STOCHASTIC DUAL DYNAMIC PROGRAMMING AND ITS VARIANTS – A REVIEW

CHRISTIAN FÜLLNER* AND STEFFEN REBENNACK*

Abstract. We provide a tutorial-type review on stochastic dual dynamic programming (SDDP), as one of the state-of-the-art solution methods for large-scale multistage stochastic programs. Since introduced about 30 years ago for solving large-scale multistage stochastic linear programming problems in energy planning, SDDP has been applied to practical problems from several fields and is enriched by various improvements and enhancements to broader problem classes. We begin with a detailed introduction to SDDP, with special focus on its motivation, its complexity and required assumptions. Then, we present and discuss in depth the existing enhancements as well as current research trends, allowing for an alleviation of those assumptions.

Key words. stochastic dual dynamic programming, dynamic programming, multistage stochastic programming, sequential decision problems, large-scale optimization, linear programming, nested Benders decomposition, sampling-based optimization, global optimization

1. Introduction.

In many decision-making situations at least some of the data are uncertain. While this uncertainty is often disregarded, the importance of taking it into account during the decision process was already recognized in 1955 by George Dantzig [44]. In stochastic programming, a common approach to achieve this is to split up this process into two different stages: At the first stage, decisions have to be taken before any uncertain data are revealed and to hedge against the existing uncertainty (so-called *here-and-now* decisions). At the second stage, corrective actions, called *recourse* or *wait-and-see* decisions, can be taken, once the realization of the uncertain data is known [26]. Typically, the aim is to determine an optimal decision rule *in expectation* or with respect to some risk measure.

In many practical applications, not only two, but multiple subsequent decisions have to be taken [7]. If these decisions cannot be taken independently, but are coupled by their effects on a system state, *e.g.*, hydroelectric generation affecting the water level of a reservoir, or orders affecting the size of an inventory stock, this can be modeled as a multistage stochastic problem with several subsequent recourse decisions (this is also referred to as *dynamic programming*, and was recently coined *sequential decision problem* in [169]). In such a problem, trade-offs have to be made between using an existing resource immediately or saving it up for later stages, taking into account the future uncertainty.

Stochastic dual dynamic programming (SDDP) is an algorithm to tackle such multistage stochastic problems in order to compute, or at least approximate, an optimal *policy*, that is, a strategy or decision rule providing the best here-and-now decision as well as the best wait-and-see decisions for any stage and any given realization of the uncertain data. It was first proposed by Pereira and Pinto in 1991 in [152].

Historically, SDDP has its roots in two separate research streams dealing with sequential decision problems. The first one is *stochastic dynamic programming* (SDP), which is closely related to stochastic optimal control and Markov decision processes. Here, a crucial assumption is that the uncertain data on different stages of the decision process are independent of each other (or at least Markovian). In this case, multistage stochastic problems can be expressed by dynamic programming equations (DPE), which decompose the large-scale problem by stages into several smaller subproblems.

*Institute for Operations Research (IOR), Stochastic Optimization (SOP), Karlsruhe Institute of Technology (KIT), Germany (christian.fuellner@kit.edu, steffen.rebennack@kit.edu)

These DPE exploit the famous optimality principle by Bellman [13], which allows one to express the optimal objective value from some stage t onwards, given some state x_{t-1}, recursively by means of some stage-t objective function and a so-called *expected value function* $\mathcal{Q}_t(\cdot)$, modeling the expected optimal objective value from stage $t+1$ onwards, given the new state x_t. We formally introduce these concepts in Section 2.4.

The DPE can be solved exactly by SDP solution methods, such as value iteration [13]. Basically, this method is based on traversing the stages backwards and evaluating the expected value functions $\mathcal{Q}_t(\cdot)$ for all possible states x_{t-1} (concept of a lookup table). Each such evaluation requires solving an optimization problem for all possible realizations of the uncertain data, which, in turn, requires finding an optimal decision over all possible actions. For this evaluation to be possible, it is assumed that the state space, the action space and the scenario space are finite – otherwise they have to be discretized. However, even in the discrete case, enumerating all possible combinations is computationally intractable for all but low dimensions, as the number of evaluations suffers from combinatorial explosion. This phenomenon is known as the *curse of dimensionality* of SDP [168]. In order to circumvent this, *approximate dynamic programming* (ADP) methods have been developed, where expected value functions are approximated instead of being evaluated exactly (or where optimal policies are approximated using different strategies) [168, 169]. SDDP can be regarded as one such method. Due to its close relation with SDP it also heavily relies on the assumption of stagewise independence.

