University of Groningen

Pseudo-Valid Cutting Planes for Two-Stage Mixed-Integer Stochastic Programs with Right-Hand-Side Uncertainty

Romeijnders, Ward; van der Laan, Niels

Published in:
Operations Research

DOI:
[10.1287/opre.2019.1905](https://doi.org/10.1287/opre.2019.1905)

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

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

Publication date:
2020

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Romeijnders, W., & van der Laan, N. (2020). Pseudo-Valid Cutting Planes for Two-Stage Mixed-Integer Stochastic Programs with Right-Hand-Side Uncertainty. *Operations Research*, 68(4), 1199-1217. <https://doi.org/10.1287/opre.2019.1905>

Copyright

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

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

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

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

Operations Research

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Pseudo-Valid Cutting Planes for Two-Stage Mixed-Integer Stochastic Programs with Right-Hand-Side Uncertainty

Ward Romeijnnders, Niels van der Laan

To cite this article:

Ward Romeijnnders, Niels van der Laan (2020) Pseudo-Valid Cutting Planes for Two-Stage Mixed-Integer Stochastic Programs with Right-Hand-Side Uncertainty. Operations Research 68(4):1199-1217. <https://doi.org/10.1287/opre.2019.1905>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2020, INFORMS

Please scroll down for article—it is on subsequent pages

With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Methods

Pseudo-Valid Cutting Planes for Two-Stage Mixed-Integer Stochastic Programs with Right-Hand-Side Uncertainty

Ward Romeijnders,^a Niels van der Laan^a

^aDepartment of Operations, University of Groningen, 9700 AV Groningen, Netherlands

Contact: w.romeijnders@rug.nl, <https://orcid.org/0000-0002-6918-5635> (WR); n.van.der.laan@rug.nl,
<https://orcid.org/0000-0002-8045-4517> (NvdL)

Received: September 21, 2018

Revised: February 12, 2019; May 10, 2019

Accepted: June 28, 2019

Published Online in Articles in Advance:
May 7, 2020

Subject Classifications: stochastic
programming; integer programming

Area of Review: Optimization

<https://doi.org/10.1287/opre.2019.1905>

Copyright: © 2020 INFORMS

Abstract. We propose a novel way of applying cutting plane techniques to two-stage mixed-integer stochastic programs with uncertainty in the right-hand side. Instead of using cutting planes that are always valid, our idea is to apply pseudo-valid cutting planes to the second-stage feasible regions that may cut away feasible integer second-stage solutions for some scenarios and may be overly conservative for others. The advantage is that it allows us to use cutting planes that are affine in the first-stage decision variables, so that the approximation is convex and can be efficiently solved using techniques from convex optimization. We derive tight performance guarantees for using particular types of pseudo-valid cutting planes for simple integer recourse models. Moreover, we show in general that using pseudo-valid cutting planes leads to good first-stage solutions if the total variations of the one-dimensional conditional probability density functions of the random variables in the model converge to zero.

Funding: This work was supported by The Netherlands Organisation for Scientific Research (NWO) [Grant 451-17-034 4043].

Keywords: stochastic programming • integer programming • cutting plane techniques • convex approximations

1. Introduction

Many practical problems under uncertainty, for example, in energy, finance, logistics, and healthcare, involve integer decision variables. Such problems can be modeled as mixed-integer stochastic programs (MISPs), but are notoriously difficult to solve. In this paper, we do not attempt to solve these problems exactly. Instead, we introduce a novel approach to approximately solve two-stage MISPs with uncertainty in the right-hand side, and we derive performance guarantees for the resultant approximating solutions.

Traditional solution methods for MISPs combine solution approaches for continuous stochastic programs and deterministic mixed-integer programs (MIPs). See, for example, Ahmed et al. (2004) for branch-and-bound; Sen and Hingle (2005) and Ntaimo (2010) for disjunctive decomposition; Carøe and Schultz (1999) for dual decomposition; Laporte and Louveaux (1993) for the integer L-shaped method; and Zhang and Küçükyavuz (2014) for cutting plane techniques. All these solution methods aim to find the exact optimal solution for MISPs, but generally have difficulties scaling up to solve large problems. This is not surprising, because contrary to their continuous counterparts, these MISPs are nonconvex in general (Rinnooy Kan and Stougie 1988). This means that

efficient techniques from convex optimization cannot be used to solve these problems.

Based on this observation and inspired by the success of cutting plane techniques for deterministic MIPs (Marchand et al. 2002), we propose to use *cutting planes* to solve two-stage MISPs. However, we will use them in a fundamentally different way than in existing methods for both deterministic and stochastic MIPs. Instead of using exact cutting planes that are always valid, we propose to use *pseudo-valid* cutting planes for the second-stage feasible regions in such a way that the approximating problem remains convex in the first-stage decision variables, and thus efficient convex optimization techniques can be used to solve the approximation.

The disadvantage of using pseudo-valid cutting planes is that they may cut away part of the second-stage feasible region or that they may be overly conservative, so that we significantly over- or underestimate the second-stage costs, respectively. However, these disadvantages may be justified for MISPs, because our aim is to obtain good first-stage decisions, not find the exact and complete characterization of the integer hulls of the second-stage feasible regions. In fact, one of our main contributions is that we obtain good or even near-optimal *first-stage* decisions for pseudo-valid cutting plane

approximations in which the cutting planes only need to be valid for a grid of first-stage decisions.

For simple integer recourse (SIR) models, a special type of MISP, our pseudo-valid cutting plane approximation is equivalent to convex α -approximations, derived by Klein Haneveld et al. (2006) from a completely different perspective. By reinterpreting these α -approximations using pseudo-valid cutting planes, we connect two existing solution methodologies for MISPs that use convex approximations and exact cutting planes, respectively. Moreover, this reinterpretation allows us to apply existing performance guarantees derived in Romeijnders et al. (2015) for α -approximations to pseudo-valid cutting plane techniques for SIR models. Furthermore, we use results from Romeijnders et al. (2016a) to show for general MISPs with continuously distributed random right-hand-side parameters that the error of using pseudo-valid cutting plane approximations converges to zero if all total variations of the one-dimensional conditional probability density functions of these random parameters converge to zero.

Contrary to the convex approximation derived in Romeijnders et al. (2016a), our pseudo-valid cutting planes are suitable for solving large-scale MISPs. Indeed, in a numerical case study we solve large nurse scheduling problems using tight pseudo-valid cutting planes and obtain first-stage decisions that are close to optimal. Moreover, we derive pseudo-valid mixed-integer Gomory cuts for general two-stage MISPs with right-hand-side uncertainty.

Summarizing, the main contributions of our paper are as follows.

- We propose a novel solution approach for two-stage MISPs with uncertainty in the right-hand side by applying pseudo-valid cutting planes to the second-stage feasible regions.
- We reinterpret α -approximations for SIR models as pseudo-valid cutting plane approximations, connecting two existing solution methodologies for MISPs, and yielding a tight error bound for applying pseudo-valid cutting planes to SIR models.
- We derive a performance guarantee for applying pseudo-valid cutting planes to general MISPs with continuous distributions and uncertainty in the right-hand side.
- We derive pseudo-valid mixed-integer Gomory cuts for general MISPs and derive tight pseudo-valid cutting planes for a nurse scheduling problem.
- We carry out numerical experiments on a nurse scheduling problem and show that we obtain good first-stage decisions when using pseudo-valid cutting planes.

The remainder of this paper is organized as follows. In Section 2, we define MISPs and explain our pseudo-

valid cutting plane approach. In Section 3, we reinterpret α -approximations for SIR models using pseudo-valid cutting planes, and in Section 4, we derive an error bound for pseudo-valid cutting plane approximations for general MISPs. In Section 5, we derive pseudo-valid mixed-integer Gomory cuts, and apply tight pseudo-valid cutting planes to a nurse scheduling problem. We present numerical experiments for the latter type of cutting planes in Section 6, and we end with a discussion in Section 7.

2. Problem Definition and Solution Approach

2.1. Problem Definition

Two-stage MISPs can be interpreted as hierarchical planning problems. In the first stage, decisions x have to be made before some random parameters ω are known, whereas in the second stage, decisions y are made after the realizations of these random parameters ω are revealed. We assume that the probability distribution of ω is known, with F denoting the cumulative distribution function and Ω the support of ω . The MISPs that we consider are defined as

$$\eta^* := \min_{x,z} \{c^\top x + Q(z) : Ax = b, z = Tx, x \in X\}, \quad (1)$$

where $z = Tx \in \mathbb{R}^m$ represent tender variables. Moreover, the *expected value function* Q represents the expected second-stage costs

$$Q(z) := \mathbb{E}_\omega[v(\omega, z)], \quad z \in \mathbb{R}^m, \quad (2)$$

where the *second-stage value function* v is defined as

$$v(\omega, z) := \min_y \{q^\top y : Wy = \omega - z, y \in Y\}, \quad \omega \in \Omega, z \in \mathbb{R}^m. \quad (3)$$

The second-stage decisions y are also called *recourse actions*. Indeed, if $Tx = \omega$ represents random goal constraints, then the second-stage optimization problem v models all possible recourse actions y , and their corresponding costs, to compensate for infeasibilities of these goal constraints. Observe that we only consider randomness in the right-hand side of these goal constraints. Moreover, we assume that at least some of the second-stage decision variables y_i are restricted to be integer. This is captured by the feasible regions $X \subset \mathbb{R}_+^{n_1}$ and $Y \subset \mathbb{R}_+^{n_2}$ that may impose integrality restrictions on the first- and second-stage decision variables, respectively.

Throughout this paper, we make the following assumptions. The first is often referred to as the *complete recourse* assumption, meaning that there always exists a feasible recourse action y , ensuring that $v(\omega, z) < +\infty$ for all $\omega \in \Omega$ and $z \in \mathbb{R}^m$. The second is equivalent to

the dual feasible region of the LP-relaxation of v being nonempty, implying that $v(\omega, z) > -\infty$ for all $\omega \in \Omega$ and $z \in \mathbb{R}^m$. Together with the third assumption, these assumptions guarantee that $Q(z)$ is finite for every $z \in \mathbb{R}^m$.

Assumption 1. We assume that

- there exists $y \in Y$ such that $Wy = \omega - z$ for every $\omega \in \Omega$ and $z \in \mathbb{R}^m$,
- there exists $\lambda \in \mathbb{R}^m$ such that $W^\top \lambda \leq q$, and
- $\mathbb{E}_\omega[|\omega_i|] < +\infty$, for all $i = 1, \dots, m$.

2.2. Novel Solution Approach: Pseudo-Valid Cutting Planes

To solve the MISP defined in (1), we propose relaxing the integrality restrictions on the second-stage decision variables y and adding *pseudo-valid* cutting planes to the second-stage feasible region:

$$Y(\omega, z) := \{y \in Y : Wy = \omega - z\}.$$

In particular, we assume that the cutting planes are of the form $\hat{W}(\omega)y \geq \hat{h}(\omega) - \hat{T}(\omega)z$, so that they are affine in the tender variables z .

Definition 1. Consider the second-stage value function v defined in (3). Then we call $\hat{v}$ an *affine cutting plane approximation* of v if it is of the form

$$\hat{v}(\omega, z) = \min_y \{q^\top y : Wy = \omega - z, \hat{W}(\omega)y \geq \hat{h}(\omega) - \hat{T}(\omega)z, y \in \mathbb{R}_+^{n_2}\}, \quad \omega \in \Omega, z \in \mathbb{R}^m.$$

Moreover, we define the affine cutting plane approximation $\hat{Q}$ of the expected value function Q , defined in (2), as $\hat{Q}(z) := \mathbb{E}_\omega[\hat{v}(\omega, z)]$, $z \in \mathbb{R}^m$.

The main reason we use cutting planes that are *affine* in z is that the approximating value function $\hat{v}(\omega, z)$ with feasible region

$$\hat{Y}(\omega, z) := \left\{ y \in \mathbb{R}_+^{n_2} : \begin{array}{l} Wy = \omega - z \\ \hat{W}(\omega)y \geq \hat{h}(\omega) - \hat{T}(\omega)z \end{array} \right\}$$

is *convex* in z for every fixed $\omega \in \Omega$, and thus the corresponding approximating expected value function $\hat{Q}$ is convex. This means that the MISP in (1) with Q replaced by $\hat{Q}$ is much easier to solve than the original MISP.

Observation 1. Consider the affine cutting plane approximations $\hat{v}$ and $\hat{Q}$ of Definition 1. Then $\hat{Q}$ is convex, and $\hat{v}(\omega, z)$ is convex in z for every fixed $\omega \in \Omega$. $\square$

Obviously, we cannot expect to obtain a good approximation $\hat{Q}$ if we arbitrarily add invalid affine cutting planes. That is why we restrict our attention to so-called *pseudo-valid cutting planes*. Such cutting

planes are valid on a grid of points z , but may be invalid for other values of z .

Definition 2. Consider the second-stage value function v defined in (3). Then the cutting planes $\hat{W}(\omega)y \geq \hat{h}(\omega) - \hat{T}(\omega)z$ are called *pseudo-valid* for v if there exist $\alpha \in \mathbb{R}^m$ and $\beta \in \mathbb{Z}^m$ such that for all $z \in \alpha + \beta\mathbb{Z}^m$ and for all $\omega \in \Omega$,

$$\begin{aligned} & \{y \in \mathbb{Z}_+^{p_2} \times \mathbb{R}_+^{n_2-p_2} : Wy = \omega - z\} \\ & \subset \{y \in \mathbb{R}_+^{n_2} : Wy = \omega - z, \hat{W}(\omega)y \geq \hat{h}(\omega) - \hat{T}(\omega)z\}. \end{aligned} \quad (4)$$

If, in addition, $\hat{v}(\omega, z) = v(\omega, z)$ for all $z \in \alpha + \beta\mathbb{Z}^m$ and for all $\omega \in \Omega$, then we call the pseudo-valid cutting planes *tight*.

