Surveys in Operations Research and Management Science 19 (2014) 17–33

Contents lists available at ScienceDirect

Surveys in Operations Research and Management Science

journal homepage: www.elsevier.com/locate/sorms

Review

Approximation in two-stage stochastic integer programming

Ward Romeijnders[a], Leen Stougie[b,c,*], Maarten H. van der Vlerk[a]

[a] University of Groningen, The Netherlands
[b] VU University, The Netherlands
[c] CWI Amsterdam, The Netherlands

HIGHLIGHTS

- We survey complexity results for stochastic (integer) programming problems.
- We survey convex approximations of stochastic integer programming objectives.
- We survey approximation guarantees for stochastic integer programming problems.

ARTICLE INFO

Article history:
Received 15 February 2013
Received in revised form
3 January 2014
Accepted 28 April 2014

ABSTRACT

Approximation algorithms are the prevalent solution methods in the field of stochastic programming. Problems in this field are very hard to solve. Indeed, most of the research in this field has concentrated on designing solution methods that approximate the optimal solution value. However, efficiency in the complexity theoretical sense is usually not taken into account. Quality statements mostly remain restricted to convergence to an optimal solution without accompanying implications on the running time of the algorithms for attaining more and more accurate solutions.

However, over the last thirty years also some studies on performance analysis of approximation algorithms for stochastic programming have appeared. In this direction we find both probabilistic analysis and worst-case analysis.

Recently the complexity of stochastic programming problems has been addressed, indeed confirming that these problems are harder than most deterministic combinatorial optimization problems. Polynomial-time approximation algorithms and their performance guarantees for stochastic linear and integer programming problems have received increasing research attention only very recently.

Approximation in the traditional stochastic programming sense will not be discussed in this chapter. The reader interested in this issue is referred to surveys on stochastic programming, like the Handbook on Stochastic Programming by Ruszczyński and Shapiro (2003) or the textbooks by Birge and Louveaux (1997), Kall and Wallace (1994), Prékopa (1995), and Shapiro et al. (2009). We concentrate on the studies of approximation algorithms in relation to computational complexity theory.

With this survey we intend to give a flavor of the type of results existing in the literature on approximation algorithms in two-stage stochastic integer programming rather than a complete overview of the literature on the subject. We do so by exhibiting a representative selection of results, which we present in full detail. While presenting them we do not refer to the literature; these references, together with pointers to other relevant work in this field of research, are given in an extensive notes section at the end of the survey.

Contents

* Corresponding author at: VU University, The Netherlands. Tel.: +31 20 59 89391.
E-mail addresses: w.romeijnders@rug.nl (W. Romeijnders), l.stougie@vu.nl, stougie@cwi.nl (L. Stougie), m.h.van.der.vlerk@rug.nl (M.H. van der Vlerk).

http://dx.doi.org/10.1016/j.sorms.2014.04.001

1. Introduction

Stochastic programming models arise as reformulations or extensions of optimization problems with random parameters. To set the stage for our review of approximation in stochastic (integer) programming, we first introduce the models and give an overview of relevant mathematical properties.

Consider the optimization problem:

$$\begin{array}{ll} \min\limits_{x} & cx \\ \text{s.t.} & Ax = b \\ & \boldsymbol{T}x = \boldsymbol{h} \\ & x \in X, \end{array}$$

where $X \subset \mathbb{R}^n_+$ specifies nonnegativity of and possibly integrality constraints on the decision variables x. In addition to the m_1 deterministic constraints $Ax = b$, there is a set of m constraints $\boldsymbol{T}x = \boldsymbol{h}$, whose parameters $\boldsymbol{T}$ and $\boldsymbol{h}$ depend on information which becomes available only after a decision x is made. The *stochastic programming* approach to such problems is to assume that this uncertainty can be modeled by random variables with known probability distribution, and then to reformulate the model to obtain a meaningful and well-defined optimization problem. In this survey we will use bold face characters for random variables, and plain face to indicate their realizations.

1.1. Stochastic programming models

An important class of stochastic programming problems consists of *probabilistic or chance-constrained problems*, which model random constraints[1] by requiring that they should be satisfied with some prescribed reliability $\alpha \in [0, 1]$; typically, $\alpha \in (0.5, 1)$. Thus, the random constraints $\boldsymbol{T}x \geq \boldsymbol{h}$ are replaced by the *joint chance constraint*

$$\Pr\{\boldsymbol{T}x \geq \boldsymbol{h}\} \geq \alpha,$$

or by m *individual chance constraints*

$$\Pr\{\boldsymbol{T}_i x \geq \boldsymbol{h}_i\} \geq \alpha_i, \quad i = 1, \dots, m.$$

Since we will not consider chance-constrained models in our discussion of approximation results, we do not present them in more detail here. An introduction on the subject by Henrion can be found in [1], and the book by Prékopa [2] provides an excellent survey of these models.

This survey considers *recourse models*, the other main class of stochastic programming models. Recourse models are obtained by allowing additional or recourse decisions after observing the realizations of the random variables $(\boldsymbol{T}, \boldsymbol{h})$. Thus, recourse models are dynamic: time is modeled discretely by means of stages, corresponding to the available information. If all uncertainty is dissolved at the same moment, this is captured by a recourse model with two stages: 'present' and 'future'. Given a first-stage decision x, for every possible realization q, T, h of $\boldsymbol{q}, \boldsymbol{T}, \boldsymbol{h}$, infeasibilities $h - Tx$ are compensated at minimal costs by choosing second-stage decisions y as an optimal solution of the second-stage problem

$$\begin{array}{ll} \min\limits_{y} & qy \\ \text{s.t.} & Wy = h - Tx, \\ & y \in Y, \end{array}$$

where $\boldsymbol{q}$ is the (random) recourse unit cost vector, the recourse matrix W specifies the available technology, and the set $Y \subset \mathbb{R}^{n_2}_+$ is defined analogously to X. We will use $\boldsymbol{\xi} = (\boldsymbol{q}, \boldsymbol{T}, \boldsymbol{h})$ to denote the random object representing all randomness in the problem. The *value function* of this second-stage problem, specifying the minimal recourse costs as a function of the first-stage decision x and a realization ξ of $\boldsymbol{\xi}$, will be denoted by $v(x, \xi)$; its expectation $Q(x) := \mathbb{E}_{\boldsymbol{\xi}}[v(x, \boldsymbol{\xi})]$ gives the expected recourse costs associated with a first-stage decision x. Thus, the two-stage recourse model is

$$\begin{array}{ll} \min\limits_{x} & cx + Q(x) \\ \text{s.t.} & Ax = b \\ & x \in X, \end{array} \tag{1}$$

where the objective function $cx + Q(x)$ specifies the total expected costs of a decision x.

Example 1.1. Consider the following production planning problem. Using n production resources, denoted by $x \in \mathbb{R}^n_+$ with corresponding unit cost vector c, a production plan needs to be made such that the uncertain future demand for m products, denoted by $\boldsymbol{h} \in \mathbb{R}^m$, is satisfied at minimal costs. The available production technology suffers from failures: deploying resources x yields uncertain amounts of products $\boldsymbol{T}_i x$, $i = 1, \dots, m$. Restrictions on the use of x are captured by the constraints $Ax = b$.

We assume that the uncertainty about future demand and the production technology can be modeled by the random matrix $(\boldsymbol{T}, \boldsymbol{h})$, whose joint distribution is known, for example based on historical data.

A possible two-stage recourse model for this problem is based on the following extension of the model. For each of the individual products, if the demand h_i turns out to be larger than the production $T_i x$, the demand surplus $h_i - T_i x$ is bought from a competitor at unit costs q_i^1. On the other hand, a demand shortage

[1] Barring uninteresting cases, chance constraints make sense only for inequality constraints.

gives rise to storage costs of q_i^2 per unit. The corresponding second-stage problem and its value function are

$$\begin{aligned} v(x,\xi) = \min_y \quad & q^1 y^1 + q^2 y^2 \\ \text{s.t.} \quad & y^1 - y^2 = h - Tx, \\ & y = (y^1, y^2) \in \mathbb{R}_+^{2m}. \end{aligned}$$

Defining Q as the expectation of this value function, we obtain a two-stage recourse model that fits the general form (1).

This particular model type with recourse matrix $W = (I_m, -I_m)$, where I_m is the m-dimensional identity matrix, is known as a simple recourse model. The integer recourse version of this model, for example corresponding to the case that only batches of fixed size can be bought, will be discussed in Section 3. $\lhd$

So far, we have introduced the recourse concept as a modeling tool to handle random constraints by means of specifying recourse actions with corresponding recourse costs. There is however another class of problems for which the (two-stage) recourse model is a natural approach, namely *hierarchical planning models* (HPM). Such problems involve decisions at two distinct levels: *strategic* decisions which have a long-term impact, and *operational* decisions which are depending on the strategic decisions. At the time that the strategic decisions need to be made, only probabilistic information on the operational level problems is available. Hierarchical planning models fit the structure of two-stage recourse models, with strategic and operational decisions corresponding to first-stage and second-stage variables, respectively. Despite the differences in interpretation we use the generic name (two-stage) recourse model to refer to both model types, in line with the stochastic programming literature.

In many applications new information becomes available at several distinct moments, say $t = 1, \ldots, H$, where H is the planning horizon. That is, we assume that realizations of random vectors $\boldsymbol{\xi}^t = (\boldsymbol{q}^t, \boldsymbol{T}^t, \boldsymbol{h}^t)$ become known at time t. This can be modeled explicitly using a *multistage recourse* structure: for each such moment $t = 1, \ldots, H$, a time stage with corresponding recourse decisions is defined. In compact notation, the multistage recourse model is

$$\begin{aligned} \min_{x^0} \quad & cx^0 + Q^1(x^0) \\ \text{s.t.} \quad & Ax^0 = b \\ & x^0 \in X, \end{aligned}$$

where the functions Q^t, $t = 1, \ldots, H$, representing expected recourse costs, are recursively defined as

$$Q^t(x^{t-1}) := \mathbb{E}_{\boldsymbol{\xi}^t}\left[v^t(x^{t-1}, \boldsymbol{\xi}^t) \mid \xi^1, \ldots, \xi^{t-1}\right],$$

where the expectation is with respect to the conditional distribution of $\boldsymbol{\xi}^t$ given $\xi^1, \ldots, \xi^{t-1}$,

$$\begin{aligned} v^t(x^{t-1}, \xi^t) := \min_{x^t} \quad & q^t x^t + Q^{t+1}(x^t) \\ \text{s.t.} \quad & W^t x^t = h^t - T^t x^{t-1} \\ & x^t \in X^t, \end{aligned}$$

and $Q^{H+1} \equiv 0$ (or some other suitable choice). In this survey we concentrate on two-stage problems, and touch on multistage problems only in remarks.

1.2. Mathematical properties

In this subsection, we review mathematical properties of recourse models. This provides the background and motivation for the discussion of many approximation results.

First we consider properties of continuous recourse models. Some of the results will be used when we discuss the complexity of this problem class, and furthermore they facilitate the subsequent discussion of properties of mixed-integer recourse models. We state all properties here without proof. References to the proofs are given in the Notes section at the end of this survey.

Remark 1.1. As before, all models are discussed here in their canonical form, i.e., all constraints are either equalities or nonnegativities. The models in subsequent sections, which also contain inequalities and/or simple bounds, can be written in canonical form using standard transformations.

1.2.1. Continuous recourse

Properties of (two-stage) recourse models follow from those of the recourse function Q and the underlying value function v. In case all second-stage variables are continuous, properties of the value function v are well-known from duality and perturbation theory for linear programming, and are summarized here for easy reference.

Lemma 1.1. *The function v, defined for $x \in \mathbb{R}^n$ and $\xi = (q, T, h) \in \mathbb{R}^{n_2+m(n+1)}$ as*

$$v(x,\xi) = \inf\left\{qy : Wy = h - Tx,\ y \in \mathbb{R}_+^{n_2}\right\},$$

takes values in $[-\infty, \infty]$.

It is a convex polyhedral function of x for each $\xi \in \mathbb{R}^{n_2+m(n+1)}$, and it is concave polyhedral in q and convex polyhedral in (h, T) for all $x \in \mathbb{R}^n$. $\square$

If for some x the function v takes on the value $+\infty$ with positive probability, this means that x is extremely unattractive since it has infinitely high expected recourse costs $Q(x)$. From a modeling point of view this is not necessarily a problem, but in practice it may be desirable to exclude this situation.

On the other hand, the situation that $v(x, \xi)$ equals $-\infty$ with positive probability should be excluded altogether. Indeed, the value $-\infty$ indicates that the model does not adequately represent our intention, which is penalization of infeasibilities.

Finiteness of v is often guaranteed by assuming that the recourse is *complete* and *sufficiently expensive*.

Definition 1.1. The recourse is *complete* if $v < +\infty$, i.e., if for all $t \in \mathbb{R}^m$ there exists a $y \in Y$ such that $Wy = t$.

Assuming that $Y = \mathbb{R}_+^{n_2}$, completeness is a property of the recourse matrix W only. Such a matrix is called a *complete recourse matrix*.

Definition 1.2. The recourse is *sufficiently expensive* if $v > -\infty$ with probability 1, i.e., if the dual feasible region $\{\lambda \in \mathbb{R}^m : \lambda W \leq \boldsymbol{q}\}$ is nonempty with probability 1.

For example, the recourse is sufficiently expensive if $\Pr\{\boldsymbol{q} \geq 0\} = 1$.

Unless stated otherwise we assume that the recourse is complete and sufficiently expensive. Then the recourse or expected value function $Q(x)$ is finite if the distribution of ξ satisfies the following condition:

For all i, j, k the random functions $\boldsymbol{q}_j \boldsymbol{h}_i$ and $\boldsymbol{q}_j \boldsymbol{T}_{ik}$ have finite expectations.

Sufficiency of this *weak covariance condition* [3] follows from the representation of basic feasible solutions in terms of the problem parameters.

The following properties of the recourse function Q are inherited from the second-stage value function v.

Theorem 1.1. *Consider the recourse function Q, defined by*

$$Q(x) = \mathbb{E}_{\xi}\left[\inf\left\{\boldsymbol{q}y : Wy = \boldsymbol{h} - \boldsymbol{T}x,\ y \in \mathbb{R}_+^{n_2}\right\}\right], \quad x \in \mathbb{R}^n.$$

Assume that the recourse is complete and sufficiently expensive.

(a) *The function Q is convex, finite, and (Lipschitz) continuous.*
(b) *If $\boldsymbol{\xi}$ follows a finite discrete distribution, then Q is a convex polyhedral function.*
(c) *The function Q is subdifferentiable, with subdifferential*

$$\partial Q(x) = \int_{\Xi} \partial v(x, \xi)\, dF(\xi), \quad x \in \mathbb{R}^n,$$

where F is the cumulative distribution function (cdf) of the random vector $\boldsymbol{\xi}$.