A second perspective on SDDP is one from *stochastic programming*. Traditionally, in this field, multistage uncertain data are often modeled by a scenario tree, which branches at each stage and consists of finitely many possible scenarios. Scenario trees do not require the stochastic data process to be stagewise independent. Using finite scenario trees and assuming linearity, a multistage stochastic program can be reformulated as a large-scale linear programming problem [178]. However, in this extensive form such a problem usually is way too large to be solved by monolithic approaches, since the number of decision variables and constraints grows exponentially in the number of stages. To cope with this challenge, special solution techniques are required which decompose the problem. Based on the L-shaped method for solving two-stage stochastic programs [226] (a special variant of Benders decomposition [17]), one such idea is the extension of Benders-type solution methods to the multistage setting. The *nested Benders decomposition* (NBD) method by Birge [24] is such an extension. It can be interpreted as a nested sequence of solving two-stage stochastic programs while traversing the scenario tree. In contrast to SDP, in NBD the functions $\mathcal{Q}_t(\cdot)$ are not evaluated at all possible states, but iteratively approximated by linear functions called cutting-planes or *cuts*, starting from a rough initial relaxation. Such approximation is possible, since $\mathcal{Q}_t(\cdot)$ can be proven to be convex in x_{t-1} for LPs. It also allows to consider a continuous state space without discretization.

While NBD is a reasonable method to solve multistage stochastic linear programs of moderate time horizons (maximum 4 or 5 time steps), for larger problems, it is still computationally prohibitive, as the scenario tree grows exponentially in the number of stages. As a relief, several methods have been proposed to combine the cutting-plane approximations in NBD with sampling techniques from simulation [38, 53, 101]. The most prominent among these methods is SDDP. From this perspective, SDDP can be considered a sampling-based variant of NBD. In order to use the sampling step in a beneficial way, compared to NBD, SDDP comes with the additional prerequisite that the data process is stagewise independent.

Application-wise, the development of SDDP is closely related to hydrothermal

operational planning, which attempts to determine cost-optimal generation decisions for thermal and hydroelectric power plants over several stages, while ensuring system balance and satisfaction of technical constraints. Since future water availability is affected by uncertain inflows into hydro reservoirs, this optimization problem can be considered multistage, stochastic, and thus very complex.

Prior to SDDP, various solution techniques had been proposed to tackle this type of problem. Among those are simulation models, linear programming techniques (either based on assuming inflows as deterministic or based on reformulating stochastic LPs into a deterministic equivalent), special variants of dynamic programming and SDP [228]. However, all of these techniques either do not consider the uncertain nature of inflows, suffer from the aforementioned curses of dimensionality or do not guarantee convergence. For operating a large-scale power system dominated by hydro power these shortcomings are severe, as they prohibit a cost-minimal and reliable, but at the same time computationally efficient operational planning. The development of SDDP by Pereira and Pinto was directly driven by the endeavor to replace SDP with a more efficient optimization technique in operating the Brazilian power system. While it avoids *some* of the computational drawbacks of SDP or NBD (sometimes advertized as "breaking the curse of dimensionality"), SDDP comes with its own shortcomings, as we thoroughly discuss in this paper.

Since its invention in 1991 SDDP has gained enormous interest, both from a theoretical and an application perspective. To this date, it can be considered one of the state-of-the-art solution methods for large-scale multistage stochastic problems. For this reason, it is used in various practical applications to optimize decision processes, for instance hydrothermal operational planning, portfolio optimization or inventory management, see Section 9.

Several extensions and improvements of SDDP have been proposed by now, many of them attempting to relax the originally required theoretical assumptions, making SDDP applicable to broader problem classes. Others strive for improving the performance of SDDP because, despite its merits, the algorithm may take too long to converge for large problem instances.

Due to both, the sheer amount and the variety of proposed enhancements, SDDP has developed into a wide-ranging research area with several sub-branches, becoming increasingly difficult to keep track of. In this article, we give a comprehensive tutorial-type review on SDDP-related research, covering its basic principle and assumptions, strengths and weaknesses, existing extensions and current research trends.

