb{R}$. Then we have the following:

(a) **(necessity of the KKT conditions)** *If $\mathbf{x}^*$ is a local minimum point of* (Q), *then there exist $\lambda_1, \lambda_2, \dots, \lambda_m \ge 0$ and $\mu_1, \mu_2, \dots, \mu_p \in \mathbb{R}$ such that*

$$
\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \mathbf{a}_i + \sum_{j=1}^{p} \mu_j \mathbf{c}_j = \mathbf{0} \tag{10.7}
$$

and

$$
\lambda_i(\mathbf{a}_i^T \mathbf{x}^* - b_i) = 0, \quad i = 1,2,\dots,m. \tag{10.8}
$$

(b) **(sufficiency in the convex case)** *If in addition f is convex over $\mathbb{R}^n$ and $\mathbf{x}^*$ is a feasible solution of* (Q) *for which there exist $\lambda_1, \lambda_2, \dots, \lambda_m \ge 0$ and $\mu_1, \mu_2, \dots, \mu_p \in \mathbb{R}$ such that* (10.7) *and* (10.8) *are satisfied, then $\mathbf{x}^*$ is an optimal solution of* (Q).

Proof. (a). Consider the equivalent problem

$$
\text{(Q}'\text{)} \quad \begin{array}{lll} \min & f(\mathbf{x}) & \\ \text{s.t.} & \mathbf{a}_i^T \mathbf{x} \le b_i, & i = 1,2,\dots,m, \\ & \mathbf{c}_j^T \mathbf{x} \le d_j, & j = 1,2,\dots,p, \\ & -\mathbf{c}_j^T \mathbf{x} \le -d_j, & j = 1,2,\dots,p. \end{array}
$$

Then since $\mathbf{x}^*$ is an optimal solution of (Q), it is also an optimal solution of (Q′), and thus, by Theorem 10.5, it follows that there exist multipliers $\lambda_1, \lambda_2, \dots, \lambda_m \ge 0$ and $\mu_1^+, \mu_1^-, \mu_2^+, \mu_2^-, \dots, \mu_p^+, \mu_p^- \ge 0$ such that

$$
\nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \mathbf{a}_i + \sum_{j=1}^{p} \mu_j^+ \mathbf{c}_j - \sum_{j=1}^{p} \mu_j^- \mathbf{c}_j = \mathbf{0} \tag{10.9}
$$

and

$$
\lambda_i(\mathbf{a}_i^T \mathbf{x}^* - b_i) = 0, \quad i = 1,2,\dots,m, \tag{10.10}
$$

$$
\mu_j^+(\mathbf{c}_j^T \mathbf{x}^* - d_j) = 0, \quad j = 1,2,\dots,p, \tag{10.11}
$$

$$
\mu_j^-(-\mathbf{c}_j^T \mathbf{x}^* + d_j) = 0, \quad j = 1,2,\dots,p. \tag{10.12}
$$

We thus obtain that (10.7) and (10.8) are satisfied with $\mu_j = \mu_j^+ - \mu_j^-, j = 1,2,\dots,p$.

(b) To prove the second part, suppose that $\mathbf{x}^*$ satisfies (10.7) and (10.8). Then it also satisfies (10.9), (10.10), (10.11), and (10.12) with

$$
\mu_j^+ = [\mu_j]_+, \mu_j^- = [\mu_j]_- = -\min\{\mu_j, 0\},
$$

which by Theorem 10.6 implies that $\mathbf{x}^*$ is an optimal solution of (Q′) and thus also an optimal solution of (Q). $\square$

We note that a feasible point $\mathbf{x}^*$ is called *a KKT point* if there exist multipliers for which (10.7) and (10.8) are satisfied.

