# Chapter 3
# SLP models with recourse

For various SLP models with recourse, we present in this chapter properties which are relevant for the particular solution methods developed for various model types, to be discussed later on.

## 3.1 The general multi-stage SLP

As briefly sketched in Section 1.1 an SLP with recourse is a dynamic decision model with $T \geq 2$ stages, as illustrated in Fig. 3.1,

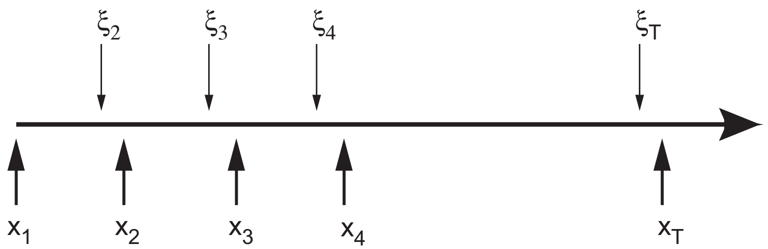


**Fig. 3.1** Dynamic decision structure.

where for feasibility sets, emerging stagewise during the horizon $\mathscr{T} = \{1, 2, \cdots, T\}$,

$$\mathscr{B}_t(x_1, \cdots, x_{t-1}; \xi_2, \cdots, \xi_t),\ t \in \mathscr{T},$$

we take successively

– a first stage decision $x_1 \in \mathscr{B}_1 \subset \mathbb{R}^{n_1}$; then, after observing the realization of a random variable (or vector) $\xi_2$,
– a second stage decision $x_2(x_1; \xi_2) \in \mathscr{B}_2(x_1; \xi_2) \subset \mathbb{R}^{n_2}$; then after observing the realization of a further random variable (or vector) $\xi_3$,

P. Kall, J. Mayer, *Stochastic Linear Programming,* 2nd edition, International Series in Operations Research & Management Science 156, DOI 10.1007/978-1-4419-7729-8_3,

- a third stage decision $x_3(x_1,x_2;\xi_2,\xi_3) \in \mathscr{B}_3(x_1,x_2;\xi_2,\xi_3) \subset \mathbb{R}^{n_3}$; and so on until, after observing the realization of $\xi_T$, finally
- a $T$-th stage decision
  $x_T(x_1,\cdots,x_{T-1};\xi_2,\cdots,\xi_T) \in \mathscr{B}_T(x_1,\cdots,x_{T-1};\xi_2,\cdots,\xi_T) \subset \mathbb{R}^{n_T}$.

Here the feasibility set $\mathscr{B}_t(x_1,\cdots,x_{t-1};\xi_2,\cdots,\xi_t)$ for $x_t$ is given by (random) linear constraints, depending on the previous decisions $x_1,\cdots,x_{t-1}$ and the observations of $\xi_2,\cdots,\xi_t$.

For each stage $t$ the decision $x_t(x_1,\cdots,x_{t-1};\xi_2,\cdots,\xi_t)$ involves the $t$-th stage objective value $c_t^{\mathrm{T}}(\xi_2,\cdots,\xi_t)x_t(x_1,\cdots,x_{t-1};\xi_2,\cdots,\xi_t)$, and the goal is to minimize the expected value of the sum of these $T$ objectives.

More precisely, with any set $\Omega \neq \emptyset$, some $\sigma$-algebra $\mathscr{G}$ of subsets of $\Omega$ and a probability measure $P : \mathscr{G} \to [0,1]$, the general model may be stated as follows: Given the probability space $(\Omega,\mathscr{G},P)$, random vectors $\xi_t : \Omega \longrightarrow \mathbb{R}^{r_t}$, and the probability distribution $\mathbb{P}_\xi$ induced by $\xi = (\xi_2^{\mathrm{T}},\cdots,\xi_T^{\mathrm{T}})^{\mathrm{T}} : \Omega \longrightarrow \mathbb{R}^R$, $R = r_2+\cdots+r_T$, on the Borel $\sigma$-field of $\mathbb{R}^R$, with $\zeta_t = (\xi_2^{\mathrm{T}},\cdots,\xi_t^{\mathrm{T}})^{\mathrm{T}}$ being the state variable at stage $t$, the multi-stage stochastic linear program (MSLP) reads as

$$\left.\begin{array}{ll}
\min\{c_1^{\mathrm{T}}x_1 + \mathbb{E}\sum_{t=2}^{T} c_t^{\mathrm{T}}(\zeta_t)x_t(\zeta_t)\} & \\
A_{11}x_1 \qquad\qquad\qquad\qquad = b_1 & \\
A_{t1}(\zeta_t)x_1 + \sum_{\tau=2}^{t} A_{t\tau}(\zeta_t)x_\tau(\zeta_\tau) = b_t(\zeta_t) & \text{a.s., } t = 2,\cdots,T, \\
\qquad x_1 \geq 0,\;\; x_t(\zeta_t) \geq 0 & \text{a.s., } t = 2,\cdots,T,
\end{array}\right\} \tag{3.1}$$

where $x_t : \mathbb{R}^{r_2+\cdots+r_t} \longrightarrow \mathbb{R}^{n_t}$ is to be Borel measurable, implying that $x_t(\zeta_t(\cdot)) : \Omega \longrightarrow \mathbb{R}^{n_t}$ is $\mathscr{F}_t$-measurable, with $\mathscr{F}_t = \sigma(\zeta_t) \subset \mathscr{G}$, the $\sigma$-algebra in $\Omega$ generated at stage $t$ by $\{\zeta_t^{-1}[M] \mid M \in \mathbb{B}^{r_2+\cdots+r_t}\}$. With $\zeta_1 \equiv \xi_1 = const$ and therefore $\mathscr{F}_1 = \{\emptyset,\Omega\}$, it follows that $\mathscr{F}_t \subset \mathscr{F}_{t+1}$ for $t = 1,\cdots,T-1$, such that $\mathscr{F} = \{\mathscr{F}_1,\mathscr{F}_2,\cdots,\mathscr{F}_T\}$ is a *filtration*. With $x_t(\zeta_t(\cdot))$ being $\mathscr{F}_t$-measurable for $t = 1,\cdots,T$, the policy $\{x_t(\zeta_t(\cdot)); t = 1,\cdots,T\}$ is said to be $\mathscr{F}$*-adapted* or else *nonanticipative*.

The $\xi_t : \Omega \longrightarrow \mathbb{R}^{r_t}$ as random vectors defined on the probability space $\{\Omega,\mathscr{G},P\}$ are obviously $\mathscr{G}$-measurable. According to the definition in (2.6) on page 73, we say that $\xi_t \in \mathscr{L}_{r_t}^2 := \mathscr{L}_{r_t}^2(\Omega,\mathscr{G},\mathbb{R}^{r_t})$ if, in addition, the $\xi_t$ are square integrable, i.e. if $\int_\Omega \|\xi_t(\omega)\|^2 P(d\omega)$ exists. In particular, for any arbitrary $\mathscr{F}_t$-simple function $\gamma_t(\omega) := \sum_{i=1}^K g_i \cdot \chi_{M_i}(\omega)$ with $g_i \in \mathbb{R}^{r_t}$, $\chi_{M_i}(\omega) = 1$ if $\omega \in M_i$ and $\chi_{M_i}(\omega) = 0$ otherwise, $M_i \in \mathscr{F}_t$, $M_i \cap M_j = \emptyset$ for $i \neq j$, and $\cup_{i=1}^K M_i = \Omega$, it obviously follows that $\gamma_t \in \mathscr{L}_{r_t}^2(\Omega,\mathbb{R}^{r_t})$.

**Assumption 3.1.** *Let*

- $\xi_t \in \mathscr{L}_{r_t}^2 := \mathscr{L}_{r_t}^2(\Omega,\mathscr{G},\mathbb{R}^{r_t})\ \forall t$,
- $A_{t\tau}(\cdot), b_t(\cdot), c_t(\cdot)$ *be linear affine in* $\zeta_t$ *(and therefore* $\mathscr{F}_t$*-measurable), where* $A_{t\tau}(\cdot)$ *is a* $m_t \times n_\tau$*-matrix.*

Due to this assumption, also the elements of $A_{t\tau}(\cdot), b_t(\cdot), c_t(\cdot)$ are square-integrable with respect to $P$. Hence, requiring that $\xi_t \in \mathscr{L}^2_{r_t}$ $\forall t$ holds, Schwarz's inequality (see e.g. Zaanen [353]) implies in particular that $\mathbb{E}[c_t^{\mathrm{T}}(\zeta_t)x_t(\zeta_t)]$, $t = 2, \cdots, T$, exist, such that problem (3.1) is well defined.
Sometimes the following reformulation of (3.1) may be convenient: Given

- a probability space $(\Omega, \mathscr{G}, P)$;
- $\mathscr{F}_t, t = 1, \cdots, T$, being $\sigma$-algebras such that $\mathscr{F}_t \subset \mathscr{G}$ $\forall t$ and $\mathscr{F}_t \subset \mathscr{F}_{t+1}$ for $t = 1, \cdots, T-1$ (i.e. $\{\mathscr{F}_t \mid t = 1, \cdots, T\}$ being a filtration);
- $\mathscr{F} := \{\mathscr{F}_1, \cdots, \mathscr{F}_T\}$, where possibly, but not necessarily, $\mathscr{F}_T = \mathscr{G}$;
- $X_t$ a linear subspace of $\mathscr{L}^2_{n_t}$ (with respect to $(\Omega, \mathscr{G}, P)$), including the set of $\mathscr{F}_t$-simple functions;
- $M_t$ the set of $\mathscr{F}_t$-measurable functions $\Omega \longrightarrow \mathbb{R}^{n_t}$ and hence, $X_t \cap M_t$ being a closed linear subspace of $X_t$;

then problem (3.1) may be restated as

$$\left.\begin{array}{l} \min \mathbb{E}\left\{\sum_{t=1}^{T} c_t^{\mathrm{T}} x_t\right\} \\ \left.\begin{array}{rl} \sum_{\tau=1}^{t} A_{t\tau} x_\tau &= b_t \text{ a.s.} \\ x_t &\geq 0 \text{ a.s.} \\ x_t &\in X_t \cap M_t \end{array}\right\} t = 1, \cdots, T, \end{array}\right\} \tag{3.2}$$

with $A_{t\tau}, b_t, c_t$ assumed to be $\mathscr{F}_t$-measurable for $1 \leq \tau \leq t$, $t = 1, \cdots, T$, and to have finite second moments, as implied by Assumption 3.1. (remember: $\mathscr{F}_1 = \{\emptyset, \Omega\}$, such that $A_{11}, b_1, c_1$ are constant).

Following S.E. Wright [347] various aggregated problems may be derived from (3.2) by using coarser information structures, chosen as subfiltrations $\widehat{\mathscr{F}} = \{\widehat{\mathscr{F}_t}\}$, $\widehat{\mathscr{F}_t} \subset \widehat{\mathscr{F}}_{t+1}$, such that $\widehat{\mathscr{F}_t} \subseteq \mathscr{F}_t$, $\forall t$, instead of the original filtration $\mathscr{F} = \{\mathscr{F}_t\}$, $\mathscr{F}_t \subset \mathscr{F}_{t+1}$, $t = 1, \cdots, T-1$.
Denoting problem (3.2) as $\mathscr{P}(\mathscr{F}, \mathscr{F})$, we then may consider

- the *decision-aggregated* problem $\mathscr{P}(\widehat{\mathscr{F}}, \mathscr{F})$,

$$\left.\begin{array}{l} \min \mathbb{E}\left\{\sum_{t=1}^{T} c_t x_t\right\} \\ \left.\begin{array}{rl} \sum_{\tau=1}^{t} A_{t\tau} x_\tau &= b_t \text{ a.s.} \\ x_t &\geq 0 \text{ a.s.} \\ x_t &\in X_t \cap \widehat{M}_t \end{array}\right\} t = 1, \cdots, T, \end{array}\right\} \tag{3.3}$$

where $\widehat{M}_t$ is the set of $\widehat{\mathscr{F}_t}$-measurable functions $\Omega \longrightarrow \mathbb{R}^{n_t}$, thus requiring that $x = (x_1^{\mathrm{T}}, \cdots, x_T^{\mathrm{T}})^{\mathrm{T}}$ is $\widehat{\mathscr{F}}$-adapted;

– the *constraint-aggregated* problem $\mathscr{P}(\mathscr{F},\widehat{\mathscr{F}})$,

$$\left.\begin{array}{l} \min \mathbb{E}\left\{\sum_{t=1}^{T} c_t x_t\right\} \\ \left.\begin{array}{ll} \mathbb{E}\left\{\sum_{\tau=1}^{t} A_{t\tau}x_\tau \,\middle|\, \widehat{\mathscr{F}_t}\right\} = \mathbb{E}\left\{b_t \,\middle|\, \widehat{\mathscr{F}_t}\right\} & \text{a.s.} \\ x_t \geq 0 & \text{a.s.} \\ x_t \in X_t \cap M_t & \end{array}\right\} t = 1,\cdots,T, \end{array}\right\} \tag{3.4}$$

i.e. $x$ is $\mathscr{F}$-adapted as in (3.2), and the constraints are stated in conditional expectation given $\widehat{\mathscr{F}_t}$;

– and the *fully aggregated* problem $\mathscr{P}(\widehat{\mathscr{F}},\widehat{\mathscr{F}})$ defined as:

$$\left.\begin{array}{lll} \min \mathbb{E}\left\{\sum_{t=1}^{T} \mathbb{E}[c_t \mid \widehat{\mathscr{F}_t}]x_t\right\} & & \\ \mathbb{E}\left\{\sum_{\tau=1}^{t} A_{t\tau}x_\tau \,\middle|\, \widehat{\mathscr{F}_t}\right\} = \mathbb{E}\left\{b_t \,\middle|\, \widehat{\mathscr{F}_t}\right\} & \text{a.s. } \forall t \\ x_t \geq 0 & \text{a.s. } \forall t \\ x_t \in X_t \cap \widehat{M}_t & \forall t\,. \end{array}\right\} \tag{3.5}$$

Observe that by Assumption 3.1. the expected values

$$\mathbb{E}\left\{\sum_{\tau=1}^{t} A_{t\tau}x_\tau\right\} \quad \text{and} \quad \mathbb{E}\{b_t\}$$

exist and hence, the conditional expectations in (3.4) and (3.5),

$$\mathbb{E}\left\{\sum_{\tau=1}^{t} A_{t\tau}x_\tau \,\middle|\, \widehat{\mathscr{F}_t}\right\} \quad \text{and} \quad \mathbb{E}\left\{b_t \,\middle|\, \widehat{\mathscr{F}_t}\right\},$$

are a.s. uniquely determined and $\widehat{\mathscr{F}_t}$-measurable due to the Radon-Nikodym theorem (see e.g. Halmos [131]).

Denoting for the above problems $\mathscr{P}(\mathscr{F},\mathscr{F})$, $\mathscr{P}(\widehat{\mathscr{F}},\mathscr{F})$, $\mathscr{P}(\mathscr{F},\widehat{\mathscr{F}})$, $\mathscr{P}(\widehat{\mathscr{F}},\widehat{\mathscr{F}})$

– their feasible sets by $\mathscr{B}(\mathscr{F},\mathscr{F})$, $\mathscr{B}(\widehat{\mathscr{F}},\mathscr{F})$, $\mathscr{B}(\mathscr{F},\widehat{\mathscr{F}})$ and $\mathscr{B}(\widehat{\mathscr{F}},\widehat{\mathscr{F}})$, and
– their optimal values by $\inf(\mathscr{P}(\mathscr{F},\mathscr{F}))$, $\inf(\mathscr{P}(\widehat{\mathscr{F}},\mathscr{F}))$, $\inf(\mathscr{P}(\mathscr{F},\widehat{\mathscr{F}}))$ and $\inf(\mathscr{P}(\widehat{\mathscr{F}},\widehat{\mathscr{F}}))$,

respectively, and with the usual convention that $\inf\{\varphi(x) \mid x \in \mathscr{B}\} = \infty$ if $\mathscr{B} = \emptyset$, the following relations between the above problems are mentioned in S.E. Wright [347]:

**Proposition 3.1.** *For the feasible sets of the above problems hold the inclusions*

$$\begin{array}{ll} \mathscr{B}(\mathscr{F},\mathscr{F}) \supseteq \mathscr{B}(\widehat{\mathscr{F}},\mathscr{F}) & \mathscr{B}(\mathscr{F},\mathscr{F}) \subseteq \mathscr{B}(\mathscr{F},\widehat{\mathscr{F}}) \\ \mathscr{B}(\widehat{\mathscr{F}},\mathscr{F}) \subseteq \mathscr{B}(\widehat{\mathscr{F}},\widehat{\mathscr{F}}) & \mathscr{B}(\mathscr{F},\widehat{\mathscr{F}}) \supseteq \mathscr{B}(\widehat{\mathscr{F}},\widehat{\mathscr{F}}), \end{array}$$

*implying for the corresponding optimal values the inequalities*

$$\begin{array}{l} \inf(\mathscr{P}(\mathscr{F},\widehat{\mathscr{F}})) \le \inf(\mathscr{P}(\mathscr{F},\mathscr{F})) \le \inf(\mathscr{P}(\widehat{\mathscr{F}},\mathscr{F})) \\ \inf(\mathscr{P}(\mathscr{F},\widehat{\mathscr{F}})) \le \inf(\mathscr{P}(\widehat{\mathscr{F}},\widehat{\mathscr{F}})) \le \inf(\mathscr{P}(\widehat{\mathscr{F}},\mathscr{F})). \end{array}$$

*Proof:* The above inclusions result from the following observations:

$\underline{\mathscr{B}(\mathscr{F},\mathscr{F}) \supseteq \mathscr{B}(\widehat{\mathscr{F}},\mathscr{F})}$: Any $\{x_t\} \in \mathscr{B}(\widehat{\mathscr{F}},\mathscr{F})$ satisfies the constraints of (3.3) and hence in particular the conditions $x_t \in X_t \cap \widehat{M}_t\ \forall t$. Since $\widehat{\mathscr{F}_t} \subseteq \mathscr{F}_t\ \forall t$, we then have $x_t \in X_t \cap M_t\ \forall t$, such that $\{x_t\} \in \mathscr{B}(\mathscr{F},\mathscr{F})$.

$\underline{\mathscr{B}(\mathscr{F},\mathscr{F}) \subseteq \mathscr{B}(\mathscr{F},\widehat{\mathscr{F}})}$: Any $\{x_t\} \in \mathscr{B}(\mathscr{F},\mathscr{F})$ is $\mathscr{F}$-adapted and satisfies all other constraints in (3.2), in particular the random vectors $\sum_{\tau=1}^{t} A_{t\tau}x_\tau$ and $b_t$, measurable w.r.t. $\mathscr{F}_t$, coincide almost surely, such that for any sub–$\sigma$–algebras $\widehat{\mathscr{F}_t} \subseteq \mathscr{F}_t$ their conditional expectations $\mathbb{E}\left\{\sum_{\tau=1}^{t} A_{t\tau}x_\tau \,\middle|\, \widehat{\mathscr{F}_t}\right\}$ and $\mathbb{E}\left\{b_t \,\middle|\, \widehat{\mathscr{F}_t}\right\}$, being a.s. uniquely determined and $\widehat{\mathscr{F}_t}$-measurable as mentioned above, coincide a.s. as well. Hence we have $\{x_t\} \in \mathscr{B}(\mathscr{F},\widehat{\mathscr{F}})$.

The two remaining inclusions,

$$\underline{\mathscr{B}(\widehat{\mathscr{F}},\mathscr{F}) \subseteq \mathscr{B}(\widehat{\mathscr{F}},\widehat{\mathscr{F}})} \text{ and } \underline{\mathscr{B}(\mathscr{F},\widehat{\mathscr{F}}) \supseteq \mathscr{B}(\widehat{\mathscr{F}},\widehat{\mathscr{F}})},$$

as well as the inequalities for the optimal values, are now obvious. □

**Remark 3.1.** *Concerning the fully aggregated problem (3.5) we have the following facts:*

- *If $\mathscr{F}$ is infinite, i.e. at least one of the $\sigma$-algebras $\mathscr{F}_t = \sigma(\zeta_t)$, $t = 1,\cdots,T$, is not finitely generated (equivalently, at least one random vector $\zeta_t$ has not a finite discrete distribution), and $\widehat{\mathscr{F}}$ is finite, then $\mathscr{P}(\widehat{\mathscr{F}},\widehat{\mathscr{F}})$ with finitely many constraints and variables is clearly simpler to deal with than $\mathscr{P}(\mathscr{F},\mathscr{F})$;*
- *for a sequence $\{\widehat{\mathscr{F}^\nu}\}$ of (finite) filtrations with successive refinements, i.e. $\widehat{\mathscr{F}_t^\nu} \subseteq \widehat{\mathscr{F}_t^{\nu+1}}\ \forall t$, under appropriate assumptions, e.g. for a corresponding sequence of measures $P_\nu$ on $\widehat{\mathscr{F}_T^\nu}$ converging weakly to $P$ (see Billingsley [20]), we may expect convergence of the optimal values of (3.5) to that one of (3.2);*
- *according to Prop. 3.1., in general there is no definite relationship between the optimal values of (3.5) and of (3.2), as remarked for instance by Wright [347] (p. 900); however there are special problem classes—in particular in the two-stage case—and particular assumptions for the multi-stage case implying that*

$\inf(\mathscr{P}(\widehat{\mathscr{F}},\widehat{\mathscr{F}}))$ *yields a lower bound for* $\inf(\mathscr{P}(\mathscr{F},\mathscr{F}))$*, which can be used in designing solution methods, as we shall see later.* □

First we shall deal with two-stage SLP's. Under various assumptions on the model structure and the underlying probability distributions, we shall reveal properties of the recourse function and its expectation which turn out to be useful when designing solution methods. Unfortunately, not all of these results can be generalized to corresponding statements for multi-stage SLP's in general.

## 3.2 The two-stage SLP: Properties and solution appraoches

In the previous section, for the $T$-stage SLP we had the following general probabilistic setup: On some probability space $(\Omega,\mathscr{G},P)$ a sequence of random vectors $\xi_t : \Omega \longrightarrow \mathbb{R}^{r_t}$, $t = 2,\cdots,T$, was defined, such that $\xi = (\xi_2^{\mathrm{T}},\cdots,\xi_T^{\mathrm{T}})^{\mathrm{T}}$ induced the probability distribution $\mathbb{P}_\xi$ on the Borel $\sigma$-field of $\mathbb{R}^{r_2+\cdots+r_T}$. Then the random vectors $\zeta_t = (\xi_2^{\mathrm{T}},\cdots,\xi_t^{\mathrm{T}})^{\mathrm{T}}$, $t = 2,\cdots,T$, implied the filtration $\mathscr{F} = \{\mathscr{F}_2,\cdots,\mathscr{F}_T\}$ in $\mathscr{G}$ with $\mathscr{F}_t = \sigma(\zeta_t)$. Restricting ourselves in this section to the case $T = 2$ allows for the following simplification of this setup.

Assume some probability space $(\Omega,\mathscr{F},P)$ together with a random vector $\xi : \Omega \to \mathbb{R}^r$ to be given, such that $\mathscr{F} = \sigma(\xi)$. Then $\xi$ induces the probability measure $\mathbb{P}_\xi$ on $\mathbb{B}^r$, the Borel $\sigma$-algebra in $\mathbb{R}^r$, according to $\mathbb{P}_\xi(B) = P(\xi^{-1}[B])\ \forall B \in \mathbb{B}^r$.

Besides deterministic arrays $A \in \mathbb{R}^{m_1\times n_1}$, $b \in \mathbb{R}^{m_1}$, and $c \in \mathbb{R}^{n_1}$, for the first stage, let the random arrays $T(\xi) \in \mathbb{R}^{m_2\times n_1}$, $W(\xi) \in \mathbb{R}^{m_2\times n_2}$, $h(\xi) \in \mathbb{R}^{m_2}$, and $q(\xi) \in \mathbb{R}^{n_2}$, be defined for the second stage as:

$$\left.\begin{aligned}
T(\xi) &= T + \sum_{j=1}^{r} T^j \xi_j;\quad T,\ T^j \in \mathbb{R}^{m_2\times n_1} \text{ deterministic,}\\
W(\xi) &= W + \sum_{j=1}^{r} W^j \xi_j;\ W,\ W^j \in \mathbb{R}^{m_2\times n_2} \text{ deterministic,}\\
h(\xi) &= h + \sum_{j=1}^{r} h^j \xi_j;\quad h,\ h^j \in \mathbb{R}^{m_2} \text{ deterministic,}\\
q(\xi) &= q + \sum_{j=1}^{r} q^j \xi_j;\quad q,\ q^j \in \mathbb{R}^{n_2} \text{ deterministic.}
\end{aligned}\right\} \tag{3.6}$$

Then, with $\xi \in \mathscr{L}_r^2$ due to Assumption 3.1. and according to (3.2), the general two-stage SLP with random recourse is formulated as

$$\left.\begin{array}{lcll}
\min \mathbb{E}_\xi \left\{c^{\mathrm{T}}x + q^{\mathrm{T}}(\xi)y(\xi)\right\} & & & \\
Ax & = & b & \\
T(\xi)x + W(\xi)y(\xi) & = & h(\xi) & \text{a.s.}\\
x & \geq & 0 & \\
y(\xi) & \geq & 0 & \text{a.s.}\\
y(\cdot) & \in & Y \cap M, &
\end{array}\right\} \tag{3.7}$$

where $Y$—corresponding to (3.2)—is a linear subspace of $\mathscr{L}^2_{n_2}$ (with respect to $(\Omega, \mathscr{F}, P)$), including the set of $\mathscr{F}$-simple functions; and $M$ is the set of $\mathscr{F}$-measurable functions $\Omega \longrightarrow \mathbb{R}^{n_2}$. To avoid unnecessary formalism, we may just assume, that $Y = \mathscr{L}^2_{n_2}$ which obviously contains the $\mathscr{F}$-simple functions and satisfies $Y \subset M$.

Hence problem (3.7) is equivalent to

$$\left.\begin{array}{rll} \min \mathbb{E}_\xi \; \{c^{\mathrm{T}}x + q^{\mathrm{T}}(\xi)y(\xi)\} & & \\ Ax \phantom{+ W(\xi)y(\xi)} & = b & \\ T(\xi)x + W(\xi)y(\xi) & = h(\xi) & \text{a.s.} \\ x \phantom{+ W(\xi)y(\xi)} & \geq 0 & \\ y(\xi) & \geq 0 & \text{a.s.} \\ y(\cdot) & \in Y\,. & \end{array}\right\} \tag{3.8}$$

A brief sketch on modeling situations leading to variants of the general two-stage SLP (3.8) is given in Chapter 1 on page 4.

**Remark 3.2.** *Instead of the constraints $\{Ax = b,\; x \geq 0\}$ in (3.8) we also could consider constraints of the form $\{Ax \propto b,\; l \leq x \leq u\}$ as in (1.1) on page 1, and the constraints $\{W(\xi)y(\xi) = h(\xi) - T(\xi)x,\; y(\xi) \geq 0 \text{ a.s.}\}$ of (3.8) could be replaced as well by $\{W(\xi)y(\xi) \propto h(\xi) - T(\xi)x,\; \tilde{l} \leq y(\xi) \leq \tilde{u} \text{ a.s.}\}$. However, in order to have a unified presentation, for two-stage programs we stay with the formulation chosen in (3.8).* □

Except for particular cases where it is stated explicitly otherwise, instead of (3.6) we shall restrict ourselves to $W(\cdot) \equiv W$, i.e. to *fixed recourse*. In general, problem (3.8) contains implicitly the *recourse function*

$$\left.\begin{array}{rll} Q(x;T(\xi),h(\xi),W(\xi),q(\xi)) := \inf_y q^{\mathrm{T}}(\xi)y(\xi) & & \\ T(\xi)x + W(\xi)y(\xi) = h(\xi) & \text{a.s.} & \\ y(\xi) \geq 0 & \text{a.s.} & \\ y(\cdot) \in Y\,. & & \end{array}\right\} \tag{3.9}$$

To simplify the notation, we shall enter into the recourse function $Q(x;\cdot)$ of (3.9), in addition to the first stage decision variable $x$, only those parameter arrays being random in the model under consideration. For instance, $Q(x;T(\xi),h(\xi))$ indicates that $T(\cdot)$, $h(\cdot)$ are random arrays defined according to (3.6) whereas $W(\cdot) \equiv W$, $q(\cdot) \equiv q$; and $Q(x;h(\xi))$ stands for $h(\cdot)$ being a random vector due to (3.6) and $T(\cdot) \equiv T$, $W(\cdot) \equiv W$, $q(\cdot) \equiv q$ being deterministic data.

Furthermore, in applications of this model, the selection of a decision $\hat{x}$ feasible for the first stage constraints $Ax = b,\; x \geq 0$, appears to be meaningful only if it allows almost surely to satisfy the second stage constraints $W(\xi)y(\xi) = h(\xi) - T(\xi)\hat{x},\; y(\xi) \geq 0$ a.s., since otherwise, according to the usual convention, we should get for the recourse function

$$\begin{aligned}&Q(\hat{x};T(\xi),h(\xi),W(\xi),q(\xi)) = \\ &= \inf_{y\in Y}\{q^{\mathrm{T}}(\xi)y(\xi) \mid W(\xi)y(\xi) = h(\xi) - T(\xi)\hat{x},\ y(\xi) \geq 0 \text{ a.s.}\} = +\infty\end{aligned}$$

with some positive probability. This implies

- either $\mathscr{Q}(\hat{x}) := \mathbb{E}_{\xi}\left[Q(\hat{x};T(\xi),h(\xi),W(\xi),q(\xi))\right] = +\infty$,
- or else the expected recourse $\mathscr{Q}(\hat{x})$ to be undefined if with positive probability $Q(\hat{x};T(\xi),h(\xi),W(\xi),q(\xi)) = -\infty$ results simultaneously.

Clearly in anyone of these situations $\hat{x}$ is not to be chosen since neither an infinite nor an undefined objective value corresponds to our aim to minimize the objective of (3.8). Hence, in general we may be faced with so-called *induced constraints* on $x$, meaning that we require

$$\hat{x} \in K := \{x \mid x \in \mathbb{R}^{n_1};\ Q(x;T(\xi),h(\xi),W(\xi),q(\xi)) < +\infty \text{ a.s.}\}.$$

For $\Xi = \operatorname{supp}\mathbb{P}_{\xi}$—the support of $\mathbb{P}_{\xi}$, i.e. the smallest closed set in $\mathbb{R}^r$ such that $\mathbb{P}_{\xi}(\Xi) = 1$—being an infinite set, $K$ is described in general by an infinite set of constraints, which is not easy to deal with. If however $\Xi$ is either finite, i.e. $\Xi = \{\xi^1,\cdots,\xi^{\rho}\}$, or else a convex polyhedron given by finitely many points as $\Xi = \operatorname{conv}\{\xi^1,\cdots,\xi^{\rho}\}$ (see Chapter 1, Def. 1.3. on page 10), then the induced constraints imply $x \in K$ with

$$K := \{x \mid T(\xi^j)x + W(\xi^j)y^j = h(\xi^j),\ y^j \geq 0,\ j = 1,\cdots,\rho\},$$

and, with $\mathscr{B}_1 := \{x \mid Ax = b,\ x \geq 0\} \subset \mathbb{R}^{n_1}$, the first stage decisions have to satisfy $x \in \mathscr{B}_1 \cap K$. A more detailed discussion of induced constraints may be found in Rockafellar–Wets [285] and in Walkup–Wets [340] (see also Kall [154], Ch. III).

### *3.2.1 The complete fixed recourse problem (CFR)*

If for a particular application it does not seem appropriate, that the future outcomes of $\xi$ affect the set of feasible first stage decisions, given as

$$\mathscr{B}_1 = \{x \mid Ax = b,\ x \geq 0\}, \tag{3.10}$$

we might require at least *relatively complete recourse:*

$$\forall x \in \mathscr{B}_1 \Longrightarrow \{y \mid W(\xi)y = h(\xi) - T(\xi)x,\ y \geq 0\} \neq \emptyset \ \text{ a.s.}. \tag{3.11}$$

Due to the Farkas lemma, Chapter 1, Prop. 1.13. on page 15, condition (3.11) is equivalent to:

$$\forall x \in \mathscr{B}_1 \text{ holds}:\ \left[W^{\mathrm{T}}(\xi)u \leq 0 \Longrightarrow (h(\xi) - T(\xi)x)^{\mathrm{T}}u \leq 0 \ \text{ a.s.}\right].$$

Hence the requirement of relatively complete recourse is a joint restriction on $\mathscr{B}_1$ and on the range of $h(\xi), T(\xi), W(\xi)$ for $\xi \in \Xi$, simultaneously, which may be difficult to verify, in general.

Therefore, in applications it is often preferred to assume *complete fixed recourse* (CFR), which requires for $W(\xi) \equiv W$ the following condition:

$$\{z \mid z = Wy,\ y \geq 0\} = \mathbb{R}^{m_2}. \tag{3.12}$$

If this condition is satisfied, then for any $\hat{x}$ feasible according to an arbitrary set of first stage constraints in (3.8), and for any realization $\hat{\xi}$ of the random vector $\xi$, the second stage constraints in (3.9) are feasible. Furthermore, complete fixed recourse is a condition on the matrix $W$ only, and may easily be checked due to

**Lemma 3.1.** *A matrix $W \in \mathbb{R}^{m_2 \times n_2}$ satisfies the complete recourse condition (3.12) if and only if*

- $\operatorname{rank}(W) = m_2$, *and*
- *for an arbitrary set $\{W_{i_1}, W_{i_2}, \cdots, W_{i_{m_2}}\}$ of linearly independent columns of $W$, the linear constraints*

$$\left.\begin{aligned} Wy &= 0 \\ y_{i_k} &\geq 1,\ k = 1, \cdots, m_2, \\ y &\geq 0 \end{aligned}\right\} \tag{3.13}$$

*are feasible.*

*Proof:* Assume that $W$ is a complete recourse matrix. Then from (3.12) follows that $\operatorname{rank}(W) = m_2$ necessarily holds.

Furthermore, for some selection $\{W_{i_1}, W_{i_2}, \cdots, W_{i_{m_2}}\}$ of linearly independent columns of $W$, let

$$\hat{z} = -\sum_{k=1}^{m_2} W_{i_k}.$$

By our assumption on $W$, we have $\{y \mid Wy = \hat{z},\ y \geq 0\} \neq \emptyset$. Hence, with the index set $\{j_1, \cdots, j_{n_2-m_2}\}$ chosen such that

$$\begin{aligned} \{i_1, i_2, \cdots, i_{m_2}\} \cap \{j_1, \cdots, j_{n_2-m_2}\} &= \emptyset \\ \text{and } \{i_1, i_2, \cdots, i_{m_2}\} \cup \{j_1, \cdots, j_{n_2-m_2}\} &= \{1, \cdots, n_2\}, \end{aligned}$$

there exists a feasible solution $\hat{y}$ of

$$\begin{aligned} \sum_{k=1}^{m_2} W_{i_k} \hat{y}_{i_k} + \sum_{l=1}^{n_2-m_2} W_{j_l} \hat{y}_{j_l} &= \hat{z} \\ &= -\sum_{k=1}^{m_2} W_{i_k} \\ \hat{y}_i \geq 0,\ i = 1, \cdots, n_2. \end{aligned}$$

Hence, with

$$y_\nu = \begin{cases} \hat{y}_\nu + 1\,, & \nu = i_1, i_2, \cdots, i_{m_2}\,, \\ \hat{y}_\nu\,, & \nu = j_1, j_2, \cdots, j_{n_2-m_2}\,, \end{cases}$$

the constraints (3.13) are necessarily satisfied.

Assume now that the conditions of this lemma hold. Choose an arbitrary $\bar{z} \in \mathbb{R}^{m_2}$. Then the linear equation

$$\sum_{k=1}^{m_2} W_{i_k} y_{i_k} = \bar{z}$$

has a unique solution $\{\bar{y}_{i_1}, \cdots, \bar{y}_{i_{m_2}}\}$. If $\bar{y}_{i_k} \geq 0$ for $k = 1, \cdots, m_2$, we have a feasible solution for the recourse equation $Wy = \bar{z}$. Otherwise, set $\gamma := \min\{\bar{y}_{i_1}, \cdots, \bar{y}_{i_{m_2}}\} < 0$. Let $\tilde{y}$ be a feasible solution of (3.13). Then for

$$\hat{y}_\nu = \begin{cases} \bar{y}_\nu - \gamma\tilde{y}_\nu\,, & \nu = i_1, i_2, \cdots, i_{m_2}\,, \\ -\gamma\tilde{y}_\nu\,, & \nu = j_1, j_2, \cdots, j_{n_2-m_2}\,, \end{cases}$$

follows

$$\begin{aligned} W\hat{y} &= \sum_{k=1}^{m_2} W_{i_k}\hat{y}_{i_k} + \sum_{l=1}^{n_2-m_2} W_{j_l}\hat{y}_{j_l} \\ &= \sum_{k=1}^{m_2} W_{i_k} \underbrace{(\bar{y}_{i_k} - \gamma\tilde{y}_{i_k})}_{\geq 0} + \sum_{l=1}^{n_2-m_2} W_{j_l} \underbrace{(-\gamma\tilde{y}_{j_l})}_{\geq 0} \\ &= \bar{z} - \gamma \underbrace{\sum_{r=1}^{n_2} W_r \tilde{y}_r}_{=0} \end{aligned}$$

such that $\hat{y}$ is a feasible solution of $Wy = \bar{z}$, $y \geq 0$. □

Hence, to verify complete fixed recourse, we only have to determine rank$(W)$ and—if rank$(W) = m_2$ is satisfied—to check the feasibility of (3.13) by applying any algorithm for finding a feasible basic solution of this system, as e.g. the method described in Section 1.2.4 on page 19. Throughout our discussion of two-stage SLP's we shall make the

**Assumption 3.2.** *The recourse matrix W satisfies the complete fixed recourse condition (3.12).*

Even for the complete fixed recourse case if, with $\mathscr{C}^P$ being the polar cone of $\mathscr{C} = \{y \mid Wy = 0, y \geq 0\}$, it happens that

$$\Xi \cap \{\xi \mid -q(\xi) \in \mathscr{C}^P\} \neq \Xi\,,$$

then, due to Prop. 1.6. in Chapter 1 (p. 11) $\{\xi \mid -q(\xi) \in \mathscr{C}^P\} \neq \emptyset$ is closed, such that the definition of the support $\Xi$ implies $\mathbb{P}_\xi(\Xi \cap \{\xi \mid -q(\xi) \in \mathscr{C}^P\}) < 1$.

Hence, with $\Xi_0 = \Xi \setminus \{\xi \mid -q(\xi) \in \mathscr{C}^P\}$, by Prop. 1.7. in Chapter 1 (p. 12) follows $Q(x; T(\xi), h(\xi), q(\xi)) = -\infty$ for $\xi \in \Xi_0$ with probability $\mathbb{P}_\xi(\Xi_0) > 0$, yielding $\mathscr{Q}(x) = -\infty\ \forall x \in \mathscr{B}_1$.

Therefore, for allowing the objective of (3.8) to discriminate among various first stage feasible solutions, we need to assume that $-q(\xi) \in \mathscr{C}^P \; \forall \xi \in \Xi$, i.e. using the Farkas lemma (Chapter 1, Prop. 1.13. on page 15) we add to Assumption 3.2. the further

**Assumption 3.3.** *The recourse matrix $W$ together with $q(\cdot)$ satisfy*

$$\{u \mid W^{\mathrm{T}} u \leq q(\xi)\} \neq \emptyset \; \forall \xi \in \Xi \, . \tag{3.14}$$

Observe that due to (3.14) the requirement that $-q(\xi) \in \mathscr{C}^P \, \forall \xi \in \Xi$ is equivalent to dual feasibility of the recourse problem, a.s.

**Lemma 3.2.** *Given Assumptions 3.2. and 3.3., for any $x \in \mathbb{R}^{n_1}$ there exists an optimal recourse $y(\cdot) \in Y$ such that $Q(x; T(\xi), h(\xi), q(\xi)) = q^{\mathrm{T}}(\xi) y(\xi)$.*

*Proof:* Due to Assumptions 3.2. and 3.3. the LP

$$\left.\begin{array}{rcl} \min q^{\mathrm{T}}(\xi) y & & \\ \text{s.t. } Wy & = & h(\xi) - T(\xi)x \\ y & \geq & 0 \end{array}\right\} \tag{3.15}$$

is solvable for all $\xi \in \Xi$. Let $B^{(\nu)}$, $\nu = 1, \cdots, K$, denote all bases out of $W$ (i.e. all the regular $m_2 \times m_2$-submatrices of $W$). Partitioning $W$ into the basic part $B^{(\nu)}$ and the nonbasic part $N^{(\nu)}$ and correspondingly restating $q(\xi) \cong (q_{B^{(\nu)}}(\xi), q_{N^{(\nu)}}(\xi))$ and $y \cong (y_{B^{(\nu)}}, y_{N^{(\nu)}})$, we know from Prop. 1.3. in Chapter 1 (p. 9) that with the convex polyhedral set

$$\mathscr{A}_\nu := \{\xi \mid {B^{(\nu)}}^{-1}(h(\xi) - T(\xi)x) \geq 0, \; q^{\mathrm{T}}_{B^{(\nu)}}(\xi) {B^{(\nu)}}^{-1} N^{(\nu)} - q^{\mathrm{T}}_{N^{(\nu)}}(\xi) \leq 0\}$$

$y(\xi) \cong \left(y_{B^{(\nu)}}(\xi) = {B^{(\nu)}}^{-1}(h(\xi) - T(\xi)x), y_{N^{(\nu)}}(\xi) = 0\right)$ solves (3.15) for any $\xi \in \mathscr{A}_\nu$. Furthermore, from (3.6) follows $y(\cdot) \in \mathscr{L}^2_{n_2}(\mathscr{A}_\nu, \mathbb{B}^r, \mathbb{R}^{n_2})$ for $\nu = 1, \cdots . K$. Since—due to the solvability of (3.15) for all $\xi \in \Xi$—we have that $\bigcup_{\nu=1}^{K} \mathscr{A}_\nu \supset \Xi$, this inclusion also holds for $\bigcup_{\nu=1}^{K} \hat{\mathscr{A}_\nu}$ with the sets $\hat{\mathscr{A}_\nu}$ being defined as $\hat{\mathscr{A}_1} = \mathscr{A}_1$ and $\hat{\mathscr{A}_\nu} = \mathscr{A}_\nu \setminus \bigcup_{\mu=1}^{\nu-1} \mathscr{A}_\mu$ for $\nu = 2, \cdots, K$.