1.1. Structure. The structure of this review is summarized in Table 1. The review can be divided into four major parts. In the first part (Sections 2 to 8), we discuss the basic mechanism of SDDP. This includes formal preliminaries to formulate multistage stochastic decision problems, but also the main algorithmic steps of SDDP and a complexity analysis. In particular, we point out crucial assumptions for standard SDDP to work. In the second part (Sections 9 and 10), we discuss applications, which underline the practical relevance of SDDP, but also the requirement to relax some of the standard assumptions. In the third part (Sections 11 to 20), we discuss various extensions of SDDP to cases where the standard assumptions are relaxed. These extensions comprise modifications of SDDP itself as well as modifications or reformulations of the considered decision problems. Finally, in the fourth part (Section 21), we discuss approaches to improve the computational performance of SDDP.

Table 1: Table of contents.

Part I: The Mechanism of SDDP		
Section 2	Preliminaries	p. 5
Section 3	Standard SDDP	p. 13
Section 4	Convergence & Complexity	p. 21
Section 5	Comparison with Related Methods	p. 24
Section 6	Sampling	p. 27
Section 7	Stopping Criteria	p. 31
Section 8	Exact Upper Bounds and Dual SDDP	p. 33
Part II: Applications of SDDP		
Section 9	Applications	p. 39
Section 10	Software	p. 43
Part III: Extensions of SDDP		
Section 11	Handling Continuous Uncertainty	p. 44
Section 12	Handling Risk-Aversion	p. 49
Section 13	Handling Distributional Uncertainty	p. 66
Section 14	Handling Stagewise Dependent Uncertainty	p. 71
Section 15	Handling Nonlinear Convex Problems	p. 84
Section 16	Handling Mixed-integer and Non-convex Problems	p. 86
Section 17	Handling Infeasibility	p. 89
Section 18	Handling Non-block-diagonal Problems	p. 92
Section 19	Handling Infinite Horizons	p. 92
Section 20	Handling Random Horizons	p. 93
Part IV: Accelerating SDDP		
Section 21	Performance Improvements	p. 94
Section 22	Outlook	p. 106

1.2. Terminology and Notation. As already mentioned, SDDP is linked to several different research fields and communities, such as stochastic programming, dynamic programming, Markov decision processes, optimal control or reinforcement learning, each using different terminology and notation. This aggravates a presentation of SDDP in a form that is familiar and accessible to all those interested.

To our knowledge, the majority of active research on SDDP is conducted by researchers from the stochastic programming community. For this reason, in many sections we resort to stochastic programming language and notation. On the other hand, this review is also dedicated to offer an access to SDDP for practitioners and researchers from fields in which different perspectives and notation are standard. Therefore, we address these differences if required for the understanding of SDDP, and attempt to avoid heavy mathematical programming notation whenever possible, especially in early sections introducing SDDP.

For a general, not SDDP-specific, attempt at unifying different disciplines related to optimization under uncertainty and sequential decision processes into a common framework, we refer to the excellent book [169].

In the following, we denote random variables by bold letters, *e.g.*, $\boldsymbol{\xi}$, and their realizations by letters in normal font, *e.g.*, ξ. To enhance readability, we summarize

Table 2: Abbreviations that are used throughout the text.

(P)AR	(Periodic) Autoregressive process
DPE	Dynamic programming equations
LP	Linear program
MI(N)LP	Mixed-integer (non-)linear program
MSLP	Multistage stochastic linear programming problem
NBD	Nested Benders Decomposition
RHS	Right-hand side
SDP	Stochastic Dynamic Programming
SDDP	Stochastic Dual Dynamic Programming

some recurring abbreviations in Table 2.

2. Preliminaries for SDDP. In order to present SDDP in its standard form, we start by formally introducing the considered decision problem. In particular, we point out assumptions which are crucial for the presented SDDP method to work.

We consider a multistage decision process where decisions x_t have to be taken over some horizon $[T] := \{1, \dots, T\}$ consisting of T stages, with the aim to minimize some objective function subject to constraints. For now, the horizon T is assumed to satisfy the following condition:

ASSUMPTION 1 (Finite and deterministic horizon). *The number $T \in \mathbb{N}$ of stages is finite and deterministic.*

We discuss later how SDDP may be applied to cases where this is not satisfied, see Sections 19 and 20.