If $\boldsymbol{\xi}$ follows a continuous distribution, then Q is continuously differentiable. □

Consider the special case that $\boldsymbol{\xi}$ follows a finite discrete distribution specified by $\Pr\left\{\boldsymbol{\xi} = (q^k, T^k, h^k)\right\} = p^k, k = 1, \ldots, K$. The finitely many possible realizations (q^k, T^k, h^k) of the random parameters are also called *scenarios*. It is easy to see that in this case the two-stage recourse model is equivalent to the large-scale linear programming problem

$$\begin{aligned} \min \quad & cx + \sum_{k=1}^{K} p^k q^k y^k \\ \text{s.t.} \quad & Ax = b \\ & T^k x + W y^k = h^k, \quad k = 1, \ldots, K \\ & x \in \mathbb{R}^n_+, \qquad y^k \in \mathbb{R}^{n_2}_+. \end{aligned} \tag{2}$$

Analogously, a mixed-integer recourse problem with finite discrete distribution can be represented as a deterministic large-scale mixed-integer programming problem.

1.2.2. Mixed-integer recourse

Mixed-integer recourse models do not possess such nice mathematical properties; in particular, convexity of the recourse function Q is not guaranteed. Indeed, the underlying second-stage value function v is only lower semicontinuous (assuming rationality of the recourse matrix W), and discontinuous in general.

Also in this setting we are mostly interested in the case that v is finite. To have $v < +\infty$ we will assume complete recourse, see Definition 1.1. For example, this condition is satisfied if $\bar{W}$ is a complete recourse matrix, where $\bar{W}$ consists of the columns of W corresponding to the continuous second-stage variables. On the other hand, $v > -\infty$ if the recourse is sufficiently expensive, see Definition 1.2, i.e., if the dual of the LP relaxation of the second-stage problem is feasible with probability 1.

Theorem 1.2. *Consider the mixed-integer recourse function Q, defined by*

$$Q(x) = \mathbb{E}_{\boldsymbol{\xi}}\left[\inf\{\boldsymbol{q}y : Wy = \boldsymbol{h} - \boldsymbol{T}x,\ y \in Y\}\right], \quad x \in \mathbb{R}^n,$$

where $Y := \mathbb{Z}^p_+ \times \mathbb{R}^{n_2-p}_+$. Assume that the recourse is complete and sufficiently expensive, and that ξ satisfies the weak covariance condition. Then

(a) *The function Q is lower semicontinuous on $\mathbb{R}^n$.*
(b) *Let $D(x), x \in \mathbb{R}^n$, denote the set containing all (h, T) such that $h - Tx$ is a discontinuity point of the mixed-integer value function v. Then Q is continuous at x if $\Pr\{(\boldsymbol{h}, \boldsymbol{T}) \in D(x)\} = 0$.*

In particular, if ξ is continuously distributed, then Q is continuous on $\mathbb{R}^n$. □

1.3. Outline

As mentioned above, solving stochastic programming problems is very difficult in general. Indeed, such problems are defined in terms of expectations of value functions of linear (mixed-integer) programming problems or indicator functions (in the case of chance constraints). This calls for the evaluation of multi-dimensional integrals, which is computationally challenging already if the underlying random vector ξ has low dimension, and becomes a formidable task for problems of realistic size. Even if the underlying distribution is discrete, the typically huge number of possible realizations may render the frequent evaluation of function values impracticable. In Section 2 the computational complexity of two-stage recourse models is addressed, showing that stochastic programming problems are considerably more difficult to solve than their deterministic counterparts.

It is therefore not surprising that much of the stochastic programming literature is devoted to approximation of some sorts. For example, a key issue for recourse models is the construction of suitable discrete approximations of the distribution of the underlying random vector. Such an approximation should have a relatively small number of possible realizations, and at the same time result in a good approximation of the recourse function, at least in a neighborhood of an optimal solution. For chance-constrained problems such discrete approximations of the distribution would destroy convexity of the problem, even if it exists. In this context, fast and accurate approximation of high-dimensional (normal) distribution functions receives much research attention.

We do not discuss these ‘typical’ stochastic programming approximation issues here. They, as well as related subjects such as convergence and stability, are covered in the Handbook on Stochastic Programming [4]. Instead, we consider approximations as they appear in a number of other ways in stochastic programming and which are in spirit closer to approximation in deterministic combinatorial optimization, i.e., their analysis aiming at worst-case performance guarantees of polynomial-time algorithms.

Section 3 deals with convex approximations for integer recourse problems. Here the problems themselves are approximated by perturbing the distributions such as to achieve convex expected value functions. The strength of this approximation is that a bound on the *absolute error* of the approximate solution value can be derived, making this an example of *worst-case analysis of approximation algorithms*. This research remains closest in nature to the classical research in the stochastic programming community, deriving algorithms based on mathematical properties of approximations of the expected value function.

In Section 4 we switch to research that is in spirit closer to the theoretical computer science and combinatorial optimization community. We will give some examples of polynomial-time approximation algorithms for two-stage stochastic combinatorial optimization problems and provide bounds on their *worst-case performance ratio* between optimal and approximate solution value. There will be a crucial distinction in the way the stochastic information is represented in the model. A basic technique for designing approximation algorithms for such problems is formulating them as stochastic (mixed) integer programming problems, finding approximate solutions for the continuous relaxations of these problems and applying an appropriate rounding technique to arrive at integer solutions.

Section 5 contains bibliographical notes on approximation in stochastic programming as reviewed in this survey. It also addresses some interesting open problems and new research directions in this field, major parts of which are still unexplored.

2. Complexity of two-stage stochastic programming problems

In this section we study the complexity of two-stage stochastic programming problems. The complexity of a problem, in terms of time or space to solve it, is related to its input size. For each instance, a bound on the number of elementary computer operations or storage units required to solve the problem instance as a function of the size of its input indicates, respectively, the time or space complexity of the problem. For an easily accessible introduction into computational complexity theory we refer to the first chapters of [5]. We will see that the way in which the random parameters in stochastic programming problems are described may have a crucial impact on the complexity. In this survey we consider three models of describing randomness.

We start by studying problem (2), the deterministic equivalent LP formulation of the two-stage stochastic programming problem. This model is also sometimes referred to as *the scenario-based model*. If in the input of the problem each scenario (q^k, T^k, h^k) and its corresponding probability p^k is specified *separately*, then the input size of the problem is just the size of the binary encoding of all the parameters in this (large-scale) deterministic equivalent problem and hence the problem is polynomially solvable in case the decision variables are continuous and NP-complete in general if there are integrality constraints on decision variables.

In *the independent model*, all parameters are independently distributed random variables. For example, if in this case each parameter has value a_1 with probability p and a_2 with probability $1 - p$, then there are $K = 2^{n_1+mn+m}$ possible scenarios. Hence, the size of the deterministic equivalent problem is exponential in the dimension of the parameter space, which is essentially the size required to encode the input. The complexity changes correspondingly, as will become clear below. Although it will not surprise anybody that an exponential number of scenarios indeed leads to higher complexity classes, the mere fact that the number of scenarios is exponential in the input size is not sufficient to conclude the correct complexity right away. Indeed, Section 4 presents algorithms for problems in this class that in polynomial time are yielding more or less good quality approximations in spite of the exponential number of scenarios.

The third and most extreme representation of uncertainty in stochastic programming is based on the definition of a sample oracle. In this case the probability distribution of the parameters is not specified and as such not counted as part of the input. The only way to get information on these parameters is by random sampling through an oracle. Each oracle call is regarded as a basic computer operation in determining the time complexity. In the literature this model is often referred to as the *black box model*.

We will concentrate in this section on models of the second type, having all random (second-stage) parameters independently and discretely distributed. We will establish $\sharp P$-hardness of the evaluation of the second-stage expected value function $Q(x)$ for fixed x. The class $\sharp P$ consists of counting problems, for which membership to the set of items to be counted can be verified in polynomial time. Examples of such problems are counting the number of Hamiltonian circuits in a given graph, or counting the number of basic feasible solutions to a given linear programming problem. We notice that strictly following this definition of $\sharp P$, none of the stochastic programming problems can belong to this complexity class. However, we will use the term $\sharp P$-hard for an optimization problem in the same way as NP-hardness is used for optimization problems whose recognition version is NP-complete. For an exposition of the definitions and structures of the various complexity classes we refer to [6].

To prove $\sharp P$-hardness of the evaluation of the second-stage expected value function $Q(x)$ we use a reduction from the $\sharp P$-complete problem GRAPH RELIABILITY.

Definition 2.1 (GRAPH RELIABILITY)**.** Given a directed graph with m arcs and n vertices, determine the reliability of the graph, defined as the probability that two given vertices u and v are connected if each edge fails independently with probability $1/2$.

This is equivalent to the problem of counting the number of subgraphs, from among all 2^m possible subgraphs, that contain a path from u to v.

Theorem 2.1. *Two-stage stochastic programming with discretely distributed parameters is $\sharp P$-hard.*

Proof. Take any instance of GRAPH RELIABILITY, i.e., a network $G = (V, A)$ with two prefixed nodes u and v in V. Introduce an extra arc from v to u, and introduce for each arc $(i, j) \in A$ a variable y_{ij}. Give each arc a random weight $\mathbf{q}_{ij}$ except for the arc (v, u) that gets a deterministic weight of 1. Let the weights be independent and identically distributed (i.i.d.) with distribution $Pr\{\mathbf{q} = -2\} = Pr\{\mathbf{q} = 0\} = 1/2$. The event $\{q = -2\}$ corresponds to failure of the arc in the GRAPH RELIABILITY instance. If for a realization of the failures of the arcs the network has a path from u to v, then it implies that there is a path from u to v consisting of arcs with weight 0 only and vice versa.

Denote $A' = A \cup (v, u)$. Now define the two-stage stochastic programming problem:

$$\max\{-cx + Q(x) \mid 0 \le x \le 1\},$$

where c is a parameter and

$$Q(x) = \mathbb{E}_{\mathbf{q}}[\varphi(x, \mathbf{q})],$$

with

$$\begin{aligned} \varphi(x, q) = \max_{y} \quad & \sum_{(i,j)\in A} q_{ij} y_{ij} + y_{vu} \\ \text{s.t.} \quad & \sum_{i:(i,j)\in A'} y_{ij} - \sum_{k:(j,k)\in A'} y_{jk} = 0, \quad \forall j \in V, \\ & y_{ij} \le x, \quad \forall (i,j) \in A. \end{aligned}$$

Suppose that for a realization of the failures of the arcs there is a path from u to v in the network. As we argued the costs $q_{ij} = 0$ for those coefficients (i, j) that correspond to arcs on the path. For such a realization, the optimal solution of the second-stage problem, is obtained by setting all y_{ij}'s corresponding to arcs (i, j) on this path and y_{vu} equal to x, their maximum feasible value, and setting $y_{ij} = 0$ for all (i, j) not on the path. This yields solution value x for this realization.

Suppose that for a realization the graph does not have a path from u to v, implying in the reduced instance that on each path there is an arc with weight -2 and vice versa, then the optimal solution of the realized second-stage problem is obtained by setting all y_{ij}'s equal to 0, and henceforth also $y_{vu} = 0$, yielding solution value 0.

Therefore, the network has reliability R if and only if $Q(x) = Rx$ and hence the objective function of the two-stage problem is $(R - c)x$.

Thus, if $c \le R$, then the optimal solution is $x = 1$ with value $(R-c)$, and if $c \ge R$, then the optimal solution is $x = 0$ with value 0. Since R can take only 2^m possible values, by performing a bisection search we can compute the exact value of R by solving only m two-stage stochastic programming problems to know the exact value of R. Thus, solving this two-stage stochastic programming problem implies solving the $\sharp P$-hard GRAPH RELIABILITY problem. □

By total unimodularity of the restriction coefficients matrix [7] in the proof, the same reduction shows that the TWO-STAGE STOCHASTIC INTEGER PROGRAMMING problem with discretely distributed parameters, i.e., the problem in which second-stage decision variables are restricted to have integer values is $\sharp P$-hard. We notice that imposing restrictions $y_{ij} \in \{0, 1\}$ on all second-stage variables will give $Q(x) = R\lfloor x \rfloor$. Thus, evaluating $Q(x)$ is only $\sharp P$-hard when $x = 1$, but finding the optimal value for x is still $\sharp P$-hard.

In the two-stage linear programming problem, evaluation of Q at any point x is $\sharp P$-*easy* since for any realization of the second-stage random parameters a linear program remains to be solved. Given a $\sharp P$-oracle for evaluating Q at any point x, solving two-stage stochastic linear programming problems (with discretely distributed random variables) will require a polynomial number of consultations of the oracle, since Q is a concave function in x, and maximizing a concave function over a convex set is known to be easy [8]. Thus, two-stage stochastic linear programming is in the class $P^{\sharp P} = \sharp P$ [6].

Given a $\sharp P$-oracle for evaluating Q at any point x, a two-stage stochastic integer programming problem lies in NP. In this case the expected value function is not convex (see [9]), but there are a finite number of points x that are candidates for optimality. Thus, two-stage stochastic integer programming is in the class $NP^{\sharp P}$.

In case the random parameters of the two-stage stochastic programming problem are continuously distributed, the evaluation of the function Q in a single point of its domain requires the computation of a multi-dimensional integral. Most of the stochastic programming literature on this subclass of problems is concerned with how to get around this obstacle. We give the complexity of this class of problems without proof.

Theorem 2.2. *Two-stage stochastic programming with continuously distributed parameters is $\sharp P$-hard.* □

The membership of this problem to $\sharp P$ requires additional conditions on the input distributions, since exact computation may not even be in *PSPACE*.

3. Convex approximations for integer recourse problems

In this section we consider convex approximations for pure integer recourse problems. For such problems, the second-stage integer program is necessarily defined using only inequalities. Moreover, in all integer recourse models that we consider here, only the right-hand side vector $\boldsymbol{h}$ is random. The second-stage value function is thus

$$\begin{aligned} v(x,h) := \min_y \quad & qy \\ \text{s.t.} \quad & Wy \geq h - Tx, \\ & y \in \mathbb{Z}_+^{n_2}, \quad x \in \mathbb{R}^n,\ h \in \mathbb{R}^m, \end{aligned}$$

where the components of W are assumed to be integers. Assuming complete and sufficiently expensive recourse, v is a finite, discontinuous, piecewise constant function; in particular, v is non-convex. It follows from Theorem 1.2 that the integer recourse function $Q(x) = \mathbb{E}_{\boldsymbol{h}}[v(x,\boldsymbol{h})]$, $x \in \mathbb{R}^n$, is continuous if $\boldsymbol{h}$ is continuously distributed, but in general Q is non-convex.

However, for certain integer recourse models, characterized by their recourse matrices W, a class of distributions of $\boldsymbol{h}$ is known such that the corresponding recourse function Q is convex. Thus, for such integer recourse models we can construct convex approximations of the function Q by approximating any given distribution of $\boldsymbol{h}$ by a distribution belonging to this special class.

Below we first apply this approach to the simple integer recourse model. Subsequently, we consider general complete integer recourse models, starting with the case of totally unimodular recourse matrices.

