Chapter 2

Two-Stage Problems

Andrzej Ruszczyński and Alexander Shapiro

2.1 Linear Two-Stage Problems

2.1.1 Basic Properties

In this section we discuss two-stage stochastic linear programming problems of the form

$$\begin{aligned} \operatorname*{Min}_{x \in \mathbb{R}^n} \ & c^{\mathsf{T}} x + \mathbb{E}[Q(x, \xi)] \\ \text{s.t.} \ & Ax = b, \ x \geq 0, \end{aligned} \tag{2.1}$$

where $Q(x, \xi)$ is the optimal value of the second-stage problem

$$\begin{aligned} \operatorname*{Min}_{y \in \mathbb{R}^m} \ & q^{\mathsf{T}} y \\ \text{s.t.} \ & Tx + Wy = h, \ y \geq 0. \end{aligned} \tag{2.2}$$

Here $\xi := (q, h, T, W)$ are the data of the second-stage problem. We view some or all elements of vector ξ as random, and the expectation operator at the first-stage problem (2.1) is taken with respect to the probability distribution of ξ. Often, we use the same notation ξ to denote a random vector and its particular realization. Which of these two meanings will be used in a particular situation will usually be clear from the context. If there is doubt, then we write $\xi = \xi(\omega)$ to emphasize that ξ is a random vector defined on a corresponding probability space. We denote by $\Xi \subset \mathbb{R}^d$ the support of the probability distribution of ξ.

If for some x and $\xi \in \Xi$ the second-stage problem (2.2) is infeasible, then by definition $Q(x, \xi) = +\infty$. It could also happen that the second-stage problem is unbounded from below and hence $Q(x, \xi) = -\infty$. This is somewhat pathological situation, meaning that for some value of the first-stage decision vector and a realization of the random data, the value of the second-stage problem can be improved indefinitely. Models exhibiting such properties should be avoided. (We discuss this later.)

The second-stage problem (2.2) is a linear programming problem. Its dual problem can be written in the form

$$\begin{aligned} &\operatorname*{Max}_{\pi} \ \pi^{\mathsf{T}}(h - Tx) \\ &\text{s.t. } W^{\mathsf{T}}\pi \le q. \end{aligned} \tag{2.3}$$

By the theory of linear programming, the optimal values of problems (2.2) and (2.3) are equal to each other, unless both problems are infeasible. Moreover, if their common optimal value is finite, then each problem has a nonempty set of optimal solutions.

Consider the function

$$s_q(\chi) := \inf\left\{q^{\mathsf{T}} y : Wy = \chi,\ y \ge 0\right\}. \tag{2.4}$$

Clearly, $Q(x, \xi) = s_q(h - Tx)$. By the duality theory of linear programming, if the set

$$\Pi(q) := \left\{\pi : W^{\mathsf{T}}\pi \le q\right\} \tag{2.5}$$

is nonempty, then

$$s_q(\chi) = \sup_{\pi \in \Pi(q)} \pi^{\mathsf{T}}\chi, \tag{2.6}$$

i.e., $s_q(\cdot)$ is the support function of the set $\Pi(q)$. The set $\Pi(q)$ is convex, closed, and polyhedral. Hence, it has a finite number of extreme points. (If, moreover, $\Pi(q)$ is bounded, then it coincides with the convex hull of its extreme points.) It follows that if $\Pi(q)$ is nonempty, then $s_q(\cdot)$ is a positively homogeneous *polyhedral* function. If the set $\Pi(q)$ is empty, then the infimum on the right-hand side of (2.4) may take only two values: $+\infty$ or $-\infty$. In any case it is not difficult to verify directly that the function $s_q(\cdot)$ is convex.

Proposition 2.1. *For any given ξ, the function $Q(\cdot, \xi)$ is convex. Moreover, if the set $\{\pi : W^{\mathsf{T}}\pi \le q\}$ is nonempty and problem* (2.2) *is feasible for at least one x, then the function $Q(\cdot, \xi)$ is polyhedral.*

Proof. Since $Q(x, \xi) = s_q(h - Tx)$, the above properties of $Q(\cdot, \xi)$ follow from the corresponding properties of the function $s_q(\cdot)$. $\square$

Differentiability properties of the function $Q(\cdot, \xi)$ can be described as follows.

Proposition 2.2. *Suppose that for given $x = x_0$ and $\xi \in \Xi$, the value $Q(x_0, \xi)$ is finite. Then $Q(\cdot, \xi)$ is subdifferentiable at x_0 and*

$$\partial Q(x_0, \xi) = -T^{\mathsf{T}}\mathfrak{D}(x_0, \xi), \tag{2.7}$$

where

$$\mathfrak{D}(x, \xi) := \arg\max_{\pi \in \Pi(q)} \pi^{\mathsf{T}}(h - Tx)$$

is the set of optimal solutions of the dual problem (2.3).

Proof. Since $Q(x_0, \xi)$ is finite, the set $\Pi(q)$ defined in (2.5) is nonempty, and hence $s_q(\chi)$ is its support function. It is straightforward to see from the definitions that the support function $s_q(\cdot)$ is the conjugate function of the indicator function

$$\mathbb{I}_q(\pi) := \begin{cases} 0 & \text{if } \pi \in \Pi(q), \\ +\infty & \text{otherwise.} \end{cases}$$

Since the set $\Pi(q)$ is convex and closed, the function $\mathbb{I}_q(\cdot)$ is convex and lower semicontinuous. It follows then by the Fenchel–Moreau theorem (Theorem 7.5) that the conjugate of $s_q(\cdot)$ is $\mathbb{I}_q(\cdot)$. Therefore, for $\chi_0 := h - Tx_0$, we have (see (7.24))

$$\partial s_q(\chi_0) = \arg\max_{\pi} \left\{\pi^{\mathsf{T}}\chi_0 - \mathbb{I}_q(\pi)\right\} = \arg\max_{\pi\in\Pi(q)} \pi^{\mathsf{T}}\chi_0. \tag{2.8}$$

Since the set $\Pi(q)$ is polyhedral and $s_q(\chi_0)$ is finite, it follows that $\partial s_q(\chi_0)$ is nonempty. Moreover, the function $s_0(\cdot)$ is piecewise linear, and hence formula (2.7) follows from (2.8) by the chain rule of subdifferentiation. □

It follows that if the function $Q(\cdot, \xi)$ has a finite value in at least one point, then it is subdifferentiable at that point and hence is proper. Its domain can be described in a more explicit way.

The *positive hull* of a matrix W is defined as

$$\text{pos}\, W := \{\chi : \chi = Wy,\ y \geq 0\}. \tag{2.9}$$

It is a convex polyhedral cone generated by the columns of W. Directly from the definition (2.4) we see that $\text{dom}\, s_q = \text{pos}\, W$. Therefore,

$$\text{dom}\, Q(\cdot, \xi) = \{x : h - Tx \in \text{pos}\, W\}.$$

Suppose that x is such that $\chi = h - Tx \in \text{pos}\, W$, and let us analyze formula (2.7). The recession cone of $\Pi(q)$ is equal to

$$\Pi_0 := \Pi(0) = \left\{\pi : W^{\mathsf{T}}\pi \leq 0\right\}. \tag{2.10}$$

Then it follows from (2.6) that $s_q(\chi)$ is finite iff $\pi^{\mathsf{T}}\chi \leq 0$ for every $\pi \in \Pi_0$, that is, iff χ is an element of the polar cone to Π_0. This polar cone is nothing else but $\text{pos}\, W$, i.e.,

$$\Pi_0^* = \text{pos}\, W. \tag{2.11}$$

If $\chi_0 \in \text{int}(\text{pos}\, W)$, then the set of maximizers in (2.6) must be bounded. Indeed, if it was unbounded, there would exist an element $\pi_0 \in \Pi_0$ such that $\pi_0^{\mathsf{T}}\chi_0 = 0$. By perturbing χ_0 a little to some χ, we would be able to keep χ within $\text{pos}\, W$ and get $\pi_0^{\mathsf{T}}\chi > 0$, which is a contradiction, because $\text{pos}\, W$ is the polar of Π_0. Therefore the set of maximizers in (2.6) is the convex hull of the vertices v of $\Pi(q)$ for which $v^{\mathsf{T}}\chi = s_q(\chi)$. Note that $\Pi(q)$ must have vertices in this case, because otherwise the polar to Π_0 would have no interior.

If χ_0 is a boundary point of $\text{pos}\, W$, then the set of maximizers in (2.6) is unbounded. Its recession cone is the intersection of the recession cone Π_0 of $\Pi(q)$ and of the subspace $\{\pi : \pi^{\mathsf{T}}\chi_0 = 0\}$. This intersection is nonempty for boundary points χ_0 and is equal to the normal cone to $\text{pos}\, W$ at χ_0. Indeed, let π_0 be normal to $\text{pos}\, W$ at χ_0. Since both χ_0 and $-\chi_0$ are feasible directions at χ_0, we must have $\pi_0^{\mathsf{T}}\chi_0 = 0$. Next, for every $\chi \in \text{pos}\, W$ we have $\pi_0^{\mathsf{T}}\chi = \pi_0^{\mathsf{T}}(\chi - \chi_0) \leq 0$, so $\pi_0 \in \Pi_0$. The converse argument is similar.

2.1.2 The Expected Recourse Cost for Discrete Distributions

Let us consider now the expected value function

$$\phi(x) := \mathbb{E}[Q(x,\xi)]. \tag{2.12}$$

As before, the expectation here is taken with respect to the probability distribution of the random vector ξ. Suppose that the distribution of ξ has finite support. That is, ξ has a finite number of realizations (called *scenarios*) $\xi_k = (q_k, h_k, T_k, W_k)$ with respective (positive) probabilities p_k, $k = 1, \dots, K$, i.e., $\Xi = \{\xi_1, \dots, \xi_K\}$. Then

$$\mathbb{E}[Q(x,\xi)] = \sum_{k=1}^{K} p_k Q(x, \xi_k). \tag{2.13}$$

For a given x, the expectation $\mathbb{E}[Q(x,\xi)]$ is equal to the optimal value of the linear programming problem

$$\begin{aligned} \operatorname*{Min}_{y_1,\dots,y_K} \ & \sum_{k=1}^{K} p_k q_k^{\mathsf{T}} y_k \\ \text{s.t. } & T_k x + W_k y_k = h_k, \\ & y_k \geq 0, \quad k = 1, \dots, K. \end{aligned} \tag{2.14}$$

If for at least one $k \in \{1, \dots, K\}$ the system $T_k x + W_k y_k = h_k$, $y_k \geq 0$, has no solution, i.e., the corresponding second-stage problem is infeasible, then problem (2.14) is infeasible, and hence its optimal value is $+\infty$. From that point of view, the sum in the right-hand side of (2.13) equals $+\infty$ if at least one of $Q(x, \xi_k) = +\infty$. That is, we assume here that $+\infty + (-\infty) = +\infty$.

The whole two stage-problem is equivalent to the following large-scale linear programming problem:

$$\begin{aligned} \operatorname*{Min}_{x,y_1,\dots,y_K} \ & c^{\mathsf{T}} x + \sum_{k=1}^{K} p_k q_k^{\mathsf{T}} y_k \\ \text{s.t. } & T_k x + W_k y_k = h_k, \quad k = 1, \dots, K, \\ & Ax = b, \\ & x \geq 0, \quad y_k \geq 0, \quad k = 1, \dots, K. \end{aligned} \tag{2.15}$$

Properties of the expected recourse cost follow directly from properties of parametric linear programming problems.

Proposition 2.3. *Suppose that the probability distribution of ξ has finite support $\Xi = \{\xi_1, \dots, \xi_K\}$ and that the expected recourse cost $\phi(\cdot)$ has a finite value in at least one point $\bar{x} \in \mathbb{R}^n$. Then the function $\phi(\cdot)$ is polyhedral, and for any $x_0 \in \operatorname{dom} \phi$,*

$$\partial \phi(x_0) = \sum_{k=1}^{K} p_k \partial Q(x_0, \xi_k). \tag{2.16}$$

Proof. Since $\phi(\bar{x})$ is finite, all values $Q(\bar{x}, \xi_k)$, $k = 1, \ldots, K$, are finite. Consequently, by Proposition 2.2, every function $Q(\cdot, \xi_k)$ is polyhedral. It is not difficult to see that a linear combination of polyhedral functions with positive weights is also polyhedral. Therefore, it follows that $\phi(\cdot)$ is polyhedral. We also have that $\operatorname{dom} \phi = \bigcap_{k=1}^{K} \operatorname{dom} Q_k$, where $Q_k(\cdot) := Q(\cdot, \xi_k)$, and for any $h \in \mathbb{R}^n$, the directional derivatives $Q'_k(x_0, h) > -\infty$ and

$$\phi'(x_0, h) = \sum_{k=1}^{K} p_k Q'_k(x_0, h). \tag{2.17}$$

Formula (2.16) then follows from (2.17) by duality arguments. Note that equation (2.16) is a particular case of the Moreau–Rockafellar theorem (Theorem 7.4). Since the functions Q_k are polyhedral, there is no need here for an additional regularity condition for (2.16) to hold true. □

The subdifferential $\partial Q(x_0, \xi_k)$ of the second-stage optimal value function is described in Proposition 2.2. That is, if $Q(x_0, \xi_k)$ is finite, then

$$\partial Q(x_0, \xi_k) = -T_k^{\mathsf{T}} \arg\max \left\{ \pi^{\mathsf{T}}(h_k - T_k x_0) : W_k^{\mathsf{T}} \pi \leq q_k \right\}. \tag{2.18}$$

It follows that the expectation function ϕ is differentiable at x_0 iff for every $\xi = \xi_k$, $k = 1, \ldots, K$, the maximum in the right-hand side of (2.18) is attained at a unique point, i.e., the corresponding second-stage dual problem has a unique optimal solution.

Example 2.4 (Capacity Expansion). We have a directed graph with node set $\mathcal{N}$ and arc set $\mathcal{A}$. With each arc $a \in \mathcal{A}$, we associate a decision variable x_a and call it the *capacity* of a. There is a cost c_a for each unit of capacity of arc a. The vector x constitutes the vector of first-stage variables. They are restricted to satisfy the inequalities $x \geq x^{\min}$, where $x^{\min}$ are the existing capacities.

At each node n of the graph, we have a random demand ξ_n for shipments to n. (If ξ_n is negative, its absolute value represents shipments from n and we have $\sum_{n \in \mathcal{N}} \xi_n = 0$.) These shipments have to be sent through the network, and they can be arbitrarily split into pieces taking different paths. We denote by y_a the amount of the shipment sent through arc a. There is a unit cost q_a for shipments on each arc a.

Our objective is to assign the arc capacities and to organize the shipments in such a way that the expected total cost, comprising the capacity cost and the shipping cost, is minimized. The condition is that the capacities have to be assigned *before* the actual demands ξ_n become known, while the shipments can be arranged *after* that.

Let us define the second-stage problem. For each node n, denote by $\mathcal{A}_+(n)$ and $\mathcal{A}_-(n)$ the sets of arcs entering and leaving node i. The second-stage problem is the network flow problem

$$\operatorname{Min} \sum_{a \in \mathcal{A}} q_a y_a \tag{2.19}$$

$$\text{s.t.} \sum_{a \in \mathcal{A}_+(n)} y_a - \sum_{a \in \mathcal{A}_-(n)} y_a = \xi_n, \quad n \in \mathcal{N}, \tag{2.20}$$

$$0 \leq y_a \leq x_a, \quad a \in \mathcal{A}. \tag{2.21}$$

This problem depends on the random demand vector ξ and on the arc capacities, x. Its optimal value is denoted by $Q(x, \xi)$.