2.1. Modeling the Uncertainty. The data in the considered decision process can be subject to uncertainty, which is revealed over time. To this end, we consider a filtered probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with sample space Ω, σ-algebra $\mathscr{F}$ and probability measure $\mathbb{P}$, which models the uncertainty over the horizon $[T]$. Further let $\mathscr{F}_1, \dots, \mathscr{F}_T$ with $\mathscr{F}_T := \mathscr{F}$ be a sequence of σ-algebras containing the events observable up to time t, thus defining a filtration with $\mathscr{F}_1 \subseteq \mathscr{F}_2 \cdots \subseteq \mathscr{F}_T$, and let Ω_t be the sample space restricted to stage $t \in [T]$. We then define a stochastic process $(\boldsymbol{\xi}_t)_{t\in[T]}$ with random vectors $\boldsymbol{\xi}_t : \Omega_t \to \mathbb{R}^{\kappa_t}, \kappa_t \in \mathbb{N}$, over the probability space. These random vectors are assumed to be $\mathscr{F}_t$-measurable functions. We denote their support by $\Xi_t \subseteq \mathbb{R}^{\kappa_t}$ for all $t \in [T]$. For the first stage, the data are assumed deterministic, *i.e.*, Ξ_1 is a singleton. For each random vector $\boldsymbol{\xi}_t$, we denote a specific realization by ξ_t.

As a crucial ingredient for SDDP to work, we assume that the uncertainty on different stages does not depend on each other.

ASSUMPTION 2 (Stagewise independence). *For all $t \in [T]$, the random vector $\boldsymbol{\xi}_t$ is independent of the history $\xi_{[t-1]} := (\xi_1, \dots, \xi_{t-1})$ of the data process.*

Under Assumption 2, the random vectors $\boldsymbol{\xi}_t$ are often referred to as *noises*. This assumption is common in dynamic programming, but not standard in stochastic programming. In practical applications it may not be satisfied. We address how to apply SDDP to problems with stagewise dependent uncertainty in Section 14.

Additionally, we take the following assumptions for the stochastic process.

ASSUMPTION 3 (Known distribution). *The probability distribution $F_{\boldsymbol{\xi}}$ of the data*

process $(\boldsymbol{\xi}_t)_{t\in[T]}$ *is known.*

ASSUMPTION 4 (Exogeneity). *The random variables* $\boldsymbol{\xi}_t$ *are exogeneous, i.e., the distribution* $F_{\boldsymbol{\xi}}$ *of the data process* $(\boldsymbol{\xi}_t)_{t\in[T]}$ *is independent of decisions* $(x_t)_{t\in[T]}$.

ASSUMPTION 5 (Finite randomness). *The support* Ξ_t *of* $\boldsymbol{\xi}_t$ *is finite for all* $t \in [T]$. *The number of noise realizations at stage* $t \in [T]$ *is given by* $q_t \in \mathbb{N}$ *with* $q_1 = 1$.

We discuss how to apply SDDP if Assumption 3 is not satisfied in Section 13. If Assumption 4 is not satisfied, the problem is said to have *decision-dependent* uncertainty [114]. As this case is not covered in the literature on SDDP so far, we do not discuss the relaxation of this assumption.

Assumption 5 is a key assumption for SDDP and standard in dynamic programming and stochastic programming in order to obtain computationally tractable problems. We discuss possible ways to relax it in Section 11. As $\boldsymbol{\xi}_t$ is a discrete and finite random variable for all $t \in [T]$, its distribution $F_{\boldsymbol{\xi}}$ is defined by finitely many realizations $\xi_{tj}, j = 1, \dots, q_t$, and assigned probabilities p_{tj}.

The stagewise independent and finite data process $(\boldsymbol{\xi}_t)_{t\in[T]}$ can be illustrated by a *recombining scenario tree* [178], also called *scenario lattice* [128]. On each stage $t \in [T]$, its nodes represent the possible noise realizations $\xi_{tj}, j = 1, \dots, q_t$. Due to stagewise independence (Assumption 2) all nodes at the same stage have an identical set of child nodes with the same noise realizations and associated probabilities. We call paths $\xi = (\xi_t)_{t\in[T]}$ through the complete tree (stage-T) *scenarios* and index them by $s \in \mathcal{S}$. Note that for each scenario ξ^s, there exists some $j_s \in \{1, \dots, q_t\}$ such that $\xi_t^s = \xi_{tj_s}$. The total number of different scenarios modeled by the tree is $|\mathcal{S}| = \prod_{t\in[T]} q_t$. An example of a recombining scenario tree is presented in Figure 1.