Therefore, $\{\Xi \cap \hat{\mathscr{A}_\nu} \mid \nu = 1, \cdots, K\}$ is a (disjoint) partition of $\Xi$ with $y(\cdot)$ according to

$$y(\xi) \cong \left(y_{B^{(\nu)}}(\xi) = {B^{(\nu)}}^{-1}(h(\xi) - T(\xi)x), y_{N^{(\nu)}}(\xi) = 0\right) \quad \text{for } \xi \in \hat{\mathscr{A}_\nu}$$

a solution of (3.15), being piecewise linear in $\xi$ and hence belonging to $Y$, and yielding $Q(x;T(\xi),h(\xi),q(\xi)) = q^{\mathrm{T}}(\xi)y(\xi)$. □

The above convex polyhedral sets $\mathscr{A}_\nu$ depend, by definition, on $x$, and so do the pairwise disjoint sets $\hat{\mathscr{A}}_\nu$, which we may indicate by denoting them as $\hat{\mathscr{A}}_\nu(x)$. Then for some given $x^{(i)}$, $i = 1,2$, and any $\xi \in \Xi$ there exist $\nu_i \in \{1,\cdots,K\}$ such that $\xi \in \hat{\mathscr{A}}_{\nu_i}(x^{(i)})$ and hence

$$\left.\begin{aligned} Q(x^{(i)};T(\xi),h(\xi),q(\xi)) &= q^{\mathrm{T}}_{B^{(\nu_i)}}(\xi)B^{(\nu_i)^{-1}}(h(\xi)-T(\xi)x^{(i)}) \\ &= \alpha_{\nu_i}(\xi)+d^{(\nu_i)^{\mathrm{T}}}(\xi)x^{(i)}, \\ \text{where}\quad \alpha_{\nu_i}(\xi) &= q^{\mathrm{T}}_{B^{(\nu_i)}}(\xi)B^{(\nu_i)^{-1}}h(\xi) \in L^1 \\ \text{and}\ -d^{(\nu_i)}(\xi) &= (q^{\mathrm{T}}_{B^{(\nu_i)}}(\xi)B^{(\nu_i)^{-1}}T(\xi))^{\mathrm{T}} \in L^1 . \end{aligned}\right\} \tag{3.16}$$

Since, due to the simplex criterion, $u^{(\nu_i)} = B^{(\nu_i)^{-1^{\mathrm{T}}}}q_{B^{(\nu_i)}}(\xi)$, $i = 1,2$, are dual feasible with respect to (3.15), it follows for $i \neq j$

$$\left.\begin{aligned} \alpha_{\nu_i}(\xi)+d^{(\nu_i)^{\mathrm{T}}}(\xi)x^{(j)} &= (h(\xi)-T(\xi)x^{(j)})^{\mathrm{T}}u^{(\nu_i)} \\ &\leq (h(\xi)-T(\xi)x^{(j)})^{\mathrm{T}}u^{(\nu_j)} \\ &= \alpha_{\nu_j}(\xi)+d^{(\nu_j)^{\mathrm{T}}}(\xi)x^{(j)} \\ &= Q(x^{(j)};T(\xi),h(\xi),q(\xi)) . \end{aligned}\right\} \tag{3.17}$$

Now we are ready to show that (3.8) under appropriate assumptions is a meaningful optimization problem.

**Theorem 3.1.** *Let the Assumptions 3.2. and 3.3. be satisfied. Then the recourse function $Q(x;T(\xi),h(\xi),q(\xi))$ is*

*a) finitely valued $\forall x \in \mathscr{B}_1$, $\xi \in \Xi$,*

*b) convex in $x$ $\forall \xi \in \Xi$, and*

*c) Lipschitz continuous in $x$ $\forall \xi \in \Xi$ with a Lipschitz constant $D(\xi) \in \mathscr{L}^1_1$.*

*Proof:*

a) The LP defining the recourse function $Q(x;T(\xi),h(\xi),q(\xi))$ is given by (3.15) as

$$\min\{q^{\mathrm{T}}(\xi)y \mid Wy = h(\xi)-T(\xi)x,\ y \geq 0\},$$

which due to Assumption 3.2. is primal feasible for arbitrary $x \in \mathbb{R}^{n_1}$ and $\xi \in \mathbb{R}^r$, and according to Assumption 3.3. is also dual feasible $\forall \xi \in \Xi$; therefore it is solvable for all $x \in \mathscr{B}_1$ and for all $\xi \in \Xi$, such that

$$Q(x;T(\xi),h(\xi),q(\xi)) \text{ is finitely valued } \forall x \in \mathscr{B}_1 \text{ and } \forall \xi \in \Xi .$$

b) Hence for an arbitrary $\hat{\xi} \in \Xi$ and some $x^{(1)},x^{(2)} \in \mathscr{B}_1$ there exist $y^{(i)}$ for $i = 1,2$ such that

$$Q(x^{(i)};T(\hat{\xi}),h(\hat{\xi}),q(\hat{\xi})) = q^{\mathrm{T}}(\hat{\xi})y^{(i)}, \quad \text{where}$$
$$Wy^{(i)} = h(\hat{\xi}) - T(\hat{\xi})x^{(i)},\ y^{(i)} \geq 0.$$

Then for $\tilde{x} = \lambda x^{(1)} + (1-\lambda)x^{(2)}$ with some $\lambda \in (0,1)$ it follows that

$$\tilde{y} = \lambda y^{(1)} + (1-\lambda)y^{(2)} \text{ is feasible for } Wy = h(\hat{\xi}) - T(\hat{\xi})\tilde{x},\ y \geq 0.$$

Hence

$$Q(\tilde{x};T(\hat{\xi}),h(\hat{\xi}),q(\hat{\xi})) \leq q^{\mathrm{T}}(\hat{\xi})\tilde{y} = \lambda q^{\mathrm{T}}(\hat{\xi})y^{(1)} + (1-\lambda)q^{\mathrm{T}}(\hat{\xi})y^{(2)},$$

showing the convexity of $Q(x;T(\hat{\xi}),h(\hat{\xi}),q(\hat{\xi}))$ in $x$.

c) For any two $x^{(1)} \neq x^{(2)}$ and any $\xi \in \Xi$, according to (3.16) there exist $\nu_i \in \{1,\cdots,K\}$, $i = 1,2$, such that

$$Q(x^{(i)};T(\xi),h(\xi),q(\xi)) = \alpha_{\nu_i}(\xi) + {d^{(\nu_i)}}^{\mathrm{T}}(\xi)x^{(i)},$$

and due to (3.17) holds

$$\begin{aligned}
&[\alpha_{\nu_1}(\xi) + {d^{(\nu_1)}}^{\mathrm{T}}(\xi)x^{(2)}] - [\alpha_{\nu_1}(\xi) + {d^{(\nu_1)}}^{\mathrm{T}}(\xi)x^{(1)}] \\
&= {d^{(\nu_1)}}^{\mathrm{T}}(\xi)(x^{(2)} - x^{(1)}) \\
&\leq Q(x^{(2)};T(\xi),h(\xi),q(\xi)) - Q(x^{(1)};T(\xi),h(\xi),q(\xi)) \\
&\leq [\alpha_{\nu_2}(\xi) + {d^{(\nu_2)}}^{\mathrm{T}}(\xi)x^{(2)}] - [\alpha_{\nu_2}(\xi) + {d^{(\nu_2)}}^{\mathrm{T}}(\xi)x^{(1)}] \\
&= {d^{(\nu_2)}}^{\mathrm{T}}(\xi)(x^{(2)} - x^{(1)}),
\end{aligned}$$

such that

$$\begin{aligned}
&|Q(x^{(2)};T(\xi),h(\xi),q(\xi)) - Q(x^{(1)};T(\xi),h(\xi),q(\xi))| \\
&\leq \max_{i\in\{1,2\}} |{d^{(\nu_i)}}^{\mathrm{T}}(\xi)(x^{(2)} - x^{(1)})| \leq \max_{i\in\{1,2\}} \|d^{(\nu_i)}(\xi)\| \|(x^{(2)} - x^{(1)})\|.
\end{aligned}$$

Hence, with $D(\xi) = \max_{i\in\{1,\cdots,K\}} \|d^{(\nu_i)}(\xi)\| \in L^1$—due to (3.16)—follows the proposition. □

Due to Chapter 1, Def. 1.10. (p. 54) a vector $g \in \mathbb{R}^n$ is a subgradient of a convex function $\varphi : \mathbb{R}^n \longrightarrow \mathbb{R}$ at a point $x$ if it satisfies

$$g^{\mathrm{T}}(z-x) \leq \varphi(z) - \varphi(x) \ \ \forall z,$$

and the subdifferential $\partial\varphi(x)$ is the set of all subgradients of $\varphi$ at $x$. In particular for linear programs we have

**Lemma 3.3.** *Assume that the LP*

$$\min\{c^{\mathrm{T}}x \mid Ax = b, x \geq 0\}$$

*is solvable* $\forall b \in \mathbb{R}^m$*. Then its optimal value* $\varphi(b)$ *(obviously convex in* $b$*) is subdifferentiable at any* $b$*, and the subdifferential is given as* $\partial\varphi(b) = \arg\max\{b^{\mathrm{T}}u \mid A^{\mathrm{T}}u \leq c\}$*, the set of optimal dual solutions at* $b$*.*

*Proof:* For a given $\widehat{b}$ let $\hat{u} \in \arg\max\{\widehat{b}^{\mathrm{T}}u \mid A^{\mathrm{T}}u \leq c\}$, such that $\varphi(\widehat{b}) = \widehat{b}^{\mathrm{T}}\hat{u}$. Hence $\hat{u}$ is also feasible for the LP $\varphi(\widetilde{b}) = \max\{\widetilde{b}^{\mathrm{T}}u \mid A^{\mathrm{T}}u \leq c\}$ for an arbitrary $\widetilde{b}$ such that $\widetilde{b}^{\mathrm{T}}\hat{u} \leq \varphi(\widetilde{b})$ holds. Hence

$$\hat{u}^{\mathrm{T}}(\widetilde{b} - \widehat{b}) \leq \varphi(\widetilde{b}) - \varphi(\widehat{b})$$

showing that $\arg\max\{\widehat{b}^{\mathrm{T}}u \mid A^{\mathrm{T}}u \leq c\} \subset \partial\varphi(\widehat{b})$.

Assume now that $g \in \partial\varphi(\widehat{b})$ for some $\widehat{b}$. Therefore, for any $b$ holds

$$g^{\mathrm{T}}(b - \widehat{b}) \leq \varphi(b) - \varphi(\widehat{b}).$$

With $\hat{x} \in \arg\min\{c^{\mathrm{T}}x \mid Ax = \widehat{b}, x \geq 0\}$ and $x^{(i)} = \hat{x} + e_i (\geq 0)$, $i = 1, \cdots, n$, ($e_i$ the $i$-th unit vector), by our assumption, for all $b^{(i)} = Ax^{(i)}$, the LP's $\varphi(b^{(i)}) = \min\{c^{\mathrm{T}}x \mid Ax = b^{(i)}, x \geq 0\}$ are solvable. Obviously we have $\varphi(b^{(i)}) \leq c^{\mathrm{T}}x^{(i)}$ such that

$$\begin{aligned} g^{\mathrm{T}}Ae_i = g^{\mathrm{T}}A(x^{(i)} - \hat{x}) &= g^{\mathrm{T}}(b^{(i)} - \widehat{b}) \\ &\leq \varphi(b^{(i)}) - \varphi(\widehat{b}) \\ &\leq c^{\mathrm{T}}x^{(i)} - c^{\mathrm{T}}\hat{x} = c^{\mathrm{T}}e_i, \quad i = 1, \cdots, n, \end{aligned}$$

implying $A^{\mathrm{T}}g \leq c$, the dual feasibility of $g$. Then, due to the weak duality theorem (Chapter 1, Prop. 1.9., page 13), we have $g^{\mathrm{T}}\widehat{b} - \varphi(\widehat{b}) \leq 0$. Assume that with some $\alpha < 0$ holds $g^{\mathrm{T}}\widehat{b} - \varphi(\widehat{b}) \leq \alpha$. For $\widetilde{b} = 0$ obviously follows $\varphi(\widetilde{b}) = 0$ such that the subgradient inequality, valid for all $b$, yields

$$0 = g^{\mathrm{T}}\widetilde{b} - \varphi(\widetilde{b}) \leq g^{\mathrm{T}}\widehat{b} - \varphi(\widehat{b}) \leq \alpha < 0.$$

This contradiction, implied by the assumption $g^{\mathrm{T}}\widehat{b} - \varphi(\widehat{b}) \leq \alpha < 0$, shows that $g^{\mathrm{T}}\widehat{b} = \varphi(\widehat{b})$ and hence $\partial\varphi(\widehat{b}) \subset \arg\max\{\widehat{b}^{\mathrm{T}}u \mid A^{\mathrm{T}}u \leq c\}$. □

Now we get immediately

**Theorem 3.2.** *Let the Assumptions 3.2. and 3.3. be satisfied. Then the recourse function* $Q(x; T(\xi), h(\xi), q(\xi))$ *is subdifferentiable in* $x$ *for any* $\xi \in \Xi$*. For any* $\hat{x}$ *holds (the subscript at* $\partial$ *indicating the variable of subdifferentiation)*

$$\begin{aligned} &\partial_x Q(\hat{x}; T(\xi), h(\xi), q(\xi)) = \\ &= \{-T^{\mathrm{T}}(\xi)\hat{u} \mid \hat{u} \in \arg\max\{(h(\xi) - T(\xi)\hat{x})^{\mathrm{T}}u \mid W^{\mathrm{T}}u \leq q(\xi)\}\} \ \forall \xi \in \Xi. \end{aligned}$$

*Proof:* For an arbitrary $\xi \in \Xi$ define $b(x;\xi) := h(\xi) - T(\xi)x$. Introducing

$$\psi(b(x;\xi);\xi) := Q(x; T(\xi), h(\xi), q(\xi))$$

$$= \min\{q^{\mathrm{T}}(\xi)y \mid Wy = b(x;\xi),\, y \geq 0\},$$

from Lemma 3.3 follows for the subdifferential of $\psi(\cdot;\xi)$ at $b(\hat{x};\xi)$

$$\partial_b \psi(b(\hat{x};\xi);\xi) = \arg\max\{b^{\mathrm{T}}(\hat{x};\xi)u \mid W^{\mathrm{T}}u \leq q(\xi)\}.$$

Then from Prop. 1.25. in Chapter 1 (p. 55) we know that

$$\begin{aligned} \partial_x Q(\hat{x};T(\xi),h(\xi),q(\xi)) &= -T^{\mathrm{T}}(\xi)\partial_b \psi(b(\hat{x};\xi);\xi) \\ &= -T^{\mathrm{T}}(\xi)\arg\max\{b^{\mathrm{T}}(\hat{x};\xi)u \mid W^{\mathrm{T}}u \leq q(\xi)\}. \end{aligned}$$

□

**Theorem 3.3.** *Since $\xi \in \mathscr{L}_r^2$ (i.e. $\xi$ square-integrable with respect to $\mathbb{P}_\xi$), the expected recourse $\mathscr{Q}(x)$ is*

*a) finitely valued $\forall x \in \mathscr{B}_1$, and*
*b) a convex and Lipschitz continuous function in x.*

*Hence, (3.8) is a convex optimization problem with a Lipschitz continuous objective function.*

*Proof:*

a) Let $\hat{x} \in \mathbb{R}^{n_1}$ be fixed. Due to Assumptions 3.2. and 3.3., for any $\xi \in \Xi$ there exists an optimal feasible basic solution of the recourse program (3.15), i.e. there is an $(m_2 \times m_2)$-submatrix $B$ of $W$ such that

$$\left.\begin{aligned} B^{-1}(h(\xi) - T(\xi))\hat{x} \quad &\geq 0 \quad \text{and} \\ Q(\hat{x};T(\xi),h(\xi),q(\xi)) &= q_B(\xi)^{\mathrm{T}}B^{-1}(h(\xi) - T(\xi)\hat{x}) \end{aligned}\right\}, \tag{3.18}$$

where the components of the $m_2-$subvector $q_B(\xi)$ of $q(\xi)$ correspond to the columns in $B$ selected from $W$, as mentioned in Chapter 1, Prop. 1.2. (p. 9). Together with the simplex criterion, Prop. 1.3. in Chapter 1 (p. 9), such a particular basis is feasible and optimal on a polyhedral subset $\Xi_B \subset \Xi$, a so-called decision region (also: stability region).
According to (3.6) and (3.18), the recourse function $Q(\hat{x};T(\xi),h(\xi),q(\xi))$ is, in general, a quadratic function in $\xi$ for $\xi \in \Xi_B$, such that, due to the assumption that $\xi \in L_2$, the integral $\int_{\Xi_B} Q(\hat{x};T(\xi),h(\xi),q(\xi))\mathbb{P}_\xi(d\xi)$ exists. By the Assumptions 3.2. and 3.3., the support $\Xi$ is contained in the union of finitely many decision regions, which implies that also

$$\mathscr{Q}(\hat{x}) = \int_{\Xi} Q(\hat{x};T(\xi),h(\xi),q(\xi))\mathbb{P}_\xi(d\xi) \quad \text{exists.}$$

b) In Theorem 3.1., for any $\xi \in \Xi$, the recourse function $Q(x;T(\xi),h(\xi),q(\xi))$ has been shown to be convex and Lipschitz continuous in $x$, with a Lipschitz constant $D(\xi) \in \mathscr{L}_1^1$.

Hence the convexity of $\mathscr{Q}(x) = \int_\Xi Q(x;T(\xi),h(\xi),q(\xi))\mathbb{P}_\xi(d\xi)$ is obvious. And for any two $x^{(1)}$ and $x^{(2)}$ we have

$$
\begin{aligned}
&|\mathscr{Q}(x^{(1)}) - \mathscr{Q}(x^{(2)})| \\
&\leq \left| \int_\Xi \{Q(x^{(1)};T(\xi),h(\xi),q(\xi)) - Q(x^{(2)};T(\xi),h(\xi),q(\xi))\}\mathbb{P}_\xi(d\xi) \right| \\
&\leq \int_\Xi \left| Q(x^{(1)};T(\xi),h(\xi),q(\xi)) - Q(x^{(2)};T(\xi),h(\xi),q(\xi)) \right| \mathbb{P}_\xi(d\xi) \\
&\leq \int_\Xi D(\xi)\|x^{(1)} - x^{(2)}\|\mathbb{P}_\xi(d\xi) = D\|x^{(1)} - x^{(2)}\|
\end{aligned}
$$

with the Lipschitz constant $D = \int_\Xi D(\xi)\mathbb{P}_\xi(d\xi)$. □

**Corollary 3.1.** *Given that the random entries $q(\xi)$ and $(h(\xi),T(\xi))$ are stochastically independent, then with $\xi \in \mathscr{L}_r^1$ (instead of $\xi \in \mathscr{L}_r^2$ as before), the conclusions of Th. 3.3. hold true, as well.*

*Proof:* Only the existence of $\mathscr{Q}(x) = \int_\Xi Q(x;T(\xi),h(\xi),q(\xi))\mathbb{P}_\xi(d\xi)$ has to be proved, which follows, with $\xi \in \mathscr{L}_r^1(\Omega,\mathbb{R}^r)$, from the independence of $q(\xi)$ and $(h(\xi),T(\xi))$ according to

$$
\begin{aligned}
&\int_{\Xi_B} Q(x;T(\xi),h(\xi),q(\xi))\mathbb{P}_\xi(d\xi) = \\
&= \int_{\Xi_B} q_B(\xi)^{\mathrm{T}} B^{-1}(h(\xi) - T(\xi)x)\mathbb{P}_\xi(d\xi) \\
&= \left( \int_{\Xi_B} q_B(\xi)\mathbb{P}_\xi(d\xi) \right)^{\mathrm{T}} \left( \int_{\Xi_B} B^{-1}(h(\xi) - T(\xi)x)\mathbb{P}_\xi(d\xi) \right).
\end{aligned}
$$

□

**Remark 3.3.** *In Theorem 3.2. the subdifferential of the recourse function at any $\hat{x}$ under the Assumptions 3.2. and 3.3. was derived as*

$$
\begin{aligned}
&\partial_x Q(\hat{x};T(\xi),h(\xi),q(\xi)) = \\
&= \{-T^{\mathrm{T}}(\xi)\hat{u} \mid \hat{u} \in \arg\max\{(h(\xi) - T(\xi)\hat{x})^{\mathrm{T}} u \mid W^{\mathrm{T}} u \leq q(\xi)\}\} \ \forall \xi \in \Xi\,.
\end{aligned}
$$

*It can be shown, that then $\mathscr{Q}(\cdot)$ is subdifferentiable at $\hat{x}$ and*

$$
\partial \mathscr{Q}(\hat{x}) = \int_\Xi \partial_x Q(\hat{x};T(\xi),h(\xi),q(\xi))\mathbb{P}_\xi(d\xi), \tag{3.19}
$$

*where this integral is understood as the set* $\left\{ \int_{\Xi} G(\xi) \mathbb{P}_{\xi}(d\xi) \right\}$ *for all functions* $G(\cdot)$ *being measurable selections from* $\partial_x Q(\hat{x}; T(\cdot), h(\cdot), q(\cdot))$ *such that the integral* $\int_{\Xi} \|G(\xi)\| \mathbb{P}_{\xi}(d\xi)$ *exists.*

*Finally,* $\mathscr{Q}(\cdot)$ *is differentiable at* $\hat{x}$ *if and only if* $\partial_x Q(\hat{x}; T(\cdot), h(\cdot), q(\cdot))$ *is a singleton a.s. with respect to* $\mathbb{P}_{\xi}$*.*

*To prove statements of this type involves several technicalities, like the existence of measurable selections from subdifferentials or equivalently, from solution sets of optimization problems, integrability statements like Lebesgue's bounded convergence theorem, and so on. Under specific assumptions, these problems were considered for instance in Kall [152], Kall–Oettli [170], Rockafellar [280] (see also Kall [154]), and the general case is dealt with in Ch. 2 of Ruszczyński–Shapiro [295], where a sketch of a proof is presented.*

*Due to the fact that (sub)gradient methods will—in general—not be a central part of our discussion of solution approaches for recourse problems later on, we omit a proof of the interchangeability of subdifferentiation and integration, as stated in (3.19).* □

#### 3.2.1.1 CFR: Direct bounds for the expected recourse $\mathscr{Q}(x)$

Finally, assume that $q(\xi) \equiv q$, i.e. $q(\cdot)$ is deterministic. Then we have

**Proposition 3.2.** *Given the Assumptions 3.2. and 3.3. (the latter one now reading as* $\{u \mid W^{\mathrm{T}} u \leq q\} \neq \emptyset$*),* $Q(x; T(\cdot), h(\cdot))$ *is a convex function in* $\xi$ *for any* $x \in \mathbb{R}^{n_1}$*.*

*Proof:* According to (3.6) for any fixed $x \in \mathbb{R}^{n_1}$ the right–hand–side of the LP

$$Q(x; T(\xi), h(\xi)) := \min\{q^{\mathrm{T}} y \mid Wy = h(\xi) - T(\xi)x, \, y \geq 0\}$$

is linear in $\xi$, which implies the asserted convexity. □

In this case we have a lower bound for $\mathscr{Q}(x)$, frequently used in solution methods, which is based on *Jensen's inequality* [148]:

**Lemma 3.4.** *Let* $\xi \in \mathbb{R}^r$ *be a random vector with probability distribution* $\mathbb{P}_{\xi}$ *such that* $\mathbb{E}_{\xi}[\xi]$ *exists, and assume* $\varphi : \mathbb{R}^r \longrightarrow \mathbb{R}$ *to be a convex function. Then the following inequality holds true:*

$$\varphi(\mathbb{E}_{\xi}[\xi]) \leq \mathbb{E}_{\xi}[\varphi(\xi)]. \tag{3.20}$$

*Proof:* Due to Chapter 1, Prop. 1.25. (p. 55), at any $\hat{\xi} \in \mathbb{R}^r$ there exists a nonempty, convex, compact subdifferential $\partial\varphi(\hat{\xi})$. Hence for any linear affine function $\ell(\cdot)$ out of the family $\widetilde{\mathscr{L}}_{\hat{\xi}}$ for some $\hat{\xi} \in \mathbb{R}^r$ with

$$\widetilde{\mathscr{L}_{\hat{\xi}}} := \{l(\cdot) \mid l(\xi) := \varphi(\hat{\xi}) + g_{\hat{\xi}}^{\mathrm{T}}(\xi - \hat{\xi}),\, g_{\hat{\xi}} \in \partial\varphi(\hat{\xi})\}\,,\ \hat{\xi} \in \mathbb{R}^r\,,$$

the set of linear support functions to $\varphi(\cdot)$ at $\hat{\xi}$, we have the subgradient inequality

$$\ell(\xi) = \varphi(\hat{\xi}) + g_{\hat{\xi}}^{\mathrm{T}}(\xi - \hat{\xi}) \leq \varphi(\xi) \ \ \forall \xi \in \mathbb{R}^r\,.$$

By integration with respect to $\mathbb{P}_\xi$ follows

$$\mathbb{E}_\xi\,[\ell(\xi)] = \ell(\mathbb{E}_\xi\,[\xi]) = \varphi(\hat{\xi}) + g_{\hat{\xi}}^{\mathrm{T}}(\mathbb{E}_\xi\,[\xi] - \hat{\xi}) \leq \mathbb{E}_\xi\,[\varphi(\xi)]$$

such that $\ell(\mathbb{E}_\xi\,[\xi])$ yields a lower bound for $\mathbb{E}_\xi\,[\varphi(\xi)]$.

Since $\mathbb{E}_\xi\,[\xi] \in \mathbb{R}^r$, due to the subgradient inequality, at any $\hat{\xi} \in \mathbb{R}^r$ holds

$$\ell(\mathbb{E}_\xi\,[\xi]) = \varphi(\hat{\xi}) + g_{\hat{\xi}}^{\mathrm{T}}(\mathbb{E}_\xi\,[\xi] - \hat{\xi}) \leq \varphi(\mathbb{E}_\xi\,[\xi]) \ \ \forall \ell(\cdot) \in \widetilde{\mathscr{L}_{\hat{\xi}}}\,.$$

Hence, in $\{\widetilde{\mathscr{L}_{\hat{\xi}}},\ \hat{\xi} \in \mathbb{R}^r\}$, the set of all possible linear support functions to $\varphi(\cdot)$, we get

$$\arg\max_{\hat{\xi}}\{\ell(\mathbb{E}_\xi\,[\xi]) \mid \ell(\cdot) \in \widetilde{\mathscr{L}_{\hat{\xi}}},\ \hat{\xi} \in \mathbb{R}^r\} = \mathbb{E}_\xi\,[\xi]\,.$$

Therefore, among all linear support functions to $\varphi(\cdot)$ we get the greatest lower bound for $\mathbb{E}_\xi\,[\varphi(\xi)]$ by choosing $\hat{\xi} = \mathbb{E}_\xi\,[\xi]$, i.e. $\ell(\xi) = \varphi(\mathbb{E}_\xi\,[\xi]) + g_{\mathbb{E}_\xi\,[\xi]}^{\mathrm{T}}(\xi - \mathbb{E}_\xi\,[\xi])$, yielding

$$\ell(\mathbb{E}_\xi\,[\xi]) = \varphi(\mathbb{E}_\xi\,[\xi]) \leq \mathbb{E}_\xi\,[\varphi(\xi)]\,.$$

□

Whereas under the assumptions of Lemma 3.4 we know for sure that the integral $\int_{\mathbb{R}^r} \varphi(\xi)\mathbb{P}_\xi(d\xi)$ is bounded below, it cannot be excluded in general that $\mathbb{E}_\xi\,[\varphi(\xi)] = +\infty$ holds. In contrast, under our assumptions for Prop. 3.2. we know from Cor. 3.1. that $\mathscr{Q}(x) = \mathbb{E}_\xi\,[Q(x;T(\xi),h(\xi))]$ is finite for all $x \in \mathbb{R}^{n_1}$. From Prop. 3.2. and Lemma 3.4 follows immediately the *Jensen lower bound* for the expected recourse:

**Theorem 3.4.** *Given the Assumptions 3.2. and 3.3., with $\bar{\xi} = \mathbb{E}_\xi\,[\xi]$, the expected recourse $\mathscr{Q}(x) = \mathbb{E}_\xi\,[Q(x;T(\xi),h(\xi))]$ is bounded below due to*

$$Q(x;T(\bar{\xi}),h(\bar{\xi})) \leq \mathscr{Q}(x)\,. \tag{3.21}$$

Observe that in this case the lower bound for the expected recourse is defined by the one-point distribution $\mathbb{P}_\eta$ with $\mathbb{P}_\eta(\{\eta \mid \eta = \bar{\xi}\}) = 1$, which does not depend on the particular recourse function, since

$$\int Q(x;T(\eta),h(\eta))\mathbb{P}_\eta(d\eta) = Q(x;T(\bar{\xi}),h(\bar{\xi})) \leq \mathscr{Q}(x)$$

holds true for any function $Q(x;T(\cdot),h(\cdot))$ being convex in $\xi$.

Concerning upper bounds for the expected recourse, the situation is more difficult. The first attempts to derive upper bounds for the expectation of convex functions of random variables are assigned to Edmundson [83] and Madansky [210]. Hence, the basic relation is referred to as *Edmundson–Madansky inequality* (E–M):

**Lemma 3.5.** *Let $\tau$ be a random variable with* $\operatorname{supp}\mathbb{P}_\tau \subseteq [\alpha,\beta] \subset \mathbb{R}$ *such that the expectation* $\mu = \mathbb{E}_\tau[\tau] \in [\alpha,\beta]$. *Then, for any convex function* $\psi : [\alpha,\beta] \longrightarrow \mathbb{R}$ *holds*

$$\mathbb{E}_\tau[\psi(\tau)] \le \mathbb{E}_{\hat{\tau}}[\psi(\hat{\tau})], \tag{3.22}$$

*where $\hat{\tau}$ is the discrete random variable with the two-point distribution*

$$\mathbb{P}_{\hat{\tau}}(\{\hat{\tau} \mid \hat{\tau} = \alpha\}) = \frac{\beta-\mu}{\beta-\alpha}, \quad \mathbb{P}_{\hat{\tau}}(\{\hat{\tau} \mid \hat{\tau} = \beta\}) = \frac{\mu-\alpha}{\beta-\alpha}. \tag{3.23}$$

*Proof:* With $\lambda_\tau = \dfrac{\beta-\tau}{\beta-\alpha}$ we have $\lambda_\tau\alpha + (1-\lambda_\tau)\beta = \tau \;\forall\tau \in [\alpha,\beta]$ and $\lambda_\tau \in [0,1]$. Due to the convexity of $\psi$ follows

$$\psi(\tau) = \psi(\lambda_\tau\alpha + (1-\lambda_\tau)\beta) \le \lambda_\tau\psi(\alpha) + (1-\lambda_\tau)\psi(\beta) \;\; \forall\tau \in [\alpha,\beta]$$

and therefore, integrating both sides of this inequality with respect to $\mathbb{P}_\tau$,

$$\mathbb{E}_\tau[\psi(\tau)] \le \frac{\beta-\mu}{\beta-\alpha}\cdot\psi(\alpha) + \frac{\mu-\alpha}{\beta-\alpha}\cdot\psi(\beta) = \mathbb{E}_{\hat{\tau}}[\psi(\hat{\tau})].$$

□

#### 3.2.1.2 CFR: Moment problems and bounds for $\mathscr{Q}(x)$

It is worthwhile to observe the following relation to the theory of *moment problems* and *semi-infinite programs*.

Under the assumptions of Lemma 3.5 consider, with $\mathscr{P}$ the set of probability measures on $[\alpha,\beta]$, as primal (P) the problem

$$\sup_{\mathbb{P}\in\mathscr{P}} \left\{ \int_\alpha^\beta \psi(\xi)\mathbb{P}(d\xi) \;\middle|\; \int_\alpha^\beta \xi\,\mathbb{P}(d\xi) = \mu,\; \int_\alpha^\beta \mathbb{P}(d\xi) = 1 \right\}, \tag{3.24}$$

a so-called moment problem, and as its dual problem (D)

$$\inf_{y\in\mathbb{R}^2} \{y_1 + \mu y_2 \mid y_1 + \xi y_2 \ge \psi(\xi) \;\; \forall\xi \in [\alpha,\beta]\}, \tag{3.25}$$

the corresponding semi-infinite program.

Since, as required by the constraints of (D), a linear affine function majorizes a convex function on an interval if and only if it does so on the endpoints, (D) is equivalent to

$$\min_{y\in\mathbb{R}^2}\{y_1+\mu y_2 \mid y_1+\alpha y_2 \geq \psi(\alpha),\ y_1+\beta y_2 \geq \psi(\beta)\}.$$

Due to the fact that $\alpha \leq \mu \leq \beta$ this LP is solvable, and hence so is its dual (P), which now reads as

$$\max_{p_\alpha,p_\beta}\{\psi(\alpha)p_\alpha+\psi(\beta)p_\beta \mid p_\alpha+p_\beta=1,\ \alpha p_\alpha+\beta p_\beta=\mu;\ p_\alpha,p_\beta\geq 0\}$$

and has, as the unique solution of its constraints, the distribution of $\hat{\tau}$ as given in (3.23). It is worth mentioning that in this case the solution of the moment problem (P), i.e. the E–M distribution yielding the upper bound, is independent of the particular choice of the convex function $\psi : [\alpha,\beta] \longrightarrow \mathbb{R}$.

Suppose now that we have a random vector $\xi \in \mathbb{R}^r$. Then, as mentioned in Kall–Stoyan [171], Lemma 3.5 can immediately be generalized as follows:

**Lemma 3.6.** *Let* $\operatorname{supp}\mathbb{P}_\xi \subset \Xi = \prod_{i=1}^r[\alpha_i,\beta_i] \subset \mathbb{R}^r$ *and assume the components of* $\xi$ *to be stochastically independent. With* $\mu = \mathbb{E}_\xi[\xi] \in \Xi$ *let* $\mathbb{P}_{\eta_i}$, $i=1,\cdots,r$, *be the two-point distributions defined on* $[\alpha_i,\beta_i]$ *as*

$$\mathbb{P}_{\eta_i}(\{\eta_i \mid \eta_i=\alpha_i\}) = \frac{\beta_i-\mu_i}{\beta_i-\alpha_i},\ \ \mathbb{P}_{\eta_i}(\{\eta_i \mid \eta_i=\beta_i\}) = \frac{\mu_i-\alpha_i}{\beta_i-\alpha_i}. \tag{3.26}$$

*Then for the random vector* $\eta \in \mathbb{R}^r$ *with the probability distribution given as*

$$\mathbb{P}_\eta = \mathbb{P}_{\eta_1}\times\mathbb{P}_{\eta_2}\times\cdots\times\mathbb{P}_{\eta_r} \quad on \quad \Xi=\prod_{i=1}^r[\alpha_i,\beta_i] \tag{3.27}$$

*it follows for any convex function* $\varphi : \Xi \longrightarrow \mathbb{R}$ *that*

$$\mathbb{E}_\xi[\varphi(\xi)] \leq \mathbb{E}_\eta[\varphi(\eta)]. \tag{3.28}$$

*Proof:* With $\mathbb{P}_{\xi_i}$ the marginal distribution of $\mathbb{P}_\xi$ for $\xi_i \in [\alpha_i,\beta_i]$, the assumed stochastic independence of the components of $\xi$ implies that

$$\mathbb{P}_\xi = \mathbb{P}_{\xi_1}\times\mathbb{P}_{\xi_2}\times\cdots\times\mathbb{P}_{\xi_r}.$$

Hence the asserted inequality (3.28) follows immediately from Lemma 3.5 by induction to $r$, using the fact that the product measures $\mathbb{P}_\xi$ and $\mathbb{P}_\eta$ allow for iterated integration, as known from Fubini's theorem (see Halmos [131]). □

Also in this case we may assign a moment problem, with $\mathscr{P}$ the set of all product measures on $\Xi=\prod_{i=1}^r \Xi_i = \prod_{i=1}^r[\alpha_i,\beta_i]$, stated as (P)

$$\sup_{\mathbb{P}\in\mathscr{P}}\left\{\int_\Xi \varphi(\xi)P_{\xi_1}(d\xi_1)\cdots P_{\xi_r}(d\xi_r)\ \middle|\ \begin{array}{ll}\int_{\Xi_i}\xi_i P_{\xi_i}(d\xi_i) = \mu_i, & \\ \int_{\Xi_i} P_{\xi_i}(d\xi_i) \quad = 1, & \end{array}\forall i\right\}, \tag{3.29}$$

and its dual semi-infinite program (D)

$$\inf_{y\in\mathbb{R}^{2r}}\left\{\sum_{i=1}^{r}(y_1^i+\mu_i y_2^i)\mid y_1^i+\xi_i y_2^i\geq\tilde{\varphi}_i(\xi_i)\;\;\forall\xi_i\in\Xi_i\,\forall i\right\}\tag{3.30}$$

where with $\Xi/\Xi_i := \Xi_1\times\cdots\times\Xi_{i-1}\times\Xi_{i+1}\times\cdots\times\Xi_r$

$$\tilde{\varphi}_i(\xi_i)=$$
$$\int_{\Xi/\Xi_i}\varphi(\xi_1,\cdots,\xi_r)\mathbb{P}_{\xi_1}(d\xi_1)\cdots\mathbb{P}_{\xi_{i-1}}(d\xi_{i-1})\mathbb{P}_{\xi_{i+1}}(d\xi_{i+1})\cdots\mathbb{P}_{\xi_r}(d\xi_r)$$

is obviously a convex function in $\xi_i$. Therefore again, the constraints of (D) are satisfied if and only if they hold in the endpoints $\alpha_i$ and $\beta_i$ of all intervals $\Xi_i$. Hence (D) is equivalent to

$$\inf_{y\in\mathbb{R}^{2r}}\left\{\sum_{i=1}^{r}(y_1^i+\mu_i y_2^i)\mid y_1^i+\alpha_i y_2^i\geq\tilde{\varphi}_i(\alpha_i),\;y_1^i+\beta_i y_2^i\geq\tilde{\varphi}_i(\beta_i)\;\;\forall i\right\},$$

which due to $\mu_i\in[\alpha_i,\beta_i]$ is solvable again and hence so is its dual, the moment problem

$$\begin{aligned}&\max\left\{\sum_{i=1}^{r}(\tilde{\varphi}_i(\alpha_i)p^i_{\alpha_i}+\tilde{\varphi}_i(\beta_i)p^i_{\beta_i})\right\}\\ &\text{s.t.}\qquad \alpha_i p^i_{\alpha_i}+\beta_i p^i_{\beta_i}=\mu_i,\;p^i_{\alpha_i}+p^i_{\beta_i}=1\;\forall i.\end{aligned}$$

Since the only feasible solution of its constraints coincides with the two-point measures (3.26), the product measure (3.27) solving the moment problem (P) is independent of the particular convex function $\varphi$, again.