Suppose that for a given $x = x_0$ the second-stage problem (2.19)–(2.21) is feasible. Denote by μ_n, $n \in \mathcal{N}$, the optimal Lagrange multipliers (node potentials) associated with the node balance equations (2.20), and denote by π_a, $a \in \mathcal{A}$, the (nonnegative) Lagrange multipliers associated with the constraints (2.21). The dual problem has the form

$$
\begin{aligned}
\text{Max}\ & -\sum_{n\in\mathcal{N}} \xi_n \mu_n - \sum_{(i,j)\in\mathcal{A}} x_{ij}\pi_{ij} \\
\text{s.t.}\ & -\pi_{ij} + \mu_i - \mu_j \le q_{ij}, \quad (i,j) \in \mathcal{A}, \\
& \pi \ge 0.
\end{aligned}
$$

As $\sum_{n\in\mathcal{N}} \xi_n = 0$, the values of μ_n can be translated by a constant without any change in the objective function, and thus without any loss of generality we can assume that $\mu_{n_0} = 0$ for some fixed node n_0. For each arc $a = (i, j)$, the multiplier π_{ij} associated with the constraint (2.21) has the form

$$
\pi_{ij} = \max\{0, \mu_i - \mu_j - q_{ij}\}.
$$

Roughly, if the difference of node potentials $\mu_i - \mu_j$ is greater than q_{ij}, the arc is saturated and the capacity constraint $y_{ij} \le x_{ij}$ becomes relevant. The dual problem becomes equivalent to

$$
\text{Max}\ -\sum_{n\in\mathcal{N}} \xi_n \mu_n - \sum_{(i,j)\in\mathcal{A}} x_{ij} \max\{0, \mu_i - \mu_j - q_{ij}\}. \tag{2.22}
$$

Let us denote by $\mathcal{M}(x_0, \xi)$ the set of optimal solutions of this problem satisfying the condition $\mu_{n_0} = 0$. Since $T^{\mathsf{T}} = [0\ \ -I]$ in this case, formula (2.18) provides the description of the subdifferential of $Q(\cdot, \xi)$ at x_0:

$$
\partial Q(x_0, \xi) = -\left\{ \left(\max\{0, \mu_i - \mu_j - q_{ij}\}\right)_{(i,j)\in\mathcal{A}} : \mu \in \mathcal{M}(x_0, \xi) \right\}.
$$

The first-stage problem has the form

$$
\underset{x \ge x^{\min}}{\text{Min}} \sum_{(i,j)\in\mathcal{A}} c_{ij} x_{ij} + \mathbb{E}[Q(x, \xi)]. \tag{2.23}
$$

If ξ has finitely many realizations ξ^k attained with probabilities p_k, $k = 1, \dots, K$, the subdifferential of the overall objective can be calculated by (2.16):

$$
\partial f(x_0) = c + \sum_{k=1}^{K} p_k \partial Q(x_0, \xi^k). \qquad \blacksquare
$$

2.1.3 The Expected Recourse Cost for General Distributions

Let us discuss now the case of a general distribution of the random vector $\xi \in \mathbb{R}^d$. The recourse cost $Q(\cdot, \cdot)$ is the minimum value of the integrand which is a random lower semicontinuous function (see section 7.2.3). Therefore, it follows by Theorem 7.37 that $Q(\cdot, \cdot)$

is measurable with respect to the Borel sigma algebra of $\mathbb{R}^n \times \mathbb{R}^d$. Also for every ξ the function $Q(\cdot, \xi)$ is lower semicontinuous. It follows that $Q(x, \xi)$ is a random lower semicontinuous function. Recall that in order to ensure that the expectation $\phi(x)$ is well defined, we have to verify two conditions:

(i) $Q(x, \cdot)$ is measurable (with respect to the Borel sigma algebra of $\mathbb{R}^d$);

(ii) either $\mathbb{E}[Q(x, \xi)_+]$ or $\mathbb{E}[(-Q(x, \xi))_+]$ is finite.

The function $Q(x, \cdot)$ is measurable as the optimal value of a linear programming problem. We only need to verify condition (ii). We describe below some important particular situations where this condition is satisfied.

The two-stage problem (2.1)–(2.2) is said to have *fixed* recourse if the matrix W is fixed (not random). Moreover, we say that the recourse is *complete* if the system $Wy = \chi$ and $y \geq 0$ has a solution for every χ. In other words, the positive hull of W is equal to the corresponding vector space. By duality arguments, the fixed recourse is complete iff the feasible set $\Pi(q)$ of the dual problem (2.3) is bounded (in particular, it may be empty) for every q. Then its recession cone, $\Pi_0 = \Pi(0)$, must contain only the point 0, provided that $\Pi(q)$ is nonempty. Therefore, another equivalent condition for complete recourse is that $\pi = 0$ is the only solution of the system $W^{\mathsf{T}}\pi \leq 0$.

A particular class of problems with fixed and complete recourse are *simple recourse* problems, in which $W = [I; -I]$, the matrix T and the vector q are deterministic, and the components of q are positive.

It is said that the recourse is *relatively complete* if for every x in the set

$$X = \{x : Ax = b,\ x \geq 0\},$$

the feasible set of the second-stage problem (2.2) is nonempty for almost everywhere (a.e.) $\omega \in \Omega$. That is, the recourse is relatively complete if for every feasible first-stage point x the inequality $Q(x, \xi) < +\infty$ holds true for a.e. $\xi \in \Xi$, or in other words, $Q(x, \xi(\omega)) < +\infty$ w.p. 1. This definition is in accordance with the general principle that an event which happens with zero probability is irrelevant for the calculation of the corresponding expected value. For example, the capacity expansion problem of Example 2.4 is *not* a problem with relatively complete recourse, unless $x^{\min}$ is so large that every demand $\xi \in \Xi$ can be shipped over the network with capacities $x^{\min}$.

The following condition is sufficient for relatively complete recourse:

$$\text{for every } x \in X \text{ the inequality } Q(x, \xi) < +\infty \text{ holds true for all } \xi \in \Xi. \tag{2.24}$$

In general, condition (2.24) is not necessary for relatively complete recourse. It becomes necessary and sufficient in the following two cases:

(i) the random vector ξ has a finite support, or

(ii) the recourse is fixed.

Indeed, sufficiency is clear. If ξ has a finite support, i.e., the set Ξ is finite, then the necessity is also clear. To show the necessity in the case of fixed recourse, suppose the recourse is relatively complete. This means that if $x \in X$, then $Q(x, \xi) < +\infty$ for all ξ in Ξ, except possibly for a subset of Ξ of probability zero. We have that $Q(x, \xi) < +\infty$ iff

$h - Tx \in \text{pos}\, W$. Let $\Xi_0(x) = \{(h, T, q) : h - Tx \in \text{pos}\, W\}$. The set pos W is convex and closed and thus $\Xi_0(x)$ is convex and closed as well. By assumption, $P[\Xi_0(x)] = 1$ for every $x \in X$. Thus $\bigcap_{x \in X} \Xi_0(x)$ is convex, closed, and has probability 1. The support of ξ must be its subset.

Example 2.5. Consider

$$Q(x, \xi) := \inf\{y : \xi y = x,\ \ y \geq 0\}$$

with $x \in [0, 1]$ and ξ being a random variable whose probability density function is $p(z) := 2z$, $0 \leq z \leq 1$. For all $\xi > 0$ and $x \in [0, 1]$, $Q(x, \xi) = x/\xi$, and hence

$$\mathbb{E}[Q(x, \xi)] = \int_0^1 \left(\frac{x}{z}\right) 2z dz = 2x.$$

That is, the recourse here is relatively complete and the expectation of $Q(x, \xi)$ is finite. On the other hand, the support of $\xi(\omega)$ is the interval [0, 1], and for $\xi = 0$ and $x > 0$ the value of $Q(x, \xi)$ is $+\infty$, because the corresponding problem is infeasible. Of course, probability of the event "$\xi = 0$" is zero, and from the mathematical point of view the expected value function $\mathbb{E}[Q(x, \xi)]$ is well defined and finite for all $x \in [0, 1]$. Note, however, that arbitrary small perturbation of the probability distribution of ξ may change that. Take, for example, some discretization of the distribution of ξ with the first discretization point $t = 0$. Then, no matter how small the assigned (positive) probability at $t = 0$ is, $Q(x, \xi) = +\infty$ with positive probability. Therefore, $\mathbb{E}[Q(x, \xi)] = +\infty$ for all $x > 0$. That is, the above problem is extremely unstable and is not well posed. As discussed above, such behavior cannot occur if the recourse is fixed. ■

Let us consider the support function $s_q(\cdot)$ of the set $\Pi(q)$. We want to find sufficient conditions for the existence of the expectation $\mathbb{E}[s_q(h - Tx)]$. By Hoffman's lemma (Theorem 7.11), there exists a constant κ, depending on W, such that if for some q_0 the set $\Pi(q_0)$ is nonempty, then for every q the following inclusion is satisfied:

$$\Pi(q) \subset \Pi(q_0) + \kappa \|q - q_0\| B, \tag{2.25}$$

where $B := \{\pi : \|\pi\| \leq 1\}$ and $\|\cdot\|$ denotes the Euclidean norm. This inclusion allows us to derive an upper bound for the support function $s_q(\cdot)$. Since the support function of the unit ball B is the norm $\|\cdot\|$, it follows from (2.25) that if the set $\Pi(q_0)$ is nonempty, then

$$s_q(\cdot) \leq s_{q_0}(\cdot) + \kappa \|q - q_0\|\, \|\cdot\|. \tag{2.26}$$

Consider $q_0 = 0$. The support function $s_0(\cdot)$ of the cone Π_0 has the form

$$s_0(\chi) = \begin{cases} 0 & \text{if } \chi \in \text{pos}\, W, \\ +\infty & \text{otherwise.} \end{cases}$$

Therefore, (2.26) with $q_0 = 0$ implies that if $\Pi(q)$ is nonempty, then $s_q(\chi) \leq \kappa \|q\|\, \|\chi\|$ for all $\chi \in \text{pos}\, W$, and $s_q(\chi) = +\infty$ for all $\chi \notin \text{pos}\, W$. Since $\Pi(q)$ is polyhedral, if it is nonempty, then $s_q(\cdot)$ is piecewise linear on its domain, which coincides with pos W, and

$$\left| s_q(\chi_1) - s_q(\chi_2) \right| \leq \kappa \|q\|\, \|\chi_1 - \chi_2\|, \quad \forall \chi_1, \chi_2 \in \text{pos}\, W. \tag{2.27}$$

Proposition 2.6. *Suppose that the recourse is fixed and*

$$\mathbb{E}\big[\|q\|\,\|h\|\big] < +\infty \;\; \textit{and} \;\; \mathbb{E}\big[\|q\|\,\|T\|\big] < +\infty. \tag{2.28}$$

Consider a point $x \in \mathbb{R}^n$. Then $\mathbb{E}[Q(x,\xi)_+]$ is finite iff the following condition holds w.p. 1:

$$h - Tx \in \operatorname{pos} W. \tag{2.29}$$

Proof. We have that $Q(x,\xi) < +\infty$ iff condition (2.29) holds. Therefore, if condition (2.29) does not hold w.p. 1, then $Q(x,\xi) = +\infty$ with positive probability, and hence $\mathbb{E}[Q(x,\xi)_+] = +\infty$.

Conversely, suppose that condition (2.29) holds w.p. 1. Then $Q(x,\xi) = s_q(h - Tx)$ with $s_q(\cdot)$ being the support function of the set $\Pi(q)$. By (2.26) there exists a constant κ such that for any χ,

$$s_q(\chi) \le s_0(\chi) + \kappa \|q\|\,\|\chi\|.$$

Also for any $\chi \in \operatorname{pos} W$ we have that $s_0(\chi) = 0$, and hence w.p. 1,

$$s_q(h - Tx) \le \kappa \|q\|\,\|h - Tx\| \le \kappa \|q\| \big(\|h\| + \|T\|\,\|x\|\big).$$

It follows then by (2.28) that $\mathbb{E}\left[s_q(h - Tx)_+\right] < +\infty$. $\square$

Remark 2. If q and (h, T) are independent and have finite first moments,[4] then

$$\mathbb{E}\big[\|q\|\,\|h\|\big] = \mathbb{E}\big[\|q\|\big]\,\mathbb{E}\big[\|h\|\big] \;\; \text{and} \;\; \mathbb{E}\big[\|q\|\,\|T\|\big] = \mathbb{E}\big[\|q\|\big]\,\mathbb{E}\big[\|T\|\big],$$

and hence condition (2.28) follows. Also condition (2.28) holds if (h, T, q) has finite second moments.

We obtain that, under the assumptions of Proposition 2.6, the expectation $\phi(x)$ is well defined and $\phi(x) < +\infty$ iff condition (2.29) holds w.p. 1. If, moreover, the recourse is complete, then (2.29) holds for any x and ξ, and hence $\phi(\cdot)$ is well defined and is less than $+\infty$. Since the function $\phi(\cdot)$ is convex, we have that if $\phi(\cdot)$ is less than $+\infty$ on $\mathbb{R}^n$ and is finite valued in at least one point, then $\phi(\cdot)$ is finite valued on the entire space $\mathbb{R}^n$.

Proposition 2.7. *Suppose that* (i) *the recourse is fixed,* (ii) *for a.e. q the set $\Pi(q)$ is nonempty, and* (iii) *condition* (2.28) *holds.*

Then the expectation function $\phi(x)$ is well defined and $\phi(x) > -\infty$ for all $x \in \mathbb{R}^n$. Moreover, ϕ is convex, lower semicontinuous and Lipschitz continuous on $\operatorname{dom}\phi$, *and its domain is a convex closed subset of $\mathbb{R}^n$ given by*

$$\operatorname{dom}\phi = \left\{x \in \mathbb{R}^n : h - Tx \in \operatorname{pos} W \text{ w.p.1}\right\}. \tag{2.30}$$

Proof. By assumption (ii), the feasible set $\Pi(q)$ of the dual problem is nonempty w.p. 1. Thus $Q(x,\xi)$ is equal to $s_q(h - Tx)$ w.p. 1 for every x, where $s_q(\cdot)$ is the support function of the set $\Pi(q)$. Let $\pi(q)$ be the element of the set $\Pi(q)$ that is closest to 0. It exists

[4] We say that a random variable $Z = Z(\omega)$ has a finite rth moment if $\mathbb{E}\left[|Z|^r\right] < +\infty$. It is said that $\xi(\omega)$ has finite rth moments if each component of $\xi(\omega)$ has a finite rth moment.

because $\Pi(q)$ is closed. By Hoffman's lemma (see (2.25)) there is a constant κ such that $\|\pi(q)\| \leq \kappa\|q\|$. Then for every x the following holds w.p. 1:

$$s_q(h - Tx) \geq \pi(q)^{\mathsf{T}}(h - Tx) \geq -\kappa\|q\|\big(\|h\| + \|T\|\,\|x\|\big). \tag{2.31}$$

Owing to condition (2.28), it follows from (2.31) that $\phi(\cdot)$ is well defined and $\phi(x) > -\infty$ for all $x \in \mathbb{R}^n$. Moreover, since $s_q(\cdot)$ is lower semicontinuous, the lower semicontinuity of $\phi(\cdot)$ follows by Fatou's lemma. Convexity and closedness of $\operatorname{dom}\phi$ follow from the convexity and lower semicontinuity of ϕ. We have by Proposition 2.6 that $\phi(x) < +\infty$ iff condition (2.29) holds w.p. 1. This implies (2.30).