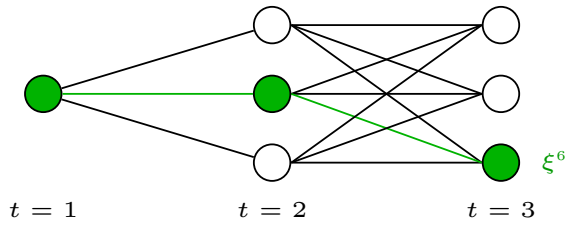


Fig. 1: Recombining tree with 3 realizations per stage and highlighted scenario ξ^6.

2.2. The Decision Process. With the stochastic process in mind, we can now turn to the decision process. At stage 1, the *here-and-now* decision x_1 is taken to hedge against the uncertainty in the following stages. At those stages, recourse decisions $\boldsymbol{x}_t \in \mathbb{R}^{n_t}, n_t \in \mathbb{N}$, can be taken under knowledge of the realization of the data process at stage t. This decision process is illustrated in Figure 2.

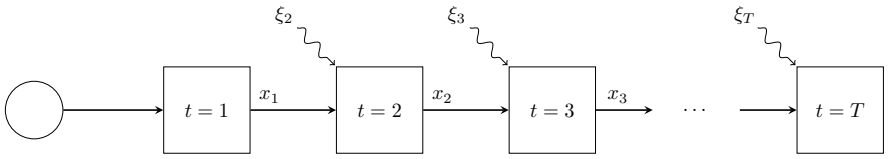


Fig. 2: Multistage decision process with uncertainty.

In other words, the paradigm is that decisions can be taken *after* the uncertainty corresponding to stage t has unfolded (so-called *wait-and-see* decisions), making $\boldsymbol{x}_t(\xi_t)$ a function of $\boldsymbol{\xi}_t$, and by that a random variable. We account for that using a bold symbol. Importantly, $\boldsymbol{x}_t(\cdot)$ does only depend on realizations up to stage t, but does not anticipate future events or decisions. Future events are only considered using distributional information. Therefore, $\boldsymbol{x}_t(\cdot)$ is $\mathscr{F}_t$-measurable [200]. As we will see, $\boldsymbol{x}_t(\cdot)$ may also depend on the choice for $\boldsymbol{x}_{t-1}(\cdot)$ and so on, so that despite stagewise independence (Assumption 2), $\boldsymbol{x}_t(\cdot)$ is actually a function of the whole history $\xi_{[t]}$ of the data process.

A sequence of decision functions $\left(\boldsymbol{x}_t(\xi_{[t]})\right)_{t\in[T]}$ is called a *policy* and provides a decision rule for all stages $t \in [T]$ and any realization of the data process. By the previous arguments, such a policy is *non-anticipative*, modeling a sequence of nested conditional decisions. The aim of the decision process is to determine an *optimal* policy with respect to a given objective function and a given set of constraints.

In this context, the following assumptions are standard for SDDP.

ASSUMPTION 6 (Linearity). *All functions occurring in the objective and the constraints are linear.*

ASSUMPTION 7 (Consecutive coupling). *Only decisions on consecutive stages can be linked by constraints.*

ASSUMPTION 8 (Risk-neutral policy). *The aim is to determine an optimal risk-neutral policy.*

As not all of these assumptions are guaranteed to be satisfied for an arbitrary problem in practice, we discuss possible ways to relax them in Sections 15 and 16 (for Assumption 6), Section 18 (for Assumption 7) and Section 12 (for Assumption 8).

Under Assumptions 6 and 8, the optimization objective can be expressed as

$$\min_{x_1, \boldsymbol{x_2}, \ldots, \boldsymbol{x_T}} \mathbb{E}\left[\sum_{t\in[T]} \left(\boldsymbol{c}_t(\xi_t)\right)^\top \boldsymbol{x}_t(\xi_{[t]})\right], \tag{2.1}$$

with data vectors $c_t \in \mathbb{R}^{n_t}$ for all $t \in [T]$ and $\mathbb{E}[\cdot]$ denoting the expected value.

Under Assumptions 6 and 7, for all $t \in [T]$, the constraints on the decisions can be expressed using the $\mathscr{F}_t$-measurable set-valued mappings $\mathcal{X}_t(\cdot)$, which for any x_{t-1} and any $\xi_t \in \Xi_t$ are defined by

$$\mathcal{X}_t(x_{t-1}, \xi_t) := \Big\{ x_t \in X_t \subset \mathbb{R}^{n_t} \ : \ T_{t-1}(\xi_t)x_{t-1} + W_t(\xi_t)x_t = h_t(\xi_t) \Big\}. \tag{2.2}$$

Here, $h_t \in \mathbb{R}^{m_t}$ are data vectors (for $m_t \in \mathbb{N}$), T_t and W_t are $(m_{t+1} \times n_t)$ and $(m_t \times n_t)$ data matrices and X_t is a non-empty polyhedron, *e.g.*, modeling non-negativity constraints.

