Total variation error bounds for convex approximations of two-stage mixed-integer recourse models
Romeijnders, Ward

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 5

Assessing the quality of convex approximations using sampling

Abstract. We consider two types of convex approximations of two-stage totally unimodular integer recourse models. Although worst-case error bounds are available for these approximations, their actual performance has not yet been investigated, mainly because this requires solving the original recourse model. In this chapter we assess the quality of the approximating solutions using Monte Carlo sampling, or more specifically, using the so-called multiple replications procedure. Based on numerical experiments for an integer newsvendor problem and a fleet allocation and routing problem, we conclude that the actual performance is much better than the error bounds suggest, especially if the variability of the random parameters in the model is medium to large. In case this variability is small, the performance of the approximations is not so good. However, these are precisely the cases for which sampling methods, with modest sample sizes, may perform best. In this sense, the convex approximations and sampling methods can be considered as complementary solution methods. Finally, for a fleet allocation and routing problem, we derive a new error bound dealing with deterministic second-stage side constraints and relatively complete recourse.

This chapter is submitted for publication as [58].
5.1 Introduction

We consider the stochastic optimization problem

\[ \eta^* := \min_x \left\{ \mathbb{E}_\omega[g(x, \omega)] : x \in X \right\}, \]  

where the uncertainty is explicitly modeled using the random vector \( \omega \) (with known distribution function \( F \)) and the objective is to find optimal here-and-now decisions \( x \in X \) to minimize the expected value function \( G(x) := \mathbb{E}_\omega[g(x, \omega)] \).

In its general form, model (5.1) can represent many stochastic programming problems (see, e.g., [6, 55, 76, 89]). However, throughout this chapter we restrict attention to two-stage integer recourse models, where \( X \subset \mathbb{R}^{n_1}_+ \), \( Y \subset \mathbb{R}^{n_2}_+ \), and \( g \) is defined for every \( x \in X \) and \( \omega \in \Omega \) as

\[ g(x, \omega) = cx + \min_y \left\{ q(\omega)y : Wy \geq \zeta(\omega) - T(\omega)x, \ y \in Y \cap \mathbb{Z}^{n_2} \right\}. \]  

Here, the cost vector \( q(\omega) \), right-hand side vector \( \zeta(\omega) \), and technology matrix \( T(\omega) \) are random and depend on the underlying random vector \( \omega \). We introduce the notation in (5.1) since several of our ideas, methods, and results also hold in this more general setting.

The function \( g \) in (5.2) is called an integer recourse function because of the integer-constrained recourse variables \( y \). Such decision variables arise often in practice to model indivisibilities or on-off decisions. With the corresponding model (5.1) in mind, these recourse variables \( y \) are determined after realization of the random vector \( \omega \), and can be used to compensate for infeasibilities of the underlying random goal constraints \( T(\omega)x \geq \zeta(\omega) \).

The integer recourse function \( g \) is clearly non-convex in \( x \) for every \( \omega \in \Omega \) because of the integer programming problem involved. As a consequence, the expected value function \( G \) is generally non-convex as well [56]; see [42] for exceptions in the simple integer recourse case. This lack of convexity is the main reason why integer recourse models are much harder to solve than their continuous counterparts. Indeed, for the latter type of problems, efficient algorithms such as the L-shaped method [87], regularized decomposition [66], and stochastic decomposition [32] are available that explicitly use convexity (see [94] for a recent numerical study comparing several algorithms).
A possible approach for solving integer recourse models is to replace $g$ in (5.1) by a function $\hat{g} : X \times \Omega \rightarrow \mathbb{R}$ that is convex in $x$ for every $\omega \in \Omega$. Then, the approximating model

$$\hat{\eta} := \min\limits_x \{ \hat{G}(x) : x \in X \} \quad (5.3)$$

with $\hat{G}(x) := \mathbb{E}_\omega[\hat{g}(x, \omega)]$, $x \in X$, can be solved using tools from convex optimization (yielding an approximating solution $\hat{x}$). For the special case of totally unimodular (TU) integer recourse models, i.e., for TU recourse matrices $W$, with $Y = \mathbb{R}^n_{\geq}$ and deterministic $q$ and $T$, such convex approximations have been developed [83, 62]. In fact, these references show that model (5.3) corresponding to these approximations can be represented as a continuous recourse model and can thus be solved efficiently using one of the methods mentioned above. The question that remains, and the main topic of this chapter, concerns the quality of the approximating solution, $\hat{x}$, of model (5.3).

In Chapters 3 and 4 we measure the performance of the convex approximations by upper bounds $U(G, \hat{G})$ on $\|G - \hat{G}\|_{\infty} := \sup\{ |G(x) - \hat{G}(x)| : x \in X \}$. Such upper bounds are useful since

$$|\hat{\eta} - \eta^*| \leq \|G - \hat{G}\|_{\infty} \leq U(G, \hat{G})$$

and

$$G(\hat{x}) - \eta^* \leq 2\|G - \hat{G}\|_{\infty} \leq 2U(G, \hat{G}). \quad (5.4)$$

See, e.g., Chapter 3 for a proof of these results for the TU integer recourse case. Numerical experiments in Section 3.5.2 for several (small) examples suggest that the second inequality in (5.4) is reasonably tight if the upper bounds $U(G, \hat{G})$ of Chapters 3 and 4 are used. The sharpness of the first inequality in (5.4), however, has not yet been investigated. The main difficulty of doing so, and in fact the motivation for deriving convex approximations of $g$, is that it is very hard to solve the original integer recourse model (5.1) to obtain $\eta^*$, especially for larger problem instances.

In this chapter we assess the quality of $\hat{x}$, and thus the sharpness of the first inequality in (5.4), using sampling. In particular, we will use the multiple replications procedure (MRP) developed in [51]. We carry out numerical experiments for
an integer newsvendor problem and for a fleet allocation and routing problem. These experiments show that the convex approximations are much better than their error bounds suggest, especially if the ‘variability’ of the random parameters in the models is medium to large. If this ‘variability’ is small then the performance of the convex approximations is not so good. However, these are precisely the cases in which sampling methods can work quite well with modest sample sizes, and in this sense we may view convex approximations and sampling methods as complementary approaches for approximately solving TU integer recourse models.

The remainder of this chapter is organized as follows. In Section 5.2 we discuss the literature on convex approximations for integer recourse models and the literature on sampling methods for assessing the quality of candidate solutions in stochastic programs. In Section 5.3 we show numerical experiments for an integer newsvendor problem and in Section 5.4 for a fleet allocation and routing problem. For this latter problem we have to extend the analysis of Chapter 4 to derive an error bound for the convex approximation to deal with deterministic second-stage side constraints and relatively complete recourse instead of complete recourse. Finally, Section 5.5 comprises a summary and conclusions.

5.2 Literature review

We review the literature on both solution methods for integer recourse models (Section 5.2.1) and sampling methods for assessing the quality of candidate solutions in stochastic programming problems (Section 5.2.2). Our focus is on convex approximations of integer recourse models, their error bounds, and the multiple replications procedure (MRP) to be used in Sections 5.3 and 5.4.

5.2.1 Solution methods for integer recourse models

During the last decades a variety of solution methods have been developed for integer recourse models, including the integer L-shaped method [46], dual decomposition [11], branch-and-bound [1], and disjunctive decomposition [73]. These solution methods typically combine ideas from deterministic integer programming and stochastic continuous programming, and are aimed at finding (near-)optimal solutions. This is the main reason why these methods, in general, have difficulties solving very large problem instances, motivating the development of convex approximations.
In the remainder of this section we restrict our attention to the literature on these convex approximations and their error bounds. For readers interested in other solution methods for integer recourse models we refer to the survey papers of Klein Haneveld and Van der Vlerk [43], Louveaux and Schultz [47], Schultz [70], and Sen [72].

**Convex approximations for integer recourse models**

Klein Haneveld et al. [42] were the first to develop a class of convex approximations for the special case of simple integer recourse models. These so-called $\alpha$-approximations were later extended by van der Vlerk to the cases of TU integer recourse [83] and simple mixed-integer recourse [85]. The recurring idea in these approximations (see also Section 3 of the survey paper [60]) is to simultaneously relax the integrality constraints in the model defining $g$ and perturb the distribution of the random right-hand side $\omega$. For $g$ defined in (5.2) with $Y = \mathbb{R}^n_+$, $\zeta(\omega) = \omega$, and deterministic $q$ and $T$, this, for every $\alpha \in \mathbb{R}^m$, yields

$$g_\alpha(x, \omega) := cx + \min_y \left\{ qy : Wy \geq \lceil \omega \rceil - Tx, \ y \in \mathbb{R}^n_+ \right\}, \quad x \in X, \ \omega \in \Omega,$$

where $\lceil \omega \rceil := \lceil \omega - \alpha \rceil + \alpha$ is a discrete random vector with support contained in $\alpha + \mathbb{Z}^m$. As already mentioned in the introduction, the resulting approximating problem with $g$ replaced by $g_\alpha$ in (5.1) corresponds to a (convex) continuous recourse model with a discrete distribution that, with existing algorithms, is more computationally tractable.

An alternative convex approximation that also can be represented as a continuous recourse model is the so-called shifted LP-relaxation approximation developed in Chapter 4, defined as

$$\hat{g}(x, \omega) = cx + \min_y \left\{ qy : Wy \geq \omega + 1/2e_m - Tx, \ y \in \mathbb{R}^n_+ \right\}, \quad x \in X, \ \omega \in \Omega,$$

where $e_m$ is the $m$-dimensional all-one vector. The error bound of this approximation (see Theorem 5.1 below) improves the bound of the $\alpha$-approximation by a factor 2. Moreover, in Chapter 4 we show that the bound is tight in a worst-case sense.
Error bounds for convex approximations of TU integer recourse models

Error bounds, i.e., upper bounds on \( \|G - G_\alpha\|_\infty \) and \( \|G - \hat{G}\|_\infty \) with \( G_\alpha(x) := \mathbb{E}_\omega[g_\alpha(x, \omega)] \) and \( \hat{G}(x) := \mathbb{E}_\omega[g(x, \omega)] \), \( x \in X \), are derived under several assumptions:

(A1) Complete recourse: \( g(x, \omega) < +\infty \) for every \( x \in \mathbb{R}^{n_1} \) and \( \omega \in \mathbb{R}^{m_1} \).

(A2) Sufficiently expensive (or dual feasible) recourse: \( \Lambda := \{ \lambda \in \mathbb{R}^m_+ : \lambda W \leq q \} \neq \emptyset \).

(A3) Finite expectations: \( \mathbb{E}_\omega[|\omega_i|] < +\infty \) for every \( i = 1, \ldots, m \).

In Section 5.4 we consider a problem where the complete recourse assumption is violated. Instead, a relaxation of this assumption holds:

(A1') Relatively complete recourse: \( g(x, \omega) < +\infty \) for every \( x \in X \) and \( \omega \in \Omega \).

As we show in Section 5.4 this has consequences for the type of convex approximation to use and its corresponding error bound.

Theorem 5.1 below shows error bounds for \( \alpha \)-approximations and the shifted LP-relaxation approximation. They correspond to upper bounds \( U(G, G_\alpha) \) and \( U(G, \hat{G}) \) that can appear on the right-hand side of (5.4), and in Sections 5.3 and 5.4 we compare them with the optimality gaps \( G(x_\alpha) - \eta^* \) and \( \hat{G}(x) - \eta^* \). A detailed proof of Theorem 5.1 is omitted here and can be found in Chapter 4. We do discuss the main line of the proof because it helps facilitate our proof in Section 5.4 of an error bound for a convex approximation of the fleet allocation and routing problem.