Consider two points $x, x' \in \operatorname{dom}\phi$. Then by (2.30) the following holds true w.p. 1:

$$h - Tx \in \operatorname{pos} W \quad\text{and}\quad h - Tx' \in \operatorname{pos} W. \tag{2.32}$$

By (2.27), if the set $\Pi(q)$ is nonempty and (2.32) holds, then

$$\big|s_q(h - Tx) - s_q(h - Tx')\big| \leq \kappa\|q\|\,\|T\|\,\|x - x'\|.$$

It follows that

$$|\phi(x) - \phi(x')| \leq \kappa\,\mathbb{E}\big[\|q\|\,\|T\|\big]\,\|x - x'\|.$$

With condition (2.28) this implies the Lipschitz continuity of ϕ on its domain. ∎

Denote by Σ the support[5] of the probability distribution (measure) of (h, T). Formula (2.30) means that a point x belongs to $\operatorname{dom}\phi$ iff the probability of the event $\{h - Tx \in \operatorname{pos} W\}$ is one. Note that the set $\{(h, T) : h - Tx \in \operatorname{pos} W\}$ is convex and polyhedral and hence is closed. Consequently x belongs to $\operatorname{dom}\phi$ iff for every $(h, T) \in \Sigma$ it follows that $h - Tx \in \operatorname{pos} W$. Therefore, we can write formula (2.30) in the form

$$\operatorname{dom}\phi = \bigcap_{(h,T)\in\Sigma} \{x : h - Tx \in \operatorname{pos} W\}. \tag{2.33}$$

It should be noted that we assume that the recourse is fixed.

Let us observe that for any set $\mathcal{H}$ of vectors h, the set $\cap_{h\in\mathcal{H}}(-h + \operatorname{pos} W)$ is convex and polyhedral. Indeed, we have that $\operatorname{pos} W$ is a convex polyhedral cone and hence can be represented as the intersection of a finite number of half spaces $A_i = \{\chi : a_i^{\mathsf{T}}\chi \leq 0\}$, $i = 1, \dots, \ell$. Since the intersection of any number of half spaces of the form $b + A_i$, with $b \in B$, is still a half space of the same form (provided that this intersection is nonempty), we have that the set $\cap_{h\in\mathcal{H}}(-h + \operatorname{pos} W)$ can be represented as the intersection of half spaces of the form $b_i + A_i$, $i = 1, \dots, \ell$, and hence is polyhedral. It follows that if T and W are fixed, then the set at the right-hand side of (2.33) is convex and polyhedral.

Let us discuss now the differentiability properties of the expectation function $\phi(x)$. By Theorem 7.47 and formula (2.7) of Proposition 2.2 we have the following result.

[5]Recall that the support of the probability measure is the smallest closed set such that the probability (measure) of its complement is zero.

Proposition 2.8. *Suppose that the expectation function $\phi(\cdot)$ is proper and its domain has a nonempty interior. Then for any $x_0 \in \operatorname{dom}\phi$,*

$$\partial\phi(x_0) = -\mathbb{E}\left[T^{\mathsf T}\mathfrak{D}(x_0,\xi)\right] + \mathcal{N}_{\operatorname{dom}\phi}(x_0), \tag{2.34}$$

where

$$\mathfrak{D}(x,\xi) := \arg\max_{\pi\in\Pi(q)} \pi^{\mathsf T}(h - Tx).$$

Moreover, ϕ is differentiable at x_0 iff x_0 belongs to the interior of $\operatorname{dom}\phi$ *and the set $\mathfrak{D}(x_0,\xi)$ is a singleton w.p.* 1.

As discussed earlier, when the distribution of ξ has a finite support (i.e., there is a finite number of scenarios), the expectation function ϕ is piecewise linear on its domain and is differentiable everywhere only in the trivial case if it is linear.[6] In the case of a continuous distribution of ξ, the expectation operator smoothes the piecewise linear function $Q(\cdot,\xi)$.

Proposition 2.9. *Suppose the assumptions of Proposition* 2.7 *are satisfied and the conditional distribution of h, given (T,q), is absolutely continuous for almost all (T,q). Then ϕ is continuously differentiable on the interior of its domain.*

Proof. By Proposition 2.7, the expectation function $\phi(\cdot)$ is well defined and greater than $-\infty$. Let x be a point in the interior of $\operatorname{dom}\phi$. For fixed T and q, consider the multifunction

$$\mathfrak{Z}(h) := \arg\max_{\pi\in\Pi(q)} \pi^{\mathsf T}(h - Tx).$$

Conditional on (T,q), the set $\mathfrak{D}(x,\xi)$ coincides with $\mathfrak{Z}(h)$. Since $x \in \operatorname{dom}\phi$, relation (2.30) implies that $h - Tx \in \operatorname{pos} W$ w.p. 1. For every $h - Tx \in \operatorname{pos} W$, the set $\mathfrak{Z}(h)$ is nonempty and forms a face of the polyhedral set $\Pi(q)$. Moreover, there exists a set A given by the union of a finite number of linear subspaces of $\mathbb{R}^m$ (where m is the dimension of h), which are perpendicular to the faces of sets $\Pi(q)$, such that if $h - Tx \in (\operatorname{pos} W)\setminus A$, then $\mathfrak{Z}(h)$ is a singleton. Since an affine subspace of $\mathbb{R}^m$ has Lebesgue measure zero, it follows that the Lebesgue measure of A is zero. As the conditional distribution of h, given (T,q), is absolutely continuous, the probability that $\mathfrak{Z}(h)$ is not a singleton is zero. By integrating this probability over the marginal distribution of (T,q), we obtain that probability of the event "$\mathfrak{D}(x,\xi)$ is not a singleton" is zero. By Proposition 2.8, this implies the differentiability of $\phi(\cdot)$. Since $\phi(\cdot)$ is convex, it follows that for every $x \in \operatorname{int}(\operatorname{dom}\phi)$ the gradient $\nabla\phi(x)$ coincides with the (unique) subgradient of ϕ at x and that $\nabla\phi(\cdot)$ is continuous at x. □

Of course, if h and (T,q) are independent, then the conditional distribution of h given (T,q) is the same as the unconditional (marginal) distribution of h. Therefore, if h and (T,q) are independent, then it suffices to assume in the above proposition that the (marginal) distribution of h is absolutely continuous.

[6] By linear, we mean here that it is of the form $a^{\mathsf T}x + b$. It is more accurate to call such a function affine.

2.1.4 Optimality Conditions

We can now formulate optimality conditions and duality relations for linear two-stage problems. Let us start from the problem with discrete distributions of the random data in (2.1)–(2.2). The problem takes on the form

$$\begin{aligned} &\operatorname*{Min}_{x}\ c^{\mathsf T}x + \sum_{k=1}^{K} p_k Q(x,\xi_k) \\ &\text{s.t. } Ax = b,\ x \ge 0, \end{aligned} \tag{2.35}$$

where $Q(x,\xi)$ is the optimal value of the second-stage problem, given by (2.2).

Suppose the expectation function $\phi(\cdot) := \mathbb{E}[Q(\cdot,\xi)]$ has a finite value in at least one point $\bar{x} \in \mathbb{R}^n$. It follows from Propositions 2.2 and 2.3 that for every $x_0 \in \operatorname{dom}\phi$,

$$\partial\phi(x_0) = -\sum_{k=1}^{K} p_k T_k^{\mathsf T}\mathfrak{D}(x_0,\xi_k), \tag{2.36}$$

where

$$\mathfrak{D}(x_0,\xi_k) := \arg\max\left\{\pi^{\mathsf T}(h_k - T_k x_0) : W_k^{\mathsf T}\pi \le q_k\right\}.$$

As before, we denote $X := \{x : Ax = b,\ x \ge 0\}$.

Theorem 2.10. *Let $\bar{x}$ be a feasible solution of problem* (2.1)–(2.2)*, i.e., $\bar{x} \in X$ and $\phi(\bar{x})$ is finite. Then $\bar{x}$ is an optimal solution of problem* (2.1)–(2.2) *iff there exist $\pi_k \in \mathfrak{D}(\bar{x},\xi_k)$, $k = 1,\dots,K$, and $\mu \in \mathbb{R}^m$ such that*

$$\begin{aligned} &\sum_{k=1}^{K} p_k T_k^{\mathsf T}\pi_k + A^{\mathsf T}\mu \le c, \\ &\bar{x}^{\mathsf T}\left(c - \sum_{k=1}^{K} p_k T_k^{\mathsf T}\pi_k - A^{\mathsf T}\mu\right) = 0. \end{aligned} \tag{2.37}$$

Proof. Necessary and sufficient optimality conditions for minimizing $c^{\mathsf T}x + \phi(x)$ over $x \in X$ can be written as

$$0 \in c + \partial\phi(\bar{x}) + \mathcal{N}_X(\bar{x}), \tag{2.38}$$

where $\mathcal{N}_X(\bar{x})$ is the normal cone to the feasible set X. Note that condition (2.38) implies that the sets $\mathcal{N}_X(\bar{x})$ and $\partial\phi(\bar{x})$ are nonempty and hence $\bar{x} \in X$ and $\phi(\bar{x})$ is finite. Note also that there is no need here for additional regularity conditions since $\phi(\cdot)$ and X are convex and polyhedral. Using the characterization of the subgradients of $\phi(\cdot)$, given in (2.36), we conclude that (2.38) is equivalent to existence of $\pi_k \in \mathfrak{D}(\bar{x},\xi_k)$ such that

$$0 \in c - \sum_{k=1}^{K} p_k T_k^{\mathsf T}\pi_k + \mathcal{N}_X(\bar{x}).$$

Observe that

$$\mathcal{N}_X(\bar{x}) = \{A^{\mathsf{T}}\mu - h : h \geq 0,\ h^{\mathsf{T}}\bar{x} = 0\}. \tag{2.39}$$

The last two relations are equivalent to conditions (2.37). $\square$

Conditions (2.37) can also be obtained directly from the optimality conditions for the large-scale linear programming formulation

$$\begin{aligned} \operatorname*{Min}_{x, y_1, \dots, y_K} \ & c^{\mathsf{T}}x + \sum_{k=1}^{K} p_k q_k^{\mathsf{T}} y_k \\ \text{s.t. } & T_k x + W_k y_k = h_k, \quad k = 1, \dots, K, \\ & Ax = b, \\ & x \geq 0, \\ & y_k \geq 0, \quad k = 1, \dots, K. \end{aligned} \tag{2.40}$$

By minimizing, with respect to $x \geq 0$ and $y_k \geq 0$, $k = 1, \dots, K$, the Lagrangian

$$\begin{aligned} & c^{\mathsf{T}}x + \sum_{k=1}^{K} p_k q_k^{\mathsf{T}} y_k - \mu^{\mathsf{T}}(Ax - b) - \sum_{k=1}^{K} p_k \pi_k^{\mathsf{T}}(T_k x + W_k y_k - h_k) \\ & = \left(c - A^{\mathsf{T}}\mu - \sum_{k=1}^{K} p_k T_k^{\mathsf{T}}\pi_k\right)^{\mathsf{T}} x + \sum_{k=1}^{K} p_k \left(q_k - W_k^{\mathsf{T}}\pi_k\right)^{\mathsf{T}} y_k + b^{\mathsf{T}}\mu + \sum_{k=1}^{K} p_k h_k^{\mathsf{T}}\pi_k, \end{aligned}$$

we obtain the following dual of the linear programming problem (2.40):

$$\begin{aligned} \operatorname*{Max}_{\mu, \pi_1, \dots, \pi_K} \ & b^{\mathsf{T}}\mu + \sum_{k=1}^{K} p_k h_k^{\mathsf{T}}\pi_k \\ \text{s.t. } & c - A^{\mathsf{T}}\mu - \sum_{k=1}^{K} p_k T_k^{\mathsf{T}}\pi_k \geq 0, \\ & q_k - W_k^{\mathsf{T}}\pi_k \geq 0,\ k = 1, \dots, K. \end{aligned}$$

Therefore, optimality conditions of Theorem 2.10 can be written in the following equivalent form:

$$\begin{aligned} & \sum_{k=1}^{K} p_k T_k^{\mathsf{T}}\pi_k + A^{\mathsf{T}}\mu \leq c, \\ & \bar{x}^{\mathsf{T}}\left(c - \sum_{k=1}^{K} p_k T_k^{\mathsf{T}}\pi_k - A^{\mathsf{T}}\mu\right) = 0, \\ & q_k - W_k^{\mathsf{T}}\pi_k \geq 0,\ k = 1, \dots, K, \\ & \bar{y}_k^{\mathsf{T}}\left(q_k - W_k^{\mathsf{T}}\pi_k\right) = 0,\ k = 1, \dots, K. \end{aligned}$$

The last two of the above conditions correspond to feasibility and optimality of multipliers π_k as solutions of the dual problems.

If we deal with general distributions of the problem's data, additional conditions are needed to ensure the subdifferentiability of the expected recourse cost and the existence of Lagrange multipliers.

Theorem 2.11. *Let $\bar{x}$ be a feasible solution of problem* (2.1)–(2.2). *Suppose that the expected recourse cost function $\phi(\cdot)$ is proper,* $\operatorname{int}(\operatorname{dom}\phi) \cap X$ *is nonempty, and $\mathcal{N}_{\operatorname{dom}\phi}(\bar{x}) \subset \mathcal{N}_X(\bar{x})$. Then $\bar{x}$ is an optimal solution of problem* (2.1)–(2.2) *iff there exist a measurable function $\pi(\omega) \in \mathfrak{D}(x, \xi(\omega))$, $\omega \in \Omega$, and a vector $\mu \in \mathbb{R}^m$ such that*

$$\mathbb{E}\big[T^{\mathsf{T}}\pi\big] + A^{\mathsf{T}}\mu \le c,$$
$$\bar{x}^{\mathsf{T}}\left(c - \mathbb{E}\big[T^{\mathsf{T}}\pi\big] - A^{\mathsf{T}}\mu\right) = 0.$$

Proof. Since $\operatorname{int}(\operatorname{dom}\phi) \cap X$ is nonempty, we have by the Moreau–Rockafellar theorem that

$$\partial\left(c^{\mathsf{T}}\bar{x} + \phi(\bar{x}) + \mathbb{I}_X(\bar{x})\right) = c + \partial\phi(\bar{x}) + \partial\mathbb{I}_X(\bar{x}).$$

Also, $\partial\mathbb{I}_X(\bar{x}) = \mathcal{N}_X(\bar{x})$. Therefore, we have here that (2.38) is necessary and sufficient optimality conditions for minimizing $c^{\mathsf{T}}x + \phi(x)$ over $x \in X$. Using the characterization of the subdifferential of $\phi(\cdot)$ given in (2.8), we conclude that (2.38) is equivalent to existence of a measurable function $\pi(\omega) \in \mathfrak{D}(x_0, \xi(\omega))$ such that

$$0 \in c - \mathbb{E}\big[T^{\mathsf{T}}\pi\big] + \mathcal{N}_{\operatorname{dom}\phi}(\bar{x}) + \mathcal{N}_X(\bar{x}). \tag{2.41}$$

Moreover, because of the condition $\mathcal{N}_{\operatorname{dom}\phi}(\bar{x}) \subset \mathcal{N}_X(\bar{x})$, the term $\mathcal{N}_{\operatorname{dom}\phi}(\bar{x})$ can be omitted. The proof can be completed now by using (2.41) together with formula (2.39) for the normal cone $\mathcal{N}_X(\bar{x})$. $\square$

The additional technical condition $\mathcal{N}_{\operatorname{dom}\phi}(\bar{x}) \subset \mathcal{N}_X(\bar{x})$ was needed in the above derivations in order to eliminate the term $\mathcal{N}_{\operatorname{dom}\phi}(\bar{x})$ in (2.41). In particular, this condition holds if $\bar{x} \in \operatorname{int}(\operatorname{dom}\phi)$, in which case $\mathcal{N}_{\operatorname{dom}\phi}(\bar{x}) = \{0\}$, or in the case of relatively complete recourse, i.e., when $X \subset \operatorname{dom}\phi$. If the condition of relatively complete recourse is not satisfied, we may need to take into account the normal cone to the domain of $\phi(\cdot)$. In general, this requires application of techniques of functional analysis, which are beyond the scope of this book. However, in the special case of a deterministic matrix T we can carry out the analysis directly.