Remark 1. With a slight abuse of notation, we use $\alpha + \beta\mathbb{Z}^m$ to represent the grid of points:

$$\alpha + \beta\mathbb{Z}^m := \{(\alpha_1 + \beta_1 l_1, \dots, \alpha_m + \beta_m l_m) : l \in \mathbb{Z}^m\}.$$

Different classes of exact cutting planes can be applied as pseudo-valid cutting planes to two-stage MISPs. For example, in Section 5 we derive pseudo-valid mixed-integer Gomory cuts and tight pseudo-valid cutting planes for a nurse scheduling problem. However, our main focus is not on how to obtain the pseudo-valid cutting planes. Instead, we assume that they are given or can be iteratively generated by an algorithm, and we consider the performance of using such pseudo-valid cutting planes.

The performance of these pseudo-valid cutting planes may be surprisingly good, even if they cut away feasible integer second-stage solutions or admit second-stage solutions outside the integer hull $\tilde{Y}(\omega, z)$ of the second-stage feasible region $Y(\omega, z)$; in these cases, $\hat{v}(\omega, z)$ may significantly over- or underestimate $v(\omega, z)$, respectively. However, to obtain good first-stage decisions x , we do not require $\hat{v}(\omega, z)$ to be a good approximation of $v(\omega, z)$ for every $\omega \in \Omega$ and $z \in \mathbb{R}^m$, but merely require $\hat{v}(\omega, z)$ to be a good approximation of $v(\omega, z)$ on average for every $z \in \mathbb{R}^m$. This explains why applying pseudo-valid cutting planes may work for stochastic MIPs, but not for deterministic MIPs.

Using a one-dimensional example, we illustrate a class of pseudo-valid cutting planes.

Example 1. Consider a special case of the second-stage value function defined in (3), given by

$$\begin{aligned} v(\omega, z) &= \min_{y, u_1, u_2} qy + ru_1 + ru_2 \\ \text{s.t.} \quad & y - u_1 + u_2 = \omega - z \\ & y \in \mathbb{Z}_+, u_1, u_2 \in \mathbb{R}_+, \end{aligned} \quad (5)$$

where $0 < q < r$. By rewriting the equality in (5) as $u_2 = \omega - z - y + u_1$, we can eliminate the variable u_2 from the second-stage value function to obtain

$$\begin{aligned} v(\omega, z) &= r(\omega - z) + \min_{y, u_1} (q - r)y + 2ru_1 \quad (6) \\ \text{s.t. } &y - u_1 \leq \omega - z \\ &y \in \mathbb{Z}_+, u_1 \in \mathbb{R}_+. \end{aligned}$$

Because the minimization problem in (6) only has two decision variables, y and u_1 , we can graphically depict its feasible region $Y(\omega, z)$. The left panel in Figure 1 shows this feasible region for $\omega = 2.5$ and $z = 1$, and also depicts the feasible region of the LP-relaxation of $v(\omega, z)$. Clearly, the latter is larger than the integer hull $\tilde{Y}(\omega, z)$ of $Y(\omega, z)$.

It is well known that the integer hull $\tilde{Y}(\omega, z)$ can be obtained by adding a mixed-integer rounding (MIR) inequality, so that for every $\omega \in \Omega$ and $z \in \mathbb{R}$, the integer hull $\tilde{Y}(\omega, z)$ equals

$$\begin{aligned} \tilde{Y}(\omega, z) := \left\{ (y, u_1) \in \mathbb{R}_+^2 : y - u_1 \leq \omega - z, \right. \\ \left. y - \frac{1}{1 - (\omega - z) + \lfloor \omega - z \rfloor} u_1 \leq \lfloor \omega - z \rfloor \right\}. \end{aligned}$$

The right panel in Figure 1 shows $\tilde{Y}(\omega, z)$ and this MIR inequality.

Observe that the MIR inequality is not affine in z , which means that it will be hard to use for optimization purposes. However, if $z \in \mathbb{Z}$, then it reduces to

$$y - \frac{1}{1 - \omega + \lfloor \omega \rfloor} u_1 \leq \lfloor \omega \rfloor - z, \quad (7)$$

which means it is of the form of the affine cutting planes in Definition 1. Thus, a natural idea is to use the

cutting planes in (7), also when $z \notin \mathbb{Z}$. In the latter case, the cutting planes are not always valid. However, it is not hard to show that these cutting planes are *pseudo-valid* and *tight*.

Figure 2 shows the approximating feasible region

$$\begin{aligned} \hat{Y}(\omega, z) = \left\{ (y, u_1) \in \mathbb{R}_+^2 : y - u_1 \leq \omega - z, \right. \\ \left. y - \frac{1}{1 - \omega + \lfloor \omega \rfloor} u_1 \leq \lfloor \omega \rfloor - z \right\}, \end{aligned}$$

for $z = 0.5$ and $\omega = 1.5, 1.75, 2, 2.25$. We observe that for $\omega = 2$, the approximating MIR inequality coincides with the constraint $y - u \leq \omega - z$, so that $\hat{Y}(\omega, z)$ is equal to the feasible region of the LP-relaxation of $v(\omega, z)$, and thus admits solutions outside the integer hull $\tilde{Y}(\omega, z)$. For $\omega = 1.5$, on the other hand, the approximating MIR inequality cuts away feasible integer solutions. For $\omega = 1.75$ and $\omega = 2.25$, we see a combination of both.

In Section 5.2, we numerically assess the performance of the pseudo-valid cutting plane approximation,

$$\begin{aligned} \hat{v}(\omega, z) := r(\omega - z) + \min_{y, u_1} \{ (q - r)y + 2ru_1 : (y, u_1) \in \hat{Y}(\omega, z) \}, \\ \omega \in \Omega, z \in \mathbb{R}, \end{aligned} \quad (8)$$

and show that for a normally distributed random variable $\omega \sim N(\mu, \sigma^2)$, $\hat{Q}$ is a good approximation of Q for medium to large values of the standard deviation σ .

3. Pseudo-Valid Cutting Planes for Simple Integer Recourse Models

In this section, we show that existing convex approximations for simple integer recourse (SIR) models can be interpreted as pseudo-valid cutting plane

Figure 1. Illustration of the Feasible Region of $v(\omega, z)$ of Example 1 with $\omega = 2.5$ and $z = 1$

Notes. The feasible region $Y(\omega, z)$ is represented by the black dots and the thick black lines. In the left panel, the shaded region corresponds to the feasible region of the LP-relaxation of v , whereas in the right panel, the MIR inequality is added, and the dark shaded region represents the integer hull $\tilde{Y}(\omega, z)$ of the feasible region $Y(\omega, z)$ of $v(\omega, z)$.

Figure 2. Illustration of the Feasible Region of $v(\omega, z)$ of Example 1 with $z = 0.5$ and $\omega = 1.5, 1.75, 2$, and 2.25

Notes. The feasible region $Y(\omega, z)$ is represented by the black dots and the thick black lines. The dotted line represents the pseudo-valid MIR inequality defined in (7), and the shaded regions the approximating feasible region $\tilde{Y}(\omega, z)$.

approximations. SIR models are introduced by Louveaux and van der Vlerk, (1993) and can be considered the most-simple version of a MISP as defined in (1). For ease of exposition, we consider here the one-sided and one-dimensional version of SIR, where the second-stage value function v is defined as

$$v(\omega, z) = \min_y \{qy : y \geq \omega - z, y \in \mathbb{Z}_+\}, \quad \omega \in \Omega, z \in \mathbb{R}.$$

Observe that we can derive a closed-form expression for v , because for every $\omega \in \Omega$ and $z \in \mathbb{R}$, the optimal solution is $y^* = \lceil \omega - z \rceil^+ := \max\{0, \lceil \omega - z \rceil\}$, and thus $v(\omega, z) = q\lceil \omega - z \rceil^+$. Clearly, $v(\omega, z)$ is a nonconvex function of z because of the round-up operator.

We, however, focus on the feasible region $Y(\omega, z) = \{y \in \mathbb{Z}_+ : y \geq \omega - z\}$ and its integer hull

$$\tilde{Y}(\omega, z) = \{y \in \mathbb{R}_+ : y \geq \lceil \omega - z \rceil\}, \quad \omega \in \Omega, z \in \mathbb{R}.$$

Here, the cutting plane $y \geq \lceil \omega - z \rceil$ makes the original constraint $y \geq \omega - z$ redundant. Similar to Example 1,

this exact cutting plane is not affine in z and thus not suitable for optimization purposes. However, if $z \in \mathbb{Z}$, then the cutting plane is equivalent to $y \geq \lceil \omega \rceil - z$, which we can use as a pseudo-valid cutting plane for $z \notin \mathbb{Z}$. In fact, $y \geq \lceil \omega \rceil - z$ is a *tight* pseudo-valid cutting plane. We define a family of tight pseudo-valid cutting plane approximations $\hat{v}_\alpha$, each of them using the cutting plane $y \geq \lceil \omega - \alpha \rceil + \alpha - z$ that is exact for $z \in \alpha + \mathbb{Z}$.

Definition 3. For every $\alpha \in \mathbb{R}$, define the pseudo-valid cutting plane approximation $\hat{v}_\alpha$ for the SIR second-stage value function v as

$$\begin{aligned} \hat{v}_\alpha(\omega, z) &= \min_y \{qy : y \geq \lceil \omega - \alpha \rceil + \alpha - z, y \in \mathbb{R}_+\} \\ &= q(\lceil \omega - \alpha \rceil + \alpha - z)^+, \quad \omega \in \Omega, z \in \mathbb{R}. \end{aligned}$$

Moreover, define the corresponding pseudo-valid cutting plane approximation $\hat{Q}_\alpha$ for the SIR expected value function Q as $\hat{Q}_\alpha(z) = q\mathbb{E}_\omega[(\lceil \omega - \alpha \rceil + \alpha - z)^+]$, $z \in \mathbb{R}$.

Surprisingly, the pseudo-valid cutting plane approximation $\hat{Q}_\alpha$ equals the α -approximations of Klein Haneveld et al. (2006), derived from a completely different perspective. They first identify all probability distributions of ω for which the expected value function Q is convex. This is all continuous distributions with probability density function f satisfying $f(s) = G(s+1) - G(s)$, $s \in \mathbb{R}$, for some cumulative distribution function G with finite mean. For all other distributions, they use this condition to generate an approximating density function $\hat{f}$, resulting in a convex approximation $\hat{Q}$ of Q . Selecting $G(s+1) = F(\lceil s - \alpha \rceil + \alpha)$, $s \in \mathbb{R}$, yields the α -approximation $\hat{Q}_\alpha(z) := q\mathbb{E}_\omega[(\lceil \omega - \alpha \rceil + \alpha - z)^+]$, $z \in \mathbb{R}$, equivalent to the pseudo-valid cutting plane approximation of Definition 3.

In this paper, we reinterpret $\hat{Q}_\alpha$ as a pseudo-valid cutting plane approximation, connecting the convex approximation solution philosophy, introduced by van der Vlerk (1995) and continued by, among others, Klein Haneveld et al. (2006), Romeijnders et al. (2015, 2016a, b), van der Laan et al. (2018), and van der Vlerk (2004, 2010), with exact cutting plane techniques for MISP, studied in, for example, Carøe and Tind (1997), Ntamo and Sen (2005), Sen and Sherali (2006), Sherali and Zhu (2006), Guan et al. (2009), Gade et al. (2014), Zhang M and Küçükyavuz (2014), Bodur et al. (2017), Qi and Sen (2017), Bansal et al. (2018). This is particularly relevant, because performance guarantees are available for using convex approximations that may be used for pseudo-valid cutting plane approximations. In fact, for SIR models, Romeijnders et al. (2015) derive an upper bound on $\|Q - \hat{Q}_\alpha\|_\infty := \sup_{z \in \mathbb{R}} |Q(z) - \hat{Q}_\alpha(z)|$ for every $\alpha \in \mathbb{R}$, that depends on the *total variation* of the probability density function f of the random variable ω .

Definition 4. Let $f: \mathbb{R} \rightarrow \mathbb{R}$ be a real-valued function and let $I \subset \mathbb{R}$ be an interval. Let $\Pi(I)$ denote the set of all finite ordered sets $P = \{x_1, \dots, x_{N+1}\}$ with $x_1 < \dots < x_{N+1}$ in I . Then the *total variation* of f on I , denoted $|\Delta|f(I)$, is defined as

$$|\Delta|f(I) = \sup_{P \in \Pi(I)} V_f(P),$$

where $V_f(P) = \sum_{i=1}^N |f(x_{i+1}) - f(x_i)|$. We write $|\Delta|f := |\Delta|f(\mathbb{R})$.

Theorem 1. Consider the SIR expected value function $Q(z) = q\mathbb{E}_\omega[(\omega - z)^+]$, $z \in \mathbb{R}$, and its pseudo-valid cutting plane approximation $\hat{Q}_\alpha(z) = q\mathbb{E}_\omega[(\lceil \omega - \alpha \rceil - \alpha - z)^+]$, $z \in \mathbb{R}$, for $\alpha \in \mathbb{R}$. Then, for every continuous random variable ω with probability density function f , we have

$$\|Q - \hat{Q}_\alpha\|_\infty \leq qh(|\Delta|f),$$

where $h: [0, \infty) \mapsto \mathbb{R}$ is defined as

$$h(|\Delta|f) = \begin{cases} |\Delta|f/8, & |\Delta|f \leq 4, \\ 1 - 2/|\Delta|f, & |\Delta|f \geq 4. \end{cases}$$

Proof. See Romeijnders et al. (2015). $\square$

Remark 2. Romeijnders et al. (2015) actually derive an error bound for the α -approximations of van der Vlerk (2010) for integer recourse models with totally unimodular recourse matrix W . Also, these α -approximations can be interpreted as tight pseudo-valid cutting plane approximations.