A very popular representation of the KKT conditions is via the Lagrangian function, which we will present in the setting of general nonlinear programming problems:

$$
\text{(NLP)} \quad \begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & g_i(\mathbf{x}) \le 0, \quad i = 1, 2, \ldots, m, \\ & h_j(\mathbf{x}) = 0, \quad j = 1, 2, \ldots, p. \end{array}
$$

Here $f, g_1, g_2, \ldots, g_m, h_1, h_2, \ldots, h_p$ are all continuously differentiable functions over $\mathbb{R}^n$. The associated Lagrangian function takes the form

$$
L(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^{p} \mu_j h_j(\mathbf{x}).
$$

In the linearly constrained case of problem (Q), the condition (10.7) is the same as

$$
\nabla_{\mathbf{x}} L(\mathbf{x}, \lambda, \mu) = \nabla f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\mathbf{x}) + \sum_{j=1}^{p} \mu_j \nabla h_j(\mathbf{x}) = \mathbf{0}.
$$

Back to problem (Q), if in addition we define the matrices $\mathbf{A}$ and $\mathbf{C}$ and the vectors $\mathbf{b}$ and $\mathbf{d}$ by

$$
\mathbf{A} = \begin{pmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_m^T \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} \mathbf{c}_1^T \\ \mathbf{c}_2^T \\ \vdots \\ \mathbf{c}_p^T \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_p \end{pmatrix},
$$

then the constraints of problem (Q) can be written as

$$
\mathbf{A}\mathbf{x} \le \mathbf{b}, \qquad \mathbf{C}\mathbf{x} = \mathbf{d}.
$$

The Lagrangian can be also written as

$$
L(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \lambda^T (\mathbf{A}\mathbf{x} - \mathbf{b}) + \mu^T (\mathbf{C}\mathbf{x} - \mathbf{d}),
$$

and condition (10.7) takes the form

$$
\nabla_{\mathbf{x}} L(\mathbf{x}, \lambda, \mu) = \nabla f(\mathbf{x}) + \mathbf{A}^T \lambda + \mathbf{C}^T \mu = \mathbf{0}.
$$

Example 10.8. Consider the problem

$$
\begin{array}{ll} \min & \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) \\ \text{s.t.} & x_1 + x_2 + x_3 = 3. \end{array}
$$

Since the problem is convex, the KKT conditions are necessary and sufficient. The Lagrangian of the problem is

$$
L(x_1, x_2, x_3, \mu) = \frac{1}{2}(x_1^2 + x_2^2 + x_3^2) + \mu(x_1 + x_2 + x_3 - 3).
$$

The KKT conditions are (we also incorporate feasibility within the KKT system)

$$\begin{aligned}
\frac{\partial L}{\partial x_1} &= x_1 + \mu = 0,\\
\frac{\partial L}{\partial x_2} &= x_2 + \mu = 0,\\
\frac{\partial L}{\partial x_3} &= x_3 + \mu = 0,\\
x_1 + x_2 + x_3 &= 3.
\end{aligned}$$

By the first three equalities we obtain that $x_1 = x_2 = x_3 = -\mu$. Substituting this in the last equation yields $\mu = -1$, and we obtained that the unique solution of the KKT system is $x_1 = x_2 = x_3 = 1, \mu = -1$. Hence, the unique optimal solution of the problem is $(x_1, x_2, x_3) = (1, 1, 1)$. ∎

Example 10.9. Consider the problem

$$\begin{array}{ll}
\min & x_1^2 + 2x_2^2 + 4x_1x_2\\
\text{s.t.} & x_1 + x_2 = 1,\\
& x_1, x_2 \geq 0.
\end{array}$$

The problem is nonconvex, since the matrix associated with the quadratic objective function $\mathbf{A} = \left(\begin{smallmatrix} 1 & 2\\ 2 & 2\end{smallmatrix}\right)$ is indefinite. However, the KKT conditions are still necessary optimality conditions. The Lagrangian of the problem is

$$L(x_1, x_2, \mu, \lambda_1, \lambda_2) = x_1^2 + 2x_2^2 + 4x_1x_2 + \mu(x_1 + x_2 - 1) - \lambda_1 x_1 - \lambda_2 x_2, \quad \lambda_1, \lambda_2 \in \mathbb{R}_+, \mu \in \mathbb{R}.$$

The KKT conditions are

$$\begin{aligned}
\frac{\partial L}{\partial x_1} &= 2x_1 + 4x_2 + \mu - \lambda_1 = 0,\\
\frac{\partial L}{\partial x_2} &= 4x_2 + 4x_1 + \mu - \lambda_2 = 0,\\
\lambda_1 x_1 &= 0,\\
\lambda_2 x_2 &= 0,\\
x_1 + x_2 &= 1,\\
x_1, x_2 &\geq 0,\\
\lambda_1, \lambda_2 &\geq 0.
\end{aligned}$$

We will split the analysis into 4 cases.