Theorem 2.12. *Let $\bar{x}$ be a feasible solution of problem* (2.1)–(2.2). *Suppose that the assumptions of Proposition* 2.7 *are satisfied,* $\operatorname{int}(\operatorname{dom}\phi) \cap X$ *is nonempty, and the matrix T is deterministic. Then $\bar{x}$ is an optimal solution of problem* (2.1)–(2.2) *iff there exist a measurable function $\pi(\omega) \in \mathfrak{D}(x, \xi(\omega))$, $\omega \in \Omega$, and a vector $\mu \in \mathbb{R}^m$ such that*

$$T^{\mathsf{T}}\mathbb{E}[\pi] + A^{\mathsf{T}}\mu \le c,$$
$$\bar{x}^{\mathsf{T}}\big(c - T^{\mathsf{T}}\mathbb{E}[\pi] - A^{\mathsf{T}}\mu\big) = 0.$$

Proof. Since T is deterministic, we have that $\mathbb{E}[T^{\mathsf T}\pi] = T^{\mathsf T}\mathbb{E}[\pi]$, and hence the optimality conditions (2.41) can be written as

$$0 \in c - T^{\mathsf T}\mathbb{E}[\pi] + \mathcal{N}_{\mathrm{dom}\,\phi}(\bar{x}) + \mathcal{N}_X(\bar{x}).$$

Now we need to calculate the cone $\mathcal{N}_{\mathrm{dom}\,\phi}(\bar{x})$. Recall that under the assumptions of Proposition 2.7 (in particular, that the recourse is fixed and $\Pi(q)$ is nonempty w.p. 1), we have that $\phi(\cdot) > -\infty$ and formula (2.30) holds true. We have here that only q and h are random while both matrices W and T are deterministic, and (2.30) simplifies to

$$\mathrm{dom}\,\phi = \Big\{x : -Tx \in \bigcap_{h \in \Sigma} \big(- h + \mathrm{pos}\, W\big)\Big\},$$

where Σ is the support of the distribution of the random vector h. The tangent cone to $\mathrm{dom}\,\phi$ at $\bar{x}$ has the form

$$\begin{aligned}\mathcal{T}_{\mathrm{dom}\,\phi}(\bar{x}) &= \Big\{d : -Td \in \bigcap_{h \in \Sigma} \big(\mathrm{pos}\, W + \mathrm{lin}(-h + T\bar{x})\big)\Big\}\\ &= \Big\{d : -Td \in \mathrm{pos}\, W + \bigcap_{h \in \Sigma} \mathrm{lin}(-h + T\bar{x})\Big\}.\end{aligned}$$

Defining the linear subspace

$$L := \bigcap_{h \in \Sigma} \mathrm{lin}(-h + T\bar{x}),$$

we can write the tangent cone as

$$\mathcal{T}_{\mathrm{dom}\,\phi}(\bar{x}) = \{d : -Td \in \mathrm{pos}\, W + L\}.$$

Therefore the normal cone equals

$$\mathcal{N}_{\mathrm{dom}\,\phi}(\bar{x}) = \big\{ - T^{\mathsf T}v : v \in (\mathrm{pos}\, W + L)^*\big\} = -T^{\mathsf T}\big[(\mathrm{pos}\, W)^* \cap L^{\perp}\big].$$

Here we used the fact that $\mathrm{pos}\, W$ is polyhedral and no interior condition is needed for calculating $(\mathrm{pos}\, W + L)^*$. Recalling equation (2.11) we conclude that

$$\mathcal{N}_{\mathrm{dom}\,\phi}(\bar{x}) = -T^{\mathsf T}\big(\Pi_0 \cap L^{\perp}\big).$$

Observe that if $\nu \in \Pi_0 \cap L^{\perp}$, then ν is an element of the recession cone of the set $\mathfrak{D}(\bar{x}, \xi)$ for all $\xi \in \Xi$. Thus $\pi(\omega) + \nu$ is also an element of the set $\mathfrak{D}(x, \xi(\omega))$ for almost all $\omega \in \Omega$. Consequently,

$$\begin{aligned}-T^{\mathsf T}\mathbb{E}\big[\mathfrak{D}(\bar{x}, \xi)\big] + \mathcal{N}_{\mathrm{dom}\,\phi}(\bar{x}) &= -T^{\mathsf T}\mathbb{E}\big[\mathfrak{D}(\bar{x}, \xi)\big] - T^{\mathsf T}\big(\Pi_0 \cap L^{\perp}\big)\\ &= -T^{\mathsf T}\mathbb{E}\big[\mathfrak{D}(\bar{x}, \xi)\big],\end{aligned}$$

and the result follows. $\square$

Example 2.13 (Capacity Expansion, continued). Let us return to Example 2.13 and suppose the support Ξ of the random demand vector ξ is compact. Only the right-hand side ξ in the second-stage problem (2.19)–(2.21) is random, and for a sufficiently large x the second-stage problem is feasible for all $\xi \in \Xi$. Thus conditions of Theorem 2.11 are satisfied. It follows from Theorem 2.11 that $\bar{x}$ is an optimal solution of problem (2.23) iff there exist measurable functions $\mu_n(\xi)$, $n \in \mathcal{N}$, such that for all $\xi \in \Xi$ we have $\mu(\xi) \in \mathcal{M}(\bar{x}, \xi)$, and for all $(i, j) \in \mathcal{A}$ the following conditions are satisfied:

$$c_{ij} \geq \int_{\Xi} \max\{0, \mu_i(\xi) - \mu_j(\xi) - q_{ij}\}\, P(d\xi), \tag{2.42}$$

$$\left(\bar{x}_{ij} - x_{ij}^{\min}\right)\left(c_{ij} - \int_{\Xi} \max\{0, \mu_i(\xi) - \mu_j(\xi) - q_{ij}\}\, P(d\xi)\right) = 0. \tag{2.43}$$

In particular, for every $(i, j) \in \mathcal{A}$ such that $\bar{x}_{ij} > x_{ij}^{\min}$ we have equality in (2.42). Each function $\mu_n(\xi)$ can be interpreted as a random potential of node $n \in \mathcal{N}$. ■

2.2 Polyhedral Two-Stage Problems

2.2.1 General Properties

Let us consider a slightly more general formulation of a two-stage stochastic programming problem,

$$\operatorname*{Min}_{x}\ f_1(x) + \mathbb{E}[Q(x, \omega)], \tag{2.44}$$

where $Q(x, \omega)$ is the optimal value of the second-stage problem

$$\begin{aligned} &\operatorname*{Min}_{y}\ f_2(y, \omega) \\ &\text{s.t. } T(\omega)x + W(\omega)y = h(\omega). \end{aligned} \tag{2.45}$$

We assume in this section that the above two-stage problem is *polyhedral*. That is, the following holds:

- The function $f_1(\cdot)$ is *polyhedral* (compare with Definition 7.1). This means that there exist vectors c_j and scalars α_j, $j = 1, \dots, J_1$, vectors a_k and scalars b_k, $k = 1, \dots, K_1$, such that $f_1(x)$ can be represented as follows:

$$f_1(x) = \begin{cases} \max\limits_{1 \leq j \leq J_1} \alpha_j + c_j^{\mathsf{T}} x & \text{if } a_k^{\mathsf{T}} x \leq b_k, \quad k = 1, \dots, K_1, \\ +\infty & \text{otherwise,} \end{cases}$$

  and its domain $\operatorname{dom} f_1 = \left\{x : a_k^{\mathsf{T}} x \leq b_k,\ k = 1, \dots, K_1\right\}$ is nonempty. (Note that any polyhedral function is convex and lower semicontinuous.)

- The function f_2 is *random polyhedral*. That is, there exist random vectors $q_j = q_j(\omega)$ and random scalars $\gamma_j = \gamma_j(\omega)$, $j = 1, \dots, J_2$, random vectors $d_k = d_k(\omega)$, and random scalars $r_k = r_k(\omega)$, $k = 1, \dots, K_2$, such that $f_2(y, \omega)$ can be represented as follows:

$$f_2(y,\omega) = \begin{cases} \max\limits_{1 \le j \le J_2} \gamma_j(\omega) + q_j(\omega)^{\mathsf T} y & \text{if } d_k(\omega)^{\mathsf T} y \le r_k(\omega), \quad k = 1, \dots, K_2, \\ +\infty & \text{otherwise,} \end{cases}$$

and for a.e. ω the domain of $f_2(\cdot, \omega)$ is nonempty.

Note that (linear) constraints of the second-stage problem which are independent of x, for example, $y \ge 0$, can be absorbed into the objective function $f_2(y, \omega)$. Clearly, the linear two-stage model (2.1)–(2.2) is a special case of a polyhedral two-stage problem. The converse is also true, that is, every polyhedral two-stage model can be reformulated as a linear two-stage model. For example, the second-stage problem (2.45) can be written as follows:

$$\begin{aligned} \operatorname*{Min}_{y,v} \ & v \\ \text{s.t. } & T(\omega)x + W(\omega)y = h(\omega), \\ & \gamma_j(\omega) + q_j(\omega)^{\mathsf T} y \le v, \quad j = 1, \dots, J_2, \\ & d_k(\omega)^{\mathsf T} y \le r_k(\omega), \quad k = 1, \dots, K_2. \end{aligned}$$

Here, both v and y play the role of the second stage variables, and the data (q, T, W, h) in (2.2) have to be redefined in an appropriate way. In order to avoid all these manipulations and unnecessary notational complications that come with such a conversion, we shall address polyhedral problems in a more abstract way. This will also help us to deal with multistage problems and general convex problems.

Consider the Lagrangian of the second-stage problem (2.45):

$$L(y, \pi; x, \omega) := f_2(y, \omega) + \pi^{\mathsf T}\big(h(\omega) - T(\omega)x - W(\omega)y\big).$$

We have

$$\begin{aligned} \inf_y L(y, \pi; x, \omega) &= \pi^{\mathsf T}\big(h(\omega) - T(\omega)x\big) + \inf_y \big[f_2(y,\omega) - \pi^{\mathsf T} W(\omega) y\big] \\ &= \pi^{\mathsf T}\big(h(\omega) - T(\omega)x\big) - f_2^*(W(\omega)^{\mathsf T}\pi, \omega), \end{aligned}$$

where $f_2^*(\cdot, \omega)$ is the conjugate[7] of $f_2(\cdot, \omega)$. We obtain that the dual of problem (2.45) can be written as

$$\operatorname*{Max}_{\pi} \ \big[\pi^{\mathsf T}\big(h(\omega) - T(\omega)x\big) - f_2^*(W(\omega)^{\mathsf T}\pi, \omega)\big]. \tag{2.46}$$

By the duality theory of linear programming, if, for some (x, ω), the optimal value $Q(x, \omega)$ of problem (2.45) is less than $+\infty$ (i.e., problem (2.45) is feasible), then it is equal to the optimal value of the dual problem (2.46).

Let us denote, as before, by $\mathfrak{D}(x, \omega)$ the set of optimal solutions of the dual problem (2.46). We then have an analogue of Proposition 2.2.

[7]Note that since $f_2(\cdot, \omega)$ is polyhedral, so is $f_2^*(\cdot, \omega)$.

Proposition 2.14. *Let $\omega \in \Omega$ be given and suppose that $Q(\cdot, \omega)$ is finite in at least one point $\bar{x}$. Then the function $Q(\cdot, \omega)$ is polyhedral (and hence convex). Moreover, $Q(\cdot, \omega)$ is subdifferentiable at every x at which the value $Q(x, \omega)$ is finite, and*

$$\partial Q(x, \omega) = -T(\omega)^{\mathsf{T}} \mathfrak{D}(x, \omega). \tag{2.47}$$

Proof. Let us define the function $\psi(\pi) := f_2^*(W^{\mathsf{T}}\pi)$. (For simplicity we suppress the argument ω.) We have that if $Q(x, \omega)$ is finite, then it is equal to the optimal value of problem (2.46), and hence $Q(x, \omega) = \psi^*(h - Tx)$. Therefore, $Q(\cdot, \omega)$ is a polyhedral function. Moreover, it follows by the Fenchel–Moreau theorem that

$$\partial \psi^*(h - Tx) = \mathfrak{D}(x, \omega),$$

and the chain rule for subdifferentiation yields formula (2.47). Note that we do not need here additional regularity conditions because of the polyhedricity of the considered case. $\square$

If $Q(x, \omega)$ is finite, then the set $\mathfrak{D}(x, \omega)$ of optimal solutions of problem (2.46) is a nonempty convex closed polyhedron. If, moreover, $\mathfrak{D}(x, \omega)$ is bounded, then it is the convex hull of its finitely many vertices (extreme points), and $Q(\cdot, \omega)$ is finite in a neighborhood of x. If $\mathfrak{D}(x, \omega)$ is unbounded, then its recession cone (which is polyhedral) is the normal cone to the domain of $Q(\cdot, \omega)$ at the point x.

2.2.2 Expected Recourse Cost

Let us consider the expected value function $\phi(x) := \mathbb{E}[Q(x, \omega)]$. Suppose that the probability measure P has a finite support, i.e., there exists a finite number of scenarios ω_k with respective (positive) probabilities p_k, $k = 1, \dots, K$. Then

$$\mathbb{E}[Q(x, \omega)] = \sum_{k=1}^{K} p_k Q(x, \omega_k).$$

For a given x, the expectation $\mathbb{E}[Q(x, \omega)]$ is equal to the optimal value of the problem

$$\begin{aligned} \operatorname*{Min}_{y_1, \dots, y_K} \; & \sum_{k=1}^{K} p_k f_2(y_k, \omega_k) \\ \text{s.t. } & T_k x + W_k y_k = h_k, \; k = 1, \dots, K, \end{aligned} \tag{2.48}$$

where $(h_k, T_k, W_k) := (h(\omega_k), T(\omega_k), W(\omega_k))$. Similarly to the linear case, if for at least one $k \in \{1, \dots, K\}$ the set

$$\operatorname{dom} f_2(\cdot, \omega_k) \cap \{y : T_k x + W_k y = h_k\}$$

is empty, i.e., the corresponding second-stage problem is infeasible, then problem (2.48) is infeasible, and hence its optimal value is $+\infty$.

Proposition 2.15. *Suppose that the probability measure P has a finite support and that the expectation function $\phi(\cdot) := \mathbb{E}[Q(\cdot, \omega)]$ has a finite value in at least one point $x \in \mathbb{R}^n$.*

Then the function $\phi(\cdot)$ is polyhedral, and for any $x_0 \in \operatorname{dom}\phi$,

$$\partial\phi(x_0) = \sum_{k=1}^{K} p_k \partial Q(x_0, \omega_k). \tag{2.49}$$

The proof is identical to the proof of Proposition 2.3. Since the functions $Q(\cdot, \omega_k)$ are polyhedral, formula (2.49) follows by the Moreau–Rockafellar theorem.

The subdifferential $\partial Q(x_0, \omega_k)$ of the second-stage optimal value function is described in Proposition 2.14. That is, if $Q(x_0, \omega_k)$ is finite, then

$$\partial Q(x_0, \omega_k) = -T_k^{\mathsf{T}} \arg\max \left\{ \pi^{\mathsf{T}}\big(h_k - T_k x_0\big) - f_2^*(W_k^{\mathsf{T}}\pi, \omega_k) \right\}. \tag{2.50}$$

It follows that the expectation function ϕ is differentiable at x_0 iff for every ω_k, $k = 1, \dots, K$, the maximum at the right-hand side of (2.50) is attained at a unique point, i.e., the corresponding second-stage dual problem has a unique optimal solution.

Let us now consider the case of a general probability distribution P. We need to ensure that the expectation function $\phi(x) := \mathbb{E}[Q(x, \omega)]$ is well defined. General conditions are complicated, so we resort again to the case of fixed recourse.