For later use we just mention the following fact, which due to the above results is evident:

**Corollary 3.2.** *Let* $\operatorname{supp}\mathbb{P}_\xi\subset\Xi=\prod_{i=1}^r[\alpha_i,\beta_i]\subset\mathbb{R}^r$ *with* $\mu=\mathbb{E}_\xi[\xi]$ *and assume the function* $\varphi:\Xi\longrightarrow\mathbb{R}$ *to be convex separable, i.e.* $\varphi(\xi)=\sum_{i=1}^{r}\varphi_i(\xi_i)$. *Then, with the distributions* $\mathbb{P}_{\eta_i}$ *given in (3.26), it follows that*

$$\mathbb{E}_\xi[\varphi(\xi)]=\sum_{i=1}^{r}\mathbb{E}_\xi[\varphi_i(\xi_i)]\leq\sum_{i=1}^{r}\mathbb{E}_{\eta_i}[\varphi_i(\eta_i)].\tag{3.31}$$

We shall refer to (3.22), (3.28) and (3.31) as the E–M inequality. For the expected recourse we then get the *E–M upper bound*:

**Theorem 3.5.** *Assume that the components of* $\xi$ *are stochastically independent and that* $\operatorname{supp}\mathbb{P}_\xi\subset\Xi=\prod_{i=1}^r[\alpha_i,\beta_i]$ *with* $\mu=\mathbb{E}_\xi[\xi]\in\Xi$. *Given the Assumptions 3.2. and 3.3., with the E–M distribution defined by (3.26) and (3.27) the expected recourse* $\mathscr{Q}(x)=\mathbb{E}_\xi[Q(x;T(\xi),h(\xi))]$ *is bounded above according to*

$$\mathscr{Q}(x) \leq \mathbb{E}_\eta[Q(x;T(\eta),h(\eta))]. \tag{3.32}$$

According to Lemma 3.6 and Cor. 3.2. we have the E–M inequality for multidimensional distributions either for random vectors with independent components or for convex integrands being separable. However this upper bound does not remain valid for arbitrary integrands and dependent components, in general, as shown by the following example:

**Example 3.1.** *Let $\xi$ be the discrete random vector in $\mathbb{R}^2$ with the*

*distribution of $\xi$:*

| | | | | |
|---|---|---|---|---|
| *realizations:* | (0,0) | (1,0) | (0,1) | (1,1) |
| *probabilities:* | 0.1 | 0.2 | 0.1 | 0.6 |

*yielding the expectation $\bar{\xi} = (0.8, 0.7)$. This implies the*

*marginal distributions of $\xi_1$ and $\xi_2$:*

| | | |
|---|---|---|
| *realizations:* | 0 | 1 |
| *probabilities* $\mathbb{P}_{\xi_1}$ : | 0.2 | 0.8 |
| *probabilities* $\mathbb{P}_{\xi_2}$ : | 0.3 | 0.7 |

*being obviously stochastically dependent. Using these marginal distributions to compute the E–M distribution according to Th. 3.5., we get the*

*E–M distribution of $\eta$:*

| | | | | |
|---|---|---|---|---|
| *realizations:* | (0,0) | (1,0) | (0,1) | (1,1) |
| *probabilities:* | 0.06 | 0.24 | 0.14 | 0.56 |

*with the expectation $\bar{\eta} = (0.8, 0.7)$. Then for any convex function $\varphi(\cdot,\cdot)$ such that*

$$\varphi(0,0) = \varphi(1,0) = \varphi(0,1) = 0 \;\; \textit{and} \;\; \varphi(1,1) = 1$$

*we get $\mathbb{E}_\xi[\varphi(\xi)] = 0.6$ and $\mathbb{E}_\eta[\varphi(\eta)] = 0.56$. Hence, in this case, with the E–M distribution (3.27) as derived for the independent case, the E–M inequality (3.28) does not hold.* □

To generalize the E–M inequality for random vectors with dependent components and $\operatorname{supp}\mathbb{P}_\xi \subset \Xi = \prod_{i=1}^r [\alpha_i, \beta_i]$, and for arbitrary convex integrands, according to Frauendorfer [102] we may proceed as follows:

Assume first that for some $\xi \in \Xi$ we have the random vector $\zeta$ with the one-point distribution $\mathbb{P}_\zeta(\{\zeta \mid \zeta = \xi\}) = 1$. Obviously the components of $\zeta$ are stochastically independent, and for $\eta_i(\xi_i)$ with the two-point distributions

$$\left.\begin{aligned}\mathbb{P}_{\eta_i(\xi_i)}(\{\eta_i \mid \eta_i = \alpha_i\}) &= \frac{\beta_i - \xi_i}{\beta_i - \alpha_i} \\ \mathbb{P}_{\eta_i(\xi_i)}(\{\eta_i \mid \eta_i = \beta_i\}) &= \frac{\xi_i - \alpha_i}{\beta_i - \alpha_i}\end{aligned}\right\} \tag{3.33}$$

holds

$$\mathbb{E}_{\eta_i(\xi_i)}[\eta_i] = \xi_i = \mathbb{E}_{\zeta_i}[\zeta_i]. \tag{3.34}$$

Hence for the probability measure

$$\mathbb{P}_{\eta(\xi)} = \mathbb{P}_{\eta_1(\xi_1)} \times \mathbb{P}_{\eta_2(\xi_2)} \times \cdots \times \mathbb{P}_{\eta_r(\xi_r)} \quad \text{on} \quad \Xi = \prod_{i=1}^{r} [\alpha_i, \beta_i], \tag{3.35}$$

defined on the vertices $v^\nu$ of $\Xi$, $\nu = 1, \cdots, 2^r$, we have the probabilities

$$\mathbb{P}_{\eta(\xi)}(v^\nu) = \prod_{i \in I_\nu} \frac{\beta_i - \xi_i}{\beta_i - \alpha_i} \cdot \prod_{i \in J_\nu} \frac{\xi_i - \alpha_i}{\beta_i - \alpha_i},$$

where $I_\nu = \{i \mid v_i^\nu = \alpha_i\}$ and $J_\nu = \{1, \cdots, r\} \setminus I_\nu$ (with $\prod_{i \in \emptyset} \{\cdot\} = 1$). Thus we get immediately

**Lemma 3.7.** *For any convex function $\varphi : \Xi \longrightarrow \mathbb{R}$, Jensen's inequality implies*

$$\left.\begin{aligned}\varphi(\mathbb{E}_\zeta[\zeta]) = \varphi(\xi) &\le \int_\Xi \varphi(\eta(\xi)) \mathbb{P}_{\eta(\xi)}(d\eta) \\ &= \sum_{\nu=1}^{2^r} \varphi(v^\nu) \mathbb{P}_{\eta(\xi)}(v^\nu).\end{aligned}\right\} \tag{3.36}$$

*Hence, with the probability measure $\mathbb{Q}$ defined on the vertices $v^\nu$ of $\Xi$ by*

$$\left.\begin{aligned}\mathbb{Q}(v^\nu) &= \int_\Xi \mathbb{P}_{\eta(\xi)}(v^\nu) \mathbb{P}_\xi(d\xi) \\ &= \int_\Xi \prod_{i \in I_\nu} \frac{\beta_i - \xi_i}{\beta_i - \alpha_i} \cdot \prod_{i \in J_\nu} \frac{\xi_i - \alpha_i}{\beta_i - \alpha_i} \mathbb{P}_\xi(d\xi),\end{aligned}\right\} \tag{3.37}$$

*we get the generalized E–M inequality*

$$\mathbb{E}_\xi[\varphi(\xi)] \le \sum_{\nu=1}^{2^r} \varphi(v^\nu) \mathbb{Q}(v^\nu). \tag{3.38}$$

**Remark 3.4.** *Observe that for stochastically independent components of $\xi$, due to (3.37) we get for the generalized E–M distribution*

$$\mathbb{Q}(v^\nu) = \prod_{i \in I_\nu} \frac{\beta_i - \mu_i}{\beta_i - \alpha_i} \cdot \prod_{i \in J_\nu} \frac{\mu_i - \alpha_i}{\beta_i - \alpha_i},$$

*such that in this case $\mathbb{Q}$ coincides with the E–M distribution $\mathbb{P}_\eta$ for the independent case as derived in (3.26) and (3.27).* □

Hence Theorem 3.5. may be generalized as follows:

**Theorem 3.6.** *Assume that* $\operatorname{supp}\mathbb{P}_\xi \subset \Xi = \prod_{i=1}^r [\alpha_i, \beta_i]$ *such that also* $\mu = \mathbb{E}_\xi[\xi] \in \Xi$. *Under the Assumptions 3.2. and 3.3. and with the generalized E–M distribution $\mathbb{Q}$ as defined in (3.37), according to (3.38) the expected recourse* $\mathscr{Q}(x) = \mathbb{E}_\xi[Q(x;T(\xi),h(\xi))]$ *is bounded above as*

$$\left.\begin{aligned} \mathscr{Q}(x) &\leq \int_\Xi Q(x;T(\eta),h(\eta))\,\mathbb{Q}(d\eta) \\ &= \sum_{\nu=1}^{2^r} Q(x;T(v^\nu),h(v^\nu))\,\mathbb{Q}(v^\nu)\,. \end{aligned}\right\} \tag{3.39}$$

For any $\Lambda \subset \{1,\cdots,r\}$ let $\widehat{m}_\Lambda(\xi) := \prod_{k\in\Lambda} \xi_k$ and denote the joint mixed moments of $\{\xi_k \mid k \in \Lambda\}$ as $\mu_\Lambda := \int_\Xi \widehat{m}_\Lambda(\xi)\mathbb{P}_\xi(d\xi)$ for all $\Lambda \subset \{1,\cdots,r\}$ (with $\widehat{m}_\emptyset(\xi) \equiv 1$ and $\mu_\emptyset = 1$).

Then we have, for any vertex $v^\nu$ of $\Xi$, that $\widehat{m}_\Lambda(v^\nu) = \prod_{k\in\Lambda\cap I_\nu} \alpha_k \cdot \prod_{k\in\Lambda\cap J_\nu} \beta_k$, and from (3.34) and (3.35) follows

$$\int_\Xi \widehat{m}_\Lambda(\eta)\mathbb{P}_{\eta(\xi)}(d\eta) = \widehat{m}_\Lambda(\xi) = \sum_{\nu=1}^{2^r} \widehat{m}_\Lambda(v^\nu)\cdot\mathbb{P}_{\eta(\xi)}(v^\nu)\,, \tag{3.40}$$

such that (3.37) and (3.40) imply

$$\left.\begin{aligned} \sum_{\nu=1}^{2^r} \widehat{m}_\Lambda(v^\nu)\mathbb{Q}(v^\nu) &= \int_\Xi \sum_{\nu=1}^{2^r} \widehat{m}_\Lambda(v^\nu)\mathbb{P}_{\eta(\xi)}(v^\nu)\mathbb{P}_\xi(d\xi) \\ &= \int_\Xi \widehat{m}_\Lambda(\xi)\mathbb{P}_\xi(d\xi) = \mu_\Lambda\,. \end{aligned}\right\} \tag{3.41}$$

Hence the upper bound distribution $\mathbb{Q}$ of Lemma 3.7 preserves all joint moments of the original distribution $\mathbb{P}_\xi$, suggesting to consider, for $\mathscr{P}$ being the set of all probability measures on $\Xi$, the moment problem (P)

$$\begin{aligned} &\gamma(P) := \\ &\sup_{\mathbb{P}\in\mathscr{P}} \left\{ \int_\Xi \varphi(\xi)\mathbb{P}(d\xi) \;\middle|\; \int_\Xi \widehat{m}_\Lambda(\xi)\mathbb{P}(d\xi) = \mu_\Lambda \;\forall \Lambda \subset \{1,\cdots,r\} \right\}. \end{aligned} \tag{3.42}$$

For the dual of this problem we assign the variables $y_0$ to $\Lambda = \emptyset$ ($\mu_\emptyset = 1$) and $y_\Lambda$ to any nonempty subset $\Lambda \subset \{1,\cdots,r\}$. This yields the semi-infinite program (D)

$$\delta(D) := \inf\left\{ y_0 + \sum_{\Lambda \neq \emptyset} \mu_\Lambda y_\Lambda \,\middle|\, y_0 + \sum_{\Lambda \neq \emptyset} \widehat{m}_\Lambda(\xi) y_\Lambda \geq \varphi(\xi)\ \forall \xi \in \Xi \right\}. \tag{3.43}$$

Requiring the constraints of (D) to hold only at the vertices of $\Xi$ yields the modified problem ($\tilde{\text{D}}$)

$$\delta(\tilde{D}) := \inf\left\{ y_0 + \sum_{\Lambda \neq \emptyset} \mu_\Lambda y_\Lambda \,\middle|\, y_0 + \sum_{\Lambda \neq \emptyset} \widehat{m}_\Lambda(v^\nu) y_\Lambda \geq \varphi(v^\nu),\ \nu = 1, \cdots, 2^r \right\}$$

and its dual ($\tilde{\text{P}}$), the moment problem searching for a measure $\mathbb{P}$ in $\mathscr{P}_{\text{ext}\,\Xi}$, the set of probability distributions on the vertices of $\Xi$, becomes

$$\gamma(\tilde{P}) := \sup_{\mathscr{P}_{\text{ext}\,\Xi}} \left\{ \sum_{\nu=1}^{2^r} \varphi(v^\nu) p_\nu \,\middle|\, \sum_{\nu=1}^{2^r} \widehat{m}_\Lambda(v^\nu) p_\nu = \mu_\Lambda\ \forall \Lambda \subset \{1, \cdots, r\} \right\}.$$

Due to (3.41) the upper bound distribution $\mathbb{Q}$ of Lemma 3.7 is feasible for this moment problem ($\tilde{\text{P}}$). Furthermore, since the matrix of the system of linear constraints of ($\tilde{\text{P}}$), i.e.

$$H := (\widehat{m}_\Lambda(v^\nu);\ \nu = 1, \cdots, 2^r,\ \Lambda \subset \{1, \cdots, r\}),$$

is regular, as shown in Kall [157], the generalized E–M distribution $\mathbb{Q}$ is the unique solution of ($\tilde{\text{P}}$) and independent of $\varphi$. Finally, according to linear programming duality and since $\mathscr{P}_{\text{ext}\,\Xi} \subset \mathscr{P}$ we have

$$\delta(\tilde{D}) = \gamma(\tilde{P}) \leq \gamma(P).$$

On the other hand for any $\xi \in \Xi$, given the regularity of $H$, the linear system

$$\sum_{\nu=1}^{2^r} \widehat{m}_\Lambda(v^\nu) q_\nu(\xi) = \widehat{m}_\Lambda(\xi),\ \Lambda \subset \{1, \cdots, r\} \tag{3.44}$$

has the unique solution $\{q_\nu(\xi) = \mathbb{P}_{\eta(\xi)}(v^\nu);\ \nu = 1, \cdots, 2^r\}$ due to (3.40), being continuous in $\xi$. Then for any $\mathbb{P}$ feasible in (P) follows

$$\begin{aligned}
\forall \Lambda \subset \{1, \cdots, r\}:\ \ \mu_\Lambda &= \int_\Xi \widehat{m}_\Lambda(\xi) \mathbb{P}(d\xi) \\
&= \int_\Xi \sum_{\nu=1}^{2^r} \widehat{m}_\Lambda(v^\nu) q_\nu(\xi) \mathbb{P}(d\xi) \\
&= \sum_{\nu=1}^{2^r} \widehat{m}_\Lambda(v^\nu) \hat{q}_\nu \ \text{ with } \hat{q}_\nu = \int_\Xi q_\nu(\xi) \mathbb{P}(d\xi).
\end{aligned}$$

Hence $\{\hat{q}_\nu;\ \nu = 1, \cdots, 2^r\}$ is a probability distribution on the vertices of $\Xi$ which is feasible for the moment problem (P). Since (3.44) also includes $\sum_{\nu=1}^{2^r} v^\nu q_\nu(\xi) = \xi$, by the convexity of $\varphi$ follows for the objective of (P)

$$\sum_{\nu=1}^{2^r} \varphi(v^\nu)\hat{q}_\nu = \int_\Xi \sum_{\nu=1}^{2^r} \varphi(v^\nu) q_\nu(\xi)\mathbb{P}(d\xi) \geq \int_\Xi \varphi(\xi)\mathbb{P}(d\xi)\,.$$

Therefore we have

$$\gamma(\tilde{P}) \geq \gamma(P) \Longrightarrow \gamma(\tilde{P}) = \gamma(P)\,,$$

such that the generalized E–M distribution $\mathbb{Q}$ solves the moment problem (P), and as shown in Kall [157], it is the unique solution of (P).

**Remark 3.5.** *In the above cases we could reduce particular moment problems* (P), *as e.g. (3.42), stated on $\mathscr{P}$, the set of all probability measures on some support $\Xi$, to moment problems* $(\tilde{\mathrm{P}})$ *on $\mathscr{P}_d$, some sets of probability measures with finite discrete supports $\Xi_d \subset \Xi$, such that a solution of* $(\tilde{\mathrm{P}})$ *was simultaneously a solution of* (P).

*This observation is not surprising in view of a very general result, mentioned in Kemperman [181] and assigned to Richter [276] and Rogosinski [289], stated as follows:*

*"Let $f_1,\cdots,f_N$ be integrable functions on the probability space $(\Omega,\mathscr{G},P)$. Then there exists a probability measure $\tilde{P}$ with finite support in $\Omega$ such that*

$$\int_\Omega f_i(\omega)P(d\omega) = \int_\Omega f_i(\omega)\tilde{P}(d\omega)\,,\; i = 1,\cdots,N.$$

*Even* $\mathrm{card}\,(\mathrm{supp}\,\tilde{P}) \leq N+1$ *may be achieved."*

*Hence we can take advantage of the theory of semi-infinite programming. With*

*S, an arbitrary (usually infinite) index set, and*
$a : S \longrightarrow \mathbb{R}^n$, $b : S \longrightarrow \mathbb{R}$, $c \in \mathbb{R}^n$ *arbitrary,*

*the problem*

$$v(P) := \inf\{c^{\mathrm{T}}y \mid a^{\mathrm{T}}(s)y \geq b(s) \;\; \forall s \in S\}$$

*is called a (primal) semi-infinite program. Its dual program requires, for some $s_i \in S$, $i = 1,\cdots,q \geq 1$, to determine a positive finite discrete measure $\mu$ with $\mu(s_i) = x_i$ as a solution of the generalized moment problem*

$$v(D) := \sup\left\{\sum_{i=1}^q b(s_i)x_i \mid \sum_{i=1}^q a(s_i)x_i = c,\, x_i \geq 0,\, s_i \in S,\, q \geq 1\right\}.$$

*Whereas weak duality, i.e. $v(D) \leq v(P)$, is evident, a detailed discussion of statements on (strong) duality as well as on existence of solutions for these two problems under various regularity assumptions may be found in textbooks like Glasshoff–Gustafson [126] and Goberna–López [128] (or in reviews as e.g. in Kall [158]).*

*Moment problems have been considered in detail in probability theory (see e.g. Krein–Nudel'man [196]) and in other areas of applied mathematics (like e.g.*

*Karlin–Studden [176]), and a profound geometric approach was presented in Kemperman [181].*

*In connection with stochastic programs with recourse moment problems were investigated to find upper bounds for the expected recourse, also under assumptions on the set $\Xi$ containing $\operatorname{supp}\mathbb{P}_\xi$ and moment conditions being different from those mentioned above.*

*For instance, for a convex function $\varphi$, $\Xi$ being a (bounded) convex polyhedron, and the feasible set of probability measures $\mathscr{P}$ given by the moment conditions $\int_\Xi \xi\,\mathbb{P}(d\xi) = \bar{\xi}\ (= \mathbb{E}_\xi[\xi])$, the moment problem $\sup_{\mathbb{P}\in\mathscr{P}} \int_\Xi \varphi(\xi)\mathbb{P}(d\xi)$ turns out to be the linear program to determine an optimal discrete measure on the vertices of $\Xi$ where, in contrast to the above E–M measures, the solution depends on $\varphi$ in general (see e.g. Dupačová [74, 75]).*

*Furthermore, for a lower semi-continuous proper convex function $\varphi$ and $\Xi$ being an arbitrary closed convex set, and again with*

$$\mathscr{P} = \left\{ \mathbb{P} \,\middle|\, \int_\Xi \xi\,\mathbb{P}(d\xi) = \bar{\xi} \right\},$$

*the moment problem $\sup_{\mathbb{P}\in\mathscr{P}} \int_\Xi \varphi(\xi)\mathbb{P}(d\xi)$, considered by Birge–Wets [28], amounts to determine a finite discrete probability measure $\mathbb{P}$ on $\operatorname{ext}\Xi$ and a finite discrete nonnegative measure $\nu$ on $\operatorname{ext}\operatorname{rc}\Xi$ (with $\operatorname{rc}\Xi$ the recession cone of $\Xi$, see Rockafellar [281]), which for infinite sets $\operatorname{ext}\Xi$ and $\operatorname{ext}\operatorname{rc}\Xi$ appears to be a difficult task, whereas it seems to become somewhat easier if $\Xi$ is assumed to be a convex polyhedral set as discussed e.g. in Edirisinghe–Ziemba [81], Gassmann–Ziemba [122], Huang–Ziemba–Ben-Tal [144]). Also in these cases, the solutions of the moment problems, i.e. the optimal measures, depend on $\varphi$, in general. For the special situation where $\varphi$ is convex and $\Xi$ is a regular simplex, i.e.*

$$\Xi = \operatorname{conv}\{v^0, v^1, \cdots, v^r\} \subset \mathbb{R}^r,\ \operatorname{rank}(v^1 - v^0, v^2 - v^0, \cdots, v^r - v^0) = r,$$

*mentioned in Birge–Wets [27] and later investigated and used extensively by Frauendorfer [103], the moment problem under the above first order moment conditions has the unique solution of a regular system of linear equations, independent of $\varphi$ again.*

*Finally, for $\Xi = \mathbb{R}^r$ with $\int_\Xi \xi\,\mathbb{P}_\xi(d\xi) = \mu$ and $\int_\Xi \|\xi\|^2\mathbb{P}_\xi(d\xi) = \rho$, (with $\|\cdot\|$ the Euclidean norm) moment problems with the nonlinear moment conditions*

$$\int_\Xi \xi\,\mathbb{P}(d\xi) = \mu \quad \textit{and} \quad \int_\Xi \|\xi\|^2\,\mathbb{P}(d\xi) = \rho$$

*have been discussed, first for simplicial recourse functions $\varphi$ by Dulá [73], and then for more general nonlinear recourse functions in Kall [159]. In these cases, the solutions of the moment problems depend on $\varphi$, in general. Under appropriate assumptions on the recourse functions these moment problems turn out to be non-*

*smooth optimization problems, solvable with bundle-trust methods as described in Schramm–Zowe [299], for instance.*

*We have sketched possibilities to derive upper bounds for the expected recourse using results from the theory on semi-infinite programming and moment problems. Similarly, the theory on partial orderings of spaces of probability measures, as described in Stoyan [313] and Müller–Stoyan [237], could be used. Attempts in this direction may be found e.g. in Frauendorfer [103] and in Kall–Stoyan [171].* □

#### 3.2.1.3 CFR: Approximation by successive discretization

Assuming that, for the given random vector $\xi$, we have $\operatorname{supp}\mathbb{P}_\xi \subset \Xi = \prod_{i=1}^r [\alpha_i, \beta_i]$, due to Jensen and Edmundson–Madansky there follow for any convex function $\varphi$ and $\bar{\xi} = \mathbb{E}_\xi[\xi]$ the bounds

$$\varphi(\bar{\xi}) \le \mathbb{E}_\xi[\varphi(\xi)] \le \mathbb{E}_\eta[\varphi(\eta)] = \int_\Xi \varphi(\eta)\mathbb{Q}(d\eta)\,, \tag{3.45}$$

where $\eta$ has the discrete distribution $\mathbb{Q}$ defined on the vertices of $\Xi$, as described in Lemma 3.7. Hence these bounds result from finitely many arithmetic operations provided the joint moments $\mu_\Lambda := \int_\Xi \widehat{m}_\Lambda(\xi)\mathbb{P}_\xi(d\xi) = \mathbb{E}_\xi[\widehat{m}_\Lambda(\xi)]$ are known for all $\Lambda \subset \{1,\cdots,r\}$.

The following observation is the basis of a method of discrete approximations (of the distribution) to solve complete recourse problems.

Assume that,with half-open or closed intervals $\Xi_k$ as the *cells*, a partition $\mathscr{X}$ of the interval $\Xi$ is given satisfying

$$\mathscr{X} = \{\Xi_k;\ k = 1,\cdots,K\},\ \text{such that}\ \Xi_k \cap \Xi_\ell = \emptyset,\ k \ne \ell,\ \text{and}\ \bigcup_{k=1}^K \Xi_k = \Xi\,. \tag{3.46}$$

Then there follows

**Lemma 3.8.** *Under the above assumptions holds, with $\pi_k = \mathbb{P}_\xi(\Xi_k)$, for the lower bounds of $\mathbb{E}_\xi[\varphi(\xi)]$*

$$\left.\begin{aligned}\varphi(\bar{\xi}) &\le \sum_{k=1}^K \pi_k\varphi(\mathbb{E}_\xi[\xi \mid \xi \in \Xi_k])\\ &\le \sum_{k=1}^K \pi_k\mathbb{E}_\xi[\varphi(\xi) \mid \xi \in \Xi_k]\\ &= \mathbb{E}_\xi[\varphi(\xi)]\end{aligned}\right\} \tag{3.47}$$

*whereas for the upper bounds we get the inequalities*

$$\left.\begin{aligned}\mathbb{E}_\xi\left[\varphi(\xi)\right] &= \sum_{k=1}^{K}\pi_k\mathbb{E}_\xi\left[\varphi(\xi)\mid\xi\in\Xi_k\right]\\ &\le \sum_{k=1}^{K}\pi_k\int_{\Xi_k}\varphi(\eta)\mathbb{Q}_k(d\eta)\\ &\le \int_{\Xi}\varphi(\eta)\mathbb{Q}(d\eta),\end{aligned}\right\} \tag{3.48}$$

*where $\mathbb{Q}_k$ is the E–M distribution on $\Xi_k$ yielding $\mu_\Lambda^k := \mathbb{E}_\xi\left[\widehat{m}_\Lambda(\xi)\mid\xi\in\Xi_k\right]$ for all $\Lambda\subset\{1,\cdots,r\}$ and $k=1,\cdots,K$, and $\mathbb{Q}$ is the E–M distribution on $\Xi$ as described in Lemma 3.7.*

*Proof:* For any $\mathbb{P}_\xi$-integrable function $\psi:\Xi\longrightarrow\mathbb{R}^p,\ p\in\mathbb{N}$, we have the equality

$$\sum_{k=1}^{K}\pi_k\mathbb{E}_\xi\left[\psi(\xi)\mid\xi\in\Xi_k\right]=\mathbb{E}_\xi\left[\psi(\xi)\right]. \tag{3.49}$$

Hence, with $\psi$ the identity, we have $\sum_{k=1}^{K}\pi_k\mathbb{E}_\xi\left[\xi\mid\xi\in\Xi_k\right]=\bar{\xi}$. Then, the convexity of $\varphi$ implies the first inequality of (3.47), whereas the second one follows from the fact that Jensen's inequality holds true for conditional expectations, as well (see Pfanzagl [253]).

The first equation in (3.48) follows from (3.49) with $\psi=\varphi$. The following inequality holds true due to the fact, that the E–M inequality is valid for conditional expectations, as well. For the probability measure $\mathbb{Q}_k$ holds for all $\Lambda\subset\{1,\cdots,r\}$

$$\int_{\Xi_k}\widehat{m}_\Lambda(\xi)\mathbb{Q}_k(d\xi)=\mu_\Lambda^k=\mathbb{E}_\xi\left[\widehat{m}_\Lambda(\xi)\mid\xi\in\Xi_k\right],\ k=1,\cdots,K,$$

such that with $\psi=\widehat{m}_\Lambda$ due to (3.49)

$$\sum_{k=1}^{K}\int_{\Xi_k}\pi_k\widehat{m}_\Lambda(\xi)\mathbb{Q}_k(d\xi)=\sum_{k=1}^{K}\pi_k\mathbb{E}_\xi\left[\widehat{m}_\Lambda(\xi)\mid\xi\in\Xi_k\right]=\mathbb{E}_\xi\left[\widehat{m}_\Lambda(\xi)\right]=\mu_\Lambda\,.$$

Hence, the probability measure $\sum_{k=1}^{K}\pi_k\mathbb{Q}_k$ is feasible for the moment problem (3.42) which is solved by $\mathbb{Q}$, thus implying the last inequality of (3.48). □

Hence, with any arbitrary convex function $\varphi:\Xi\longrightarrow\mathbb{R}$ on the interval $\Xi\subset\mathbb{R}^r$, for any probability distribution $\mathbb{P}_\xi$ on $\Xi$ and for each choice of a partition $\mathscr{X}=\{\Xi_k;\ k=1,\cdots,K\}$ of $\Xi$, we have bounds on $\mathbb{E}_\xi\left[\varphi(\xi)\right]$ by

– a discrete random vector $\eta$ with distribution $\mathbb{P}_{\eta_\mathscr{X}}$ yielding

$$\int_{\Xi}\varphi(\eta)\mathbb{P}_{\eta_\mathscr{X}}(d\eta)\le\mathbb{E}_\xi\left[\varphi(\xi)\right],$$

the Jensen lower bound due to (3.47), and

– a discrete random vector $\eta$ with distribution $\mathbb{Q}_{\eta_{\mathscr{X}}}$ yielding

$$\mathbb{E}_\xi\left[\varphi(\xi)\right] \le \int_\Xi \varphi(\eta)\mathbb{Q}_{\eta_{\mathscr{X}}}(d\eta)\,,$$

the (generalized) E–M upper bound according to (3.48) (with the measure $\mathbb{Q}_{\eta_{\mathscr{X}}} = \sum_{k=1}^{K} \pi_k \mathbb{Q}_k$ in the above notation).

Let a further partition $\mathscr{Y} = \{\Upsilon_l;\ l = 1,\cdots,L\}$ of $\Xi$ be a refinement of $\mathscr{X}$, i.e. each cell of $\mathscr{X}$ is the union of one or several cells of $\mathscr{Y}$, then as an immediate consequence of Lemma 3.8 follows

**Corollary 3.3.** *Under the above assumptions, the partition $\mathscr{Y}$ of $\Xi$ being a refinement of the partition $\mathscr{X}$ implies*

$$\int_\Xi \varphi(\eta)\mathbb{P}_{\eta_{\mathscr{X}}}(d\eta) \le \int_\Xi \varphi(\eta)\mathbb{P}_{\eta_{\mathscr{Y}}}(d\eta) \le \mathbb{E}_\xi\left[\varphi(\xi)\right]$$

*and*

$$\mathbb{E}_\xi\left[\varphi(\xi)\right] \le \int_\Xi \varphi(\eta)\mathbb{Q}_{\eta_{\mathscr{Y}}}(d\eta) \le \int_\Xi \varphi(\eta)\mathbb{Q}_{\eta_{\mathscr{X}}}(d\eta)$$

*and hence an increasing lower and a decreasing upper bound.*

*Proof:* Since $\mathscr{Y}$ is a refinement of $\mathscr{X}$, for $\mathscr{Y} \neq \mathscr{X}$ there is at least one cell $\Xi_k$ of $\mathscr{X}$ being partitioned into some cells $\Upsilon_{l_k 1},\cdots,\Upsilon_{l_k s_k}$ of $\mathscr{Y}$, such that $s_k > 1$ and $\bigcup_{\nu=1}^{s_k} \Upsilon_{l_k \nu} = \Xi_k$. Observing that with $p_{l_k \nu} = \mathbb{P}_\xi(\Upsilon_{l_k \nu})$ holds

$$\mathbb{E}_\xi\left[\xi \mid \xi \in \Xi_k\right] = \frac{1}{\pi_k}\sum_{\nu=1}^{s_k} p_{l_k \nu}\mathbb{E}_\xi\left[\xi \mid \xi \in \Upsilon_{l_k \nu}\right],$$

due to $\sum_{\nu=1}^{s_k} p_{l_k \nu} = \pi_k$ the convexity of $\varphi$ implies

$$\varphi(\mathbb{E}_\xi\left[\xi \mid \xi \in \Xi_k\right]) \le \frac{1}{\pi_k}\sum_{\nu=1}^{s_k} p_{l_k \nu}\varphi(\mathbb{E}_\xi\left[\xi \mid \xi \in \Upsilon_{l_k \nu}\right])\,.$$

Therefore, this increases in (3.47) the $k$-th term

$$\pi_k\varphi(\mathbb{E}_\xi\left[\xi \mid \xi \in \Xi_k\right]) \quad \text{to} \quad \sum_{\nu=1}^{s_k} p_{l_k \nu}\varphi(\mathbb{E}_\xi\left[\xi \mid \xi \in \Upsilon_{l_k \nu}\right])\,.$$

In a similar way, the monotone decreasing of the upper bound may be shown, following the arguments in the proof of Lemma 3.8. □

Hence, refining the partitions of $\Xi$ successively improves the approximation of $\mathbb{E}_\xi[\varphi(\xi)]$, by the Jensen bound from below and by the E–M bound from above. Defining in some partition $\mathscr{X} = \{\Xi_k;\ k = 1, \cdots, K\}$ of $\Xi$ the diameter of any cell $\Xi_k \in \mathscr{X}$ as

$$\operatorname{diam} \Xi_k := \sup\{\|\xi - \eta\| \mid \xi, \eta \in \Xi_k\}$$

and then introducing the grid width of this partition $\mathscr{X}$ as

$$\operatorname{grid} \mathscr{X} := \max_{k=1,\cdots,K} \operatorname{diam} \Xi_k\,,$$

we may prove convergence of the above bounds to $\mathbb{E}_\xi[\varphi(\xi)]$ under appropriate assumptions (see Kall [153]).

**Lemma 3.9.** *Let* $\operatorname{supp} \mathbb{P}_\xi \subseteq \Xi = \prod_{i=1}^{r}[\alpha_i, \beta_i]$ *and* $\varphi : \Xi \longrightarrow \mathbb{R}$ *be continuous. Assume a sequence* $\{\mathscr{X}^\nu\}$ *of successively refined partitions of* $\Xi$ *to be given such that* $\lim_{\nu\to\infty} \operatorname{grid} \mathscr{X}^\nu = 0$*. Then, for* $\{\mathbb{P}_{\eta_{\mathscr{X}^\nu}}\}$ *and* $\{\mathbb{Q}_{\eta_{\mathscr{X}^\nu}}\}$ *the corresponding sequences of Jensen distributions and E–M distributions, respectively, follows*

$$\lim_{\nu\to\infty} \int_\Xi \varphi(\xi)\mathbb{P}_{\eta_{\mathscr{X}^\nu}}(d\xi) = \lim_{\nu\to\infty} \int_\Xi \varphi(\xi)\mathbb{Q}_{\eta_{\mathscr{X}^\nu}}(d\xi) = \int_\Xi \varphi(\xi)\mathbb{P}_\xi(d\xi)\,.$$

*Proof:* Due to our assumptions $\varphi$ is uniformly continuous on $\Xi$ implying

$$\forall \varepsilon > 0\ \exists \delta_\varepsilon > 0 \text{ such that } |\varphi(\xi) - \varphi(\eta)| < \varepsilon\ \forall \xi, \eta \in \Xi :\ \|\xi - \eta\| < \delta_\varepsilon\,.$$

According to the assumptions on $\{\mathscr{X}^\nu\}$ there exists some $\nu(\delta_\varepsilon)$ such that $\operatorname{grid} \mathscr{X}^\nu < \delta_\varepsilon\ \forall \nu > \nu(\delta_\varepsilon)$. Hence, for $\nu > \nu(\delta_\varepsilon)$ and any cell $\Xi_k^\nu \in \mathscr{X}^\nu$ holds $|\varphi(\xi) - \varphi(\eta)| < \varepsilon\ \forall \xi, \eta \in \Xi_k^\nu$. The Jensen distribution $\mathbb{P}_{\eta_{\mathscr{X}^\nu}}$ assigns the probability $\pi_k^\nu = \mathbb{P}_\xi(\Xi_k^\nu) = \int_{\Xi_k^\nu} \mathbb{P}_\xi(d\xi)$ to the realization $\bar{\xi}_k^\nu = \mathbb{E}_\xi[\xi \mid \xi \in \Xi_k^\nu]$. Hence we get

$$\begin{aligned}
&\Big| \int_\Xi \varphi(\xi)\mathbb{P}_{\eta_{\mathscr{X}^\nu}}(d\xi) - \int_\Xi \varphi(\xi)\mathbb{P}_\xi(d\xi) \Big| \\
&= \Big| \sum_{k=1}^{K^\nu} \int_{\Xi_k^\nu} (\varphi(\bar{\xi}_k^\nu) - \varphi(\xi))\mathbb{P}_\xi(d\xi) \Big| \\
&\le \sum_{k=1}^{K^\nu} \int_{\Xi_k^\nu} |\varphi(\bar{\xi}_k^\nu) - \varphi(\xi)|\mathbb{P}_\xi(d\xi) \le \sum_{k=1}^{K^\nu} \varepsilon \cdot \pi_k^\nu = \varepsilon
\end{aligned}$$

such that $\int_\Xi \varphi(\xi)\mathbb{P}_{\eta_{\mathscr{X}^\nu}}(d\xi) \longrightarrow \int_\Xi \varphi(\xi)\mathbb{P}_\xi(d\xi)\,.$

The convergence of the E–M bound may be shown similarly. □

This result gives rise to introduce the following convergence concepts:

**Definition 3.1.** *A sequence of probability measures $\mathbb{P}_\xi^\nu$ on $\mathbb{B}^r$ (the Borel $\sigma$-algebra on $\mathbb{R}^r$) is said to converge weakly to the measure $\mathbb{P}_\xi$ if for the corresponding distribution functions $F_\nu$ and $F$, respectively, holds*

$$\lim_{\nu\to\infty} F_\nu(\xi) = F(\xi) \quad \text{for every continuity point } \xi \text{ of } F.$$

**Definition 3.2.** *Let $\{\psi;\, \psi_\nu,\ \nu \in \mathbb{N}\}$ be a set of functions on $\mathbb{R}^r$. The sequence $\{\psi_\nu,\ \nu \in \mathbb{N}\}$ is said to epi-converge to $\psi$ if for any $\xi \in \mathbb{R}^r$*

- *there exists a sequence $\{\eta_\nu \longrightarrow \xi\}$ such that $\limsup_{\nu\to\infty} \psi_\nu(\eta_\nu) \le \psi(\xi)$,*
- *for all sequences $\{\eta_\nu \longrightarrow \xi\}$ holds $\psi(\xi) \le \liminf_{\nu\to\infty} \psi_\nu(\eta_\nu)$.*

Lemma 3.9 ensures that the sequences of measures $\{\mathbb{P}_{\eta_{\mathscr{X}^\nu}}\}$ and $\{\mathbb{Q}_{\eta_{\mathscr{X}^\nu}}\}$ converge weakly to $\mathbb{P}_\xi$, as shown in Billingsley [20, 21]. Under the Assumptions 3.2. and 3.3., for the recourse function $Q(x;T(\xi),h(\xi))$ (with $\xi \in \Xi$, the above interval) and for any sequence of probability measures $\mathbb{P}_\xi^\nu$ on $\Xi$ converging weakly to $\mathbb{P}_\xi$, it follows that the approximating expected recourse functions $\mathscr{Q}^\nu(x) = \int_\Xi Q(x;T(\xi),h(\xi))\mathbb{P}_\xi^\nu(d\xi)$ epi-converge to the true expected recourse $\mathscr{Q}(x) = \int_\Xi Q(x;T(\xi),h(\xi))\mathbb{P}_\xi(d\xi)$, as has been shown e.g. in Wets [343]; related investigations are found in Robinson–Wets [279] and Kall [156]. The epi-convergence of the $\mathscr{Q}^\nu$ has the following desirable consequence:

**Theorem 3.7.** *Assume that $\{\mathscr{Q}^\nu\}$ epi-converges to $\mathscr{Q}$. Then, with some convex polyhedral set $X \subset \mathbb{R}^n$, for the two-stage SLP with recourse we have*

$$\limsup_{\nu\to\infty}[\inf_X\{c^{\mathrm{T}}x + \mathscr{Q}^\nu(x)\}] \le \inf_X\{c^{\mathrm{T}}x + \mathscr{Q}(x)\}.$$

*If*

$$\hat{x}^\nu \in \arg\min_X\{c^{\mathrm{T}}x + \mathscr{Q}^\nu(x)\}\ \forall \nu \in \mathbb{N},$$

*then for any accumulation point $\hat{x}$ of $\{\hat{x}^\nu\}$ it follows that*

$$c^{\mathrm{T}}\hat{x} + \mathscr{Q}(\hat{x}) = \min_X\{c^{\mathrm{T}}x + \mathscr{Q}(x)\};$$

*and for any subsequence $\{\hat{x}^{\nu_\kappa}\} \subset \{\hat{x}^\nu\}$ with $\lim_{\kappa\to\infty} \hat{x}^{\nu_\kappa} = \hat{x}$ we have*

$$c^{\mathrm{T}}\hat{x} + \mathscr{Q}(\hat{x}) = \lim_{\kappa\to\infty}\{c^{\mathrm{T}}\hat{x}^{\nu_\kappa} + \mathscr{Q}(\hat{x}^{\nu_\kappa})\}.$$

A proof of this statement may be found for instance in Wets [343] (see also Kall [155]).

Due to this result discrete approximation algorithms (DAPPROX) for the solution of two-stage SLP's with recourse may be designed, based on successive partitions $\{\mathscr{X}^\nu\}$ of $\Xi$, yielding lower bounds

$$\mathscr{Q}^\nu_{LB}(x) = \int_\Xi Q(x;T(\xi),h(\xi))\mathbb{P}_{\eta_{\mathscr{X}^\nu}}(d\xi) \tag{3.50}$$

and upper bounds

$$\mathscr{Q}^\nu_{UB}(x) = \int_\Xi Q(x;T(\xi),h(\xi))\mathbb{Q}_{\eta_{\mathscr{X}^\nu}}(d\xi) \tag{3.51}$$

for $\mathscr{Q}(\cdot) = \mathbb{E}_\xi\,[Q(\cdot;T(\xi),h(\xi))]$ due to Jensen and Edmundson–Madansky, respectively. In other words, a solution of

$$\hat{\gamma} := \min_X\{c^{\mathrm{T}}x + \mathscr{Q}(x)\} \tag{3.52}$$

may be approximated by an approach like

**DAPPROX: Approximating CFR solutions**

With $\Xi$ an interval and $\operatorname{supp}\mathbb{P}_\xi \subset \Xi$, let $\mathscr{X}^1 := \{\Xi\}$ be the first (trivial) partition of $\Xi$ and $\mathbb{P}_{\eta_{\mathscr{X}^1}}$, $\mathbb{Q}_{\eta_{\mathscr{X}^1}}$ the corresponding Jensen– and E–M – distributions determining, due to (3.50) and (3.51), the approximating expected recourse functions $\mathscr{Q}^1_{LB}(x)$ and $\mathscr{Q}^1_{UB}(x)$.

With $\nu = 1$ iterate the cycle of the following steps I.–III. until achieving the required accuracy $\varepsilon > 0$ of an approximate solution.