3.1. Simple integer recourse

The simple integer recourse second-stage problem is defined as

$$\begin{aligned} \min_y \quad & q^+y^+ + q^-y^- \\ \text{s.t.} \quad & y^+ \geq h - Tx, \\ & y^- \geq -(h - Tx), \\ & y^+,\quad y^- \in \mathbb{Z}_+^m, \end{aligned}$$

where the indices $\cdot^+$ and $\cdot^-$ are conventionally used to indicate surplus and shortage, respectively. This recourse structure is obviously complete, and it is sufficiently expensive if $q^+ \geq 0$ and $q^- \geq 0$ (componentwise), as will be assumed from now on.

It is trivial to find a closed form for the simple integer recourse value function. Due to the simple recourse structure, this function is separable in the *tender variables* $z := Tx$:

$$v(z,h) = \sum_{i=1}^m v_i(z_i,h_i), \quad z,h \in \mathbb{R}^m,$$

where

$$v_i(z_i,h_i) = q_i^+\lceil h_i - z_i\rceil^+ + q_i^-\lfloor h_i - z_i\rfloor^-, \tag{3}$$

with $\lceil s\rceil^+ := \max\{0, \lceil s\rceil\}$ and $\lfloor s\rfloor^- := \max\{0, -\lfloor s\rfloor\}$, $s \in \mathbb{R}$. Since all functions v_i have the same structure, we restrict the presentation to one such function, and drop the index. It is straightforward to translate the results below back to the full-dimensional case.

Given the closed form (3), it follows that the one-dimensional generic simple integer recourse function Q equals

$$Q(z) = q^+\mathbb{E}_{\boldsymbol{h}}\left[\lceil \boldsymbol{h} - z\rceil^+\right] + q^-\mathbb{E}_{\boldsymbol{h}}\left[\lfloor \boldsymbol{h} - z\rfloor^-\right], \quad z \in \mathbb{R}, \tag{4}$$

where $\boldsymbol{h} \in \mathbb{R}$ is a random variable. Throughout we assume that $\mathbb{E}_{\boldsymbol{h}}[|\boldsymbol{h}|]$ is finite, which is necessary and sufficient for finiteness of the function Q. We state the following results without proof.

Lemma 3.1. *Consider the one-dimensional simple integer recourse function Q defined in (4).*

(a) *For all $z \in \mathbb{R}$,*

$$Q(z) = q^+\sum_{k=0}^{\infty}\Pr\{\boldsymbol{h} > z + k\} + q^-\sum_{k=0}^{\infty}\Pr\{\boldsymbol{h} < z - k\}.$$

(b) *Assume that $\boldsymbol{h}$ has a probability density function (pdf) f that is of bounded variation. Then the right derivative Q'_+ exists everywhere, and is given by*

$$Q'_+(z) = -q^+\sum_{k=0}^{\infty}f_+(z+k) + q^-\sum_{k=0}^{\infty}f_+(z-k), \quad z \in \mathbb{R},$$

where f_+ is the right-continuous version of f. □

The function Q is generally non-convex because the right derivative Q'_+ may be decreasing on some interval. However, there exists a class of distributions of $\boldsymbol{h}$ such that Q is convex. This class of distributions is defined in the theorem below.

Theorem 3.1. *The one-dimensional simple recourse function Q is convex if and only if the underlying random variable $\boldsymbol{h}$ is continuously distributed with a pdf f that is of bounded variation, such that*

$$f_+(s) = G(s+1) - G(s), \quad s \in \mathbb{R}, \tag{5}$$

where G is an arbitrary cdf with finite mean value. □

Sufficiency of (5) is easy to see, since it implies that

$$Q'_+(z) = -q^+(1 - G(z)) + q^-G(z+1), \quad z \in \mathbb{R}, \tag{6}$$

is non-decreasing. Below we will make extensive use of the following special case.

Corollary 3.1. *Assume that $\boldsymbol{h}$ is continuously distributed with a pdf f whose right-continuous version is constant on every interval $[\alpha + k, \alpha + k + 1)$, $k \in \mathbb{Z}$, for some $\alpha \in [0,1)$. Then the function Q is piecewise linear and convex, with knots contained in $\alpha + \mathbb{Z}$.*

Proof. Immediate from Theorem 3.1 and (6), since $f_+(s) = G(s+1) - G(s)$ where G is the cdf of a discrete distribution with support contained in $\alpha + \mathbb{Z}$. □

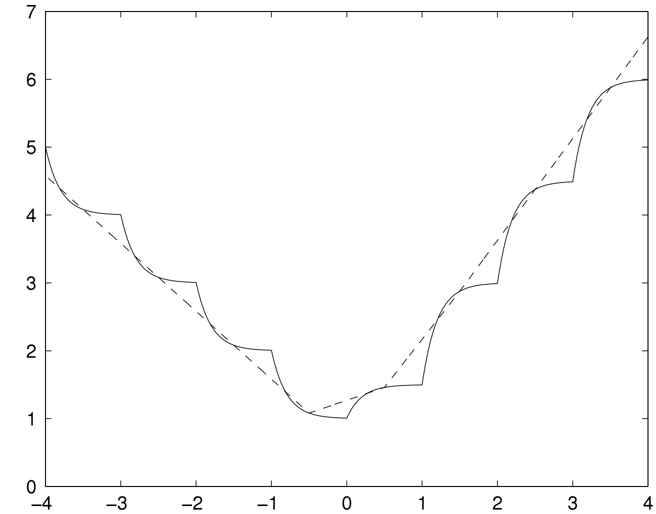

Fig. 1. The function Q and its α-approximation Q_α (dashed) in case $\boldsymbol{h}$ is exponentially distributed with parameter 5, $q^+ = 1$, $q^- = 1.5$, and $\alpha = 0.5$.

To arrive at convex approximations of the function Q, we will use Corollary 3.1 to construct suitable approximations of the distribution of the random variable $\boldsymbol{h}$. For future reference, we present the multivariate definition of the approximations that we have in mind.

Definition 3.1. Let $\boldsymbol{h} \in \mathbb{R}^m$ be a random vector with arbitrary continuous or discrete distribution, and let $\alpha = (\alpha_1, \dots, \alpha_m) \in [0, 1)^m$. Define the *$\alpha$-approximation* $\boldsymbol{h}_\alpha$ as the random vector with joint pdf f_α that is constant on every hypercube $C_\alpha^k := \prod_{i=1}^m (\alpha_i + k_i - 1, \alpha_i + k_i]$, $k \in \mathbb{Z}^m$, such that $\Pr\left\{\boldsymbol{h}_\alpha \in C_\alpha^k\right\} = \Pr\left\{\boldsymbol{h} \in C_\alpha^k\right\}$, $k \in \mathbb{Z}^m$.

Returning to the one-dimensional case, it is easy to see that the α-approximations $\boldsymbol{h}_\alpha$, $\alpha \in [0, 1)$, of an arbitrary random variable $\boldsymbol{h}$, satisfy the assumptions of Corollary 3.1. It follows that the *α-approximations* of the function Q, defined for $\alpha \in [0, 1)$,

$$Q_\alpha(z) := q^+ \mathbb{E}_{\boldsymbol{h}_\alpha}\left[\lceil \boldsymbol{h}_\alpha - z\rceil^+\right] + q^- \mathbb{E}_{\boldsymbol{h}_\alpha}\left[\lfloor \boldsymbol{h}_\alpha - z\rfloor^-\right], \quad z \in \mathbb{R},$$

are piecewise linear convex approximations of Q, with knots contained in $\alpha + \mathbb{Z}$. Moreover, it follows from Lemma 3.1(a) and Definition 3.1 that

$$Q_\alpha(z) = Q(z), \quad z \in \alpha + \mathbb{Z}.$$

We conclude that, for each $\alpha \in [0, 1)$, Q_α is the piecewise linear convex function generated by the restriction of Q to $\alpha + \mathbb{Z}$. See Fig. 1 for an example of the function Q and one of its α-approximations.

In the discussion above, no assumptions were made on the type of distribution of $\boldsymbol{h}$. However, to establish a non-trivial bound on the approximation error, we need to assume that $\boldsymbol{h}$ is continuously distributed. This loss of generality is acceptable, because it is possible to construct the convex envelope of the function Q if $\boldsymbol{h}$ is discretely distributed [10].

Theorem 3.2. *Assume that $\boldsymbol{h}$ is continuously distributed with a pdf f that is of bounded variation. Then, for all $\alpha \in [0, 1)$,*

$$\|Q_\alpha - Q\|_\infty \le (q^+ + q^-)\frac{|\Delta|f}{4},$$

where $|\Delta|f$ denotes the total variation of f.

Proof. We will sketch a proof for the special case that $q^+ = 1$ and $q^- = 0$. The proof for the general case is analogous.

Assume that $q^+ = 1$ and $q^- = 0$. Then the function Q reduces to the expected surplus function $g(z) := \mathbb{E}_{\boldsymbol{h}}\left[\lceil \boldsymbol{h} - z\rceil^+\right]$, $z \in \mathbb{R}$, with α-approximations $g_\alpha(z) := \mathbb{E}_{\boldsymbol{h}_\alpha}\left[\lceil \boldsymbol{h}_\alpha - z\rceil^+\right]$, $\alpha \in [0, 1)$. Since $g(z) = g_\alpha(z)$ if $z \in \alpha + \mathbb{Z}$, consider an arbitrary fixed $z \notin \alpha + \mathbb{Z}$, and let $\underline{z} \in \alpha + \mathbb{Z}$ be such that $\underline{z} < z < \underline{z} + 1$.

Using Lemma 3.1(b) and the fact that g is continuous if $\boldsymbol{h}$ is continuously distributed we find that

$$g(\underline{z}) - g(z) = \int_{\underline{z}}^{z} \sum_{k=0}^{\infty} f(t+k)dt.$$

It follows from Lemma 2.5 in [11] that

$$1 - F(\underline{z}) - \frac{|\Delta|f}{2} \le \sum_{k=0}^{\infty} f(t+k) \le 1 - F(\underline{z}) + \frac{|\Delta|f}{2}, \quad t \in (\underline{z}, \underline{z}+1),$$

so that

$$\left(1 - F(\underline{z}) - \frac{|\Delta|f}{2}\right)(z - \underline{z}) \le g(\underline{z}) - g(z) \le \left(1 - F(\underline{z}) + \frac{|\Delta|f}{2}\right)(z - \underline{z}). \quad (7)$$

On the other hand, using Lemma 3.1(a) we see that

$$g(s+1) = g(s) - (1 - F(s)), \quad s \in \mathbb{R}.$$

Since the function g_α coincides with g on $\alpha + \mathbb{Z}$, and moreover g_α is linear on the interval $[\underline{z}, \underline{z}+1]$, it follows that

$$g(\underline{z}) - g_\alpha(z) = (1 - F(\underline{z}))(z - \underline{z}). \quad (8)$$

Together, (7) and (8) imply

$$|g_\alpha(z) - g(z)| \le (z - \underline{z})\frac{|\Delta|f}{2}, \quad z \in [\underline{z}, \underline{z}+1]. \quad (9)$$

Similarly, by comparing $g(z)$ and $g_\alpha(z)$ to $g(\underline{z}+1)$, one obtains

$$|g_\alpha(z) - g(z)| \le (\underline{z} + 1 - z)\frac{|\Delta|f}{2}, \quad z \in [\underline{z}, \underline{z}+1]. \quad (10)$$

For α-approximations of expected surplus function g, the claimed error bound now follows from (9), (10), and the observation that $\min\{(z - \underline{z}), (\underline{z} + 1 - z)\} \le 1/2$.

Analogously, the same error bound can be derived for the special case with $q^+ = 0$ and $q^- = 1$. The claim for the general case then follows trivially. □

Example 3.1. Suppose that $\boldsymbol{h}$ is normally distributed with mean μ and variance σ^2. Then, the pdf f is a unimodal function with maximum value $f(x) = (2\pi)^{-1/2}\sigma^{-1}$ at $x = \mu$, and thus $|\Delta|f = (\pi/2)^{-1/2}\sigma^{-1}$. We conclude from Theorem 3.2 that for all $\alpha \in [0, 1)$,

$$\|Q_\alpha - Q\|_\infty \le (q^+ + q^-)(8\pi)^{-1/2}\sigma^{-1}.$$

The uniform error bound of Theorem 3.2 can be reduced by a factor 2 if the following combination of α-approximations is used. For $\alpha \in [0, 0.5)$ and $\beta = \alpha + 0.5$, define the pdf

$$f_{\alpha\beta}(s) := \frac{f_\alpha(s) + f_\beta(s)}{2}, \quad s \in \mathbb{R},$$

where f_α and f_β are density functions of α-approximations as before. The resulting convex approximations $Q_{\alpha\beta}$ of Q satisfy

$$\|Q_{\alpha\beta} - Q\|_\infty \le (q^+ + q^-)\frac{|\Delta|f}{8}. \quad (11)$$

It can be shown that this error bound cannot be reduced by using other convex combinations of pdf of type f_α.

The error bound presented above is proportional to the total variation of the pdf f of $\boldsymbol{h}$. For many distributions, e.g. with unimodal densities, the total variation of a pdf decreases as the variance of the distribution increases. We may therefore expect that the approximation Q_α becomes better as the variance of such distributions becomes higher.

Finally, we remark that convex approximations of the function Q can be represented as (one-dimensional) *continuous* simple recourse functions. The latter functions are defined like (4), except that no rounding operations are involved. In the case of α-approximations, the corresponding modification of the underlying distribution is known in closed form [11].

Lemma 3.2. *Let $\boldsymbol{h}$ be a continuous random variable with cdf F with finite mean value, and $\alpha \in [0, 1)$. Then*

$$Q_\alpha(z) = q^+\mathbb{E}_{\boldsymbol{\varphi}_\alpha}\left[(\boldsymbol{\varphi}_\alpha - z)^+\right] + q^-\mathbb{E}_{\boldsymbol{\varphi}_\alpha}\left[(\boldsymbol{\varphi}_\alpha - z)^-\right] + \frac{q^+q^-}{q^+ + q^-}, \quad z \in \mathbb{R},$$

where $\boldsymbol{\varphi}_\alpha$ is a discrete random variable with support in $\alpha + \mathbb{Z}$ and, for $k \in \mathbb{Z}$,

$$\Pr\left\{\boldsymbol{\varphi}_\alpha = \alpha + k\right\} = \frac{q^+}{q^+ + q^-}\Pr\left\{\boldsymbol{h} \in C_\alpha^k\right\} + \frac{q^-}{q^+ + q^-}\Pr\left\{\boldsymbol{h} \in C_\alpha^{k+1}\right\}. \quad \square$$

We conclude that simple *integer* recourse functions can be approximated by *continuous* simple recourse functions with *discretely distributed* right-hand side parameters, simply by dropping the integrality restrictions and a modification of the distribution according to Lemma 3.2. The resulting convex problem can be solved using existing algorithms for continuous simple recourse problems with discrete underlying distributions.

3.2. Complete integer recourse

We now turn to the much more general class of complete integer recourse models. In addition to completeness and sufficiently expensive recourse, so that v is finite, we assume that the recourse matrix W is integer (or rational, so that integrality of W can be obtained by scaling). We will see that also in this case α-approximations of the distribution of $\boldsymbol{h}$ lead to convex approximations of the recourse function Q. In fact, Van der Vlerk [12] claims that this approach leads to the convex envelope of the recourse function Q if the recourse matrix is *totally unimodular* (TU). However, it turns out that this result only holds for a special class of distributions of $\boldsymbol{h}$. Below we derive the results for this special case.