We say that the two-stage polyhedral problem has *fixed recourse* if the matrix W and the set[8] $\mathcal{Y} := \operatorname{dom} f_2(\cdot, \omega)$ are fixed, i.e., do not depend on ω. In that case,

$$f_2(y, \omega) = \begin{cases} \max\limits_{1 \le j \le J_2} \ \gamma_j(\omega) + q_j(\omega)^{\mathsf{T}} y & \text{if } y \in \mathcal{Y}, \\ +\infty & \text{otherwise.} \end{cases}$$

Denote $W(\mathcal{Y}) := \{Wy : y \in \mathcal{Y}\}$. Let x be such that

$$h(\omega) - T(\omega)x \in W(\mathcal{Y}) \quad \text{w.p. } 1. \tag{2.51}$$

This means that for a.e. ω the system

$$y \in \mathcal{Y}, \quad Wy = h(\omega) - T(\omega)x \tag{2.52}$$

has a solution. Let for some $\omega_0 \in \Omega$, y_0 be a solution of the above system, i.e., $y_0 \in \mathcal{Y}$ and $h(\omega_0) - T(\omega_0)x = Wy_0$. Since system (2.52) is defined by linear constraints, we have by Hoffman's lemma that there exists a constant κ such that for almost all ω we can find a solution $\bar{y}(\omega)$ of the system (2.52) with

$$\|\bar{y}(\omega) - y_0\| \le \kappa \|(h(\omega) - T(\omega)x) - (h(\omega_0) - T(\omega_0)x)\|.$$

Therefore the optimal value of the second-stage problem can be bounded from above as follows:

$$\begin{aligned} Q(x, \omega) &\le \max_{1 \le j \le J_2} \left\{ \gamma_j(\omega) + q_j(\omega)^{\mathsf{T}} \bar{y}(\omega) \right\} \\ &\le Q(x, \omega_0) + \sum_{j=1}^{J_2} |\gamma_j(\omega) - \gamma_j(\omega_0)| \\ &\quad + \kappa \sum_{j=1}^{J_2} \|q_j(\omega)\| \big(\|h(\omega) - h(\omega_0)\| + \|x\| \, \|T(\omega) - T(\omega_0)\| \big). \end{aligned} \tag{2.53}$$

[8]Note that since it is assumed that $f_2(\cdot, \omega)$ is polyhedral, it follows that the set $\mathcal{Y}$ is nonempty and polyhedral.

Proposition 2.16. *Suppose that the recourse is fixed and*

$$\mathbb{E}|\gamma_j| < +\infty, \ \mathbb{E}\big[\|q_j\|\,\|h\|\big] < +\infty \text{ and } \mathbb{E}\big[\|q_j\|\,\|T\|\big] < +\infty, \ j = 1, \dots, J_2. \tag{2.54}$$

Consider a point $x \in \mathbb{R}^n$. Then $\mathbb{E}[Q(x, \omega)_+]$ is finite iff condition (2.51) *holds.*

Proof. The proof uses (2.53), similar to the proof of Proposition 2.6. □

Let us now formulate conditions under which the expected recourse cost is bounded from below. Let $\mathcal{C}$ be the recession cone of $\mathcal{Y}$ and let $\mathcal{C}^*$ be its polar. Consider the conjugate function $f_2^*(\cdot, \omega)$. It can be verified that

$$\operatorname{dom} f_2^*(\cdot, \omega) = \operatorname{conv}\big\{q_j(\omega), \ j = 1, \dots, J_2\big\} + \mathcal{C}^*. \tag{2.55}$$

Indeed, by the definition of the function $f_2(\cdot, \omega)$ and its conjugate, we have that $f_2^*(z, \omega)$ is equal to the optimal value of the

$$\begin{aligned} &\operatorname*{Max}_{y,v} \ v \\ &\text{s.t. } z^{\mathsf{T}} y - \gamma_j(\omega) - q_j(\omega)^{\mathsf{T}} y \ge v, \ j = 1, \dots, J_2, \quad y \in \mathcal{Y}. \end{aligned}$$

Since it is assumed that the set $\mathcal{Y}$ is nonempty, the above problem is feasible, and since $\mathcal{Y}$ is polyhedral, it is linear. Therefore, its optimal value is equal to the optimal value of its dual. In particular, its optimal value is less than $+\infty$ iff the dual problem is feasible. Now the dual problem is feasible iff there exist $\pi_j \ge 0$, $j = 1, \dots, J_2$, such that $\sum_{j=1}^{J_2} \pi_j = 1$ and

$$\sup_{y \in \mathcal{Y}} y^{\mathsf{T}} \left(z - \sum_{j=1}^{J_2} \pi_j q_j(\omega) \right) < +\infty.$$

The last condition holds iff $z - \sum_{j=1}^{J_2} \pi_j q_j(\omega) \in \mathcal{C}^*$, which completes the argument.

Let us define the set

$$\Pi(\omega) := \big\{\pi : W^{\mathsf{T}}\pi \in \operatorname{conv}\big\{q_j(\omega), \ j = 1, \dots, J_2\big\} + \mathcal{C}^*\big\}.$$

We may remark that in the case of a linear two-stage problem, the above set coincides with the one defined in (2.5).

Proposition 2.17. *Suppose that* (i) *the recourse is fixed,* (ii) *the set $\Pi(\omega)$ is nonempty w.p.* 1*, and* (iii) *condition* (2.54) *holds.*

Then the expectation function $\phi(x)$ is well defined and $\phi(x) > -\infty$ for all $x \in \mathbb{R}^n$. Moreover, ϕ is convex, lower semicontinuous and Lipschitz continuous on $\operatorname{dom}\phi$*, its domain* $\operatorname{dom}\phi$ *is a convex closed subset of $\mathbb{R}^n$, and*

$$\operatorname{dom}\phi = \big\{x \in \mathbb{R}^n : h - Tx \in W(\mathcal{Y}) \text{ w.p.1}\big\}. \tag{2.56}$$

Furthermore, for any $x_0 \in \operatorname{dom}\phi$,

$$\partial\phi(x_0) = -\mathbb{E}\left[T^{\mathsf{T}}\mathfrak{D}(x_0, \omega)\right] + \mathcal{N}_{\operatorname{dom}\phi}(x_0), \tag{2.57}$$

Proof. Note that the dual problem (2.46) is feasible iff $W^{\mathsf{T}}\pi \in \operatorname{dom} f_2^*(\cdot, \omega)$. By formula (2.55), assumption (ii) means that problem (2.46) is feasible, and hence $Q(x, \omega)$ is equal to the optimal value of (2.46) for a.e. ω. The remainder of the proof is similar to the linear case (Propositions 2.7 and 2.8). □

2.2.3 Optimality Conditions

The optimality conditions for polyhedral two-stage problems are similar to those for linear problems. For completeness we provide the appropriate formulations. Let us start from the problem with finitely many elementary events ω_k occurring with probabilities p_k, $k = 1, \dots, K$.

Theorem 2.18. *Suppose that the probability measure P has a finite support. Then a point $\bar{x}$ is an optimal solution of the first-stage problem* (2.44) *iff there exist $\pi_k \in \mathfrak{D}(\bar{x}, \omega_k)$, $k = 1, \dots, K$, such that*

$$0 \in \partial f_1(\bar{x}) - \sum_{k=1}^{K} p_k T_k^{\mathsf{T}} \pi_k. \tag{2.58}$$

Proof. Since $f_1(x)$ and $\phi(x) = \mathbb{E}[Q(x, \omega)]$ are convex functions, a necessary and sufficient condition for a point $\bar{x}$ to be a minimizer of $f_1(x) + \phi(x)$ reads

$$0 \in \partial\big[f_1(\bar{x}) + \phi(\bar{x})\big]. \tag{2.59}$$

In particular, the above condition requires $f_1(\bar{x})$ and $\phi(\bar{x})$ to be finite valued. By the Moreau–Rockafellar theorem we have that $\partial\big[f_1(\bar{x}) + \phi(\bar{x})\big] = \partial f_1(\bar{x}) + \partial\phi(\bar{x})$. Note that there is no need here for additional regularity conditions because of the polyhedricity of functions f_1 and ϕ. The proof can be completed now by using the formula for $\partial\phi(\bar{x})$ given in Proposition 2.15. □

In the case of general distributions, the derivation of optimality conditions requires additional assumptions.

Theorem 2.19. *Suppose that* (i) *the recourse is fixed and relatively complete,* (ii) *the set $\Pi(\omega)$ is nonempty w.p.* 1*, and* (iii) *condition* (2.54) *holds.*

Then a point $\bar{x}$ is an optimal solution of problem (2.44)–(2.45) *iff there exists a measurable function $\pi(\omega) \in \mathfrak{D}(\bar{x}, \omega)$, $\omega \in \Omega$, such that*

$$0 \in \partial f_1(\bar{x}) - \mathbb{E}\big[T^{\mathsf{T}}\pi\big]. \tag{2.60}$$

Proof. The result follows immediately from the optimality condition (2.59) and formula (2.57). Since the recourse is relatively complete, we can omit the normal cone to the domain of $\phi(\cdot)$. □

If the recourse is not relatively complete, the analysis becomes complicated. The normal cone to the domain of $\phi(\cdot)$ enters the optimality conditions. For the domain described in (2.56), this cone is rather difficult to describe in a closed form. Some simplification can be achieved when T is deterministic. The analysis then mirrors the linear case, as in Theorem 2.12.

2.3 General Two-Stage Problems

2.3.1 Problem Formulation, Interchangeability

In a general way, two-stage stochastic programming problems can be written in the following form:

$$\operatorname*{Min}_{x\in X} \big\{ f(x) := \mathbb{E}[F(x,\omega)] \big\}, \tag{2.61}$$

where $F(x,\omega)$ is the optimal value of the second-stage problem

$$\operatorname*{Min}_{y\in \mathcal{G}(x,\omega)} g(x,y,\omega). \tag{2.62}$$

Here $X \subset \mathbb{R}^n$, $g : \mathbb{R}^n \times \mathbb{R}^m \times \Omega \to \mathbb{R}$, and $\mathcal{G} : \mathbb{R}^n \times \Omega \rightrightarrows \mathbb{R}^m$ is a multifunction. In particular, the linear two-stage problem (2.1)–(2.2) can be formulated in the above form with $g(x,y,\omega) := c^{\mathsf{T}}x + q(\omega)^{\mathsf{T}}y$ and

$$\mathcal{G}(x,\omega) := \{y : T(\omega)x + W(\omega)y = h(\omega),\ y \geq 0\}.$$

We also use the notation $g_\omega(x,y) = g(x,y,\omega)$ and $\mathcal{G}_\omega(x) = \mathcal{G}(x,\omega)$.

Of course, the second-stage problem (2.62) also can be written in the following equivalent form:

$$\operatorname*{Min}_{y\in \mathbb{R}^m} \bar{g}(x,y,\omega), \tag{2.63}$$

where

$$\bar{g}(x,y,\omega) := \begin{cases} g(x,y,\omega) & \text{if } y \in \mathcal{G}(x,\omega), \\ +\infty & \text{otherwise.} \end{cases} \tag{2.64}$$

We assume that the function $\bar{g}(x,y,\omega)$ is random lower semicontinuous. Recall that if $g(x,y,\cdot)$ is measurable for every $(x,y) \in \mathbb{R}^n \times \mathbb{R}^m$ and $g(\cdot,\cdot,\omega)$ is continuous for a.e. $\omega \in \Omega$, i.e., $g(x,y,\omega)$ is a Carathéodory function, then $g(x,y,\omega)$ is random lower semicontinuous. Random lower semicontinuity of $\bar{g}(x,y,\omega)$ implies that the optimal value function $F(x,\cdot)$ is measuarable (see Theorem 7.37). Moreover, if for a.e. $\omega \in \Omega$ function $F(\cdot,\omega)$ is continuous, then $F(x,\omega)$ is a Carathéodory function and hence is random lower semicontinuous. The indicator function $\mathbb{I}_{\mathcal{G}_\omega(x)}(y)$ is random lower semicontinuous if for every $\omega \in \Omega$ the multifunction $\mathcal{G}_\omega(\cdot)$ is closed and $\mathcal{G}(x,\omega)$ is measurable with respect to the sigma algebra of $\mathbb{R}^n \times \Omega$ (see Theorem 7.36). Of course, if $g(x,y,\omega)$ and $\mathbb{I}_{\mathcal{G}_\omega(x)}(y)$ are random lower semicontinuous, then their sum $\bar{g}(x,y,\omega)$ is also random lower semicontinuous.

Now let $\mathfrak{Y}$ be a linear *decomposable* space of measurable mappings from Ω to $\mathbb{R}^m$. For example, we can take $\mathfrak{Y} := \mathcal{L}_p(\Omega, \mathcal{F}, P; \mathbb{R}^m)$ with $p \in [1, +\infty]$. Then by the interchangeability principle we have

$$\underbrace{\mathbb{E}\left[\inf_{y\in\mathbb{R}^m} \bar{g}(x, y, \omega)\right]}_{F(x,\omega)} = \inf_{\boldsymbol{y}\in\mathfrak{Y}} \mathbb{E}\big[\bar{g}(x, \boldsymbol{y}(\omega), \omega)\big], \tag{2.65}$$

provided that the right-hand side of (2.65) is less than $+\infty$ (see Theorem 7.80). This implies the following *interchangeability principle for two-stage programming*.

Theorem 2.20. *The two-stage problem* (2.61)–(2.62) *is equivalent to the following problem:*

$$\begin{aligned} &\operatorname*{Min}_{x\in\mathbb{R}^n,\, \boldsymbol{y}\in\mathfrak{Y}} \ \mathbb{E}\,[g(x, \boldsymbol{y}(\omega), \omega)] \\ &\quad \text{s.t. } x \in X,\ \ \boldsymbol{y}(\omega) \in \mathcal{G}(x, \omega) \text{ a.e. } \omega \in \Omega. \end{aligned} \tag{2.66}$$

The equivalence is understood in the sense that optimal values of problems (2.61) *and* (2.66) *are equal to each other, provided that the optimal value of problem* (2.66) *is less than* $+\infty$. *Moreover, assuming that the common optimal value of problems* (2.61) *and* (2.66) *is finite, we have that if* $(\bar{x}, \bar{\boldsymbol{y}})$ *is an optimal solution of problem* (2.66)*, then* $\bar{x}$ *is an optimal solution of the first-stage problem* (2.61) *and* $\bar{y} = \bar{\boldsymbol{y}}(\omega)$ *is an optimal solution of the second-stage problem* (2.62) *for* $x = \bar{x}$ *and a.e.* $\omega \in \Omega$*; conversely, if* $\bar{x}$ *is an optimal solution of the first-stage problem* (2.61) *and for* $x = \bar{x}$ *and a.e.* $\omega \in \Omega$ *the second-stage problem* (2.62) *has an optimal solution* $\bar{y} = \bar{\boldsymbol{y}}(\omega)$ *such that* $\bar{\boldsymbol{y}} \in \mathfrak{Y}$*, then* $(\bar{x}, \bar{\boldsymbol{y}})$ *is an optimal solution of problem* (2.66)*.*

Note that optimization in the right-hand side of (2.65) and in (2.66) is performed over mappings $\boldsymbol{y} : \Omega \to \mathbb{R}^m$ belonging to the space $\mathfrak{Y}$. In particular, if $\Omega = \{\omega_1, \dots, \omega_K\}$ is finite, then by setting $y_k := \boldsymbol{y}(\omega_k)$, $k = 1, \dots, K$, every such mapping can be identified with a vector $(y_1, \dots, y_K)$ and the space $\mathfrak{Y}$ with the finite dimensional space $\mathbb{R}^{mK}$. In that case, problem (2.66) takes the form (compare with (2.15))

$$\begin{aligned} &\operatorname*{Min}_{x, y_1,\dots,y_K} \ \sum_{k=1}^{K} p_k g(x, y_k, \omega_k) \\ &\quad \text{s.t. } x \in X,\ \ y_k \in \mathcal{G}(x, \omega_k),\ \ k = 1, \dots, K. \end{aligned} \tag{2.67}$$

2.3.2 Convex Two-Stage Problems

We say that the two-stage problem (2.61)–(2.62) is *convex* if the set X is convex (and closed) and for every $\omega \in \Omega$ the function $\bar{g}(x, y, \omega)$, defined in (2.64), is convex in $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$. We leave this as an exercise to show that in such case the optimal value function $F(\cdot, \omega)$ is convex, and hence (2.61) is a convex problem. It could be useful to understand what conditions will guarantee convexity of the function $\bar{g}_\omega(x, y) = \bar{g}(x, y, \omega)$. We have that $\bar{g}_\omega(x, y) = g_\omega(x, y) + \mathbb{I}_{\mathcal{G}_\omega(x)}(y)$. Therefore $\bar{g}_\omega(x, y)$ is convex if $g_\omega(x, y)$ is convex and the indicator function $\mathbb{I}_{\mathcal{G}_\omega(x)}(y)$ is convex in (x, y). It is not difficult to see that the indicator function $\mathbb{I}_{\mathcal{G}_\omega(x)}(y)$ is convex iff the following condition holds for any $t \in [0, 1]$:

$$y \in \mathcal{G}_\omega(x),\ \ y' \in \mathcal{G}_\omega(x') \ \Rightarrow ty + (1-t)y' \in \mathcal{G}_\omega(tx + (1-t)x'). \tag{2.68}$$

Equivalently this condition can be written as

$$t\mathcal{G}_\omega(x) + (1-t)\mathcal{G}_\omega(x') \subset \mathcal{G}_\omega(tx + (1-t)x'), \quad \forall x, x' \in \mathbb{R}^n,\ \forall t \in [0, 1]. \tag{2.69}$$

The multifunction $\mathcal{G}_\omega$ satisfying the above condition (2.69) is called *convex*. By taking $x = x'$ we obtain that if the multifunction $\mathcal{G}_\omega$ is convex, then it is convex valued, i.e., the set $\mathcal{G}_\omega(x)$ is convex for every $x \in \mathbb{R}^n$.