I. Analyze the approximating problems

$$\text{a)}\ \hat{\gamma}_{LB} := \min_X\{c^{\mathrm{T}}x + \mathscr{Q}^\nu_{LB}(x)\} \quad \text{and} \quad \text{b)}\ \hat{\gamma}_{UB} := \min_X\{c^{\mathrm{T}}x + \mathscr{Q}^\nu_{UB}(x)\}\,.$$

{*Observe that, since* $\mathbb{P}_{\eta_{\mathscr{X}^\nu}}$ *and* $\mathbb{Q}_{\eta_{\mathscr{X}^\nu}}$ *are finite discrete distributions, problems* a) *and* b) *are LP's with decomposition structures.*}

II. If the prescribed accuracy is achieved, stop the procedure; otherwise, go on to step III.
{*As an example, with a solution* $\hat{x}^\nu$ *of problem* I.a) *and its optimal value* $\hat{\gamma}^\nu_{LB}$*, the error estimate* $\hat{\gamma} - \hat{\gamma}^\nu_{LB} \le c^{\mathrm{T}}\hat{x} + \mathscr{Q}^\nu_{UB}(\hat{x}^\nu) - \hat{\gamma}^\nu_{LB} =: \delta^\nu$ *might be used to check whether* $\delta^\nu \le \varepsilon$*, if this corresponds to the required accuracy.*}

III. To improve the approximation, choose a partition $\mathscr{X}^{\nu+1}$ as an appropriate refinement of $\mathscr{X}^\nu$. With the corresponding Jensen– and E–M – distributions $\mathbb{P}_{\eta_{\mathscr{X}^{\nu+1}}}$ and $\mathbb{Q}_{\eta_{\mathscr{X}^{\nu+1}}}$, defining $\mathscr{Q}^{\nu+1}_{LB}(x)$ and $\mathscr{Q}^{\nu+1}_{UB}(x)$ due to (3.50) and (3.51), let $\nu := \nu + 1$ and return to step I. above.

{*As mentioned above, due to the finite discrete Jensen– and E–M – distributions the problems* I.a) *and* I.b) *generated in each cycle are LP's with decomposition structure as shown in (1.10) on page 4. For the error estimate mentioned at* II. *the LP* I.a) *has to be solved, suggesting to apply an appropriate decomposition algorithm. In general and due to many experiments, QDECOM (see p. 47) can be considered as a proven reliable solver for this purpose. Nevertheless, keep in mind Remark 1.2. (p. 48).*} □

Obviously, this conceptual description of DAPPROX gives rise to quite a variety of algorithms, depending on various strategies of refining the partitions. For instance, the selection of the particular cells to be refined is relevant for the effectiveness of the method. Or for a cell $\Xi_k^\nu \in \mathscr{X}^\nu$ to be refined, in order to maintain the assumed interval structure of the successive partitions, through this cell we need a cut being perpendicular to one of the coordinate axes; but which coordinate axis is to be preferred, and where is the cut to be located? These and further strategies, playing a significant role for the efficiency of DAPPROX–solvers implementations, will be discussed later in Section 4.7.2.

## Exercises

**3.1.** Consider the two–stage SLP

$$\begin{array}{llllll}
\max\{2x_1 + x_2 + \sum_{k=1}^{3} p_k q^{\mathrm{T}} y^{(k)}\} & & & & & \\
x_1 \ + x_2 & & & \leq 10 & & \\
x_1 + 2x_2 \ - y_1^{(k)} & + y_2^{(k)} & + 2y_3^{(k)} & = h_1^{(k)} & & \\
x_1 \ - x_2 + 2y_1^{(k)} & + 3y_2^{(k)} & + y_3^{(k)} & = h_2^{(k)} & & k = 1,2,3, \\
 & & x_j, y_\nu^{(k)} & \geq 0 & & \forall\, j, k, \nu
\end{array}$$

with $q = (-2,-3,-2)^{\mathrm{T}}$, $h^{(1)} = (5,4)^{\mathrm{T}}$, $h^{(2)} = (3,5)^{\mathrm{T}}$, $h^{(3)} = (2,2)^{\mathrm{T}}$.

(a) Has the problem the (relatively) complete recourse property?
(b) If not, determine the induced constraints for $(x_1, x_2)^{\mathrm{T}}$ (see page 198).
(c) Compute the first stage solution and its first stage objective of the problem.

You may verify your answers to items (a) and (c) using SLP-IOR.

**3.2.** Change the recourse matrix $W$ of exercise **3.1** to the new recourse matric

$$\widetilde{W} = \begin{pmatrix} -1 & 1 & -1 \\ 2 & 3 & -5 \end{pmatrix};$$

(a) check the complete recourse property;
(b) compute the first stage solution (including its objective) and compare it to the result of exercise **3.1**.

You may use SLP-IOR to confirm your answers.

**3.3.** Let the function $\psi : [-1,1] \times [0,2] \longrightarrow \mathbb{R}$ be defined as $\psi(\xi,\eta) := \varphi(\xi) + \theta(\eta)$ with $\varphi(\xi) := \begin{cases} -\xi & \text{for } -1 \le \xi \le -\frac{1}{4} \\ 0.5+\xi & \text{for } -\frac{1}{4} \le \xi \le 1 \end{cases}$ and $\theta(\eta) := 2\eta$. Assume $\xi$ and $\eta$ to be independent random variables with the densities $g_\xi(\zeta) \equiv \dfrac{1}{2}$ for $\zeta \in [-1,1]$ and $h_\eta(\zeta) = \dfrac{e^{-\zeta}}{1-e^{-2}}$. Due to the independence of $\xi$ and $\eta$ holds

$$\begin{aligned}\mathbb{E}[\psi(\xi,\eta)] &= \int_0^2 \int_{-1}^1 \psi(\xi,\eta) d\xi d\eta \\ &= \int_{-1}^1 \varphi(\xi) d\xi + \int_0^2 \theta(\eta) d\eta = 0.781250 + 1.373929 = 2.155179\,.\end{aligned}$$

(a) Compute $(\bar{\xi},\bar{\eta}) = \mathbb{E}[(\xi,\eta)]$ and Jensen's bound $\psi(\bar{\xi},\bar{\eta}) = \varphi(\bar{\xi}) + \theta(\bar{\eta})$.
(b) Compute the E–M upper bound of $\mathbb{E}[\psi(\xi,\eta)]$.
(c) Subdivide the support $\Xi = [-1,1] \times [0,2]$ into two rectangles,

(c1) either by dividing the $\xi$-interval $I^{(\xi)} = [-1,1]$ at $\bar{\xi}$ into $I_1^{(\xi)}$ and $I_2^{(\xi)}$
(c2) or by dividing the $\eta$-interval $I^{(\eta)} = [0,2]$ at $\bar{\eta}$ into $I_1^{(\eta)}$ and $I_2^{(\eta)}$.

Compute alternatively the two new Jensen bounds of $\mathbb{E}[\psi(\xi,\eta)]$, as either $lb_{|\bar{\xi}} := \mathbb{P}_\xi(I_1^{(\xi)}) \cdot \varphi(\mathbb{E}_\xi[\xi \mid I_1^{(\xi)}]) + \mathbb{P}_\xi(I_2^{(\xi)}) \cdot \varphi(\mathbb{E}_\xi[\xi \mid I_2^{(\xi)}]) + \theta(\mathbb{E}_\eta[\eta])$, or else
$lb_{|\bar{\eta}} := \varphi(\mathbb{E}_\xi[\xi]) + \mathbb{P}_\eta(I_1^{(\eta)}) \cdot \theta(\mathbb{E}_\eta[\eta \mid I_1^{(\eta)}]) + \mathbb{P}_\eta(I_2^{(\eta)}) \cdot \theta(\mathbb{E}_\eta[\eta \mid I_2^{(\eta)}])$.
(d) How do the new Jensen bounds $lb_{|\bar{\xi}}$ and $lb_{|\bar{\eta}}$ compare to the first bound $\psi(\bar{\xi},\bar{\eta})$, and how much does the above error estimate decrease at best?

**3.4.** Concerning the moment problem (3.24) and its dual, the semi-infinite program (3.25), it was claimed, that

(a) a linear affine function majorizes a convex function $\psi(\cdot)$ on an interval, if and only if it does so at the endpoints of the interval;
(b) due to the relation $\alpha \le \mu \le \beta$ (with the natural assumption that $\alpha < \beta$) the LP corresponding to (3.25) (due to (a)) is solvable and hence its dual, the LP equivalent to the moment problem (3.24), is uniquely solvable.

Show that these claims hold true.

**3.5.** Let $F$ be a convex function on a convex polyhedron $\mathscr{B} = \operatorname{conv}\{z^{(1)},\cdots,z^{(k)}\}$, the support of some distribution $\mathbb{P}_\zeta$ with expectation $EX[\zeta] = \mu \in \mathscr{B}$. There is a lower bound for $\bar{F} = \int_{\mathscr{B}} F(z)\mathbb{P}_\zeta(dz)$ given by $F(\mu)$ due to Jensen, and as mentioned on page 217 referring to Dupačová, an upper bound can be determined by solving an LP, which maximizes $EX[F]$ on the class $\tilde{P}$ of dicrete distributions on the vertices of $\mathscr{B}$, satisfying the moment conditions $\sum_i p_i \cdot z^{(i)} = \mu,\ \sum_i p_i = 1,\ p_i \ge 0\ \forall i$.

As an example, define $\mathscr{B} := \{\xi \mid \xi_1 + 2\xi_2 \leq 10\,,\ 2\xi_1 + \xi_2 \leq 8\,,\ \xi_i \geq 0\}$. Assume some distribution $\mathbb{P}_\xi$ on $\mathscr{B}$ with first moments $\mu = \mathbb{E}[\xi] = (2;2)^{\mathrm{T}}$. Finally define on $\mathscr{B}$ a function $F(\xi) = \xi^{\mathrm{T}} M \xi + c^{\mathrm{T}}\xi$ with $M = \begin{pmatrix} 3 & 2 \\ 2 & 7 \end{pmatrix}$ and $c = (-18; -46)^{\mathrm{T}}$.

(a) Is $F(\xi)$ a convex function on $\mathscr{B}$ (and why)? If so:
(b) Compute the Jensen lower bound of $EX[F(\xi)]$ re $\mathbb{P}_\xi$.
(c) Find an upper bound of $EX[F(\xi)]$ re $\mathbb{P}_\xi$ as an LP-solution as described above.

**3.6.** Consider the recourse problem

$$\begin{array}{l} \min_x \{c^{\mathrm{T}}x + \mathbb{E}[Q(x;\xi,\eta)] \mid Ax \leq b,\ x \geq 0\} \ \text{ where} \\ Q(x;\xi,\eta) := \min_y \{q^{\mathrm{T}}y \mid Wy = h(\xi,\eta) - Tx,\ y \geq 0\} \ \text{ with} \end{array}$$

the data: $c = (3,5)^{\mathrm{T}}$; $b = (18,18)^{\mathrm{T}}$; $q = (2,3,2,1)^{\mathrm{T}}$; $h = (12+\xi, 22+\eta)^{\mathrm{T}}$; the arrays $A = \begin{pmatrix} 1 & 3 \\ 3 & 2 \end{pmatrix}$ ; $W = \begin{pmatrix} 1 & 1 & 1 & -3 \\ 2 & 1 & -4 & 2 \end{pmatrix}$ ; $T = \begin{pmatrix} 2 & 2 \\ 5 & 3 \end{pmatrix}$; and the random variables $\xi$ and $\eta$, with $\xi$ distributed with density $\varphi(\zeta) = \lambda \cdot e^{-\lambda\zeta}/(1-e^{9\lambda})$ (exponential, conditional to the interval [0,9], or else truncated at the confidence interval of $\hat{p} = 0.95$), and $\eta$ distributed as $\mathscr{U}[-10,10]$ (uniform).

(a) Is this problem of complete fixed recourse?
(b) Compute on the support $\Xi = [0,9] \times [-10,10]$ of $(\xi,\eta)$ the lower (Jensen) and upper (E–M) bound for the optimal value and the resulting error estimate.
(c) Compute (e.g with SLP-IOR) the corresponding bounds for partitioning the support $\Xi$ into two (dividing the $\eta$-interval) and four subintervals (dividing the $\xi$– and the $\eta$-interval once, each).

### 3.2.2 *The simple recourse case*

For the special complete recourse case with $q(\xi) \equiv (q^{+\mathrm{T}}, q^{-\mathrm{T}})^{\mathrm{T}}$ and $W = (I, -I)$, we get the *generalized simple recourse* (GSR) function

$$\left.\begin{array}{rl} Q^G(x,\xi) := \min\ q^{+\mathrm{T}}y^+ + q^{-\mathrm{T}}y^- & \\ Iy^+ - \quad Iy^- &= h(\xi) - T(\xi)x \\ y^+, \quad y^- &\geq 0\,. \end{array}\right\} \tag{3.53}$$

Given that $\xi$ is a random vector in $\mathbb{R}^R$ such that $\mathbb{E}_\xi[\xi]$ exists, we have the *expected generalized simple recourse* (EGSR)

$$\mathscr{Q}^G(x) := \mathbb{E}_\xi\left[Q^G(x,\xi)\right], \tag{3.54}$$

yielding the two-stage SLP with *generalized simple recourse* (GSR)

$$
\left.\begin{array}{l} \min\{c^{\mathrm{T}}x + \mathscr{Q}^{G}(x)\} \\ Ax = b \\ x \geq 0 . \end{array}\right\} \tag{3.55}
$$

Before dealing with GSR problems, it is meaningful to discuss first the original version of simple recourse problems, as first analyzed in detail by Wets [342].

#### 3.2.2.1 The standard simple recourse problem (SSR)

In contrast to (3.53) it is now assumed in addition that $T(\xi) \equiv T$. Then it is obviously meaningful to let $h(\xi) \equiv \xi \in \mathbb{R}^{m_2}$ such that instead of (3.53) the *standard simple recourse* (SSR) function is given as

$$
\left.\begin{array}{rl} Q(x;\xi) := \min & {q^{+}}^{\mathrm{T}} y^{+} + {q^{-}}^{\mathrm{T}} y^{-} \\ & Iy^{+} - \quad Iy^{-} = \xi - Tx \\ & y^{+}, \qquad y^{-} \geq 0 . \end{array}\right\} \tag{3.56}
$$

This implies the expected simple recourse $\mathscr{Q}(x) = \mathbb{E}_{\xi}\left[Q(x;\xi)\right]$.

Obviously, problem (3.56) is always feasible; and it is solvable iff its dual program

$$
\left.\begin{array}{l} \max(\xi - Tx)^{\mathrm{T}} u \\ u \leq \quad q^{+} \\ u \geq -q^{-} \end{array}\right\} \tag{3.57}
$$

is feasible, which in turn is true iff $q^{+} + q^{-} \geq 0$. Considering (3.57), we get immediately the optimal recourse value as

$$
Q(x,\xi) = \sum_{i=1}^{m_2} [(\xi - Tx)_i]^{+} q_i^{+} + \sum_{i=1}^{m_2} [(\xi - Tx)_i]^{-} q_i^{-} \tag{3.58}
$$

where, for $\rho \in \mathbb{R}$,

$$
[\rho]^{+} = \begin{cases} \rho \ \text{if}\ \rho > 0 \\ 0 \ \text{else} \end{cases} \quad \text{and} \quad [\rho]^{-} = \begin{cases} -\rho \ \text{if}\ \rho < 0 \\ 0 \quad \text{else.} \end{cases}
$$

This optimal recourse value $Q(x,\xi)$ is achieved in (3.56) by choosing

$$
y_i^{+} = [(\xi - Tx)_i]^{+} \quad \text{and} \quad y_i^{-} = [(\xi - Tx)_i]^{-}, \ \ i = 1, \cdots, m_2. \tag{3.59}
$$

Introducing $\chi := Tx$, we get from (3.59) the optimal value of (3.56) as

$$
\left.\begin{array}{rl} \tilde{Q}(\chi,\xi) := & \sum_{i=1}^{m_2} \left\{ q_i^{+} \left[\xi_i - \chi_i\right]^{+} + q_i^{-} \left[\xi_i - \chi_i\right]^{-} \right\} \\ =: & \sum_{i=1}^{m_2} \tilde{Q}_i(\chi_i,\xi_i) \end{array}\right\} \tag{3.60}
$$

with

$$\begin{aligned}\tilde{Q}_i(\chi_i, \xi_i) &= q_i^+ [\xi_i - \chi_i]^+ + q_i^- [\xi_i - \chi_i]^- \\ &= \min\{q_i^+ y_i^+ + q_i^- y_i^- \mid y_i^+ - y_i^- = \xi_i - \chi_i;\ y_i^+, y_i^- \geq 0\}.\end{aligned} \tag{3.61}$$

Hence the recourse function $Q(x, \xi)$ of (3.56) may be rewritten as a function $\tilde{Q}(\chi, \xi)$ being separable in $(\chi_i, \xi_i)$, implying the expected recourse $\mathscr{Q}(x)$ to be equivalent to a function $\widetilde{\mathscr{Q}}(\chi)$, separable in $\chi_i$ (see Wets [342]), according to

$$\left.\begin{aligned}\widetilde{\mathscr{Q}}(\chi) &= \sum_{i=1}^{m_2} \widetilde{\mathscr{Q}}_i(\chi_i), \quad \text{where} \\ \widetilde{\mathscr{Q}}_i(\chi_i) &:= \mathbb{E}_\xi [\tilde{Q}_i(\chi_i, \xi_i)] = \mathbb{E}_{\xi_i} [\tilde{Q}_i(\chi_i, \xi_i)],\ i = 1, \cdots, m_2,\end{aligned}\right\} \tag{3.62}$$

such that (3.55) may now be rewritten as

$$\left.\begin{array}{l}\min\{c^{\mathrm{T}} x + \sum_{i=1}^{m_2} \widetilde{\mathscr{Q}}_i(\chi_i)\} \\ Ax \qquad = b \\ Tx - \chi = 0 \\ x \qquad \geq 0\end{array}\right\} \tag{3.63}$$

with

In this case, as indicated by the operator $\mathbb{E}_{\xi_i}$, to compute the expected simple recourse we may restrict ourselves to the marginal distributions of the single components $\xi_i$ instead of the joint distribution of $\xi = (\xi_1, \cdots, \xi_{m_2})^{\mathrm{T}}$. From (3.61) obviously follows that $\tilde{Q}_i(\cdot, \xi_i)$ is a convex function in $\chi_i$ (and hence in $x$) for any fixed value of $\xi_i$. Hence, the expected recourse $\widetilde{\mathscr{Q}}_i(\cdot)$ is convex in $\chi_i$ as well.

If $\mathbb{P}_\xi$ happens to be a finite discrete distribution with the marginal distribution of any component given by $p_{ij} = \mathbb{P}_\xi(\{\xi \mid \xi_i = \hat{\xi}_{ij}\}),\ j = 1, \cdots, k_i$, then (3.63) is equivalent to the linear program

$$\left.\begin{array}{l}\min\{c^{\mathrm{T}} x + \sum_{i=1}^{m_2} \sum_{j=1}^{k_i} p_{ij}(q_i^+ y_{ij}^+ + q_i^- y_{ij}^-)\} \\ Ax \qquad\qquad = b \\ Tx - \chi \qquad\quad = 0 \\ \qquad\quad y_{ij}^+ - y_{ij}^- = \hat{\xi}_{ij} - \chi_i \quad \forall i, j \\ x, \quad y_{ij}^+, \quad y_{ij}^- \geq 0\end{array}\right\} \tag{3.64}$$

which due to its special data structure can easily be solved.

If, on the other hand, $\mathbb{P}_\xi$ or at least some of its marginal distributions $\mathbb{P}_{\xi_i}$ are of the continuous type, the corresponding expected recourse $\widetilde{\mathscr{Q}}_i(\cdot)$ and hence the program (3.63) may be expected to be nonlinear. Nevertheless, the simple recourse functions $\tilde{Q}_i(\chi_i, \xi_i)$ and their expectations $\widetilde{\mathscr{Q}}_i(\chi_i)$ have some special properties, advantageous in solution procedures and not shared by complete recourse functions in general. To point out these particular properties we introduce *simple recourse type* functions (referred to as SRT functions) and discuss some of their properties advantageous for their approximation.

**Definition 3.3.** *For a real variable z, a random variable $\xi$ with distribution $\mathbb{P}_\xi$, and real constants $\alpha, \beta, \gamma$ with $\alpha + \beta \geq 0$, the function $\varphi(\cdot,\cdot)$ given by*

$$\varphi(z,\xi) := \alpha \cdot [\xi - z]^+ + \beta \cdot [\xi - z]^- - \gamma$$

*is called a simple recourse type function (see Fig. 3.2).*

*Then, $\mathbb{E}_\xi[\xi]$ provided to exist,*

$$\Phi(z) := \mathbb{E}_\xi[\varphi(z,\xi)] = \int_{-\infty}^{\infty} (\alpha \cdot [\xi - z]^+ + \beta \cdot [\xi - z]^-) \mathbb{P}_\xi(d\xi) - \gamma$$

*is the expected SRT function (ESRT function).*

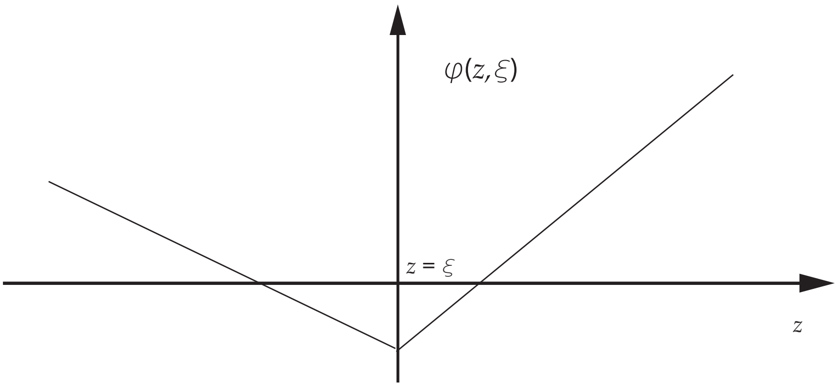


**Fig. 3.2** SRT function.

Obviously, the functions $\tilde{Q}_i(\chi_i, \xi_i)$ and $\widetilde{\mathcal{Q}}_i(\chi_i)$ considered above are SRT and ESRT functions, respectively; however, SRT functions may also appear in models different from (3.61)–(3.63), as we shall see later.

From Definition 3.3. follows immediately

**Lemma 3.10.** *Let $\varphi(\cdot,\cdot)$ be a SRT function and $\Phi(\cdot)$ the corresponding expected SRT function. Then*

- $\varphi(z,\cdot)$ *is convex in $\xi$ for any fixed $z \in \mathbb{R}$;*
- $\varphi(\cdot,\xi)$ *is convex in z for any fixed $\xi \in \mathbb{R}$;*
- $\Phi(\cdot)$ *is convex in z.*

Since (3.61)–(3.63) describes a particular complete fixed recourse problem, we know already from Section 3.2.1 that, $\xi$ provided to be integrable and $q^+ + q^- \geq 0$, the functions $\tilde{Q}_i(\chi_i, \xi_i)$ and $\widetilde{\mathcal{Q}}_i(\chi_i)$ are SRT and ESRT functions, respectively.

Assuming $\mu := \mathbb{E}_\xi[\xi]$ to exist, Jensen's inequality for SRT functions obviously holds:

$$\varphi(z,\mu) = \varphi(z, \mathbb{E}_\xi[\xi]) \le \mathbb{E}_\xi[\varphi(z,\xi)] = \Phi(z).$$

Furthermore, for $\xi$ being integrable (with $F_\xi$ the distribution function of $\xi$), the asymptotic behaviour of the ESRT function may immediately be derived:

**Lemma 3.11.** *For*

$$\begin{aligned}\Phi(z) &:= \mathbb{E}_\xi[\varphi(z,\xi)] \\ &= \int_{-\infty}^{\infty} (\alpha \cdot [\xi - z]^+ + \beta \cdot [\xi - z]^-)\, dF_\xi(\xi) - \gamma \\ &= \left\{ \alpha \cdot \int_z^{\infty} [\xi - z]\, dF_\xi(\xi) + \beta \cdot \int_{-\infty}^{z} [z - \xi]\, dF_\xi(\xi) \right\} - \gamma\end{aligned}$$

*holds:*

$$\Phi(z) - \varphi(z,\mu) = \Phi(z) - [\alpha \cdot (\mu - z) - \gamma] \longrightarrow 0 \;\; as \;\; z \to -\infty$$

*and analogously*

$$\Phi(z) - \varphi(z,\mu) = \Phi(z) - [\beta \cdot (z - \mu) - \gamma] \longrightarrow 0 \;\; as \;\; z \to +\infty.$$

*In particular follows:*

$$\text{If} \begin{cases} \mathbb{P}_\xi(\xi < a) = 0 \\ \mathbb{P}_\xi(\xi > b) = 0 \end{cases} \text{then} \;\; \Phi(z) = \begin{cases} \alpha \cdot (\mu - z) - \gamma = \varphi(z,\mu) \text{ for } z \le a \\ \beta \cdot (z - \mu) - \gamma = \varphi(z,\mu) \text{ for } z \ge b. \end{cases}$$

Hence we have, as mentioned above,

$$\varphi(z,\mu) \le \Phi(z) \;\forall z$$

and, furthermore (see Fig. 3.3),

$$\begin{aligned} a &:= \inf \operatorname{supp} \mathbb{P}_\xi > -\infty \implies \Phi(z) = \varphi(z,\mu) \;\forall z \le a \\ b &:= \sup \operatorname{supp} \mathbb{P}_\xi < +\infty \implies \Phi(z) = \varphi(z,\mu) \;\forall z \ge b. \end{aligned}$$

Consider now an interval $I = \{\xi \mid a < \xi \le b\} \not\supseteq \operatorname{supp} \mathbb{P}_\xi$—implying at least one of the bounds $a, b$ to be finite—with $\mathbb{P}_\xi(I) > 0$. Then Jensen's inequality holds as well for the corresponding conditional expectations.

**Lemma 3.12.** *With* $\mu_{|I} = \mathbb{E}_\xi[\xi \mid \xi \in I]$ *and* $\Phi_{|I}(z) = \mathbb{E}_\xi[\varphi(z,\xi) \mid \xi \in I]$*, for all* $z \in \mathbb{R}$ *holds*

$$\varphi(z, \mu_{|I}) \le \Phi_{|I}(z) = \frac{1}{\mathbb{P}_\xi(I)} \int_a^b \varphi(z,\xi)\, dF_\xi(\xi).$$

As shown in Kall-Stoyan [171], in analogy to Lemma 3.11 follows also

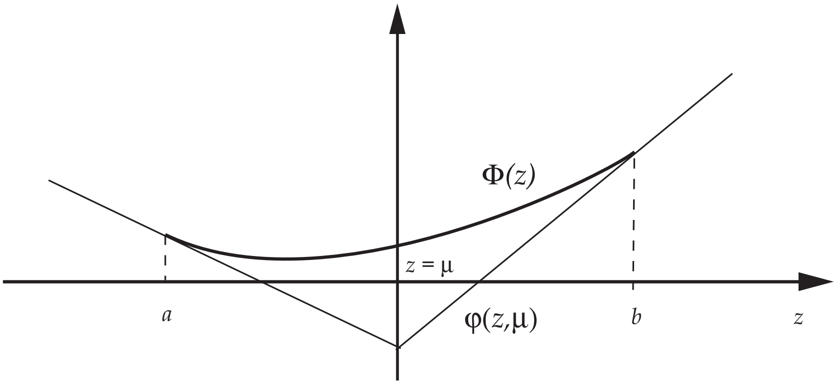


**Fig. 3.3** SRT and expected SRT function (supp $\mathbb{P}_\xi$ bounded).

**Lemma 3.13.** *For any finite a and/or b, for* $I = (a,b]$ *holds*

$$\Phi_{|I}(z) = \begin{cases} \varphi(z,\mu_{|I}) \text{ for } z \le a \\ \varphi(z,\mu_{|I}) \text{ for } z \ge b\,. \end{cases}$$

If in particular $J := \operatorname{supp} \mathbb{P}_\xi = [a,b]$ is a finite interval, then Lemma 3.11 yields

$$\Phi(z) = \Phi_{|J}(z) = \varphi(z,\mu_{|J}) = \varphi(z,\mu) \quad \text{for } z \le a \text{ or } z \ge b\,, \tag{3.65}$$

and for $z \in (a,b)$ Jensen's inequality implies $\varphi(z,\mu) \le \Phi(z)$. To get an upper bound for $z \in (a,b)$ and hence an estimate for $\Phi(z)$, the E–M inequality may be used:

$$\Phi_{|J}(z) = \Phi(z) \le \frac{b-\mu}{b-a}\,\varphi(z,a) + \frac{\mu-a}{b-a}\,\varphi(z,b) = \frac{b-\mu_{|J}}{b-a}\,\varphi(z,a) + \frac{\mu_{|J}-a}{b-a}\,\varphi(z,b)\,.$$

Analogously, for an interval $I = \{\xi \mid a < \xi \le b\} \not\supseteq \operatorname{supp} \mathbb{P}_\xi$ and $z \in \operatorname{int} I$ follows

$$\varphi(z,\mu_{|I}) \le \Phi_{|I}(z) \le \frac{b-\mu_{|I}}{b-a}\,\varphi(z,a) + \frac{\mu_{|I}-a}{b-a}\,\varphi(z,b)\,. \tag{3.66}$$

If $\varphi(z,\cdot)$ happens to be linear on $I$, the lower and upper bounds of these inequalities coincide such that $\Phi_{|I}(z) = \varphi(z,\mu_{|I})\ \forall z$. If, on the other hand, $\varphi(z,\cdot)$ is nonlinear (convex) in $I$, the approximation of $\Phi_{|I}(\hat{z})$ for any $\hat{z} \in (a,b)$ due to (3.66) can be improved as follows: Partition $I = (a,b]$ at $a_1 := \hat{z}$ into the two intervals $I_1 := (a_0,a_1]$ and $I_2 := (a_1,a_2]$, where $a_0 := a$ and $a_2 := b$. Observing that, with $\pi_I = \mathbb{P}_\xi(I)$ and $p_\nu := \mathbb{P}_\xi(I_\nu),\ \nu = 1,2$, we have $\dfrac{p_1}{\pi_I}\cdot\mu_{|I_1} + \dfrac{p_2}{\pi_I}\cdot\mu_{|I_2} = \mu_{|I}$ as well as for arbitrary $\mathbb{P}_\xi$-integrable functions $\psi(\cdot)$ the relation

$$\mathbb{E}_\xi\left[\psi(\xi) \mid \xi \in I\right] = \frac{p_1}{\pi_I}\cdot\mathbb{E}_\xi\left[\psi(\xi) \mid \xi \in I_1\right] + \frac{p_2}{\pi_I}\cdot\mathbb{E}_\xi\left[\psi(\xi) \mid \xi \in I_2\right], \tag{3.67}$$

Lemma 3.12 implies

**Lemma 3.14.** *Due to the convexity of* $\varphi(z,\cdot)$*, we have*

a) *for arbitrary* $z \in (a_0, a_2)$

$$\left.\begin{aligned} \varphi(z,\mu_{|I}) &= \varphi(z, \frac{p_1}{\pi_I}\cdot\mu_{|I_1} + \frac{p_2}{\pi_I}\cdot\mu_{|I_2}) \\ &\le \frac{p_1}{\pi_I}\cdot\varphi(z,\mu_{|I_1}) + \frac{p_2}{\pi_I}\cdot\varphi(z,\mu_{|I_2}) \\ &\le \frac{p_1}{\pi_I}\cdot\Phi_{|I_1}(z) + \frac{p_2}{\pi_I}\cdot\Phi_{|I_2}(z) \\ &= \Phi_{|I}(z)\,; \end{aligned}\right\} \tag{3.68}$$

b) *for* $a_\kappa \in \{a_0, a_1, a_2\}$

$$\left.\begin{aligned} \Phi_{|I_\nu}(a_\kappa) &= \varphi(a_\kappa, \mu_{|I_\nu}) \qquad \text{for } \nu = 1,2 \\ \Phi_{|I}(a_\kappa) &= \sum_{\nu=1}^{2}\frac{p_\nu}{\pi_I}\Phi_{|I_\nu}(a_\kappa) = \sum_{\nu=1}^{2}\frac{p_\nu}{\pi_I}\varphi(a_\kappa,\mu_{|I_\nu}). \end{aligned}\right\} \tag{3.69}$$

*Proof:* The above relations are consequences of previously mentioned facts:

a) The two equations reflect (3.67), the first inequality follows from the convexity of $\varphi(z,\cdot)$, and the second inequality applies Lemma 3.12.
b) The first two equations apply Lemma 3.13, the last equation uses (3.67) again.

□

Based on Lemmas 3.12–3.14, similar to the general complete recourse case, approximation schemes with successively refined discrete distributions may be designed.

#### 3.2.2.2 SSR: Approximation by successive discretization

Eq. (3.68) yields with $\varphi^\star(z,\mu_{|I_1},\mu_{|I_2}) = \frac{p_1}{\pi_I}\varphi(z,\mu_{|I_1}) + \frac{p_2}{\pi_I}\varphi(z,\mu_{|I_2}) \le \Phi_{|I}(z)$ an increased lower bound of $\Phi_{|I}(z)$ as

$$\varphi^\star(z,\mu_{|I_1},\mu_{|I_2}) = \begin{cases} \varphi(z,\mu_{|I}) \text{ for } z \in (-\infty,\mu_{|I_1}) \cup (\mu_{|I_2},\infty) \\ \left\{\left(\frac{p_1}{\pi_I}\beta - \frac{p_2}{\pi_I}\alpha\right)z - \frac{p_1}{\pi_I}\beta\mu_{|I_1} + \frac{p_2}{\pi_I}\alpha\mu_{|I_2} - \gamma\right\} \\ \qquad\qquad \text{for } z \in [\mu_{|I_1},\mu_{|I_2}], \end{cases} \tag{3.70}$$

and instead of the general upper bound (3.66) of $\Phi_{|I}(z)$, with $\hat{z} = a_1$ (3.69) yields the exact value $\Phi_{|I}(\hat{z}) = \frac{p_1}{\pi_I}\varphi(\hat{z},\mu_{|I_1}) + \frac{p_2}{\pi_I}\varphi(\hat{z},\mu_{|I_2}) = \varphi^\star(\hat{z},\mu_{|I_1},\mu_{|I_2})$ (see Fig. 3.4).

If, on the other hand, the partition $I = (a_0,a_1] \cup (a_1,a_2] = I_1 \cup I_2$ is given, from Lemma 3.13 and 3.14 together with (3.67) follows

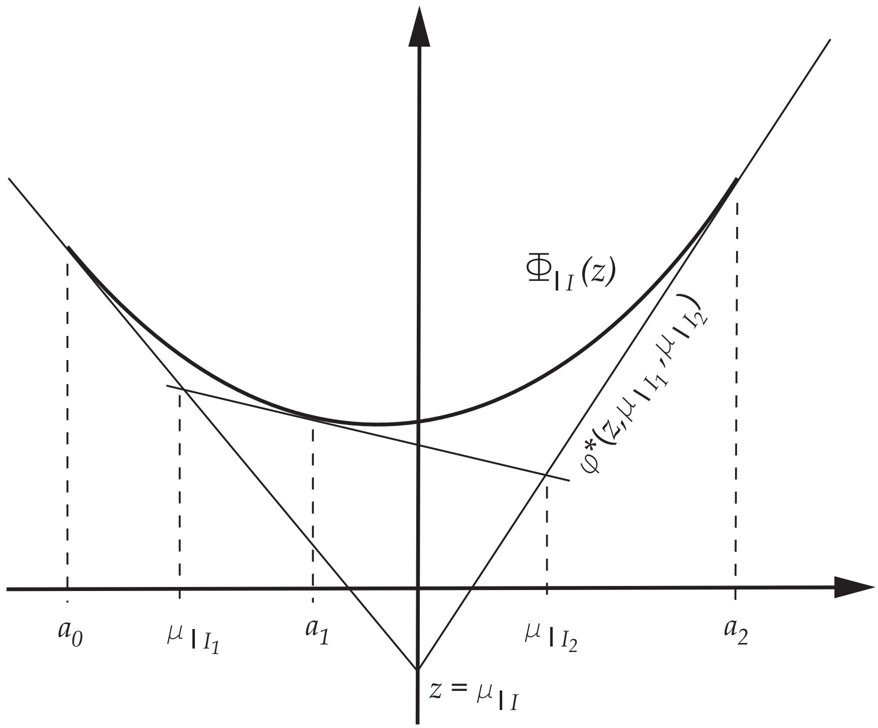


**Fig. 3.4** Expected SRT function: Increasing lower bounds.

$$\Phi_{|I}(z) = \frac{p_1}{\pi_I}\varphi(z,\mu_{|I_1}) + \frac{p_2}{\pi_I}\varphi(z,\mu_{|I_2}) \text{ for } z \le a_0 \text{ or } z \ge a_2 \text{ or } z = a_1\,; \qquad (3.71)$$

hence $\Phi_{|I}(z) > \frac{p_1}{\pi_I}\varphi(z,\mu_{|I_1}) + \frac{p_2}{\pi_I}\varphi(z,\mu_{|I_2})$ may occur only if $z \in \operatorname{int} I_1 \cup \operatorname{int} I_2$, which implies that $\Phi_{|I_\nu}(z) > \varphi(z,\mu_{|I_\nu})$ for $z \in \operatorname{int} I_\nu$ with $\nu = 1$ or $\nu = 2$. Then we may derive the following rather rough error estimate:

**Lemma 3.15.** *For $z \in \operatorname{int} I_\nu$, $\nu = 1,2$, we have the parameter-free error estimate $\Delta_\nu(z)$ satisfying*

$$0 \le \Delta_\nu(z) := \Phi_{|I_\nu} - \varphi(z,\mu_{|I_\nu}) \le \frac{1}{2}(\alpha+\beta)\frac{a_\nu - a_{\nu-1}}{2}\,.$$

*Proof:* Using the relations $\varphi(z,\mu_{|I_\nu}) = \alpha[\mu_{|I_\nu} - z]^+ + \beta[\mu_{|I_\nu} - z]^- - \gamma$ from Definition 3.3. as well as the relations

$$\Phi_{|I_\nu}(a_{\nu-1}) = \varphi(a_{\nu-1},\mu_{|I_\nu}) \text{ and } \Phi_{|I_\nu}(a_\nu) = \varphi(a_\nu,\mu_{|I_\nu})$$

from Lemma 3.14, and furthermore the convexity of $\Phi_{|I_\nu}$ according to Lemma 3.10, we get for $z = \lambda a_{\nu-1} + (1-\lambda)a_\nu$ with $\lambda \in (0,1)$

$$
\begin{aligned}
\Delta_\nu(z) &= \Phi_{|I_\nu}(z) - \varphi(z, \mu_{|I_\nu}) \\
&\le \lambda \Phi_{|I_\nu}(a_{\nu-1}) + (1-\lambda)\Phi_{|I_\nu}(a_\nu) - \varphi(z, \mu_{|I_\nu}) \\
&= \lambda \Phi_{|I_\nu}(a_{\nu-1}) + (1-\lambda)\Phi_{|I_\nu}(a_\nu) \\
&\qquad - \begin{cases} [\alpha(\mu_{|I_\nu} - z) - \gamma] & \text{if } z < \mu_{|I_\nu} \\ [\beta(z - \mu_{|I_\nu}) - \gamma] & \text{if } z \ge \mu_{|I_\nu} \end{cases} \\
&= \lambda[\alpha(\mu_{|I_\nu} - a_{\nu-1}) - \gamma] + (1-\lambda)[\beta(a_\nu - \mu_{|I_\nu}) - \gamma] \\
&\qquad - \begin{cases} [\alpha(\mu_{|I_\nu} - z) - \gamma] & \text{if } z < \mu_{|I_\nu} \\ [\beta(z - \mu_{|I_\nu}) - \gamma] & \text{if } z \ge \mu_{|I_\nu}\,. \end{cases}
\end{aligned}
$$

Assuming

$$
z \le \mu_{|I_\nu} \Longleftrightarrow \lambda \ge \frac{a_\nu - \mu_{|I_\nu}}{a_\nu - a_{\nu-1}} \;\text{ and }\; 1-\lambda \le \frac{\mu_{|I_\nu} - a_{\nu-1}}{a_\nu - a_{\nu-1}}
$$

it follows that

$$
\begin{aligned}
\Delta_\nu(z) &\le \lambda[\alpha(\mu_{|I_\nu} - a_{\nu-1}) - \gamma] + (1-\lambda)[\beta(a_\nu - \mu_{|I_\nu}) - \gamma] \\
&\qquad -[\alpha(\mu_{|I_\nu} - \lambda a_{\nu-1} - (1-\lambda)a_\nu) - \gamma] \\
&= (1-\lambda)(\alpha+\beta)(a_\nu - \mu_{|I_\nu}) \\
&\le \frac{\mu_{|I_\nu} - a_{\nu-1}}{a_\nu - a_{\nu-1}}(\alpha+\beta)(a_\nu - \mu_{|I_\nu}),
\end{aligned}
$$

the maximum of the last term being assumed for $\mu_{|I_\nu} = \dfrac{a_{\nu-1} + a_\nu}{2}$ such that

$$
\Delta_\nu(z) \le \frac{1}{2}(\alpha+\beta)\frac{a_\nu - a_{\nu-1}}{2}\,.
$$

For $z \ge \mu_{|I_\nu}$ the result follows analogously. □

Taking the probabilities $p_\nu$ associated with the partition intervals $I_\nu$ into account yields an improved (global) error estimate:

**Lemma 3.16.** *Given the interval partition* $\{I_\nu;\ \nu = 1,2\}$ *of* $I$ *and* $z \in I_\kappa$, *then the (global) error estimate* $\Delta(z)$ *satisfies*

$$
0 \le \Delta(z) = \Phi_{|I}(z) - \sum_{\nu=1}^{2} \frac{p_\nu}{\pi_I}\varphi(z, \mu_{|I_\nu}) \le \frac{1}{2}\frac{p_\kappa}{\pi_I}(\alpha+\beta)\frac{a_\kappa - a_{\kappa-1}}{2}
$$

*for* $z \in \operatorname{int} I_\kappa$, *whereas for* $z \notin (\operatorname{int} I_1 \cup \operatorname{int} I_2)$ *we have* $\Delta(z) = 0$.