**Theorem 5.1.** Consider the totally unimodular integer recourse function

\[
g(x, \omega) := cx + \min_y \left\{ qy : Wy \geq \omega - Tx, y \in \mathbb{Z}^{n_2}_+ \right\}, \quad x \in X, \omega \in \Omega,
\]

where \( \omega \) is a continuous random vector with joint pdf \( f \) and with independently distributed components. Let \( g_\alpha \) and \( \hat{g} \) denote the \( \alpha \)-approximation and shifted LP-relaxation approximation defined in (5.5) and (5.6), respectively, with \( G_\alpha \) and \( \hat{G} \) denoting their expected value functions. Then, under assumptions (A1)–(A3) we have for every \( \alpha \in \mathbb{R}^m \),

\[
\|G - G_\alpha\|_\infty \leq \sum_{i=1}^m \lambda^*_i h(|\Delta| f_i) \quad \text{and} \quad \|G - \hat{G}\|_\infty \leq \frac{1}{2} \sum_{i=1}^m \lambda^*_i h(|\Delta| f_i),
\]
where $\lambda^*_i := \max\{\lambda_i : \lambda W \leq q, \lambda \in \mathbb{R}_+^m\}$, $|\Delta|f_i$ denotes the total variation of the $i$-th marginal density function, $f_i$, and $h : (0, \infty) \mapsto \mathbb{R}$ is defined as

$$h(t) = \begin{cases} 
\frac{t}{8}, & t \leq 4 \\
1 - \frac{2}{t}, & t \geq 4
\end{cases}$$

(5.7)

similar as in (3.31) and (4.23) in Chapters 3 and 4, respectively. □

**Remark 5.1.** The assumption in Theorem 5.1 that the components of $\omega$ are independently distributed is not necessary. Indeed, in Chapters 3 and 4 bounds for the dependent case are derived involving conditional density functions instead of marginal ones. However, since we do not consider the latter case in our numerical experiments in Sections 5.3 and 5.4, we present Theorem 5.1 in its current form for ease of exposition.

The error bounds in Theorem 5.1 are smaller if the total variations $|\Delta|f_i$ of the marginal densities $f_i$ are smaller. For example, for a normally distributed random vector, $\omega$, this implies that we expect the performance of the convex approximations to be better if the standard deviations are larger. We will confirm this conjecture by numerical experiments in Sections 5.3 and 5.4.

**Main line of the proof of Theorem 5.1**

Since the derivation of the error bounds in Theorem 5.1 is very similar for both the $\alpha$-approximation and the shifted LP-relaxation approximation, we only discuss the proof for the $\alpha$-approximation.

First, we derive a dual representation of the optimization problems in $g$ and $g_\alpha$. Since $W$ is TU, we have for every $x \in X$ and $\omega \in \Omega$,

$$g(x, \omega) = cx + \min_y \left\{ qy : W y \geq [\omega - Tx], y \in \mathbb{R}_+^{n_2} \right\}$$

(5.8)

$$= cx + \max_\lambda \left\{ \lambda [\omega - Tx] : \lambda \in \Lambda \right\},$$

(5.9)

where the second equality follows from strong LP duality and where the dual feasible region $\Lambda := \{ \lambda \in \mathbb{R}_+^m : \lambda W \leq q \}$ is non-empty and bounded by assumptions (A1) and (A2). Similarly, for the $\alpha$-approximation we have for every $\alpha \in \mathbb{R}^m$,

$$g_\alpha(x, \omega) = cx + \max_\lambda \left\{ \lambda ([\omega]_\alpha - Tx) : \lambda \in \Lambda \right\}, \quad x \in X, \omega \in \Omega.$$  

(5.10)
Suppose for the moment that the dual feasible region \( \Lambda \) contains only a single point. Then, for every fixed \( x \in X \), \( \omega \in \Omega \), and defining tender variables \( z := Tx \),

\[
g(x, \omega) - g_\alpha(x, \omega) = \lambda \left( [\omega - Tx] + Tx - [\omega]_\alpha \right) \\
= \sum_{i=1}^{m} \lambda_i \left( [\omega_i - z_i] + z_i - [\omega_i]_{\alpha_i} \right) \\
= \sum_{i=1}^{m} \lambda_i \left( [\omega_i]_{z_i} - [\omega_i]_{\alpha_i} \right).
\]

Thus, for fixed \( z \) and \( \alpha \) the difference \( g - g_\alpha \) can be decomposed componentwise in \( \omega_i \). Moreover, all properties of \( g - g_\alpha \) follow directly from those of the one-dimensional function \( \bar{\varphi}_{z_i, \alpha_i} \) given in Definition 5.1.

**Definition 5.1.** For every \( z_i \in \mathbb{R} \) and \( \alpha_i \in \mathbb{R} \) we define the function \( \bar{\varphi}_{z_i, \alpha_i} : \mathbb{R} \rightarrow \mathbb{R} \) as

\[
\bar{\varphi}_{z_i, \alpha_i}(t) = \left[ t \right]_{z_i} - \left[ t \right]_{\alpha_i} = \left( \left[ t - z_i \right] + z_i \right) - \left( \left[ t - \alpha_i \right] + \alpha_i \right), \quad t \in \mathbb{R}.
\]

Moreover, for every \( z_i \in \mathbb{R} \) we define \( \hat{\varphi}_{z_i} : \mathbb{R} \rightarrow \mathbb{R} \) as

\[
\hat{\varphi}_{z_i}(t) = \left[ t \right]_{z_i} - \left( t + \frac{1}{2} \right) = \left( \left[ t - z_i \right] + z_i \right) - \left( t + \frac{1}{2} \right), \quad t \in \mathbb{R}.
\]

The function \( \hat{\varphi}_{z_i} \) can be interpreted as the underlying difference function of the shifted LP-relaxation approximation, which we use in Section 5.4. Both functions \( \bar{\varphi}_{z_i, \alpha_i} \) and \( \hat{\varphi}_{z_i} \) are periodic in \( t \) with period \( p = 1 \) and mean value \( p^{-1} \int_0^p \bar{\varphi}_{z_i, \alpha_i}(t) dt = p^{-1} \int_0^p \hat{\varphi}_{z_i}(t) dt = 0 \). We use these properties to bound \( E_{\omega_i}[\bar{\varphi}_{z_i, \alpha_i}(\omega_i)] \), yielding an upper bound on \( ||G - G_\alpha||_{\infty} \) since

\[
G(x) - G_\alpha(x) = \sum_{i=1}^{m} \lambda_i E_{\omega_i} \left[ \bar{\varphi}_{z_i, \alpha_i}(\omega_i) \right].
\]

Surprisingly, in the general case (without the assumption that \( \Lambda \) is a singleton) the analysis above is still helpful since it turns out that we are allowed to ‘round up’ the \( \lambda \)'s to a single vector \( \lambda^* \) with \( \lambda \leq \lambda^* \) for every \( \lambda \in \Lambda \). Below we illustrate this idea by deriving an upper bound for \( G(x) - G_\alpha(x), \ x \in X \). A lower bound can be
Let $x \in X$ be given and let $\lambda(\omega) \in \Lambda$ denote maximizers in model (5.9) for each $\omega \in \Omega$. Since $\lambda(\omega)$ is feasible but not necessarily optimal in model (5.10), we have

$$g(x, \omega) - g_\alpha(x, \omega) \leq \sum_{i=1}^m \lambda_i(\omega) \left( \lfloor \omega_i \rfloor z_i - \lceil \omega_i \rceil \alpha_i \right) = \sum_{i=1}^m \lambda_i(\omega) \bar{\phi}_{z_i, \alpha_i}(\omega).$$

Moreover, it is not hard to show that $\lambda_i(\omega)$ is monotone non-decreasing in $\omega_i$ for every $\omega(i) \in \mathbb{R}^{m-1}$, where $\omega(i)$ denotes $\omega$ without its $i$-th component. This monotonicity property is one of the assumptions in Proposition 5.1, which is key to ‘round up’ $\lambda_i(\omega)$ to $\lambda^*_i$.

The proof of Proposition 5.1 can be found in Chapter 4. Observe that we use $\mathcal{F}$ to denote the set of probability density functions $f$ of bounded variation (i.e., $|\Delta f| < +\infty$).

**Proposition 5.1.** Let $\lambda : \mathbb{R} \mapsto \mathbb{R}$ be a real-valued monotone function such that $0 \leq \lambda(x) \leq \lambda^*$ for all $x \in \mathbb{R}$, and let $\varphi : \mathbb{R} \mapsto \mathbb{R}$ be a bounded periodic function with period $p$ and mean value $\nu := p^{-1} \int_0^p \varphi(x)dx = 0$. Then, for every continuous random variable $\omega$ with probability density function $f \in \mathcal{F}$,

$$-\lambda^* M(-\varphi, |\Delta f|) \leq \mathbb{E}_\omega[\lambda(\omega)\varphi(\omega)] \leq \lambda^* M(\varphi, |\Delta f|),$$

where for every $B \in \mathbb{R}$ with $B > 0$,

$$M(\varphi, B) := \sup_{f \in \mathcal{F}} \left\{ \mathbb{E}_\omega[\varphi(\omega)] : |\Delta f| \leq B \right\}. \quad (5.11)$$

Using Proposition 5.1 we are able to reduce the problem of finding an upper bound on $\|G - G_\alpha\|_\infty$ to the bound of (5.11) since for every $x \in X$ and $\alpha \in \mathbb{R}^m$,

$$G(x) - G_\alpha(x) \leq \sum_{i=1}^m \lambda_i^* M(\bar{\phi}_{z_i, \alpha_i}, |\Delta f_i|), \quad (5.12)$$

with $\lambda^*_i = \max\{\lambda_i : \lambda \in \Lambda\}$. It turns out that for periodically monotone functions $\varphi$ (including $\bar{\phi}_{z_i, \alpha_i}$ and $\hat{\phi}_{z_i}$) exact expressions of $M(\varphi, B)$ can be obtained; in all other cases an upper bound is available. Moreover, as shown in Examples 4.1 and 4.2 of...
Chapter 4 we have for every $z_i, \alpha_i \in \mathbb{R}$ and for every $B \in \mathbb{R}$ with $B > 0$ that
\[ M(\bar{\varphi}_{z_i, \alpha_i}, B) \leq h(B) \quad \text{and} \quad M(\hat{\varphi}_{z_i}, B) = M(-\hat{\varphi}_{z_i}, B) = \frac{1}{2} h(B) \quad (5.13) \]
with $h$ defined in (5.7). Combining (5.12) and the first inequality in (5.13) we obtain the error bound from Theorem 5.1. Moreover, observe that the difference of a factor 2 between $M(\bar{\varphi}_{z_i, \alpha_i}, B)$ and $M(\hat{\varphi}_{z_i}, B)$ in (5.13) causes the factor 2 difference between the error bounds of the $\alpha$-approximations and the shifted LP-relaxation approximation.

Since the error bounds in Theorem 5.1 are determined using worst-case analysis, among others in the form of (5.11), the question arises how sharp these error bounds actually are. As already mentioned in the introduction they are reasonably tight when compared with $\|G - G_\alpha\|_\infty$ and $\|G - \hat{G}\|_\infty$. In Sections 5.3 and 5.4, however, we will compare the error bounds with $G(x_\alpha) - \eta^*$ and $G(\hat{x}) - \eta^*$ and show that the quality of the convex approximations may in fact be much better than the error bounds of Theorem 5.1 suggest.

### 5.2.2 Assessing the quality of candidate solutions using sampling

In this section we review sampling methods for assessing the quality of candidate solutions, $x \in X$, for model (5.1). In particular, we discuss the multiple replications procedure (MRP) of [51]. This procedure is easy to implement and works under very general assumptions, which are satisfied by the integer recourse function $g$ defined in (5.2), at least if $g(x, \omega)$ has finite variance for all $x \in X$. On the other hand, we note that the single- and two-replications procedures in [3], for example, assume that $g(\cdot, \omega)$ is continuous for every $\omega \in \Omega$, which is not the case when integer decision variables are involved in the second stage.