In the remainder of this section we assume that the multifunction $\mathcal{G}(x, \omega)$ is defined in the form

$$\mathcal{G}(x, \omega) := \{y \in Y : T(x, \omega) + W(y, \omega) \in -C\}, \tag{2.70}$$

where Y is a nonempty convex closed subset of $\mathbb{R}^m$ and $T = (t_1, \dots, t_\ell) : \mathbb{R}^n \times \Omega \to \mathbb{R}^\ell$, $W = (w_1, \dots, w_\ell) : \mathbb{R}^m \times \Omega \to \mathbb{R}^\ell$, and $C \subset \mathbb{R}^\ell$ is a closed convex cone. Cone C defines a partial order, denoted "$\preceq_C$", on the space $\mathbb{R}^\ell$. That is, $a \preceq_C b$ iff $b - a \in C$. In that notation the constraint $T(x, \omega) + W(y, \omega) \in -C$ can be written as $T(x, \omega) + W(y, \omega) \preceq_C 0$. For example, if $C := \mathbb{R}^\ell_+$, then the constraint $T(x, \omega) + W(y, \omega) \preceq_C 0$ means that $t_i(x, \omega) + w_i(y, \omega) \le 0$, $i = 1, \dots, \ell$. We assume that $t_i(x, \omega)$ and $w_i(y, \omega)$, $i = 1, \dots, \ell$, are Carathéodory functions and that for every $\omega \in \Omega$, mappings $T_\omega(\cdot) = T(\cdot, \omega)$ and $W_\omega(\cdot) = W(\cdot, \omega)$ are convex with respect to the cone C. A mapping $G : \mathbb{R}^n \to \mathbb{R}^\ell$ is said to be *convex with respect* to C if the multifunction $\mathcal{M}(x) := G(x) + C$ is convex. Equivalently, mapping G is convex with respect to C if

$$G\big(tx + (1-t)x'\big) \preceq_C tG(x) + (1-t)G(x'), \quad \forall x, x' \in \mathbb{R}^n,\ \forall t \in [0, 1].$$

For example, mapping $G(\cdot) = (g_1(\cdot), \dots, g_\ell(\cdot))$ is convex with respect to $C := \mathbb{R}^\ell_+$ iff all its components $g_i(\cdot)$, $i = 1, \dots, \ell$, are convex functions. Convexity of T_ω and W_ω implies convexity of the corresponding multifunction $\mathcal{G}_\omega$.

We assume, further, that $g(x, y, \omega) := c(x) + q(y, \omega)$, where $c(\cdot)$ and $q(\cdot, \omega)$ are real valued convex functions. For $\mathcal{G}(x, \omega)$ of the form (2.70), and given x, we can write the second-stage problem, up to the constant $c(x)$, in the form

$$\begin{aligned} &\operatorname*{Min}_{y \in Y}\ q_\omega(y) \\ &\text{s.t. } W_\omega(y) + \chi_\omega \preceq_C 0 \end{aligned} \tag{2.71}$$

with $\chi_\omega := T(x, \omega)$. Let us denote by $\vartheta(\chi, \omega)$ the optimal value of problems (2.71). Note that $F(x, \omega) = c(x) + \vartheta(T(x, \omega), \omega)$. The (Lagrangian) dual of problem (2.71) can be written in the form

$$\operatorname*{Max}_{\pi \succeq_C 0} \big\{\pi^{\mathsf{T}} \chi_\omega + \inf_{y \in Y} L_\omega(y, \pi)\big\}, \tag{2.72}$$

where

$$L_\omega(y, \pi) := q_\omega(y) + \pi^{\mathsf{T}} W_\omega(y)$$

is the Lagrangian of problem (2.71). We have the following results (see Theorems 7.8 and 7.9).

Proposition 2.21. *Let $\omega \in \Omega$ and χ_ω be given and suppose that the specified above convexity assumptions are satisfied. Then the following statements hold true:*

(i) *The functions $\vartheta(\cdot, \omega)$ and $F(\cdot, \omega)$ are convex.*

(ii) *Suppose that problem (2.71) is subconsistent. Then there is no duality gap between problem (2.71) and its dual (2.72) iff the optimal value function $\vartheta(\cdot, \omega)$ is lower semicontinuous at χ_ω.*

(iii) *There is no duality gap between problems (2.71) and (2.72) and the dual problem (2.72) has a nonempty set of optimal solutions iff the optimal value function $\vartheta(\cdot, \omega)$ is subdifferentiable at χ_ω.*

(iv) *Suppose that the optimal value of (2.71) is finite. Then there is no duality gap between problems (2.71) and (2.72) and the dual problem (2.72) has a nonempty and bounded set of optimal solutions iff $\chi_\omega \in \operatorname{int}(\operatorname{dom} \vartheta(\cdot, \omega))$.*

The regularity condition $\chi_\omega \in \operatorname{int}(\operatorname{dom} \vartheta(\cdot, \omega))$ means that for all small perturbations of χ_ω the corresponding problem (2.71) remains feasible.

We can also characterize the differentiability properties of the optimal value functions in terms of the dual problem (2.72). Let us denote by $\mathfrak{D}(\chi, \omega)$ the set of optimal solutions of the dual problem (2.72). This set may be empty, of course.

Proposition 2.22. *Let $\omega \in \Omega$, $x \in \mathbb{R}^n$ and $\chi = T(x, \omega)$ be given. Suppose that the specified convexity assumptions are satisfied and that problems (2.71) and (2.72) have finite and equal optimal values. Then*

$$\partial\vartheta(\chi, \omega) = \mathfrak{D}(\chi, \omega). \tag{2.73}$$

Suppose, further, that functions $c(\cdot)$ and $T_\omega(\cdot)$ are differentiable, and

$$0 \in \operatorname{int}\left\{T_\omega(x) + \nabla T_\omega(x)\mathbb{R}^\ell - \operatorname{dom} \vartheta(\cdot, \omega)\right\}. \tag{2.74}$$

Then

$$\partial F(x, \omega) = \nabla c(x) + \nabla T_\omega(x)^{\mathsf{T}} \mathfrak{D}(\chi, \omega). \tag{2.75}$$

Corollary 2.23. *Let $\omega \in \Omega$, $x \in \mathbb{R}^n$ and $\chi = T(x, \omega)$ and suppose that the specified convexity assumptions are satisfied. Then $\vartheta(\cdot, \omega)$ is differentiable at χ iff $\mathfrak{D}(\chi, \omega)$ is a singleton. Suppose, further, that the functions $c(\cdot)$ and $T_\omega(\cdot)$ are differentiable. Then the function $F(\cdot, \omega)$ is differentiable at every x at which $\mathfrak{D}(\chi, \omega)$ is a singleton.*

Proof. If $\mathfrak{D}(\chi, \omega)$ is a singleton, then the set of optimal solutions of the dual problem (2.72) is nonempty and bounded, and hence there is no duality gap between problems (2.71) and (2.72). Thus formula (2.73) holds. Conversely, if $\partial\vartheta(\chi, \omega)$ is a singleton and hence is nonempty, then again there is no duality gap between problems (2.71) and (2.72), and hence formula (2.73) holds.

Now if $\mathfrak{D}(\chi, \omega)$ is a singleton, then $\vartheta(\cdot, \omega)$ is continuous at χ and hence the regularity condition (2.74) holds. It follows then by formula (2.75) that $F(\cdot, \omega)$ is differentiable at x and formula

$$\nabla F(x, \omega) = \nabla c(x) + \nabla T_\omega(x)^{\mathsf{T}} \mathfrak{D}(\chi, \omega) \tag{2.76}$$

holds true. ∎

Let us focus on the expectation function $f(x) := \mathbb{E}[F(x, \omega)]$. If the set Ω is finite, say, $\Omega = \{\omega_1, \dots, \omega_K\}$ with corresponding probabilities p_k, $k = 1, \dots, K$, then $f(x) = \sum_{k=1}^K p_k F(x, \omega_k)$ and subdifferentiability of $f(x)$ is described by the Moreau–Rockafellar theorem (Theorem 7.4) together with formula (2.75). In particular, $f(\cdot)$ is differentiable at a point x if the functions $c(\cdot)$ and $T_\omega(\cdot)$ are differentiable at x and for every $\omega \in \Omega$ the corresponding dual problem (2.72) has a unique optimal solution.

Let us consider the general case, when Ω is not assumed to be finite. By combining Proposition 2.22 and Theorem 7.47 we obtain that, under appropriate regularity conditions ensuring for a.e. $\omega \in \Omega$ formula (2.75) and interchangeability of the subdifferential and expectation operators, it follows that $f(\cdot)$ is subdifferentiable at a point $\bar{x} \in \operatorname{dom} f$ and

$$\partial f(\bar{x}) = \nabla c(\bar{x}) + \int_\Omega \nabla T_\omega(\bar{x})^{\mathsf{T}} \mathfrak{D}(T_\omega(\bar{x}), \omega)\, dP(\omega) + \mathcal{N}_{\operatorname{dom} f}(\bar{x}). \tag{2.77}$$

In particular, it follows from the above formula (2.77) that $f(\cdot)$ is differentiable at $\bar{x}$ iff $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$ and $\mathfrak{D}(T_\omega(\bar{x}), \omega) = \{\pi(\omega)\}$ is a singleton w.p. 1, in which case

$$\nabla f(\bar{x}) = \nabla c(\bar{x}) + \mathbb{E}\left[\nabla T_\omega(\bar{x})^{\mathsf{T}} \pi(\omega)\right]. \tag{2.78}$$

We obtain the following conditions for optimality.

Proposition 2.24. *Let $\bar{x} \in X \cap \operatorname{int}(\operatorname{dom} f)$ and assume that formula* (2.77) *holds. Then $\bar{x}$ is an optimal solution of the first-stage problem* (2.61) *iff there exists a measurable selection $\pi(\omega) \in \mathfrak{D}(T(\bar{x}, \omega), \omega)$ such that*

$$-c(\bar{x}) - \mathbb{E}\big[\nabla T_\omega(\bar{x})^{\mathsf{T}} \pi(\omega)\big] \in \mathcal{N}_X(\bar{x}). \tag{2.79}$$

Proof. Since $\bar{x} \in X \cap \operatorname{int}(\operatorname{dom} f)$, we have that $\operatorname{int}(\operatorname{dom} f) \neq \emptyset$ and $\bar{x}$ is an optimal solution iff $0 \in \partial f(\bar{x}) + \mathcal{N}_X(\bar{x})$. By formula (2.77) and since $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$, this is equivalent to condition (2.79). ∎

2.4 Nonanticipativity

2.4.1 Scenario Formulation

An additional insight into the structure and properties of two-stage problems can be gained by introducing the concept of *nonanticipativity*. Consider the first-stage problem (2.61). Assume that the number of scenarios is finite, i.e., $\Omega = \{\omega_1, \dots, \omega_K\}$ with respective (positive) probabilities $p_1, \dots, p_K$. Let us relax the first-stage problem by replacing vector

x with K vectors $x_1, x_2, \dots, x_K$, one for each scenario. We obtain the following relaxation of problem (2.61):

$$\operatorname*{Min}_{x_1,\dots,x_K} \sum_{k=1}^{K} p_k F(x_k, \omega_k) \text{ subject to } x_k \in X,\ k = 1, \dots, K. \tag{2.80}$$

We observe that problem (2.80) is separable in the sense that it can be split into K smaller problems, one for each scenario,

$$\operatorname*{Min}_{x_k \in X} F(x_k, \omega_k), \quad k = 1, \dots, K, \tag{2.81}$$

and that the optimal value of problem (2.80) is equal to the weighted sum, with weights p_k, of the optimal values of problems (2.81), $k = 1, \dots, K$. For example, in the case of the two-stage linear program (2.15), relaxation of the form (2.80) leads to solving K smaller problems,

$$\begin{aligned} &\operatorname*{Min}_{x_k \geq 0,\, y_k \geq 0} c^{\mathsf T} x_k + q_k^{\mathsf T} y_k \\ &\text{s.t. } Ax_k = b,\ T_k x_k + W_k y_k = h_k. \end{aligned}$$

Problem (2.80), however, is not suitable for modeling a two-stage decision process. This is because the first-stage decision variables x_k in (2.80) are now allowed to depend on a realization of the random data at the second stage. This can be fixed by introducing the additional constraint

$$(x_1, \dots, x_K) \in \mathfrak{L}, \tag{2.82}$$

where $\mathfrak{L} := \{\boldsymbol{x} = (x_1, \dots, x_K) : x_1 = \dots = x_K\}$ is a linear subspace of the nK-dimensional vector space $\mathfrak{X} := \mathbb{R}^n \times \dots \times \mathbb{R}^n$. Due to the constraint (2.82), all realizations x_k, $k = 1, \dots, K$, of the first-stage decision vector are equal to each other, that is, they do not depend on the realization of the random data. The constraint (2.82) can be written in different forms, which can be convenient in various situations, and will be referred to as the *nonanticipativity constraint*. Together with the nonanticipativity constraint (2.82), problem (2.80) becomes

$$\begin{aligned} &\operatorname*{Min}_{x_1,\dots,x_K} \sum_{k=1}^{K} p_k F(x_k, \omega_k) \\ &\text{s.t. } x_1 = \dots = x_K,\ x_k \in X,\ k = 1, \dots, K. \end{aligned} \tag{2.83}$$

Clearly, the above problem (2.83) is equivalent to problem (2.61). Such nonanticipativity constraints are especially important in multistage modeling, which we discuss later.