*Proof:* For $z \in I_\kappa$ Lemma 3.13 yields $\Phi_{|I_\nu}(z) - \varphi(z, \mu_{|I_\nu}) = 0$ for $\nu \neq \kappa$; hence from Lemmas 3.14 and 3.15 follows for $z \in \operatorname{int} I_\kappa$

$$\begin{aligned}\Delta(z) &= \Phi_{|I}(z) - \sum_{\nu=1}^{2} \frac{p_\nu}{\pi_I}\, \varphi(z, \mu_{|I_\nu}) \\ &= \sum_{\nu=1}^{2} \frac{p_\nu}{\pi_I}\, (\Phi_{|I_\nu}(z) - \varphi(z, \mu_{|I_\nu})) \\ &= \frac{p_\kappa}{\pi_I}\, (\Phi_{|I_\kappa}(z) - \varphi(z, \mu_{|I_\kappa})) \\ &\leq \frac{1}{2}\, \frac{p_\kappa}{\pi_I}\, (\alpha + \beta)\, \frac{a_\kappa - a_{\kappa-1}}{2}\,,\end{aligned}$$

and for $z \notin (\operatorname{int} I_1 \cup \operatorname{int} I_2)$ from (3.71) follows that $\Delta(z) = 0$. □

Due to (3.60) and (3.63) the simple recourse function $\tilde{Q}(\chi, \xi) = \sum_{i=1}^{m_2} \tilde{Q}^{(i)}(\chi_i, \xi_i)$ as well as the expected simple recourse function $\widetilde{\mathscr{Q}}(\chi) = \sum_{i=1}^{m_2} \widetilde{\mathscr{Q}}^{(i)}(\chi_i)$ are separable, and their additive components $\tilde{Q}^{(i)}(\chi_i, \xi_i)$ and $\widetilde{\mathscr{Q}}^{(i)}(\chi_i)$ are SRT and ESRT functions, respectively. Therefore, the properties derived for these functions allow for modifications of the discrete approximation algorithms of the type DAPPROX, as described on page 223 for the more general complete recourse case. This leads for the standard simple recourse case to special algorithms (named SRAPPROX), being more efficient than the general DAPPROX approach since, for an interval $I^{(i)} \supset \operatorname{supp} \mathbb{P}_{\xi_i}$, at any partitioning point $\hat{\xi}_i := \hat{\chi}_i \in \operatorname{int} I^{(i)}$, instead of the E–M upper bound of $\widetilde{\mathscr{Q}}^{(i)}_{|I}(\hat{\chi})$ its exact value is—due to (3.71) and Lemma 3.14—easily computed.

**SRAPPROX: Approximating SSR solutions**

Assume that $\operatorname{supp} \mathbb{P}_\xi \subset \Xi := \prod_{i=1}^{m_2} I^{(i)}$ for some intervals $I^{(i)} = (a^{(i)}, b^{(i)}]$, $i = 1, \cdots, m_2$. For each component $\xi_i$ of $\xi$ choose as a first partition $\mathscr{X}^{(i)} = \{I^{(i)}\}$ corresponding for $\Xi \subset \mathbb{R}^{m_2}$ to the first partition $\mathscr{X} = \{\mathscr{X}^{(1)} \times \mathscr{X}^{(2)} \times \cdots \times \mathscr{X}^{(m_2)}\}$. For the trivial discrete ditribution $\pi_{i1} = \mathbb{P}_{\xi_i}(\{\xi_i \in I_1^{(i)}\}) = 1\ \forall i$, with $I_1^{(i)} = (a_0^{(i)}, a_1^{(i)}] = I^{(i)}$ and with $\tilde{Q}^{(i)}(\chi_i, \xi_i) = q_i^+ [\xi_i - \chi_i]^+ + q_i^- [\xi_i - \chi_i]^-$ due to (3.61), it follows for $\mu_{|I_1^{(i)}} := \mathbb{E}_{\xi_i}[\xi_i \mid \xi_i \in I_1^{(i)}]$ that $\tilde{Q}^{(i)}(\chi_i, \mu_{|I_1^{(i)}}) \leq \widetilde{\mathscr{Q}}^{(i)}_{|I_1^{(i)}}(\chi_i) = \mathbb{E}_{\xi_i}[\tilde{Q}_i(\chi_i, \xi_i) \mid \xi_i \in I_1^{(i)}]$. With $K_i = 1\ \forall i$ iterate the following cycle:

I. Find a solution $(\hat{x}, \hat{\chi})$ of

$$\min\{c^{\mathrm{T}}x + \sum_{i=1}^{m_2}\sum_{\nu=1}^{K_i} \pi_{i\nu}\tilde{Q}^{(i)}(\chi_i, \mu_{|I_\nu^{(i)}}) \mid Ax = b,\, Tx - \chi = 0,\, x \geq 0\}.$$

If $\hat{\chi}_i \notin \cup_{\nu=1}^{K_i} \mathrm{int} I_\nu^{(i)}$ for all $i \in \{1, \cdots, m_2\}$, then $(\hat{x}, \hat{\chi})$ solves problem (3.63) due to Lemma 3.13 and (3.67) and hence stop; otherwise continue.

II. With $I_\nu^{(i)} = (a_{\nu-1}^{(i)}, a_\nu^{(i)}]$, $\nu = 1, \cdots, K_i$, let $\Lambda := \{i \mid \hat{\chi}_i \in \mathrm{int} I_{\nu_i}^{(i)}$ for one $\nu_i\}$. For $i \in \Lambda$ split up $I_{\nu_i}^{(i)}$ as $I_{\nu_i}^{(i)} = (a_{\nu_i-1}^{(i)}, \hat{\chi}_i] \cup (\hat{\chi}_i, a_{\nu_i}^{(i)}] =: I_{\nu_i 1}^{(i)} \cup I_{\nu_i 2}^{(i)}$ and determine the conditional expectations $\mu_{|I_{\nu_i j}^{(i)}} := \mathbb{E}_{\xi_i}[\xi_i \mid \xi_i \in I_{\nu_i j}^{(i)}]$, $j = 1, 2$. Due to Lemma 3.14 this implies for $\widetilde{\mathscr{Q}}_{|I_{\nu_i}^{(i)}}^{(i)}(\hat{\chi}_i)$ a lower bound $\ell_{\nu_i}^{(i)}$ and the exact value, respectively, to be given as

$$\ell_{\nu_i}^{(i)} = \tilde{Q}^{(i)}(\hat{\chi}_i, \mu_{|I_{\nu_i}^{(i)}}) \leq \widetilde{\mathscr{Q}}_{|I_{\nu_i}^{(i)}}^{(i)}(\hat{\chi}_i)$$

$$\widetilde{\mathscr{Q}}_{|I_{\nu_i}^{(i)}}^{(i)}(\hat{\chi}_i) = \sum_{j=1}^{2} \frac{p_{ij}}{\pi_{\nu_i}} \tilde{Q}^{(i)}(\hat{\chi}_i, \mu_{|I_{\nu_i j}}^{(i)})$$

with $p_{ij} = \mathbb{P}_{\xi_i}(\{\xi_i \in I_{\nu_i j}^{(i)}\})$, $j = 1, 2$.
If for $\delta^{(i)} := \pi_{\nu_i} \cdot (\widetilde{\mathscr{Q}}_{|I_{\nu_i}^{(i)}}^{(i)}(\hat{\chi}_i) - \ell_{\nu_i}^{(i)})$ follows that $\delta^{(i)} < \varepsilon\ \forall i \in \Lambda$ with $\varepsilon$ a prescribed tolerance, then stop with the required accuracy achieved; otherwise continue with $\tilde{\Lambda} := \{i \in \Lambda \mid \delta^{(i)} \geq \varepsilon\}$.

III. For (some) $i \in \tilde{\Lambda}$ extend $\mathscr{X}^{(i)}$ to the new partition $\mathscr{Y}^{(i)}$ by splitting up the interval $I_{\nu_i}^{(i)}$ into the two subintervals $I_{\nu_i j}^{(i)}$ with $\pi_{\nu_i j} := p_{ij}$, $j = 1, 2$, and adjust $K_i := K_i + 1$. With the new data $I_{\nu_i j}^{(i)}$, $\pi_{\nu_i j}$, $\mu_{|I_{\nu_i j}^{(i)}}$ (for $j = 1, 2$) and $K_i$, update the extended partitions to $\mathscr{X}^{(i)} := \mathscr{Y}^{(i)}$ and return to step I.

□

This conceptual algorithm does, in contrast to DAPPROX, leave no choice of where to split an interval $I_{\nu_i}^{(i)}$, $i \in \tilde{\Lambda}$, as long as the true value $\widetilde{\mathscr{Q}}_{|I_{\nu_i}^{(i)}}^{(i)}(\hat{\chi}_i)$, and thus also the exact value of $\widetilde{\mathscr{Q}}^{(i)}(\hat{\chi}_i)$, are of interest. On the other hand there are various strategies for the selection of components $i \in \tilde{\Lambda}$, for which the respective subintervals $I_{\nu_i}^{(i)}$ are splitted up. A detailed description of an executable version of SRAPPROX, including the presentation of the implemented algorithm, can be found in Section 4.7.2 of the next chapter.

#### 3.2.2.3 The multiple simple recourse problem

The simple recourse function (3.56) was extended by Klein Haneveld [188] to the multiple simple recourse function. Here, instead of (3.61), for any single recourse constraint the following value is to be determined:

$$\left.\begin{array}{rcll}\psi(z,\xi) := \min \left\{ \sum_{k=1}^{K} q_k^+ y_k^+ + \sum_{k=1}^{K} q_k^- y_k^- \right\} & & & \\ \sum_{k=1}^{K} y_k^+ - \sum_{k=1}^{K} y_k^- & = & \xi - z & \\ \left.\begin{array}{rcl} y_k^+ & \leq & u_k - u_{k-1} \\ y_k^- & \leq & l_k - l_{k-1} \end{array}\right\}, & & & k = 1,\cdots,K-1, \\ y_k^+, y_k^- & \geq & 0, & k = 1.\cdots,K, \end{array}\right\} \tag{3.72}$$

where

$$\begin{array}{l} u_0 = 0 < u_1 < \cdots < u_{K-1} \\ l_0 \ = 0 < l_1 \ < \cdots < l_{K-1}\,, \end{array}$$

and

$$q_k^+ \geq q_{k-1}^+\,,\; q_k^- \geq q_{k-1}^-\,,\; k = 2,\cdots,K\,,$$

with $q_1^+ \geq -q_1^-$ and $q_K^+ + q_K^- > 0$ (to ensure convexity and prevent from linearity of this modified recourse function).

According to these assumptions, for any value of $\tau := \xi - z$ it is obvious to specify a feasible solution of (3.72), namely for any $\kappa \in \{1,\cdots,K\}$ (with $u_K = \infty$ and $l_K = \infty$)

$$\tau \in [u_{\kappa-1}, u_\kappa) \Longrightarrow \begin{cases} y_k^+ = u_k - u_{k-1}\,, & 1 \leq k \leq \kappa - 1 \\ y_\kappa^+ = \tau - u_{\kappa-1} & \\ y_k^+ = 0 & \forall k > \kappa \\ y_k^- = 0 & k = 1,\cdots,K; \end{cases}$$

$$\tau \in (-l_\kappa, -l_{\kappa-1}] \Longrightarrow \begin{cases} y_k^+ = 0 & k = 1,\cdots,K \\ y_k^- = l_k - l_{k-1}\,, & 1 \leq k \leq \kappa - 1 \\ y_\kappa^- = \tau - l_{\kappa-1} & \\ y_k^- = 0 & \forall k > \kappa\,. \end{cases}$$

Furthermore, this feasible solution is easily seen to be optimal along the following arguments:

- Due to the increasing marginal costs (for surplus as well as for shortage), assuming $\tau \in [u_{\kappa-1}, u_\kappa)$ and $y_k^- = 0\ \forall k$, it is certainly meaningful to exhaust the available capacities for the variables $y_1,\cdots,y_{\kappa-1}$ first. The same argument holds true if $\tau \in (-l_\kappa, -l_{\kappa-1}]$ and $y_k^+ = 0\ \forall k$.
- Assuming a feasible solution of (3.72) with some $y_{k_1}^+$ as well as some $y_{k_2}^-$ simultaneously being greater than some $\delta > 0$, allows to reduce these variables to $\widehat{y}_{k_1}^+ = y_{k_1}^+ - \delta$ and $\widehat{y}_{k_2}^- = y_{k_2}^- - \delta$, yielding a new feasible solution with the objec-

tive changed by $(-\delta)\cdot(q_{k_1}^{+}+q_{k_2}^{-})$ with $(q_{k_1}^{+}+q_{k_2}^{-})\geq 0$ due to the assumptions. Therefore, the modified feasible solution is at least as good as the original one as far as minimization of the objective is concerned.

Hence, for $\tau=\xi-z\in[u_{\kappa-1},u_\kappa)$ with $\kappa\in\{1,\cdots,K\}$ we get

$$\begin{aligned}\psi(z,\xi) &= \left\{\sum_{k=1}^{K} q_k^{+}y_k^{+}+\sum_{k=1}^{K} q_k^{-}y_k^{-}\right\}\\ &= \sum_{k=1}^{\kappa-1} q_k^{+}(u_k-u_{k-1})+q_\kappa^{+}(\tau-u_{\kappa-1})\\ &= \sum_{k=1}^{\kappa-1} q_k^{+}u_k-\sum_{k=0}^{\kappa-2} q_{k+1}^{+}u_k+q_\kappa^{+}(\tau-u_{\kappa-1})\\ &= \sum_{k=1}^{\kappa-2}(q_k^{+}-q_{k+1}^{+})u_k+q_{\kappa-1}^{+}u_{\kappa-1}-q_1^{+}u_0+q_\kappa^{+}(\tau-u_{\kappa-1})\\ &= \sum_{k=0}^{\kappa-1}(q_k^{+}-q_{k+1}^{+})u_k+q_\kappa^{+}\tau \quad \text{with} \quad u_0=0,\; q_0^{+}=0\,.\end{aligned}$$

Defining

$$\alpha_0:=q_1^{+},\; \alpha_k:=q_{k+1}^{+}-q_k^{+},\, k=1,\cdots,K-1\,,$$

it follows immediately that

$$q_k^{+}=\sum_{\nu=1}^{k}\alpha_{\nu-1} \quad \text{for} \quad k=1,\cdots,K$$

such that

$$\psi(z,\xi)=-\sum_{k=0}^{\kappa-1}\alpha_k u_k+\sum_{k=0}^{\kappa-1}\alpha_k\cdot\tau=\sum_{k=0}^{\kappa-1}\alpha_k(\tau-u_k)=\sum_{k=0}^{K-1}\alpha_k[\tau-u_k]^{+}\,.$$

Analogously, for $\tau=\xi-z\in(-l_\kappa,-l_{\kappa-1}]$ with $\kappa\in\{1,\cdots,K\}$ we get

$$\psi(z,\xi)=\sum_{k=0}^{K-1}\beta_k[\tau+l_k]^{-}$$

with $\beta_0:=q_1^{-}$, $\beta_k:=q_{k+1}^{-}-q_k^{-}$, $k=1,\cdots,K-1$, such that in general

$$\psi(z,\xi)=\sum_{k=0}^{K-1}\alpha_k[\tau-u_k]^{+}+\sum_{k=0}^{K-1}\beta_k[\tau+l_k]^{-}\,.$$

Due to the assumptions on (3.72), we have $\alpha_0+\beta_0\geq 0$ as well as

$$\alpha_k \geq 0\,,\ \beta_k \geq 0\,,\ \forall k \in \{1,\cdots,K-1\}\ \text{ and } \sum_{k=0}^{K-1}(\alpha_k+\beta_k) = q_K^+ + q_K^- > 0\,.$$

Hence, whereas the SRT function

$$\varphi(z,\xi) := \alpha\cdot[\xi - z]^+ + \beta\cdot[\xi - z]^- - \gamma$$

according to Definition 3.3. represents the optimal objective value with a simple recourse constraint and implies for some application constant marginal costs for shortage and surplus, respectively, we now have the objective's optimal value for a so-called multiple simple recourse constraint, allowing to model increasing marginal costs for shortage and surplus, respectively, which may be more appropriate for particular real life problems.

To study properties of this model in more detail it is meaningful to introduce *multiple simple recourse type* functions (referred to as MSRT functions) as follows.

**Definition 3.4.** *For real constants* $\{\alpha_k,\beta_k,u_k,l_k; k=0,\cdots,K-1\}$ *and* $\gamma$*, such that* $\alpha_0+\beta_0 \geq 0$ *and*

$$\begin{aligned}
&\alpha_k \geq 0,\ \beta_k \geq 0\ \ \textit{for}\ k=1,\cdots,K-1\ \ \textit{with}\ \sum_{k=0}^{K-1}(\alpha_k+\beta_k) > 0\,,\\
&u_0\ = 0 < u_1\ < \cdots < u_{K-1}\,,\\
&l_0\ \ = 0 < l_1\ \ < \cdots < l_{K-1}\,,
\end{aligned}$$

*the function* $\psi(\cdot,\cdot)$ *given by*

$$\psi(z,\xi) := \sum_{k=0}^{K-1}\{\alpha_k\cdot[\xi - z - u_k]^+ + \beta_k\cdot[\xi - z + l_k]^-\} - \gamma$$

*is called a multiple simple recourse type function (see Fig. 3.5).*

$$\begin{aligned}
\Psi(z) &= \mathbb{E}_\xi\,[\psi(z,\xi)]\\
&= \int_{-\infty}^{\infty}\sum_{k=0}^{K-1}\{\alpha_k\cdot[\xi - z - u_k]^+ + \beta_k\cdot[\xi - z + l_k]^-\}\,dF_\xi\,(\xi) - \gamma
\end{aligned}$$

*is the expected MSRT function.*

**Remark 3.6.** *In this definition the number of "shortage pieces" and of "surplus pieces" is assumed to coincide (with K). Obviously this is no restriction. If, for instance, we had for the number L of "surplus pieces" that* $L < K$*, with the trivial modification*

$$l_k = l_{k-1}+1,\ \beta_k = 0\ \ \textit{for}\ \ k = L,\cdots,K-1$$

*we would have that*

$$\begin{aligned}\psi(z,\xi) &:= \sum_{k=0}^{K-1} \alpha_k \cdot [\xi - z - u_k]^+ + \sum_{k=0}^{L-1} \beta_k \cdot [\xi - z + l_k]^- \} - \gamma \\ &= \sum_{k=0}^{K-1} \{\alpha_k \cdot [\xi - z - u_k]^+ + \beta_k \cdot [\xi - z + l_k]^- \} - \gamma.\end{aligned}$$

□

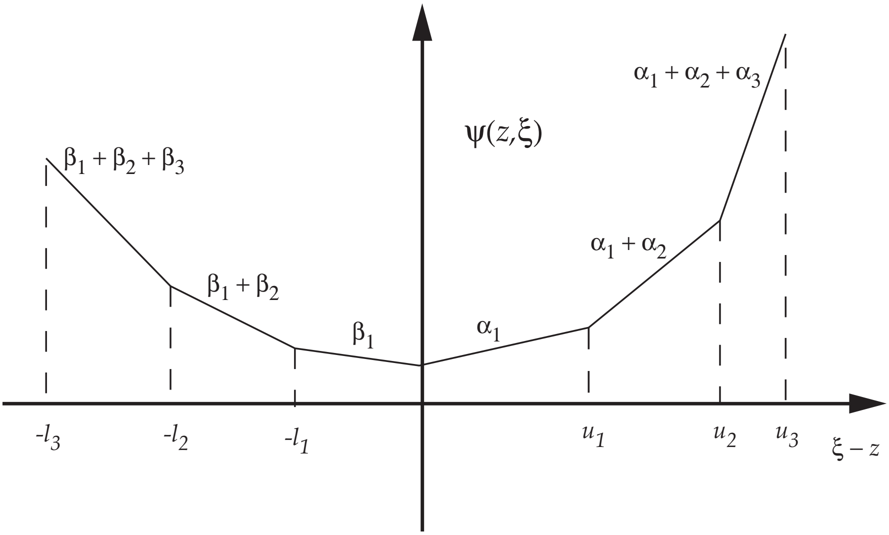


**Fig. 3.5** MSRT function.

For the expected MSRT function we have

$$\begin{aligned}&\Psi(z) + \gamma = \\ &= \sum_{k=0}^{K-1} \left\{ \alpha_k \int_{-\infty}^{\infty} [\xi - z - u_k]^+ dF_\xi(\xi) + \beta_k \int_{-\infty}^{\infty} [\xi - z + l_k]^- dF_\xi(\xi) \right\} \\ &= \sum_{k=0}^{K-1} \left\{ \alpha_k \int_{z+u_k}^{\infty} (\xi - z - u_k)\, dF_\xi(\xi) + \beta_k \int_{-\infty}^{z-l_k} (z - l_k - \xi)\, dF_\xi(\xi) \right\} \\ &= \sum_{k=0}^{K-1} \alpha_k \int_{z}^{\infty} (\eta - z)\, dF_\xi(\eta + u_k) + \sum_{k=0}^{K-1} \beta_k \int_{-\infty}^{z} (z - \zeta)\, dF_\xi(\zeta - l_k) \\ &= \sum_{k=0}^{K-1} \alpha_k \int_{-\infty}^{\infty} (\xi - z)^+ \, dF_\xi(\xi + u_k) + \sum_{k=0}^{K-1} \beta_k \int_{-\infty}^{\infty} (\xi - z)^- \, dF_\xi(\xi - l_k),\end{aligned}$$

using the substitutions $\eta = \xi - u_k$ and $\zeta = \xi + l_k$ (and $\xi = \eta$ and $\xi = \zeta$ in the last expression).

The last one of the above relations for $\Psi(z)+\gamma$, i.e.

$$\begin{aligned}\Psi(z)+\gamma = \sum_{k=0}^{K-1}\alpha_k\int_{-\infty}^{\infty}(\xi-z)^+\,dF_\xi\,(\xi+u_k) \\ +\sum_{k=0}^{K-1}\beta_k\int_{-\infty}^{\infty}(\xi-z)^-\,dF_\xi\,(\xi-l_k)\,,\end{aligned} \tag{3.73}$$

indicates a formal similarity with an expected SRT function using a positive mixture of the distribution functions $F_\xi\,(\xi+u_k)$ and $F_\xi\,(\xi-l_k)$, $k=0,\cdots,K-1$,

$$H(\xi)=\sum_{k=0}^{K-1}\alpha_k F_\xi\,(\xi+u_k)+\sum_{k=0}^{K-1}\beta_k F_\xi\,(\xi-l_k)\,.$$

Due to Definition 3.4., $H(\cdot)$ is monotonically increasing, right-continuous, and satisfies

$$H(\xi)\ge 0\ \forall\xi,\quad \lim_{\xi\to-\infty}H(\xi)=0,\quad\text{and}\quad \lim_{\xi\to\infty}H(\xi)=\sum_{k=0}^{K-1}(\alpha_k+\beta_k)>0\,,$$

such that standardizing $H(\cdot)$, i.e. dividing by $W:=\sum_{k=0}^{K-1}(\alpha_k+\beta_k)$, yields a new distribution function as the mixture

$$G(\xi):=\frac{H(\xi)}{W}=\frac{\sum_{k=0}^{K-1}\alpha_k F_\xi\,(\xi+u_k)+\sum_{k=0}^{K-1}\beta_k F_\xi\,(\xi-l_k)}{W}\,. \tag{3.74}$$

Assuming now that $\Psi(\cdot)$ may be represented as an expected SRT function using the distribution function $G(\cdot)$ we get, with constants $A$, $B$ and $C$ to be determined later, using the trivial relations $\rho^+=\rho+\rho^-$ and $\rho^-=-\rho+\rho^+$, and writing $\int$ instead of $\int_{-\infty}^{\infty}$ for simplicity,

$$\begin{aligned}&\Psi(z)+C=A\int(\xi-z)^+\,dG(\xi)+B\int(\xi-z)^-\,dG(\xi)\\
&=\frac{A}{W}\left\{\sum_{k=0}^{K-1}\alpha_k\int(\xi-z)^+\,dF_\xi\,(\xi+u_k)+\sum_{k=0}^{K-1}\beta_k\int(\xi-z)^+\,dF_\xi(\xi-l_k)\right\}\\
&\quad+\frac{B}{W}\left\{\sum_{k=0}^{K-1}\alpha_k\int(\xi-z)^-\,dF_\xi\,(\xi+u_k)+\sum_{k=0}^{K-1}\beta_k\int(\xi-z)^-\,dF_\xi\,(\xi-l_k)\right\}\\
&=\frac{A}{W}\left\{\sum_{k=0}^{K-1}\alpha_k\int(\xi-z)^+\,dF_\xi\,(\xi+u_k)+\sum_{k=0}^{K-1}\beta_k\int(\xi-z)\,dF_\xi(\xi-l_k)\right.\\
&\qquad\qquad\left.+\sum_{k=0}^{K-1}\beta_k\int(\xi-z)^-\,dF_\xi(\xi-l_k)\right\}\end{aligned}$$

$$
\begin{aligned}
&+\frac{B}{W}\left\{\sum_{k=0}^{K-1}\alpha_k\int(z-\xi)\,dF_\xi\,(\xi+u_k)+\sum_{k=0}^{K-1}\beta_k\int(\xi-z)^-\,dF_\xi(\xi-l_k)\right.\\
&\qquad\qquad\left.+\sum_{k=0}^{K-1}\alpha_k\int(\xi-z)^+\,dF_\xi\,(\xi+u_k)\right\}\\
&=\frac{A}{W}\left\{\sum_{k=0}^{K-1}\alpha_k\int(\xi-z)^+\,dF_\xi\,(\xi+u_k)+\sum_{k=0}^{K-1}\beta_k(\mu+l_k-z)\right.\\
&\qquad\qquad\left.+\sum_{k=0}^{K-1}\beta_k\int(\xi-z)^-\,dF_\xi(\xi-l_k)\right\}\\
&+\frac{B}{W}\left\{\sum_{k=0}^{K-1}\alpha_k(z-\mu+u_k)+\sum_{k=0}^{K-1}\beta_k\int(\xi-z)^-\,dF_\xi(\xi-l_k)\right.\\
&\qquad\qquad\left.+\sum_{k=0}^{K-1}\alpha_k\int(\xi-z)^+\,dF_\xi\,(\xi+u_k)\right\}.
\end{aligned}
$$

Hence we have

$$
\begin{aligned}
&\Psi(z)+C=\\
&=\frac{A+B}{W}\sum_{k=0}^{K-1}\left\{\alpha_k\int(\xi-z)^+\,dF_\xi\,(\xi+u_k)+\beta_k\int(\xi-z)^-\,dF_\xi\,(\xi-l_k)\right\}\\
&\quad+\frac{A}{W}\sum_{k=0}^{K-1}\beta_k(\mu+l_k-z)+\frac{B}{W}\sum_{k=0}^{K-1}\alpha_k(z-\mu+u_k)\,.
\end{aligned}
$$

To get coincidence with equation (3.73) we ought to have, with $W_\alpha=\sum_{k=0}^{K-1}\alpha_k$ and $W_\beta=\sum_{k=0}^{K-1}\beta_k$,

$$
\begin{aligned}
&\frac{A+B}{W}=1\quad\text{and}\\
&\frac{A}{W}\left(W_\beta\,(\mu-z)+\sum_{k=0}^{K-1}\beta_k l_k\right)+\frac{B}{W}\left(W_\alpha\,(z-\mu)+\sum_{k=0}^{K-1}\alpha_k u_k\right)=C\,.
\end{aligned}
$$

To assure that the left-hand side of the last equation is constant (in $z$), we have the condition

$$A\cdot W_\beta-B\cdot W_\alpha=0\,,$$

which together with $A+B=W=W_\alpha+W_\beta$ implies that

$$A=W_\alpha\quad\text{and}\quad B=W_\beta\,,$$

such that

$$C = \frac{W_\alpha \sum_{k=0}^{K-1} \beta_k l_k + W_\beta \sum_{k=0}^{K-1} \alpha_k u_k}{W}.$$

Hence, for the multiple simple recourse problem (with one recourse constraint)

$$\left.\begin{array}{ll} \min\{c^{\mathrm{T}}x + \Psi(z)\} & \\ Ax & = b \\ t^{\mathrm{T}}x - z & = 0 \\ x & \geq 0 \end{array}\right\} \tag{3.75}$$

we have derived in an elementary way the following result, deduced first in Van der Vlerk [334], based on a statement proved in Klein Haneveld–Stougie–Van der Vlerk [189]:

**Theorem 3.8.** *The multiple simple recourse problem (3.75) with the expected MSRT function*

$$\Psi(z) = \tag{3.76}$$
$$= \sum_{k=0}^{K-1} \alpha_k \int (\xi - z)^+ \, dF_\xi\,(\xi + u_k) + \sum_{k=0}^{K-1} \beta_k \int (\xi - z)^- \, dF_\xi\,(\xi - l_k)$$

*is equivalent to the simple recourse problem with the expected SRT function*

$$\Psi(z) = \tag{3.77}$$
$$\left(\sum_{k=k}^{K-1} \alpha_k\right) \int (\xi - z)^+ \, dG(\xi) + \left(\sum_{k=k}^{K-1} \beta_k\right) \int (\xi - z)^- \, dG(\xi) - C$$

*using the distribution function*

$$G(\xi) = \frac{\sum_{k=0}^{K-1} \alpha_k F_\xi\,(\xi + u_k) + \sum_{k=0}^{K-1} \beta_k F_\xi\,(\xi - l_k)}{\sum_{k=0}^{K-1} (\alpha_k + \beta_k)} \tag{3.78}$$

*and the constant*

$$C = \frac{\left(\sum_{k=0}^{K-1} \alpha_k\right) \sum_{k=0}^{K-1} \beta_k l_k + \left(\sum_{k=0}^{K-1} \beta_k\right) \sum_{k=0}^{K-1} \alpha_k u_k}{\sum_{k=0}^{K-1} (\alpha_k + \beta_k)}. \tag{3.79}$$

As shown in Van der Vlerk [334], if $F_\xi$ represents a finite discrete distribution

$$\{(\xi_\nu, p_\nu);\ \nu = 1, \cdots, N\} \quad \text{with} \quad p_\nu > 0\ \forall \nu,\ \sum_{\nu=1}^{N} p_\nu = 1, \tag{3.80}$$

then $G$ corresponds to a finite discrete distribution with at most $N \cdot (2K-1)$ pairwise different realizations (with positive probabilities). This distribution, disregarding possible coincidences of some of its realizations, according to (3.78) and (3.80) is given by the following set of realizations and their corresponding probabilities

$$\left.\begin{array}{ll}
\xi_\nu, & \pi_{\nu 0} = \dfrac{(\alpha_0 + \beta_0) p_\nu}{\gamma};\ \nu = 1, \cdots, N;\ (\kappa = 0); \\
\xi_\nu - u_\kappa, & \pi^-_{\nu\kappa} = \dfrac{\alpha_\kappa p_\nu}{\gamma};\ \nu = 1, \cdots, N;\ \kappa = 1, \cdots, K-1; \\
\xi_\nu + l_\kappa, & \pi^+_{\nu\kappa} = \dfrac{\beta_\kappa p_\nu}{\gamma};\ \nu = 1, \cdots, N;\ \kappa = 1, \cdots, K-1; \\
 & \text{with } \gamma = \displaystyle\sum_{k=0}^{K-1} (\alpha_k + \beta_k).
\end{array}\right\} \tag{3.81}$$

#### 3.2.2.4 The generalized simple recourse problem (GSR)

GSR functions according to (3.53) on page 226 are defined as

$$\left.\begin{array}{rl}
Q^G(x,\xi) := \min & q^{+\mathrm{T}} y^+ + q^{-\mathrm{T}} y^- \\
 & I y^+ - I y^- = h(\xi) - T(\xi) x \\
 & y^+, \quad y^- \geq 0.
\end{array}\right\}$$

In contrast to (3.60) and (3.62) on page 227, neither GSR functions nor the corresponding EGSR functions $\mathcal{Q}^G(x) := \mathbb{E}_\xi\left[Q^G(x,\xi)\right]$ can be converted in a similar manner into separable functions in $(\chi_i, \xi_i)$ and in $(\chi_i)$, respectively.

Requiring Assumption 3.3., and hence in this case presuming that $q^+ + q^- \geq 0$, implies problem (3.53) to have the optimal value

$$\left.\begin{array}{l}
Q^G(x,\xi) = \displaystyle\sum_{i=1}^{m_2} Q_i^G(x, \xi^{(i)}) \quad \text{with} \\
Q_i^G(x,\xi^{(i)}) = q_i^+ [(\eta_i(x,\xi^{(i)})]^+ + q_i^- [(\eta_i(x,\xi^{(i)})]^-, \quad i = 1, \cdots, m_2,
\end{array}\right\} \tag{3.82}$$

where $\eta(x,\xi) = h(\xi) - T(\xi)x$, and $\xi^{(i)}$ is the subvector of $\xi$ with those components (of $\xi$) affecting $(h_i(\xi) - T_i(\xi)x)$, the $i$-th row of $(h(\xi) - T(\xi)x)$.

Observing that

$$\eta_i(x,\xi) = [\eta_i(x,\xi)]^+ - [\eta_i(x,\xi)]^- \Longrightarrow [\eta_i(x,\xi)]^+ = \eta_i(x,\xi) + [\eta_i(x,\xi)]^-$$

and denoting by $\mathbb{E}_{\xi^{(i)}}$ integration with respect to the marginal distribution $\mathbb{P}_{\xi^{(i)}}$ of $\xi^{(i)}$, it follows with $\overline{q} = q^+ + q^-$ and $(\overline{h}, \overline{T}) = \mathbb{E}_\xi[(h(\xi), T(\xi)]$, that

$$\mathscr{Q}^G(x) = \mathbb{E}_\xi[Q^G(x,\xi)] = \sum_{i=1}^{m_2} \mathbb{E}_{\xi^{(i)}}[Q_i^G(x,\xi^{(i)})] = \sum_{i=1}^{m_2} \mathscr{Q}_i^G(x), \tag{3.83}$$

where

$$\left.\begin{aligned}
\mathscr{Q}_i^G(x) &= \mathbb{E}_{\xi^{(i)}}[Q_i^G(x,\xi^{(i)})] \\
&= q_i^+ \mathbb{E}_{\xi^{(i)}}\left[[(\eta_i(x,\xi^{(i)})]^+\right] + q_i^- \mathbb{E}_{\xi^{(i)}}\left[[(\eta_i(x,\xi^{(i)})]^-\right] \\
&= q_i^+ \mathbb{E}_{\xi^{(i)}}[(\eta_i(x,\xi^{(i)})] \\
&\quad + q_i^+ \mathbb{E}_{\xi^{(i)}}\left[[(\eta_i(x,\xi^{(i)})]^-\right] + q_i^- \mathbb{E}_{\xi^{(i)}}\left[[(\eta_i(x,\xi^{(i)})]^-\right] \\
&= q_i^+(\overline{h}_i - \overline{T}_i x) + \overline{q}_i \mathbb{E}_{\xi^{(i)}}\left[[(\eta_i(x,\xi^{(i)})]^-\right].
\end{aligned}\right\}$$

As shown in Corollary 3.1. (p. 206) the expected recourse $\mathscr{Q}_i^G(x)$ is a convex function $\forall \overline{q} \geq 0$, and hence also $\mathbb{E}_{\xi^{(i)}}\left[[(\eta_i(x,\xi^{(i)})]^-\right]$ is convex in $x$.

By defining $S^{(i)}(x) := \{\xi^{(i)} \mid \eta_i(x,\xi^{(i)}) < 0\}$ for arbitrary $x \in \mathbb{R}^n$, it follows

$$\begin{aligned}
\mathbb{E}_{\xi^{(i)}}\left[[(\eta_i(x,\xi^{(i)})]^-\right] &= \int_{S^{(i)}(x)} -\eta_i(x,\xi^{(i)})\, \mathbb{P}_{\xi^{(i)}}(d\xi^{(i)}) \\
&= \int_{S^{(i)}(x)} (T_i(\xi)x - h_i(\xi))\, \mathbb{P}_{\xi^{(i)}}(d\xi^{(i)})
\end{aligned}$$

and, with fixed $\tilde{x}$, arbitrary $x$, and with $\overline{S}^{(i)}(x)$ the complement of $S^{(i)}(x)$, for

$$\left.\begin{aligned}
L_i(x;\tilde{x}) &:= \int_{S^{(i)}(\tilde{x})} -\eta_i(x,\xi^{(i)})\, \mathbb{P}_{\xi^{(i)}}(d\xi^{(i)}) \\
&= \int_{S^{(i)}(\tilde{x}) \cap S^{(i)}(x)} -\eta_i(x,\xi^{(i)})\, \mathbb{P}_{\xi^{(i)}}(d\xi^{(i)}) \\
&\quad + \underbrace{\int_{S^{(i)}(\tilde{x}) \cap \overline{S}^{(i)}(x)} -\eta_i(x,\xi^{(i)})\, \mathbb{P}_{\xi^{(i)}}(d\xi^{(i)})}_{\leq 0} \\
&\leq \int_{S^{(i)}(x)} -\eta_i(x,\xi^{(i)})\, \mathbb{P}_{\xi^{(i)}}(d\xi^{(i)}) = \mathbb{E}_{\xi^{(i)}}\left[[(\eta_i(x,\xi^{(i)})]^-\right].
\end{aligned}\right\} \tag{3.84}$$

Hence, the function

$$\left.\begin{aligned} L_i(x;\tilde{x}) &= \int_{S^{(i)}(\tilde{x})} -\eta_i(x,\xi^{(i)})\,\mathbb{P}_{\xi^{(i)}}(d\xi^{(i)}) \\ &= \int_{S^{(i)}(\tilde{x})} (T_i(\xi^{(i)})x - h_i(\xi^{(i)}))\,\mathbb{P}_{\xi^{(i)}}(d\xi^{(i)}) \\ &= \left\{\int_{S^{(i)}(\tilde{x})} T_i(\xi^{(i)})\,\mathbb{P}_{\xi^{(i)}}(d\xi^{(i)})\right\}x - \int_{S^{(i)}(\tilde{x})} h_i(\xi^{(i)})\,\mathbb{P}_{\xi^{(i)}}(d\xi^{(i)}) \end{aligned}\right\} \tag{3.85}$$

is a lower bound for $\mathbb{E}_{\xi^{(i)}}\left[[(\eta_i(x,\xi^{(i)})]^-\right]$, due to (3.85) linear affine in $x$, and sharp for $x = \tilde{x}$, since $L_i(\tilde{x};\tilde{x}) = \mathbb{E}_{\xi^{(i)}}\left[[(\eta_i(\tilde{x},\xi^{(i)})]^-\right]$ by (3.84). Due to $\overline{q} \geq 0$ follows that

$$\left.\begin{aligned} \mathscr{L}_i(x;\tilde{x}) &:= q_i^+(\overline{h}_i - \overline{T}_i x) + \overline{q}_i L_i(x;\tilde{x}) \\ &\leq q_i^+(\overline{h}_i - \overline{T}_i x) + \overline{q}_i \mathbb{E}_{\xi^{(i)}}\left[[(\eta_i(x,\xi^{(i)})]^-\right] = \mathscr{Q}_i^G(x)\,. \end{aligned}\right\} \tag{3.86}$$

Furthermore $\mathscr{L}_i(\tilde{x};\tilde{x}) = \mathscr{Q}_i^G(\tilde{x})$, since $L_i(\tilde{x};\tilde{x}) = \mathbb{E}_{\xi^{(i)}}\left[[(\eta_i(\tilde{x},\xi^{(i)})]^-\right]$, such that

$$\left.\begin{aligned} &\mathscr{Q}_i^G(x) - \mathscr{Q}_i^G(\tilde{x}) \geq \\ &\geq \mathscr{L}_i(x;\tilde{x}) - \mathscr{L}_i(\tilde{x};\tilde{x}) \\ &= q_i^+\overline{T}_i(\tilde{x}-x) + \overline{q}_i\left\{L_i(x;\tilde{x}) - L_i(\tilde{x};\tilde{x})\right\} \\ &= \left\{-q_i^+\overline{T}_i + \overline{q}_i \int_{S^{(i)}(\tilde{x})} T_i(\xi^{(i)})\,\mathbb{P}_{\xi^{(i)}}(d\xi^{(i)})\right\}(x-\tilde{x}), \end{aligned}\right\} \tag{3.87}$$

thus yielding a linear support function of $\mathscr{Q}_i^G(\cdot)$ at $\tilde{x}$ as

$$\mathscr{L}_i(x;\tilde{x}) = \mathscr{Q}_i^G(\tilde{x}) + g_i(\tilde{x})(x-\tilde{x}) = \mathscr{L}_i(\tilde{x};\tilde{x}) + g_i(\tilde{x})(x-\tilde{x}) \tag{3.88}$$

with $g_i(\tilde{x})$ a subgradient (row vector) of $\mathscr{Q}_i^G(\cdot)$ at $\tilde{x}$ given as

$$g_i(\tilde{x}) := -q_i^+\overline{T}_i + \overline{q}_i \int_{S^{(i)}(\tilde{x})} T_i(\xi^{(i)})\,\mathbb{P}_{\xi^{(i)}}(d\xi^{(i)}) \in \partial\mathscr{Q}_i^G(\tilde{x}).$$

Assume the first stage feasible set

$$\mathscr{B}_1 := \{x \mid Ax = b\,,\ x \geq 0\}$$

to be nonempty and compact. As mentioned above $\mathscr{Q}_i^G(\cdot)$, and thus also the related EGSR function $\mathscr{Q}^G(\cdot) = \sum_{i=1}^{m_2} \mathscr{Q}_i^G(\cdot)$, are convex functions and hence, according to Prop. 1.24. (p. 54), continuous. Therefore, $\hat{\Theta} := \min_{x\in\mathscr{B}_1} \mathscr{Q}^G(x)$ exists.