3.2.1. Totally unimodular complete integer recourse

Because W is TU, the extreme points of the feasible set $\{y \in \mathbb{R}_+^{n_2} : Wy \geq h\}$ are integral for any integer right-hand side h. However, in our recourse problem the right-hand side $\boldsymbol{h} - Tx$ is not an integer vector in general. But since Wy is integral for all $y \in \mathbb{Z}^{n_2}$ we may round up the right-hand-side. Due to the assumption that W is TU, we may now relax the integrality restrictions on y, without changing the optimal value of the problem. That is,

$$\begin{aligned} v(x, h) &:= \min_y \left\{qy : Wy \geq h - Tx,\ y \in \mathbb{Z}_+^{n_2}\right\} \\ &= \min_y \left\{qy : Wy \geq \lceil h - Tx \rceil,\ y \in \mathbb{R}_+^{n_2}\right\} && (12) \\ &= \max_\lambda \left\{\lambda \lceil h - Tx \rceil : \lambda W \leq q,\ \lambda \in \mathbb{R}_+^m\right\}, && (13) \end{aligned}$$

where the last equality follows from (strong) LP duality.

Since the recourse structure is complete and sufficiently expensive, it follows that the dual feasible region $\Lambda := \{\lambda \in \mathbb{R}_+^m : \lambda W \leq q\}$ is a bounded, non-empty polyhedral set. Hence,

$$v(x, h) = \max_{k=1,\ldots,K} \lambda^k \lceil h - Tx \rceil, \quad x \in \mathbb{R}^n, h \in \mathbb{R}^m, \tag{14}$$

where λ^k, $k = 1, \ldots, K$, are the finitely many extreme points of the dual feasible region Λ.

Thus, v is the maximum of finitely many round-up functions, and hence non-convex. However, as we will see below, the expectation of a round-up function is convex if the underlying distribution of $\boldsymbol{h}$ is of a certain type. Analogous to the simple recourse case, this allows the construction of convex approximations of Q by means of special purpose approximations of the distribution.

To set the stage, we first study the expected round-up function

$$R(z) := \lambda\mathbb{E}_{\boldsymbol{h}}\left[\lceil \boldsymbol{h} - z \rceil\right], \quad z \in \mathbb{R}^m,$$

defined for any fixed $\lambda \in \mathbb{R}_+^m$.

If $m = 1$, $\lambda = 1$, and $\boldsymbol{h}$ is continuously distributed, then

$$R(z) = \mathbb{E}_{\boldsymbol{h}}\left[\lceil \boldsymbol{h} - z \rceil^+\right] - \mathbb{E}_{\boldsymbol{h}}\left[\lfloor \boldsymbol{h} - z + 1 \rfloor^-\right], \quad z \in \mathbb{R}, \tag{15}$$

since $\lceil s \rceil = \lceil s \rceil^+ - \lceil s \rceil^-$, $s \in \mathbb{R}$, and $\lceil s \rceil^- = \lfloor s + 1 \rfloor^-$ for all $s \notin \mathbb{Z}$. The right-hand side of (15) is very similar to the one-dimensional simple recourse function with $q^+ = 1$ and $q^- = -1$. Hence, in view of Corollary 3.1 it is not surprising that this one-dimensional function R is convex if $\boldsymbol{h}$ has a piecewise constant pdf of the type specified there. This result can be generalized to m-dimensional round-up functions.

Lemma 3.3. *Let $\boldsymbol{h} \in \mathbb{R}^m$ be a continuous random vector with joint pdf f that is constant on every hypercube $C_\alpha^k := \prod_{i=1}^m (\alpha_i + k_i - 1, \alpha_i + k_i]$, $k \in \mathbb{Z}^m$, for an arbitrary but fixed $\alpha = (\alpha_1, \ldots, \alpha_m) \in [0, 1)^m$. Then*

$$\mathbb{E}_{\boldsymbol{h}}\left[\lceil \boldsymbol{h} - z \rceil\right] = \mathbb{E}_{\boldsymbol{\varphi}_\alpha}\left[\boldsymbol{\varphi}_\alpha - z\right] = \mu_\alpha - z, \quad z \in \mathbb{R}^m,$$

where $\boldsymbol{\varphi}_\alpha := \lceil \boldsymbol{h} - \alpha \rceil + \alpha$ is a discrete random vector with mean value μ_α and support in $\alpha + \mathbb{Z}^m$, with

$$\Pr\left\{\boldsymbol{\varphi}_\alpha = \alpha + k\right\} = \Pr\left\{\boldsymbol{h} \in C_\alpha^k\right\}, \quad k \in \mathbb{Z}^m.$$

Hence, in this case the round-up function $R(z) = \lambda\mathbb{E}_{\boldsymbol{h}}\left[\lceil \boldsymbol{h} - z \rceil\right]$, $z \in \mathbb{R}^m$, is affine with gradient $-\lambda$, and thus convex.

Proof. We use that

$$\mathbb{E}_{\boldsymbol{h}}\left[\lceil \boldsymbol{h} - z \rceil\right] = \sum_{k \in \mathbb{Z}^m} \Pr\left\{\boldsymbol{h} \in C_\alpha^k\right\} \mathbb{E}_{\boldsymbol{h}}\left[\lceil \boldsymbol{h} - z \rceil \mid \boldsymbol{h} \in C_\alpha^k\right], \quad z \in \mathbb{R}^m. \tag{16}$$

For each fixed $k \in \mathbb{Z}^m$ for which $\Pr\left\{\boldsymbol{h} \in C_\alpha^k\right\} > 0$, the conditional distribution of $\boldsymbol{h}$ given $\boldsymbol{h} \in C_\alpha^k$ is uniform on C_α^k. It follows that conditioned on the event $\boldsymbol{h} \in C_\alpha^k$, the components of the vector $\boldsymbol{h}$ are independent random variables, with each $\boldsymbol{h}_i$ uniformly distributed on $(\alpha_i + k_i - 1, \alpha_i + k_i]$, $i = 1, \ldots, m$. Hence, writing each component as in (15) and applying Lemma 3.2 to each term individually, it follows that

$$\mathbb{E}_{\boldsymbol{h}}\left[\lceil \boldsymbol{h} - z \rceil \mid \boldsymbol{h} \in C_\alpha^k\right] = \alpha + k - z, \quad z \in \mathbb{R}^m. \tag{17}$$

Substitution of (17) in (16) proves the first claim.

The second claim follows trivially from the first one. $\square$

Based on Lemma 3.3, we define α-approximations of the round-up function R as

$$R_\alpha(z) := \lambda\mathbb{E}_{\boldsymbol{h}_\alpha}\left[\lceil \boldsymbol{h}_\alpha - z \rceil\right], \quad z \in \mathbb{R}^m,$$

for $\alpha \in [0, 1)^m$, with h_α as introduced in Definition 3.1.

In general, an α-approximation is neither a lower bound nor an upper bound of R. However, since $R(z+k) = R(z) - \lambda k$, for every $k \in \mathbb{Z}^m$ and $z \in \mathbb{R}^m$, we see that $R(z) + \lambda z$ is a periodic function, which repeats itself on every set C_α^k. Thus, defining

$$\alpha^\star \in \operatorname{argmin}\left\{R(z) + \lambda z : z \in [0,1)^m\right\}, \tag{18}$$

$R_{\alpha^\star}$ is a lower bound for R, which is sharp at every $z \in \alpha^\star + \mathbb{Z}^m$. By construction, the affine function $R_{\alpha^\star}$ is actually the convex envelope of R. Moreover, the components $\alpha_i^\star$, $i = 1, \ldots, m$, of the parameter vector $\alpha^\star$ can be determined independently.

Now we are ready to prove the main result for this class of models with TU recourse matrix. Using the dual representation (14) of the value function v, we have

$$Q(x) = \mathbb{E}_{\boldsymbol{h}}\left[\max_{k=1,\ldots,K} \lambda^k \lceil \boldsymbol{h} - Tx \rceil\right], \quad x \in \mathbb{R}^n.$$

Note that Q is not simply the pointwise maximum of a number of expected round-up functions R. However, the results above for the function R play a major role in the proof of Theorem 3.3.

Theorem 3.3. *Consider the integer recourse expected value function Q, defined as*

$$Q(x) = \mathbb{E}_{\boldsymbol{h}}\left[\min_y qy : Wy \geq \boldsymbol{h} - Tx,\ y \in \mathbb{Z}_+^{n_2}\right], \quad x \in \mathbb{R}^n. \tag{19}$$

Assume that

(i) *the recourse is complete and sufficiently expensive,*
(ii) *the recourse matrix W is totally unimodular, and*
(iii) *$\boldsymbol{h}$ is uniformly distributed on $\prod_{i=1}^m (a_i, b_i]$ for $a_i < b_i \in \mathbb{R}$, $i = 1, \ldots, m$.*

If

(iv) *the matrix T is of full row rank,*

then the convex envelope of Q is the continuous recourse expected value function $Q_{\alpha^\star}$, defined as

$$Q_{\alpha^\star}(x) = \mathbb{E}_{\varphi_{\alpha^\star}}\left[\min_y qy : Wy \geq \boldsymbol{\varphi}_{\alpha^\star} - Tx,\ y \in \mathbb{R}_+^{n_2}\right],$$
$$x \in \mathbb{R}^n, \tag{20}$$

where $\alpha^\star$ is defined by (18), and $\boldsymbol{\varphi}_{\alpha^\star}$ is the discrete random vector $\boldsymbol{\varphi}_{\alpha^\star} = \lceil \boldsymbol{h} - \alpha^\star \rceil + \alpha^\star$ with support in $\alpha^\star + \mathbb{Z}^m$, and

$$\Pr\left\{\boldsymbol{\varphi}_{\alpha^\star} = \alpha^\star + k\right\} = \Pr\left\{\boldsymbol{h} \in C_{\alpha^\star}^k\right\}, \quad k \in \mathbb{Z}^m.$$

If condition (iv) *is not satisfied, then $Q_{\alpha^\star}$ is a lower bound for Q.*

Proof. We will prove that $Q_{\alpha^\star}$ is the convex envelope of Q if T is of full row rank. The other case then follows from basic analysis.

Assuming that T is of full row rank, we may conveniently consider Q as a function of the tender variables $z := Tx \in \mathbb{R}^m$.

First we will prove that $Q_{\alpha^\star}(z) = Q(z)$ for all $z \in \alpha^\star + \mathbb{Z}^m$, and subsequently that $Q_{\alpha^\star}$ is a lower bound for Q. This completes the proof, since all vertices of the polyhedral function $Q_{\alpha^\star}$ are contained in $\alpha^\star + \mathbb{Z}^m$.

Using the dual representation (14) of the value function v, we have

$$Q(z) = \mathbb{E}_{\boldsymbol{h}}\left[\max_{k=1,\ldots,K} \lambda^k \lceil \boldsymbol{h} - z \rceil\right], \quad z \in \mathbb{R}^m,$$

and, analogously,

$$Q_{\alpha^\star}(z) = \mathbb{E}_{\varphi_{\alpha^\star}}\left[\max_{k=1,\ldots,K} \lambda^k (\varphi_{\alpha^\star} - z)\right], \quad z \in \mathbb{R}^m.$$

Consider a fixed $\bar{z} \in \alpha^\star + \mathbb{Z}^m$ and a fixed $l \in \mathbb{Z}^m$. Then $\lceil h - \bar{z} \rceil = \alpha^\star + l - \bar{z}$ is constant for all $h \in C_{\alpha^\star}^l$, and by definition $\Pr\left\{\boldsymbol{h} \in C_{\alpha^\star}^l\right\} = \Pr\left\{\boldsymbol{\varphi}_{\alpha^\star} = \alpha^\star + l\right\}$. Since this is true for every $l \in \mathbb{Z}^m$, it follows that

$$\begin{aligned}
Q(\bar{z}) &= \sum_{l \in \mathbb{Z}^m} \Pr\left\{\boldsymbol{h} \in C_{\alpha^\star}^l\right\} \mathbb{E}_{\boldsymbol{h}}\left[\max_{k=1,\ldots,K} \lambda^k \lceil \boldsymbol{h} - \bar{z} \rceil \mid \boldsymbol{h} \in C_{\alpha^\star}^l\right] \\
&= \sum_{l \in \mathbb{Z}^m} \Pr\left\{\boldsymbol{h} \in C_{\alpha^\star}^l\right\} \max_{k=1,\ldots,K} \lambda^k (\alpha^\star + l - \bar{z}) \\
&= \sum_{l \in \mathbb{Z}^m} \Pr\left\{\boldsymbol{\varphi}_{\alpha^\star} = \alpha^\star + l\right\} \\
&\quad \times \mathbb{E}_{\varphi_{\alpha^\star}}\left[\max_{k=1,\ldots,K} \lambda^k (\boldsymbol{\varphi}_{\alpha^\star} - \bar{z}) \mid \boldsymbol{\varphi}_{\alpha^\star} = \alpha^\star + l\right] \\
&= Q_{\alpha^\star}(\bar{z}).
\end{aligned}$$

It remains to prove that $Q_{\alpha^\star}$ is a lower bound for Q. Again conditioning on the events $\boldsymbol{h} \in C_{\alpha^\star}^l$, $l \in \mathbb{Z}^m$, we obtain, for $z \in \mathbb{R}^m$,

$$\begin{aligned}
Q(z) &= \sum_{l \in \mathbb{Z}^m} \Pr\left\{\boldsymbol{h} \in C_{\alpha^\star}^l\right\} \mathbb{E}_{\boldsymbol{h}}\left[\max_{k=1,\ldots,K} \lambda^k \lceil \boldsymbol{h} - z \rceil \mid \boldsymbol{h} \in C_{\alpha^\star}^l\right] \\
&\geq \sum_{l \in \mathbb{Z}^m} \Pr\left\{\boldsymbol{h} \in C_{\alpha^\star}^l\right\} \max_{k=1,\ldots,K} \lambda^k \mathbb{E}_{\boldsymbol{h}}\left[\lceil \boldsymbol{h} - z \rceil \mid \boldsymbol{h} \in C_{\alpha^\star}^l\right] \\
&\geq \sum_{l \in \mathbb{Z}^m} \Pr\left\{\boldsymbol{h} \in C_{\alpha^\star}^l\right\} \max_{k=1,\ldots,K} \lambda^k \mathbb{E}_{\boldsymbol{h}_{\alpha^\star}}\left[\lceil \boldsymbol{h}_{\alpha^\star} - z \rceil \mid \boldsymbol{h}_{\alpha^\star} \in C_{\alpha^\star}^l\right] \\
&= \sum_{l \in \mathbb{Z}^m} \Pr\left\{\boldsymbol{h} \in C_{\alpha^\star}^l\right\} \max_{k=1,\ldots,K} \lambda^k \left(\alpha^\star + l - z\right) \\
&= \sum_{l \in \mathbb{Z}^m} \Pr\left\{\boldsymbol{\varphi}_{\alpha^\star} = \alpha^\star + l\right\} \max_{k=1,\ldots,K} \lambda^k \left(\alpha^\star + l - z\right) \\
&= Q_{\alpha^\star}(z).
\end{aligned}$$

The second inequality is valid because each λ^k is nonnegative, so that for every $C_{\alpha^\star}^l$ the α-approximation $\lambda^k \mathbb{E}_{\boldsymbol{h}_{\alpha^\star}}\left[\lceil \boldsymbol{h}_{\alpha^\star} - z \rceil \mid \boldsymbol{h}_{\alpha^\star} \in C_{\alpha^\star}^l\right]$ is a lower bound for $\lambda^k \mathbb{E}_{\boldsymbol{h}}\left[\lceil \boldsymbol{h} - z \rceil \mid \boldsymbol{h} \in C_{\alpha^\star}^l\right]$. Note that this assumption only holds under (iii), see [13]. □

Assumption (iii) in Theorem 3.3 ensures that $Q_{\alpha^\star}$ is a lower bound for Q. For other distributions this is not the case, because the second inequality in the proof above does not hold in general. For general distributions the α-approximations require a performance guarantee similar to the one in Theorem 3.2 for simple integer recourse models. Obtaining such a performance guarantee is a topic of current research.