A way to write the nonanticipativity constraint is to require that

$$x_k = \sum_{i=1}^{K} p_i x_i, \quad k = 1, \dots, K, \tag{2.84}$$

which is convenient for extensions to the case of a continuous distribution of problem data. Equations (2.84) can be interpreted in the following way. Consider the space $\mathfrak{X}$ equipped with the scalar product

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \sum_{i=1}^{K} p_i x_i^{\mathsf T} y_i. \tag{2.85}$$

Define linear operator $\boldsymbol{P} : \mathfrak{X} \to \mathfrak{X}$ as

$$\boldsymbol{P}\boldsymbol{x} := \left(\sum_{i=1}^{K} p_i x_i, \dots, \sum_{i=1}^{K} p_i x_i \right).$$

Constraint (2.84) can be compactly written as

$$\boldsymbol{x} = \boldsymbol{P}\boldsymbol{x}.$$

It can be verified that $\boldsymbol{P}$ is the orthogonal projection operator of $\mathfrak{X}$, equipped with the scalar product (2.85), onto its subspace $\mathfrak{L}$. Indeed, $\boldsymbol{P}(\boldsymbol{P}\boldsymbol{x}) = \boldsymbol{P}\boldsymbol{x}$, and

$$\langle \boldsymbol{P}\boldsymbol{x}, \boldsymbol{y} \rangle = \left(\sum_{i=1}^{K} p_i x_i \right)^{\mathsf{T}} \left(\sum_{k=1}^{K} p_k y_k \right) = \langle \boldsymbol{x}, \boldsymbol{P}\boldsymbol{y} \rangle. \tag{2.86}$$

The range space of $\boldsymbol{P}$, which is the linear space $\mathfrak{L}$, is called the nonanticipativity subspace of $\mathfrak{X}$.

Another way to algebraically express nonanticipativity, which is convenient for numerical methods, is to write the system of equations

$$\begin{aligned} x_1 &= x_2, \\ x_2 &= x_3, \\ &\ \vdots \\ x_{K-1} &= x_K. \end{aligned} \tag{2.87}$$

This system is very sparse: each equation involves only two variables, and each variable appears in at most two equations, which is convenient for many numerical solution methods.

2.4.2 Dualization of Nonanticipativity Constraints

We discuss now a dualization of problem (2.80) with respect to the nonanticipativity constraints (2.84). Assigning to these nonanticipativity constraints Lagrange multipliers $\lambda_k \in \mathbb{R}^n$, $k = 1, \dots, K$, we can write the Lagrangian

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) := \sum_{k=1}^{K} p_k F(x_k, \omega_k) + \sum_{k=1}^{K} p_k \lambda_k^{\mathsf{T}} \left(x_k - \sum_{i=1}^{K} p_i x_i \right).$$

Note that since $\boldsymbol{P}$ is an orthogonal projection, $\boldsymbol{I} - \boldsymbol{P}$ is also an orthogonal projection (onto the space orthogonal to $\mathfrak{L}$), and hence

$$\sum_{k=1}^{K} p_k \lambda_k^{\mathsf{T}} \left(x_k - \sum_{i=1}^{K} p_i x_i \right) = \langle \boldsymbol{\lambda}, (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{x} \rangle = \langle (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{\lambda}, \boldsymbol{x} \rangle.$$

Therefore, the above Lagrangian can be written in the following equivalent form:

$$L(\boldsymbol{x},\boldsymbol{\lambda}) = \sum_{k=1}^{K} p_k F(x_k,\omega_k) + \sum_{k=1}^{K} p_k \left(\lambda_k - \sum_{j=1}^{K} p_j \lambda_j\right)^{\mathsf{T}} x_k.$$

Let us observe that shifting the multipliers λ_k, $k = 1,\dots,K$, by a constant vector does not change the value of the Lagrangian, because the expression $\lambda_k - \sum_{j=1}^{K} p_j \lambda_j$ is invariant to such shifts. Therefore, with no loss of generality we can assume that

$$\sum_{j=1}^{K} p_j \lambda_j = 0.$$

or, equivalently, that $\boldsymbol{P\lambda} = 0$. Dualization of problem (2.80) with respect to the nonanticipativity constraints takes the form of the following problem:

$$\operatorname*{Max}_{\boldsymbol{\lambda}} \left\{ D(\boldsymbol{\lambda}) := \inf_{\boldsymbol{x}} L(\boldsymbol{x},\boldsymbol{\lambda}) \right\} \text{ s.t. } \boldsymbol{P\lambda} = 0. \tag{2.88}$$

By general duality theory we have that the optimal value of problem (2.61) is greater than or equal to the optimal value of problem (2.88). These optimal values are equal to each other under some regularity conditions; we will discuss a general case in the next section. In particular, if the two-stage problem is linear and since the nonanticipativity constraints are linear, we have in that case that there is no duality gap between problem (2.61) and its dual problem (2.88) unless both problems are infeasible.

Let us take a closer look at the dual problem (2.88). Under the condition $\boldsymbol{P\lambda} = 0$, the Lagrangian can be written simply as

$$L(\boldsymbol{x},\boldsymbol{\lambda}) = \sum_{k=1}^{K} p_k \big(F(x_k,\omega_k) + \lambda_k^{\mathsf{T}} x_k \big).$$

We see that the Lagrangian can be split into K components:

$$L(\boldsymbol{x},\boldsymbol{\lambda}) = \sum_{k=1}^{K} p_k L_k(x_k,\lambda_k),$$

where $L_k(x_k,\lambda_k) := F(x_k,\omega_k) + \lambda_k^{\mathsf{T}} x_k$. It follows that

$$D(\boldsymbol{\lambda}) = \sum_{j=1}^{K} p_k \mathcal{D}_k(\lambda_k),$$

where

$$\mathcal{D}_k(\lambda_k) := \inf_{x_k \in X} L_k(x_k,\lambda_k).$$

For example, in the case of the two-stage linear program (2.15), $\mathfrak{D}_k(\lambda_k)$ is the optimal value of the problem

$$\begin{aligned} \operatorname*{Min}_{x_k, y_k} \; & (c + \lambda_k)^{\mathsf{T}} x_k + q_k^{\mathsf{T}} y_k \\ \text{s.t. } & A x_k = b, \\ & T_k x_k + W_k y_k = h_k, \\ & x_k \geq 0, \quad y_k \geq 0. \end{aligned}$$

We see that value of the dual function $D(\boldsymbol{\lambda})$ can be calculated by solving K independent *scenario subproblems*.

Suppose that there is no duality gap between problem (2.61) and its dual (2.88) and their common optimal value is finite. This certainly holds true if the problem is linear, and both problems, primal and dual, are feasible. Let $\bar{\boldsymbol{\lambda}} = (\bar{\lambda}_1, \dots, \bar{\lambda}_K)$ be an optimal solution of the dual problem (2.88). Then the set of optimal solutions of problem (2.61) is contained in the set of optimal solutions of the problem

$$\operatorname*{Min}_{x_k \in X} \sum_{k=1}^{K} p_k L_k(x_k, \bar{\lambda}_k) \tag{2.89}$$

This inclusion can be strict, i.e., the set of optimal solutions of (2.89) can be larger than the set of optimal solutions of problem (2.61). (See an example of linear program defined in (7.32).) Of course, if problem (2.89) has *unique* optimal solution $\bar{\boldsymbol{x}} = (\bar{x}_1, \dots, \bar{x}_K)$, then $\bar{\boldsymbol{x}} \in \mathfrak{L}$, i.e., $\bar{x}_1 = \cdots = \bar{x}_K$, and this is also the optimal solution of problem (2.61) with $\bar{x}$ being equal to the common value of $\bar{x}_1, \dots, \bar{x}_K$. Note also that the above problem (2.89) is separable, i.e., $\bar{\boldsymbol{x}}$ is an optimal solution of (2.89) iff for every $k = 1, \dots, K$, $\bar{x}_k$ is an optimal solution of the problem

$$\operatorname*{Min}_{x_k \in X} L_k(x_k, \bar{\lambda}_k).$$

2.4.3 Nonanticipativity Duality for General Distributions

In this section we discuss dualization of the first-stage problem (2.61) with respect to nonanticipativity constraints in the general (not necessarily finite-scenarios) case. For the sake of convenience we write problem (2.61) in the form

$$\operatorname*{Min}_{x \in \mathbb{R}^n} \left\{ \bar{f}(x) := \mathbb{E}[\bar{F}(x, \omega)] \right\}, \tag{2.90}$$

where $\bar{F}(x, \omega) := F(x, \omega) + \mathbb{I}_X(x)$, i.e., $\bar{F}(x, \omega) = F(x, \omega)$ if $x \in X$ and $\bar{F}(x, \omega) = +\infty$ otherwise. Let $\mathfrak{X}$ be a linear *decomposable* space of measurable mappings from Ω to $\mathbb{R}^n$. Unless stated otherwise we use $\mathfrak{X} := \mathcal{L}_p(\Omega, \mathcal{F}, P; \mathbb{R}^n)$ for some $p \in [1, +\infty]$ such that for every $\boldsymbol{x} \in \mathfrak{X}$ the expectation $\mathbb{E}[\bar{F}(\boldsymbol{x}(\omega), \omega)]$ is well defined. Then we can write problem (2.90) in the equivalent form

$$\operatorname*{Min}_{\boldsymbol{x} \in \mathfrak{L}} \mathbb{E}[\bar{F}(\boldsymbol{x}(\omega), \omega)], \tag{2.91}$$

where $\mathfrak{L}$ is a linear subspace of $\mathfrak{X}$ formed by mappings $\boldsymbol{x} : \Omega \to \mathbb{R}^n$ which are constant almost everywhere, i.e.,

$$\mathfrak{L} := \left\{ \boldsymbol{x} \in \mathfrak{X} : \boldsymbol{x}(\omega) \equiv x \text{ for some } x \in \mathbb{R}^n \right\},$$

where $\boldsymbol{x}(\omega) \equiv x$ means that $\boldsymbol{x}(\omega) = x$ for a.e. $\omega \in \Omega$.

Consider the dual[9] $\mathfrak{X}^* := \mathcal{L}_q(\Omega, \mathcal{F}, P; \mathbb{R}^n)$ of the space $\mathfrak{X}$ and define the scalar product (bilinear form)

$$\langle \boldsymbol{\lambda}, \boldsymbol{x} \rangle := \mathbb{E}\left[\boldsymbol{\lambda}^{\mathsf{T}} \boldsymbol{x}\right] = \int_{\Omega} \boldsymbol{\lambda}(\omega)^{\mathsf{T}} \boldsymbol{x}(\omega) dP(\omega), \quad \boldsymbol{\lambda} \in \mathfrak{X}^*, \ \boldsymbol{x} \in \mathfrak{X}.$$

Also, consider the projection operator $\boldsymbol{P} : \mathfrak{X} \to \mathfrak{L}$ defined as $[\boldsymbol{P}\boldsymbol{x}](\omega) \equiv \mathbb{E}[\boldsymbol{x}]$. Clearly the space $\mathfrak{L}$ is formed by such $\boldsymbol{x} \in \mathfrak{X}$ that $\boldsymbol{P}\boldsymbol{x} = \boldsymbol{x}$. Note that

$$\langle \boldsymbol{\lambda}, \boldsymbol{P}\boldsymbol{x} \rangle = \mathbb{E}\left[\boldsymbol{\lambda}\right]^{\mathsf{T}} \mathbb{E}\left[\boldsymbol{x}\right] = \langle \boldsymbol{P}^* \boldsymbol{\lambda}, \boldsymbol{x} \rangle,$$

where $\boldsymbol{P}^*$ is a projection operator $[\boldsymbol{P}^*\boldsymbol{\lambda}](\omega) \equiv \mathbb{E}[\boldsymbol{\lambda}]$ from $\mathfrak{X}^*$ onto its subspace formed by constant a.e. mappings. In particular, if $p = 2$, then $\mathfrak{X}^* = \mathfrak{X}$ and $\boldsymbol{P}^* = \boldsymbol{P}$.

With problem (2.91) is associated the following Lagrangian:

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) := \mathbb{E}[\bar{F}(\boldsymbol{x}(\omega), \omega)] + \mathbb{E}\left[\boldsymbol{\lambda}^{\mathsf{T}}(\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}])\right].$$

Note that

$$\mathbb{E}\left[\boldsymbol{\lambda}^{\mathsf{T}}(\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}])\right] = \langle \boldsymbol{\lambda}, \boldsymbol{x} - \boldsymbol{P}\boldsymbol{x} \rangle = \langle \boldsymbol{\lambda} - \boldsymbol{P}^*\boldsymbol{\lambda}, \boldsymbol{x} \rangle,$$

and $\boldsymbol{\lambda} - \boldsymbol{P}^*\boldsymbol{\lambda}$ does not change by adding a constant to $\boldsymbol{\lambda}(\cdot)$. Therefore we can set $\boldsymbol{P}^*\boldsymbol{\lambda} = 0$, in which case

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = \mathbb{E}\left[\bar{F}(\boldsymbol{x}(\omega), \omega) + \boldsymbol{\lambda}(\omega)^{\mathsf{T}} \boldsymbol{x}(\omega)\right] \quad \text{for } \mathbb{E}[\boldsymbol{\lambda}] = 0. \tag{2.92}$$

This leads to the following dual of problem (2.90):

$$\operatorname*{Max}_{\boldsymbol{\lambda} \in \mathfrak{X}^*} \left\{ D(\boldsymbol{\lambda}) := \inf_{\boldsymbol{x} \in \mathfrak{X}} L(\boldsymbol{x}, \boldsymbol{\lambda}) \right\} \text{ s.t. } \mathbb{E}[\boldsymbol{\lambda}] = 0. \tag{2.93}$$

In case of finitely many scenarios, the above dual is the same as the dual problem (2.88).

By the interchangeability principle (Theorem 7.80) we have

$$\inf_{\boldsymbol{x} \in \mathfrak{X}} \mathbb{E}\left[\bar{F}(\boldsymbol{x}(\omega), \omega) + \boldsymbol{\lambda}(\omega)^{\mathsf{T}} \boldsymbol{x}(\omega)\right] = \mathbb{E}\left[\inf_{x \in \mathbb{R}^n} \left(\bar{F}(x, \omega) + \boldsymbol{\lambda}(\omega)^{\mathsf{T}} x\right)\right].$$

Consequently,

$$D(\boldsymbol{\lambda}) = \mathbb{E}[\mathfrak{D}_\omega(\boldsymbol{\lambda}(\omega))],$$

where $\mathfrak{D}_\omega : \mathbb{R}^n \to \overline{\mathbb{R}}$ is defined as

$$\mathfrak{D}_\omega(\lambda) := \inf_{x \in \mathbb{R}^n} \left(\lambda^{\mathsf{T}} x + \bar{F}_\omega(x)\right) = -\sup_{x \in \mathbb{R}^n} \left(-\lambda^{\mathsf{T}} x - \bar{F}_\omega(x)\right) = -\bar{F}^*_\omega(-\lambda). \tag{2.94}$$

That is, in order to calculate the dual function $D(\boldsymbol{\lambda})$ one needs to solve for every $\omega \in \Omega$ the finite dimensional optimization problem (2.94) and then to integrate the optimal values obtained.

[9]Recall that $1/p + 1/q = 1$ for $p, q \in (1, +\infty)$. If $p = 1$, then $q = +\infty$. Also for $p = +\infty$ we use $q = 1$. This results in a certain abuse of notation since the space $\mathfrak{X} = \mathcal{L}_\infty(\Omega, \mathcal{F}, P; \mathbb{R}^n)$ is not reflexive and $\mathfrak{X}^* = \mathcal{L}_1(\Omega, \mathcal{F}, P; \mathbb{R}^n)$ is smaller than its dual. Note also that if $\boldsymbol{x} \in \mathcal{L}_p(\Omega, \mathcal{F}, P; \mathbb{R}^n)$, then its expectation $\mathbb{E}[\boldsymbol{x}] = \int_\Omega \boldsymbol{x}(\omega) dP(\omega)$ is well defined and is an element of vector space $\mathbb{R}^n$.