Then problem (3.55) (see p. 227) can be written as

$$\min\{c^{\mathrm{T}}x + \mathscr{Q}^G(x) \mid x \in \mathscr{B}_1\}$$

or equivalently as

$$\min\{c^{\mathrm{T}}x + \Theta \mid x \in \mathscr{B}_1\,,\ \mathscr{Q}^G(x) - \Theta \leq 0\}. \tag{3.89}$$

Obviously, $\Theta \geq \hat{\Theta}$, and a fortiori $\Theta \geq \hat{\Theta} - \gamma$ with some $\gamma > 0$, holds for all $(x, \Theta)$ being feasible in (3.89). On the other hand, to add the constraint $\Theta \leq \hat{\Theta} + \gamma$ has no impact on the solution of problem (3.89). Thus

$$\mathscr{B}^{+} := \{(x^{\mathrm{T}}, \Theta)^{\mathrm{T}} \mid x \in \mathscr{B}_1 \,,\ \hat{\Theta} - \gamma \leq \hat{\Theta} \leq \Theta \leq \hat{\Theta} + \gamma\} \subset \mathbb{R}^{n+1},$$

instead of $\mathscr{B}_1 \subset \mathbb{R}^n$, is nonempty and compact again. Hence with $z := (x^{\mathrm{T}}, \Theta)^{\mathrm{T}}$, the contraint function $F(z) := \mathscr{Q}^G(x) - \Theta$ (convex in $z$ as well), and with the objective $d^{\mathrm{T}} z := (c^{\mathrm{T}}, 1) z = c^{\mathrm{T}} x + \Theta$, the program (3.89) has the same set of solutions as

$$\min\{d^{\mathrm{T}} z \mid z \in \mathscr{B}^{+} \,,\ F(z) \leq 0\}. \tag{3.90}$$

Finally, since $\mathscr{Q}^G(\cdot) = \sum_{i=1}^{m_2} \mathscr{Q}_i^G(\cdot)$ due to (3.83), from (3.88) follows obviously that $g(\tilde{x}) = \sum_{i=1}^{m_2} g_i(\tilde{x}) \in \partial \mathscr{Q}^G(\tilde{x})$ at any arbitrary $\tilde{x}$, such that the function $F(\cdot)$ has at any $\tilde{z} = (\tilde{x}^{\mathrm{T}}, \tilde{\Theta})^{\mathrm{T}}$ a subgradient (row vector), given as

$$\left.\begin{aligned} f(\tilde{z}) &:= (g(\tilde{x}), -1) \\ &= \left(\sum_{i=1}^{m_2} g_i(\tilde{x}), -1\right) \\ &= \left(\sum_{i=1}^{m_2} \left\{-q_i^{+} \overline{T}_i + \overline{q}_i \int_{S^{(i)}(\tilde{x})} T_i(\xi^{(i)})\, \mathbb{P}_{\xi^{(i)}}(d\xi^{(i)})\right\}, -1\right) \in \partial F(\tilde{z}) \,. \end{aligned}\right\} \tag{3.91}$$

With these requisites the following procedure can be formulated:

**GSR-CUT: Approximating GSR solutions by successive cuts**

Find a solution $\hat{x}$ of the LP $\min\{c^{\mathrm{T}} x \mid x \in \mathscr{B}_1\}$.

With $\hat{u}^{(1)} := \hat{x}$ and $J := 1$ go to Step I..

I. Find a solution $(\hat{x}^{\mathrm{T}}, \hat{\Theta})^{\mathrm{T}}$ of the LP

$$\left.\begin{array}{c} \min\{c^{\mathrm{T}} x + \Theta\} \\ x \in \mathscr{B}_1 \\ \sum_{i=1}^{m_2} \mathscr{L}_i(x; \hat{u}^{(j)}) \leq \Theta, \ \ j = 1, \cdots, J, \end{array}\right\} \tag{3.92}$$

and denote this solution as $\hat{z}^{(J)} := (\hat{x}^{\mathrm{T}}, \hat{\Theta})^{(J)\mathrm{T}}$.

II. If

$$\Delta := F(\hat{z}^{(J)}) = \sum_{i=1}^{m_2} \mathscr{Q}_i^G(\hat{x}) - \hat{\Theta} = \sum_{i=1}^{m_2} \mathscr{L}_i(\hat{x}; \hat{x}) - \hat{\Theta} \leq 0,$$

stop (in practice: if $\Delta \le \varepsilon$ with a prescribed tolerance $\varepsilon$, stop);

else, with $J := J+1$ and $\hat{u}^{(J)} := \hat{x}$, return to Step I..

□

**Remark 3.7.** *The following observations on the above procedure GSR-CUT may be useful:*

1) *Due to (3.87), the $\mathscr{L}_i(x;\hat{u}^{(j)})$ are linear support functions of $\mathscr{Q}_i^G(x)$ at $\hat{u}^{(j)}$, and their gradients $\nabla_x \mathscr{L}_i(x;\hat{u}^{(j)})$ coincide due to (3.88) with the subgradients of $\mathscr{Q}_i^G(\hat{u}^{(j)})$ given as $g_i(\hat{u}^{(j)}) \in \partial \mathscr{Q}_i^G(\hat{u}^{(j)})$.*
2) *It follows immediately that, due to (3.91),*

$$\begin{aligned}
&\sum_{i=1}^{m_2} \mathscr{L}_i(x;\hat{u}^{(j)}) - \Theta = \\
&\quad = \{\mathscr{Q}^G(\hat{u}^{(j)}) - \hat{\Theta}^{(j)}\} + g(\hat{u}^{(j)})(x-\hat{u}^{(j)}) + (-1)(\Theta - \hat{\Theta}^{(j)}) \\
&\quad = F(\hat{z}^{(j)}) + f(\hat{z}^{(j)})(z-\hat{z}^{(j)})
\end{aligned}$$

*is a linear support function of $F(z) = \mathscr{Q}^G(x) - \Theta$ at $\hat{z}^{(j)} = (\hat{x}^{(j)\mathrm{T}}, \hat{\Theta}^{(j)})^{\mathrm{T}}$, and since by (3.92) obviously holds $\hat{\Theta}^{(J)} = \max_{1\le j\le J} \mathscr{L}(\hat{x}^{(J)};\hat{u}^{(j)})$ for all J, from the compactness of $\mathscr{B}_1$, the continuity of $\mathscr{Q}^G(\cdot)$ as well as the uniform boundedness of the subgradients $g(\cdot) \in \partial \mathscr{Q}^G(\cdot)$ (see the proof of Prop. 1.29., p. 61), follows the existence of an appropriate compact (polyhedral) set $\mathscr{B}^+ \subset \mathbb{R}^{n+1}$ such that $\hat{z}^{(j)} \in \mathscr{B}^+$ for all solutions of (3.92) generated within the above iteration. In other words, in the above iteration we deal simultaneously with problem (3.89) as well as with problem (3.90).*
3) *The standard convergence statements—convergence of $\varphi_J = c^{\mathrm{T}}\hat{x}^{(J)} + \hat{\Theta}^{(J)}$, the optima of (3.92), to the optimal value of (3.89), and any accumulation point of iterates $\{\hat{z}^{(J)}\}$, generated by (3.92), being a solution of (3.89)—follow immediately from Prop. 1.29. (p. 61), observing that procedure GSR-CUT is just the application of Kelley's cutting plane method (on page 61) to problem (3.90).*
4) *In (3.92) the evaluation of $\mathscr{L}_i(x;\hat{u}^{(j)}) = q_i^+\overline{h}_i - (q_i^+\overline{T}_i)x + \overline{q}_i L_i(x;\hat{u}^{(j)})$ requires according to (3.6) (p. 196) for $\overline{h}_i$ and $\overline{T}_i$ the expectations $\mathbb{E}_{\xi^{(i)}}[\xi^{(i)}]$ and due to (3.85) in particular the computation of the integrals*

$$\left\{ \int_{S^{(i)}(\hat{u}^{(j)})} T_i(\xi^{(i)})\, \mathbb{P}_{\xi^{(i)}}(d\xi^{(i)}) \right\} \quad \textit{and} \quad \left\{ -\int_{S^{(i)}(\hat{u}^{(j)})} h_i(\xi^{(i)})\, \mathbb{P}_{\xi^{(i)}}(d\xi^{(i)}) \right\}.$$

*Since in general for multivariate distributions of continuous type (described by densities) there is no algebraic formula for these integrals, they need to be approximated by some simulation approach, e.g. an appropriate variant of the Monte Carlo method.*
*For a finite discrete distribution $\mathbb{P}_{\xi^{(i)}}(\xi^{(i)} = \xi^{(i)\nu}) = p_\nu^{(i)},\ \nu = 1,\cdots,N^{(i)}$, the sets $S^{(i)}(\hat{u}^{(j)}) := \{\xi^{(i)} \mid \eta_i(\hat{u}^{(j)},\xi^{(i)}) < 0\}$ are replaced by the index sets*

$K^{(i)}(\hat{u}^{(j)}) := \{\nu \mid \eta_i(\hat{u}^{(j)}, \xi^{(i)\nu}) < 0\}$, *thus yielding* $\mathbb{E}_{\xi^{(i)}}[\xi^{(i)}] = \sum_{\nu=1}^{N^{(i)}} p_\nu^{(i)} \xi^{(i)\nu}$ *and*

$$\mathscr{L}_i(x;\hat{u}^{(j)}) = q_i^+ \overline{h}_i - (q_i^+ \overline{T}_i)x + \overline{q}_i \sum_{\nu \in K^{(i)}(\hat{u}^{(j)})} p_\nu^{(i)} \{h_i(\xi^{(i)\nu}) - T_i(\xi^{(i)\nu})x\}.$$

□

## Exercises

**3.7.** Consider the following two simple recourse problems:

$$\left.\begin{array}{llll}
\min\{c^{\mathrm{T}}x + \mathbb{E}[q^{\mathrm{T}}y(\zeta)]\} & \text{with} & c = (3,1,2,4)^{\mathrm{T}}, & q = (2,1,1,3,2,1,2,1)^{\mathrm{T}} \\
Ax \qquad \le b & \text{with} & A = \begin{pmatrix} 2 & 1 & 3 & 5 \\ 3 & 4 & 3 & 2 \end{pmatrix}, & b = \begin{pmatrix} 32 \\ 35 \end{pmatrix} \\
Tx + Wy(\zeta) = h(\zeta) \ \text{a.s.} & \text{with} & T = \begin{pmatrix} 2 & 0 & 3 & 2 \\ 3 & 5 & 0 & 2 \\ 0 & 2 & 4 & 0 \\ 2 & 1 & 0 & 3 \end{pmatrix}, & h(\zeta) = \begin{pmatrix} 25+\xi_1 \\ 15+\xi_2 \\ 17+\xi_3 \\ 23+\xi_4 \end{pmatrix} \\
\quad x, y(\zeta) \ge 0 \quad \text{a.s.} & & & \\
\multicolumn{4}{l}{\text{where } \zeta \text{ with independent components has either a uniform or a normal distribution as}} \\
\multicolumn{4}{l}{\mathscr{U}\{[-5,5]\times[-7,7]\times[-3,3]\times[-8,8]\} \ \text{or}} \\
\multicolumn{4}{l}{\mathscr{N}\{(0;2),(0;1.5),(0;2.2),(0;1.7)\} \ \text{with truncation probabilities of 0.999, each.}}
\end{array}\right\} \quad (3.93)$$

Solve both problems using SLP-IOR, applying SRAPPROX as well as DAPPROX.

(a) For each of the two problems compare, as indicators for the efficiency of the two solvers, the number of iterations as well as the number of splits (or subintervals, respectively) used by SLP-IOR to get the solutions with the pre-set accuracy.

(b) How do you explain the difference with respect to the above indicators, in particular the remarkable discrepancy of the numbers of splits/subintervals?

**3.8.** For the SRT function $\varphi(z,\xi) := \alpha[\xi - z]^+ + \beta[\xi - z]^-$, $\alpha + \beta \ge 0$, and the corresponding ESRT function $\Phi(z) = \alpha(\bar{\xi} - z) - (\alpha+\beta)\int_{-\infty}^{z}(\xi - z)\mathbb{P}_\xi(d\xi)$ assume the distribution $\mathbb{P}_\xi$ to be bounded to the interval $[a,b]$.

(a) Show that then the integral $\Psi(z) := \int_{-\infty}^{z}(z-\xi)\mathbb{P}_\xi(d\xi)$ may be computed with $\Theta(z_1,z_2,z_3) := z_3 + \int_{\xi \le a+z_2}(z_2 + a - \xi)\mathbb{P}_\xi(d\xi)$ as

$$\left.\begin{array}{l}
\Psi(z) := \min \Theta(z_1,z_2,z_3) \text{ subject to:} \\
-z_1 + z_2 + z_3 = z - \bar{\xi},\ z_1 \ge \bar{\xi} - a,\ z_2 \le b - a,\ (z_1,z_2,z_3) \ge 0
\end{array}\right\}. \quad (3.94)$$

(b) How does $\Psi(z)$ and hence the ESRT function $\Phi(z)$ look like, if $\mathbb{P}_\xi$ is given as $\mathscr{U}\{[a,b]\}$?

**3.9.** Assume for (3.93) of Exercise 3.7 the uniform distributions mentioned for the right-hand sides $h_i(\zeta_i)$ and formulate problem (3.93) according to the result of the previous exercise as a quadratic program. If you have access to any convex programming software package, than solve the quadratic program and compare the solution with that one you have got with SLP-IOR applying SRAPPROX (and/or DAPPROX).

### *3.2.3 CVaR and recourse problems*

Assume the result of some process to be a loss, modelled as a random variable $\vartheta \in \mathscr{L}_1^1$ with a distribution function $F_\vartheta(z)$. As mentioned in Section 2.1, an example from finance could be a portfolio optimization problem with $t^{\mathrm{T}}(\xi)x$ as random return of a portfolio $x \in \mathbb{R}^n$ (usually represented as the mixture of different assets) compared to $h(\xi)$, the random return of some benchmark portfolio. In this case $\vartheta := (t^{\mathrm{T}}(\xi)x - h(\xi))$ is considered as loss if $\vartheta^- > 0$. With the $\alpha$–VaR (value at risk) $v_\alpha$, defined in Section 2.3 (p. 137) as $v_\alpha := v_\alpha(\vartheta) := \min\{z \mid F_\vartheta(z) \geq \alpha\},\ \alpha \in (0,1)$, the $\alpha$–CVaR (conditional value at risk) $v_\alpha^c := v_\alpha^c(\vartheta)$was introduced in Section 2.4.3 (p. 152) as

$$v_\alpha^c(\vartheta) := v_\alpha + \frac{1}{1-\alpha}\mathbb{E}_\vartheta[(\vartheta - v_\alpha)^+] = \min_z \left\{ z + \frac{1}{1-\alpha}\mathbb{E}_\vartheta[(\vartheta - z)^+] \right\}. \quad (3.95)$$

It is well known, that—in spite of the naming—for the $\alpha$–CVaR holds the inequality $v_\alpha^c(\vartheta) \geq \mathbb{E}_\vartheta[\vartheta \mid \vartheta \geq v_\alpha]$, where equality can only be ensured for continuous distribution functions $F_\vartheta(\cdot)$. Nevertheless, $v_\alpha^c(\vartheta)$ is widely used in finance applications as risk measure. Whereas the VaR $v_\alpha(\vartheta)$ is by definition the (smallest) threshold for a realization $\hat{\vartheta}$ not being exceeded with a probability of at least $\alpha$, for continuous distributions the $\alpha$–CVaR $v_\alpha^c(\vartheta)$ is then the conditional expectation of $\vartheta$ given that $\vartheta \geq \hat{\vartheta}$. Moreover, due to Prop. 2.48. the $\alpha$–CVaR satisfies the axioms for coherent risk measures presented in Artzner, Delbaen et al. [7], which is in general not true for the $\alpha$–VaR. A more detailed discussion of the concept of CVaR can be found in Rockafellar–Uryasev [283].

Due to (3.95), computing the $\alpha$–CVaR $v_\alpha^c(\vartheta)$ can be considered as solving a single-stage stochastic program. However, $v_\alpha^c(\vartheta)$ can also be considered as the optimal value of a particular two-stage stochastic program with simple recourse.

**Proposition 3.3.** *The $\alpha$–CVaR as defined in (3.95) is the optimal value of the SSR problem*

$$v_\alpha^c = \min_z (z + \mathbb{E}_\vartheta[Q(z;\vartheta)], \quad (3.96)$$

*where*

$$Q(z;\vartheta) = \min_{\eta}\left\{ \frac{1}{1-\alpha}\eta \,\middle|\, z+\eta \geq \vartheta,\ \eta \geq 0 \right\}.$$

*Proof:* Obviously holds

$$(\vartheta - z)^+ = \min_{\eta}\{\eta \mid \eta \geq \vartheta - z,\ \eta \geq 0\} = (1-\alpha)Q(z;\vartheta),$$

thus yielding the proposition and allowing for the interpretation, that after the first-stage decision on $z$ a realization of $\vartheta$ has to be observed before taking the second-stage decision on $\eta$. □

Assuming now that, instead of $\vartheta : \Omega \longrightarrow \mathbb{R}$, a random vector $\xi : \Omega \longrightarrow \Xi \subset \mathbb{R}^r$ is given with $\Xi = \operatorname{supp}\xi$, and $f(x,\xi) : X \times \Xi \longrightarrow \mathbb{R}$ is defined as decision-dependent (loss) function, where

- $X \subset \mathbb{R}^n$ is a closed convex set of feasible decisions,
- $f(\cdot,\xi)$ is continuous in $x \ \forall \xi \in \Xi$,
- $f(x,\cdot)$ is $\xi$–measurable $\forall x \in X$, and
- $\mathbb{E}_\xi[|f(x,\xi)|] < \infty \ \forall x \in X$.

With the distribution function $\Phi(x,z) := \mathbb{P}(\{\xi \mid f(x,\xi) \leq z\})$ the $\alpha$–VaR of $f(x,\xi)$ is $\nu_\alpha(x) = \min\{z \mid \Phi(x,z) \geq \alpha\}$, yielding in analogy to (3.95) the $\alpha$–CVaR of $f(x,\xi)$ as

$$\nu_\alpha^c(x) := \min_z \left\{ z + \frac{1}{1-\alpha}\mathbb{E}_\xi[(f(x,\xi)-z)^+] \right\}.$$

If in addition to the above assumptions $f(\cdot,\xi)$ is convex in $x \ \forall \xi \in \Xi$, then it follows that $\nu_\alpha^c(x)$ is convex in $x$ as well. In this case Prop. 3.3. is modified to

**Proposition 3.4.** *For $f(\cdot,\xi)$ convex $\forall \xi \in \Xi$ the $\alpha$–CVaR denoted as $\nu_\alpha^c(x)$ is a convex function in x, computable as the optimal value of the convex CFR program*

$$\left.\begin{array}{rl} \nu_\alpha^c(x) := & \min\limits_{z\in\mathbb{R}}\{z + \mathbb{E}_\xi[Q(x,z;\xi)]\} \\ \textit{with}\quad Q(x,z;\xi) := & \min\limits_{\eta\in\mathbb{R}}\left\{ \frac{1}{1-\alpha}\eta \,\middle|\, z+\eta \geq f(x,\xi),\ \eta \geq 0 \right\}, \end{array}\right\} \tag{3.97}$$

*or equivalently, the optimum of*

$$\left.\begin{array}{rl} \nu_\alpha^c(x) := & \min\limits_{z\in\mathbb{R},\eta(x,z;\cdot)\in\mathscr{L}_1^1} \left\{ z + \mathbb{E}_\xi\left[\frac{1}{1-\alpha}\eta(x,z;\xi)\right] \right\} \\ & z + \eta(x,z;\xi) \geq f(x,\xi) \ \textit{a.s.} \\ & \eta(x,z;\xi) \geq 0 \qquad \textit{a.s.} \end{array}\right\} \tag{3.98}$$

*Proof:* Introducing $x$ as a parameter and replacing $\vartheta$ by $f(x;\xi)$, (3.97) follows immediately from (3.96). The integrability of $f(x;\xi)$ with respect ot $\xi$ implies the well known fact, that for each $(x,z)$ there exists an $\eta(x,z;\cdot) \in \mathscr{L}_1^1$ such that $\eta(x,z;\xi) =$

$(1-\alpha)Q(x,z;\xi)$ a.s. Finally, the convexity of $Q(x,z;\xi)$ in $(x,z)$ $\forall \xi \in \Xi$ follows trivially from (3.97), implying the convexity of $\mathbb{E}_{\xi}[Q(x,z;\xi)]$ in $x$ for any fixed $z$ and thus the convexity of $v_{\alpha}^{c}(x)$ in $x$. □

To be more specific, assume that $X := \{x \mid Ax = b,\ x \geq 0\} \neq \emptyset$ is compact, and that the loss function is defined as $f(x,\xi) := \lambda(h(\xi) - t^{\mathrm{T}}(\xi)x)$ with some coefficient $\lambda > 0$. As an interpretation, think of a linear production function, transforming a vector $x$ of input factors with a random vector $t(\xi)$ of productivities into a random output $t^{\mathrm{T}}(\xi)x$; on the other hand let $h(\xi)$ be a random demand to be covered by that output, such that, given that $h(\xi) - t^{\mathrm{T}}(\xi)x > 0$, the above loss function $f(x,\xi)$ is just proportional to this excess demand.

Different types of models may be set up in this situation, as for instance:

1) In addition to the linear constraints of an LP a further constraint, restricting the $\alpha$–CVaR $v_{\alpha}^{c}(x) := v_{\alpha}^{c}(f(x,\xi)) = \min\limits_{z}\left\{z + \frac{1}{1-\alpha}\mathbb{E}_{\xi}[(f(x,\xi)-z)^{+}]\right\}$ of the above loss function, may be inserted yielding the model

$$\begin{array}{rc} & \min c^{\mathrm{T}}x \\ \text{s.t.} & Ax = b \\ & v_{\alpha}^{c}(x) \leq \gamma \\ & x \geq 0\,, \end{array}$$

which according to (2.152), (2.153) on pages157/157 coincides with the convex NLP

$$\begin{array}{rr} & \min c^{\mathrm{T}}x \\ \text{s.t.} & Ax = b \\ & z + \frac{1}{1-\alpha}\mathbb{E}_{\xi}[(\lambda(h(\xi) - t^{\mathrm{T}}(\xi)x) - z)^{+}] \leq \gamma \\ & x \geq 0, \end{array}$$

a single stage problem.

2) Extending instead the linear term of an LP's objective by adding the $\alpha$–CVaR $v_{\alpha}^{c}(x) := v_{\alpha}^{c}(f(x,\xi))$ of the loss $f(x;\xi)$ yields the NLP

$$\begin{array}{rl} & \min\{c^{\mathrm{T}}x + v_{\alpha}^{c}(x)\} \\ \text{s.t.} & Ax = b \\ & x \geq 0\,, \end{array}$$

which due to Prop. 3.4. can be restated as

$$\left.\begin{aligned} &\min\{c^{\mathrm{T}}x+z+\mathbb{E}_{\xi}[Q(x,z;\xi)]\} \\ &\quad Ax=b \\ &\quad x\in\mathbb{R}^n_+,\ z\in\mathbb{R} \\ &\text{where} \\ &Q(x,z;\xi):=\min_{\eta\in\mathbb{R}}\left\{\frac{1}{1-\alpha}\eta\,\middle|\, z+\eta\geq f(x,\xi),\ \eta\geq 0\right\} \\ &\qquad\quad =\min_{\eta\in\mathbb{R}}\left\{\frac{1}{1-\alpha}\eta\,\middle|\,\lambda t^{\mathrm{T}}(\xi)x+z+\eta\geq\lambda h(\xi),\ \eta\geq 0\right\}, \end{aligned}\right\} \tag{3.99}$$

a particular two-stage generalized simple recourse SLP with the first stage variables $x\in\mathbb{R}^n$ and $z\in\mathbb{R}$ and the recourse variable $\eta$ (which, as mentioned above, can be chosen for each $(x,z)$ as a function $\eta(x,z;\cdot)\in\mathscr{L}_1^1$); in (2.150) on page 156 this model was derived for the special case of a finite discrete probability distribution.
Solution methods of the GSR-CUT type were considered for this problem by Klein Haneveld and van der Vlerk [191] and by Künzi–Bay and Mayer [198], assuming $\xi$ to have a finite discrete distribution; since in this case (3.99) is a special LP with decomposition structure, in accordance with Remark 1.2. (p. 48) the main concern of the authors was to find appropriate cut generation strategies for the corresponding decomposition algorithm to be as efficient as possible.

Due to the above discussion on general GSR-Cut procedures, for continuous distributions the cutting plane methods described on page 247 can be designed to solve (3.99) as well.

3) For the two-stage model (3.99) it is assumed that the *first-stage* decision on $x$ implies the deterministic *first-stage* outcome $c^{\mathrm{T}}x$, and that the loss $f(x,\xi)$, given the first-stage decision $x$, is the random *second-stage* outcome determined by the realization of $\xi$ (unknown when deciding on $x$). To take into account the risk (due to the random loss $f(x,\xi)$), this model aims at determining a minimizer $\hat{x}$ for the overall objective given as the sum of the first-stage outcome $c^{\mathrm{T}}x$ with the $\alpha$–CVaR of the second-stage outcome $f(x,\xi)$.
Another two-stage model is based on the following view: With a convex polyhedral set $X\subset\mathbb{R}^n_+$ of feasible first-stage decisions $x$, causing $c^{\mathrm{T}}x$ as the deterministic part of of the first-stage outcome, and with a very general recourse function

$$Q(x;\xi):=\min_y\{q^{\mathrm{T}}(\xi)y \mid T(\xi)x+W(\xi)y=h(\xi),\ y\geq 0\}, \tag{3.100}$$

as the random part of the first-stage outcome, the *overall first-stage objective* is defined as

$$f(x;\xi):=c^{\mathrm{T}}x+Q(x;\xi).$$

Now the decision maker wants to find any $\hat{x}\in X$ which minimizes some mixture of the mean of this outcome and some risk measure of it, e.g. the $\alpha$–CVaR of $f$. Thus, observing that due to Prop. 2.48. (page 177) the $\alpha$–CVaR is translation invariant, with some $\lambda > 0$ the problem to solve would be

$$\begin{aligned}&\min_{x\in X}\{\mathbb{E}_\xi[f(x;\xi)]+\lambda v_\alpha^c(f(x;\xi))\} = \\ &= \min_{x\in X}\{(1+\lambda)c^{\mathrm{T}}x+\mathbb{E}_\xi[Q(x;\xi)]+\lambda v_\alpha^c(Q(x;\xi))\}.\end{aligned}$$

This can be rewritten as the two-stage SLP

$$\left.\begin{array}{ll}\min\limits_{x,\eta,y,\theta}\left[(1+\lambda)c^{\mathrm{T}}x+\mathbb{E}_\xi[q^{\mathrm{T}}(\xi)y(\xi)]+\lambda\left(\eta+\dfrac{1}{1-\alpha}\mathbb{E}_\xi[\theta(\xi)]\right)\right] & \\ x\in X,\ \eta\in\mathbb{R} & \\ y(\cdot)\in\mathscr{L}_{n_2}^2,\ \theta(\cdot)\in\mathscr{L}_1^1 & \\ W(\xi)y(\xi)=h(\xi)-T(\xi)x & \text{a.s.} \\ \theta(\xi)\ge q^{\mathrm{T}}(\xi)y(\xi)-\eta & \text{a.s.} \\ y(\xi)\ge 0 & \text{a.s.} \\ \theta(\xi)\ge 0 & \text{a.s.}\end{array}\right\} \quad (3.101)$$

with first-stage variables $(x,\eta)$ and second-stage decisions $(y(\xi),\theta(\xi))$, or more precisely $y(\xi)$, since due to the objective of (3.101) automatically $\theta(\xi)=(q^{\mathrm{T}}(\xi)y(\xi)-\eta)^+$ a.s. will result. At present, it seems unlikely to find an efficient solver for the above problem in this generality for continuous distributions $\mathbb{P}_\xi$. Obviously one might think of approximating solutions via constructing sequences of discrete distributions $\mathbb{P}_\xi^\nu$, weakly converging to $\mathbb{P}_\xi$ and thus taking advantage of known results on the use of epi-convergence in optimization, as presented for instance in Pennanen [251, 252], Robinson and Wets [279], Wets [343], and Kall [156]. However, since in this generality the recourse function $Q(x;\xi)$ is not convex in $\xi$, neither Jensen's inquality nor the Edmundson–Madansky inequality apply. Hence, there seems to be no efficient tool to check the approximation error (as e.g. in DAPPROX) and thus to verify a prescribed accuracy during such an iterative procedure. Obviously this would change substantially, if for the recourse (3.100) holds $q(\xi)\equiv q\in\mathbb{R}^{n_2}$ and $W(\xi)\equiv W$, a constant $(m_2\times n_2)$–matrix, thus allowing for an approximation via successive discretization.

In the general case the situation becomes much better manageable for finite discrete distributions of $\xi$ given by $\mathbb{P}_\xi(\xi=\xi_i)=p_i,\ i=1,\cdots,N$. Then (3.101) reads as

$$\left.\begin{array}{ll}\min\limits_{x,\eta,y,\theta}\left[(1+\lambda)c^{\mathrm{T}}x+\sum\limits_{i=1}^N p_i\cdot {q_i}^{\mathrm{T}}y_i+\lambda\left(\eta+\dfrac{1}{1-\alpha}\sum\limits_{i=1}^N p_i\cdot\theta_i\right)\right] & \\ x\in X,\ \eta\in\mathbb{R} & \\ W_iy_i=h_i-T_ix & i=1,\cdots,N \\ \theta_i\ge {q_i}^{\mathrm{T}}y_i-\eta & i=1,\cdots,N \\ y_i\ge 0 & i=1,\cdots,N \\ \theta_i\ge 0 & i=1,\cdots,N.\end{array}\right\} \quad (3.102)$$

This model, a linear program with decomposition structure, was recently analyzed by Noyan [246], providing two variants of appropriate cuts within decomposition procedures for solving this problem efficiently.

### 3.2.4 Some characteristic values for two-stage SLP's

Among various paradigms of modeling two-stage stochastic linear programs we have discussed so far the general (two-stage) stochastic program with recourse with the optimal value $RS$ given due to (3.8), (3.9) as

$$\left.\begin{array}{rl} RS := \min\limits_{x} \left\{c^{\mathrm{T}}x + \mathbb{E}_{\xi}\left[Q(x;T(\xi),h(\xi),W(\xi),q(\xi))\right]\right\} & \\ \text{s.t. } Ax = b & \\ x \geq 0\,, & \end{array}\right\} \tag{3.103}$$

where

$$\begin{array}{rll} Q(x;T(\xi),h(\xi),W(\xi),q(\xi)) := \inf q^{\mathrm{T}}(\xi)y(\xi) & & \\ \text{s.t. } W(\xi)y(\xi) = h(\xi) - T(\xi)x & \text{a.s.} \\ y(\xi) \geq 0 & \text{a.s.} \\ y(\cdot) \in Y & \end{array}$$

with $Y = \mathscr{L}^2_{n_2}$. As in (3.6), we assume that the random parameters in these problems are defined as linear affine mappings on $\Xi = \mathbb{R}^r$ by

$$\begin{array}{ll} T(\xi) = T + \sum_{j=1}^{r} T^j\xi_j\,; & T,\, T^j \in \mathbb{R}^{m_2\times n_1} \text{ deterministic,} \\ W(\xi) = W + \sum_{j=1}^{r} W^j\xi_j\,; & W,\, W^j \in \mathbb{R}^{m_2\times n_2} \text{ deterministic,} \\ h(\xi) = h + \sum_{j=1}^{r} h^j\xi_j\,; & h,\, h^j \in \mathbb{R}^{m_2} \text{ deterministic,} \\ q(\xi) = q + \sum_{j=1}^{r} q^j\xi_j\,; & q,\, q^j \in \mathbb{R}^{n_2} \text{ deterministic.} \end{array}$$

**Remark 3.8.** *Whereas by the modeling paradigm of problem (3.103), the second stage decision on $y(\xi)$ is to be taken after observing the realization of $\xi$, and knowing the first stage decision on x, which was taken before having knowledge of the realization of $\xi$—one possible interpretation being (see Fig. 3.1, p. 191) that, in time, the decision on the first stage variables x is taken before the observation of a realization of $\xi$, and the second stage variables $y(\xi)$ are determined afterwards—other paradigms could be either to replace the random vector $\xi$ in advance by its expectation $\overline{\xi}$, thus yielding the expected value problem (3.104), or else to delay the first stage decision until a realization of $\xi$ is known, such that now the second stage* *decision $y(\cdot)$ as well as the first stage decision $x(\cdot)$ depend on $\xi$, which leads to the wait-and-see model (3.105).* □

As just mentioned, replacing the random vector $\xi$ by its expectation $\bar{\xi} = \mathbb{E}_{\xi}[\xi]$, yields instead of *RS* the optimal value *EV* of the *expected value problem*,

$$\left.\begin{array}{rl} EV := \min\limits_{x,y}\{c^{\mathrm{T}}x + q^{\mathrm{T}}(\bar{\xi})y\} & \\ \text{s.}t. \quad Ax \qquad\qquad = b & \\ T(\bar{\xi})x + W(\bar{\xi})y = h(\bar{\xi}) & \\ x, \qquad\quad y \geq 0\,. & \end{array}\right\} \tag{3.104}$$

Except for the first moment $\bar{\xi}$, this model does not take at all into account the distribution of $\xi$. Hence the solution will always be the same, no matter of the distribution being discrete or continuous, skew or symmetric, flat or concentrated, as long as the expectation remains the same. In other words, the randomness of $\xi$ does not play an essential role in this model.

In contrast to the recourse model (3.103), in the *wait-and-see* model both, the decisions on the first stage variables $x$ and the second stage variables y, are taken simultaneously only when the outcome of $\xi$ is known, with the optimal values of the family of LP's

$$\left.\begin{array}{rl} \forall \xi \in \Xi : \gamma(\xi) := & \min\limits_{x,y}\{c^{\mathrm{T}}x + q^{\mathrm{T}}(\xi)y\} \\ & \text{s.}t. \quad Ax \qquad\qquad = b \\ & T(\xi)x + W(\xi)y = h(\xi) \\ & x, \qquad\quad y \geq 0 \end{array}\right\} \tag{3.105}$$

the so-called *wait-and-see* value $WS$ is the expected value

$$WS := \mathbb{E}_{\xi}[\gamma(\xi)]\,. \tag{3.106}$$

Finally, with the first stage solution fixed as any optimal first stage solution $\hat{x}$ of the *EV* problem (3.104), we may ask for the objective's value of (3.103), the *expected result of the EV solution*

$$\begin{aligned} EEV &:= \\ &= c^{\mathrm{T}}\hat{x} + \mathbb{E}_{\xi}[\min_y\{q^{\mathrm{T}}(\xi)y \mid W(\xi)y = h(\xi) - T(\xi)\hat{x}, y \geq 0\}]\,. \end{aligned} \tag{3.107}$$

Observe that, in contrast to the values *RS*, *EV*, and *WS*, the value *EEV* may not be uniquely determined by (3.107): If the expected value problem (3.104) happens to have two different solutions $\hat{x} \neq \tilde{x}$, this may lead to $EEV(\hat{x}) \neq EEV(\tilde{x})$.

For the above values assigned in various ways to the two-stage stochastic programming situations mentioned, several relations are known which, essentially, can be traced back to Madansky [211].

**Proposition 3.5.** *For an arbitrary recourse problem (3.103) and the associated problems (3.106) and (3.107) the following inequalities hold:*

$$WS \leq RS \leq EEV\,. \tag{3.108}$$

*Furthermore, with the recourse function $Q(x;T(\xi),h(\xi))$, allowing only for the matrix $T(\cdot)$ and the right–hand–side $h(\cdot)$ to contain random data, it follows that*

$$EV \leq RS \leq EEV\,. \tag{3.109}$$

*Proof:* Let $x^\star$ be an optimal first stage solution of (3.103). Then obviously the inequality

$$\gamma(\xi) \leq c^{\mathrm{T}}x^\star + Q(x^\star;T(\xi),h(\xi),W(\xi),q(\xi)) \;\; \forall \xi \in \Xi$$

holds, and therefore

$$WS = \mathbb{E}_\xi\,[\gamma(\xi)] \leq \{c^{\mathrm{T}}x^\star + \mathbb{E}_\xi\,[Q(x^\star;T(\xi),h(\xi),W(\xi),q(\xi))]\} = RS\,.$$

The second inequality in (3.108) is obvious.

To show the second part, for any fixed $\tilde{x}$ the recourse function

$$Q(\tilde{x};T(\xi),h(\xi)) = \min\{q^{\mathrm{T}}y \mid Wy = h(\xi) - T(\xi)\tilde{x},\; y \geq 0\}$$

is convex in $\xi$. In particular, for the optimal first stage solution $x^\star$ of (3.103) follows with Jensen's inequality and the definition (3.104) of $EV$, that

$$\begin{aligned} RS &= c^{\mathrm{T}}x^\star + \mathbb{E}_\xi\,[Q(x^\star;T(\xi),h(\xi))] \\ &\geq c^{\mathrm{T}}x^\star + Q(x^\star;T(\bar{\xi}),h(\bar{\xi})) \\ &\geq EV \end{aligned}$$

which implies (3.109). □

**Proposition 3.6.** *Given the recourse function $Q(x;h(\xi))$ (i.e. only the right–hand–side $h(\cdot)$ is random) it follows that*

$$EV \leq WS\,.$$

*Proof:* For the wait-and-see situation we have

$$\gamma(\xi) = \min_{x,y}\{c^{\mathrm{T}}x + q^{\mathrm{T}}y \mid Ax = b,\; Tx + Wy = h(\xi);\; x,y \geq 0\}\,,$$

which is obviously convex in $\xi$. Then by Jensen's inequality follows

$$\gamma(\bar{\xi}) = EV \leq \mathbb{E}_\xi\,[\gamma(\xi)] = WS\,.$$

□

For more general recourse functions the inequality of Prop. 3.6. cannot be expected to hold true; for a counterexample see Birge–Louveaux [26].

Furthermore, in Avriel–Williams [9] the *expected value of perfect information* $EVPI$ was introduced as

$$EVPI := RS - WS \tag{3.110}$$

and may be understood in applications as the maximal amount a decision maker would be willing to pay for the exact information on future outcomes of the random vector $\xi$. Obviously due to Prop. 3.5. we have $EVPI \geq 0$. However, to compute this value exactly would require by (3.110) to solve the original recourse problem (3.103) as well as the wait-and-see problem (3.106), both of which may turn out to be hard tasks. Hence the question of easier computable and still sufficiently tight bounds on the $EVPI$ was widely discussed. As may be expected, the results on bounding the expected recourse function mentioned earlier are used for this purpose as well as approaches especially designed for bounding the $EVPI$ as presented e.g. in Huang–Vertinsky–Ziemba [143] and some of the references therein.

Finally, the *value of the stochastic solution* was introduced in Birge [22] as the quantity

$$VSS := EEV - RS\,, \tag{3.111}$$

which in applications may be given the interpretation of the expected loss for neglecting stochasticity in determining the first stage decision, as mentioned with the $EV$ solution of (3.104). Obviously it measures the extra cost for using, instead of the "true" first stage solution for the recourse problem (3.103), the first stage solution of the expected value problem (3.104). Also in this case Prop. 3.5. implies $VSS \geq 0$.

If in the problem at hand there is no randomness around, in other words if with some fixed $\hat{\xi} \in \mathbb{R}^r$ we have $\mathbb{P}_\xi\,(\xi = \hat{\xi}) = 1$, then obviously follows $EVPI = VSS = 0$. In turn, if one of these characteristic values is strictly positive, it is often considered as a "measure of the degree of stochasticity" of the recourse problem. However, one must be careful with this interpretation; it should be observed that examples can be given for which either $EVPI = 0$ and $VSS > 0$ or, on the other side, $EVPI > 0$ and $VSS = 0$ (see Birge–Louveaux [26]). Hence, the impact of stochasticity to the $EVPI$ and the $VSS$ may be rather different. Although these values are not comparable in general, there are at least some joint bounds:

**Proposition 3.7.** *With the recourse function $Q(x; T(\xi), h(\xi))$, allowing only for the matrix $T(\cdot)$ and the right–hand–side $h(\cdot)$ to contain random data, the value of the stochastic solution has the upper bound*

$$VSS \leq EEV - EV\,. \tag{3.112}$$

*With the recourse function $Q(x; h(\xi))$, i.e. with only the right–hand–side $h(\cdot)$ being random, the expected value of perfect information is bounded above as*

$$EVPI \leq EEV - EV\,. \tag{3.113}$$

*Proof:* Due to (3.109) in Prop. 3.5., we have $RS \geq EV$ and therefore

$$VSS = EEV - RS \leq EEV - EV\,.$$

From Prop. 3.6. we know that with the recourse function $Q(x;h(\xi))$ holds $EV \leq WS$. Hence, together with Prop. 3.5. we get

$$EVPI = RS - WS \leq EEV - EV\,.$$

□

The above bounds are due to Avriel–Williams [9] for the $EVPI$ and Birge [22] for the $VSS$.