The MRP can be applied to any candidate solution, $x \in X$, independent of the method by which $x$ is obtained. So, for the TU integer recourse models that we consider in Sections 5.3 and 5.4 we can use the MRP with $x := x_\alpha$ and $x := \hat{x}$, where $x_\alpha$ and $\hat{x}$ denote solutions of the $\alpha$-approximations and shifted LP-relaxation approximation, respectively.

Other sampling methods for assessing solution quality in stochastic programming problems include [27, 32, 33, 77]; see also the tutorial of Bayraksan and Morton [4].
Although descriptions of MRP can be found in this tutorial and in, e.g., [3], we discuss it here to set notation for what follows.

### Multiple replications procedure

We measure the quality of a candidate solution, \( x \in X \), of (5.1) by its optimality gap \( \theta(x) := G(x) - \eta^* \). This gap cannot be obtained by straightforward computation, since it is typically impossible to calculate \( \eta^* \) exactly and because evaluating \( G(x) \) can require computing a higher-dimensional integral. Nonetheless, the optimality gap may be estimated using sampling.

Let \( \omega^1, \ldots, \omega^n \) denote an i.i.d. sample from the distribution of \( \omega \). Then, \( n^{-1} \sum_{j=1}^{n} g(x, \omega^j) \) is a consistent estimator of \( G(x) \). We can estimate \( \eta^* \) by solving

\[
(SP_n) \quad \eta_n^* := \min_x \left\{ \frac{1}{n} \sum_{j=1}^{n} g(x, \omega^j) : x \in X \right\}.
\]  

Model (5.14) is of the same form as model (5.1), but will be computationally tractable if the sample size is small enough. The estimator \( \eta_n^* \) of \( \eta^* \) has a negative bias, i.e., \( \mathbb{E}[\eta_n^*] \leq \eta^* \); see [51]. In this way, \( \theta_n(x) := n^{-1} \sum_{j=1}^{n} g(x, \omega^j) - \eta_n^* \) will be a conservative estimate of the optimality gap \( \theta(x) \) in the sense that \( \mathbb{E}[\theta_n(x)] \geq \theta(x) \).

The distribution of \( \theta_n(x) \) may be asymptotically non-normal, making it more difficult to derive probabilistic statements on \( \theta_n(x) \). This issue is circumvented by replicating the procedure \( N_r \) times and applying the central limit theorem (CLT). A complete description of the MRP is given below.

**MRP:**

**Step 1:** For \( i = 1, \ldots, N_r \),

(i) Sample (i.i.d.) observations \( \omega^{i1}, \ldots, \omega^{in} \) from the distribution of \( \omega \).

(ii) Solve \( (SP_n) \) in (5.14) using the sample \( \omega^{i1}, \ldots, \omega^{in} \) of (i), yielding objective \( \eta_n^* \) and solution \( x_n^* \).

(iii) Calculate \( \theta_n^*(x) := n^{-1} \sum_{j=1}^{n} (g(x, \omega^{ij}) - g(x_n^*, \omega^{ij})) \).
Step 2: Calculate the gap estimate $\hat{\theta}_n(x)$ and sample variance $s^2_\theta(x)$ by

$$\hat{\theta}_n(x) := \frac{1}{N_r} \sum_{i=1}^{N_r} \theta_i^n(x) \quad \text{and} \quad s^2_\theta(x) = \frac{1}{N_r - 1} \sum_{i=1}^{N_r} (\theta_i^n(x) - \hat{\theta}_n(x))^2.$$ 

Step 3: Let $\epsilon_\theta := t_{N_r - 1, \gamma} \cdot s_\theta(x) / \sqrt{N_r}$, where $t_{N_r - 1, \gamma}$ denotes the $(1 - \gamma)$% quantile of the $t$ distribution with $N_r - 1$ degrees of freedom. Then, the one-sided $(1 - \gamma)$% confidence interval on $\theta(x) = G(x) - \eta^*$ is given by $[0, \hat{\theta}_n(x) + \epsilon_\theta]$. That is, if the CLT were to hold exactly for finite sample size, $N_r$, we would have

$$\mathbb{P}\left\{ G(x) - \eta^* \in [0, \hat{\theta}_n(x) + \epsilon_\theta] \right\} = 1 - \gamma.$$ 

Step 1(i) of the MRP need not use i.i.d. sample; only i.i.d. samples over replications $i$ are required. For example, throughout this chapter we use Latin hypercube sampling (LHS) in this step, which reduces variance and also often decreases the bias. This is important because the width of the confidence interval of $\theta(x)$ may be large since (i) $x$ is suboptimal, (ii) the negative bias of $\hat{\theta}_n(x)$ is large, or (iii) the sample variance, and thus $\epsilon_\theta$, is large. Using LHS we reduce the effect of (ii) and (iii) so that we can better assess the quality of the candidate solution $x$; see, e.g., [24].

Although the purpose of the MRP is to assess the quality of a candidate solution, $x \in X$, it also calculates potential candidate solutions $x_n^*$ in Step 1(ii), and can thus also be considered a sampling (solution) method. The candidate solutions will most likely be suboptimal, particularly when the sample size $n$ is small, but we can obtain the best among them using an out-of-sample evaluation or by averaging them [74] if $X$ is convex. In Sections 5.3 and 5.4 we compare the solution of this sampling method with the solutions obtained from the convex approximations.

5.3 Integer newsvendor problem

5.3.1 Problem definition and analysis

In this section we consider an integer newsvendor problem. This problem, which is an example of a model with simple integer recourse, is the simplest version of a TU
Assessing the quality of convex approximations using sampling

integer recourse problem with \( g \) as in (5.2), and is defined as

\[
\eta^* = \min_{x \geq 0} \left\{ cx + rE_\omega[\lceil \omega - x \rceil] \right\},
\]

(5.15)

with \( 0 < c < r \) and \( \lceil s \rceil^+ := \max\{0, \lfloor s \rfloor\} \), \( s \in \mathbb{R} \). We have substituted the exact expression

\[
g(x, \omega) = cx + \min_y \{ ry : y \geq \omega - x, y \in \mathbb{Z}_+ \} = cx + r \lceil \omega - x \rceil^+, \quad x \geq 0, \omega \in \mathbb{R},
\]

in the objective function of (5.15). Moreover, observe that the problem is generally non-convex because of the round-up operator.

The approximating models corresponding to the \( \alpha \)-approximations and shifted LP-relaxation approximation defined in (5.5) and (5.6), respectively, reduce to

\[
\eta_\alpha := \min_{x \geq 0} \left\{ cx + rE_\omega[(\lceil\omega\rceil_\alpha - x\rceil^+] \right\},
\]

(5.16)

and

\[
\hat{\eta} := \min_{x \geq 0} \left\{ cx + rE_\omega[(\omega + 1/2 - x)\rceil^+] \right\}.
\]

(5.17)

Both models are newsvendor problems for which closed-form solutions can be obtained. For example, for the shifted LP-relaxation approximation we have \( \hat{x} = (1/2 + F^{-1}(\frac{\omega}{2}))^+ \) with \( F^{-1} \) denoting the quantile function of \( \omega \). The quality of these solutions is guaranteed by the error bounds in Theorem 5.1. Combining those with (5.4) we have for this integer newsvendor problem that

\[
G(x_\alpha) - \eta^* \leq 2rh(|\Delta|_f) \quad \text{and} \quad G(\hat{x}) - \eta^* \leq rh(|\Delta|_f),
\]

where \( h \) is defined in (5.7). In the next section we analyze the sharpness of these bounds using numerical experiments. Below we compute these bounds for both normal and lognormal random variables \( \omega \).

**Example 5.1.** Let \( \omega \sim N(\mu, \sigma^2) \) be a normally distributed random variable with pdf \( f \). Then, \( f \) is unimodal with maximum \( 1/\sqrt{2\pi\sigma^2} \) at \( x = \mu \), so that \( |\Delta|_f = \ldots \)
\[2/\sqrt{2\pi \sigma^2} = \sigma^{-1}\sqrt{2/\pi}, \text{ and thus}\]

\[G(x_\alpha) - \eta^* \leq 2rh(\sigma^{-1}\sqrt{2/\pi}) \quad \text{and} \quad G(\hat{x}) - \eta^* \leq rh(\sigma^{-1}\sqrt{2/\pi}),\]

where

\[h(\sigma^{-1}\sqrt{2/\pi}) = \begin{cases} 
1 - \sigma\sqrt{2\pi}, & \sigma \leq 1/\sqrt{8\pi}, \\
(8\sigma)^{-1}\sqrt{2/\pi}, & \sigma \geq 1/\sqrt{8\pi}.
\end{cases}\]

Similarly, we have \(\|G - G_\alpha\|_\infty \leq rh(\sigma^{-1}\sqrt{2/\pi})\) and \(\|G - \hat{G}\|_\infty \leq \frac{1}{2}rh(\sigma^{-1}\sqrt{2/\pi})\).

Figure 5.1 compares the actual values of the supremum norms with their upper bound for the case \(r = 1\). It is the same figure as in Example 3.2 of Chapter 3, but now with the values of \(\|G - \hat{G}\|_\infty\) included. Observe that indeed the upper bound is reasonably tight, and that the shifted LP-relaxation approximation is better than the \(\alpha\)-approximations.

Example 5.2. Let \(\omega\) be lognormally distributed, i.e., \(\ln \omega \sim N(\mu, \sigma^2)\). In this case, \(E_\omega[\omega] = \exp\{\mu + \sigma^2/2\}\) and \(\text{Var}(\omega) = (\exp\{\sigma^2\} - 1)\exp\{2\mu + \sigma^2\}\), and moreover \(\omega\) has pdf

\[f(x) = \begin{cases} 
\frac{1}{x\sqrt{2\pi \sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right\}, & x > 0 \\
0, & \text{otherwise}.
\end{cases}\]

The pdf \(f\) is unimodal with mode \(\exp\{\mu - \sigma^2\}\). It follows immediately that

\[|\Delta|f = \frac{2}{\pi \sigma^2} \exp\left\{\frac{1}{2}\sigma^2 - \mu\right\} \cdot \frac{1}{x\sqrt{2\pi \sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right\}, \quad x > 0\]

In contrast to the normal case, the total variation \(|\Delta|f\) of \(f\) depends on the mean \(\mu\), and it decreases as \(\mu\) increases. Moreover, for large values of \(\sigma\) the total variation \(|\Delta|f\) is also large in the lognormal case, even though the variance of \(\omega\) is large. This illustrates that there is not necessarily a one-to-one relation between the variance of \(\omega\) and the total variation \(|\Delta|f\) of the pdf, \(f\), of \(\omega\), as is the case when \(\omega\) is normally distributed.

Remark 5.2. In general it is hard to compute \(\|G - G_\alpha\|_\infty\) and \(\|G - \hat{G}\|_\infty\), but for this integer newsvendor problem it is possible using brute force computation. In fact,
Assessing the quality of convex approximations using sampling

Figure 5.1: The supremum norms $\|G - \hat{G}\|_\infty$ and $\|G - G_\alpha\|_\infty$, and their upper bound $rh(|\Delta f|)$, of Example 5.1 (with $r = 1$) as a function of $\sigma$, the standard deviation of the random variable $\omega \sim N(0, \sigma^2)$. The dotted line corresponds to $h(|\Delta f|)$, the dashed lines to $\|G - \hat{G}\|_\infty$, and the solid lines to $\|G - G_\alpha\|_\infty$ for $\alpha = 0, 0.5, 0.75, 0.99$. The solid lines can be identified by their limit points at $\sigma \to 0$. For $\alpha = 0.75$ and $\alpha = 0.99$, this limit point equals $(0, \alpha)$. For both $\alpha = 0$ and $\alpha = 0.5$, this limit point is $(0, 0.5)$ and the graph of $\|G - G_\alpha\|_\infty$ corresponding to $\alpha = 0.5$ lies below the one corresponding to $\alpha = 0$.

values of $G(x_\alpha) - \eta^*$ and $G(\hat{x}) - \eta^*$ might also be obtained in a similar way. However, we prefer to use the MRP here instead for comparison with the fleet allocation and routing problem of Section 5.4 for which brute force computations are intractable.