Theorem 1 shows that the error bound is smaller for lower values of the total variation $|\Delta|f$ of the probability density function f of ω . For unimodal density functions, such as the normal density function in Example 2 below, this total variation $|\Delta|f$ is small if the variance of the random variable ω is large. Thus, in general we conclude from Theorem 1 that the larger the variability of the random variable in the model, the better the pseudo-valid cutting plane approximation.

Example 2. Let ω be a normal random variable with mean μ and standard deviation σ . Then the probability density function f of ω is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\}, \quad x \in \mathbb{R},$$

which is unimodal with mode μ , and thus has total variation $|\Delta|f = 2f(\mu) = \sigma^{-1}\sqrt{2/\pi}$. Hence, if the standard deviation σ increases, then the total variation $|\Delta|f$ of f decreases, and thus the upper bound on $\|Q - \hat{Q}_\alpha\|_\infty$ in Theorem 1 decreases. In other words, if the standard deviation is large, then $\hat{Q}_\alpha$ is a close approximation of Q , and thus the resultant approximating first-stage decision $\hat{x}_\alpha$ will be good.

4. Pseudo-Valid Cutting Plane Approximations for General MISP

In this section, we consider tight pseudo-valid cutting plane approximations $\hat{Q}$ for general MISP, and we derive an upper bound on $\|Q - \hat{Q}\|_\infty$ for the case that the random right-hand-side vector is continuously distributed. We use $\|Q - \hat{Q}\|_\infty$ to measure the error of the approximation $\hat{Q}$, because it can be used to bound the optimality gap, $c^\top \hat{x} + Q(\hat{x}) - \eta^*$, of the approximating solution $\hat{x}$ obtained by solving (1) with Q replaced by $\hat{Q}$. Indeed, Romeijnders et al. (2015) show that for any approximation $\hat{Q}$, we have

$$c^\top \hat{x} + Q(\hat{x}) - \eta^* \leq 2\|Q - \hat{Q}\|_\infty.$$

We prove that for tight pseudo-valid cutting plane approximations $\hat{Q}$, the error $\|Q - \hat{Q}\|_\infty$ vanishes if the

total variations of the one-dimensional conditional probability density functions of the random vector ω in the model go to zero. For example, for normally distributed ω this means that the solutions obtained by using tight pseudo-valid cutting planes are good if the variance of ω is large enough. The final result is Theorem 2, which is conveniently stated below. If the pseudo-valid cuts are not tight, then we use $\hat{Q}$ to derive a lower bound for Q in Corollary 1.

Definition 5. For every $i = 1, \dots, m$ and $t \in \mathbb{R}^m$, we let $t_{-i} \in \mathbb{R}^{m-1}$ denote the vector t without its i th component.

Definition 6. For every $i = 1, \dots, m$ and $t_{-i} \in \mathbb{R}^{m-1}$, define the i th conditional density function $f_i(\cdot|t_{-i})$ of the m -dimensional joint pdf f as

$$f_i(t_i|t_{-i}) = \begin{cases} \frac{f(t)}{f_{-i}(t_{-i})}, & f_{-i}(t_{-i}) > 0, \\ 0, & f_{-i}(t_{-i}) = 0, \end{cases}$$

where f_{-i} represents the joint density function of ω_{-i} , the random vector obtained by removing the i th element of ω .

Definition 7. Let $\mathcal{H}^m$ denote the set of all m -dimensional joint pdfs f whose conditional density functions $f_i(\cdot|t_{-i})$ are of bounded variation.

Theorem 2. Consider the mixed-integer recourse function Q and its tight pseudo-valid cutting plane approximation $\hat{Q}$. Then, under Assumption 1, there exists a constant $C \in \mathbb{R}$ with $C > 0$ such that for all ω with pdf $f \in \mathcal{H}^m$,

$$\|Q - \hat{Q}\|_\infty \leq C \sum_{i=1}^m \mathbb{E}_{\omega_{-i}}[|\Delta|f_i(\cdot|\omega_{-i})].$$

If all components of the random vector ω are independent, then $|\Delta|f_i(\cdot|\omega_{-i}) = |\Delta|f_i$ for every $i = 1, \dots, m$ and $\omega_{-i} \in \mathbb{R}^{m-1}$, and thus the error bound in Theorem 2 reduces to $C \sum_{i=1}^m |\Delta|f_i$, where $|\Delta|f_i$ is the total variation of the marginal probability density function f_i of ω_i , $i = 1, \dots, m$. Hence, if the components of ω are independent and normally distributed, then the error bound in Theorem 2 is small if the standard deviations of the components of ω are large, see Example 2.

In the next corollary, we derive a lower bound for Q that holds irrespective of whether or not the pseudo-valid cuts are tight.

Corollary 1. Consider the mixed-integer recourse function Q and its pseudo-valid cutting plane approximation $\hat{Q}$. Then, under Assumption 1, there exists a constant $C \in \mathbb{R}$ with $C > 0$ such that for all ω with pdf $f \in \mathcal{H}^m$ and $z \in \mathbb{R}^m$,

$$Q(z) \geq \hat{Q}(z) - C \sum_{i=1}^m \mathbb{E}_{\omega_{-i}}[|\Delta|f_i(\cdot|\omega_{-i})].$$

Because $\hat{Q}$ is typically strictly larger than the LP-relaxation Q_{LP} of Q , it follows that Corollary 1 defines a better lower bound on Q than Q_{LP} , at least if the second right-hand-side term vanishes. Such a better lower bound may be useful, for example, in a branch-and-bound algorithm, if some of the first-stage decision variables are integer or binary.

Remark 3. Theorem 2 and Corollary 1 provide a performance guarantee for pseudo-valid cutting plane approximations of MISPs with continuous distributions only. However, for practical computations these continuous distributions are typically discretized, for example, using a sample average approximation (SAA) (Kleywegt et al. 2002), to deal with the high-dimensional integrals in the expected value function. Because $\hat{Q}$ is essentially a continuous stochastic program, the additional error that we incur by discretizing continuous distributions is thus the same as for standard continuous stochastic programs with continuous distributions. Moreover, if the discretization is fine enough, then the discretized version of $\hat{Q}$ will be sufficiently close to the original pseudo-valid cutting plane approximation $\hat{Q}$ with continuous distributions (Han and Chen 2015).

The proofs of Theorem 2 and Corollary 1 are postponed to Section 4.4. First, however, we discuss preliminary results required for these proofs. In particular, in Section 4.1, we review properties of the mixed-integer value function $v(\omega, z)$, in Section 4.2, we show that a tight pseudo-valid cutting plane approximation $\hat{v}(\omega, z)$ is affine in z on parts of its domain, and in Section 4.3, we derive bounds on $\hat{v}$. The proofs of our auxiliary lemmas and propositions in these sections are postponed to the appendix.

4.1. Properties of Mixed-Integer Value Functions

Let B be a dual feasible basis matrix of the LP-relaxation v_{LP} of v . Then we can rewrite v_{LP} as

$$\begin{aligned} v_{LP}(\omega, z) &= \min_{y_B, y_N} q_B^\top y_B + q_N^\top y_N \\ \text{s.t.} \quad & B y_B + N y_N = \omega - z \\ & y_B \in \mathbb{R}_+^m, \quad y_N \in \mathbb{R}_+^{n_2-m}, \end{aligned} \quad (9)$$

where y_B denote the basic variables and y_N the non-basic variables. Using the equality in (9) to solve for the basic variables y_B , we obtain the equivalent representation

$$\begin{aligned} v_{LP}(\omega, z) &= q_B^\top B^{-1}(\omega - z) + \min_{y_N} \bar{q}_N^\top y_N \\ \text{s.t.} \quad & B^{-1}(\omega - z) - B^{-1}N y_N \geq 0 \\ & y_N \in \mathbb{R}_+^{n_2-m}, \end{aligned} \quad (10)$$

with reduced costs $\bar{q}_N^\top := q_N^\top - q_B^\top B^{-1}N \geq 0$. Obviously, it is optimal to select the nonbasic variables y_N equal to zero in the minimization problem in (10) if $B^{-1}(\omega - z) \geq 0$. The latter condition can be conveniently rewritten as $\omega - z \in \Lambda$, where the simplicial cone Λ is defined as $\Lambda := \{t \in \mathbb{R}^m : B^{-1}t \geq 0\}$. Thus, if $\omega - z \in \Lambda$, then

$$v_{LP}(\omega, z) = q_B^\top B^{-1}(\omega - z).$$

This result holds for every dual feasible basis matrix B . In fact, the basis decomposition theorem of Walkup and Wets (1969) shows that there exist basis matrices B_k and corresponding simplicial cones $\Lambda^k := \{t \in \mathbb{R}^m : B_k^{-1}t \geq 0\}$, $k = 1, \dots, K$, such that these cones Λ^k cover $\mathbb{R}^m$, the interiors of these cones Λ^k are mutually disjoint, and $v_{LP}(\omega, z) = q_{B_k}^\top B_k^{-1}(\omega - z)$ for $\omega - z \in \Lambda^k$ for every $k = 1, \dots, K$.

Romeijnders et al. (2016a) prove a similar result for the mixed-integer value function v , involving the same basis matrices B_k and simplicial cones Λ^k , $k = 1, \dots, K$. They show that there exist distances $d_k \geq 0$ such that if $\omega - z \in \Lambda^k$ and $\omega - z$ has at least Euclidean distance d_k to the boundary of Λ^k , then

$$v(\omega, z) = q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - z),$$

where ψ^k is a B_k -periodic function (see Definition 8). The first term is the same as the LP-relaxation v_{LP} , and thus the second term can be interpreted as the additional costs of having integer variables instead of continuous ones. Theorem 3 summarizes these results.

Definition 8. Let $B \in \mathbb{Z}^{m \times m}$ be an integer matrix. Then a function $\psi : \mathbb{R}^m \mapsto \mathbb{R}$ is called B -periodic if and only if $\psi(z) = \psi(z + Bl)$ for every $z \in \mathbb{R}^m$ and $l \in \mathbb{Z}^m$.

Definition 9. Let $\Lambda \subset \mathbb{R}^m$ be a closed convex cone and let $d \in \mathbb{R}$ with $d > 0$ be given. Then we define $\Lambda(d)$ as

$$\Lambda(d) := \{s \in \Lambda : \mathcal{B}(s, d) \subset \Lambda\},$$

where $\mathcal{B}(s, d) := \{t \in \mathbb{R}^m : \|t - s\| \leq d\}$ is the closed ball centered on s with radius d . We can interpret $\Lambda(d)$ as the set of points in Λ with at least Euclidean distance d to the boundary of Λ .

Theorem 3. Consider the mixed-integer value function

$$v(\omega, z) = \min\{q^\top y : Wy = \omega - z, y \in \mathbb{Z}_+^{p_2} \times \mathbb{R}_+^{n_2 - p_2}\}, z \in \mathbb{R}^m,$$

where W is an integer matrix, and $v(\omega, z)$ is finite for all $\omega \in \Omega$ and $z \in \mathbb{R}^m$ by Assumption 1. Then there exist dual feasible basis matrices B_k of v_{LP} , $k = 1, \dots, K$, simplicial cones $\Lambda^k := \{t \in \mathbb{R}^m : B_k^{-1}t \geq 0\}$, distances $d_k \geq 0$, and bounded B_k -periodic functions ψ^k such that

- $\bigcup_{k=1}^K \Lambda^k = \mathbb{R}^m$,

- $(\text{int } \Lambda^k) \cap (\text{int } \Lambda^l) = \emptyset$ for every $k, l \in \{1, \dots, K\}$ with $k \neq l$, and
- $v(\omega, z) = q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - z)$ for every $\omega - z \in \Lambda^k(d_k)$.

Proof. See Romeijnders et al. (2016a). $\square$

4.2. Linearity Regions of Pseudo-Valid Cutting Plane Approximations

Consider a tight pseudo-valid cutting plane approximation $\hat{v}(\omega, z)$. Moreover, let $k = 1, \dots, K$ be given and consider a fixed $\omega \in \Omega$. Theorem 3 shows that for all $z \in \mathbb{R}^m$ with $\omega - z \in \Lambda^k(d_k)$, that is, for all $z \in \omega - \Lambda^k(d_k)$, the mixed-integer value function v is given by

$$v(\omega, z) = q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - z).$$

Because ψ^k is B_k -periodic, there exist values of β for which $\psi^k(\omega - z) = \psi^k(\omega - \alpha)$ for all $z \in \alpha + \beta\mathbb{Z}^m$ (see the proof of Proposition 1). Let β be such a value. Then, for all $z \in \omega - \Lambda^k(d_k)$ and $z \in \alpha + \beta\mathbb{Z}^m$, we have $\hat{v}(\omega, z) = v(\omega, z)$, and, thus, the tight pseudo-valid cutting plane approximation $\hat{v}$ equals

$$\hat{v}(\omega, z) = q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - \alpha). \quad (11)$$

Thus, for a fixed $\omega \in \Omega$, the tight pseudo-valid cutting plane approximation $\hat{v}(\omega, z)$ is affine in z over a grid of points in $\omega - \Lambda^k(d_k)$. Because $\hat{v}(\omega, z)$ is convex in z , we intuitively expect $\hat{v}(\omega, z)$ to satisfy (11) for points outside the grid in $\omega - \Lambda^k(d_k)$ as well. Lemma 1 confirms our intuition.

Lemma 1. Let $v : \mathbb{R}^m \mapsto \mathbb{R}$ be a convex function and let $C \subset \mathbb{R}^m$ be a compact convex set with extreme points $z^j \in C$, $j = 1, \dots, J$, and interior point $z^0 \in C$. Suppose that there exist $a \in \mathbb{R}^m$ and $b \in \mathbb{R}$ such that $v(z^j) = a^\top z^j + b$ for all $j = 0, \dots, J$. Then $v(z) = a^\top z + b$ for all $z \in C$.