- $\lambda_1 = \lambda_2 = 0$. In this case we obtain the three equations

$$\begin{aligned}
2x_1 + 4x_2 + \mu &= 0,\\
4x_2 + 4x_1 + \mu &= 0,\\
x_1 + x_2 &= 1,
\end{aligned}$$

  whose solution is $x_1 = 0, x_2 = 1, \mu = -4$. We thus obtain that $(x_1, x_2) = (0, 1)$ is a KKT point.

- $\lambda_1, \lambda_2 > 0$. In this case, by the complementary slackness conditions we obtain that $x_1 = x_2 = 0$, which contradicts the constraint $x_1 + x_2 = 1$.
- $\lambda_1 > 0, \lambda_2 = 0$. In this case, by the complementary slackness conditions we have $x_1 = 0$ and consequently $x_2 = 1$, which was already shown to be a KKT point.
- $\lambda_1 = 0, \lambda_2 > 0$. Here $x_2 = 0$ and $x_1 = 1$. The first two equations in the KKT system reduce to

$$\begin{aligned} 2 + \mu &= 0, \\ 4 + \mu - \lambda_2 &= 0. \end{aligned}$$

The solution of this system is $\mu = -2, \lambda_2 = 2$. We thus obtain that $(1,0)$ is also a KKT point.

To summarize, there are two KKT points: $(0,1)$ and $(1,0)$. Since the problem consists of minimizing a continuous function over a compact set it follows by the Weierstrass theorem (Theorem 2.30) that it has a global optimal solution. The KKT conditions are necessary optimality conditions, and hence the optimal solution is either $(0,1)$ or $(1,0)$. Since the respective objective function values are 2 and 1, it follows that $(1,0)$ is the global optimal solution of the problem. ∎

Example 10.10 (orthogonal projection onto an affine space). Let C be the affine space

$$C = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}\},$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$. We assume that the rows of $\mathbf{A}$ are linearly independent. Given $\mathbf{y} \in \mathbb{R}^n$, the optimization problem associated with the problem of finding $P_C(\mathbf{y})$ is

$$\begin{array}{ll} \min & \|\mathbf{x}-\mathbf{y}\|^2 \\ \text{s.t.} & \mathbf{A}\mathbf{x} = \mathbf{b}. \end{array}$$

This is a convex optimization problem, so the KKT conditions are necessary and sufficient. The Lagrangian function is

$$L(\mathbf{x}, \lambda) = \|\mathbf{x}-\mathbf{y}\|^2 + 2\lambda^T(\mathbf{A}\mathbf{x}-\mathbf{b}) = \|\mathbf{x}\|^2 - 2(\mathbf{y}-\mathbf{A}^T\lambda)^T\mathbf{x} - 2\lambda^T\mathbf{b} + \|\mathbf{y}\|^2, \quad \lambda \in \mathbb{R}^m.$$

Therefore, the KKT conditions are

$$\begin{aligned} 2\mathbf{x} - 2(\mathbf{y}-\mathbf{A}^T\lambda) &= 0, \\ \mathbf{A}\mathbf{x} &= \mathbf{b}. \end{aligned}$$

The first equation can be written as

$$\mathbf{x} = \mathbf{y} - \mathbf{A}^T\lambda. \tag{10.13}$$

Substituting this expression for $\mathbf{x}$ in the second equation yields the equation

$$\mathbf{A}(\mathbf{y} - \mathbf{A}^T\lambda) = \mathbf{b},$$

which is the same as

$$\mathbf{A}\mathbf{A}^T\lambda = \mathbf{A}\mathbf{y} - \mathbf{b}.$$

Thus,

$$\lambda = (\mathbf{A}\mathbf{A}^T)^{-1}(\mathbf{A}\mathbf{y} - \mathbf{b}),$$

where here we used the fact that $\mathbf{A}\mathbf{A}^T$ is nonsingular since the rows of $\mathbf{A}$ are linearly independent. Plugging the latter expression for λ into (10.13), we obtain that

$$P_C(\mathbf{y}) = \mathbf{y} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}(\mathbf{A}\mathbf{y} - \mathbf{b}).$$