5.3.2 Numerical experiments

Here, we carry out numerical experiments for the integer newsvendor problem. We compare the performance of $x_\alpha$ and $\hat{x}$, the approximating solutions of the $\alpha$-approximation and shifted LP-relaxation approximation, respectively. We also apply the MRP to estimate the optimality gaps $\theta(x_\alpha) = G(x_\alpha) - \eta^*$ and $\theta(\hat{x}) = G(\hat{x}) - \eta^*$, and we compare these optimality gaps, or rather their estimates, with the upper bounds
We consider two types of distributions, the normal \( \omega \sim N(\mu, \sigma^2) \) and lognormal \((\ln \omega \sim N(\mu, \sigma^2))\). The latter is a distribution with heavy tails, whereas the tails of the normal distribution decrease exponentially. The value of \( c \) is standardized to 1 and we use \( r \in \{0.1, 0.3, 2, 4, 20\} \) in our experiments. The values of \( r \) are chosen such that (approximately) \( r - c \in \{0.05, 0.25, 0.5, 0.75\} \), and thus the solution \( \hat{x} \) is obtained by computing very different quantiles of the distribution of \( \omega \).

Table 5.1: Comparison of the shifted LP-relaxation approximation and the \( \alpha \)-approximation for the integer newsvendor problem (5.15) when \( \omega \) is normally distributed. Exact objective values \( G(\hat{x}) \) and \( G(x_\alpha) \) with \( \alpha = 0, 0.25, 0.5, \) and 0.75 are given; for each experiment the minimum of these objective values is displayed in bold.

<table>
<thead>
<tr>
<th>Exp.</th>
<th>( \mu )</th>
<th>( \sigma )</th>
<th>( r )</th>
<th>( G(\hat{x}) )</th>
<th>( G(x_0) )</th>
<th>( G(x_{0.25}) )</th>
<th>( G(x_{0.5}) )</th>
<th>( G(x_{0.75}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.1</td>
<td>1.05</td>
<td>1.336</td>
<td>1.526</td>
<td>1.257</td>
<td>1.500</td>
<td>1.750</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.1</td>
<td>1.3</td>
<td>1.433</td>
<td>1.667</td>
<td>1.258</td>
<td>1.500</td>
<td>1.750</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.1</td>
<td>2</td>
<td>1.500</td>
<td>2.000</td>
<td>1.262</td>
<td>1.500</td>
<td>1.750</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.1</td>
<td>4</td>
<td>1.567</td>
<td>2.000</td>
<td>1.275</td>
<td>1.500</td>
<td>1.750</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.1</td>
<td>20</td>
<td>1.664</td>
<td>2.000</td>
<td>1.374</td>
<td>1.500</td>
<td>1.750</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0.5</td>
<td>1.05</td>
<td>1.550</td>
<td>1.550</td>
<td>1.564</td>
<td>1.554</td>
<td>1.548</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0.5</td>
<td>1.3</td>
<td>1.673</td>
<td>1.697</td>
<td>1.670</td>
<td>1.713</td>
<td>1.761</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0.5</td>
<td>2</td>
<td>1.820</td>
<td>2.046</td>
<td>1.880</td>
<td>1.820</td>
<td>1.884</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>0.5</td>
<td>4</td>
<td>2.026</td>
<td>2.091</td>
<td>2.275</td>
<td>2.140</td>
<td>2.018</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>0.5</td>
<td>20</td>
<td>2.404</td>
<td>2.456</td>
<td>2.374</td>
<td>2.527</td>
<td>2.755</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>1</td>
<td>1.05</td>
<td>1.604</td>
<td>1.604</td>
<td>1.611</td>
<td>1.604</td>
<td>1.604</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1</td>
<td>1.3</td>
<td>1.906</td>
<td>1.910</td>
<td>1.943</td>
<td>1.931</td>
<td>1.908</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2.264</td>
<td>2.366</td>
<td>2.290</td>
<td>2.264</td>
<td>2.290</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2.717</td>
<td>2.731</td>
<td>2.724</td>
<td>2.793</td>
<td>2.829</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>1</td>
<td>20</td>
<td>3.481</td>
<td>3.482</td>
<td>3.506</td>
<td>3.629</td>
<td>3.613</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>3</td>
<td>1.05</td>
<td>2.198</td>
<td>2.198</td>
<td>2.198</td>
<td>2.198</td>
<td>2.198</td>
</tr>
<tr>
<td>17</td>
<td>1</td>
<td>3</td>
<td>1.3</td>
<td>2.785</td>
<td>2.785</td>
<td>2.785</td>
<td>2.785</td>
<td>2.785</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3.883</td>
<td>3.916</td>
<td>3.891</td>
<td>3.883</td>
<td>3.891</td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>5.296</td>
<td>5.344</td>
<td>5.311</td>
<td>5.296</td>
<td>5.307</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>3</td>
<td>20</td>
<td>7.660</td>
<td>7.723</td>
<td>7.669</td>
<td>7.662</td>
<td>7.697</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>10</td>
<td>1.05</td>
<td>5.034</td>
<td>5.034</td>
<td>5.034</td>
<td>5.034</td>
<td>5.034</td>
</tr>
<tr>
<td>25</td>
<td>1</td>
<td>10</td>
<td>20</td>
<td>22.119</td>
<td>22.119</td>
<td>22.128</td>
<td>22.139</td>
<td>22.122</td>
</tr>
</tbody>
</table>

Table 5.1 compares the shifted LP-relaxation approximation and the \( \alpha \)-approximation with \( \alpha = 0, 0.25, 0.5, \) and 0.75 for \( \omega \) normally distributed. For large values of \( \sigma \), i.e., \( \sigma = 10 \), the difference between the approximations is very small, whereas for
small values of $\sigma$, i.e., $\sigma = 0.1$, the approximations differ significantly. In the latter case, the solution $x_{0.25}$ is best. In some sense, this can be considered a coincidence because by construction the optimal solution to (5.16) is either $x_\alpha = 0$ or $x_\alpha \in \alpha + \mathbb{Z}$, and $x_{0.25} = 1.25$ is close to the optimal solution of the integer newsvendor problem (5.15). On the other hand, for medium to large values of $\sigma$, the shifted LP-relaxation approximation outperforms all $\alpha$-approximations, in line with the fact that its error bound is better by a factor of 2. We thus prefer the shifted LP-relaxation approximation, also because in contrast to the $\alpha$-approximation it does not require specification of parameter $\alpha$. For the $\alpha$-approximations the experiments suggest that it is important to select a good value of $\alpha$, in particular if $\sigma$ is small, but this value of $\alpha$ depends on the fractional value of the unknown optimal solution $x^*$ of (5.15). The analog of Table 5.1 when $\omega$ is lognormal is similar, and so we do not include those results here.

Table 5.1 does not give any information on how close to optimal the approximations are. So, we use the MRP to evaluate the optimality gaps. We restrict our attention here to the shifted LP-relaxation approximation. Results for the $\alpha$-approximations are very similar.

| $\mu$ | $\sigma$ | $|\Delta|/f$ | $\rho_1(\hat{x})$ | $\rho_2(\hat{x})$ |
|-------|---------|-------------|------------------|------------------|
| 1     | 0.1     | 7.9         | 1.05             | 1.05             |
| 1     | 0.5     | 1.5         | 15.2             | 14.9             |
| 1     | 1       | 0.8         | 2.5              | 2.5              |
| 1     | 3       | 0.26        | 2.2              | 2.2              |
| 1     | 10      | 0.08        | 9.4              | 9.4              |

For the MRP we use $N_r = 30$, $n = 1000$, and $\gamma = 0.05$, and we use LHS in Step 1(i) of the procedure. The approximating solutions $\hat{x}$ are obtained by solving (5.17) exactly. Alternatively, this solution could also have been obtained using a sample average approximation of (5.17) with a large sample. We report three performance measures $\rho_1$, $\rho_2$, and $\rho_3$. 

Table 5.2: Numerical results for the shifted LP-relaxation approximation applied to the integer newsvendor problem (5.15) with $\omega$ normally distributed. The MRP is applied with $N_r = 30$, $n = 1000$, and $\gamma = 0.05$. 

| $\mu$ | $\sigma$ | $|\Delta|/f$ | $\rho_1(\hat{x})$ | $\rho_2(\hat{x})$ |
|-------|---------|-------------|------------------|------------------|
| 1     | 0.1     | 7.9         | 1.05             | 1.05             |
| 1     | 0.5     | 1.5         | 15.2             | 14.9             |
| 1     | 1       | 0.8         | 2.5              | 2.5              |
| 1     | 3       | 0.26        | 2.2              | 2.2              |
| 1     | 10      | 0.08        | 9.4              | 9.4              |

Optimality gap compared with upper bound (in %):

$\rho_1(\hat{x})$

Optimality gap compared with optimal objective (in %):

$\rho_2(\hat{x})$
The first,
\[ \rho_1(\hat{x}) := \frac{\hat{\theta}_n(\hat{x}) + \epsilon_\theta}{\tau h(|\Delta|f)} \times 100\%, \]
compares the optimality gap, or more precisely the width of the MRP’s 95% confidence interval on the optimality gap, \( G(\hat{x}) - \eta^* \), with the error bound of Theorem 5.1. If \( \rho_1(\hat{x}) \) is small, then the actual performance of the shifted LP-relaxation approximation is better than its error bound suggests.

The second performance measure,
\[ \rho_2(\hat{x}) := \frac{\hat{\theta}_n(\hat{x}) + \epsilon_\theta}{N_r^{-1} \sum_{i=1}^{N_r} \eta^*_n} \times 100\%, \]
compares the same estimate of the optimality gap with an estimate of the optimal objective value, \( \eta^* \). Obviously, the approximating solution \( \hat{x} \) is estimated to be a good solution if \( \rho_2(\hat{x}) \) is small. Note that because \( \mathbb{E}[\eta^*_n] \leq \eta^* \) and \( \mathbb{E}[\hat{\theta}_n(\hat{x}) + \epsilon_\theta] \geq G(\hat{x}) - \eta^* \) the performance measure \( \rho_2(\hat{x}) \) is a conservative estimate of \( \frac{(G(\hat{x}) - \eta^*)/\eta^* \times 100\%}{\eta^*} \).

Furthermore, we compare the approximating solution \( \hat{x} \) with a sampling solution \( x^S \). This sampling solution is the best of the solutions \( x^*_n, i = 1, \ldots, N_r \), obtained during the MRP, and their average \( N_r^{-1} \sum_{i=1}^{N_r} x^*_n \). To assess the quality of the sampling solution we report \( \rho_2(x^S) \), and to compare \( \hat{x} \) and \( x^S \) we use the performance measure,
\[ \rho_3(\hat{x}, x^S) := \frac{G(\hat{x}) - G(x^S)}{N_r^{-1} \sum_{i=1}^{N_r} \eta^*_n} \times 100\%. \]

We note that we use the MRP twice; first to obtain the sampling solution, \( x^S \), and second to assess the performance of \( \hat{x} \) and \( x^S \).