To apply Lemma 1 to $\hat{v}(\omega, z)$, we introduce hyperrectangles $C^l(\alpha, \beta)$ that have extreme points on the grid $\alpha + \beta\mathbb{Z}^m$.

Definition 10. Let $\alpha \in \mathbb{R}^m$ and $\beta \in \mathbb{R}^m$ be given. For every $l \in \mathbb{Z}^m$, we define the hyperrectangle $C^l(\alpha, \beta)$ as

$$C^l(\alpha, \beta) := \prod_{i=1}^m [\alpha_i + \beta_i(l_i - 1), \alpha_i + \beta_i(l_i + 1)].$$

For every value of $\alpha, \beta \in \mathbb{R}^m$ and $l \in \mathbb{Z}^m$, the hyperrectangle $C^l(\alpha, \beta) \subset \mathbb{R}^m$ is convex. Moreover, all its extreme points and the interior point $(\alpha_1 + \beta_1 l_1, \dots, \alpha_m + \beta_m l_m)$ are on the grid $\alpha + \beta\mathbb{Z}^m$. Thus, if $C^l(\alpha, \beta) \subset \omega - \Lambda^k(d_k)$, then we can apply Lemma 1 to $\hat{v}(\omega, \cdot)$ with $C := C^l(\alpha, \beta)$ to conclude that $\hat{v}(\omega, z)$ satisfies (11) for all $z \in C^l(\alpha, \beta)$, and thus $\hat{v}(\omega, z)$ is affine in z over $C^l(\alpha, \beta)$. Applying Lemma 1 for all $C^l(\alpha, \beta)$ that are completely contained in $\omega - \Lambda^k(d_k)$, we can show that $\hat{v}(\omega, z)$ is

affine in z over at least $\omega - \Lambda^k(d_k + 2\|\beta\|)$. This is true, because the diameter of $C^l(\alpha, \beta)$ is $2\|\beta\|$, and $\Lambda^k(d_k + 2\|\beta\|)$ represents all points in Λ^k with at least Euclidean distance $d_k + 2\|\beta\|$ to the boundary of Λ^k . Thus, for every $z \in \omega - \Lambda^k(d_k + 2\|\beta\|)$ there exists a hyperrectangle $C^l(\alpha, \beta) \subset \omega - \Lambda^k(d_k)$ that contains z . Here, the diameter of $C^l(\alpha, \beta)$ is defined as

$$\max_{z_1, z_2} \{\|z_1 - z_2\| : z_1, z_2 \in C^l(\alpha, \beta)\} = 2\|\beta\|.$$

Proposition 1 shows all linearity regions of $\hat{v}(\omega, z)$ for fixed $\omega \in \Omega$. These are subsets of the domain of $\hat{v}(\omega, \cdot)$ on which $\hat{v}(\omega, z)$ is affine in z .

Proposition 1. Consider a tight pseudo-valid cutting plane approximation $\hat{v}(\omega, z)$ as defined in Definition 2, and let Λ^k , $k = 1, \dots, K$, denote the simplicial cones from Theorem 3. Then, under Assumption 1, for every $k = 1, \dots, K$, there exists a distance $d'_k \geq 0$ such that if $\omega - z \in \Lambda^k(d'_k)$, then

$$\hat{v}(\omega, z) = q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - \alpha). \quad (12)$$

4.3. Bounds on the Value Function of a Pseudo-Valid Cutting Plane Approximation

Proposition 1 defines $\hat{v}(\omega, z)$ on the linearity regions $\Lambda^k(d'_k)$. In fact, on these linearity regions, $v(\omega, z) = q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - z)$ and $\hat{v}(\omega, z) = q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - \alpha)$, so that the difference between the two equals

$$v(\omega, z) - \hat{v}(\omega, z) = \psi^k(\omega - z) - \psi^k(\omega - \alpha), \\ z \in \omega - \Lambda^k(d'_k).$$

This difference is B_k -periodic and bounded, because ψ^k is a bounded B_k -periodic function by Theorem 3. These properties will be exploited to derive an error bound for tight pseudo-valid cutting plane approximations $\hat{Q}$ in Section 4.4.

Outside the linearity regions, that is, on $\mathcal{N} := \mathbb{R}^m \setminus \bigcup_{k=1}^K \Lambda^k(d'_k)$, we cannot prove such properties for $v(\omega, z)$ and $\hat{v}(\omega, z)$. However, we can show that the difference between the two is bounded. That is, there exists $R \in \mathbb{R}$ such that

$$\|v - \hat{v}\|_\infty := \sup_{\omega, z} |v(\omega, z) - \hat{v}(\omega, z)| \leq R.$$

To prove this result, we use that $\mathcal{N}$ can be covered by finitely many hyperslices H_j , $j \in \mathcal{J}$, see Romeijnders et al. (2016a).

Definition 11. Let $\delta > 0$ and normal vector $a \in \mathbb{R}^m \setminus \{0\}$ be given. Then the hyperslice $H(a, \delta)$ is defined as

$$H(a, \delta) := \{z \in \mathbb{R}^m : 0 \leq a^\top z \leq \delta\}.$$

However, before we derive an upper bound on $\|v - \hat{v}\|_\infty$, we first derive a lower bound and an upper

bound on the value function $\hat{v}(\omega, z)$ of the tight pseudo-valid cutting plane approximation. The lower bound directly follows from Proposition 1 and the fact that $\hat{v}(\omega, z)$ is convex in z for every fixed $\omega \in \Omega$.

Lemma 2. Consider a tight pseudo-valid cutting plane approximation $\hat{v}(\omega, z)$ as defined in Definition 2. Then, under Assumption 1, we have for every $\omega \in \Omega$ and $z \in \mathbb{R}^m$ that

$$\hat{v}(\omega, z) \geq \max_{k=1, \dots, K} \{q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - \alpha)\}.$$

The lower bound of $\hat{v}(\omega, z)$ in Lemma 2 is valid not only on the linearity regions of $\hat{v}(\omega, \cdot)$ but also on $\omega - \mathcal{N}$. We will show that the difference between $\hat{v}(\omega, z)$ and this lower bound is bounded. Again, we use the fact that $\hat{v}(\omega, z)$ is convex in z for every fixed $\omega \in \Omega$.

Lemma 3. Consider a tight pseudo-valid cutting plane approximation $\hat{v}(\omega, z)$ as defined in Definition 2. Then, under Assumption 1, there exists $R' \in \mathbb{R}$ such that

$$\hat{v}(\omega, z) - \max_{k=1, \dots, K} \{q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - \alpha)\} \leq R'. \quad (13)$$

Now we are ready to prove an upper bound on $\|v - \hat{v}\|_\infty$. The idea of the proof is that we can use Lemmas 2 and 3 to bound $\|\hat{v} - v_{LP}\|_\infty$, where the LP-relaxation $v_{LP}(\omega, z)$ of $v(\omega, z)$ is equal to

$$v_{LP}(\omega, z) = \max_{k=1, \dots, K} \{q_{B_k}^\top B_k^{-1}(\omega - z)\}, \quad (14)$$

and the maximum difference between v and v_{LP} is known.

Proposition 2. Consider a tight pseudo-valid cutting plane approximation $\hat{v}(\omega, z)$ as defined in Definition 2. Then, under Assumption 1, there exists $R \in \mathbb{R}$ such that

$$\|v - \hat{v}\|_\infty \leq R.$$

4.4. Proof of Error Bound

In this section, we give the proofs of Theorem 2 and Corollary 1. Whereas the focus in Sections 4.2 and 4.3 was on $\hat{v}(\omega, z)$ as a function of z for fixed $\omega \in \Omega$, we now consider the difference $v(\omega, z) - \hat{v}(\omega, z)$ as a function of ω for fixed $z \in \mathbb{R}^m$. This is because $Q(z) - \hat{Q}(z) = \mathbb{E}_\omega[v(\omega, z) - \hat{v}(\omega, z)]$, $z \in \mathbb{R}^m$, and thus $v(\omega, z) - \hat{v}(\omega, z)$ can be interpreted as the underlying difference function for fixed $z \in \mathbb{R}^m$. Based on Propositions 1 and 2, we know for tight pseudo-valid cutting plane approximations $\hat{v}(\omega, z)$ that for $\omega \in z + \Lambda^k(d'_k)$, $k = 1, \dots, K$,

$$v(\omega, z) - \hat{v}(\omega, z) = \psi^k(\omega - z) - \psi^k(\omega - \alpha),$$

and for $\omega \in z + \mathcal{N}$,

$$|v(\omega, z) - \hat{v}(\omega, z)| \leq R.$$

We will use these two main properties to derive an upper bound for $\|Q - \hat{Q}\|_\infty$ that depends on the total variations of the one-dimensional conditional probability density functions of the random variables in the model.

Proof of Theorem 2. Combining Theorem 3 and Proposition 1, there exist basis matrices B_k , corresponding simplicial cones Λ^k , distances $d'_k \geq 0$, and bounded B_k -periodic functions ψ^k such that for $\omega - z \in \Lambda^k(d'_k)$,

$$v(\omega, z) - \hat{v}(\omega, z) = \psi^k(\omega - z) - \psi^k(\omega - \alpha).$$

Moreover, by Proposition 2, there exists $R \in \mathbb{R}$ such that $\|v - \hat{v}\|_\infty \leq R$.

Fix $z \in \mathbb{R}^m$ and consider the difference $v(\omega, z) - \hat{v}(\omega, z)$ as a function of ω . We will use that for $\omega \in z + \Lambda^k(d'_k)$, this difference is B_k -periodic, and for $\omega \in z + \mathcal{N}$, it is bounded by R . In fact, using theorems 4.6 and 4.13 in Romeijnders et al. (2016a), we can show that there exist constants $D > 0$ and $C'_k > 0$, $k = 1, \dots, K$, such that

$$\mathbb{P}\{\omega \in z + \mathcal{N}\} \leq D \sum_{i=1}^m \mathbb{E}_{\omega_{-i}}[|\Delta|f_i(\cdot|\omega_{-i})|], \quad (15)$$

and for every $k = 1, \dots, K$,

$$\begin{aligned} & \left| \int_{z + \Lambda^k(d'_k)} (\psi^k(t - z) - \psi^k(t - \alpha))f(t)dt \right| \\ & \leq C'_k \sum_{i=1}^m \mathbb{E}_{\omega_{-i}}[|\Delta|f_i(\cdot|\omega_{-i})|]. \end{aligned} \quad (16)$$

Then

$$\begin{aligned} |Q(z) - \hat{Q}(z)| &= \left| \int_{\mathbb{R}^m} (v(t, z) - \hat{v}(t, z))f(t)dt \right| \\ &\leq \left| \int_{z + \mathcal{N}} (v(t, z) - \hat{v}(t, z))f(t)dt \right| \\ &\quad + \sum_{k=1}^K \left| \int_{z + \Lambda^k(d'_k)} (v(t, z) - \hat{v}(t, z))f(t)dt \right| \\ &\leq R\mathbb{P}\{\omega \in z + \mathcal{N}\} \\ &\quad + \sum_{k=1}^K \left| \int_{z + \Lambda^k(d'_k)} (v(t, z) - \hat{v}(t, z))f(t)dt \right|. \end{aligned}$$

Applying the bound in (15) to the first term and the bounds in (16) to the second term, we obtain

$$\begin{aligned} |Q(z) - \hat{Q}(z)| &\leq RD \sum_{i=1}^m \mathbb{E}_{\omega_{-i}}[|\Delta|f_i(\cdot|\omega_{-i})|] \\ &\quad + \sum_{k=1}^K C'_k \sum_{i=1}^m \mathbb{E}_{\omega_{-i}}[|\Delta|f_i(\cdot|\omega_{-i})|] \\ &= C \sum_{i=1}^m \mathbb{E}_{\omega_{-i}}[|\Delta|f_i(\cdot|\omega_{-i})|], \end{aligned}$$

where the constant C is defined as $C := RD + \sum_{k=1}^K C'_k$. $\square$

Proof of Corollary 1. Because (4) holds for the pseudo-valid cutting plane approximation $\hat{v}(\omega, z)$, it immediately follows that there exist $\alpha \in \mathbb{R}^m$ and $\beta \in \mathbb{Z}^m$ such that for all $z \in \alpha + \beta\mathbb{Z}^m$ and $\omega \in \Omega$,

$$\hat{v}(\omega, z) \leq v(\omega, z).$$

This implies, for example, that (12) holds with “ $\leq$ ” instead of “ $=$.” Thus, analogously to the proof of Theorem 2, we are able to show that there exists a constant $C \in \mathbb{R}$ with $C > 0$ such that for all ω with pdf $f \in \mathcal{H}^m$, and for all $z \in \mathbb{R}^m$,

$$\hat{Q}(z) - Q(z) \leq C \sum_{i=1}^m \mathbb{E}_{\omega_{-i}}[|\Delta|f_i(\cdot|\omega_{-i})|]. \quad (17)$$

Notice that this upper bound does not necessarily apply to $Q(z) - \hat{Q}(z)$, because the pseudo-valid cuts are not necessarily tight. The claim follows by rearranging terms in (17). $\square$

5. Examples of Pseudo-Valid Cutting Planes

In this section, we consider examples of pseudo-valid cutting plane approximations. We derive pseudo-valid mixed-integer Gomory cuts in Section 5.1 and a tight pseudo-valid cutting plane approximation for a nurse scheduling problem in Section 5.2.