We conclude that if the recourse matrix W is totally unimodular, then the integer complete recourse problem with recourse function Q can be approximated by the continuous complete recourse problem with recourse function $Q_{\alpha^\star}$. To construct this approximation, the integer restrictions on the second-stage variables are dropped, and the distribution of the right-hand side parameters is modified according to Theorem 3.3. The resulting continuous complete recourse problem with discretely distributed right-hand side parameters can be solved by existing special purpose algorithms [14,15].

3.2.2. General complete integer recourse

Finally, we drop the assumption that W is TU. We will prove that $Q_{\alpha^\star}$ is an upper bound for the approximation obtained by using the LP relaxation of the second-stage problem. The latter convex function will be denoted by Q^{LP}, defined as

$$Q^{LP}(x) := \mathbb{E}_{\boldsymbol{h}}\left[\min_y \left\{qy : Wy \geq \boldsymbol{h} - Tx,\ y \in \mathbb{R}_+^{n_2}\right\}\right],$$
$$x \in \mathbb{R}^n. \tag{21}$$

Theorem 3.4. *Consider the functions $Q_{\alpha^\star}$ and Q^{LP}, defined by (20) and (21) respectively, which are convex approximations for the integer* *recourse expected value function Q, defined by (19).*

(a) $Q_{\alpha^\star}(x) \geq Q^{LP}(x)$ *for all* $x \in \mathbb{R}^n$.
(b) *Assume*
   (i) $q \geq 0$, *so that* 0 *is a trivial lower bound for* v *and* Q, *and*
   (ii) $\boldsymbol{h}$ *is continuously distributed.*
   Then, $Q_{\alpha^\star}(x) > 0$ *implies* $Q_{\alpha^\star}(x) > Q^{LP}(x)$.

Proof. As before, we condition on the events $\boldsymbol{h} \in C^l_{\alpha^\star}$, $l \in \mathbb{Z}^m$, to obtain, for $x \in \mathbb{R}^n$,

$$Q_{\alpha^\star}(x) = \sum_{l\in\mathbb{Z}^m} \Pr\left\{\boldsymbol{h} \in C^l_{\alpha^\star}\right\} \max_{k=1,\dots,K} \lambda^k\left(\alpha^\star + l - Tx\right) \tag{22}$$

and

$$Q^{LP}(x) = \sum_{l\in\mathbb{Z}^m} \Pr\left\{\boldsymbol{h} \in C^l_{\alpha^\star}\right\} \times \mathbb{E}_{\boldsymbol{h}}\left[\max_{k=1,\dots,K} \lambda^k\left(\boldsymbol{h} - Tx\right) \mid \boldsymbol{h} \in C^l_{\alpha^\star}\right]. \tag{23}$$

For each $l \in \mathbb{Z}^m$ it follows from the definition of $C^l_{\alpha^\star} = \prod_{i=1}^m \left(\alpha^\star_i + l_i - 1, \alpha^\star_i + l_i\right]$ that $\alpha^\star + l \geq h$ for all $h \in C^l_{\alpha^\star}$. Thus, $\lambda^k(\alpha^\star + l - Tx) \geq \lambda^k(h - Tx)$ for all $h \in C^l_{\alpha^\star}$, $k = 1, \dots, K$, since $\lambda^k \geq 0$. This implies that

$$\max_{k=1,\dots,K} \lambda^k(\alpha^\star + l - Tx) \geq \max_{k=1,\dots,K} \lambda^k(h - Tx)$$

for all $h \in C^l_{\alpha^\star}$ and substitution in (22) and (23) proves that $Q_{\alpha^\star} \geq Q^{LP}$.

In order to show that $Q_{\alpha^\star}(x) > 0$ implies $Q_{\alpha^\star}(x) > Q^{LP}(x)$, it suffices to prove that if $Q_{\alpha^\star}(x) > 0$, then there exists $\bar{l} \in \mathbb{Z}^m$ such that $\Pr\left\{\boldsymbol{h} \in \mathrm{int} C^{\bar{l}}_{\alpha^\star}\right\} > 0$ and

$$\max_{k=1,\dots,K} \lambda^k(\alpha^\star + \bar{l} - Tx) > \max_{k=1,\dots,K} \lambda^k(h - Tx) \tag{24}$$

for all $h \in \mathrm{int} C^{\bar{l}}_{\alpha^\star}$.

From (22) it follows that if $Q_{\alpha^\star}(x) > 0$, then there exists $\bar{l} \in \mathbb{Z}^m$ such that $\max_{k=1,\dots,K} \lambda^k(\alpha^\star + \bar{l} - Tx) > 0$ and $\Pr\left\{\boldsymbol{h} \in C^{\bar{l}}_{\alpha^\star}\right\} > 0$. Since $\boldsymbol{h}$ is continuously distributed, $\Pr\left\{\boldsymbol{h} \in \mathrm{int} C^{\bar{l}}_{\alpha^\star}\right\} > 0$. Moreover, for all $h \in \mathrm{int} C^{\bar{l}}_{\alpha^\star}$ and every $k = 1, \dots, K$, with $\lambda^k \neq 0$, we have $\lambda^k(\alpha^\star + \bar{l} - Tx) > \lambda^k(h - Tx)$ since $\lambda^k \geq 0$ and $\alpha^\star + \bar{l} > h$. This observation combined with $\max_{k=1,\dots,K} \lambda^k(\alpha^\star + \bar{l} - Tx) > 0$ implies that (24) holds for all $h \in \mathrm{int} C^{\bar{l}}_{\alpha^\star}$, proving that $Q_{\alpha^\star}(x) > 0$ implies $Q_{\alpha^\star}(x) > Q^{LP}(x)$. □

It follows immediately from Theorem 3.4 that if $Q_{\alpha^\star}$ is a lower bound for Q, then it is a strictly better convex approximation of Q than Q^{LP}.

Corollary 3.2. *Consider the functions* $Q_{\alpha^\star}$ *and* Q^{LP}, *defined by (20) and (21) respectively, which are convex approximations for the integer recourse expected value function* Q, *defined by (19).*

Assume

(i) $q \geq 0$, *so that* 0 *is a trivial lower bound for* v *and* Q, *and*
(ii) $\boldsymbol{h}$ *is uniformly distributed on* $\prod_{i=1}^m (a_i, b_i]$ *for* $a_i < b_i \in \mathbb{R}$, $i = 1, \dots, m$.

Then, the function $Q_{\alpha^\star}$ *is a strictly better convex approximation of* Q *than* Q^{LP}, *in the sense that* $Q_{\alpha^\star}$ *is a lower bound for* Q, *and* $Q_{\alpha^\star}(x) > 0$ *implies* $Q_{\alpha^\star}(x) > Q^{LP}(x)$.

Proof. If we do not assume that W is TU, then relaxing the integrality restrictions in (12) yields a lower bound. This combined with Theorem 3.3 proves that $Q_{\alpha^\star}$ is a lower bound for Q. The second claim follows immediately from Theorem 3.4. □

Note that the distribution of $\boldsymbol{\varphi}_{\alpha^\star}$ as defined in Theorem 3.3 is always discrete, no matter what kind of distribution $\boldsymbol{h}$ follows. Thus, in particular if $\boldsymbol{h}$ is continuously distributed, then $Q_{\alpha^\star}$ is *computationally more tractable* than Q^{LP} which in this case is defined as an m-dimensional integral. Moreover, if $\boldsymbol{h}$ is distributed as in Corollary 3.2, then $Q_{\alpha^\star}$ is not only computationally more tractable than Q^{LP}, it is also a better approximation of Q.

4. Worst-case performance ratio analysis

Worst-case performance analysis of approximation algorithms for stochastic combinatorial optimization problems has seen an increasing research interest over the last ten years. The two-stage stochastic optimization problems studied are mainly recourse type versions of classical deterministic combinatorial problems, like the two-stage vertex cover problem, where in the first stage a subset of vertices can be bought for vertex specific prices, and in the second stage, when a random subset of the edges has to be covered, extra vertices can be bought against a higher price. All of them exhibit the *relative complete recourse* property: every first-stage decision can be extended to a feasible second-stage solution for every realization of the random parameters. This is clearly true in the two-stage vertex cover problem. We will come back to this type of problem later.

The strength of the approximation results for two-stage stochastic combinatorial optimization problems studied varies according to the assumptions made on the randomness of the stochastic input.

We start in Section 4.1 with an example with the most severe restriction on the randomness: the random input parameters are assumed to come in a number of scenarios which is finite or polynomially bounded by the other problem parameters. We call this the *scenario-based model*. The example is in fact to the best of our knowledge also the first example of worst-case performance ratio analysis of stochastic combinatorial optimization algorithms.

At the end of Section 4.1 we briefly mention some more results along this line, where we present how some results can be extended to the stochastic version in which input parameters are not given in scenarios but are independently distributed. We call the latter the *independent model*. As noted in Section 2 this gives rise to deterministic equivalent problems of exponential size.

The most extreme form of randomness is the so-called *black box model*. Here the randomness is not regarded as part of the input at all, but merely the assumption is made that random scenarios can be sampled by an oracle; each sample taking only one oracle call. This type of model is the subject of Section 4.2.

4.1. Scenario-based model

Although the example below resembles a two-stage stochastic knapsack problem, we prefer to present it as a service provision problem. It actually diverges from most problems studied in this research direction, since it is a hierarchical planning problem rather than a recourse problem.

Consider a service provider that can provide a variety of services. For each service the provider receives requests from its customers. However, in order to actually provide a service, the service has to be installed first. Both installation and provision of a service require capacity from the same resource, which has limited capacity. Since only served requests yield a profit, the problem is to select a subset of the services to be installed and to decide which customer requests to serve, such as to maximize the total profit.

If all demands for services are known in advance, the problem is NP-hard *in the ordinary sense* and a fully polynomial-time approximation scheme exists [16].

We study the problem with uncertain demand for services, where services have to be installed before demand is observed. The uncertainty is expressed by a number of scenarios, i.e., demand vectors with corresponding probabilities. The stochastic service provision problem is to select services to be installed such as to maximize *expected* profit of meeting service requests. It has been shown that this stochastic problem is *strongly* NP-hard [17]. Thus, the complexity of the problem increases by introducing stochasticity.

We will exhibit the performance analysis of an approximation algorithm for this problem under the restriction that the resource has enough capacity to install all services. However, it may not be optimal to install all of them since it may leave too little capacity for serving the requests.

A typical approach for constructing an approximation algorithm that is frequently used in stochastic combinatorial optimization problems is the following. First formulate the problem as a two-stage stochastic integer programming problem. Then solve the LP-relaxation and round its optimal solution to obtain a feasible solution for the original stochastic integer problem. Obviously, the quality of our approximation will rely heavily on the way we round the solution of the LP-relaxation. Below we describe such an approximation algorithm with a worst-case performance ratio of $5+\sqrt{3}$. First we introduce some notation.

Let n be the number of services and s the capacity of the single resource. Let q_j be the profit obtained from allocating one resource unit to meeting demand for service j. Each service j requires a resource capacity r_j to be installed, which is independent of the demand met. Demand is denoted by the random vector $\boldsymbol{d} \in \mathbb{R}^n$, with $\boldsymbol{d}_j$ denoting the demand for service j. Binary first-stage decision variables z_j are used to indicate whether service j is installed ($z_j = 1$) or not ($z_j = 0$), $j = 1, \dots, n$. The second-stage decision variable x_j gives the amount of resource used to meet demand for service j. The two-stage stochastic programming formulation becomes:

$$\begin{aligned}
\max \quad & \mathbb{E}_{\boldsymbol{d}}\left[v(z, \boldsymbol{d})\right] \\
\text{s.t.} \quad & \sum_{j=1}^{n} r_j z_j \le s \\
& z_j \in \{0, 1\} \quad j = 1, \dots, n,
\end{aligned}$$

with

$$\begin{aligned}
v(z, d) = \max \quad & \sum_{j=1}^{n} q_j x_j \\
\text{s.t.} \quad & \sum_{j=1}^{n} x_j \le s - \sum_{j=1}^{n} r_j z_j \\
& x_j \le d_j z_j \quad j = 1, \dots, n, \\
& x_j \ge 0 \quad j = 1, \dots, n.
\end{aligned}$$

The second-stage problem is to determine the amount x_j of requests for service j to be served, taking into account the *capacity constraint* and *demand constraints*. These constraints ensure that the resource capacity is not exceeded, that the number of service requests provided does not exceed the demand, and that requests are only met if the corresponding services are installed. Observe that the recourse is relatively complete due to the first-stage constraint and because the demands d_j are non-negative.

Let K be the number of scenarios describing the probability distribution of demand, p^k be the probability that scenario k occurs, and d_j^k be the demand for service j in scenario k. Given the scenarios the following deterministic equivalent linear mixed-integer program can be formulated, in which we use x_{jk} to denote the resource allocated to providing service j in scenario k (we use a subscript instead of superscript for k here because of notational convenience later on).

$$\begin{aligned}
\max \quad & \sum_{k=1}^{K} p^k \sum_{j=1}^{n} q_j x_{jk} \\
\text{s.t.} \quad & \sum_{j=1}^{n} (r_j z_j + x_{jk}) \le s \quad k = 1, \dots, K, \\
& d_j^k z_j - x_{jk} \ge 0 \quad j = 1, \dots, n,\ k = 1, \dots, K, \\
& z_j \in \{0, 1\},\ x_{jk} \ge 0 \quad j = 1, \dots, n,\ k = 1, \dots, K.
\end{aligned} \tag{25}$$

Though integrality conditions only apply to the first-stage variables z_j, if the data, resource capacity, installation requirements, and demands are integral, the second stage will have an integer solution in every scenario.

As stated, we assume that $\sum_{j=1}^{n} r_j \le s$. Moreover, to facilitate the exposition we assume that no demand is larger than the node capacity minus the corresponding installation requirement: for any service j in any scenario k, $d_j^k \in [0, s - r_j]$. If necessary, this can be ensured by preprocessing.