In the literature, you may occasionally find statements claiming that the bounds given in (3.112) and (3.113) hold true without the restrictions made in Prop. 3.7.. There are obvious reasons to doubt those claims. Concerning $VSS$ the above argument for (3.109) using Jensen's inequality fails as soon as we loose the convexity of the recourse function in $\xi$ for any fixed $\tilde{x}$. For the $EVPI$ we present again the following example (as mentioned in Kall [154]):

**Example 3.2.** *With $X = \mathbb{R}_+$ let $c = 2$, $W = (1,-1)$, $q = (1,0)^{\mathrm{T}}$ and*

$$\mathbb{P}_\xi\,\{(T^{(1)},h^{(1)}) = (1,2)\} = \mathbb{P}_\xi\,\{(T^{(2)},h^{(2)}) = (3,12)\} = \frac{1}{2}\,.$$

*Then we have $\bar{T} = 2$, $\bar{h} = 7$ and*

$$EV = \min\{2x + y_1 \mid 2x + y_1 - y_2 = 7;\ x \geq 0, y \geq 0\} = 7 \ \text{ with } \hat{x} = \frac{7}{2}\,.$$

*With*

$$Q(\hat{x};T^{(1)},h^{(1)}) = \min_y\{y_1 \mid y_1 - y_2 = 2 - \hat{x}, y \geq 0\} = 0$$

*and*

$$Q(\hat{x};T^{(2)},h^{(2)}) = \min_y\{y_1 \mid y_1 - y_2 = 12 - 3\hat{x}, y \geq 0\} = \frac{3}{2}$$

*follows*

$$EEV = 2\cdot\frac{7}{2} + \frac{1}{2}\cdot\frac{3}{2} = 7.75$$

*and hence $EEV - EV = 0.75$. On the other hand we get $RS$ as optimal value from*

$$\begin{array}{llll}
\min\{2\cdot x + 0.5\cdot y_1^{(1)} + 0.5\cdot y_1^{(2)}\} & & & \\
1\cdot x + 1\cdot y_1^{(1)} - 1\cdot y_2^{(1)} & & = & 2 \\
3\cdot x + & 1\cdot y_1^{(2)} - 1\cdot y_2^{(2)} & = & 12 \\
 & x,\ y^{(1)},\ y^{(2)} & \geq & 0,
\end{array}$$

*yielding $RS = 7$ with $x^\star = 2$, $y_1^{(2)} = 6$. To get the $WS$ we compute*

$$\gamma_1 := \min\{2\cdot x + y_1 \mid 1\cdot x + 1\cdot y_1 - 1\cdot y_2 = 2;\ x, y \geq 0\} = 2$$

*and*

$$\gamma_2 := \min\{2\cdot x + y_1 \mid 3\cdot x + 1\cdot y_1 - 1\cdot y_2 = 12;\ x, , y \geq 0\} = 8$$

*yielding* $WS = 0.5\cdot 2 + 0.5\cdot 8 = 5$ *such that*

$$EVPI = RS - WS = 2 > EEV - EV = 0.75\,.$$

□

## 3.3 The multi-stage SLP

According to (3.1) on page 192 the general MSLP may be stated as

$$\left.\begin{array}{ll} \min\{c_1^{\mathrm{T}} x_1 + \mathbb{E}\sum_{t=2}^{T} c_t^{\mathrm{T}}(\zeta_t)x_t(\zeta_t)\} & \\ A_{11}x_1 \qquad\qquad\qquad\qquad = b_1 & \\ A_{t1}(\zeta_t)x_1 + \sum_{\tau=2}^{t} A_{t\tau}(\zeta_t)x_\tau(\zeta_\tau) = b_t(\zeta_t) & \text{a.s.}, t = 2,\cdots,T, \\ \qquad x_1 \geq 0,\ x_t(\zeta_t) \geq 0 & \text{a.s.}, t = 2,\cdots,T, \end{array}\right\} \tag{3.114}$$

where on a given probability space $(\Omega, \mathscr{G}, P)$ random vectors $\xi_t : \Omega \longrightarrow \mathbb{R}^{r_t}$ are defined, with $\xi = (\xi_2^{\mathrm{T}}, \cdots, \xi_T^{\mathrm{T}})^{\mathrm{T}}$ inducing the probability distribution $\mathbb{P}_\xi$ on $\mathbb{R}^{r_2+\cdots+r_T}$, and $\zeta_t = (\xi_2^{\mathrm{T}}, \cdots, \xi_t^{\mathrm{T}})^{\mathrm{T}}$ the state variable at stage $t$.

**Remark 3.9.** *Not to overload the notation, for the remainder of this section, instead of* $\xi = (\xi_2^{\mathrm{T}}, \cdots, \xi_T^{\mathrm{T}})^{\mathrm{T}}$ *and* $\zeta_t = (\xi_2^{\mathrm{T}}, \cdots, \xi_t^{\mathrm{T}})^{\mathrm{T}}$, *we shall write* $\xi = (\xi_2, \cdots, \xi_T)$ *and* $\zeta_t = (\xi_2, \cdots, \xi_t)$, *understanding that* $\xi = (\xi_2, \cdots, \xi_T) \in \mathbb{R}^{r_2+\cdots+r_T}$ *and* $\zeta_t = (\xi_2, \cdots, \xi_t) \in \mathbb{R}^{r_2+\cdots+r_t}$, *as before.* □

Furthermore, the (random) decisions $x_t(\cdot)$ are required to be $\mathscr{F}_t$-measurable, with $\mathscr{F}_t = \sigma(\zeta_t) \subset \mathscr{G}$. Since $\{\mathscr{F}_1, \cdots, \mathscr{F}_T\}$ is a filtration, this implies the nonanticipativity of the feasible policies $\{x_1(\cdot), \cdots, x_T(\cdot)\}$. Finally, Assumption 3.1., page 192, prescribes the square-integrability of $\xi_t(\cdot)$ w.r.t. $P$ for $t = 1, \cdots, T$, and $A_{t\tau}(\cdot), b_t(\cdot), c_t(\cdot)$ are assumed to be linear affine in $\zeta_t$. In addition, we have required the square-integrability of the decisions $x_t(\cdot)$.

Obviously, for $\xi$ having a non-discrete distribution, to solve problem (3.114) means to determine decision functions $x_t(\cdot)$ (instead of decision variables) satisfying infinitely many constraints, which appears to be a very hard task to achieve, in general. The problem becomes more tractable for the case of $\xi$ having a finite discrete distribution, a situation found or assumed in most applications of this model.

### *3.3.1 MSLP with finite discrete distributions*

Let $\xi : \Omega \longrightarrow \mathbb{R}^R$, $R = \sum_{t=2}^{T} r_t$, be a random vector with a finite discrete distribution, having the realizations $\widehat{\xi}^1, \widehat{\xi}^2, \cdots, \widehat{\xi}^S$ with the probabilities $q_1, q_2, \cdots, q_S$, respectively.

Anyone of these realizations is also denoted as a *scenario* $\widehat{\xi}^s = (\widehat{\xi}_2^s, \cdots, \widehat{\xi}_T^s)$ with the probability $\mathbb{P}_\xi\{\xi = \widehat{\xi}^s\} = q_s$, $s \in \mathscr{S} := \{1, \cdots, S\}$. Then the time discrete stochastic process $\{\zeta_t;\ t = 2, \cdots, T\}$ with discretely distributed state variables $\zeta_t$ may be assigned to a *scenario tree* as follows:

- The (deterministic) state of the system at stage 1 is assigned to node 1, the unique root of the tree.
- Among all scenarios $\widehat{\xi}^s$, $s = 1, \cdots, S$, there are a finite number $k_2$ having pairwise different realizations $\widehat{\zeta}_2^s$ of the stage 2 state variables, denoted as $\widehat{\zeta}_2^{\rho(n)} = \widehat{\xi}_2^{\rho(n)}$, $n = 2, \cdots, 1 + k_2$, and assigned to the nodes numbered as $n = 2, \cdots, 1 + k_2 =: K_2$. Here $\rho(n)$ refers to the first of the scenarios $\widehat{\xi}^s$, $s = 1, \cdots, S$, passing through the particular state $\widehat{\zeta}_2^s$. Node 1 is connected by an arc to each of the $k_2$ nodes in stage 2 due to the fact, that the corresponding states in stage 2 are realized by at least one scenario.
- Having assigned, according to all scenarios, up and until stage $t < T$ the nodes and arcs to all states and implied transitions between consecutive states (i.e. given a scenario $\widehat{\xi}^s = (\widehat{\xi}_2^s, \cdots, \widehat{\xi}_{t-1}^s, \widehat{\xi}_t^s, \cdots, \widehat{\xi}_T^s)$, implies a transition from state $\widehat{\zeta}_{t-1}^s = (\widehat{\xi}_2^s, \cdots, \widehat{\xi}_{t-1}^s)$ to $\widehat{\zeta}_t^s = (\widehat{\xi}_2^s, \cdots, \widehat{\xi}_t^s)$ at least once), we consider for each scenario $\widehat{\xi}^s$ the state $\widehat{\zeta}_{t+1}^s = (\widehat{\xi}_2^s, \cdots, \widehat{\xi}_{t+1}^s)$. Again, in stage $t+1$ there is a finite number $k_{t+1}$ of different states denoted as $\widehat{\zeta}_{t+1}^{\rho(n)}$, $n = K_t + 1, \cdots, K_t + k_{t+1} =: K_{t+1}$, and assigned to the nodes $K_t + 1, \cdots, K_t + k_{t+1} =: K_{t+1}$ (with $\rho(n)$ referring again to the first scenario passing through this particular state). Finally, we insert the arcs from stage $t$ to stage $t+1$ according to the implied transitions.

With this scenario tree, representing graphically the possible developments of the stochastic process $\{\xi_2, \cdots, \xi_T\}$ over time, we may combine probabilistic information to get a complete description of the process (see Fig. 3.6).

To this end, we may identify the leaf nodes of the tree (the stage T nodes) $K_{T-1} + 1, \cdots, K_T$ with the scenarios $\widehat{\xi}^s$, $s = 1, \cdots, S$, and assign to these nodes the probabilities $q_s$ of the respective scenario. Hence we have first the probabilities to reach the leaf nodes $n = K_{T-1} + 1, \cdots, K_T$ as $p_n = q_{n-K_{T-1}}$.

For all other nodes, i.e. for $n \leq K_{T-1}$, we then compute the probabilities $p_n$ to pass through these nodes: Given node $n$, by the above construction of the scenario tree we know the stage $t_n$ of this node as well as its corresponding state $\widehat{\zeta}_{t_n}^{\rho(n)}$; then with $\mathscr{S}(n) = \{s \mid \widehat{\zeta}_{t_n}^s = \widehat{\zeta}_{t_n}^{\rho(n)}\}$ we have $\{\widehat{\xi}^s \mid s \in \mathscr{S}(n)\}$, the set of scenarios passing through this state, called the *scenario bundle* of node $n$, and we get $p_n$, the total probability of this scenario bundle, as $p_n = \sum_{s \in \mathscr{S}(n)} q_s$.

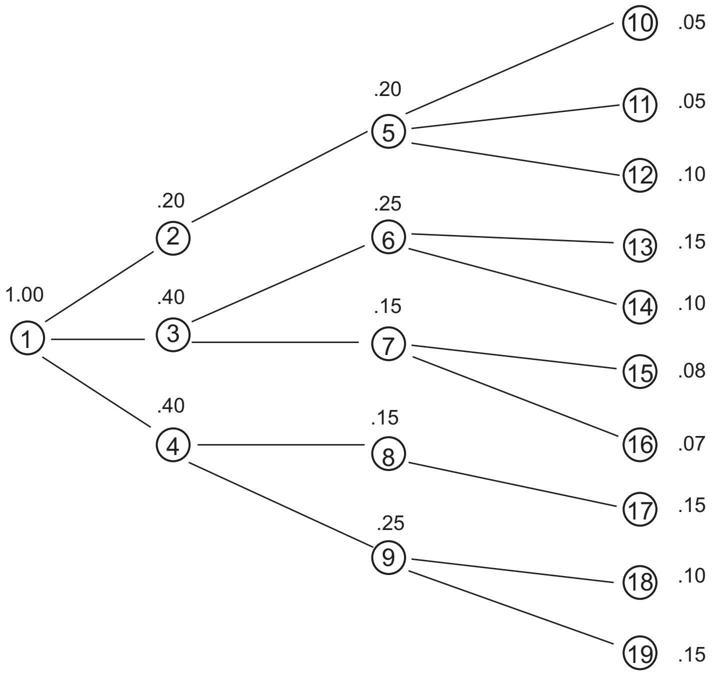


**Fig. 3.6** Four-stage scenario tree representing a stochastic process.

After the above description of a scenario tree it seems to be meaningful to introduce the following collection of specific variables and sets for discussing various issues on scenario trees. These entities have shown to be useful when dealing with rather complex problems defined on scenario trees, like e.g. multi-stage SLP's with finite discrete distributions, as to be discussed next. There we shall make use of the following

Notation for scenario trees:

$(\mathscr{N}, \mathscr{A})$ : rooted tree with nodes $\mathscr{N} \subset \mathbb{N}$ ($n = 1$ the unique root), and $\mathscr{A}$ the set of arcs.
The nodes $n \in \mathscr{N}$ are assigned to stages $t = 1, \cdots, T$, with $n = 1$ in stage $t = 1$, and with $k_t > 0$ nodes for $t = 2, \cdots, T$, and $|\mathscr{N}| = 1 + \sum_{t=2}^{T} k_t$ .
The arcs in $\mathscr{A}$ connect selected nodes of stage $t$ and stage $t+1$, $t = 1, \cdots, T-1$, such that each node in

some stage $t < T$ has at least one immediate successor, and each node in some stage $t > 1$ has exactly one immediate predecessor.
Any path $n_1, \cdots, n_T$, with $n_1 = 1, t_{n_t} = t$ for $t \geq 2$, and $(n_t, n_{t+1}) \in \mathscr{A}$ for $t = 1, \cdots, T-1$, corresponds one-to-one to the scenario $\widehat{\xi}^s$, $s \in \mathscr{S} = \{1, \cdots, S\}$, identified with the leaf node $n_T$ .

$q_s, s \in \mathscr{S}$ : $q_s = \mathbb{P}_\xi\{\xi = \widehat{\xi}^s\}$, the probability of scenario $\widehat{\xi}^s$, and hence the probability to reach the leaf node identified with this scenario;

$t_n$ : the stage of node $n \in \mathscr{N}$;

$\rho(n)$ : the smallest $s \in \mathscr{S}$ such that scenario $\widehat{\xi}^{\rho(n)}$ passes through the state $\widehat{\zeta}^s_{t_n}$ assigned to node $n$;

$\widehat{\zeta}^n$ : $\widehat{\zeta}^n := \widehat{\zeta}^{\rho(n)}_{t_n}$, the state in stage $t_n$ uniquely assigned to $n$;

$\mathscr{D}(t) \subset \mathscr{N}$ : the set of nodes in stage $t$ with $|\mathscr{D}(t)| = k_t$;

$h_n$ : parent node (immediate predecessor) of $n \in \mathscr{N}$, $n \geq 2$;

$\mathscr{H}(n) \subset \mathscr{N}$ : set of nodes in the unique path from $n \in \mathscr{N}$ through the successive predecessors back to the root, ordered by stages, the *history* of $n$ (including $n$);

$\mathscr{S}(n)$ : $\mathscr{S}(n) = \{s \mid \widehat{\zeta}^s_{t_n} = \widehat{\zeta}^{\rho(n)}_{t_n}\}$, the index set identifying the scenario bundle of node $n$;

$p_n$ : $p_n = \sum_{s \in \mathscr{S}(n)} q_s$, the probability to pass node $n$;

$\mathscr{C}(n) \subset \mathscr{N}$ : the set of children (immediate successors) of node $n$;

$\mathscr{G}_s(n) \subseteq \mathscr{N}$ : the future of node $n$ along scenario $\widehat{\xi}^s : \ s \in \mathscr{S}(n)$, including node $n$, i.e. the nodes $n_{t_n} = n, \cdots, n_T$ provided the path $\{n_1, \cdots, n_{t_n}, \cdots, n_T\}$ corresponds to scenario $\widehat{\xi}^s$ (hence $\mathscr{G}_s(n) = \emptyset$ if $s \notin \mathscr{S}(n)$);

$\mathscr{G}(n) \subseteq \mathscr{N}$ : the future of $n \in \mathscr{N}$, $\mathscr{G}(n) = \bigcup_{s \in \mathscr{S}(n)} \mathscr{G}_s(n)$;

$q_{n \to m}$ : $q_{n \to m} = \dfrac{p_m}{p_n} \ \forall m \in \mathscr{G}(n)$, the conditional probability to reach node $m$ given node $n$ (provided that $p_n > 0$).

To keep the following problem formulations simple, we introduce

**Assumption 3.4.** *For any MSLP with a finite discrete distribution of the scenarios $\xi$ holds*

$$q_s = \mathbb{P}_\xi\{\xi = \widehat{\xi}^s\} > 0 \ \ \forall s \in \mathscr{S}. \tag{3.115}$$

By construction the following facts are obvious:

- Through each node passes at least one scenario, i.e. $\mathscr{S}(n) \neq \emptyset \; \forall n \in \mathscr{N}$ ;
- given any stage $t$, each scenario passes through exactly one node in stage $t$, i.e. $\bigcup_{n \in \mathscr{D}(t)} \mathscr{S}(n) = \mathscr{S}$ and $\mathscr{S}(n) \cap \mathscr{S}(m) = \emptyset \; \forall n, m \in \mathscr{D}(t) : n \neq m$.

Hence, it follows in general that

$$\sum_{n \in \mathscr{D}(t)} p_n = 1, \; t = 1, \cdots, T, \tag{3.116}$$

and due to Assumption 3.4. holds

$$p_n = \sum_{s \in \mathscr{S}(n)} q_s > 0 \;\; \forall n \in \mathscr{N}. \tag{3.117}$$

For the general MSLP (3.114), the decisions $x_t(\zeta_t)$ in stage $t$ are required to be $\mathscr{F}_t$-measurable with $\mathscr{F}_t = \sigma(\zeta_t) \subset \mathscr{G}$. For $\xi$ having a finite discrete distribution, $\sigma(\zeta_t)$ is generated by the $k_t$ atoms $\zeta_t^{-1}[\widehat{\zeta}_{t_n}^{\rho(n)}]$, $n = K_{t-1}+1, \cdots, K_t$. Then $x_t(\cdot)$ has to be constant on each of these atoms or equivalently, to each node $n$ we have to determine the decision vector $x_n := x_{t_n}(\widehat{\zeta}^{\,n})$. Observing that the expected values $\mathbb{E}\,[c_t^{\mathrm{T}}(\zeta_t) x_t(\zeta_t)]$ may now be written as $\sum_{n=K_{t-1}+1}^{K_t} p_n c_{t_n}^{\mathrm{T}}(\widehat{\zeta}^{\,n}) x_n$, problem (3.114) for a discrete distribution reads as

$$\left.\begin{aligned} \min \sum_{m \in \mathscr{N}} p_m c_{t_m}^{\mathrm{T}}(\widehat{\zeta}^{\,m}) x_m & \\ \sum_{m \in \mathscr{H}(n)} A_{t_n t_m}(\widehat{\zeta}^{\,n}) x_m \;=\; b_{t_n}(\widehat{\zeta}^{\,n}) \;\; & \forall n \in \mathscr{N} \\ x_m \;\geq\; 0 \qquad\quad & \forall m \in \mathscr{N} \end{aligned}\right\} \tag{3.118}$$

with $p_1 = 1$ and $c_{t_1}^{\mathrm{T}}(\widehat{\zeta}^{\,1}) = c_1$, $A_{t_1 t_1}(\widehat{\zeta}^{\,1}) = A_{11}$, $b_{t_1}(\widehat{\zeta}^{\,1}) = b_1$ being constant.
With an obvious simplification of the notation problem (3.118) may be rewritten equivalently as

$$\left.\begin{aligned} \min \sum_{m \in \mathscr{N}} p_m c_{t_m}^{\mathrm{T}}(m) x_m & \\ \sum_{m \in \mathscr{H}(n)} A_{t_n t_m}(n) x_m \;=\; b_n \;\; & \forall n \in \mathscr{N} \\ x_m \;\geq\; 0 \;\; & \forall m \in \mathscr{N}. \end{aligned}\right\} \tag{3.119}$$

As the dual LP of (3.119) we have

$$\left.\begin{aligned} \max \sum_{n \in \mathscr{N}} b_n^{\mathrm{T}} u_n & \\ \sum_{n \in \mathscr{G}(m)} A_{t_n t_m}^{\mathrm{T}}(n) u_n \leq p_m c_{t_m}(m) \;\; & \forall m \in \mathscr{N}. \end{aligned}\right\} \tag{3.120}$$

**Remark 3.10.** *If in particular, $\forall n \in \mathscr{N} \setminus \{1\}$ and for each node $m \in \mathscr{H}(n) : t_m < t_n - 1$, we have that $A_{t_n t_m}(n) = A_{t_n t_m}(\widehat{\zeta}^n) = 0$, then with $W_1 := A_{11}$ and*

$$T_n := A_{t_n t_n - 1}(n) \text{ and } W_n := A_{t_n t_n}(n) \quad \forall n \in \mathscr{N} \setminus \{1\}$$

*problem (3.119) reads as*

$$\left.\begin{array}{rl} \min \sum_{m \in \mathscr{N}} p_m c_{t_m}^{\mathrm{T}}(m) x_m & \\ W_1 x_1 & = b_1 \\ T_n x_{h_n} + W_n x_n & = b_n \ \forall n \in \mathscr{N} \setminus \{1\} \\ x_n & \geq \ 0 \ \ \forall n \in \mathscr{N} . \end{array}\right\} \tag{3.121}$$

*Hence we have the same problem structure as assumed when discussing the nested decomposition in section 1.2.7 of Chapter 1, in particular the structure of problem (1.29) on page 33.*

*The general MSLP problem (3.114) can always be transformed to an equivalent problem where $A_{t\tau} = 0$ holds for $\tau < t - 1$, thus assuming the following staircase form*

$$\left.\begin{array}{llll} \min\{c_1^{\mathrm{T}} x_1 + \mathbb{E} \sum_{t=2}^{T} c_t^{\mathrm{T}}(\zeta_t) z_t(\zeta_t)\} & & & \\ W_1 z_1 & = & b_1 & \\ T_t(\zeta_t) z_{t-1}(\zeta_{t-1}) \ + W_t(\zeta_t) z_t(\zeta_t) & = & b_t(\zeta_t) & a.s.,\ t = 2, \cdots, T, \\ x_1 \geq 0, \ \ x_t(\zeta_t) & \geq & 0 & a.s.,\ t = 2, \cdots, T, \end{array}\right\} \tag{3.122}$$

*formally corresponding to (3.121), where now $z_t$ is an $n_1 + \ldots + n_t$–dimensional variable and $T_t$ and $W_t$ have $m_t + n_1 + \ldots + n_t$ rows. For specifying the transformation which maps (3.114) into (3.122) we will employ double indices. The transformation is as follows. Let*

$$z_t^{\mathrm{T}}(\zeta_t) = (z_{t,1}(\zeta_t), \cdots, z_{t,t-1}(\zeta_t), z_{t,t}(\zeta_t))$$

*with $z_{t,\tau}$ being an $n_\tau$–dimensional variable, $\tau = 1, \ldots, t$, and with $z_{tt}$ corresponding to $x_t$ in (3.114). The matrices are defined as follows. Let $W_1 = A_{1,1}$. For $1 < t < T$ we define*

$$T_t(\zeta_t) = \begin{pmatrix} A_{t,1}(\zeta_t) \ \ldots \ A_{t,t-1}(\zeta_t) \\ I & & \\ & \ddots & \\ & & I \end{pmatrix}$$

*and*

$$W_t(\zeta_t) = \begin{pmatrix} 0 & \dots & 0 & A_{t,t}(\zeta_t) \\ -I & & & 0 \\ & \ddots & & \vdots \\ & & -I & 0 \end{pmatrix}$$

*and for* $t = T$ *let*

$$T_T(\zeta_T) = (A_{T,1}(\zeta_T), \dots A_{T,T-1}(\zeta_T)) \ \text{ and } \ W_T(\zeta_T) = A_{TT}(\zeta_T).$$

*Loosely speaking, the auxiliary variables* $(z_{t,1}, \dots, z_{t,t-1})$ *serve for "forwarding" the solution to later stages. As an example let us consider an MSLP with* $T = 4$ *and let us drop in the notation the dependency on* $\zeta_t$*. The original structure is*

$$\left.\begin{array}{lllll} A_{11}x_1 & & & & = b_1 \\ A_{21}x_1 & +A_{22}x_2 & & & = b_2 \\ A_{31}x_1 & +A_{32}x_2 & +A_{33}x_3 & & = b_3 \\ A_{41}x_1 & +A_{42}x_2 & +A_{43}x_3 & +A_{44}x_4 & = b_4 \end{array}\right\}$$

*which transforms into*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $A_{11}z_{11}$ | | | | | | $=$ | $b_1$ |
| $A_{21}z_{11}$ | | $+A_{22}z_{22}$ | | | | $=$ | $b_2$ |
| $z_{11}$ | $-z_{21}$ | | | | | $=$ | $0$ |
| | $A_{31}z_{21}$ | $+A_{32}z_{22}$ | | $+A_{33}z_{33}$ | | $=$ | $b_3$ |
| | $z_{21}$ | | $-z_{31}$ | | | $=$ | $0$ |
| | | $z_{22}$ | $-z_{32}$ | | | $=$ | $0$ |
| | | | $A_{41}z_{31}$ $+A_{42}z_{32}$ | $+A_{43}z_{33}$ | $+A_{44}z_{44}$ | $=$ | $b_4$ |

*In the literature, multi-stage SLP's are often presented just in the so-called staircase formulation (3.121). Although problems of this form, at the first glance, look simpler than problems in the lower block triangular formulation like (3.119), this does not imply a computational advantage in general. Indeed, if the staircase formulation results from the above transformation of (3.114) into (3.122), then the numbers of variables and of constraints are increased.* □

### 3.3.2 MSLP with non-discrete distributions

In Section 3.2.1 we have discussed two-stage SLP's with complete fixed recourse and with bounded distributions, i.e. with $\operatorname{supp} \mathbb{P}_\xi \subseteq \Xi = \prod_{i=1}^r [\alpha_i, \beta_i]$. In particular, we considered the recourse function $Q(x; T(\xi), h(\xi))$, which according to our notation (see page 197) implies for the second stage problem (3.9) that only $T(\cdot)$ and $h(\cdot)$ (or some elements of these arrays) are random. In this case, we could apply Jensen's inequality to get in Theorem 3.4. a lower bound for the expected recourse $\mathscr{Q}(x) = \int_{\Xi} Q(x;T(\xi),h(\xi))\mathbb{P}_{\xi}(d\xi)$ as $Q(x;T(\bar{\xi}),h(\bar{\xi})) \leq \mathscr{Q}(x)$, where $\bar{\xi} := \mathbb{E}_{\xi}[\xi]$. In other words, introducing the Jensen distribution $\mathbb{P}_{\eta}$ as the one-point distribution with $\mathbb{P}_{\eta}\{\eta = \mathbb{E}_{\xi}[\xi]\} = 1$, the Jensen inequality can formally be written as

$$\int_{\Xi} Q(x;T(\eta),h(\eta))\mathbb{P}_{\eta}(d\eta) \leq \mathscr{Q}(x).$$

On the other hand, we have derived particular discrete probability distributions $\mathbb{Q}_{\eta}$ on the vertices $v^{\nu}$ of $\Xi$, the E–M distribution for stochastically independent components of $\xi$ in Lemma 3.6 and the generalized E–M distribution for stochastically dependent components of $\xi$ in Lemma 3.7, respectively, which were shown to solve two special types of moment problems. According to Theorems 3.5. and 3.6., using these distributions the E–M inequality provides an upper bound for the expected recourse as

$$\begin{aligned}
\mathscr{Q}(x) &\leq \int_{\Xi} Q(x;T(\eta),h(\eta))\,\mathbb{Q}_{\eta}(d\eta) \\
&= \sum_{\nu=1}^{2^r} Q(x;T(v^{\nu}),h(v^{\nu}))\,\mathbb{Q}_{\eta}(v^{\nu}).
\end{aligned}$$

For any disjoint interval partition $\mathscr{X} = \{\Xi_k;\ k = 1,\cdots,K\}$ of $\Xi$, we apply Jensen's inequality for the conditional expectations, meaning to introduce on the set of conditional expectations $\{\bar{\xi}_k := \mathbb{E}_{\xi}[\xi \mid \xi \in \Xi_k] \mid k = 1,\cdots,K\}$ the corresponding discrete distribution $\mathbb{P}_{\eta_{\mathscr{X}}}$, defined by $\mathbb{P}_{\eta_{\mathscr{X}}}\{\bar{\xi}_k\} = \mathbb{P}_{\xi}\{\Xi_k\}$, and to compute $\int_{\Xi} Q(x;T(\eta),h(\eta))\mathbb{P}_{\eta_{\mathscr{X}}}(d\eta)$ to get a lower bound for $\mathscr{Q}(x)$. Similarly, we apply the E–M inequality using the distribution $\mathbb{Q}_{\eta_{\mathscr{X}}} = \sum_{k=1}^{K} \mathbb{P}_{\xi}\{\Xi_k\} \cdot \mathbb{Q}_{\eta_{\Xi_k}}$, where $\mathbb{Q}_{\eta_{\Xi_k}}$ is either the E–M distribution or else the generalized E–M distribution solving the corresponding conditional moment problems, conditioned with respect to the cell $\Xi_k \in \mathscr{X}$. This way, according to Lemma 3.8 we get an increased lower bound as well as a decreased upper bound.

For any sequence of appropriately refined interval partitions $\{\mathscr{X}^{\nu}\}$ the corresponding sequences of discrete distributions $\{\mathbb{P}_{\eta_{\mathscr{X}^{\nu}}}\}$ and $\{\mathbb{Q}_{\eta_{\mathscr{X}^{\nu}}}\}$ of Jensen distributions and E–M distributions, respectively, are shown in Lemma 3.9 to converge weakly to the original distribution $\mathbb{P}_{\xi}$. For the corresponding sequences $\{\widetilde{\mathscr{Q}}^{\nu}\}$ and $\{\widehat{\mathscr{Q}}^{\nu}\}$ of Jensen lower bounds and E–M upper bounds, respectively, of the expected recourse function $\mathscr{Q}$, this implies epi-convergence of both sequences to $\mathscr{Q}$. This convergence behaviour, however, provides due to Theorem 3.7. promising conditions to design approximation schemes for the solution of two-stage SLP's with complete fixed recourse.

The question arises whether we may expect a similar approach to be applicable for the solution of multi-stage SLP's with more than two stages. To get a first impression let us take a look at a rather simple three-stage example.

**Example 3.3.** *Consider the complete fixed recourse problem*

$$\begin{aligned} V^\star := \min\{2x + \mathbb{E}[y_1(\xi_2) + 2y_2(\xi_2)] + \mathbb{E}[z_1(\xi_2,\xi_3) + z_2(\xi_2,\xi_3)]\} & \\ \text{s.t. } x + y_1(\xi_2) - y_2(\xi_2) \qquad\qquad\qquad\qquad &= \xi_2 \\ x + y_1(\xi_2) - y_2(\xi_2) + z_1(\xi_2,\xi_3) - z_2(\xi_2,\xi_3) &= \xi_3 \\ x, y_1, y_2, z_1, z_2 &\geq 0 \end{aligned}$$

*with* $\zeta := (\xi_2,\xi_3)^{\mathrm{T}}$ *having the (joint) probability distribution* $\mathbb{P}_\zeta$ *on* $\operatorname{supp}\mathbb{P}_\zeta := \Xi = [0,1]\times[0,1]$, *given by the density*

$$f(\xi_2,\xi_3) = \begin{cases} 1+\varepsilon \text{ for } & 0 \le \xi_2,\xi_3 \le 0.5 \\ 1+\varepsilon \text{ for } & 0.5 \le \xi_2,\xi_3 \le 1 \\ 1-\varepsilon \text{ for } & 0 \le \xi_2 < 0.5 < \xi_3 \le 1 \\ 1-\varepsilon \text{ for } & 0 \le \xi_3 < 0.5 < \xi_2 \le 1 \\ 0 \qquad \text{else}, & \end{cases}$$

*where* $\varepsilon$ *is some constant such that* $\varepsilon \in (-1,+1)$.

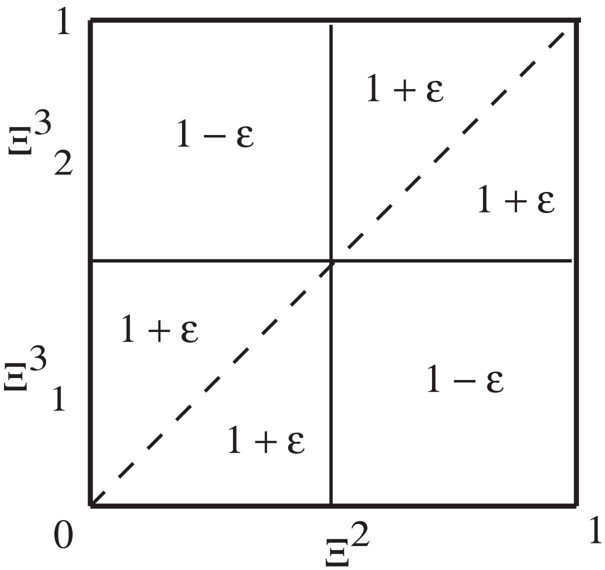


**Fig. 3.7** $\operatorname{supp}\mathbb{P}_\zeta = \Xi^2 \times \Xi^3 = \Xi^2 \times (\Xi^3_1 \cup \Xi^3_2)$ with density $f(\xi_2,\xi_3)$.

*For the marginal distribution of* $\xi_2$ *we obviously get the marginal density as*

$$f_2(\xi_2) = \int_0^1 f(\xi_2,\xi_3)d\xi_3 = \begin{cases} 1 \; for \; \xi_2 \in [0,1] \\ 0 \; else, \end{cases}$$

*such that the corresponding distribution* $\mathbb{P}_{\xi_2}$ *is* $\mathscr{U}[0,1]$*, the uniform distribution on the interval* $[0,1]$*. According to the definition of* $f(\xi_2,\xi_3)$*, for* $\xi_3$ *follows the same marginal distribution.*

*Considering, for instance, the interval* $\widetilde{\Xi} := \{[0,0.5]\times[0,0.5]\} \subset \mathbb{R}^2$*, we get*

$$\mathbb{P}_\zeta(\widetilde{\Xi}) = \int_{\widetilde{\Xi}} f(\zeta)d\zeta = \frac{1}{4}(1+\varepsilon)\,,$$

*whereas for the marginal distributions in* $\mathscr{U}[0,1]$ *follows*

$$\mathbb{P}_{\xi_2}([0,0.5])\cdot\mathbb{P}_{\xi_3}([0,0.5]) = \frac{1}{4}\,.$$

*Hence, for* $\varepsilon \neq 0$ *the random variables* $\xi_2$ *and* $\xi_3$ *are dependent.*

*Due to the objective of our recourse problem, for any given first stage solution* $x \geq 0$ *the second stage solution* $y_i(\xi_2), i=1,2$*, minimizing the second stage objective* $y_1(\xi_2)+2y_2(\xi_2)$*, has to satisfy the rules*

a) $\xi_2 < x \Longrightarrow y_1(\xi_2)=0,\, y_2(\xi_2) = x-\xi_2$
b) $\xi_2 \geq x \Longrightarrow y_1(\xi_2) = \xi_2 - x,\, y_2(\xi_2)=0.$

*Minimizing the third stage objective* $z_1(\xi_2,\xi_3)+z_2(\xi_2,\xi_3)$ *then yields, for both of the cases* a) *and* b) *above,*

$$\begin{aligned} x+y_1(\xi_2)-y_2(\xi_2) \leq \xi_3 &\Longrightarrow z_1(\xi_2,\xi_3)=\xi_3-\xi_2,\, z_2(\xi_2,\xi_3)=0 \\ x+y_1(\xi_2)-y_2(\xi_2) > \xi_3 &\Longrightarrow z_1(\xi_2,\xi_3)=0,\, z_2(\xi_2,\xi_3)=\xi_2-\xi_3\,. \end{aligned}$$

*Observe that a first stage decision* $x<0$ *is not feasible. On the other hand,* $x>1$ *cannot be optimal, since this would increase unnecessarily the overall objective, more precisely the first stage cost* $2x$ *plus the expected second stage cost* $\mathbb{E}[y_1(\xi_2)+2y_2(\xi_2)]$ *due to* a) *by at least* $2(x-1)+2\mathbb{E}[(x-\xi_2)] > 4(x-1)$*. Hence we compute the objective value, for* $0 \leq x \leq 1$*, as*

$$\begin{aligned} V(x) &= 2x + \int_{\xi_2=0}^{x} 2(x-\xi_2)d\xi_2 + \int_{\xi_2=x}^{1}(\xi_2-x)d\xi_2 + \\ &\quad + \int_{\Xi} |\xi_3-\xi_2| f(\xi_2,\xi_3)d\xi_2 d\xi_3 \\ &= 2x + \frac{3}{2}x^2 - x + \frac{1}{2} + \int_{\Xi} |\xi_3-\xi_2| f(\xi_2,\xi_3)d\xi_2 d\xi_3\,. \end{aligned}$$

*For the last integral we get*

$$\int_{\Xi} |\xi_3 - \xi_2| f(\xi_2, \xi_3) d\xi_2 d\xi_3 = \underbrace{\int_{\xi_2=0}^{1} \int_{\xi_3=\xi_2}^{1} (\xi_3 - \xi_2) f(\xi_2, \xi_3) d\xi_2 d\xi_3}_{A} + \underbrace{\int_{\xi_3=0}^{1} \int_{\xi_2=\xi_3}^{1} (\xi_2 - \xi_3) f(\xi_2, \xi_3) d\xi_2 d\xi_3}_{B},$$

*where $A = B$ for symmetry reasons (see Fig. 3.7). For A, the integral taken over the triangle above the line $\xi_3 = \xi_2$ in Fig. 3.7, we get by integration of $(\xi_3 - \xi_2) f(\xi_2, \xi_3)$*

$$A = \frac{1}{2}(1+\varepsilon)\frac{1}{24} + \frac{1}{2}(1-\varepsilon)\frac{1}{4} + \frac{1}{2}(1+\varepsilon)\frac{1}{24} = \frac{2-\varepsilon}{12}$$

*such that $A + B = \dfrac{2-\varepsilon}{6}$ and hence*

$$V(x) = \frac{3}{2}x^2 + x + \frac{1}{2} + \frac{2-\varepsilon}{6} = \frac{3}{2}x^2 + x + \frac{5-\varepsilon}{6}.$$

*Obviously, $\min_{x \geq 0} V(x)$ is achieved at $\hat{x} = 0$ such that the optimal value of our problem turns out to be*

$$V^\star = \min_{x \geq 0} V(x) = \frac{5-\varepsilon}{6}.$$

*Let us now discretize the distributions of $\xi_2$ in stage two and $\zeta = (\xi_2, \xi_3)^{\mathrm{T}}$ in stage three by choosing the partitions $\mathscr{X}^2$ of $\Xi^2$ and $\mathscr{X}^3$ of $\Xi^2 \times \Xi^3$, respectively, as follows:*

Stage 2: $\mathscr{X}^2 = \{\Xi^2\}$ *yielding for $\xi_2$ the realization*

$$\bar{\xi}_2 = \mathbb{E}_{\xi_2}[\xi_2] = \frac{1}{2} \ \text{ with } \ p_2 = \mathbb{P}(\{\xi_2 \in [0,1]\}) = 1;$$

Stage 3: $\mathscr{X}^3 = \left\{ \Xi^2 \times \left[0, \frac{1}{2}\right), \Xi^2 \times \left[\frac{1}{2}, 1\right] \right\}$ *yielding for $\xi_3$ the realizations*

$$\bar{\xi}_{31} = \mathbb{E}\left[\xi_3 \mid \xi_2 \in [0,1], \xi_3 \in \left[0, \frac{1}{2}\right)\right] = \mathbb{E}\left[\xi_3 \mid \xi_3 \in \left[0, \frac{1}{2}\right)\right] = \frac{1}{4}$$

$$\bar{\xi}_{32} = \mathbb{E}\left[\xi_3 \mid \xi_2 \in [0,1], \xi_3 \in \left[\frac{1}{2}, 1\right]\right] = \mathbb{E}\left[\xi_3 \mid \xi_3 \in \left[\frac{1}{2}, 1\right]\right] = \frac{3}{4}$$

*with*

$$p_{31} = \mathbb{P}\left(\left\{\xi_3 \in \left[0, \frac{1}{2}\right)\right\}\right) = \frac{1}{2} \ \text{ and }$$

$$p_{32} = \mathbb{P}\left(\left\{\xi_3 \in \left[\frac{1}{2}, 1\right]\right\}\right) = \frac{1}{2}.$$

*Then the discretized problem reads as*

$$
\begin{array}{rl}
\overline{V} := \min \left\{ 2x + y_1 + 2y_2 + \frac{1}{2}(z_1^1 + z_2^1) + \frac{1}{2}(z_1^2 + z_2^2) \right\} & \\
\textit{s.t. } x + y_1 - y_2 & = \frac{1}{2} \\
x + y_1 - y_2 + z_1^1 - z_2^1 & = \frac{1}{4} \\
x + y_1 - y_2 + z_1^2 - z_2^2 & = \frac{3}{4} \\
x, y_1, y_2, z_1^1, z_2^1, z_1^2, z_2^2 & \geq 0 .
\end{array}
$$

*Also in this case the optimum is achieved for* $\tilde{x} = 0$ *with* $\overline{V} = \frac{3}{4}$*. Comparing this value with the optimum* $V^\star = \frac{5-\varepsilon}{6}$ *of the original problem, we see that*

$$
\overline{V} \begin{cases} \leq V^\star \text{ if } \varepsilon \leq \frac{1}{2} \\ > V^\star \text{ if } \varepsilon > \frac{1}{2} . \end{cases}
$$

*In conclusion, even for a rather simple situation like three stages, randomness in the right–hand–sides only, and complete fixed recourse, we cannot expect in general to get a lower bound of the optimum by discretization of the distributions in an analogous manner as in the two-stage case.* □

This example as well as the following considerations are essentially based on discussions related to an idea, originally due to S. Sen, concerning refinements of discretizations in order to improve discrete approximations for MSLP problems. The outcome of these endeavours was reported in Fúsek–Kall–Mayer–Sen–Siegrist [108].