5.1. Pseudo-Valid Mixed-Integer Gomory Cuts

Consider the second-stage value function

$$\begin{aligned} v(\omega, z) &:= \min_{y_B, y_N} \{q_B^\top y_B + q_N^\top y_N : By_B + Ny_N = \omega - z, \\ &\quad y_B \in Y_B, y_N \in Y_N\}, \quad \omega \in \Omega, z \in \mathbb{R}^m, \end{aligned}$$

where similar to Section 4.1, we let B denote a dual feasible basis matrix of the LP-relaxation of v . Multiplying the equality constraint in $v(\omega, z)$ by $e_i^\top B^{-1}$, where e_i is the i th unit vector, we obtain

$$y_{B_i} + e_i^\top B^{-1}Ny_N = e_i^\top B^{-1}(\omega - z), \quad (18)$$

where y_{B_i} denotes the i th basic variable. Let $\bar{w}_{ij}$ denote the j th component of the vector $e_i^\top B^{-1}N$, let y_{N_j} denote the j th nonbasic variable, and let $r_i(\omega, z) := e_i^\top B^{-1}(\omega - z) - \lfloor e_i^\top B^{-1}(\omega - z) \rfloor$. If the i th basic variable y_{B_i} is restricted to be integer, then we can derive from (18) the exact mixed-integer Gomory cut

$$\begin{aligned} & \sum_{j \in J_1} \min \left\{ \frac{\bar{w}_{ij} - \lfloor \bar{w}_{ij} \rfloor}{r_i(\omega, z)}, \frac{1 - \bar{w}_{ij} + \lfloor \bar{w}_{ij} \rfloor}{1 - r_i(\omega, z)} \right\} y_{N_j} \\ & + \sum_{j \in J_2} \max \left\{ \frac{\bar{w}_{ij}}{r_i(\omega, z)}, \frac{-\bar{w}_{ij}}{1 - r_i(\omega, z)} \right\} y_{N_j} \geq 1, \end{aligned} \quad (19)$$

where J_1 denotes the index set of integer nonbasic variables y_{N_1} and J_2 the index set of continuous nonbasic variables y_{N_2} (see, e.g., Balas and Jeroslow 1980).

Obviously, the exact mixed-integer Gomory cut in (19) is not affine in z , among others, because $r_i(\omega, z)$ is not affine in z . However, if $z \in \beta\mathbb{Z}^m$, where $\beta := |\det(B)|e$ with e the m -dimensional all-one vector, then under the assumption that W is integer, we can show that $r_i(\omega, z) = r_i(\omega, 0)$, and thus the mixed-integer Gomory cut in (19) does not depend on z . This is true, because for such z , we have

$$e_i^\top B^{-1}z = e_i^\top (\det(B)^{-1} \text{adj}(B))z \in \mathbb{Z},$$

and thus

$$\begin{aligned} r_i(\omega, z) &= e_i^\top B^{-1}(\omega - z) - \lfloor e_i^\top B^{-1}(\omega - z) \rfloor \\ &= e_i^\top B^{-1}\omega - \lfloor e_i^\top B^{-1}\omega \rfloor = r_i(\omega, 0). \end{aligned}$$

Similarly, if $z \in \alpha + \beta\mathbb{Z}^m$ with $\beta := |\det(B)|e$, then $r_i(\omega, z) = r_i(\omega, \alpha)$. Thus, replacing $r_i(\omega, z)$ by $r_i(\omega, \alpha)$ in (19) yields an approximate mixed-integer Gomory cut that does not depend on z and is valid for all ω and for all z on a grid of points $\alpha + \beta\mathbb{Z}^m$. Hence, the approximate mixed-integer Gomory cut

$$\begin{aligned} \sum_{j \in J_1} \min \left\{ \frac{\bar{w}_{ij} - \lfloor \bar{w}_{ij} \rfloor}{r_i(\omega, \alpha)}, \frac{1 - \bar{w}_{ij} + \lfloor \bar{w}_{ij} \rfloor}{1 - r_i(\omega, \alpha)} \right\} y_{N_1} \\ + \sum_{j \in J_2} \max \left\{ \frac{\bar{w}_{ij}}{r_i(\omega, \alpha)}, \frac{-\bar{w}_{ij}}{1 - r_i(\omega, \alpha)} \right\} y_{N_2} \geq 1, \end{aligned} \quad (20)$$

is pseudo-valid and thus satisfies the assumptions of Corollary 1. This implies that we can use these pseudo-valid mixed-integer Gomory cuts to obtain a strictly better (approximate) lower bound of Q than Q_{LP} , as we illustrate in Example 3.

Example 3. Consider an adjusted version of the MISP in Schultz et al. (1998), given by

$$\min_{x \in X} -\frac{3}{2}x_1 - 4x_2 + Q(x),$$

where $X = [0, 5]^2$ and $Q(x) = \mathbb{E}_\omega[v(\omega, x)]$ with

$$\begin{aligned} v(\omega, x) &= \min_{y, u, v} -16y_1 - 19y_2 - 23y_3 - 28y_4 + 50v_1 + 50v_2 \\ \text{s.t.} \quad &2y_1 + 3y_2 + 4y_3 + 5y_4 + u_1 - v_1 = \omega_1 - x_1 \\ &6y_1 + y_2 + 3y_3 + 2y_4 + u_2 - v_2 = \omega_2 - x_2 \\ &y \in \mathbb{Z}_+^4, \quad u, v \in \mathbb{R}_+^2. \end{aligned}$$

Similar to Schultz et al. (1998), the random vector ω follows a discrete uniform distribution on $\{5, 5.5, \dots, 15\} \times \{5, 5.5, \dots, 15\}$, and thus the support Ω of ω consists of 441 scenarios. Different variants of this

problem are considered in, for example, Carøe and Tind (1997), Sherali and Zhu (2006), Gade et al. (2014), and Qi and Sen (2017).

We use (20) to derive pseudo-valid mixed-integer Gomory cuts for Q . To be precise, we derive pseudo-valid mixed-integer Gomory cuts from those dual feasible basis matrices B that are optimal for the LP-relaxation v_{LP} of v for at least one pair of $\omega \in \Omega$ and $x \in X$. These basis matrices turn out to be

$$B_1 = \begin{pmatrix} 2 & 3 \\ 6 & 1 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 2 & 0 \\ 6 & 1 \end{pmatrix}, \quad \text{and } B_3 = \begin{pmatrix} 3 & 1 \\ 1 & 0 \end{pmatrix}.$$

Because both columns of B_1 and only the first column of B_2 and B_3 correspond to integer second-stage variables, we derive four pseudo-valid mixed-integer Gomory cuts: two corresponding to B_1 and one corresponding to B_2 and B_3 .

Because the first-stage decision vector x is two-dimensional, we are able to depict both Q and Q_{LP} graphically in Figure 3. Moreover, we also show four pseudo-valid cutting plane approximations $\hat{Q}$, obtained by iteratively adding an additional pseudo-valid mixed-integer Gomory cut. From Figure 3 we observe that already by adding a single pseudo-valid mixed-integer Gomory cut we significantly improve the LP-relaxation lower bound Q_{LP} of Q . Moreover, this lower bound improves further if we add additional pseudo-valid mixed-integer Gomory cuts. Notice that $\hat{Q}$ is indeed an *approximate* lower bound, because for $\hat{Q}$ with three or four cuts, the function $\hat{Q}$ may exceed the original expected value function Q . However, Corollary 1 guarantees that $\sup_{x \in X} \{\hat{Q}(x) - Q(x)\}$ cannot be too large.

Finally, in Figure 4 we show the approximation error $\|Q - \hat{Q}\|_\infty := \sup_{x \in X} |Q(x) - \hat{Q}(x)|$ for the four pseudo-valid cutting plane approximations $\hat{Q}$ with respect to the integrality gap $\|Q - Q_{LP}\|_\infty$ of Q_{LP} . We observe that this integrality gap reduces by almost 20% if we add a single pseudo-valid mixed-integer Gomory cut. Moreover, by adding all four cuts, the approximation error of $\hat{Q}$ reduces to approximately 50% of that of Q_{LP} .

5.2. Nurse Scheduling Problem

In this section, we apply tight pseudo-valid cutting planes to a nurse scheduling problem, introduced by Kim and Mehrotra (2015). In this problem, a regular work schedule for the nurses is determined in the first stage, resulting in a number z_t of available nurses per time period $t = 1, \dots, T$. This regular work schedule is determined before the random demand ω_t for nurses per time period is known. Thus, we may have a shortage or surplus of nurses in some of the time periods. In this case, it is possible to add or subtract nurse shifts, consisting of several consecutive time periods, after the

Figure 3. Expected Value Function Q of Example 3, Its LP-Relaxation Q_{LP} , and Four Pseudo-Valid Cutting Plane Approximations $\hat{Q}$ with a Different Number of Pseudo-Valid Mixed-Integer Gomory Cuts

demands ω_t are known. Moreover, we penalize any remaining nurse shortages and nurse surpluses using unit penalty costs per time period. The corresponding second-stage value function v is given by

$$\begin{aligned} v(\omega, z) = \min_{y, u_1, u_2} \quad & q^\top y + r_1^\top u_1 + r_2^\top u_2 \\ \text{s.t.} \quad & Wy - u_1 + u_2 = \omega - z \\ & y \in \mathbb{Z}_+^{n_2}, \quad u_1, u_2 \in \mathbb{R}_+^T, \end{aligned} \quad (21)$$

where $y \in \mathbb{Z}_+^{n_2}$ represents the possibility to add or subtract nurse shifts, and W is a $\{-1, 0, 1\}$ -matrix, modelling which time periods are contained in which shift. Kim and Mehrotra (2015) show that W is a totally unimodular matrix. Moreover, they show that if $z \in \mathbb{Z}^T$, then the cutting planes $Wy - \hat{D}(\omega)u_1 \leq \lfloor \omega \rfloor - z$, with $\hat{D}(\omega)$ the diagonal matrix with t th diagonal component $\hat{D}_{tt}(\omega)$ equal to

$$\hat{D}_{tt}(\omega) = \frac{1}{1 - \omega_t + \lfloor \omega_t \rfloor}, \quad t = 1, \dots, T,$$

are valid for all $\omega \in \Omega$. In particular, combined with the constraints $Wy - u_1 + u_2 = \omega - z$, they completely

define the integer hull $\tilde{Y}(\omega, z)$ of the feasible region $Y(\omega, z)$ of $v(\omega, z)$. That is, for every $\omega \in \Omega$ and $z \in \mathbb{Z}^T$,

$$\begin{aligned} \tilde{Y}(\omega, z) = \{ (y, u_1, u_2) \in \mathbb{R}_+^{n_2+2T} : & Wy - u_1 + u_2 = \omega - z, \\ & Wy - \hat{D}(\omega)u_1 \leq \lfloor \omega \rfloor - z \}. \end{aligned}$$

If we assume, contrary to Kim and Mehrotra (2015), that z is not necessarily integral, then we may use the cutting planes $Wy - \hat{D}(\omega)u_1 \leq \lfloor \omega \rfloor - z$ to derive the tight pseudo-valid cutting plane approximation

$$\begin{aligned} \hat{v}(\omega, z) = \min_{y, u_1, u_2} \quad & q^\top y + r_1^\top u_1 + r_2^\top u_2 \\ \text{s.t.} \quad & Wy - u_1 + u_2 = \omega - z \\ & Wy - \hat{D}(\omega)u_1 \leq \lfloor \omega \rfloor - z \\ & y \in \mathbb{R}_+^{n_2}, \quad u_1, u_2 \in \mathbb{R}_+^T. \end{aligned} \quad (22)$$

Because the cutting planes are valid for all $z \in \alpha + \beta\mathbb{Z}^m$, with $\alpha = 0$ and $\beta = e$, and, for all $\omega \in \Omega$, they are indeed pseudo-valid and tight. Hence, by Theorem 2, the error of the corresponding tight pseudo-valid cutting plane approximation $\hat{Q}$ converges to zero if all total variations of one-dimensional conditional pdf of ω

Figure 4. Relative Approximation Errors of the Pseudo-Valid Cutting Plane Approximations $\hat{Q}$ of Example 3

converge to zero. In Example 4 below, we numerically show the actual performance of this approximation for the one-dimensional second-stage value function of Example 1 in Section 2, which can be considered a special case of (21).

Example 4. Consider the second-stage value function $v(\omega, z)$ of Example 1,

$$\begin{aligned} v(\omega, z) &= r(\omega - z) + \min_{y, u_1} (q - r)y + 2ru_1 \\ \text{s.t. } & y - u_1 \leq \omega - z \\ & y \in \mathbb{Z}_+, u_1 \in \mathbb{R}_+, \end{aligned}$$

and its tight pseudo-valid cutting plane approximation defined in (8). Let ω be a normal random variable with mean μ and standard deviation σ . Then, as shown in Example 2, the total variation $|\Delta|f$ of the probability density function f of ω equals $|\Delta|f = \sigma^{-1}\sqrt{2/\pi}$. Figure 5 shows $\|Q - \hat{Q}\|_\infty$, the maximum difference between the expected value function Q and its tight pseudo-valid cutting plane approximation $\hat{Q}$, as a function of the standard deviation σ for $q = 1$ and $r = 2$. We observe that this difference decreases if σ increases. This is in line with Theorem 2, because the total variation $|\Delta|f$ of a normal probability density function f decreases if the standard deviation σ increases.

6. Numerical Case Study

In this section, we evaluate the performance of the tight pseudo-valid cutting planes derived for the nurse scheduling problem in Section 5.2. This nurse scheduling problem is adapted from Kim and Mehrotra (2015) so that all first-stage decision variables are continuous instead of integer. For small problems, we compute the optimality gap of our pseudo-valid cutting plane approximation, and, for larger problems, we compare our approximation with

Figure 5. Maximum Difference between Q and Its Pseudo-Valid Cutting Plane Approximation $\hat{Q}$ of Example 4, with $q = 1$ and $r = 2$, as a Function of the Standard Deviation σ of a Normal Random Variable ω

two benchmark approximations. All computations are carried out on a single Intel Xeon E5 2680v3 core (2.5 GHz).

6.1. Experimental Design

Like in Kim and Mehrotra (2015), the regular nurse schedule that we have to determine is a weekly repeating nurse scheduling for a 12-week period in which nurses work in shifts of 8 or 12 hours. Moreover, there are three possible employment types for nurses: full-time, part-time, and casual. See Kim and Mehrotra (2015) for the details.