In Table 5.2 we show the performance measures \( \rho_1(\hat{x}) \) and \( \rho_2(\hat{x}) \) for normally distributed \( \omega \) and in Table 5.3 for lognormal \( \omega \). For the normal case, we observe that \( \rho_1(\hat{x}) \) is small in case of medium variability (i.e., \( \sigma = 0.5, 1, 3 \)). In these cases, the quality of the solution \( \hat{x} \) is much better than the error bound of Theorem 5.1 suggests. Indeed, \( \rho_2(\hat{x}) \) suggests that, with high confidence, the objective value of the solution \( \hat{x} \) is within 1% of the optimal objective function value in almost all cases. In contrast, for both low variability (\( \sigma = 0.1 \)) and high variability (\( \sigma = 10 \)), the value of \( \rho_1(\hat{x}) \) may be above 15% and range up to 60%. In the first case this is because the solution \( \hat{x} \) is not good: \( \rho_2(\hat{x}) \) may exceed 20%. In the second case, however, the value of \( \rho_2(\hat{x}) \) is
Assessing the quality of convex approximations using sampling

Table 5.3: Numerical results for the shifted LP-relaxation approximation applied to the integer newsvendor problem (5.15) with $\omega$ lognormally distributed. The MRP is applied with $N_r = 30$, $n = 1000$, and $\gamma = 0.05$.

| $\mu$ | $\sigma$ | $|\Delta f|$ (in %) | $\rho_1(\hat{x})$ | $\rho_2(\hat{x})$ |
|---|---|---|---|---|
| $\mu$ | $\sigma$ | 1.05 | 1.3 | 2 | 4 | 20 |
| 0.1 | 8.0 | 15.2 | 19.1 | 15.7 | 9.1 | 2.1 |
| 0.5 | 1.8 | 2.8 | 0.9 | 1.3 | 1.3 | 0.7 |
| 1.5 | 1.6 | 2.6 | 2.6 | 2.3 | 1.9 | 2.0 |
| 0.5 | 0.67 | 5.0 | 3.8 | 3.1 | 3.0 | 2.5 |
| 1.7 | 0.26 | 30.9 | 33.5 | 31.8 | 32.4 | 15.1 |

optimal objective (in %):

<table>
<thead>
<tr>
<th>$\rho_1(\hat{x})$</th>
<th>$\rho_2(\hat{x})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.05</td>
<td>1.3</td>
</tr>
<tr>
<td>9.7</td>
<td>15.4</td>
</tr>
<tr>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>0.03</td>
<td>0.04</td>
</tr>
</tbody>
</table>

small, and thus $\hat{x}$ is a good solution. The large values of $\rho_1(\hat{x})$ in this case are inherent to the nature of the total variation error bound and the MRP optimality gap. As the standard deviation $\sigma$ grows, the total variation error bound shrinks, whereas the MRP optimality gap remains approximately the same, explaining the large values of $\rho_1(\hat{x})$ for the high variability case ($\sigma = 10$). For the lognormal case, we obtain similar results (see Table 5.3). We have selected values of $\mu$ and $\sigma$ so that $|\Delta f|$ approximately matches those in the normal case. As detailed in Example 5.2 this does not mean that the variances of $\omega$ match, however. For example, in the lognormal case with $\mu = 0$ and $\sigma = 1.5$, we have $\text{Var}(\omega) \approx 80.5$, and for $\mu = 1$ and $\sigma = 0.5$, we have $\text{Var}(\omega) \approx 2.7$.

Comparing the shifted LP-relaxation approximation solution, $\hat{x}$, and the sampling solution, $x^S$, in Table 5.4 for normal random variables, $\omega$, we observe that $\rho_3(\hat{x}, x^S)$ is only large in the case of low variability (i.e., $\sigma = 0.1$). Indeed, in contrast to the shifted LP-relaxation approximation, the sampling solution, $x^S$, is good in the low variability case as well. In fact, $\rho_2(x^S)$ is small in all cases. This is as expected, since we use a large sample (of size $n = 1000$) to obtain the sampling solution, $x^S$. In some higher-dimensional problems, larger sample sizes are required to obtain high-quality solutions, and yet such problems are more difficult to solve and could be intractable even with modest sample sizes. With this in mind, the shifted LP-relaxation approximation performs well in case of medium to high variability. In these cases both solution methods perform approximately the same. Since we obtain similar results for the lognormal case, we omit those computational results here.
Table 5.4: A comparison between the shifted LP-relaxation approximation and a sampling solution method for the integer newsvendor problem (5.15) with $\omega$ normally distributed. The MRP is applied with $N_r = 30$, $n = 1000$, and $\gamma = 0.05$. Values reported as $\pm 0.00$ are small in magnitude while values reported as 0 are indeed zero.

| $\mu$ | $\sigma$ | $|\Delta f|$ | $\rho_3(\hat{x}, x^S)$ | $|\Delta f|$ | $\rho_3(x^S)$ |
|-------|---------|--------------|-----------------|--------------|-----------------|
| 1     | 0.1     | 7.9          | 9.8             | 1.05         | 0.08            |
| 1     | 0.5     | 1.5          | 0.15            | 1.3          | 0.1             |
| 1     | 1       | 0.8          | -0.00           | 2            | 0.1             |
| 1     | 3       | 0.26         | 0               | 4            | 0.04            |
| 1     | 10      | 0.08         | 0               | 20           | 0.02            |

**5.4 Fleet allocation and routing problem**

This section discusses a variant of the fleet allocation and routing problem introduced in [15]. Mak et al. [51] also report numerical results for this problem.

The problem may be viewed as a two-stage totally unimodular integer recourse model, but with relatively complete recourse instead of complete recourse and with deterministic side constraints in the second stage. This is why we have to reconsider what type of convex approximations are suitable for this problem, and, moreover, why we (again) have to derive an error bound for these approximations.

First, we define the problem, formulate it as a two-stage integer recourse model, and derive properties of this model in Section 5.4.1. Next, in Section 5.4.2 we construct a concave approximation $g_0$—since we are maximizing—of the recourse function $g$, and we derive an error bound for this approximation. Finally, in Section 5.4.3 we carry out numerical experiments comparing the actual error of the concave approximation with its error bound.

**5.4.1 Problem definition and model formulation**

In this section, we define the problem, formulate it as a stochastic integer program, and discusses properties of the model.
Figure 5.2: The graph $\mathcal{G} = (V, E)$ used for the numerical experiments. Nodes 1-5 are source nodes ($V_s$) and node 20 is the sink node ($t$). All arcs are directed from left to right.

Problem definition

Consider an acyclic directed graph $\mathcal{G} = (V, E)$, modeling a road network. A fleet of trucks will traverse this network starting at source nodes $V_s \subset V$ and finishing at a sink node $t \in V$. For every arc $(i, j) \in E$, the first $D_{ij}$ trucks traversing the arc receive a reward $r_{ij} > 0$ and subsequent trucks incur a cost $c_{ij} > 0$. The quantities $D_{ij}$ are “soft” demands, or customers requesting service, along arc $(i, j)$. Trucks receive profit if they serve a customer, and incur a cost otherwise. The problem is to allocate $N$ trucks to the source nodes $V_s$ and route them through the network to maximize profit.

When we allocate trucks to the source nodes, the demands $D_{ij}$ are unknown. We assume that they are in part random but may be increased by investments, or marketing actions $\beta_{ij}$, which incur unit costs $q_{ij}$. That is, the demand $D_{ij}(\omega_{ij}, \beta_{ij})$ is a function of the random variable $\omega_{ij}$ (with known probability distribution) and the investments $\beta_{ij}$. The effect of the investments may be additive ($D_{ij}(\omega_{ij}, \beta_{ij}) = \omega_{ij} + \beta_{ij}$) or multiplicative ($D_{ij}(\omega_{ij}, \beta_{ij}) = \omega_{ij}(1 + \beta_{ij})$). In either case, when the investment is zero, we have $D_{ij}(\omega_{ij}, 0) = \omega_{ij}$. Observe that we do not define demands to be integer; instead we will impose integrality restrictions on the number of trucks traversing an arc. Thus, our objective is to allocate the trucks to the source nodes...
and invest $\beta$ in the arcs at costs $q$ to maximize expected profits.

Model formulation: two-stage recourse

This problem can be formulated as a two-stage integer recourse model. In the first stage we decide the number of trucks, $n_i$, to allocate to each source node, $i \in V_s$, and on the investments, $\beta_{ij} \geq 0$, for each arc, $(i,j) \in E$. In the second stage we let $y_{ij} \in \mathbb{Z}_+$ denote the number of trucks receiving a reward $r_{ij}$ when traversing arc $(i,j)$, and we let $z_{ij} \in \mathbb{Z}_+$ denote the number of trucks incurring a cost $c_{ij}$. Then, the two-stage integer recourse model for this problem is given by

$$
\eta^* := \max_x \left\{ E_\omega[g(x,\omega)] : x \in X \right\},
$$

where $x := (n,\beta)$ with feasible region $X := \{(n,\beta) \in \mathbb{Z}_+^{\left|V_s\right|} \times \mathbb{R}_+^{\left|E\right|} : \sum_{i \in V_s} n_i = N\}$, and $g(x,\omega) := -q\beta + \pi(x,\omega)$ with

$$
\pi(x,\omega) := \max_{y,z} \quad ry - cz
$$

s.t.

$$
\sum_{j: (i,j) \in E} (y_{ij} + z_{ij}) = n_i \quad i \in V_s \quad (5.19)
$$

$$
\sum_{j: (i,j) \in E} (y_{ij} + z_{ij}) - \sum_{j: (j,i) \in E} (y_{ji} + z_{ji}) = 0 \quad i \in V \setminus (V_s \cup \{t\}) \quad (5.20)
$$

$$
\sum_{j: (i,j) \in E} (y_{ji} + z_{ji}) = N \quad i = t \quad (5.21)
$$

$$
0 \leq y_{ij} \leq D_{ij}(\omega_{ij},\beta_{ij}), \quad 0 \leq z_{ij} \quad (i,j) \in E
$$

$$
y, z \in \mathbb{Z}_+^{\left|E\right|}.
$$

Here, constraints (5.19)–(5.21) represent flow balance constraints, modeling, respectively, that $n_i$ trucks must leave source node $i \in V_s$, that every truck that enters node $i \in V \setminus (V_s \cup \{t\})$ must leave that node, and that all $N$ trucks must arrive at sink node $t$.

**Remark 5.3.** We assume without loss of generality that the source nodes only have outgoing arcs and the sink node only has incoming arcs.

**Remark 5.4.** This fleet allocation and routing model is a special case of the two-stage integer recourse model defined in (5.2), since maximization can easily be reformulated
as minimization and the flow balance constraints can be capture by the set $Y$. When the effect of investments is only additive, there is only randomness in the right-hand side vector $\zeta(\omega)$. If multiplicative effects are considered, then also the technology matrix $T(\omega)$ is random.

Let $A$ denote the node-arc incidence matrix of $G$, where the rows of $A$ correspond to the nodes of $G$ and the columns of $A$ to the arcs of $G$. The column of $A$ corresponding to arc $(i, j) \in E$ has one entry equal to $+1$ in row $i$, one entry equal to $-1$ in row $j$, and the remaining entries equal zero. Defining $b(n) = (n, 0, -N)$, the flow balance constraints can be written as $Ay + Az = b(n)$. Thus, for every $x = (n, \beta) \in X$,

$$g(x, \omega) = -q\beta + \max_{y,z} \left\{ ry - cz : Ay + Az = b(n), y \leq D(\omega, \beta), y, z \in \mathbb{Z}_+ \right\}.$$

(5.22)

Properties of the recourse function $g$

Here, we discuss properties of the recourse function $g$. We assume that the directed acyclic graph $G$ is $t$-connected; i.e., we assume that for every node $i$ in the graph there is a directed $i$-$t$ path. Under this assumption, we show that the recourse is relatively complete and that the recourse matrix corresponding to the second-stage optimization problem in $g$ is TU.

**Lemma 5.1.** Let graph $G = (V, E)$ be $t$-connected, let $g$ be the recourse function defined in (5.22), and let $G(x) := \mathbb{E}_\omega[g(x, \omega)]$, $x \in X$. Then, the recourse is relatively complete and sufficiently expensive; that is,

(i) $g(x, \omega)$ is finite for every $x \in X$ and $\omega \in \mathbb{R}_+^{|E|}$; and,

(ii) $G(x)$ is finite for every $x \in X$ and nonnegative random vector $\omega$.