Note that the projection onto an affine space is by itself an affine transformation. ∎

In the last example we multiplied the Lagrange multiplier in the Lagrangian function by 2. This was done to simplify the computations, and it is always allowed to multiply the Lagrange multipliers by a positive constant.

Example 10.11 (orthogonal projection onto hyperplanes). Consider the hyperplane

$$H = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^T\mathbf{x} = b\}.$$

where $0 \neq \mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Since a hyperplane is a spacial case of an affine space, we can use the formula obtained in the last example in order to derive an explicit expression for the projection onto H:

$$P_H(\mathbf{y}) = \mathbf{y} - \mathbf{a}(\mathbf{a}^T\mathbf{a})^{-1}(\mathbf{a}^T\mathbf{y} - b) = \mathbf{y} - \frac{\mathbf{a}^T\mathbf{y} - b}{\|\mathbf{a}\|^2}\mathbf{a}. \quad \blacksquare$$

As a consequence of the last example, we can write a result providing an explicit expression for the distance between a point and a hyperplane. (The result was already stated and not proved in Lemma 8.6.)

Lemma 10.12 (distance of a point from a hyperplane). *Let $H = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^T\mathbf{x} = b\}$, where $0 \neq \mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Then*

$$d(\mathbf{y}, H) = \frac{|\mathbf{a}^T\mathbf{y} - b|}{\|\mathbf{a}\|}.$$

Proof.

$$d(\mathbf{y}, H) = \|\mathbf{y} - P_H(\mathbf{y})\| = \left\|\mathbf{y} - \left(\mathbf{y} - \frac{\mathbf{a}^T\mathbf{y} - b}{\|\mathbf{a}\|^2}\mathbf{a}\right)\right\| = \frac{|\mathbf{a}^T\mathbf{y} - b|}{\|\mathbf{a}\|}. \quad \square$$

Example 10.13 (orthogonal projection onto half-spaces). Let

$$H^- = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^T\mathbf{x} \leq b\},$$

where $0 \neq \mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$. The corresponding optimization problem is

$$\begin{array}{ll} \min_{\mathbf{x}} & \|\mathbf{x} - \mathbf{y}\|^2 \\ \text{s.t.} & \mathbf{a}^T\mathbf{x} \leq b. \end{array}$$

The Lagrangian of the problem is

$$L(\mathbf{x}, \lambda) = \|\mathbf{x} - \mathbf{y}\|^2 + 2\lambda(\mathbf{a}^T\mathbf{x} - b), \quad \lambda \geq 0,$$

and the KKT conditions are

$$\begin{aligned} 2(\mathbf{x}-\mathbf{y})+2\lambda\mathbf{a} &= \mathbf{0}, \\ \lambda(\mathbf{a}^T\mathbf{x}-b) &= 0, \\ \mathbf{a}^T\mathbf{x} &\leq b, \\ \lambda &\geq 0. \end{aligned}$$

If $\lambda = 0$, then $\mathbf{x} = \mathbf{y}$ and the KKT conditions are satisfied when $\mathbf{a}^T\mathbf{y} \leq b$. That is, when $\mathbf{a}^T\mathbf{y} \leq b$, the optimal solution is $\mathbf{x} = \mathbf{y}$, which is not a surprise since *for any* set C, $P_C(\mathbf{y}) = \mathbf{y}$ whenever $\mathbf{y} \in C$. Now assume that $\lambda > 0$; then by the complementary slackness condition we have that

$$\mathbf{a}^T\mathbf{x} = b. \tag{10.14}$$

Plugging the first equation $\mathbf{x} = \mathbf{y} - \lambda\mathbf{a}$ into (10.14), we obtain that