The normalized costs for regular nurses are 1 per hour. In the second stage, nurse shifts can be cancelled for free, but they can be only added at a cost of 1.5 per hour. Moreover, the unit penalty costs r_1 and r_2 from (21) are 0 and 10^4 for each time period, respectively. In this way, demand for nurses should be met either by the regular workforce or by adding nurse shifts in the second stage to avoid high penalty costs.

The demand for nurses ω_t during time period t is determined by the patient volume ξ_t and the nurse-to-patient ratio. Like Kim and Mehrotra (2015), we use a nurse-to-patient ratio of 1:4 from 8 a.m. to 4 p.m. and 1:5 for the remainder of the day. For the distribution of the patient volume ξ_t , we differ from Guan et al. (2009), and consider both normal and uniform distributions with mean $\mu_t := \mathbb{E}[\xi_t] \in \{25, 75, 125\}$ and standard deviation $\sigma_t \in \{0.5, 1, 2, 5, 10, 19\}$. We assume that all ξ_t are identically and independently distributed. We include $\mu_t = 125$ and $\sigma_t = 19$, because Kim and Mehrotra (2015) report these statistics for actual patient volume data. Moreover, we note that in our experiments we truncate the patient volume if necessary, so that it is always nonnegative.

We test the performance of our tight pseudo-valid cutting plane approximation on both small and large problems of this nurse scheduling problem. In the small problems, a 3-day regular nurse schedule has to be determined, taking into account 100 scenarios for future patient volume, sampled from the abovementioned distributions. The large problems correspond to the

12-week nurse scheduling problem with 1,000 scenarios. These problems have the same size as those in Kim and Mehrotra (2015).

6.2. Solution Methods

For both small and large problems, we solve the pseudo-valid cutting plane approximation from (22). We do so using a standard L-shaped algorithm (van Slyke and Wets 1969), yielding the approximating solution $\hat{x}_0$. The solution is called $\hat{x}_0$, because the pseudo-valid cutting planes in (22) can be considered a special case of the family of pseudo-valid cutting planes

$$Wy - \hat{D}(\omega - \alpha)u_1 \leq \lfloor \omega - \alpha \rfloor + \alpha - z,$$

with $\alpha = 0$. We, however, also compute the pseudo-valid cutting plane solutions $\hat{x}_\alpha$ for 100 different values of α , and let $\hat{x}^*$ denote the best solution among those. The different values for α that we use are $\alpha = 0.01ke, k = 1, \dots, 100$, where e is the all-one vector.

For the small nurse scheduling problems, we solve their large-scale deterministic equivalent MIP, and we let x^* denote the best-found solution within 12 hours of computation time. We also obtain the MIP optimality gap corresponding to x^* . If this gap is zero, then x^* is the optimal solution. Alternatively, we could have used a state-of-the-art method, for example, Qui and Sen (2017), to obtain the exact optimal solution. However, the MIP optimality gaps that we obtain by solving the large-scale deterministic equivalent MIP are relatively small already.

The larger problems cannot be solved to optimality within reasonable time limits. For these instances we compute two benchmark approximations. The first is the LP-relaxation in which the integrality conditions on the second-stage variables y are relaxed, and the second is the expected-value model in which the random patient volumes are replaced by their mean values. We denote the resultant solutions by x_{LP} and x_μ , respectively. The solution of the LP-relaxation is obtained using a similar L-shaped algorithm as for the pseudo-valid cutting plane approximation. The expected-value model is a relatively small MIP that easily can be solved to optimality. For each problem instance, both small and large, we dedicate at most 12 hours of computation time to obtaining each candidate solution.

6.3. Performance Measures

For the small problems, we report in-sample costs, out-of-sample costs, and running times of $\hat{x}_0$ and $\hat{x}^*$. The in-sample costs are based on the 100 generated scenarios for that problem instance, whereas the out-of-sample costs are based on a new sample of 10^6 scenarios. We also report the relative difference (in %)

with the in-sample and out-of-sample costs of x^* . The in-sample difference represents the optimality gap in the 100-scenario problem. However, not all problems could be solved to optimality, so we use the MIP optimality gap of x^* to compute lower and upper bounds on the optimality gaps of $\hat{x}_0$ and $\hat{x}^*$. For the larger problems, we do not report these optimality gaps, but only the normalized in-sample costs of the pseudo-valid cutting plane solutions and the benchmark solutions.

The running times for $\hat{x}^*$ for the small problems represent the time required to solve the 100 pseudo-valid cutting plane approximations in series. For the large problems, however, we report the maximum computing time over all 100 approximations. This is the time required to obtain $\hat{x}^*$ if all pseudo-valid cutting plane approximations are solved in parallel. We actually used parallel computing over 24 cores to obtain $\hat{x}^*$.

6.4. Numerical Results

Table 1 shows the results for the small problems. We observe that the overall performance of the pseudo-valid cutting plane solutions $\hat{x}_0$ and $\hat{x}^*$ is very good, both for the normal and uniform distribution. In fact, in most of the problems the in-sample optimality gap for $\hat{x}^*$ is below 1%. The optimality gaps are smaller for larger values of the standard deviation σ . This is in line with Theorem 2, because the total variation of both the normal and the uniform probability density function is smaller if σ is larger.

For small values of the standard deviation, that is, $\sigma \leq 2$, and for $\mu = 25$, we observe that the optimality gap of $\hat{x}_0$ can be quite high. In these cases, we do not recommend using this pseudo-valid cutting plane approximation. However, the optimality gap can be significantly reduced, in particular, when σ is small, by considering multiple pseudo-valid cutting plane approximations with different values for α , and to select the best among them. Indeed, the in-sample optimality gap of $\hat{x}^*$ never exceeds 4%.

Compared with the optimal in-sample solution x^* , the out-of-sample performance of the pseudo-valid cutting plane solutions $\hat{x}_0$ and $\hat{x}^*$ is better than their in-sample performance. In fact, in some cases, the out-of-sample costs of $\hat{x}^*$ are actually lower than those for x^* . In general, we conclude that for larger values of the standard deviation, that is, $\sigma \geq 5$, the out-of-sample performance of $\hat{x}^*$ is equivalent to that of x^* .

Finally, we note from Table 1 that the pseudo-valid cutting plane solutions are obtained extremely fast for the small problems. Indeed, the running times for $\hat{x}_0$ are below 10 seconds, and to obtain $\hat{x}^*$, the best of 100 different pseudo-valid cutting plane solutions, takes at most 16 minutes. In contrast, the optimal

Table 1. Numerical Results for the Small Nurse Scheduling Problems

μ	σ	Running times (seconds)		In-sample costs (percentage gap)		Out-of-sample costs (percentage gap)	
		$\hat{x}_0$	$\hat{x}^*$	$\hat{x}_0$	$\hat{x}^*$	$\hat{x}_0$	$\hat{x}^*$
Normally distributed patient volume							
25	0.5	1.9	124.3	456 (10.9%–11.0%)	425 (3.4%–3.5%)	456 (9.9%)	426 (2.7%)
25	1.0	1.6	177.4	456 (5.9%–6.4%)	443 (2.9%–3.3%)	456 (4.8%)	444 (2.0%)
25	2.0	1.8	246.2	474 (1.8%–3.2%)	473 (1.7%–3.0%)	474 (0.7%)	475 (0.9%)
25	5.0	3.3	401.9	558 (1.3%–3.6%)	555 (0.9%–3.1%)	559 (0.5%)	557 (0.1%)
25	10.0	3.9	415.3	686 (0.6%–2.8%)	683 (0.2%–2.4%)	688 (-0.1%)	688 (-0.1%)
25	19.0	4.2	434.7	919 (0.4%–2.0%)	917 (0.2%–1.8%)	924 (0.0%)	923 (-0.2%)
75	0.5	2.0	154.5	1,230 (3.2%–3.3%)	1,196 (0.4%–0.4%)	1,230 (2.9%)	1,197 (0.2%)
75	1.0	2.9	253.8	1,248 (3.1%–3.3%)	1,214 (0.3%–0.5%)	1,248 (2.7%)	1,216 (0.1%)
75	2.0	2.1	302.9	1,252 (0.5%–1.0%)	1,249 (0.3%–0.8%)	1,253 (0.2%)	1,250 (-0.1%)
75	5.0	5.6	517.8	1,337 (0.5%–1.4%)	1,334 (0.3%–1.2%)	1,340 (0.2%)	1,337 (0.0%)
75	10.0	8.0	676.8	1,466 (0.5%–1.3%)	1,464 (0.3%–1.2%)	1,471 (0.2%)	1,468 (0.0%)
75	19.0	6.9	713.2	1,698 (0.1%–1.1%)	1,697 (0.0%–1.0%)	1,705 (0.0%)	1,705 (0.0%)
125	0.5	2.0	158.2	2,016 (2.3%–2.3%)	1,985 (0.7%–0.7%)	2,016 (2.1%)	1,986 (0.6%)
125	1.0	1.7	242.2	2,016 (1.3%–1.4%)	2,003 (0.6%–0.7%)	2,016 (1.0%)	2,004 (0.4%)
125	2.0	2.6	337.3	2,034 (0.4%–0.7%)	2,033 (0.4%–0.7%)	2,034 (0.2%)	2,035 (0.2%)
125	5.0	5.4	572.7	2,118 (0.3%–0.9%)	2,115 (0.2%–0.8%)	2,119 (0.1%)	2,118 (0.1%)
125	10.0	7.8	807.3	2,246 (0.3%–0.8%)	2,243 (0.1%–0.7%)	2,248 (0.0%)	2,248 (0.0%)
125	19.0	9.6	915.0	2,478 (0.1%–0.7%)	2,477 (0.0%–0.7%)	2,486 (0.0%)	2,485 (0.0%)
Uniformly distributed patient volume							
25	0.5	1.3	112.0	456 (13.0%–13.0%)	418 (3.6%–3.6%)	456 (12.7%)	418 (3.3%)
25	1.0	1.8	173.2	456 (9.4%–9.4%)	433 (3.9%–3.9%)	456 (9.0%)	433 (3.5%)
25	2.0	1.7	196.4	472 (6.5%–6.5%)	458 (3.3%–3.3%)	471 (5.7%)	458 (2.8%)
25	5.0	4.2	335.9	547 (4.4%–4.8%)	533 (1.6%–2.0%)	548 (3.9%)	533 (1.1%)
25	10.0	4.8	410.0	663 (1.9%–3.4%)	652 (0.3%–1.7%)	664 (1.1%)	656 (-0.1%)
25	19.0	6.7	513.4	868 (0.4%–2.1%)	866 (0.1%–1.8%)	875 (0.2%)	871 (-0.3%)
75	0.5	1.6	128.6	1,224 (3.4%–3.4%)	1,189 (0.5%–0.5%)	1,224 (3.3%)	1,189 (0.4%)
75	1.0	2.2	187.6	1,248 (4.3%–4.3%)	1,201 (0.4%–0.4%)	1,248 (4.1%)	1,201 (0.2%)
75	2.0	1.8	264.6	1,248 (2.0%–2.0%)	1,226 (0.2%–0.2%)	1,248 (1.8%)	1,226 (0.0%)
75	5.0	3.3	385.6	1,320 (1.2%–1.3%)	1,311 (0.5%–0.6%)	1,320 (1.0%)	1,311 (0.3%)
75	10.0	6.2	602.6	1,441 (0.8%–1.4%)	1,436 (0.4%–1.0%)	1,444 (0.5%)	1,438 (0.1%)
75	19.0	7.7	720.4	1,649 (0.4%–1.1%)	1,647 (0.2%–1.0%)	1,654 (0.1%)	1,653 (0.0%)
125	0.5	1.8	127.7	2,016 (2.7%–2.7%)	1,978 (0.7%–0.7%)	2,016 (2.6%)	1,978 (0.7%)
125	1.0	1.6	196.6	2,016 (2.0%–2.0%)	1,993 (0.8%–0.8%)	2,016 (1.9%)	1,993 (0.7%)
125	2.0	2.4	269.9	2,032 (1.4%–1.4%)	2,018 (0.7%–0.7%)	2,031 (1.3%)	2,018 (0.6%)
125	5.0	4.8	482.8	2,107 (1.1%–1.2%)	2,093 (0.4%–0.5%)	2,108 (1.0%)	2,093 (0.3%)
125	10.0	6.8	614.4	2,222 (0.6%–1.0%)	2,212 (0.1%–0.5%)	2,225 (0.4%)	2,216 (0.0%)
125	19.0	6.9	774.4	2,429 (0.3%–0.7%)	2,427 (0.2%–0.7%)	2,435 (0.1%)	2,432 (0.0%)

solution x^* could not be computed within 12 hours for most of the small problems.

Table 2 shows the results for the large problems. We observe that both pseudo-valid cutting plane solutions $\hat{x}_0$ and $\hat{x}^*$ consistently outperform the benchmark solutions x_{LP} and x_μ , both for the normal and uniform distribution. Similar to the small problems, $\hat{x}^*$ can be significantly better than $\hat{x}_0$, in particular for small values of the standard deviation σ . For large values of σ , we observe that the solution x_{LP} of the LP-relaxation is almost as good as the pseudo-valid cutting plane solutions $\hat{x}^*$ and $\hat{x}_0$. Because $\hat{x}^*$ and $\hat{x}_0$ are close to optimal for the small problems in these cases, we conjecture that $\hat{x}^*$, $\hat{x}_0$, and x_{LP} are all close to optimal in the corresponding large problems as well.

With respect to the running times, we observe that the pseudo-valid cutting plane approximation can be typically solved much faster than the LP-relaxation. We emphasize that both solution methods are implemented using the same standard version of the L-shaped algorithm. Further analysis of the computational results shows that computing x_{LP} requires much more master iterations than computing $\hat{x}_0$, explaining the surprising difference in running times.