By the general theory, we have that the optimal value of problem (2.91), which is the same as the optimal value of problem (2.90), is greater than or equal to the optimal value of its dual (2.93). We also have that there is no duality gap between problem (2.91) and its dual (2.93) and both problems have optimal solutions $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{\lambda}}$, respectively, iff $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\lambda}})$ is a saddle point of the Lagrangian defined in (2.92). By definition a point $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\lambda}}) \in \mathfrak{X} \times \mathfrak{X}^*$ is a saddle point of the Lagrangian iff

$$\bar{\boldsymbol{x}} \in \arg\min_{\boldsymbol{x}\in\mathfrak{L}} L(\boldsymbol{x}, \bar{\boldsymbol{\lambda}}) \quad\text{and}\quad \bar{\boldsymbol{\lambda}} \in \arg\max_{\boldsymbol{\lambda}:\mathbb{E}[\boldsymbol{\lambda}]=0} L(\bar{\boldsymbol{x}}, \boldsymbol{\lambda}). \tag{2.95}$$

By the interchangeability principle (see (7.247) of Theorem 7.80), we have that the first condition in (2.95) can be written in the following equivalent form:

$$\bar{\boldsymbol{x}}(\omega) \equiv \bar{x} \quad\text{and}\quad \bar{x} \in \arg\min_{x\in\mathbb{R}^n} \left\{ \bar{F}(x,\omega) + \bar{\boldsymbol{\lambda}}(\omega)^{\mathsf{T}} x \right\} \quad \text{a.e. } \omega \in \Omega. \tag{2.96}$$

Since $\bar{\boldsymbol{x}}(\omega) \equiv \bar{x}$, the second condition in (2.95) means that $\mathbb{E}[\bar{\boldsymbol{\lambda}}] = 0$.

Let us assume now that the considered problem is *convex*, i.e., the set X is convex (and closed) and $F_\omega(\cdot)$ is a convex function for a.e. $\omega \in \Omega$. It follows that $\bar{F}_\omega(\cdot)$ is a convex function for a.e. $\omega \in \Omega$. Then the second condition in (2.96) holds iff $\bar{\boldsymbol{\lambda}}(\omega) \in -\partial \bar{F}_\omega(\bar{x})$ for a.e. $\omega \in \Omega$. Together with condition $\mathbb{E}[\bar{\boldsymbol{\lambda}}] = 0$ this means that

$$0 \in \mathbb{E}\left[\partial \bar{F}_\omega(\bar{x})\right]. \tag{2.97}$$

It follows that the Lagrangian has a saddle point iff there exists $\bar{x} \in \mathbb{R}^n$ satisfying condition (2.97). We obtain the following result.

Theorem 2.25. *Suppose that the function $F(x,\omega)$ is random lower semicontinuous, the set X is convex and closed, and for a.e. $\omega \in \Omega$ the function $F(\cdot,\omega)$ is convex. Then there is no duality gap between problems* (2.90) *and* (2.93) *and both problems have optimal solutions iff there exists $\bar{x} \in \mathbb{R}^n$ satisfying condition* (2.97)*. In that case, $\bar{x}$ is an optimal solution of* (2.90) *and a measurable selection $\bar{\boldsymbol{\lambda}}(\omega) \in -\partial \bar{F}_\omega(\bar{x})$ such that $\mathbb{E}[\bar{\boldsymbol{\lambda}}] = 0$ is an optimal solution of* (2.93)*.*

Recall that the inclusion $\mathbb{E}\left[\partial \bar{F}_\omega(\bar{x})\right] \subset \partial \bar{f}(\bar{x})$ always holds (see (7.125) in the proof of Theorem 7.47). Therefore, condition (2.97) implies that $0 \in \partial \bar{f}(\bar{x})$, which in turn implies that $\bar{x}$ is an optimal solution of (2.90). Conversely, if $\bar{x}$ is an optimal solution of (2.90), then $0 \in \partial \bar{f}(\bar{x})$, and if in addition $\mathbb{E}\left[\partial \bar{F}_\omega(\bar{x})\right] = \partial \bar{f}(\bar{x})$, then (2.97) follows. Therefore, Theorems 2.25 and 7.47 imply the following result.

Theorem 2.26. *Suppose that* (i) *the function $F(x,\omega)$ is random lower semicontinuous,* (ii) *the set X is convex and closed,* (iii) *for a.e. $\omega \in \Omega$ the function $F(\cdot,\omega)$ is convex, and* (iv) *problem* (2.90) *possesses an optimal solution $\bar{x}$ such that $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$. Then there is no duality gap between problems* (2.90) *and* (2.93)*, the dual problem* (2.93) *has an optimal solution $\bar{\boldsymbol{\lambda}}$, and the constant mapping $\bar{\boldsymbol{x}}(\omega) \equiv \bar{x}$ is an optimal solution of the problem*

$$\operatorname*{Min}_{\boldsymbol{x}\in\mathfrak{X}} \mathbb{E}\left[\bar{F}(\boldsymbol{x}(\omega),\omega) + \bar{\boldsymbol{\lambda}}(\omega)^{\mathsf{T}} \boldsymbol{x}(\omega)\right].$$

Proof. Since $\bar{x}$ is an optimal solution of problem (2.90), we have that $\bar{x} \in X$ and $f(\bar{x})$ is finite. Moreover, since $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$ and f is convex, it follows that f is proper and $\mathcal{N}_{\mathrm{dom} f}(\bar{x}) = \{0\}$. Therefore, it follows by Theorem 7.47 that $\mathbb{E}[\partial F_\omega(\bar{x})] = \partial f(\bar{x})$. Furthermore, since $\bar{x} \in \mathrm{int}(\mathrm{dom} f)$, we have that $\partial \bar{f}(\bar{x}) = \partial f(\bar{x}) + \mathcal{N}_X(\bar{x})$, and hence $\mathbb{E}\left[\partial \bar{F}_\omega(\bar{x})\right] = \partial \bar{f}(\bar{x})$. By optimality of $\bar{x}$, we also have that $0 \in \partial \bar{f}(\bar{x})$. Consequently, $0 \in \mathbb{E}\left[\partial \bar{F}_\omega(\bar{x})\right]$, and hence the proof can be completed by applying Theorem 2.25. □

If X is a subset of $\mathrm{int}(\mathrm{dom} f)$, then any point $x \in X$ is an interior point of $\mathrm{dom} f$. In that case, condition (iv) of the above theorem is reduced to existence of an optimal solution. The condition $X \subset \mathrm{int}(\mathrm{dom} f)$ means that $f(x) < +\infty$ for every x in a neighborhood of the set X. This requirement is slightly stronger than the condition of relatively complete recourse.

Example 2.27 (Capacity Expansion Continued). Let us consider the capacity expansion problem of Examples 2.4 and 2.13. Suppose that $\bar{x}$ is the optimal first-stage decision and let $\bar{y}_{ij}(\xi)$ be the corresponding optimal second-stage decisions. The scenario problem has the form

$$\begin{aligned} \mathrm{Min} \; & \sum_{(i,j)\in\mathcal{A}} \left[(c_{ij} + \lambda_{ij}(\xi))x_{ij} + q_{ij} y_{ij}\right] \\ \text{s.t.} \; & \sum_{(i,j)\in\mathcal{A}_+(n)} y_{ij} - \sum_{(i,j)\in\mathcal{A}_-(n)} y_{ij} = \xi_n, \quad n \in \mathcal{N}, \\ & 0 \le y_{ij} \le x_{ij}, \quad (i,j) \in \mathcal{A}. \end{aligned}$$

From Example 2.13 we know that there exist random node potentials $\mu_n(\xi)$, $n \in \mathcal{N}$, such that for all $\xi \in \Xi$ we have $\mu(\xi) \in \mathcal{M}(\bar{x}, \xi)$, and conditions (2.42)–(2.43) are satisfied. Also, the random variables $g_{ij}(\xi) = -\max\{0, \mu_i(\xi) - \mu_j(\xi) - q_{ij}\}$ are the corresponding subgradients of the second stage cost. Define

$$\lambda_{ij}(\xi) = \max\{0, \mu_i(\xi) - \mu_j(\xi) - q_{ij}\} - \int_\Xi \max\{0, \mu_i(\xi) - \mu_j(\xi) - q_{ij}\}\, P(d\xi), \quad (i,j) \in \mathcal{A}.$$

We can easily verify that $x_{ij}(\xi) = \bar{x}_{ij}$ and $\bar{y}_{ij}(\xi)$, $(i,j) \in \mathcal{A}$, are an optimal solution of the scenario problem, because the first term of λ_{ij} cancels with the subgradient $g_{ij}(\xi)$, while the second term satisfies the optimality conditions (2.42)–(2.43). Moreover, $\mathbb{E}[\lambda] = 0$ by construction. ■

2.4.4 Value of Perfect Information

Consider the following relaxation of the two-stage problem (2.61)–(2.62):

$$\underset{\boldsymbol{x}\in\mathfrak{X}}{\mathrm{Min}} \; \mathbb{E}[\bar{F}(\boldsymbol{x}(\omega), \omega)]. \tag{2.98}$$

This relaxation is obtained by removing the nonanticipativity constraint from the formulation (2.91) of the first-stage problem. By the interchangeability principle (Theorem 7.80) we have that the optimal value of the above problem (2.98) is equal to $\mathbb{E}\left[\inf_{x\in\mathbb{R}^n} \bar{F}(x,\omega)\right]$. The value $\inf_{x\in\mathbb{R}^n} \bar{F}(x,\omega)$ is equal to the optimal value of the problem

$$\underset{x\in X,\; y\in\mathcal{G}(x,\omega)}{\mathrm{Min}} \; g(x, y, \omega). \tag{2.99}$$

That is, the optimal value of problem (2.98) is obtained by solving problems of the form (2.99), one for each $\omega \in \Omega$, and then taking the expectation of the calculated optimal values.

Solving problems of the form (2.99) makes sense if we have perfect information about the data, i.e., the scenario $\omega \in \Omega$ is known at the time when the first-stage decision should be made. The problem (2.99) is deterministic, e.g., in the case of two-stage linear program (2.1)–(2.2) it takes the form

$$\operatorname*{Min}_{x \geq 0, y \geq 0} c^{\mathsf{T}} x + q^{\mathsf{T}} y \ \text{ s.t. } Ax = b,\ Tx + Wy = h.$$

An optimal solution of the second-stage problem (2.99) depends on $\omega \in \Omega$ and is called the *wait-and-see* solution.

We have that for any $x \in X$ and $\omega \in \Omega$, the inequality $F(x, \omega) \geq \inf_{x \in X} F(x, \omega)$ clearly holds, and hence $\mathbb{E}[F(x, \omega)] \geq \mathbb{E}\,[\inf_{x \in X} F(x, \omega)]$. It follows that

$$\inf_{x \in X} \mathbb{E}[F(x, \omega)] \geq \mathbb{E}\left[\inf_{x \in X} F(x, \omega)\right]. \tag{2.100}$$

Another way to view the above inequality is to observe that problem (2.98) is a relaxation of the corresponding two-stage stochastic problem, which of course implies (2.100).

Suppose that the two-stage problem has an optimal solution $\bar{x} \in \arg\min_{x \in X} \mathbb{E}[F(x, \omega)]$. As $F(\bar{x}, \omega) - \inf_{x \in X} F(x, \omega) \geq 0$ for all $\omega \in \Omega$, we conclude that

$$\mathbb{E}[F(\bar{x}, \omega)] = \mathbb{E}\left[\inf_{x \in X} F(x, \omega)\right] \tag{2.101}$$

iff $F(\bar{x}, \omega) = \inf_{x \in X} F(x, \omega)$ w.p. 1. That is, equality in (2.101) holds iff

$$F(\bar{x}, \omega) = \inf_{x \in X} F(x, \omega) \ \text{ for a.e. } \omega \in \Omega. \tag{2.102}$$

In particular, this happens if $\bar{F}_\omega(x)$ has a minimizer independent of $\omega \in \Omega$. This, of course, may happen only in rather specific situations.

The difference $F(\bar{x}, \omega) - \inf_{x \in X} F(x, \omega)$ represents the *value of perfect information* of knowing ω. Consequently

$$\text{EVPI} := \inf_{x \in X} \mathbb{E}[F(x, \omega)] - \mathbb{E}\left[\inf_{x \in X} F(x, \omega)\right]$$

is called the *expected value of perfect information*. It follows from (2.100) that EVPI is always nonnegative and EVPI $= 0$ iff condition (2.102) holds.

Exercises

2.1. Consider the assembly problem discussed in section 1.3.1 in two cases:

(i) The demand which is not satisfied from the preordered quantities of parts is lost.

(ii) All demand has to be satisfied by making additional orders of the missing parts. In this case, the cost of each additionally ordered part j is $r_j > c_j$.

For each of these cases describe the subdifferential of the recourse cost and of the expected recourse cost.

2.2. A transportation company has n depots among which they send cargo. The demand for transportation between depot i and depot $j \neq i$ is modeled as a random variable D_{ij}. The total capacity of vehicles currently available at depot i is denoted s_i, $i = 1, \dots, n$. The company considers repositioning its fleet to better prepare to the uncertain demand. It costs c_{ij} to move a unit of capacity from location i to location j. After repositioning, the realization of the random vector D is observed, and the demand is served, up to the limit determined by the transportation capacity available at each location. The profit from transporting a unit of cargo from location i to location j is equal q_{ij}. If the total demand at location i exceeds the capacity available at location i, the excessive demand is lost. It is up to the company to decide how much of each demand D_{ij} be served, and which part will remain unsatisfied. For simplicity, we consider all capacity and transportation quantities as continuous variables.

(a) Formulate the problem of maximizing the expected profit as a two-stage stochastic programming problem.

(b) Describe the subdifferential of the recourse cost and the expected recourse cost.

2.3. Show that the function $s_q(\cdot)$, defined in (2.4), is convex.

2.4. Consider the optimal value $Q(x, \xi)$ of the second-stage problem (2.2). Show that $Q(\cdot, \xi)$ is differentiable at a point x iff the dual problem (2.3) has a unique optimal solution $\bar{\pi}$, in which case $\nabla_x Q(x, \xi) = -T^{\mathsf{T}}\bar{\pi}$.

2.5. Consider the two-stage problem (2.1)–(2.2) with fixed recourse. Show that the following conditions are equivalent: (i) problem (2.1)–(2.2) has complete recourse, (ii) the feasible set $\Pi(q)$ of the dual problem is bounded for every q, and (iii) the system $W^{\mathsf{T}}\pi \leq 0$ has only one solution $\pi = 0$.

2.6. Show that if random vector ξ has a finite support, then condition (2.24) is necessary and sufficient for relatively complete recourse.

2.7. Show that the conjugate function of a polyhedral function is also polyhedral.

2.8. Show that if $Q(x, \omega)$ is finite, then the set $\mathfrak{D}(x, \omega)$ of optimal solutions of problem (2.46) is a nonempty convex closed polyhedron.

2.9. Consider problem (2.63) and its optimal value $F(x, \omega)$. Show that $F(x, \omega)$ is convex in x if $\bar{g}(x, y, \omega)$ is convex in (x, y). Show that the indicator function $\mathbb{I}_{\mathcal{G}_\omega(x)}(y)$ is convex in (x, y) iff condition (2.68) holds for any $t \in [0, 1]$.

2.10. Show that equation (2.86) implies that $\langle \boldsymbol{x} - \boldsymbol{P}\boldsymbol{x}, \boldsymbol{y} \rangle = 0$ for any $\boldsymbol{x} \in \mathfrak{X}$ and $\boldsymbol{y} \in \mathfrak{L}$, i.e., that $\boldsymbol{P}$ is the orthogonal projection of $\mathfrak{X}$ onto $\mathfrak{L}$.

2.11. Derive the form of the dual problem for the linear two-stage stochastic programming problem in form (2.80) with nonanticipativity constraints (2.87).