**Proof.** Let $x \in X$ and $\omega \in \mathbb{R}_+^{|E|}$ be given. Clearly, there exists a feasible solution of the maximization problem in $g$, for example using $y = 0$ and $z$ such that $(y, z)$ is feasible. Moreover, since the graph $G$ is acyclic, it follows immediately from the flow balance constraints that for any feasible solution we have $0 \leq y_{ij}, z_{ij} \leq N$ for all $(i, j) \in E$, and thus

$$-N \sum_{(i,j)\in E} c_{ij} \leq g(x, \omega) \leq N \sum_{(i,j)\in E} r_{ij},$$

(5.22)
implying that (i) \( g(x, \omega) \) is finite. Now it is not hard to prove (ii) since for every \( x \in X \) and nonnegative random vector \( \omega \) we have

\[
-N \sum_{(i,j) \in E} c_{ij} \leq G(x) \leq N \sum_{(i,j) \in E} r_{ij}.
\]

Next, we show that the recourse matrix \( W \), defined in (5.23) below, of the maximization problem in \( g \) is TU, implying that we can represent it as a linear program. This does not imply, however, that (5.18) is as easy to solve as a two-stage continuous recourse problem since the LP representation of \( g \) involves rounding down the demands \( D_{ij}(\omega_{ij}, \beta_{ij}) \).

**Lemma 5.2.** Let \( A \) denote the node-arc incidence matrix of a directed graph \( G = (V, E) \) and define

\[
W = \begin{bmatrix}
A & A \\
I & 0
\end{bmatrix}.
\]

Then, \( W \) is totally unimodular.

**Proof.** The matrix \( A \) is TU since every node-arc incidence matrix of a directed \( G \) is TU [67]. The matrix \( [A \ A] \) is TU as well, since every square submatrix of \( [A \ A] \) is either a submatrix of \( A \) or has two identical columns. Finally, \( W \) is TU since it is an extension of a TU matrix by \([I \ 0]\).

In what follows we use Lemma 5.2 to obtain a dual representation of \( g \).

### 5.4.2 Concave approximation

Here, we derive a concave approximation of \( g \) (since here we are maximizing instead of minimizing) using the same type of approximations as in Section 5.2.1, i.e., an \( \alpha \)-approximation and a shifted LP-relaxation approximation. However, difficulties arise because the recourse is relatively complete rather than complete. Moreover, to derive an error bound we have to extend the analysis in Section 5.2.1 to be able to deal with the flow balance constraints, which we may view as deterministic second-stage side constraints. We will do so by using the same line of proof as in Section 5.2.1, exploiting the results presented there.
Definition of the concave approximation $g_0$

The main idea in the convex approximations of Section 5.2.1 is to simultaneously relax the integrality constraints and replace the right-hand side $\omega$ by $[\omega] + \alpha$ or $\omega + 1/2\epsilon_m$.

In this case, since we are maximizing instead of minimizing, and the right-hand side $D(\omega, \beta)$ is rounded down instead of up, we analogously replace $\omega$ by either $\lfloor \omega \rfloor + \alpha$ or $\omega - 1/2\epsilon_m$.

However, $\omega - 1/2\epsilon_m$ may be negative with positive probability, even if the random vector $\omega \geq 0$, which implies that for $\beta = 0$, or small, the approximating demands may be negative, and thus the approximating maximization problem infeasible. The same holds for $\lfloor \omega \rfloor$, unless $\alpha \in \mathbb{Z}^m$. For such $\alpha \in \mathbb{Z}^m$, we have that $[\omega] = \lfloor \omega \rfloor \geq 0$ if $\omega \geq 0$.

Thus, interestingly, the only reasonable concave approximation that can be used for every nonnegative random vector $\omega$ is an $\alpha$-approximation with $\alpha = 0$. This approximation, denoted $g_0$, is defined for every $x \in X$ and $\omega \in \mathbb{R}^{|E|}_+$ as

$$g_0(x, \omega) = -q\beta + \max_{y,z} \left\{ ry - cz : Ay + Az = b(n), y \leq \tilde{D}(\omega, \beta), y, z \in \mathbb{R}^{|E|}_+ \right\},$$

where for every $(i, j) \in E$,

$$\tilde{D}_{ij}(\omega_{ij}, \beta_{ij}) := \begin{cases} 
[\omega_{ij}] + \omega_{ij} \beta_{ij}, & (i, j) \in E^*, \\
[\omega_{ij}] + \beta_{ij}, & (i, j) \in E^+, \\
[\omega_{ij}], & (i, j) \in E^0,
\end{cases}$$

and where $E^*$, $E^+$, and $E^0$ partition $E$ into subsets with multiplicative, additive, and no investment effects, respectively. Observe that $g_0(x, \omega)$ is concave in $x$ for every $\omega \in \mathbb{R}^{|E|}_+$. Moreover, notice that for multiplicative investments effects we have $\tilde{D}_{ij}(\omega_{ij}, \beta_{ij}) \neq D_{ij}([\omega_{ij}], \beta_{ij})$ unless $\omega_{ij} \in \mathbb{Z}$. The latter approximation, $D_{ij}([\omega_{ij}], \beta_{ij}) = [\omega_{ij}] + [\omega_{ij}] \beta_{ij}$, would be too small for larger values of $\beta_{ij}$, and this is why we propose $\tilde{D}_{ij}$ instead.

Although the approximating model

$$\eta_0 := \max_x \left\{ \mathbb{E}_\omega [g_0(x, \omega)] : x \in X \right\},$$

yielding the approximating solution $x_0 = (n_0, \beta_0)$, has integer first-stage decision
variables, it is generally much easier to solve than the original problem (5.18) since this approximating model has a concave objective function. In Section 5.4.3 we use numerical experiments to analyze the quality of the solution \( x_0 = (\nu_0, \beta_0) \). First, however, we derive an error bound for the concave approximation \( g_0 \), similar to that in Theorem 5.1.

**Dual representation of \( g \) and \( g_0 \)**

To derive an upper bound on \( \| G - G_0 \|_\infty \) with \( G_0(x) := E_\omega[g_0(x, \omega)] \), \( x \in X \), we first derive a dual representation of \( g \) and \( g_0 \). Here, we use the fact that, by Lemma 5.2, \( g \) is a TU integer recourse function. In analogous fashion to (5.8) and (5.9) in Section 5.2.1, we round down \( D(\omega, \beta) \), relax the integrality constraints, and apply strong LP duality to obtain

\[
g(x, \omega) = -q\beta + \min_{\mu, \lambda} \left\{ \mu b(n) + \lambda [D(\omega, \beta)] : (\mu, \lambda) \in \Lambda \right\}, \quad x \in X, \omega \in \mathbb{R}^{[E]}_+, \quad (5.25)
\]

where \( \Lambda := \{ (\mu, \lambda) \in \mathbb{R}^{[V]} \times \mathbb{R}^{[E]}_+ : \mu A + \lambda \geq r, \mu A \geq -c \} \). Similarly,

\[
g_0(x, \omega) = -q\beta + \min_{\mu, \lambda} \left\{ \mu b(n) + \lambda \tilde{D}(\omega, \beta) : (\mu, \lambda) \in \Lambda \right\}, \quad x \in X, \omega \in \mathbb{R}^{[E]}_+. \quad (5.26)
\]

Next, we derive, for a fixed \( x \in X \), monotonicity properties of minimizers \( (\mu(\omega), \lambda(\omega)) \) and \( (\tilde{\mu}(\omega), \tilde{\lambda}(\omega)) \) of the optimization problems in \( g(x, \omega) \) and \( g_0(x, \omega) \), respectively, for every \( \omega \in \mathbb{R}^{[E]}_+ \). These properties are necessary to apply Proposition 5.1 to ‘round up’ \( \lambda(\omega) \) and \( \tilde{\lambda}(\omega) \) in the proof of Theorem 5.2.

**Lemma 5.3.** Let the directed acyclic graph \( G \) be \( t \)-connected. Consider the dual feasible region \( \Lambda \) defined as

\[
\Lambda = \left\{ (\mu, \lambda) \in \mathbb{R}^{[V]} \times \mathbb{R}^{[E]}_+ : \mu A + \lambda \geq r, \mu A \geq -c \right\}
\]

and let \( x \in X \) be given. Let \( H : \mathbb{R}^{[E]}_+ \mapsto \mathbb{R}^{[E]} \) be a separable nonnegative function for which \( H_{ij}(\omega_{ij}) \) is non-decreasing in \( \omega_{ij} \) for every \( (i, j) \in E \), and let \( \omega_{(i)} \in \mathbb{R}^{[E]}_+ \) denote \( \omega \) without its \( ij \)-th component. Then, there exist minimizers \( (\hat{\mu}(\omega), \hat{\lambda}(\omega)) \) of

\[
\min_{\mu, \lambda} \left\{ \mu b(n) + \lambda H(\omega) : (\mu, \lambda) \in \Lambda \right\}
\]

(5.27)
such that for every \((i, j) \in E\) and \(\omega_{ij} \geq 0\), the function \(\hat{\lambda}_{ij}(\cdot | \omega_{ij}) : \mathbb{R}_+ \rightarrow \mathbb{R}\) defined as \(\hat{\lambda}_{ij}(\omega_{ij}| \omega_{ij}) = \hat{\lambda}_{ij}(\omega)\), satisfies

(i) \(\hat{\lambda}_{ij}(\cdot | \omega_{ij})\) is monotone non-increasing, and

(ii) \(\hat{\lambda}_{ij}(\cdot | \omega_{ij}) \leq r_{ij} + c_{ij}\).

Proof. The proof of (i) is straightforward and similar to the proof of Lemma 4.11 in Chapter 4. Moreover, \(\hat{\lambda}(\omega)\) is bounded since for every fixed \(\hat{\mu}(\omega)\) an optimal solution \(\hat{\lambda}(\omega)\) of (5.27) takes values \(\lambda(\omega) = r - \hat{\mu}(\omega)A\). Here, we use the hypothesis that \(H(\omega)\) is nonnegative and the fact that \(\lambda \geq r - \mu A\) for every \((\mu, \lambda) \in \Lambda\). Since \(-\mu A \leq c\), we conclude that \(\hat{\lambda}(\omega) = r - \hat{\mu}(\omega)A \leq r + c\), proving (ii).