7. Discussion

We consider a new solution method for solving two-stage mixed-integer stochastic programs (MISPs) with uncertainty in the right-hand side. Instead of applying exact cuts to the second-stage feasible regions

Table 2. Numerical Results for the Large Nurse Scheduling Problems

μ	σ	Running times (seconds)				In-sample costs (normalized)			
		$\hat{x}^*$	$\hat{x}_0$	x_{LP}	x_μ	$\hat{x}^*$	$\hat{x}_0$	x_{LP}	x_μ
Normally distributed patient volume									
25	0.5	4,942	1,063	11,289	85	100.0	107.2	110.8	116.0
25	1.0	6,019	1,130	11,101	82	100.0	102.9	108.0	111.4
25	2.0	7,791	1,998	12,285	87	100.0	100.0	104.1	105.7
25	5.0	8,937	5,672	13,340	89	100.0	100.3	100.8	105.4
25	10.0	10,190	7,626	12,388	84	100.0	100.0	100.3	106.7
25	19.0	12,407	12,407	13,123	87	100.0	100.1	100.1	107.6
75	0.5	4,907	1,275	10,720	85	100.0	102.7	104.5	106.4
75	1.0	8,471	2,307	11,394	85	100.0	102.7	103.6	104.8
75	2.0	7,151	1,853	12,722	78	100.0	100.3	102.0	102.5
75	5.0	9,231	7,153	12,385	90	100.0	100.3	100.4	102.4
75	10.0	14,060	9,105	13,395	83	100.0	100.1	100.1	103.1
75	19.0	22,212	17,905	13,866	85	100.0	100.1	100.1	104.4
125	0.5	5,168	1,112	11,160	88	100.0	101.5	102.3	103.4
125	1.0	7,091	1,137	12,254	84	100.0	100.6	101.8	102.5
125	2.0	7,718	2,070	13,476	92	100.0	100.0	101.0	101.3
125	5.0	7,529	6,909	12,279	85	100.0	100.1	100.2	101.4
125	10.0	10,768	7,081	13,650	88	100.0	100.0	100.1	102.0
125	19.0	13,524	11,636	13,521	85	100.0	100.0	100.1	103.0
Uniformly distributed patient volume									
25	0.5	5,346	1,210	11,183	85	100.0	109.1	113.0	118.2
25	1.0	5,507	1,017	10,797	77	100.0	105.3	111.1	114.1
25	2.0	6,827	1,722	10,854	82	100.0	102.8	108.8	107.9
25	5.0	7,719	5,891	12,549	89	100.0	102.8	103.6	109.3
25	10.0	9,591	6,075	13,206	89	100.0	101.6	101.2	110.6
25	19.0	13,995	10,704	13,572	72	100.0	100.6	100.4	111.7
75	0.5	5,223	1,093	10,500	94	100.0	102.9	105.3	107.1
75	1.0	5,617	1,837	11,594	86	100.0	103.9	105.0	106.1
75	2.0	7,215	1,278	11,497	82	100.0	101.8	104.3	103.9
75	5.0	9,031	3,717	13,427	76	100.0	100.7	101.6	103.9
75	10.0	10,096	6,199	12,944	82	100.0	100.4	100.4	104.7
75	19.0	11,780	10,728	14,084	89	100.0	100.2	100.2	106.3
125	0.5	5,520	1,027	10,944	90	100.0	101.9	102.7	103.8
125	1.0	5,636	1,072	11,824	86	100.0	101.1	102.4	103.1
125	2.0	6,980	2,030	12,666	91	100.0	100.6	102.0	101.8
125	5.0	8,013	5,897	13,650	86	100.0	100.7	100.9	102.4
125	10.0	8,627	6,290	13,286	95	100.0	100.5	100.4	103.1
125	19.0	13,637	8,684	13,872	86	100.0	100.2	100.1	104.4

that are always valid, we propose to use *pseudo-valid* cutting planes that are affine in the first-stage decision variables. The advantage is that the approximating problem, that uses these pseudo-valid cuts, is convex, and thus it is much easier to solve than the original MISP.

For simple integer recourse (SIR) models, we show that the α -approximations of Klein Haneveld et al. (2006) can be interpreted as pseudo-valid cutting plane approximations. A direct consequence of this result is that we obtain an error bound on the quality of the solution obtained using pseudo-valid cutting planes for SIR models. For general MISP we derive a similar bound for so-called tight pseudo-valid cutting plane approximations. We show that the error of using such

approximations converges to zero if the total variations of the one-dimensional conditional probability density functions in the model converge to zero.

For general MISP we also derive pseudo-valid mixed-integer Gomory cuts, and for a nurse scheduling problem we derive pseudo-valid cutting planes that are tight. Numerical experiments for the latter type of cutting planes show that we are able to find good first-stage decisions, also for large-scale problems. In line with the error bound that we derive, the performance of the pseudo-valid cutting planes is better if the standard deviations of the random variables in the model are larger.

A direction for future research is to derive problem-specific pseudo-valid cutting planes for applications

of two-stage MISPs. Moreover, tighter error bounds may be derived for these problem-specific cutting plane approximations using the special structure of the problems, similar as for simple integer recourse models. Alternatively, error bounds may be derived for pseudo-valid cutting planes for two-stage MISPs with discretely distributed random right-hand-side vectors. Another future research direction is to combine exact and pseudo-valid cutting planes to obtain more accurate approximations at the expense of increasing the computational effort of solving the approximation.

Acknowledgments

The authors are grateful to Suvrajeet Sen for many beneficial discussions and for his feedback on the first version of this manuscript. Also, they thank two anonymous referees and an associate editor for their suggestions on the previous version of this paper. Finally, they thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine high-performance computing cluster.

Appendix

Proof of Lemma 1. Because C is a compact convex set, every $z \in C$ can be written as a convex combination of its extreme points:

$$z = \sum_{j=1}^J \mu_j z^j,$$

with $\sum_{j=1}^J \mu_j = 1$, and $\mu_j \geq 0, j = 1, \dots, J$. Since v is convex, this implies that for all $z \in C$

$$\begin{aligned} v(z) &= v\left(\sum_{j=1}^J \mu_j z^j\right) \leq \sum_{j=1}^J \mu_j v(z^j) \\ &= \sum_{j=1}^J \mu_j (a^\top z^j + b) = a^\top z + b. \end{aligned} \quad (23)$$

To prove that also $v(z) \geq a^\top z + b$ for all $z \in C$, assume for contradiction that there exists $\bar{z} \in C$ such that $v(\bar{z}) < a^\top \bar{z} + b$. Because C is convex and z^0 is an interior point of C , there exists $\epsilon > 0$ such that $\hat{z} := z^0 + \epsilon(z^0 - \bar{z}) \in C$. This point $\hat{z}$ is defined in such a way that z^0 can be written as a convex combination of $\bar{z}$ and $\hat{z}$:

$$z^0 = \frac{1}{1+\epsilon} \hat{z} + \frac{\epsilon}{1+\epsilon} \bar{z}.$$

Since v is convex, this implies that

$$\begin{aligned} v(z^0) &\leq \frac{1}{1+\epsilon} v(\hat{z}) + \frac{\epsilon}{1+\epsilon} v(\bar{z}) < \frac{1}{1+\epsilon} (a^\top \hat{z} + b) + \frac{\epsilon}{1+\epsilon} (a^\top \bar{z} + b) \\ &= a^\top z^0 + b, \end{aligned} \quad (24)$$

where we use that $v(\bar{z}) < a^\top \bar{z} + b$ by assumption and $v(\hat{z}) \leq a^\top \hat{z} + b$ by (23). Because (24) contradicts the assumption that $v(z^0) = a^\top z^0 + b$, we conclude that $v(z) = a^\top z + b$ for all $z \in C$. $\square$

Proof of Proposition 1. Because $\hat{v}(\omega, z) = v(\omega, z)$ for all $\omega \in \Omega$ and $z \in \alpha + \beta \mathbb{Z}^m$, it follows from Theorem 3 that for every $k = 1, \dots, K$, there exists d_k such that for all $\omega \in \Omega$ and $z \in \alpha + \beta \mathbb{Z}^m$ with $\omega - z \in \Lambda^k(d_k)$,

$$\hat{v}(\omega, z) = q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - z).$$

Fix $\omega \in \Omega$. Then $\psi^k(\omega - z)$ is B_k -periodic in z and thus $\psi^k(\omega - z) = \psi^k(\omega - z - \det(B_k)l)$ for every $l \in \mathbb{Z}^m$ by lemma 4.8 in Romeijnders et al. (2016a). Define $\delta^k := \det(B_k)\beta \in \mathbb{Z}^m$ and let $l \in \mathbb{Z}^m$ be given, and consider the hyperrectangle $C^l(\alpha, \delta^k)$. Let $z^j, j = 1, \dots, J$, denote its extreme points and let $z^0 := (\alpha_1 + \delta_1^k l_1, \dots, \alpha_m + \delta_m^k l_m)$ be an interior point. If $C^l(\alpha, \delta^k) \subset \omega - \Lambda^k(d_k)$, then we can apply Lemma 1 with $a := -q_{B_k}^\top B_k^{-1}$, $b := q_{B_k}^\top B_k^{-1}\omega + \psi^k(\omega - \alpha)$, and $C := C^l(\alpha, \delta^k)$ to conclude that $\hat{v}(\omega, z) = q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - \alpha)$ for all $z \in C^l(\alpha, \delta^k)$. Because the diameter of $C^l(\alpha, \delta^k)$ is $2\|\delta^k\|$, we conclude that the result holds for all $\omega - z \in \Lambda^k(d_k + 2\|\delta^k\|)$. Indeed, $\omega - z$ will be in $\omega - C^l(\alpha, \delta^k)$ for some $l \in \mathbb{Z}^m$. The claim now follows by defining $d'_k := d_k + 2\|\delta^k\|$. $\square$

Proof of Lemma 2. Fix $\omega \in \Omega$. Then, by Proposition 1, it follows that for every $k = 1, \dots, K$,

$$\hat{v}(\omega, z) = q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - \alpha), \quad z \in \omega - \Lambda^k(d'_k).$$

Because $\hat{v}(\omega, z)$ is convex in z and affine on $\omega - \Lambda^k(d'_k)$, we can derive a subgradient inequality for each $k = 1, \dots, K$:

$$\hat{v}(\omega, z) \geq q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - \alpha), \quad z \in \mathbb{R}^m.$$

Combining these inequalities over all $k = 1, \dots, K$, yields the desired result. $\square$

Proof of Lemma 3. Fix $\omega \in \Omega$. By Proposition 1, there exist distances $d'_k \geq 0$ such that for every $k = 1, \dots, K$,

$$\hat{v}(\omega, z) = q_{B_k}^\top B_k^{-1}(\omega - z) + \psi^k(\omega - \alpha), \quad z \in \omega - \Lambda^k(d'_k).$$

Thus, on the linearity regions $\omega - \Lambda^k(d'_k)$, $\hat{v}(\omega, z)$ equals its lower bound from Lemma 2. Therefore, we only have to show (13) for $z \in \omega - \mathcal{N}$. To this end, let $z \in \omega - \mathcal{N}$ be given. Because $\mathcal{N}$ can be covered by finitely many hyperslices, there exist $a_j \in \mathbb{R}^m \setminus \{0\}$ and $\delta_j > 0, j \in \mathcal{J}$, such that

$$\mathcal{N} \subset \bigcup_{j \in \mathcal{J}} H_j,$$

where $H_j := H(a_j, \delta_j), j \in \mathcal{J}$. We will construct points z_1 and z_2 in the linearity regions of $\hat{v}(\omega, \cdot)$, so that z is a convex combination of z_1 and z_2 . Then we can use that $\hat{v}(\omega, z)$ is convex in z to derive an upper bound on $\hat{v}(\omega, z)$ in terms of $\hat{v}(\omega, z_1)$ and $\hat{v}(\omega, z_2)$. Because z_1 and z_2 are in the linearity regions, these values are known.