$$\mathbf{a}^T(\mathbf{y}-\lambda\mathbf{a}) = b,$$

so that $\lambda = \frac{\mathbf{a}^T\mathbf{y}-b}{\|\mathbf{a}\|^2}$. The multiplier λ is indeed positive when $\mathbf{a}^T\mathbf{y} > b$. We have thus obtained that when $\mathbf{a}^T\mathbf{y} > b$, the optimal solution is

$$\mathbf{x} = \mathbf{y} - \frac{\mathbf{a}^T\mathbf{y}-b}{\|\mathbf{a}\|^2}\mathbf{a}.$$

To summarize,

$$P_H(\mathbf{y}) = \begin{cases} \mathbf{y}, & \mathbf{a}^T\mathbf{y} \leq b, \\ \mathbf{y} - \frac{\mathbf{a}^T\mathbf{y}-b}{\|\mathbf{a}\|^2}\mathbf{a}, & \mathbf{a}^T\mathbf{y} > b. \end{cases}$$

The latter expression can also be compactly written as

$$P_H(\mathbf{y}) = \mathbf{y} - \frac{[\mathbf{a}^T\mathbf{y}-b]_+}{\|\mathbf{a}\|^2}\mathbf{a}.$$

An illustration of the orthogonal projection onto a hyperplane can be found in Figure 10.2. ∎

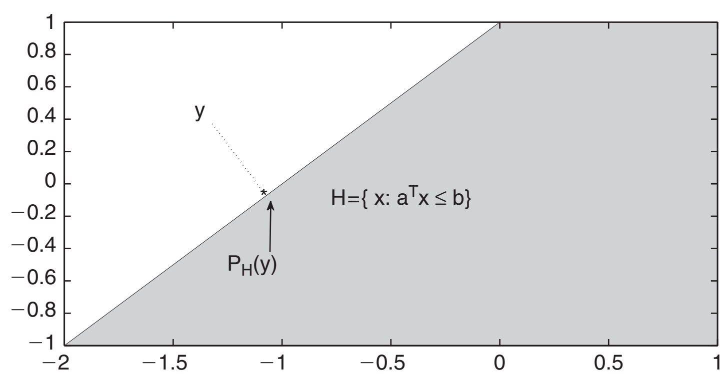

Figure 10.2. *A vector* $\mathbf{y}$ *and its orthogonal projection onto a half-space.*

Example 10.14. Consider the optimization problem

$$\min\{\mathbf{x}^T\mathbf{Q}\mathbf{x}+2\mathbf{c}^T\mathbf{x}:\mathbf{A}\mathbf{x}=\mathbf{b}\},$$

where $\mathbf{Q}\in\mathbb{R}^{n\times n}$ is a positive definite matrix, $\mathbf{c}\in\mathbb{R}^n,\mathbf{b}\in\mathbb{R}^m$, and $\mathbf{A}$ is an $m\times n$ matrix with linearly independent rows. The Lagrangian of the problem is

$$L(\mathbf{x},\lambda)=\mathbf{x}^T\mathbf{Q}\mathbf{x}+2\mathbf{c}^T\mathbf{x}+2\lambda^T(\mathbf{A}\mathbf{x}-\mathbf{b}),$$

and the KKT conditions are

$$\begin{aligned}\nabla_{\mathbf{x}}L(\mathbf{x},\lambda)=2\left[\mathbf{Q}\mathbf{x}+\mathbf{c}+\mathbf{A}^T\lambda\right]&=\mathbf{0},\\ \mathbf{A}\mathbf{x}&=\mathbf{b}.\end{aligned}$$

The first equation implies that

$$\mathbf{x}=-\mathbf{Q}^{-1}(\mathbf{c}+\mathbf{A}^T\lambda). \tag{10.15}$$

Plugging this expression of $\mathbf{x}$ into the feasibility constraint we obtain

$$-\mathbf{A}\mathbf{Q}^{-1}(\mathbf{c}+\mathbf{A}^T\lambda)=\mathbf{b},$$

so that

$$\lambda=-(\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}^T)^{-1}(\mathbf{b}+\mathbf{A}\mathbf{Q}^{-1}\mathbf{c}). \tag{10.16}$$