The approximation algorithm that we will present is based on rounding the optimal solution of the LP-relaxation of problem (25) obtained by replacing the binary restrictions on the z-variables by $0 \le z_j \le 1, j = 1, \dots, n$. We assume without loss of generality that the resource capacity s is equal to 1.

Let $(z^{\text{LP}}, x^{\text{LP}})$ be an optimal basic solution of the LP relaxation. Depending on whether $z_j^{LP} = 0, 1$ or fractional, we will use a different way of rounding. Moreover, for services with fractional z_j^{LP} we will also differentiate between services with an installation capacity rate r_j below and above some threshold $0 < w < 1$, to be specified later. Let ℓ be the number of fractional z_j^{LP} and let ℓ_w of these services have $r_j \le w$. Let Z be the set of services with $z_j^{\text{LP}} = 1$. By renumbering the services if necessary, assume that $0 < z_j^{\text{LP}} < 1$ and $r_j \le w$ for $j = 1, \dots, \ell_w$ and $0 < z_j^{\text{LP}} < 1$ and $r_j > w$ for $j = \ell_w + 1, \dots, \ell$. Write the optimal LP value as

$$\pi^{\text{LP}} = \pi_0^{\text{LP}} + \pi_1^{\text{LP}} + \pi_2^{\text{LP}} \tag{26}$$

where

$$\pi_0^{\text{LP}} = \sum_{j \in Z} \sum_{k=1}^{K} p^k q_j x_{jk}^{\text{LP}},$$

$$\pi_1^{\text{LP}} = \sum_{j=1}^{\ell_w} \sum_{k=1}^{K} p^k q_j x_{jk}^{\text{LP}},$$

and

$$\pi_2^{\text{LP}} = \sum_{j=\ell_w+1}^{\ell} \sum_{k=1}^{K} p^k q_j x_{jk}^{\text{LP}}.$$

The approximation algorithm applies rounding to the LP solution to generate various feasible solutions, as we will describe below, and selects from those solutions the best one. Each feasible solution or subset of feasible solutions allows for deriving an upper bound on either π_0^{LP}, π_1^{LP}, or π_2^{LP} in terms of the solution value of the algorithm π^R. Since $\pi^{OPT} \le \pi^{LP}$, with π^{OPT} denoting the optimal solution value of the stochastic integer program, this leads to a constant worst-case performance ratio for the approximation algorithm.

The first feasible solution is obtained by installing service j if and only if $z_j^{\text{LP}} = 1$; i.e., install all services $j \in Z$. The remaining capacity is then allocated to serve demand for the installed services in a greedy way, in order of non-increasing q_j values. Denote the resulting solution by (z^G, x^G) and its value by π^G. Then, obviously,

$$\pi_0^{\text{LP}} \le \pi^G \le \pi^R. \tag{27}$$

The next set of feasible solutions is used to bound π_1^{LP}. The idea is to group the services $j = 1, \ldots, l_w$, such that the amount of resource capacity required to install all services in a group is not low and not high. We partition the set $\{1, \ldots, \ell_w\}$ into I subsets, $\{S_i\}_{i=1}^I$, such that

$$\sum_{j \in S_i} r_j \leq \beta + w \quad i = 1, \ldots, I$$

and

$$\sum_{j \in S_i} r_j \geq \beta \quad i = 1, \ldots, I - 1. \tag{28}$$

Notice that $\sum_{j \in S_I} r_j$ is allowed to be smaller than β, for some $\beta > 0$ to be specified later such that $\beta + w < 1$. In the algorithm this partition is made in the most simple way, start filling set S_1 until addition of the next service would make the sum of installation requirements exceed $\beta + w$. This service is then the first one of S_2, etc.

Define $A = \sum_{j=1}^{\ell} r_j z_j^{\text{LP}}$ as the resource capacity used for installing services in the LP-relaxation and note that $\sum_{j=1}^{\ell_w} x_{jk}^{\text{LP}} \leq 1 - A$ for each $k = 1, \ldots, K$. Thus, in the optimal solution of the LP relaxation at most $1 - A$ units of capacity are available for the x variables. Installing only the services in one of the sets S_i, i.e. rounding z_j^{LP} to 1 for $j \in S_i$, will leave at least $1 - \beta - w$ units of capacity available. The x-variable values from the LP relaxation solution corresponding to services in S_i may be scaled down, if necessary, to use a total of no more than $1 - \beta - w$ units of capacity in each scenario.

For each $i = 1, \ldots, I$ we obtain a feasible solution (z^{H_i}, x^{H_i}) with $z_j^{H_i} = 1$ for $j \in S_i$, $z_j^{H_i} = 0$ for $j \notin S_i$, $x_{jk}^{H_i} = \gamma x_{jk}^{\text{LP}}$ for $j \in S_i$, $k = 1, \ldots, K$ and $x_{jk}^{H_i} = 0$ for $j \notin S_i$ and all k, where

$$\gamma = \begin{cases} \dfrac{1 - \beta - w}{1 - A} & \text{if } \beta + w \geq A, \\ 1 & \text{otherwise.} \end{cases} \tag{29}$$

The objective value of solution (z^{H_i}, x^{H_i}) is

$$\pi^{H_i} = \sum_{j \in S_i} \sum_{k=1}^{K} p^k q_j x_{jk}^{H_i} = \gamma \sum_{j \in S_i} \sum_{k=1}^{K} p^k q_j x_{jk}^{\text{LP}}.$$

Hence,

$$\pi_1^{\text{LP}} = \sum_{i=1}^{I} \sum_{j \in S_i} \sum_{k=1}^{K} p^k q_j x_{jk}^{\text{LP}} = \frac{1}{\gamma} \sum_{i=1}^{I} \pi^{H_i} \leq \frac{I}{\gamma} \pi^R. \tag{30}$$

By the assumption $\sum_{j=1}^{n} r_j \leq s$ and the definition of the sets S_i (specifically (28)) we have

$$1 \geq \sum_{j=1}^{n} r_j \geq \sum_{j=1}^{\ell_w} r_j = \sum_{i=1}^{I} \sum_{j \in S_i} r_j \geq (I - 1)\beta. \tag{31}$$

Thus, $I \leq 1 + 1/\beta$ which inserted in (30) implies that

$$\pi_1^{\text{LP}} \leq \frac{\beta + 1}{\beta \gamma} \pi^R. \tag{32}$$

The last set of feasible solutions considered by the algorithm consists of installing each service $j = \ell_w + 1, \ldots, \ell$ (having $r_j \geq w$) individually. Since $A = \sum_{j=1}^{\ell} r_j z_j^{\text{LP}} \geq \sum_{j=\ell_w+1}^{\ell} r_j z_j^{\text{LP}} \geq w \sum_{j=\ell_w+1}^{\ell} z_j^{\text{LP}}$, we have

$$\sum_{j=\ell_w+1}^{\ell} z_j^{\text{LP}} \leq \frac{A}{w}.$$

Just installing service j has objective value $q_j E[\boldsymbol{d}_j]$, since we have assumed that for any service j in any scenario k, $d_j^k \in [0, s - r_j]$. Satisfying the demand constraints implies that $\sum_{k=1}^{K} p^k x_{jk}^{\text{LP}} \leq E[\boldsymbol{d}_j] z_j^{\text{LP}}$. Altogether this yields the following bound.

$$\begin{aligned} \pi_2^{\text{LP}} &= \sum_{j=\ell_w+1}^{\ell} \sum_{k=1}^{K} p^k q_j x_{jk}^{\text{LP}} \\ &\leq \sum_{j=\ell_w+1}^{\ell} q_j E[\delta_j] z_j^{\text{LP}} \\ &\leq \pi^R \sum_{j=\ell_w+1}^{\ell} z_j^{\text{LP}} \leq \frac{A}{w} \pi^R. \end{aligned} \tag{33}$$

Combining (27), (32) and (33) gives

$$\pi^{\text{LP}} \leq \left(1 + \frac{\beta + 1}{\beta \gamma} + \frac{A}{w}\right) \pi^R. \tag{34}$$

Theorem 4.1. *Under the assumption that $\sum_{j=1}^{n} r_j \leq 1$, the approximation algorithm has worst-case performance ratio*

$$\frac{\pi^{OPT}}{\pi^R} \leq \left(5 + 2\sqrt{3}\right).$$

Proof. The choice of w and β depends on A in (34). When $A < \frac{1}{2}$ take $w = 1 - \frac{1}{2}\sqrt{3}$ and $\beta = -\frac{1}{2} + \frac{1}{2}\sqrt{3}$ and when $A \geq \frac{1}{2}$ take $w = \beta = \frac{1}{2}A$. In both cases $w + \beta \geq A$, and therefore $\gamma = \frac{1-\beta-w}{1-A}$. In the former case (34) leads to

$$\begin{aligned} \pi^{\text{OPT}} \leq \pi^{\text{LP}} &\leq \left(1 + \frac{2\left(1 + \sqrt{3}\right)(1 - A)}{-1 + \sqrt{3}} + \frac{A}{1 - \frac{1}{2}\sqrt{3}}\right) \pi^R \\ &= \left(1 + \left(1 + \sqrt{3}\right)^2 (1 - A) + 4\left(1 + \frac{1}{2}\sqrt{3}\right) A\right) \pi^R \\ &= \left(5 + 2\sqrt{3}\right) \pi^R. \end{aligned}$$

In the latter case (34) leads to

$$\pi^{\text{OPT}} \leq \pi^{\text{LP}} \leq \left(4 + \frac{2}{A}\right) \pi^R \leq 8\pi^R \leq \left(5 + 2\sqrt{3}\right) \pi^R. \quad \square$$

We notice that so far tightness of the bound has not been established. There exists an instance in which the ratio between the LP-bound and the optimal value is 4 and an instance for which the algorithm has ratio 2. It is worthwhile to stress once more that the deterministic counterpart of the problem, having the same number of binary decision variables, is weakly NP-hard. Thus, the complexity of the problem increases by introducing randomness, even though it only means adding continuous decision variables for each scenario of the problem.

In between the scenario-based model and the black box model treated in the next section, there is the independent model (see Section 2). For example, in the two-stage stochastic vertex cover problem, each edge will need to be covered in the second stage with probability p, independently of the other edges. In this independent model the randomness is represented by the value of p only, while the number of possible scenarios is exponential in the number of edges. Thus, in order to obtain approximation algorithms for this model with polynomial running time in terms of the size of the input, additional assumptions have to be made. Typically, the second-stage cost of buying a vertex is assumed to be inflated by a factor λ compared to the first-stage cost, and this inflation factor is the same for every vertex in every scenario. It allows to make use of the rather obvious local optimality property that a vertex is purchased in the first stage if and only if the probability that it is used in the second stage exceeds $1/\lambda$. An assumption of this kind will return in the black box models of the next section, where a uniform upper bound on this inflation factor appears to be crucial for getting desired results. In [18] a constant worst-case approximation ratio is obtained for this two-stage vertex cover model under the independent model, as well as similar results on other two-stage versions of classical combinatorial optimization problems. We do not present any of these results in further detail here, since all of them are superseded by the results that have been derived on the black box model.

4.2. The black box model

In this section we study stochastic problems under the black box model. In this model approximation algorithms are supposed to run in time polynomial in the size of the first-stage parameters only, in spite of the hardness of evaluating the objective function shown in Section 2.

Under this model the uniform inflation factor mentioned before, or more precisely a uniform bound on the inflation factor, is a crucial requirement. Thus, what follows has not been designed for general two-stage decision problems, but is restricted to recourse models assuming a uniform inflation factor. Moreover, the randomness is captured by the black box model. Information on the random vector $\boldsymbol{\xi}$ can only be obtained by sampling; a distribution is not explicitly given. Sampling one scenario is considered to be one computer operation (oracle consultation).

For this restricted class of two-stage stochastic programming models, a beautiful result shows how to transfer the performance behavior of algorithms for problems under the scenario-based model, like in the previous section, to algorithms for problems under the black box model, with only a small loss in approximation quality. It uses a technique which is well studied in the stochastic programming literature called the *sample average approximation* (SAA) method [4]. In its most basic form, this method simply draws independently a set of N scenarios from the distribution, formulates the deterministic equivalent problem, using probability $p_k = 1/N$ for each scenario, and solves this problem. We restrict our attention to models with a discrete decision space X. We will show that with high probability only a polynomial number of scenarios needs to be sampled to remain within $1 + O(\epsilon)$ from the true optimal solution value, if we are able to solve the deterministic equivalent problem optimally. Subsequently we will argue that this result also holds if we use approximation algorithms to solve the deterministic equivalent problem.

Consider the generic two-stage stochastic minimization problem

$$\min_{x\in X} f(x) := c(x) + \mathbb{E}_{\boldsymbol{\xi}}\left[v(x, \boldsymbol{\xi})\right], \tag{35}$$

where X is the set of all decision vectors. For example, X could be the set of all n-dimensional binary vectors.

The deterministic equivalent problem generated by SAA, with samples $\xi_1, \dots, \xi_N$, is formulated as

$$\min_{x\in X} \hat{f}_N(x) := c(x) + \frac{1}{N}\sum_{i=1}^{N} v(x, \xi_i). \tag{36}$$

We assume:

(A1) $c(x)$ and $v(x, \xi)$ are non-negative for every first-stage decision x and every scenario ξ;
(A2) there is an empty first-stage action $0 \in X$, with the property that for every $x \in X$ and every scenario ξ, $v(x, \xi) \le v(0, \xi)$;
(A3) a uniform bound on the inflation factor exists; i.e., there exists a smallest number $\lambda \ge 1$ such that for every $x \in X$ and every scenario ξ we have

$$v(0, \xi) - v(x, \xi) \le \lambda c(x). \tag{37}$$

Although these assumptions may seem arbitrary at first sight, they arise naturally in various applications. As an example consider again a vertex cover problem. In this problem a random set of edges needs to be covered. In the first stage, before we know the set of edges to be covered, we can already buy some vertices. If it turns out that not all edges are covered, we can buy additional vertices in the second stage. However, the cost of buying a vertex in the second stage is increased compared to the first-stage cost by an inflation factor that may depend on the vertex and the scenario, but never exceeds some upper bound λ. The objective is to minimize the expected costs of covering the edges.

Note that if the costs of buying a vertex are non-negative, then (A1) holds. Moreover, the empty first-stage action ($0 \in X$) corresponds to buying no vertices in the first stage. Obviously, the second-stage costs will not be larger for any subset of edges to be covered, if we buy some vertices ($x \in X$) in the first stage compared to buying none at all, so that (A2) holds. Finally, (A3) follows since the vertices x can be bought in the second stage at costs bounded by $\lambda c(x)$.

Next we analyze the SAA method. Fix $\epsilon > 0$, and let x^* and $\hat{x}_N$ be optimal solutions of the true problem (35) and the approximate problem (36), respectively. To facilitate the exposition we suppress from here on the subscript N and often use Z^* for denoting $f(x^*)$. For the proof of the theorem two types of scenarios are distinguished: scenario ξ is a *high* scenario if $v(0, \xi) > \lambda Z^*/\epsilon$, otherwise it is called *low*. Given the assumptions we made, the high scenarios are outliers that cause high second-stage cost, rather independently of the first-stage action. They should not affect the optimum $\hat{x}$ too much. Let Ξ_h and Ξ_ℓ denote the set of high and low scenarios, respectively.