To construct such z_1 and z_2 , let $d \in \mathbb{R}^m \setminus \{0\}$ be a direction of unit length not parallel to any of the hyperslices H_j and thus not orthogonal to any of the normal vectors $a_j, j \in \mathcal{J}$. Then $a_j^\top d \neq 0, j \in \mathcal{J}$, and $\|d\| = 1$. We consider the line through z with direction d and define the halflines L_1 and L_2 as

$$L_1 := \{z + \mu d : \mu \in \mathbb{R}_+\} \quad \text{and} \quad L_2 := \{z - \mu d : \mu \in \mathbb{R}_+\}.$$

Because the direction d is not parallel to any of the hyperslices, we have $L_1 \not\subset \bigcup_{j \in \mathcal{J}} H_j$ and $L_2 \not\subset \bigcup_{j \in \mathcal{J}} H_j$, and thus

$L_i \cap (\omega - \bigcup_{k=1}^K \Lambda^k(d'_k)) \neq \emptyset$, $i = 1, 2$. This means that it is possible to select $z^1, z^2 \in \omega - \bigcup_{k=1}^K \Lambda^k(d'_k)$ on L_1 and L_2 , respectively, with minimal distance to z :

$$z^i := \arg \min_{z'} \left\{ \|z - z'\| : z' \in L_i \cap \left(\omega - \bigcup_{k=1}^K \Lambda^k(d'_k) \right) \right\}, \quad i = 1, 2.$$

Because z is on the line segment between z^1 to z^2 , we can write z as a convex combination $z = \mu z^1 + (1 - \mu)z^2$ of z^1 and z^2 with $\mu \in [0, 1]$. We will use the convexity of $\hat{v}(\omega, \cdot)$ to derive an upper bound on $\hat{v}(\omega, z)$. Here, we will assume without loss of generality that $z^1 \in \omega - \Lambda^{k_1}(d'_{k_1})$ and $z^2 \in \omega - \Lambda^{k_2}(d'_{k_2})$ with $k_1, k_2 \in \{1, \dots, K\}$. We obtain

$$\begin{aligned} \hat{v}(\omega, z) &\leq \mu \hat{v}(\omega, z^1) + (1 - \mu) \hat{v}(\omega, z^2) \\ &= \mu \left(q_{B_{k_1}}^\top B_{k_1}^{-1} (\omega - z^1) + \psi^{k_1}(\omega - \alpha) \right) \\ &\quad + (1 - \mu) \left(q_{B_{k_2}}^\top B_{k_2}^{-1} (\omega - z^2) + \psi^{k_2}(\omega - \alpha) \right). \end{aligned}$$

To obtain the bound in (13) on the difference between $\hat{v}(\omega, z)$ and its lower bound, we subtract this lower bound from both the left- and right-hand sides of the inequality above. Defining $k^* := \arg \max_{k=1, \dots, K} \{q_{B_k}^\top B_k^{-1} (\omega - z) + \psi^k(\omega - \alpha)\}$, the difference between $\hat{v}(\omega, z)$ and its lower bound then can be bounded by

$$\begin{aligned} &\mu \left(q_{B_{k_1}}^\top B_{k_1}^{-1} (\omega - z^1) + \psi^{k_1}(\omega - \alpha) - q_{B_{k^*}}^\top B_{k^*}^{-1} (\omega - z) - \psi^{k^*}(\omega - \alpha) \right) \\ &\quad + (1 - \mu) \left(q_{B_{k_2}}^\top B_{k_2}^{-1} (\omega - z^2) + \psi^{k_2}(\omega - \alpha) - q_{B_{k^*}}^\top B_{k^*}^{-1} (\omega - z) - \psi^{k^*}(\omega - \alpha) \right). \end{aligned}$$

Because k_1 and k_2 are not necessarily the maximizing index for $\max_{k=1, \dots, K} \{q_{B_k}^\top B_k^{-1} (\omega - z) + \psi^k(\omega - \alpha)\}$, we may replace k^* by k_1 and k_2 , respectively, to obtain, after straightforward simplifications,

$$\begin{aligned} &\hat{v}(\omega, z) - \max_{k=1, \dots, K} \left\{ q_{B_k}^\top B_k^{-1} (\omega - z) + \psi^k(\omega - \alpha) \right\} \\ &\leq \mu q_{B_{k_1}}^\top B_{k_1}^{-1} (z - z^1) + (1 - \mu) q_{B_{k_2}}^\top B_{k_2}^{-1} (z - z^2). \end{aligned}$$

We bound the right-hand side in terms of the distance $\|z^1 - z^2\|$ between z^1 and z^2 . Here, we use $\lambda_i^* := \max_{k=1, \dots, K} |q_{B_k}^\top (B_k)^{-1} e_i|$, for $i = 1, \dots, m$, where e_i is the i th unit vector. We have

$$\begin{aligned} &\hat{v}(\omega, z) - \max_{k=1, \dots, K} \left\{ q_{B_k}^\top B_k^{-1} (\omega - z) + \psi^k(\omega - \alpha) \right\} \\ &\leq \mu \sum_{i=1}^m \lambda_i^* |z_i^1 - z_i| + (1 - \mu) \sum_{i=1}^m \lambda_i^* |z_i^2 - z_i| \\ &\leq \mu \sum_{i=1}^m \lambda_i^* \|z^1 - z\| + (1 - \mu) \sum_{i=1}^m \lambda_i^* \|z^2 - z\| \\ &\leq \|z^1 - z^2\| \sum_{i=1}^m \lambda_i^*, \end{aligned} \quad (25)$$

where the last inequality holds, because z is on the line segment between z_1 to z_2 , and thus $\|z^1 - z\| \leq \|z^1 - z^2\|$ and $\|z^2 - z\| \leq \|z^1 - z^2\|$.

It remains to derive an upper bound on $\|z^1 - z^2\|$. To do so, observe that $\omega - z$ is on the line segment $\omega - L$ between $\omega - z^1$ and $\omega - z^2$. Moreover, this line segment is completely

contained in the union of the hyperslices H_j , $j \in \mathcal{J}$. Hence, letting $\mathcal{L}(L)$ denote the length of line segment L , we have

$$\begin{aligned} \|z^1 - z^2\| &= \|(\omega - z^1) - (\omega - z^2)\| \\ &= \mathcal{L} \left((\omega - L) \cap \left(\bigcup_{j \in \mathcal{J}} H_j \right) \right) \\ &= \mathcal{L} \left(\bigcup_{j \in \mathcal{J}} ((\omega - L) \cap H_j) \right) \leq \sum_{j \in \mathcal{J}} \mathcal{L}((\omega - L) \cap H_j). \end{aligned}$$

To find $\mathcal{L}((\omega - L) \cap H_j)$, observe that $\hat{z} \in L$ satisfies $\hat{z} = z - \hat{\mu}d$ for some $\hat{\mu} \in \mathbb{R}$. Moreover, $\omega - \hat{z} \in H_j := H(a_j, \delta_j)$ if $0 \leq a_j^\top (\omega - z + \hat{\mu}d) \leq \delta_j$, or equivalently if

$$\begin{cases} \frac{-a_j^\top (\omega - z)}{a_j^\top d} =: \underline{\mu} \leq \hat{\mu} \leq \bar{\mu} := \frac{\delta_j - a_j^\top (\omega - z)}{a_j^\top d}, & \text{if } a_j^\top d > 0, \\ \frac{\delta_j - a_j^\top (\omega - z)}{a_j^\top d} =: \underline{\mu} \leq \hat{\mu} \leq \bar{\mu} := \frac{-a_j^\top (\omega - z)}{a_j^\top d}, & \text{if } a_j^\top d < 0. \end{cases}$$

Then $\mathcal{L}((\omega - L) \cap H_j) = (\bar{\mu} - \underline{\mu})\|d\| = \frac{\delta_j}{|a_j^\top d|}$, where we use that $\|d\| = 1$. Thus, by defining

$$R' := \left(\sum_{i=1}^m \lambda_i^* \right) \left(\sum_{j \in \mathcal{J}} \frac{\delta_j}{|a_j^\top d|} \right),$$

the claim follows from combining $\|z^1 - z^2\| \leq \sum_{j \in \mathcal{J}} \frac{\delta_j}{|a_j^\top d|}$ and (25). $\square$

Proof of Proposition 2. Consider the LP-relaxation $v_{LP}(\omega, z)$ of $v(\omega, z)$ as defined in (14). Then, by, for example, Blair and Jeroslow (1979) and Cook et al. (1986), there exists R'' such that $\|v - v_{LP}\|_\infty \leq R''$. Moreover, by combining Lemmas 2 and 3, we conclude that $\|v_{LP} - \hat{v}\|_\infty \leq R' + \max_{k=1, \dots, K} \sup_{s \in \mathbb{R}^m} |\psi^k(s)|$. If we define $R := R'' + R' + \max_{k=1, \dots, K} \sup_{s \in \mathbb{R}^m} |\psi^k(s)|$, then

$$\begin{aligned} \|v - \hat{v}\| &\leq \|v - v_{LP}\| + \|v_{LP} - \hat{v}\| \\ &\leq R'' + R' + \max_{k=1, \dots, K} \sup_{s \in \mathbb{R}^m} |\psi^k(s)| =: R, \end{aligned}$$

where the first inequality follows from the triangle inequality. Moreover, $\sup_{s \in \mathbb{R}^m} |\psi^k(s)|$ is finite for every $k = 1, \dots, K$, because by Theorem 3 the function ψ^k is bounded for every $k = 1, \dots, K$. $\square$

References

- Ahmed S, Tawarmalani M, Sahinidis NV (2004) A finite branch-and-bound algorithm for two-stage stochastic integer programs. *Math. Programming* 100(2):355–377.
- Balas E, Jeroslow RG (1980) Strengthening cuts for mixed integer programs. *Eur. J. Oper. Res.* 4(4):224–234.
- Bansal M, Huang K-L, Mehrotra S (2018) Tight second stage formulations in two-stage stochastic mixed integer programs. *SIAM J. Optim.* 28(1):788–819.
- Blair CE, Jeroslow RG (1979) The value function of a mixed integer program: II. *Discrete Math.* 25(1):7–19.
- Bodur M, Dash S, Günlük O, Luedtke J (2017) Strengthened benders cuts for stochastic integer programs with continuous recourse. *INFORMS J. Comput.* 29(1):77–91.

- Carøe CC, Schultz R (1999) Dual decomposition in stochastic integer programming. *Oper. Res. Lett.* 24(1–2):37–45.
- Carøe CC, Tind J (1997) A cutting-plane approach to mixed 0-1 stochastic integer programs. *European J. Oper. Res.* 101(2):306–316.
- Cook W, Gerards AMH, Schrijver A, Tardos É (1986) Sensitivity theorems in integer linear programming. *Math. Programming* 34(3):251–264.
- Gade D, Küçükyavuz S, Sen S (2014) Decomposition algorithms with parametric Gomory cuts for two-stage stochastic integer programs. *Math. Programming* 144(1–2):39–64.
- Guan Y, Ahmed S, Nemhauser GL (2009) Cutting planes for multistage stochastic integer programs. *Oper. Res.* 57(2):287–298.
- Han Y, Chen Z (2015) Quantitative stability of full random two-stage stochastic programs with recourse. *Optim. Lett.* 9(6):1075–1090.
- Kim K, Mehrotra S (2015) A two-stage stochastic integer programming approach to integrated staffing and scheduling with application to nurse management. *Oper. Res.* 63(6):1431–1451.
- Klein Haneveld WK, Stougie L, van der Vlerk MH (2006) Simple integer recourse models: convexity and convex approximations. *Math. Programming* 108(2–3):435–473.
- Kleywegt AJ, Shapiro A, Homem de Mello T (2002) The sample average approximation method for stochastic discrete optimization. *SIAM J. Optim.* 12(2):479–502.
- Laporte G, Louveaux FV (1993) The integer L-shaped method for stochastic integer programs with complete recourse. *Oper. Res. Lett.* 13(3):133–142.
- Louveaux FV, van der Vlerk MH (1993) Stochastic programming with simple integer recourse. *Math. Programming* 61(1–3):301–325.
- Marchand H, Martin A, Weismantel R, Wolsey L (2002) Cutting planes in integer and mixed integer programming. *Discrete Appl. Math.* 123(1–3):397–446.
- Ntaimo L (2010) Disjunctive decomposition for two-stage stochastic mixed-binary programs with random recourse. *Oper. Res.* 58(1):229–243.
- Ntaimo L, Sen S (2005) The million-variable “march” for stochastic combinatorial optimization. *J. Global Optim.* 32(3):385–400.
- Qi Y, Sen S (2017) The Ancestral Benders’ cutting plane algorithm with multi-term disjunctions for mixed-integer recourse decisions in stochastic programming. *Math. Programming* 161(1–2):193–235.
- Rinnooy Kan AHG, Stougie L (1988) Stochastic integer programming. Yu E, Wets RJ-B, eds. *Numerical Techniques for Stochastic Optimization*, vol. 10 (Springer, New York), 201–213.
- Romeijnders W, Schultz R, van der Vlerk MH, Klein Haneveld WK (2016a) A convex approximation for two-stage mixed-integer recourse models with a uniform error bound. *SIAM J. Optim.* 26(1):426–447.
- Romeijnders W, van der Vlerk MH, Klein Haneveld WK (2015) Convex approximations of totally unimodular integer recourse models: a uniform error bound. *SIAM J. Optim.* 25(1):130–158.
- Romeijnders W, van der Vlerk MH, Klein Haneveld WK (2016b) Total variation bounds on the expectation of periodic functions with applications to recourse approximations. *Math. Programming* 157(1):3–46.
- Schultz R, Stougie L, van der Vlerk MH (1998) Solving stochastic programs with integer recourse by enumeration: a framework using Gröbner basis reductions. *Math. Programming* 83(1–3):229–252.
- Sen S, Hingle JL (2005) The C^3 theorem and a D^2 algorithm for large scale stochastic mixed-integer programming: Set convexification. *Math. Programming* 104(1):1–20.
- Sen S, Sherali HD (2006) Decomposition with branch-and-cut approaches for two-stage stochastic mixed-integer programming. *Math. Programming* 106(2):203–223.
- Sherali HD, Zhu X (2006) On solving discrete two-stage stochastic programs having mixed-integer first- and second-stage variables. *Math. Programming* 108(2–3):597–616.
- van der Laan N, Romeijnders W, van der Vlerk MH (2018) Higher-order total variation bounds for expectations of periodic functions and simple integer recourse approximations. *Comput. Management Sci.* 15(3–4):325–349.
- van der Vlerk MH (1995) Stochastic programming with integer recourse. Thesis Rijksuniversiteit Groningen, Theses on Systems, Organisations, and Management, Labyrint Publications, Capelle a/d IJssel.
- van der Vlerk MH (2004) Convex approximations for complete integer recourse models. *Math. Programming* 99(2):297–310.
- van der Vlerk MH (2010) Convex approximations for a class of mixed-integer recourse models. *Ann. Oper. Res.* 177(1):139–150.
- van Slyke RM, Wets R (1969) L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM J. Appl. Math.* 17(4):638–663.
- Walkup DW, Wets RJ-B (1969) Lifting projections of convex polyhedra. *Pacific J. Math.* 28(2):465–475.
- Zhang M, Küçükyavuz S (2014) Finitely convergent decomposition algorithms for two-stage stochastic pure integer programs. *SIAM J. Optim.* 24(4):1933–1951.

Ward Romeijnders is an assistant professor in the Department of Operations at the University of Groningen. His main research interest is stochastic optimization with integer decision variables.

Niels van der Laan is a PhD student in the Department of Operations at the University of Groningen. His PhD project focuses on stochastic mixed-integer programming problems.