The optimal solution of the problem is given by (10.15) with λ as in (10.16). ∎

10.3 ▪ Orthogonal Regression

An interesting application to the formula for the distance between a point and a hyperplane given in Lemma 10.12 is in the orthogonal regression problem, which we now recall. Consider the points $\mathbf{a}_1,\ldots,\mathbf{a}_m$ in $\mathbb{R}^n$. For a given $\mathbf{0}\neq\mathbf{x}\in\mathbb{R}^n$ and $y\in\mathbb{R}$, we define the hyperplane

$$H_{\mathbf{x},y}:=\left\{\mathbf{a}\in\mathbb{R}^n:\mathbf{x}^T\mathbf{a}=y\right\}.$$

In the orthogonal regression problem we seek to find a nonzero vector $\mathbf{x}\in\mathbb{R}^n$ and $y\in\mathbb{R}$ such that the sum of squared Euclidean distances between the points $\mathbf{a}_1,\ldots,\mathbf{a}_m$ to $H_{\mathbf{x},y}$ is minimal; that is, the problem is given by

$$\min_{\mathbf{x},y}\left\{\sum_{i=1}^m d(\mathbf{a}_i,H_{\mathbf{x},y})^2:\mathbf{0}\neq\mathbf{x}\in\mathbb{R}^n,y\in\mathbb{R}\right\}. \tag{10.17}$$

An illustration of the solution to the orthogonal regression problem is given in Figure 10.3.

The optimal solution of the orthogonal regression problem is described in the next result whose proof strongly relies on the formula of the distance between a point and a hyperplane.

Proposition 10.15. *Let $\mathbf{a}_1,\ldots,\mathbf{a}_m\in\mathbb{R}^n$ and let $\mathbf{A}$ be the matrix given by*

$$\mathbf{A}=\begin{pmatrix}\mathbf{a}_1^T\\ \mathbf{a}_2^T\\ \vdots\\ \mathbf{a}_m^T\end{pmatrix}.$$

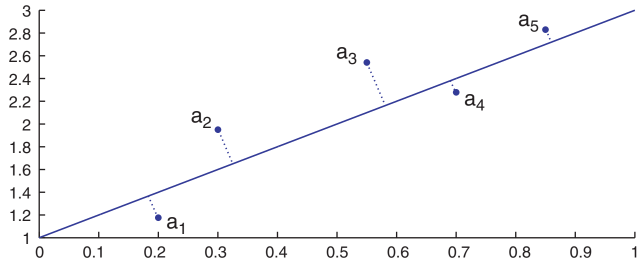

Figure 10.3. *A two-dimensional example: given 5 points* $\mathbf{a}_1,\ldots,\mathbf{a}_5$ *in the plane, the orthogonal regression problem seeks to find the line for which the sum of squared norms of the dashed lines is minimal.*

Then an optimal solution of problem (10.17) *is given by* $\mathbf{x}$ *that is an eigenvector of the matrix* $\mathbf{A}^T(\mathbf{I}_m - \frac{1}{m}\mathbf{e}\mathbf{e}^T)\mathbf{A}$ *associated with the minimum eigenvalue and* $y = \frac{1}{m}\sum_{i=1}^m \mathbf{a}_i^T\mathbf{x}$. *Here* $\mathbf{e}$ *is the* m*-length vector of ones. The optimal function value of problem* (10.17) *is* $\lambda_{\min}[\mathbf{A}^T(\mathbf{I}_m - \frac{1}{m}\mathbf{e}\mathbf{e}^T)\mathbf{A}]$.