Based on this distinction between scenarios we define the following functions for bounding the difference between f and $\hat{f}$

$$\hat{Q}_\ell(x) = \frac{1}{N}\sum_{i:\xi\in\Xi_\ell} v(x, \xi_i) \quad \text{and} \quad \hat{Q}_h(x) = \frac{1}{N}\sum_{i:\xi\in\Xi_h} v(x, \xi_i).$$

Notice that $\hat{f}(x) = c(x) + \hat{Q}_\ell(x) + \hat{Q}_h(x)$. Using $p = \Pr\{\xi \in \Xi_h\}$, we rewrite f similarly as $f(x) = c(x) + Q_\ell(x) + Q_h(x)$, with

$$Q_\ell(x) = (1-p)\mathbb{E}_{\boldsymbol{\xi}}\left[v(x, \boldsymbol{\xi}) \mid \boldsymbol{\xi} \in \Xi_\ell\right] \quad \text{and}$$
$$Q_h(x) = p\mathbb{E}_{\boldsymbol{\xi}}\left[v(x, \boldsymbol{\xi}) \mid \boldsymbol{\xi} \in \Xi_h\right].$$

We prove two preliminary lemmas that bound the differences $Q_\ell - \hat{Q}_\ell$ and $Q_h - \hat{Q}_h$, which will lead to Theorem 4.2 stating the approximation quality. In the proofs we use the Chernoff bound that we recall here for independent reading.

Lemma 4.1 (*Chernoff's Bound*). *Given independent random variables $Y_1, \dots, Y_N$ with $Y_i \in [0, 1]\ \forall i$, for every $\epsilon > 0$,*

$$\Pr\left\{\left|\sum_{i=1}^{N} Y_i - \mathbb{E}\left[\sum_{i=1}^{N} Y_i\right]\right| > \epsilon N\right\} \le 2\exp(-\epsilon^2 N). \quad \square$$

Lemma 4.2. *For $N = \Theta(\lambda^2 \frac{1}{\epsilon^4}\log|X| \log\frac{1}{\delta})$, with probability at least $1 - \delta$, $|Q_\ell(x) - \hat{Q}_\ell(x)| \le \epsilon Z^*\ \forall x \in X$.*

Proof. Fix $x \in X$. Let V_i, $i = 1, \dots, N$, denote independent identically distributed random variables defined as

$$V_i = \begin{cases} v(x, \boldsymbol{\xi}_i) & \text{if } \boldsymbol{\xi}_i \in \Xi_\ell, \\ 0 & \text{if } \boldsymbol{\xi}_i \in \Xi_h. \end{cases}$$

Then, $\hat{Q}_\ell(x) = 1/N\sum_{i=1}^{N} V_i$ and $Q_\ell(x) = \mathbb{E}\left[1/N\sum_{i=1}^{N} V_i\right]$. Since a scenario is called low if $v(0, \xi) \le \lambda Z^*/\epsilon$, we can normalize V_i to $Y_i = \frac{V_i\epsilon}{\lambda Z^*} \in [0, 1]$, $i = 1, \dots, N$. The random variables Y_i satisfy the assumptions of Chernoff's bound. Moreover, $\sum_{i=1}^{N} Y_i = \frac{N\epsilon}{\lambda Z^*}\hat{Q}_\ell(x)$ and $\mathbb{E}\left[1/N\sum_{i=1}^{N} Y_i\right] = \frac{N\epsilon}{\lambda Z^*}Q_\ell(x)$ so that

$$\begin{aligned}
&\Pr\left\{|\hat{Q}_\ell(x) - Q_\ell(x)| > \epsilon Z^*\right\} \\
&= \Pr\left\{\left|\frac{N\epsilon}{\lambda Z^*}\hat{Q}_\ell(x) - \frac{N\epsilon}{\lambda Z^*}Q_\ell(x)\right| > \frac{\epsilon^2 N}{\lambda}\right\} \\
&= \Pr\left\{\left|\sum_{i=1}^{N} Y_i - \mathbb{E}\left[1/N\sum_{i=1}^{N} Y_i\right]\right| > \frac{\epsilon^2 N}{\lambda}\right\} \\
&\le 2\exp\left(-\frac{\epsilon^4 N}{\lambda^2}\right).
\end{aligned}$$

Inserting $N = \Theta(\lambda^2 \frac{1}{\epsilon^4}\log|X|\log\frac{1}{\delta})$ makes this probability at most $\frac{\delta}{|X|}$. Obviously, we are not interested in the probability of the event $|\hat{Q}_\ell(x) - Q_\ell(x)| > \epsilon Z^*$ for a given $x \in X$, but in the probability that $|\hat{Q}_\ell(x) - Q_\ell(x)| > \epsilon Z^*$ holds for at least one $x \in X$. Using the finiteness of X and by using the union bound over all $x \in X$ we can bound this probability by δ. □

Lemma 4.3. *For $N = \Theta(\lambda^2 \frac{1}{\epsilon^4}\log|X|\log\frac{1}{\delta})$*

(a) *With probability at least $1-\delta$, $\hat{Q}_h(0) - \hat{Q}_h(x) \le 2\epsilon c(x)\ \forall x \in X$;*
(b) $Q_h(0) - Q_h(x) \le 2\epsilon c(x)\ \forall x \in X$.

Proof. For any $x \in X$, using (37) and the notation N_h for the number of high scenarios, we have

$$\hat{Q}_h(0) - \hat{Q}_h(x) = \frac{1}{N}\sum_{i:\xi\in\Xi_h}(v(0,\xi_i) - v(x,\xi_i)) \le \frac{N_h}{N}\lambda c(x), \tag{38}$$

and using (A3) and $p = \Pr\{\xi \in \Xi_h\}$, the probability that a high scenario occurs, we have for any $x \in X$,

$$\begin{aligned}
Q_h(0) - Q_h(x) &= p\left(\mathbb{E}_{\boldsymbol{\xi}}[v(0,\boldsymbol{\xi}) \mid \boldsymbol{\xi} \in \Xi_h]\right. \\
&\quad \left.- \mathbb{E}_{\boldsymbol{\xi}}[v(x,\boldsymbol{\xi}) \mid \boldsymbol{\xi} \in \Xi_h]\right) \le p\lambda c(x).
\end{aligned} \tag{39}$$

To bound p and N_h/N, we use the definition of high scenario and (37) to yield

$$\begin{aligned}
f(x^*) \ge Q_h(x^*) &= p\mathbb{E}_{\boldsymbol{\xi}}\left[v(x^*,\boldsymbol{\xi}) \mid \boldsymbol{\xi} \in \Xi_h\right] \\
&\ge p\left(\mathbb{E}_{\boldsymbol{\xi}}[v(0,\boldsymbol{\xi}) \mid \boldsymbol{\xi} \in \Xi_h] - \lambda c(x^*)\right) \\
&\ge p\left(\frac{\lambda f(x^*)}{\epsilon} - \lambda f(x^*)\right).
\end{aligned}$$

Hence, $p \le \frac{\epsilon}{(1-\epsilon)\lambda}$ if $\epsilon < 1$. Using this bound on p, another application of Chernoff's bound allows to establish that $\frac{N_h}{N} \le 2\frac{\epsilon}{\lambda}$ with probability at least $1-\delta$. Inserting the bounds on N_h/N and p in (38) and (39), respectively, completes the proof. □

Theorem 4.2. *For $N = \Theta(\lambda^2 \frac{1}{\epsilon^4}\log|X|\log\frac{1}{\delta})$, we have*

$$\frac{f(\hat{x}_N)}{f(x^*)} = 1 + O(\epsilon) \quad \text{w.p. } 1 - 2\delta.$$

Proof. The two lemmas imply that with probability at least $1-2\delta$ both of them hold simultaneously. From (A2) it follows that, for $x \in X$, $Q_h(x) \le Q_h(0)$ so that $Q_h(x) - \hat{Q}_h(x) \le Q_h(0) - \hat{Q}_h(0) + \hat{Q}_h(0) - \hat{Q}_h(x)$. This observation together with Lemmas 4.2 and 4.3(a) imply that with probability at least $1-2\delta$,

$$f(x) - \hat{f}(x) \le \epsilon Z^* + 2\epsilon c(x) + Q_h(0) - \hat{Q}_h(0). \tag{40}$$

Similarly, Lemmas 4.2 and 4.3(b) together with assumption (A2) applied to $\hat{f}$ imply that, with probability at least $1-2\delta$,

$$\hat{f}(x) - f(x) \le \epsilon Z^* + 2\epsilon c(x) + \hat{Q}_h(0) - Q_h(0). \tag{41}$$

Adding (40) with $x = \hat{x}$ and (41) with $x = x^*$, inserting $Z^* = f(x^*)$ and using that $\hat{f}(\hat{x}) \le \hat{f}(x^*)$, by definition, yields that with probability at least $1-2\delta$

$$f(\hat{x}) - 2\epsilon c(\hat{x}) \le f(x^*) + 2\epsilon Z^* + 2\epsilon c(x^*).$$

Using assumption (A1), implying that $c(x) \le f(x)$ and $c(x) \le \hat{f}(x)$ $\forall x \in X$, we have with probability at least $1-2\delta$,

$$f(\hat{x}) - 2\epsilon f(\hat{x}) \le f(x^*) + 2\epsilon f(x^*) + 2\epsilon f(x^*),$$

completing the proof. □

Thus if we have a two-stage stochastic programming problem, whose deterministic equivalent problem we can solve in polynomial time, then we have a fully polynomial-time randomized approximation scheme (FPRAS) for the black box model, assuming that the inflation factor λ is bounded by a polynomial function of the first-stage parameters, which is a minor assumption.

However, as we have seen, the deterministic equivalent formulation of the stochastic problem may have a different complexity than the deterministic counterpart of the problem. This may be even more dramatic than in the example of the previous section: e.g., the deterministic equivalent formulation of a two-stage stochastic shortest path problem in which the destination is random and edges can be bought at a cheaper price in the first stage is an NP-hard problem.

Therefore it is useful to extend the translation from the deterministic equivalent problem to the black box model to approximation algorithms, again with only a $1 + O(\epsilon)$ loss in performance ratios. Direct application of SAA to an α-approximate algorithm for minimizing $\hat{f}(x)$ does not lead to an $\alpha(1+O(\epsilon))$-approximate algorithm for minimizing $f(x)$.

A trick that does work is to minimize $c(x) + \hat{Q}_\ell(x)$, forgetting about the high scenarios. Since this requires knowledge of Z^*, instead we just select a number of samples with the lowest values $v(0,\xi_i)$. Again these values may be hard to compute, whence we assume that we have a β-approximation for evaluating $v(0,\xi_i)$. Redefine the number of samples to be $N' = N\beta^2$ and order them such that $v(0,\xi^{(1)}) \le \cdots \le v(0,\xi^{(N')})$. Then for $\gamma = 1 - \frac{2\epsilon}{\beta\gamma}$ define

$$\bar{Q}_\ell(x) = \frac{1}{N'}\sum_{i=1}^{\gamma N'} v(x,\xi^{(i)}) \quad \text{and} \quad g(x) = c(x) + \bar{Q}_\ell(x).$$

Theorem 4.3. *Choosing N' samples, any α-approximate solution for minimizing $g(x)$ gives with probability $1 - O(\delta)$ an $\alpha(1+O(\epsilon))$-approximate solution for minimizing $f(x)$.*

Proof. We only give a sketch of the proof. Working with β-approximate values for $v(0,\xi)$, we should adapt the threshold for being a high scenario at $\frac{\lambda\beta Z^*}{\epsilon}$ and adjust all the arguments above, essentially by replacing everywhere λ by $\lambda\beta$.

Then it is shown that $|\bar{Q}_\ell(x) - Q_\ell(x)| \le 3\epsilon Z^*$ with probability $1-\delta$. And eventually it is shown that an α-approximate solution of $g(x)$ is with probability $1 - O(\delta)$ an $\alpha(1+O(\epsilon))$-approximate solution for minimizing $f(x)$. □

Under linear first-stage cost and a form of Lipschitz continuity, all above results can be translated to the continuous case, implying, a.o., that under these particular conditions there exists an FPRAS for two-stage stochastic linear programming under the black box model.

5. Notes

Stochastic programming models date back to the fifties [19,20]. Several surveys on stochastic programming have appeared of which we mention here the introductory book of Kall and Wallace [15] and the comprehensive books by Prékopa [2], Shapiro et al. [21], and Birge and Louveaux [14]. For surveys specifically on stochastic integer programming we refer to the chapter by Louveaux and Schultz in the Handbook on Stochastic Programming [4], and the survey papers by Klein Haneveld and Van der Vlerk [22], Römisch and Schultz [23], Sen [24], and Stougie and Van der Vlerk [25]. Resources on the Internet are the *Stochastic Programming Community Home Page* [1] and the bibliography [26].

The focus in this survey is on the two-stage recourse model. For a detailed discussion of the multistage model and generalizations (including random recourse matrices and nonlinear models) we refer to [4,14,15,2]. More about the important class of chance-constrained problems and the related (conditional) value at risk models can be found in [4]. This class of problems is very well surveyed in [2,27].

The mathematical properties of two-stage stochastic linear programming problems have been derived by various people and at a rather early stage in the research activities on stochastic programming. In particular we refer to the overview by Wets [28] and the monograph by Kall [29].

The mathematical properties of two-stage stochastic integer programming problems have been established much more recently [9,30–32]. Schultz [33] proved the properties of the mixed-integer recourse function presented in Theorem 1.2. In addition, Schultz presented conditions for Lipschitz continuity of the function Q.

The results in Section 2 are selected from [34]. $\sharp P$-completeness of the problem GRAPH RELIABILITY has been proved in [35]. That exact evaluation of the second-stage expected value function may not even be in *PSPACE* in case random parameters are continuously distributed follows from a result in [36].

Dyer and Stougie [34] also prove *PSPACE*-hardness of a specific non-standard version of a multistage stochastic programming problem if the number of stages is considered to be part of the input. The complexity of standard multistage stochastic programming remains unsolved.

The idea in Section 3 of approximating the expected value function of a stochastic programming problem with integer recourse by a convex function through perturbing the distributions of the random right-hand sides is due to Klein Haneveld et al. [11]. They implemented this idea for the case of simple integer recourse. See Van der Vlerk [37] for a generalization to multiple simple recourse models, allowing for piecewise-linear penalty cost functions. The extension to the complete integer recourse case was done by Van der Vlerk [12], as well as the extension to mixed-integer recourse models with a single recourse constraint [38].

For the problem with simple integer recourse, the formula and properties in Lemma 3.1 have been derived by Louveaux and Van der Vlerk [31], while the characterization of all probability distributions that lead to convex expected value functions in Theorem 3.1 is due to Klein Haneveld et al. [11].