Obviously, with appropriate successive refinements of partitions $\mathscr{X}_\nu^t$ of the sets $[\Xi^2 \times \cdots \times \Xi^t] \supseteq \operatorname{supp} \mathbb{P}_{\zeta_t}$, $t = 2, \cdots, t$; $\nu = 1, 2, \cdots$, we may expect weak convergence of the associated discrete distributions $\{\mathbb{P}_{\eta_t, \mathscr{X}_\nu^t}\}$ and hence epi-convergence of the related objective functions of the general MSLP (3.114), as shown by Pennanen [251, 252]. Thus Th. 3.7. (page 222) suggests that a solution could be approximated by this kind of successive discretization of the distributions. However it seems difficult to control this procedure since, in difference to the two-stage case, for the general MSLP we do not have error bounds on the optimal value. According to Ex. 3.3., even for the much simpler problem

$$\left.\begin{array}{lrl}\min\{c_1^{\mathrm{T}}x_1 + \mathbb{E}\sum_{t=2}^{T} c_t^{\mathrm{T}} x_t(\zeta_t)\} & & \\ A_{11}x_1 & = & b_1 \\ A_{t1}x_1 \ + \sum_{\tau=2}^{t} A_{t\tau} x_\tau(\zeta_\tau) & = & b_t(\zeta_t) \text{ a.s., } t = 2,\cdots,T, \\ x_1 \geq 0,\ \ x_t(\zeta_t) & \geq & 0 \qquad \text{ a.s., } t = 2,\cdots,T, \end{array}\right\} \tag{3.123}$$

with complete fixed recourse and only the right–hand–sides being random, we cannot expect to get at least lower bounds, in general.

Nevertheless, we shall discuss first, for the purpose of defining a fully aggregated problem instead of the MSLP (3.114), how an arbitrary finite subfiltration $\widehat{\mathscr{F}}$ and the corresponding scenario tree can be generated. Again, we assume the supports of the stagewise distributions to be bounded. Hence there exist intervals $\Xi^t \subset \mathbb{R}^{r_t}$ such that $\operatorname{supp} \mathbb{P}_{\xi_t} \subseteq \Xi^t$, $t = 2,\cdots,T$. Then we proceed as follows:

<u>Subfiltration and the corresponding scenario tree</u>

- With $\Omega^{(1)} := \Omega$ and $\widehat{\mathscr{F}_1} := \{\Omega, \emptyset\}$ define $\mathscr{N}_1 := \{1\}$.
- For the stages $\nu = 1,\cdots,T-1$ repeat:

  Let $\mathscr{N}_{\nu+1} := \emptyset$.
  Then for each node $n$ in stage $\nu$ (i.e. $t_n = \nu$) and some $r_n \geq 1$:
  Define a finite set $C_n$ of children of $n$ such that $|C_n| = r_n$ and, for any $m$ with $m \neq n$, $t_m = t_n = \nu$, that $C_m \cap C_n = \emptyset$ as well as $C_n \cap \mathscr{N}_\mu = \emptyset\ \forall \mu \leq \nu$ holds. Furthermore, let $\mathscr{N}_{\nu+1} := \mathscr{N}_{\nu+1} \cup C_n$ and associate individually with the set $C_n := \{k_1^{(n)},\cdots,k_{r_n}^{(n)}\}$ a partition of $\Xi^{\nu+1}$ into subintervals as

$$\Xi^{\nu+1} = \bigcup_{l=1}^{r_n} \Xi^{\nu+1}_{k_l^{(n)}}. \tag{3.124}$$

- To generate the subfiltration, for $t = 2,\cdots,T$ repeat:
  For each $n \in \mathscr{N}_t$ and $h_n \in \mathscr{N}_{t-1}$, its unique parent node, and $\Xi^t_n$ the subinterval corresponding to node $n$ in the partition of $\Xi^t$ associated with $C_{h_n}$, let $\Omega^{(n)} := \Omega^{(h_n)} \cap \xi_t^{-1}[\Xi^t_n]$.
  Define the subfiltration $\widehat{\mathscr{F}}$ by $\widehat{\mathscr{F}_t} := \sigma\{\Omega^{(n)} \mid n \in \mathscr{N}_t\}$, $t = 2,\cdots,T$, with $\sigma\{\Omega^{(n)} \mid n \in \mathscr{N}_t\}$ the $\sigma$-algebra generated by the sets $\Omega^{(n)}$, $n \in \mathscr{N}_t$.
- The defining elements of the discretely distributed stochastic process corresponding to the above finite subfiltration, i.e. the realizations $\widehat{\zeta}^n$ at node $n$ and their probabilities $p_n$, may be assigned to the nodes as follows:
  For any $n \in \mathscr{N} \setminus \{1\}$ let $\mathscr{H}(n) = \{\ell_1 = 1,\cdots,\ell_{t_n-1},\ell_{t_n} = n\}$ be the history of node $n$. By the above construction, each node $\ell_\nu \in \mathscr{H}(n)$ corresponds uniquely to a particular subinterval $\Xi^\nu_{l_\nu}$ of $\Xi^\nu$. Then for the discrete process we choose the state $\widehat{\zeta}^n$ at node $n$ and the corresponding probability $p_n$ as

$$\left.\begin{aligned} \widehat{\zeta}^n &= \mathbb{E}\left[\zeta_{t_n} \mid \zeta_{t_n} \in \prod_{\nu=2}^{t_n} \Xi_{\ell_\nu}^\nu\right] \\ p_n &= \mathbb{P}_{\zeta_{t_n}}\left(\left\{\zeta_{t_n} \in \prod_{\nu=2}^{t_n} \Xi_{\ell_\nu}^\nu\right\}\right). \end{aligned}\right\} \tag{3.125}$$

Using this discrete process we may then replace the general MSLP (3.114), defined with respect to the filtration $\mathscr{F}$, by the fully aggregated problem with respect to the subfiltration $\widehat{\mathscr{F}}$, as represented by the LP (3.118).

Whereas, according to Ex. 3.3., for problem (3.123) we cannot expect to achieve lower bounds for the optimal value by discretization of the underlying stochastic process in general, the situation will be better if Assumption 3.1. is modified as follows:

**Assumption 3.5.** *Let*

- *only the right–hand–sides $b_t$ be random (and linear affine in $\zeta_t$);*
- *the distributions of $\xi_t$ be bounded within some intervals $\Xi^t \subset \mathbb{R}^{r_t}$, i.e.* $\operatorname{supp} \mathbb{P}_{\xi_t} \subseteq \Xi^t$*;*
- *the random vectors $\xi_2, \cdots, \xi_T$ be stochastically independent;*
- *the $A_{tt}$ be complete fixed recourse matrices $\forall t$.*

With $\mathscr{H}(n) = \{\ell_1 = 1, \cdots, \ell_{t_n - 1}, \ell_{t_n} = n\}$ the history of node $n$ as before, the assumed stochastic independence of $\xi_2, \cdots, \xi_T$ implies the distribution (3.125) to be modified to

$$\left.\begin{aligned} \widehat{\zeta}^n &= \mathbb{E}\left[\zeta_{t_n} \mid \zeta_{t_n} \in \prod_{\nu=2}^{t_n} \Xi_{\ell_\nu}^\nu\right] \\ &= \mathbb{E}\begin{bmatrix} \xi_2 & \mid \xi_\nu \in \Xi_{\ell_\nu}^\nu,\ \nu = 2, \cdots, t_n \\ \vdots & \vdots \\ \xi_{t_n} & \mid \xi_\nu \in \Xi_{\ell_\nu}^\nu,\ \nu = 2, \cdots, t_n \end{bmatrix} \\ &= \begin{pmatrix} \mathbb{E}[\xi_2 \mid \xi_2 \in \Xi_{\ell_2}^2] \\ \vdots \\ \mathbb{E}[\xi_{t_n} \mid \xi_{t_n} \in \Xi_{\ell_{t_n}}^{t_n}] \end{pmatrix} \\ p_n &= \mathbb{P}_{\zeta_{t_n}}\left(\left\{\zeta_{t_n} \in \prod_{\nu=2}^{t_n} \Xi_{\ell_\nu}^\nu\right\}\right) = \prod_{\nu=2}^{t_n} \mathbb{P}_{\xi_\nu}(\Xi_{\ell_\nu}^\nu). \end{aligned}\right\} \tag{3.126}$$

Hence we replace problem (3.123) by the fully aggregated problem

$$\left.\begin{aligned} \min \sum_{m \in \mathscr{N}} p_m c_{t_m}^{\mathrm{T}} x_m & \\ \sum_{m \in \mathscr{H}(n)} A_{t_n t_m} x_m = b_{t_n}(\widehat{\zeta}^n) & \quad \forall n \in \mathscr{N} \\ x_m \geq 0 & \quad \forall m \in \mathscr{N} \end{aligned}\right\} \tag{3.127}$$

using the distribution (3.126). Then we get

**Lemma 3.17.** *Let problem (3.123) satisfy Assumption 3.5.. Then for any subfiltration $\widehat{\mathscr{F}}$ constructed as above, the optimal value of the aggregated problem (3.127) is a lower bound of the optimum in (3.123).*

*Proof:* It is well known that problem (3.123) can be formulated as a recursive sequence of optimization problems (see Olsen [247] and Rockafellar–Wets [286]). For this purpose we use the following notation:

$z_t := \{x_1, \cdots, x_t\}$ for the sequence of decision vectors up to stage $t$;
$\zeta_t := (\xi_2, \cdots, \xi_t)$ for the state variable at stage $t$, as before;
$\widehat{\zeta}_t$ for any realization of $\zeta_t$;
$\Xi^t_n \subseteq \Xi^t$ for node $n$ in stage $t$ due to (3.124), and $\bar{\xi}^n_t := \mathbb{E}[\xi_t \mid \xi_t \in \Xi^t_n]$.

Now the above mentioned recursion may be formulated as follows:

Let $\Phi_{T+1}(z_T; \widehat{\zeta}_T) \equiv 0 \ \forall z_T, \widehat{\zeta}_T$. Determine iteratively for $t = T, T-1, \cdots, 2$, and for all nodes $n$ in stage $t_n = t$, using the assumed stagewise independence by applying Fubini's theorem (see Halmos [131]),

$$\left.\begin{array}{rcl}
r_t(z_{t-1}; \widehat{\zeta}_t) & := & \min\limits_{x_t} \{c_t^{\mathrm{T}} x_t + \Phi_{t+1}(z_t; \widehat{\zeta}_t)\} \\
 & & \text{s.t. } A_{tt} x_t = b_t(\widehat{\zeta}_t) - \sum\limits_{\tau=1}^{t-1} A_{t\tau} x_\tau, \ \ x_t \geq 0 \\
\Phi_t(z_{t-1}; \widehat{\zeta}_{t-1}) & := & \mathbb{E}[r_t(z_{t-1}; \zeta_t) \mid \zeta_{t-1} = \widehat{\zeta}_{t-1}] \\
 & = & \mathbb{E}_{\xi_t}[r_t(z_{t-1}; \widehat{\zeta}_{t-1}, \xi_t)] \\
 & = & \sum\limits_{\nu \in C_{h_n}} \mathbb{P}_{\xi_t}(\Xi^t_\nu) \mathbb{E}_{\xi_t}[r(z_{t-1}; \widehat{\zeta}_{t-1}, \xi_t) \mid \xi_t \in \Xi^t_\nu],
\end{array}\right\} \quad (3.128)$$

which finally yields

$$\begin{array}{rcl}
r_1 & = & \min\limits_{x_1} \{c_1^{\mathrm{T}} x_1 + \Phi_2(x_1; \widehat{\zeta}_1)\} \\
 & & \text{s.t. } A_{11} x_1 = b_1(\widehat{\zeta}_1) \equiv b_1, \ x_1 \geq 0,
\end{array}$$

the optimal value of (3.123), with $\widehat{\zeta}_1$ being the realization of $\xi_1 \equiv$ const due to the fact that in the first stage there is only one (deterministic) state. The notation "$\mathbb{E}_{\xi_t}$" just indicates that the integral is taken with respect to $\mathbb{P}_{\xi_t}$ only.

If $\Phi_{t+1}(z_t, \widehat{\zeta}_t)$ is jointly convex in $(z_t, \widehat{\zeta}_t)$, as is trivially true for $\Phi_{T+1}$, then it follows immediately, that

$$\begin{array}{rcl}
r_t(z_{t-1}; \widehat{\zeta}_t) & = & \min\limits_{x_t} \{c_t^{\mathrm{T}} x_t + \Phi_{t+1}(z_t; \widehat{\zeta}_t)\} \\
 & & \text{s.t. } A_{tt} x_t = b_t(\widehat{\zeta}_t) - \sum\limits_{\tau=1}^{t-1} A_{t\tau} x_\tau, \ \ x_t \geq 0
\end{array}$$

is jointly convex in $(z_{t-1};\widehat{\zeta}_t)$ (recall that $b_t(\widehat{\zeta}_t)$ is linear affine in $\widehat{\zeta}_t$). Thus, from (3.128) follows that

$$\Phi_t(z_{t-1};\widehat{\zeta}_{t-1}) = \mathbb{E}_{\xi_t}[r_t(z_{t-1};\widehat{\zeta}_{t-1},\xi_t)]$$

is jointly convex in $(z_{t-1};\widehat{\zeta}_{t-1})$ as well. Hence, by Jensen's inequality holds

$$r_t(z_{t-1};\widehat{\zeta}_{t-1},\mathbb{E}[\xi_t]) \leq \mathbb{E}_{\xi_t}[r_t(z_{t-1};\widehat{\zeta}_{t-1},\xi_t)] = \Phi_t(z_{t-1};\widehat{\zeta}_{t-1})\,. \tag{3.129}$$

In analogy to (3.128), for the discretized problem (3.127) with $\Psi_{T+1} \equiv 0$ we define for $t = T, T-1, \cdots, 2$, and for all nodes $n$ in stage $t_n = t$, the recursion

$$\left.\begin{aligned} q_t(z_{t-1};\widehat{\zeta}_{t-1},\bar{\xi}_t^n) &:= \min_{x_t}\{c_t^{\mathrm{T}}x_t + \Psi_{t+1}(z_t;\widehat{\zeta}_{t-1},\bar{\xi}_t^n)\} \\ &\quad \text{s.t. } A_{tt}x_t = b_t(\widehat{\zeta}_{t-1},\bar{\xi}_t^n) - \sum_{\tau=1}^{t-1} A_{t\tau}x_\tau,\ \ x_t \geq 0 \\ \Psi_t(z_{t-1};\widehat{\zeta}_{t-1}) &:= \sum_{\nu\in C_{h_n}} \mathbb{P}_{\xi_t}(\Xi_\nu^t)q_t(z_{t-1};\widehat{\zeta}_{t-1},\bar{\xi}_t^\nu)\,, \end{aligned}\right\} \tag{3.130}$$

we'll get

$$\begin{aligned} q_1 &:= \min_{x_1}\{c_1^{\mathrm{T}}x_1 + \Psi_2(x_1;\widehat{\zeta}_1)\} \\ &\quad \text{s.t. } A_{11}x_1 = b_1(\widehat{\zeta}_1) \equiv b_1,\ x_1 \geq 0 \end{aligned}$$

as the optimal value of (3.127).

Provided that $\Psi_{t+1}(z_t;\widehat{\zeta}_t) \leq \Phi_{t+1}(z_t;\widehat{\zeta}_t)$, as it is obviously the case for $t = T$, we conclude from (3.128) and (3.130), using Jensen's inequality (3.129) (for conditional expectations), that

$$\begin{aligned} q_t(z_{t-1};\widehat{\zeta}_{t-1},\bar{\xi}_t^n) &\leq r_t(z_{t-1};\widehat{\zeta}_{t-1},\bar{\xi}_t^n) \\ &\leq \mathbb{E}_{\xi_t}[r_t(z_{t-1};\widehat{\zeta}_{t-1},\xi_t) \mid \xi_t \in \Xi_n^t] \end{aligned}$$

and hence

$$\begin{aligned} \Psi_t(z_{t-1};\widehat{\zeta}_{t-1}) &:= \sum_{\nu\in C_{h_n}} \mathbb{P}_{\xi_t}(\Xi_\nu^t)q_t(z_{t-1};\widehat{\zeta}_{t-1},\bar{\xi}_t^\nu) \\ &\leq \sum_{\nu\in C_{h_n}} \mathbb{P}_{\xi_t}(\Xi_\nu^t)\mathbb{E}_{\xi_t}[r_t(z_{t-1};\widehat{\zeta}_{t-1},\xi_t) \mid \xi_t \in \Xi_\nu^t] \\ &= \Phi_t(z_{t-1};\widehat{\zeta}_{t-1}), \end{aligned}$$

such that finally

$$q_1 := \min_{x_1\in\mathscr{B}}\{c_1^{\mathrm{T}}x_1 + \Psi_2(x_1;\widehat{\zeta}_1)\} \leq \min_{x_1\in\mathscr{B}}\{c_1^{\mathrm{T}}x_1 + \Phi_2(x_1;\widehat{\zeta}_1)\} =: r_1$$

with $\mathscr{B} := \{x_1 \mid A_{11}x_1 = b_1(\widehat{\zeta}_1) \equiv b_1,\ x_1 \geq 0\}$. □

As seen above, with Assumption 3.5., and observing Assumption 3.4. when generating a finite subfiltration $\widehat{\mathscr{F}}$ and the corresponding scenario tree for problem (3.123), as described on page 272, we get the fully aggregated problem (see (3.127))

$$\left.\begin{aligned} \min \sum_{m \in \mathscr{N}} p_m c_{t_m}^{\mathrm{T}} x_m & \\ \sum_{m \in \mathscr{H}(n)} A_{t_n t_m} x_m &= b_n \ \forall n \in \mathscr{N} \\ x_m &\geq 0 \ \ \forall m \in \mathscr{N} \end{aligned}\right\} \tag{3.131}$$

with $b_n = b_{t_n}(\widehat{\zeta}^n)$ and $p_n > 0\ \forall n \in \mathscr{N}$.

As the dual LP of (3.131) we have

$$\left.\begin{aligned} \max \sum_{n \in \mathscr{N}} b_n^{\mathrm{T}} u_n & \\ \sum_{n \in \mathscr{G}(m)} A_{t_n t_m}^{\mathrm{T}} u_n &\leq p_m c_{t_m} \ \forall m \in \mathscr{N}\,. \end{aligned}\right\} \tag{3.132}$$

With the substitution $u_n = p_n \pi_n$ (3.132) is equivalent to

$$\left.\begin{aligned} \max \sum_{n \in \mathscr{N}} p_n b_n^{\mathrm{T}} \pi_n & \\ \sum_{n \in \mathscr{G}(m)} q_{m \to n} A_{t_n t_m}^{\mathrm{T}} \pi_n &\leq c_{t_m} \ \forall m \in \mathscr{N} \end{aligned}\right\} \tag{3.133}$$

with $q_{m \to n}$ the conditional probability to reach node $n$ given node $m$.

For $\{\hat{x}_m, \hat{\pi}_n\}$ to be a primal-dual pair of optimal solutions, according to Chapter 1, Prop. 1.12., the complementarity conditions

$$(c_{t_m} - \sum_{n \in \mathscr{G}(m)} q_{m \to n} A_{t_n t_m}^{\mathrm{T}} \hat{\pi}_n)^{\mathrm{T}} \hat{x}_m = 0 \ \ \forall m \in \mathscr{N} \tag{3.134}$$

have to hold (with $q_{m \to m} = 1$).

### Discretization under special assumptions

Under Assumption 3.5. on problem (3.123) and Assumption 3.4. on the discretized distributions (implying positive probabilities for all scenarios generated) we shall discuss now, how a successive refinement of the partitions and hence a correspondingly growing scenario tree can be designed, such that the approximation of (3.123) by the generated problem (3.131) is improved.

To begin with, let $\widehat{\mathscr{F}}$ be the coarse subfiltration with each $\widehat{\mathscr{F}_t}$ being generated by the elementary events $\{\xi_\tau^{-1}[\Xi^\tau], \emptyset \mid \tau = 1, \cdots, t\}$ i.e. by $\{\Omega, \emptyset\}$. Then for node $n = t$ holds $t_n = n = t$ and $\Xi_n^{t_n} = \Xi^t$, such that by (3.126) follows $\widehat{\zeta}^n = \widehat{\zeta}^t = \mathbb{E}[\zeta_t]$, yielding the aggregated problem

$$\left.\begin{array}{r} \min \sum_{t=1}^{T} c_t x_t \\ \sum_{\tau=1}^{t} A_{t\tau} x_\tau = b_t(\widehat{\zeta}^t) \ \ \forall t \\ x_t \geq 0 \ \ \forall t\,. \end{array}\right\} \tag{3.135}$$

The corresponding basic scenario tree is shown in Fig. 3.8.


**Fig. 3.8** Basic scenario tree.

In the coarse subfiltration, $\widehat{\mathscr{F}_t}$ was generated by $\{\Omega, \emptyset\}$ $\forall t \in \{1, \cdots, T\}$. Let this subfiltration be refined into $\widetilde{\mathscr{F}}$ by partitioning $\Xi^t$ for a particular $t > 1$ into two subintervals $\Xi_1^t, \Xi_2^t$ (whereas in all other stages the trivial partitions $\{\Xi^s,\ s \neq t\}$remain unchanged). Then it follows that

$$\widetilde{\mathscr{F}_s} \text{ is generated by } \begin{cases} \{\Omega, \emptyset\} & \text{for } s < t \\ \{\Omega, \xi_t^{-1}[\Xi_1^t], \xi_t^{-1}[\Xi_2^t], \emptyset\} & \text{for } s = t \\ \{\Omega, \xi_t^{-1}[\Xi_1^t], \xi_t^{-1}[\Xi_2^t], \xi_s^{-1}[\Xi^s], \emptyset\} & \text{for } s > t\,. \end{cases}$$

The modification of the scenario tree, corresponding to splitting node $n = t$, is shown in Fig. 3.9.

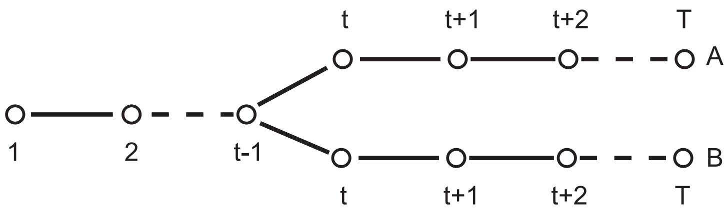


**Fig. 3.9** Basic scenario tree: First split.

Obviously we have now two branches from stage $t$ onwards, corresponding to the subintervals $\Xi_1^t$ and $\Xi_2^t$ of the partition of $\Xi^t$. Denoting the nodes of the two scenarios as $(t, A), t = 1, \cdots, T$, and $(t, B), t = 1, \cdots, T$, the respective components

$\widehat{\zeta}_\tau^{(s,A)}$ of $\widehat{\zeta}^{(s,A)}$, $s = 2, \cdots, T$, are, due to (3.126), determined as

$$\widehat{\zeta}_\tau^{(s,A)} = \begin{cases} \mathbb{E}[\xi_\tau] & \text{for } \tau \neq t \\ \mathbb{E}[\xi_t \mid \xi_t \in \Xi_1^t] & \text{for } \tau = t\,, \end{cases}$$

and analogously for $\widehat{\zeta}^{(s,B)}$, $s = 2, \cdots, T$, follows

$$\widehat{\zeta}_\tau^{(s,B)} = \begin{cases} \mathbb{E}[\xi_\tau] & \text{for } \tau \neq t \\ \mathbb{E}[\xi_t \mid \xi_t \in \Xi_2^t] & \text{for } \tau = t\,. \end{cases}$$

The corresponding node probabilities are

$$p_{(s,A)} = \begin{cases} 1 & \text{if } s < t \\ \mathbb{P}_{\xi_t}(\Xi_1^t) & \text{if } s \geq t \end{cases} \text{ and } p_{(s,B)} = \begin{cases} 1 & \text{if } s < t \\ \mathbb{P}_{\xi_t}(\Xi_2^t) & \text{if } s \geq t\,. \end{cases}$$

Hence the new aggregated problem is

$$\left.\begin{aligned} \min\left\{ \sum_{\tau=1}^{t-1} c_\tau^{\mathrm{T}} x_{(\tau,A)} + \sum_{\tau=t}^{T} c_\tau^{\mathrm{T}} \left[ p_{(\tau,A)} x_{(\tau,A)} + p_{(\tau,B)} x_{(\tau,B)} \right] \right\} & \\ \sum_{\tau=1}^{s} A_{s\tau} x_{(\tau,A)} = b_s(\widehat{\zeta}^{(s,A)}) \ \ \forall s & \\ \sum_{\tau=1}^{s} A_{s\tau} x_{(\tau,B)} = b_s(\widehat{\zeta}^{(s,B)}) \ \ \forall s \geq t & \\ x_{(s,B)} = x_{(s,A)} \ \ \forall s < t & \\ x_{(s,A)}, x_{(s,B)} \geq 0 \ \ \forall s\,. & \end{aligned}\right\} \qquad (3.136)$$

Assume now that $\mathscr{T} = (\mathscr{N}, \mathscr{A})$ is the scenario tree associated with problem (3.131). To split in this tree some node $i > 1$ into the nodes $i_1$ and $i_2$, or equivalently to subdivide the corresponding $\Xi_i^{t_i} \subseteq \Xi^{t_i}$ into two subintervals $\Xi_{i_1}^{t_i}$ and $\Xi_{i_2}^{t_i}$ (observing Assumption 3.4.), we have to run the following *node splitting procedure*:

Cut and paste

**S1** Partition $\Xi_i^{t_i}$ into $\Xi_{i_1}^{t_i}$ and $\Xi_{i_2}^{t_i}$; compute
$\widetilde{p}_{i_\nu} = \mathbb{P}_{\xi_{t_i}}(\Xi_{i_\nu}^{t_i}),\ \nu = 1, 2,$
$r_\nu = \dfrac{\widetilde{p}_{i_\nu}}{\widehat{p}_i},\ \nu = 1, 2,$ with $\widehat{p}_i = \mathbb{P}_{\xi_{t_i}}(\Xi_i^{t_i}),$
$b_{i_\nu} = \mathbb{E}_{\xi_{t_i}}[\widetilde{b}_i(\xi_{t_i}) \mid \xi_{t_i} \in \Xi_{i_\nu}^{t_i}],\ \nu = 1, 2,$ with $\widetilde{b}_i(\xi_{t_i}) := b_{t_i}(\widehat{\zeta}^{h_i}, \xi_{t_i}),$
such that $r_1 + r_2 = 1$ and $r_1 b_{i_1} + r_2 b_{i_2} = b_i\,.$

**S2** Let $\mathscr{T}_1 = (\mathscr{N}_1, \mathscr{A}_1)$ with $\mathscr{N}_1 \subset \mathscr{N}$, $\mathscr{A}_1 \subset \mathscr{A}$ be the maximal subtree of $\mathscr{T} = (\mathscr{N}, \mathscr{A})$ rooted at node $i \in \mathscr{N}$.
Let $\mathscr{T}_2 = (\mathscr{N}_2, \mathscr{A}_2)$ be a copy of $\mathscr{T}_1$, with its root denoted as $j \notin \mathscr{N}$ and all other node labels modified such that $\mathscr{N}_2 \cap \mathscr{N} = \emptyset$, $\mathscr{A}_2 \cap \mathscr{A} = \emptyset$.

Assign to the nodes of $\mathcal{T}_2$ the same quantities as associated with the corresponding nodes of $\mathcal{T}_1$.

**S3** With $\mathcal{H}(i)$ the history of node $i$ in $\mathcal{T}$, and $\widetilde{\mathcal{H}}(n)$ the history within $\mathcal{T}_\nu$ for $n \in \mathcal{N}_\nu$, $\nu = 1,2$ respectively, update the values of the subtrees $\mathcal{T}_1$ and $\mathcal{T}_2$ as follows:

$\mathcal{T}_1$ : Set $b_i^{(1)} := b_{i_1}$, and for $n \in \mathcal{G}(i) \setminus \{i\}$, the future of $i$ in $\mathcal{T}_1$, let $b_n^{(1)} := b_{t_n}(\widehat{\zeta}^n)$, with $\widehat{\zeta}^n$ computed according to (3.126), with the history of $n$ being composed as $\{\mathcal{H}(h_i), i, \widetilde{\mathcal{H}}(n)\}$;
multiply the node probabilities by $r_1$.

$\mathcal{T}_2$ : Set $b_j^{(2)} = b_{i_2}$, and for $m \in \mathcal{G}(j) \setminus \{j\}$, the future of $j$ in $\mathcal{T}_2$, let $b_m^{(2)} := b_{t_m}(\widehat{\zeta}^m)$, with $\widehat{\zeta}^m$ computed according to (3.126), with the history of $m$ being composed as $\{\mathcal{H}(h_i), j, \widetilde{\mathcal{H}}(m)\}$, implying that $b_m^{(2)}$ equals the right–hand–side for the corresponding node in $\mathcal{N}_1$;
multiply the node probabilities by $r_2$.
(Observe that $\mathcal{H}(h_i) = \mathcal{H}(h_j)$ will be enforced in step **S4**.)

**S4** Introduce a new edge from the parent node $h_i$ of $i$ to the node $j$, the root of $\mathcal{T}_2$, thus pasting $\mathcal{T}_2$ to $\mathcal{T}$ and yielding the new tree.

$$\begin{aligned} \mathcal{T}^+ &= (\mathcal{N}^+, \mathcal{A}^+), \text{ with} \\ \mathcal{N}^+ &= \mathcal{N} \cup \mathcal{N}_2 \qquad \text{and} \\ \mathcal{A}^+ &= \mathcal{A} \cup \mathcal{A}_2 \cup \{(h_i, j)\}. \end{aligned}$$

In Fig. 3.10 one cycle of this procedure is illustrated.

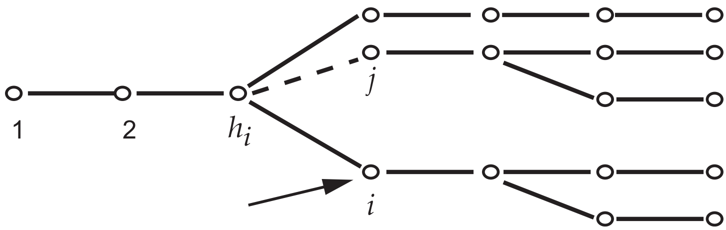


**Fig. 3.10** Cut and paste.

It is easy to see that with the above procedure of cut and paste the optimal values of the related primal LP's are non-decreasing.

**Proposition 3.8.** *With $V$ being the optimal value of the fully aggregated problem (3.131) corresponding to the scenario tree $\mathcal{T}$, and $V^+$ being the optimal value for the corresponding LP on $\mathcal{T}^+$ as generated by cut and paste, it follows that $V^+ \geq V$.*

*Proof:* Let $\{u_n,\ n \in \mathcal{N}\}$ be a solution of the dual program (3.132) associated with $\mathcal{T}$. To each node $n \in \mathcal{N}_2$ assign the vector $u_n$ as determined for the corresponding node $n \in \mathcal{N}_1$.

Now define for $n \in \mathcal{N}^+$, with $r_\nu$ from step **S1**,

$$\widetilde{u}_n := \begin{cases} r_1 u_n & \text{if } n \in \mathcal{N}_1 \\ r_2 u_n & \text{if } n \in \mathcal{N}_2 \\ u_n & \text{else.} \end{cases}$$

In order to show that $\{\widetilde{u}_n,\ n \in \mathcal{N}^+\}$ is a feasible solution to the dual program (3.132) associated with $\mathcal{T}^+$, we have to distinguish the following cases:

1) $\underline{m \in \mathcal{N}_1}$

$$\sum_{n\in\mathcal{G}(m)} A^{\mathrm{T}}_{t_n t_m} \widetilde{u}_n = r_1 \left( \sum_{n\in\mathcal{G}(m)} A^{\mathrm{T}}_{t_n t_m} u_n \right) \\ \leq r_1 p_m c_{t_m} = \widetilde{p}_m c_{t_m}$$

with $\widetilde{p}_m$ as defined in step **S3** for $m \in \mathcal{N}_1$.

2) $\underline{m \in \mathcal{N}_2}$

The analogous argument holds, with $r_2$ instead of $r_1$ .

3) $\underline{m \in \mathcal{N}^+ \setminus (\mathcal{N}_1 \cup \mathcal{N}_2) =: \Delta\mathcal{N}}$

$$\begin{aligned} &\sum_{n\in\mathcal{G}(m)} A^{\mathrm{T}}_{t_n t_m} \widetilde{u}_n = \\ &= \sum_{n\in\mathcal{G}(m)\cap\Delta\mathcal{N}} A^{\mathrm{T}}_{t_n t_m} u_n + \sum_{n\in\mathcal{G}(m)\cap\mathcal{N}_1} (r_1 + r_2) A^{\mathrm{T}}_{t_n t_m} u_n \\ &\leq p_m c_{t_m} . \end{aligned}$$

Hence, $\{\widetilde{u}_n,\ n \in \mathcal{N}^+\}$ is feasible for the dual program (3.132) corresponding to $\mathcal{T}^+$ and, with the right–hand–sides $\widetilde{b}_n$ updated according to step **S3**, yields the objective value

$$\begin{aligned} \sum_{n\in\mathcal{N}^+} \widetilde{b}^{\mathrm{T}}_n \widetilde{u}_n &= \sum_{n\in\Delta\mathcal{N}} b^{\mathrm{T}}_n u_n + \sum_{m\in\mathcal{N}_1} (r_1 b^{(1)}_m + r_2 b^{(2)}_m)^{\mathrm{T}} u_n \\ &= \sum_{n\in\mathcal{N}} b^{\mathrm{T}}_n u_n . \end{aligned}$$

This shows that the objective of the feasible solution $\{\widetilde{u}_n,\ n \in \mathcal{N}^+\}$ for $\mathcal{T}^+$ coincides with the optimal value for $\mathcal{T}$, such that $V^+ \geq V$ obviously has to hold. □

**Corollary 3.4.** *Let $\widehat{V}$ be the optimal value of problem (3.123). If Assumption 3.5. is satisfied, then each method, splitting succesively any nodes (except the root) in the scenario tree according to the cut and paste procedure, converges to a value $V^\star \leq \widehat{V}$.*

*Proof:* Under the given assumptions, the optimal objective values of the aggregated problems are

– monotonically nondecreasing according to Prop. 3.8., and

– they are lower bounds of the optimal value of (3.123) due to Lemma 3.17. □

Although this cut and paste procedure seems to have a promising behaviour, we are still left with two open questions:

1) Is there any criterion (even a heuristic one, maybe) for deciding on the next node to be split?
2) Given this criterion, may it happen that for the limit $V^\star$ in Corollary 3.4. holds $V^\star < \widehat{V}$?

As to the first question, for a fixed node $n > 1$ let $\{\hat{x}_m \mid m \in \mathscr{H}(n) \setminus \{n\}\}$ and $\{\hat{\pi}_m \mid m \in \mathscr{G}(n)\}$ be parts of solutions of (3.131) and (3.133), respectively, and consider the LP

$$\left.\begin{aligned} \varphi_n(b_n) := \min(c_{t_n} - \sum_{m\in\mathscr{G}(n)} q_{n\to m} A^{\mathrm{T}}_{t_m t_n} \hat{\pi}_m)^{\mathrm{T}} x_n \\ A_{t_n t_n} x_n = b_n - \sum_{m\in\mathscr{H}(n)\setminus\{n\}} A_{t_n t_m} \hat{x}_m \\ x_n \geq 0\,. \end{aligned}\right\} \tag{3.137}$$

Since $\{\hat{x}_k;\ k \in \mathscr{N}\}$ solves (3.131), in particular $\hat{x}_n$ is feasible in (3.137). Furthermore, the $\{\hat{\pi}_\ell;\ \ell \in \mathscr{N}\}$ being optimal in (3.133) and $\hat{x}_n \geq 0$ due to (3.137), we conclude, observing (3.134), that

$$0 \leq (c_{t_n} - \sum_{m\in\mathscr{G}(n)} q_{n\to m} A^{\mathrm{T}}_{t_m t_n} \hat{\pi}_m)^{\mathrm{T}} \hat{x}_n = 0\,,$$

showing that $\hat{x}_n$ with the optimal value $\varphi_n(b_n) = 0$ solves the LP (3.137). Using (3.126) we have that $\widehat{\zeta}^n = (\widehat{\zeta}^{h_n}, \mathbb{E}[\xi_{t_n} \mid \xi_{t_n} \in \Xi_n^{t_n}])$. Replacing $b_n = b_{t_n}(\widehat{\zeta}^n)$ by the random $\widetilde{b}_n(\xi_{t_n}) := b_{t_n}(\widehat{\zeta}^{h_n}, \xi_{t_n})$, it is obvious that the optimal value

$$\left.\begin{aligned} \varphi_n(\widetilde{b}_n(\xi_{t_n})) := \min(c_{t_n} - \sum_{m\in\mathscr{G}(n)} q_{n\to m} A^{\mathrm{T}}_{t_m t_n} \hat{\pi}_m)^{\mathrm{T}} x_n \\ A_{t_n t_n} x_n = \widetilde{b}_n(\xi_{t_n}) - \sum_{m\in\mathscr{H}(n)\setminus\{n\}} A_{t_n t_m} \hat{x}_m \\ x_n \geq 0\,. \end{aligned}\right\} \tag{3.138}$$

is a convex function in $\xi_{t_n}$, such that due to Jensen

$$\begin{aligned} \mathbb{E}[\varphi_n(\widetilde{b}_n(\xi_{t_n})) \mid \xi_{t_n} \in \Xi_n^{t_n}] &\geq \varphi_n(\widetilde{b}_n(\mathbb{E}[\xi_{t_n} \mid \xi_{t_n} \in \Xi_n^{t_n}]) \\ &= \varphi_n(b_{t_n}(\widehat{\zeta}^{h_n}, \mathbb{E}[\xi_{t_n} \mid \xi_{t_n} \in \Xi_n^{t_n}]) \\ &= \varphi_n(b_{t_n}(\widehat{\zeta}^n)) \\ &= \varphi_n(b_n) = 0\,, \end{aligned}$$

and we have the lower bound $l_n = 0$ for $\mathbb{E}[\varphi_n(\widetilde{b}_n(\xi_{t_n})) \mid \xi_{t_n} \in \Xi_n^{t_n}]$. On the other hand, according to Lemma 3.7 (on page 213), we can determine the E–M upper bound $u_n$ for $\mathbb{E}[\varphi_n(\widetilde{b}_n(\xi_{t_n})) \mid \xi_{t_n} \in \Xi_n^{t_n}]$. If, with some prescribed tolerance $\Delta > 0$, the *splitting criterion*

$$u_n - l_n > \Delta \tag{3.139}$$

is satisfied, we may decide to split node $n$ as described in the cut and paste procedure, in order to increase the lower bound and thereby to improve the approximative solution. Observe however, that this criterion ($u_n - l_n > \Delta$) to increase the lower bound and thereby to improve the solution in a particular node, is based on a heuristic argument. But it is one positive answer to the first question, at least. Moreover, test runs with this criterion did work out surprisingly well.

To come to the second question, consider the following example:

**Example 3.4.** *Assume the following problem to be given:*

$$\begin{array}{lr} \min\{x_1 + x_2 + \mathbb{E}[y_1 + y_2 + z_1 + z_2]\} & \\ x_1 - x_2 & = 0 \\ x_1 + 2x_2 + 3y_1 - 3y_2 & = \xi_2 \\ x_1 + 3x_2 + \quad y_1 - y_2 + 4z_1 - 4z_2 & = \xi_3 \\ & x_i, y_i, z_i \geq 0\,, \end{array}$$

*where $\xi_2 \sim \mathscr{U}[0,6]$ and $\xi_3 \sim \mathscr{U}[1,1.5]$, with $\mathscr{U}$ being the uniform distribution. The fully aggregated problem with $\mathbb{E}[\xi_2] = 3$ and $\mathbb{E}[\xi_3] = 1.25$ as right–hand–sides is easily seen to have the optimal solution*

$$(\hat{x}_1, \hat{x}_2, \hat{y}_1, \hat{y}_2, \hat{z}_1, \hat{z}_2) = (0, 0, 1, 0, \frac{1}{16}, 0)$$

*with the optimal value*

$$V = \frac{17}{16}$$

*and the dual solution*

$$\hat{\pi}^{\mathrm{T}} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right).$$

*Considering problem (3.138) for $n = 2$, we find that $\varphi_2(\widetilde{b}_2(\xi_{t_2})) \equiv 0$ for $\xi_2 \in [0,6]$, i.e. $\varphi_2$ is linear on $\Xi^2$ implying that $u_2 - l_2 = 0$. Analogously $\varphi_3(\widetilde{b}_3(\xi_{t_3})) \equiv 0$ for $\xi_3 \in [1, 1.5]$ such that also $\varphi_3$ is linear on $\Xi^3$ and therefore $u_3 - l_3 = 0$. Hence the above splitting criterion (3.139) cannot be satisfied, and the procedure would stop with the above solution, with $V^\star = V$.*

*However, subdividing $\Xi^2 = [0,6]$ into the intervals $[0,3)$ and $[3,6]$ and solving the corresponding LP, would yield the optimal value*

$$V^+ = \frac{18}{16} > V\,,$$

*and the same result would be achieved with splitting, instead of $\Xi^2$, the interval $\Xi^3 = [1, 1.5]$ into $[1, 1.25)$ and $[1.25, 1.5]$.* □

Hence, in this example the procedure, using the above splitting criterion (3.139), had to be finished with $u_n - l_n = 0$ for all nodes $n > 1$, although there was a substantial difference $\widehat{V} - V^{\star} > 0$. This fact could (and can in general) only be discovered by analyzing (sub)sets of nodes simultaneously in detail. In other words: For the approach using the splitting criterion (3.139) so far there is not known any simple stopping rule stating the (near-)optimality of the present iterative solution for problem (3.123).