Proof. By Lemma 10.12, the squared Euclidean distance between the point $\mathbf{a}_i$ to $H_{\mathbf{x},y}$ is given by

$$d(\mathbf{a}_i, H_{\mathbf{x},y})^2 = \frac{(\mathbf{a}_i^T\mathbf{x} - y)^2}{\|\mathbf{x}\|^2}, \quad i = 1,\ldots,m.$$

It follows that (10.17) is the same as

$$\min\left\{\sum_{i=1}^m \frac{(\mathbf{a}_i^T\mathbf{x} - y)^2}{\|\mathbf{x}\|^2} : \mathbf{0} \neq \mathbf{x} \in \mathbb{R}^n, y \in \mathbb{R}\right\}. \tag{10.18}$$

Fixing $\mathbf{x}$ and minimizing first with respect to y we obtain that the optimal y is given by

$$y = \frac{1}{m}\sum_{i=1}^m \mathbf{a}_i^T\mathbf{x} = \frac{1}{m}\mathbf{e}^T\mathbf{A}\mathbf{x}.$$

Using the latter expression for y we obtain that

$$\begin{aligned}
\sum_{i=1}^m \left(\mathbf{a}_i^T\mathbf{x} - y\right)^2 &= \sum_{i=1}^m \left(\mathbf{a}_i^T\mathbf{x} - \frac{1}{m}\mathbf{e}^T\mathbf{A}\mathbf{x}\right)^2 \\
&= \sum_{i=1}^m (\mathbf{a}_i^T\mathbf{x})^2 - \frac{2}{m}\sum_{i=1}^m (\mathbf{e}^T\mathbf{A}\mathbf{x})(\mathbf{a}_i^T\mathbf{x}) + \frac{1}{m}(\mathbf{e}^T\mathbf{A}\mathbf{x})^2 \\
&= \sum_{i=1}^m (\mathbf{a}_i^T\mathbf{x})^2 - \frac{1}{m}(\mathbf{e}^T\mathbf{A}\mathbf{x})^2 = \|\mathbf{A}\mathbf{x}\|^2 - \frac{1}{m}(\mathbf{e}^T\mathbf{A}\mathbf{x})^2 \\
&= \mathbf{x}^T\mathbf{A}^T\left(\mathbf{I}_m - \frac{1}{m}\mathbf{e}\mathbf{e}^T\right)\mathbf{A}\mathbf{x}.
\end{aligned}$$

Therefore, we arrive at the following reformulation of (10.17) as a problem consisting of minimizing a Rayleigh quotient:

$$\min_{\mathbf{x}}\left\{\frac{\mathbf{x}^T[\mathbf{A}^T(\mathbf{I}_m - \frac{1}{m}\mathbf{e}\mathbf{e}^T)\mathbf{A}]\mathbf{x}}{\|\mathbf{x}\|^2} : \mathbf{x} \neq \mathbf{0}\right\}.$$

Therefore, by Lemma 1.12 an optimal solution of the problem is an eigenvector of the matrix $\mathbf{A}^T(\mathbf{I}_m - \frac{1}{m}\mathbf{e}\mathbf{e}^T)\mathbf{A}$ corresponding to the minimum eigenvalue; the optimal function value is the minimum eigenvalue $\lambda_{\min}[\mathbf{A}^T(\mathbf{I}_m - \frac{1}{m}\mathbf{e}\mathbf{e}^T)\mathbf{A}]$. $\square$

Exercises

10.1. Show that the dual cone of

$$M = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} \geq \mathbf{0}\} \quad (\mathbf{A} \in \mathbb{R}^{m\times n})$$

is

$$M^* = \{\mathbf{A}^T\mathbf{v} : \mathbf{v} \in \mathbb{R}^m_+\}.$$

10.2. **(nonhomogenous Farkas' lemma)** Let $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{c} \in \mathbb{R}^n, \mathbf{b} \in \mathbb{R}^m$, and $d \in \mathbb{R}$. Suppose that there exists $\mathbf{y} \geq \mathbf{0}$ such that $\mathbf{A}^T\mathbf{y} = \mathbf{c}$. Prove that exactly one of the following two systems is feasible:

A. $\mathbf{A}\mathbf{x} \leq \mathbf{b}, \mathbf{c}^T\mathbf{x} > d$.

B. $\mathbf{A}^T\mathbf{y} = \mathbf{c}, \mathbf{b}^T\mathbf{y} \leq d, \mathbf{y} \geq \mathbf{0}$.

10.3. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{c} \in \mathbb{R}^n$. Show that exactly one of the following two systems is feasible:

A. $\mathbf{A}\mathbf{x} \geq \mathbf{0}, \mathbf{x} \geq \mathbf{0}, \mathbf{c}^T\mathbf{x} > 0$.

B. $\mathbf{A}^T\mathbf{y} \geq \mathbf{c}, \mathbf{y} \leq \mathbf{0}$.

10.4. Prove Motzkin's theorem of the alternative: the system

$$\text{(I)} \quad \begin{array}{rcl} \mathbf{A}\mathbf{d} & < & \mathbf{0}, \\ \mathbf{B}\mathbf{d} & \leq & \mathbf{0} \end{array}$$

has a solution if and only if the system

$$\text{(II)} \quad \begin{array}{l} \mathbf{A}^T\mathbf{u} + \mathbf{B}^T\mathbf{y} = \mathbf{0}, \\ \mathbf{u}, \mathbf{y} \geq \mathbf{0}, \qquad \mathbf{u} \neq \mathbf{0} \end{array}$$

does not have a solution (here $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{B} \in \mathbb{R}^{k\times n}$).

10.5. Prove the following nonhomogenous version of Gordan's alternative theorem: Given $\mathbf{A} \in \mathbb{R}^{m\times n}$, exactly one of these two systems is feasible.

A. $\mathbf{A}\mathbf{z} < \mathbf{b}$.

B. $\mathbf{A}^T\mathbf{y} = \mathbf{0}, \mathbf{b}^T\mathbf{y} \leq 0, \mathbf{y} \geq \mathbf{0}, \mathbf{y} \neq \mathbf{0}$.

10.6. Consider the maximization problem

$$\begin{array}{ll} \max & x_1^2 + 2x_1x_2 + 2x_2^2 - 3x_1 + x_2 \\ \text{s.t.} & x_1 + x_2 = 1 \\ & x_1, x_2 \geq 0. \end{array}$$

(i) Is the problem convex?

(ii) Find all the KKT points of the problem.

(iii) Find the optimal solution of the problem.

10.7. Consider the problem

$$\begin{array}{ll} \min & -x_1x_2x_3 \\ \text{s.t.} & x_1+3x_2+6x_3 \le 48, \\ & x_1, x_2, x_3 \ge 0. \end{array}$$

(i) Write the KKT conditions for the problem.

(ii) Find the optimal solution of the problem.

10.8. Consider the problem

$$\begin{array}{ll} \min & x_1^2+2x_2^2+x_1 \\ \text{s.t.} & x_1+x_2 \le a, \end{array}$$

where $a \in \mathbb{R}$ is a parameter.

(i) Prove that for any $a \in \mathbb{R}$, the problem has a unique optimal solution (without actually solving it).

(ii) Solve the problem (the solution will be in terms of the parameter a).

(iii) Let $f(a)$ be the optimal value of the problem with parameter a. Write an explicit expression for f and prove that it is a convex function.

10.9. Consider the problem

$$\begin{array}{ll} \min & x_1^2+x_2^2+x_3^2+x_1x_2+x_2x_3-2x_1-4x_2-6x_3 \\ \text{s.t.} & x_1+x_2+x_3 \le 1. \end{array}$$

(i) Is the problem convex?

(ii) Find all the KKT points of the problem.

(iii) Find the optimal solution of the problem.

10.10. Consider the problem

$$\begin{array}{ll} \min & x_1^2+x_2^2+x_3^2 \\ \text{s.t.} & x_1+2x_2+3x_3 \ge 4 \\ & x_3 \le 1. \end{array}$$

(i) Write down the KKT conditions.

(ii) Without solving the KKT system, prove that the problem has a unique optimal solution and that this solution satisfies the KKT conditions.

(iii) Find the optimal solution of the problem using the KKT system.

10.11. Use the KKT conditions in order to solve the problem

$$\begin{array}{ll} \min & x_1^2+x_2^2 \\ \text{s.t.} & -2x_1-x_2+10 \le 0 \\ & x_2 \ge 0. \end{array}$$