Notice that for fixed \(x \in X\), or more specifically for fixed investments, \(\beta \geq 0\), both \([D(\omega, \beta)]\) and \(\tilde{D}(\omega, \beta)\) satisfy the assumptions of \(H(\omega)\) in Lemma 5.3. Moreover, the monotonicity result in (i) is one of the assumptions in Proposition 5.1 of Section 5.2.1. We can apply this proposition if for every \((i, j) \in E\), the function \(\psi_{ij}(\omega_{ij}; \beta_{ij})\), defined as

\[
\psi_{ij}(\omega_{ij}; \beta_{ij}) = [D_{ij}(\omega_{ij}, \beta_{ij})] - \tilde{D}_{ij}(\omega_{ij}, \beta_{ij}),
\]

is periodic in \(\omega_{ij}\) for some period \(p_{ij}\) with zero mean \(\nu_{ij} := \frac{1}{p_{ij}} \int_0^{p_{ij}} \psi_{ij}(t; \beta_{ij}) \, dt = 0\). If there are no investment effects; i.e., if \((i, j) \in E^0\), then this is trivially true since \(\psi_{ij}(\omega_{ij}; \beta_{ij}) = 0\) for all \(\omega_{ij} \geq 0\). The result also holds if the investment effects are additive, i.e., if \((i, j) \in E^+\), since in this case \(\psi_{ij}(\omega_{ij}; \beta_{ij}) = [\omega_{ij} + \beta_{ij}] - [\omega_{ij}] - [\beta_{ij}] = [\omega_{ij}] - \beta_{ij} - [\omega_{ij}]\) which is the same as \(\varphi_{z_i, \alpha_i}\) defined in Section 5.2.1 with \(z_i := -\beta_{ij}\), \(\alpha_i := 0\), and the round-up operators replaced by round-down operators. However, for multiplicative investment effects, i.e., for \((i, j) \in E^*\), the function \(\psi_{ij}(\omega_{ij}; \beta_{ij})\), given for every \(\omega_{ij} \geq 0\) by

\[
\psi_{ij}(\omega_{ij}; \beta_{ij}) = [\omega_{ij}(1 + \beta_{ij})] - [\omega_{ij}] - \omega_{ij} \beta_{ij},
\]

is periodic in \(\omega_{ij}\) if and only if \(\beta_{ij}\) is rational, in which case its period is the least common multiple of \(1\) and \(1/(1 + \beta_{ij})\). If \(\beta_{ij}\) is irrational, however, this least common multiple does not exist. This implies that for multiplicative investment effects the assumptions of Proposition 5.1 are not satisfied. We can circumvent this problem by
decomposing $\psi_{ij}(\omega_{ij}; \beta_{ij})$ as the sum of two zero-mean periodic functions:

$$
\psi_{ij}(\omega_{ij}; \beta_{ij}) = \left( \lfloor \omega_{ij}(1 + \beta_{ij}) \rfloor - \omega_{ij}(1 + \beta_{ij}) + 1/2 \right) + \left( \omega_{ij} - 1/2 - |\omega_{ij}| \right),
$$

(5.29)

where the first and second periodic functions equal $\hat{\varphi}_{z_i}(\omega_{ij}(1 + \beta_{ij}))$ and $-\hat{\varphi}_{z_i}(\omega_{ij})$, respectively, with $\hat{\varphi}_{z_i}$ defined in Section 5.2.1 and in both cases with $z_i := 0$, the round-up operator replaced by a round-down operator, and the addition of 1.

Error bound

Now we are ready to derive an upper bound on $\|G - G_0\|_\infty$.

**Theorem 5.2.** Let $G = (V_E)$ be a directed acyclic graph that is $t$-connected, and let $E^+, E^0 \subset E$ denote subsets of arcs with multiplicative, additive, and no investment effects, respectively. Let $g$ denote the recourse function corresponding to the fleet allocation and routing problem defined for every $x \in X$ and $\omega \in \mathbb{R}^{|E|}_+$ as

$$
g(x, \omega) = -q\beta + \max_{y,z} \left\{ ry - cz : Ay + Az = b(n), y \leq D(\omega, \beta), y, z \in \mathbb{Z}^{|E|}_+ \right\},
$$

and let $g_0$ denote its concave approximation defined in (5.24). Then, under the assumption that $\omega$ is a continuous random vector with joint pdf $f$ and with independently distributed components, it holds for every $x \in X$, that

$$
|G(x) - G_0(x)| \leq \sum_{(i,j) \in E^+} (r_{ij} + c_{ij}) h(|\Delta f_{ij}|),
$$

where $h$ is defined in (5.7), $G(x) := \mathbb{E}_\omega[g(x, \omega)]$ and $G_0(x) := \mathbb{E}_\omega[g_0(x, \omega)]$, $x \in X$, and $|\Delta f_{ij}$ is the total variation of the marginal density function $f_{ij}$.

**Proof.** Let $x \in X$ and consider the dual representation of $g$ in (5.25). Let $(\mu(\omega), \lambda(\omega))$ be minimizers of (5.27) with $H(\omega) := [D(\omega, \beta)]$, satisfying the properties of Lemma 5.3, so that $g(x, \omega) = -q\beta + \mu(\omega)b(n) + \lambda(\omega) [D(\omega, \beta)]$ for every $\omega \in \mathbb{R}^{|E|}_+$. Since $(\mu(\omega), \lambda(\omega))$ is feasible but not necessarily optimal for (5.27) with $H(\omega) := \tilde{D}(\omega, \beta)$, it follows from the dual representation of $g_0$ in (5.26) that $g_0(x, \omega) \leq -q\beta + \mu(\omega)b(n) + $
Assessing the quality of convex approximations using sampling

\[ \lambda(\omega) \tilde{D}(\omega, \beta), \text{ and thus} \]

\[ G_0(x) - G(x) \leq \mathbb{E}_\omega \left[ \lambda(\omega) \left( \tilde{D}(\omega, \beta) - |D(\omega, \beta)| \right) \right] \]

\[ = \sum_{(i,j) \in E} \mathbb{E}_\omega \left[ \lambda_{ij}(\omega) \left( -\psi_{ij}(\omega_{ij}; \beta_{ij}) \right) \right], \]

where \( \psi_{ij}(\omega_{ij}; \beta_{ij}) \) is defined in (5.28). Similarly, with \( (\tilde{\mu}(\omega), \tilde{\lambda}(\omega)) \) denoting minimizers of (5.27) with \( H(\omega) := \tilde{D}(\omega, \beta) \), we have

\[ G(x) - G_0(x) \leq \sum_{(i,j) \in E} \mathbb{E}_\omega \left[ \tilde{\lambda}_{ij}(\omega) \psi_{ij}(\omega_{ij}; \beta_{ij}) \right]. \] (5.30)

We will derive an upper bound on \( G(x) - G_0(x) \); an upper bound for \( G_0(x) - G(x) \) can be obtained in a similar way. We obtain this upper bound by separately bounding the individual terms, \( \Psi_{ij}(\beta_{ij}) \), in (5.30), defined for each \( (i,j) \in E \) as \( \Psi_{ij}(\beta_{ij}) = \mathbb{E}_\omega \left[ \tilde{\lambda}_{ij}(\omega) \psi_{ij}(\omega_{ij}; \beta_{ij}) \right] \).

Obviously, if \( (i,j) \in E^0 \), then \( \Psi_{ij}(\beta_{ij}) = 0 \) since \( \psi_{ij}(\omega_{ij}; \beta_{ij}) = 0 \) for all \( \omega_{ij} \geq 0 \). Moreover, if \( (i,j) \in E^+ \), then

\[ \Psi_{ij}(\beta_{ij}) = \mathbb{E}_\omega \left[ \tilde{\lambda}_{ij}(\omega) \left( |\omega_{ij}| - \beta_{ij} \right) \right] = \mathbb{E}_\omega \left[ \tilde{\lambda}_{ij}(\omega) \varphi_{-\beta_{ij},0}(\omega_{ij}) \right], \]

where \( \varphi_{-\beta_{ij},0} \) is given in Definition 5.1. The second equality holds since we take the expectation with respect to a continuously distributed random vector \( \omega \), and thus it does not matter that by replacing the round-down operators by round-up operators we change the underlying difference function at countable many points \( \omega_{ij} \in \{0, -\beta_{ij}\} + \mathbb{Z} \).

Writing the expectation as an integral and using the hypothesis that the components of \( \omega \) are independent we have

\[ \Psi_{ij}(\beta_{ij}) = \int_{\mathbb{R}^{[E]-1}} \int_{\mathbb{R}} \tilde{\lambda}_{ij}(u_{ij}|u_{ij}) \varphi_{-\beta_{ij},0}(u_{ij}) f_{ij}(u_{ij}) du_{ij} f_{ij}(u_{ij}) du_{ij}. \]

Consider the inner integral for a fixed \( u_{ij} \in \mathbb{R}^{[E]}-1 \) and observe that \( \tilde{\lambda}_{ij}(|u_{ij}|) \) is monotone non-increasing and bounded by \( r_{ij} + c_{ij} \) according to Lemma 5.3. Moreover, \( \varphi_{-\beta_{ij},0} \) is periodic with zero mean value, so that all assumptions of Proposition 5.1
are satisfied. Applying this proposition to the inner integral yields

\[ \Psi_{ij}(\beta_{ij}) \leq \int_{|E|-1} (r_{ij} + c_{ij}) M(\Phi_{ij}) f_{ij} u_{ij} d\omega_{ij} \]

\[ = (r_{ij} + c_{ij}) M(\Phi_{ij}) \]

\[ \leq (r_{ij} + c_{ij}) h(|\Delta f_{ij}|), \]

where \( M \) is defined in (5.11) and the last inequality follows from (5.13).

It remains to show that \( \Psi_{ij}(\beta_{ij}) \leq (r_{ij} + c_{ij}) h(|\Delta f_{ij}|) \) for \((i,j) \in E^*\). In this case, using (5.29), we have

\[ \Psi_{ij}(\beta_{ij}) = \mathbb{E}_\omega \left[ \tilde{\lambda}_{ij}(\omega) \left( [\omega_{ij}(1 + \beta_{ij})] - [\omega_{ij}] - \omega_{ij}\beta_{ij} \right) \right] \]

\[ = \mathbb{E}_\omega \left[ \tilde{\lambda}_{ij}(\omega) \left( -\hat{\varphi}_0(\omega_{ij}) \right) \right] + \mathbb{E}_\omega \left[ \tilde{\lambda}_{ij}(\omega) \hat{\varphi}_0(\omega_{ij}(1 + \beta_{ij})) \right], \]

with \( \hat{\varphi}_0 \) given in Definition 5.1, and where again the second equality holds even though we replaced the round-down operators by round-up operators. Since \( \hat{\varphi}_0 \) is periodic with zero mean value, we can apply Proposition 5.1 twice, in a similar way as for \((i,j) \in E^+\), to obtain

\[ \Psi_{ij}(\beta_{ij}) \leq (r_{ij} + c_{ij}) M\left(-\hat{\varphi}_0, |\Delta f_{ij}|\right) + (r_{ij} + c_{ij}) M\left(\hat{\varphi}_0, \frac{|\Delta f_{ij}|}{1 + \beta_{ij}}\right). \]

(5.31)

Here we use Lemma 4.1(iii) of Chapter 4 for the second term, recognizing that pdf \( f'_{ij} \) of \( \omega'_{ij} = \omega_{ij}(1 + \beta_{ij}) \) has total variation \( |\Delta f'_{ij}| = (1 + \beta_{ij})^{-1}|\Delta f_{ij}| \). Inserting the expressions of (5.13) for \( M(\hat{\varphi}_0, B) \) and \( M(-\hat{\varphi}_0, B) \) into (5.31), we have

\[ \Psi_{ij}(\beta_{ij}) \leq \frac{1}{2} (r_{ij} + c_{ij}) h(|\Delta f_{ij}|) + \frac{1}{2} (r_{ij} + c_{ij}) h\left( \frac{|\Delta f_{ij}|}{1 + \beta_{ij}} \right) \]

\[ \leq (r_{ij} + c_{ij}) h(|\Delta f_{ij}|), \]

where the second inequality holds because \( h \) is increasing and \( \beta_{ij} \geq 0 \).

Substituting the bounds on \( \Psi_{ij}(\beta_{ij}) \) into (5.30) yields

\[ G(x) - G_0(x) \leq \sum_{(i,j) \in E^* \cup E^+} (r_{ij} + c_{ij}) h(|\Delta f_{ij}|). \]

As already mentioned, the same upper bound can be obtained for \( G_0(x) - G(x) \) using
a similar line of reasoning.

5.4.3 Computational study

We use numerical experiments to evaluate the actual performance of the concave approximation, defined in (5.24), of the fleet allocation and routing problem. All experiments are carried out on the graph instance of [15], given in Figure 5.2, and the cost and reward parameters, \( r \) and \( c \), are also taken from this reference. The computational results we report use GAMS 24.2.1 and IBM ILOG CPLEX 12.6 to solve the MILPs on a Dell Poweredge 2950 computer with two dual-core Intel (Xenon) 3.73 GHz Xeon processors and 24 GB of shared memory running Ubuntu Linux.