The uniform bounds on the absolute errors of the α-approximation in Theorem 3.2 and of the $\alpha\beta$-approximation in (11) are from [11]. There it is also shown that the latter error bound cannot be reduced by using other convex combinations of probability density functions of type f_α. The error bounds are derived in case the distributions of the random right hand sides are continuous. For the case with discretely distributed $\boldsymbol{h}$ it is possible to construct the convex envelope of the function Q, see [10].

Algorithms for continuous simple recourse problems with discretely distributed right-hand side parameters can be found in e.g. [14,15]. Using the structure of such problems, they can be represented as relatively *small* deterministic LP problems.

If the matrix W is complete but not TU, then the function $Q_{\alpha^\star}$ defined in Theorem 3.3 can be used as a convex lower bounding approximation of the function Q, allowing to approximately solve the integer recourse problem by solving a continuous complete recourse model. Although this approach is easy to implement and in many cases will give better results than using the LP lower bound Q^{LP}, no (non-trivial) bound on the approximation error is known. Indeed, in most applications the approximation will not be good enough for this purpose. On the other hand, because of the properties discussed in Section 3, the function $Q_{\alpha^\star}$ is well-suited as a building block in special-purpose algorithms for integer complete recourse models; several of these algorithms [39–42] use the LP relaxation Q^{LP} for bounding purposes.

Performance analysis of polynomial-time approximation algorithms for stochastic integer programming problems started in the early eighties, concentrating on so-called probabilistic analysis of algorithms for hierarchical planning problems, leading to asymptotic quality statements. In order to concentrate on worst-case performance analysis and on algorithms that are linear programming based, we do not cover these results here. The interested reader is referred to [9] and for a more recent result to [43].

The work on worst-case *performance ratio* analysis of polynomial-time algorithms for two-stage stochastic integer programming, described in Section 4.1, appeared in Dye et al. [17], an updated journal version of a technical report from 1999. It was the first paper of its kind, concerning a result in the scenario-based model (cf. Section 2). A more recent stream of papers using this type of analysis resulted in some important contributions. This stream has been preluded by some results on two-stage stochastic linear programming, all of them holding in fact for the black box model (cf. Section 2), even if that term had not been coined yet by then.

Though their work became recognized much later, in 2000 Nesterov and Vial [44] were the first to propose a randomized approximation scheme for two-stage stochastic linear programming based on stochastic subgradients in a subgradient-descent approach for optimization of the convex objective function. The stochastic subgradients allow to avoid expensive function evaluations (cf. Section 2). Independently, a little later Dyer et al. [45] designed a randomized approximation scheme for the two-stage stochastic linear programming problem based on a Markov Chain Monte Carlo method for approximating function values embedded in a simple randomized local search procedure. Also this work is only recently accepted by a journal. It was not until the work of Shmoys and Swamy [46] and Shapiro and Nemirovski [47] that this complexity related approximation theory of stochastic programming started to attract a lot more attention.

As a tool for using LP-rounding based approximation algorithms for two-stage stochastic integer programming problems, Shmoys and Swamy [46] designed a stochastic subgradient based approximation scheme for two-stage stochastic linear programming embedded in an ellipsoid method. For recourse models with inflation factors polynomially bounded in basic input parameters (see Section 4.2), they prove their method to be an FPRAS. They also prove that the bound on the inflation factor is unavoidable for such strong results, unless $P = NP$. They then use this FPRAS as the basis for a technique to round solutions of LP relaxations of stochastic integer problems, to obtain strong approximation results for the latter.

Shapiro and Nemirovski [47] showed how these results can be phrased in the context of the sample average algorithm for stochastic discrete optimization problems, introduced by Kleywegt et al. [48]. The strong results on this method described in Section 4.2 within the black box model (coined in [46]) are by Charikar et al. [49]. They hold for any two-stage stochastic recourse problem with a uniformly bounded inflation factor and overruled the interest in the independent model, mentioned above.

Just before the last cited results, papers on two-stage stochastic integer programming problems appeared. In the spirit of Dye et al. [17], using the scenario-based model, Ravi and Sinha [50] presented polynomial-time approximation algorithms with constant worst-case performance ratios for natural two-stage recourse versions of classical combinatorial optimization problems, like Shortest Path with stochastic destination and a higher price for buying edges in the second stage, when the destination is known, than in the first stage, or Vertex Cover with first-stage and scenario dependent cost of buying vertices to cover a random edge set. The results are obtained through formulating the deterministic equivalent stochastic integer programming problem, solving the LP relaxation and rounding its solution in a clever way.

Simultaneously, and independently, Immorlica et al. [18] studied the same problems. They presented moreover strong approximation results for the independent model (cf. Sections 2 and 4.1). To this end, they assume that the inflation factor, defined in Section 4.1, is a constant.

Among the results within the black box model, we like to mention the fundamental result of Gupta et al. [51]. In the spirit of [49], they employ *boosted sampling* to arrive at the formulation of a first-stage problem, i.e., without inflation factors, as a robust optimization problem over a set of sampled scenarios, where the sample size can be restricted to the bound on the inflation factor. It allows them to derive strong approximation results for stochastic versions of several classical combinatorial optimization problems like the Steiner Tree, Vertex Cover and Facility Location problem.

A final and more recent result we mention here is by Bertsimas and Goyal [52] who show that, under the right assumptions, the robust optimization version of the general two-stage stochastic mixed-integer linear programming problem gives a solution with worst-case performance ratio at most 2. Their analysis is more related to the one in Section 3 since they obtain their result by a careful mathematical analysis of the objective function. Among the assumptions they use are symmetry properties of the sample space and distribution, and integrality restrictions on first-stage variables only. Multistage versions of some of the above mentioned papers are found in [51,53,54].

Worst-case performance analysis in stochastic integer programming remains a rich research field. Many specific problems, including many stochastic versions of scheduling problems, have not yet been studied.

Acknowledgments

We acknowledge the suggestions of David Shmoys for Section 4. The work of Leen Stougie was partially supported by the Tinbergen Institute. The work of Ward Romeijnders was supported by The Netherlands Organisation for Scientific Research (NWO).

References

[1] COSP. Stochastic programming community home page. World Wide Web, http://stoprog.org, 2013.
[2] A. Prékopa, Stochastic Programming, Kluwer Academic Publishers, Dordrecht, 1995.
[3] R.J.-B. Wets, Stochastic programs with fixed recourse: the equivalent deterministic program, SIAM Rev. 16 (1974) 309–339.
[4] A. Ruszczyński, A. Shapiro (Eds.), Stochastic Programming, in: Handbooks in Operations Research and Management Science, vol. 10, North-Holland, 2003.
[5] M.R. Garey, D.S. Johnson, Computers and Intractability; A Guide to the Theory of NP-Completeness, Freeman, New York, 1979.
[6] C.H. Papadimitriou, Computational Complexity, Addison-Wesley, Reading, Massachusetts, 1994.
[7] A. Schrijver, Theory of Linear and Integer Programming, Wiley, Chichester, 1986.
[8] M. Grötschel, L. Lovász, A. Schrijver, Geometric Algorithms and Combinatorial Optimization, Springer-Verlag, Berlin, 1986.
[9] L. Stougie, Design and Analysis of Algorithms for Stochastic Integer Programming, in: CWI Tract, vol. 37, Centrum voor Wiskunde en Informatica, Amsterdam, 1987.
[10] W.K. Klein Haneveld, L. Stougie, M.H. van der Vlerk, An algorithm for the construction of convex hulls in simple integer recourse programming, Ann. Oper. Res. 64 (1996) 67–81.
[11] W.K. Klein Haneveld, L. Stougie, M.H. van der Vlerk, Simple integer recourse models: convexity and convex approximations, Math. Program. 108 (2006) 435–473.
[12] M.H. van der Vlerk, Convex approximations for complete integer recourse models, Math. Program. 99 (2) (2004) 297–310.
[13] W. Romeijnders, M.H. van der Vlerk, Convex hull approximation of TU integer recourse models: counterexamples, sufficient conditions, and special cases. Stochastic Programming E-Print Series. 2012.
[14] J.R. Birge, F.V. Louveaux, Introduction to Stochastic Programming, Springer Verlag, New York, 1997.
[15] P. Kall, S.W. Wallace, Stochastic Programming, Wiley, Chichester, 1994, Also available as PDF file, e.g. via http://stoprog.org/resources.html#textbooks.
[16] S. Dye, L. Stougie, A. Tomasgard, Approximation algorithms and relaxations for a service provision problem on a telecommunication network, Discrete Appl. Math. 129 (2003) 63–81.
[17] S. Dye, L. Stougie, A. Tomasgard, The stochastic single node service provision problem, Nav. Res. Logist. Q. 50 (2003) 869–887.
[18] N. Immorlica, D. Karger, M. Minkoff, V. Mirrokni, On the costs and benefits of procrastination: approximation algorithms for stochastic combinatorial optimization problems, in: Proceedings of the 15th SIAM–ACM Symposium on Discrete Algorithms (SODA), 2003, pp. 684–693.
[19] G.B. Dantzig, Linear programming under uncertainty, Manag. Sci. 1 (1955) 197–206.
[20] A. Charnes, W.W. Cooper, G.H. Symonds, Cost horizons and certainty equivalents: an approach to stochastic programming of heating oil, Manag. Sci. 4 (1958) 183–195.
[21] A. Shapiro, D. Dentcheva, A. Ruszczyński, Lectures on Stochastic Programming: Modeling and Theory, SIAM, Philadelphia, 2009.
[22] W.K. Klein Haneveld, M.H. van der Vlerk, Stochastic integer programming: general models and algorithms, Ann. Oper. Res. 85 (1999) 39–57.
[23] W. Römisch, R. Schultz, Multistage stochastic integer programs: an introduction, in: M. Grötchel, S.O. Krumke, J. Rambau (Eds.), Online Optimization of Large Scale Systems, Springer-Verlag, Berlin, 2001, pp. 581–600.
[24] S. Sen, Algorithms for stochastic mixed-integer programming models, in: K. Aardal, G.L. Nemhauser, R. Weismantel (Eds.), Discrete Optimization, in: Handbooks in Operations Research and Management Science, vol. 12, North-Holland, 2005, pp. 515–558.
[25] L. Stougie, M.H. van der Vlerk, Stochastic integer programming, in: M. Dell'Amico, F. Maffioli, S. Martello (Eds.), Annotated Bibliographies in Combinatorial Optimization, Wiley, 1997, pp. 127–141. Chapter 9.
[26] M.H. van der Vlerk, Stochastic programming bibliography. World Wide Web, http://www.eco.rug.nl/spbib.html, 1996–2007.
[27] S.P. Uryasev (Ed.), Probabilistic Constrained Optimization, Kluwer Academic Publishers, Dordrecht, 2000.
[28] R.J.-B. Wets, Stochastic programming: solution techniques and approximation schemes, in: A. Bachem, M. Grötschel, B. Korte (Eds.), Mathematical Programming, The State of the Art, Springer-Verlag, Berlin, 1982, pp. 566–603.
[29] P. Kall, Stochastic Linear Programming, Springer-Verlag, Berlin, 1976.
[30] R. Schultz, Continuity properties of expectation functions in stochastic integer programming, Math. Oper. Res. 18 (1993) 578–589.
[31] F.V. Louveaux, M.H. van der Vlerk, Stochastic programming with simple integer recourse, Math. Program. 61 (1993) 301–325.
[32] M.H. van der Vlerk, Stochastic programming with integer recourse (Ph.D. thesis), University of Groningen, The Netherlands, 1995.
[33] R. Schultz, On structure and stability in stochastic programs with random technology matrix and complete integer recourse, Math. Program. 70 (1995) 73–89.
[34] M. Dyer, L. Stougie, Computational complexity of stochastic programming problems, Math. Program. 106 (3) (2006) 423–432.
[35] L.G. Valiant, The complexity of enumeration and reliability problems, SIAM J. Comput. 8 (1979) 410–421.
[36] J. Lawrence, Polytope volume computation, Math. Comp. 57 (1991) 259–271.
[37] M.H. van der Vlerk, On multiple simple recourse, Math. Methods Oper. Res. 62 (2005) 225–242.
[38] M.H. van der Vlerk, Convex approximations for a class of mixed-integer recourse models, Ann. Oper. Res. 177 (2010) 139–150.
[39] S. Ahmed, M. Tawarmalani, N.V. Sahinidis, A finite branch-and-bound algorithm for two-stage stochastic integer programs, Math. Program. 100 (2) (2004) 355–377. Ser. A.
[40] G. Laporte, F.V. Louveaux, The integer L-shaped method for stochastic integer programs with complete recourse, Oper. Res. Lett. 13 (1993) 133–142.

[41] V.I. Norkin, Yu.M. Ermoliev, A. Ruszczyński, On optimal allocation of indivisibles under uncertainty, Oper. Res. 46 (3) (1998) 381–395.

[42] R. Schultz, L. Stougie, M.H. van der Vlerk, Solving stochastic programs with complete integer recourse: a framework using Gröbner Bases, Math. Program. 83 (2) (1998) 229–252.

[43] S. Ahmed, N.V. Sahinidis, An approximation scheme for stochastic integer programs arising in capacity expansion, Oper. Res. 51 (3) (2003) 461–471.

[44] Yu. Nesterov, J.-P. Vial, Confidence level solutions for stochastic programming, Automatica 44 (6) (2008) 1559–1568.

[45] M. Dyer, R. Kannan, L. Stougie, A simple randomised algorithm for convex optimisation, Math. Program. (2013) http://dx.doi.org/10.1007/s10107-013-0718-0.

[46] D.B. Shmoys, C. Swamy, An approximation scheme for stochastic linear programming and its application to stochastic integer programs, J. ACM 53 (6) (2006) 978–1012.

[47] A. Nemirovski, A. Shapiro, On complexity of Shmoys–Swamy class of two-stage linear stochastic programming problems, Optimization Online (2006).

[48] A.J. Kleywegt, A. Shapiro, T. Homem-de Mello, The sample average approximation method for stochastic discrete optimization, SIAM J. Optim. 12 (2) (2002) 479–502.

[49] M. Charikar, C. Chekuri, M. Pál, Sampling bounds for stochastic optimization, in: Proceedings of the 9th International Workshop on Randomization and Computation, 2005, pp. 257–269.

[50] R. Ravi, A. Sinha, Hedging uncertainty: approximation algorithms for stochastic optimization problems, Math. Program. 108 (1) (2006) 97–114.

[51] A. Gupta, M. Pál, R. Ravi, A. Sinha, Sampling and cost-sharing: approximation algorithms for stochastic optimization problems, SIAM J. Comput. 40 (5) (2011) 1361–1401.

[52] D. Bertsimas, V. Goyal, On the power of robust solutions in two-stage stochastic and adaptive optimization problems, Math. Oper. Res. 35 (2) (2010) 284–305.

[53] C. Swamy, D.B. Shmoys, Sampling-based approximation algorithms for multistage stochastic optimization, SIAM J. Comput. 41 (4) (2012) 975–1004.

[54] D. Bertsimas, V. Goyal, X.A. Sun, A geometric characterization of the power of finite adaptability in multistage stochastic and adaptive optimization, Math. Oper. Res. 36 (1) (2011) 24–54.