Experimental design

We assume that all random variables, \( \omega_{ij} \), are independently distributed and follow a truncated normal distribution; i.e., \( \omega_{ij} := [\omega_{ij} | \omega_{ij} \geq 0] \), where \( \omega_{ij} \sim N(\mu_{ij}, \sigma^2) \). The mean, \( \mu_{ij} \), is the same as in [15] and differs per arc \((i,j)\), whereas the standard deviation \( \sigma \) is the same for each arc, but varies over the experiments. We also vary the investment cost parameters \( q \) by defining \( q = \kappa_q (r + c) \) and selecting different values for the scalar inflation factor, \( \kappa_q \).

We let \( \kappa_q \in \{0.2, 0.5, 0.8\} \) and \( \sigma \in \{0.1, 1, 10\} \) so that the values of \( \sigma \) correspond to low, medium, and high variability. Moreover, we consider two settings, one with additive investment effects and one with multiplicative investment effects. In the first case \( E^+=\{(1,8),(4,9),(7,13),(11,16),(14,17)\} \) and in the second case \( E^* \) is the same arc set. (Again, see Figure 5.2.) For other experiments with different arc sets we obtained similar results.

Numerical results

We evaluate the performance of the \( \alpha \)-approximation with \( \alpha = 0 \), defined in (5.24), using the same approach as in Section 5.3.2 for the integer newsvendor problem. Here, we apply the MRP with \( n = 50 \) instead of \( n = 1000 \), because of the increased computational effort required to solve the deterministic equivalent MILP. In fact, for each MILP we stop the branch-and-bound procedure at either a relative tolerance of 0.001\% or after five minutes of computation time, which ever occurs first, and we let \( x_n^* \) be the best integer solution obtained.
To obtain the approximating solution, \( x_0 \), we solve the approximating problem using a sample average approximation with a sample size of 250. The deterministic equivalent problem of this approximation is also a MILP, but it is easier to solve because the approximating model has integer variables in the first stage only. Finally, to obtain the sampling solution, \( x^S \), we compare the solutions \( x^*_n, i = 1, \ldots, N_r \), using an out-of-sample estimation with a sample of size 1000. Note that we do not consider the average of the solutions \( x^*_n \) because this average is not necessarily feasible.

We report the same performance measures \( \rho_1(x_0), \rho_2(x_0), \rho_2(x^S) \), and \( \rho_3(x_0, x^S) \) as in Section 5.3.2. However, here the denominator in \( \rho_1(x_0) \) is Theorem 5.2’s error bound. Thus, \( \rho_1(x_0) \) compares the MRP optimality gap with the total variation error bound; \( \rho_2(x_0) \) and \( \rho_2(x^S) \) compare the MRP optimality gap with the near-optimal objective value; and, \( \rho_3(x_0, x^S) \) compares the approximating solution and the sampling solution. To estimate \( G(x_0) - G(x^S) \) we use a sample of size 10,000. Even though the MILPs are not solved to optimality, we use \( x^*_n \) in Step 1(iii) of the MRP and its objective function value \( \eta^*_n \) in the denominators of \( \rho_2 \) and \( \rho_3 \). As a result, the values of \( \rho_1(x_0), \rho_2(x_0), \) and \( \rho_2(x^S) \) may be too small when the (deterministic) MILP optimality gap is not small enough. For this reason, we also report \( \Gamma \), a bound on the (deterministic) optimality gap of these MILPs as a percentage of the objective function value. Summing \( \Gamma \) and \( \rho_2 \) has the effect of replacing the objective function value of \( x^*_n \) with the MILP relaxation bound in \( \rho_2 \)’s numerator.

Table 5.5: Numerical results for fleet allocation and routing problem. The effect of investments are additive, \( E^+ = \{(1, 8), (4, 9), (7, 13), (11, 16), (14, 17)\} \), and the \( \rho \)- and \( \Gamma \)-values are reported as percentages. Because we are maximizing, positive values for \( \rho_3 \) mean that \( x_0 \) outperforms \( x^S \) and negative values mean the opposite.

<table>
<thead>
<tr>
<th>Exp.</th>
<th>( \sigma )</th>
<th>( \kappa_q )</th>
<th>( \Gamma )</th>
<th>( \rho_1(x_0) )</th>
<th>( \rho_2(x_0) )</th>
<th>( \rho_2(x^S) )</th>
<th>( \rho_3(x_0, x^S) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1</td>
<td>0.2</td>
<td>0.00</td>
<td>5.4</td>
<td>0.69</td>
<td>0.05</td>
<td>-0.63</td>
</tr>
<tr>
<td>2</td>
<td>0.1</td>
<td>0.5</td>
<td>0.03</td>
<td>15.2</td>
<td>2.51</td>
<td>0.12</td>
<td>-2.39</td>
</tr>
<tr>
<td>3</td>
<td>0.1</td>
<td>0.8</td>
<td>0.15</td>
<td>23.1</td>
<td>4.86</td>
<td>0.21</td>
<td>-4.63</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.2</td>
<td>0.07</td>
<td>11.6</td>
<td>0.20</td>
<td>0.21</td>
<td>0.02</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.5</td>
<td>0.69</td>
<td>17.6</td>
<td>0.38</td>
<td>0.39</td>
<td>0.03</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0.8</td>
<td>1.55</td>
<td>17.0</td>
<td>0.44</td>
<td>0.53</td>
<td>0.12</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>0.2</td>
<td>0.00</td>
<td>34.6</td>
<td>0.06</td>
<td>0.07</td>
<td>0.003</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>0.5</td>
<td>0.00</td>
<td>22.3</td>
<td>0.04</td>
<td>0.05</td>
<td>0.003</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>0.8</td>
<td>0.00</td>
<td>30.8</td>
<td>0.05</td>
<td>0.06</td>
<td>0.008</td>
</tr>
</tbody>
</table>

Table 5.5’s results for additive investment effects are similar to Section 5.3.2’s
results for the integer newsvendor problem. The concave approximation is good in case of medium and high variability ($\rho_2(x_0)$ is smaller than 1%, and also the values of $\Gamma$ are fairly small) and the approximation is worse in the case of low variability. Indeed, for Experiment 3 the value of $\rho_2(x_0)$ is almost 5%. Moreover, the actual performance of the concave approximation is significantly better than what the (worst-case) error bound of Theorem 5.2 suggests. Interestingly, the values of $\rho_1(x_0)$ are higher for $\sigma = 10$ than for $\sigma = 0.1$. This is caused by the fact that the error bound decreases as $\sigma$ increases, whereas the MRP optimality gap increases due to the increased variability in the model. For medium and high variability the concave approximation is as good as the sampling solution. Keeping in mind that obtaining the sampling solution $x^S$ requires much more computational effort, we prefer to use the concave approximation under these circumstances. (The typical time to compute $x_0$ with a sample size of 250 is one second, whereas the time to compute a single $x^*_n$ with a sample size of 50 often exceeds five minutes.) For low variability, however, the sampling solution is better (as can be observed from the values of $\rho_3(x_0,x^S)$), and is in fact close to optimal (since the values of $\rho_2(x^S)$ and $\Gamma$ are small). Moreover, for low variability the sampling method typically gives good solutions even if the sample size is small, so that the sampling method can be carried out within reasonable time limits. Thus, in some sense the concave approximation and the sampling method can be considered as complementary approaches: the concave approximation can be used in case of medium and high variability, and the sampling method in case of low variability.

For multiplicative investment effects we obtain similar results; see Table 5.6. The main difference is that the performance of the concave approximation is also good in the low variability case. This may be caused by the fact that for additive investment effects and $\sigma = 0.1$, the demands, $\tilde{D}_{ij}(\omega_{ij}, \beta_{ij}) = \lfloor \omega_{ij} \rfloor + \beta_{ij}$, in the approximating model are almost deterministic, whereas for multiplicative investment effects the variability in the demands, $\tilde{D}_{ij}(\omega_{ij}, \beta_{ij}) = \lfloor \omega_{ij} \rfloor + \omega_{ij} \beta_{ij}$, is larger.

5.5 Summary and conclusions

Two-stage integer recourse models can be very difficult to solve because they are non-convex. That is why we consider convex approximations for totally unimodular integer recourse models. In particular, we consider the $\alpha$-approximations of Van der Vlerk [83] and the shifted LP-relaxation approximation developed in Chapter 4. Both
Table 5.6: Numerical results for fleet allocation and routing problem. The effect of investments are multiplicative, $E^* = \{(1, 8), (4, 9), (7, 13), (11, 16), (14, 17)\}$, and the $\rho$- and $\Gamma$-values are reported as percentages. Because we are maximizing, positive values for $\rho_3$ mean that $x_0$ outperforms $x^S$ and negative values mean the opposite.

<table>
<thead>
<tr>
<th>Exp.</th>
<th>$\sigma$</th>
<th>$\kappa_\rho$</th>
<th>$\Gamma$</th>
<th>$\rho_1(x_0)$</th>
<th>$\rho_2(x_0)$</th>
<th>$\rho_2(x^S)$</th>
<th>$\rho_3(x_0, x^S)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1</td>
<td>0.2</td>
<td>0.00</td>
<td>1.28</td>
<td>0.15</td>
<td>0.06</td>
<td>-0.09</td>
</tr>
<tr>
<td>2</td>
<td>0.1</td>
<td>0.5</td>
<td>0.25</td>
<td>1.56</td>
<td>0.20</td>
<td>0.12</td>
<td>-0.09</td>
</tr>
<tr>
<td>3</td>
<td>0.1</td>
<td>0.8</td>
<td>0.60</td>
<td>5.10</td>
<td>0.75</td>
<td>0.18</td>
<td>-0.55</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.2</td>
<td>0.02</td>
<td>15.3</td>
<td>0.25</td>
<td>0.28</td>
<td>-0.01</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.5</td>
<td>0.59</td>
<td>20.2</td>
<td>0.38</td>
<td>0.43</td>
<td>0.04</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0.8</td>
<td>1.22</td>
<td>21.4</td>
<td>0.45</td>
<td>0.47</td>
<td>-0.01</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>0.2</td>
<td>0.00</td>
<td>73.1</td>
<td>0.12</td>
<td>0.13</td>
<td>0.006</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>0.5</td>
<td>0.00</td>
<td>51.1</td>
<td>0.08</td>
<td>0.09</td>
<td>0.009</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>0.8</td>
<td>0.00</td>
<td>40.6</td>
<td>0.07</td>
<td>0.07</td>
<td>0.009</td>
</tr>
</tbody>
</table>

approximations are obtained by simultaneously relaxing the second-stage integrality constraints and perturbing the distribution of the random right-hand side vector. The resulting approximating models can be considered as continuous recourse models, and can be solved efficiently by existing solution methods.

For both $\alpha$-approximations and the shifted LP-relaxation approximation there are error bounds available that depend on the total variation of the probability density functions of the random variables in the model. The smaller these total variations, the smaller the error bounds, suggesting that the performance of the approximations is better in these cases. The actual performance, however, of these approximations had not yet been investigated.

We assess the quality of the approximating solutions using sampling. To do so, we use the multiple replications procedure of Mak et al. [51], which can be used to assess the solution quality of a candidate solution for stochastic programming problems. We carry out numerical experiments on an integer newsvendor problem and a fleet allocation and routing problem. For this latter problem we derive a new error bound to deal with the deterministic flow balance constraints in the second stage, and the fact that the recourse is relatively complete instead of complete.

From these numerical experiments we conclude that the actual performance of the convex approximations is much better than their error bounds suggest, especially if the variability of the random parameters in the model is medium to large. In case this variability is small, the performance of the approximations is not as good.
However, these are precisely the cases where sampling methods may be useful so that the convex approximations and sampling methods can be considered complementary solution methods for two-stage totally unimodular integer recourse problems.