As stated before, some (or all) of the problem data can be subject to uncertainty. Hence, for all $t \in [T]$, we consider random variables $\boldsymbol{c}_t(\xi_t), \boldsymbol{T}_{t-1}(\xi_t), \boldsymbol{W}_t(\xi_t)$ and $\boldsymbol{h}_t(\xi_t)$ depending on realizations of $\boldsymbol{\xi}_t$. X_t is considered deterministic. Note again that the first stage is assumed to be deterministic, and that $T_0 \equiv 0$ and $x_0 \equiv 0$. Hence, we define $\mathcal{X}_1 :\equiv \mathcal{X}_1(x_0, \xi_1)$.

Remark 2.1. For notational simplicity, when we deal with finite random variables $\boldsymbol{\xi}_t$ in this paper, we often index the vectors and matrices c_t, T_{t-1}, W_t and h_t with $j = 1, \ldots, q_t$ if we address specific realizations, *e.g.*, $c_{tj} := c_t(\xi_{tj})$.

Remark 2.2 (Dynamic programming perspective). In dynamic programming, Markov decision processes or optimal control, usually a slightly different perspective on sequential decision processes is chosen (see [169] for a comprehensive overview). The main difference is that the occurring variables are differentiated into *state variables* and actual *decisions*. State variables $s_t \in S_t$ model the system state at some stage t. S_t is called the state space. Importantly, state variables may not only comprise the resource state, but also the information or belief state of a system. Local decision variables model decisions on a stage t given a state s_t. In dynamic programming they are usually discrete and called *actions* $a_t \in A_t(s_t)$, in optimal control they are usually continuous and called *controls* $u_t \in U_t(s_t)$. $A_t(s_t)$ and $U_t(s_t)$ are the action space or control space, respectively. The actions or controls are what an agent actually decides on given the current state s_t, whereas the new state s_{t+1} is uniquely determined as $s_{t+1} = \mathcal{T}_t(s_t, u_t, \xi_{t+1})$ using a given transition function $\mathcal{T}_t(\cdot)$ which captures the system dynamic. Therefore, from this perspective, a policy is a sequence of mappings $\boldsymbol{\pi}_t : S_t \to U_t$ from the state space to the control (or action) space. Further note that by proper modeling of the state variable, Assumption 7 is naturally satisfied.

In our above setting, states and actions are intertwined. We can set $s_t = (x_{t-1}, \xi_t)$ and $u_t = x_t$ to switch perspectives [6]. The state space, control space and transition function are then implicitly given by (2.2) and the definition of $\boldsymbol{\xi}_t$.

Whereas our above definitions are prevalent in the literature on SDDP, sometimes also an optimal control perspective is adopted, *e.g.*, in the French community working on SDDP (see for example [78]). However, in this case usually only the resource state r_t is explicitly considered as a state variable (while not including information on $\boldsymbol{\xi}_t$). Translating our above setting, this implies that $r_t = x_{t-1}$ with state space $R_t = X_t$, $u_t = x_t$ and due to $r_{t+1} = u_t$, both the control space $U_t(r_t, \xi_t)$ and the transition function $\mathcal{T}_t(r_t, u_t, \xi_t)$ are given by the equations in (2.2).

Given the constraint sets (2.2) for all $t \in [T]$, let $\mathcal{X}_0 := \{x_0\}$ and recursively define

$$\mathcal{X}_t := \bigcup_{x_{t-1} \in \mathcal{X}_{t-1}} \bigcup_{\xi_t \in \Xi_t} \mathcal{X}_t(x_{t-1}, \xi_t)$$

for all $t \in [T]$ [71]. Using these definitions, we are able to state assumptions which we require for the feasibility of our decision problem:

ASSUMPTION 9. *(Feasibility and Compactness)*

(a) For all $t \in [T]$, all $x_{t-1} \in \mathcal{X}_{t-1}$ and almost all $\xi_t \in \Xi_t$, the set $\mathcal{X}_t(x_{t-1}, \xi_t)$ is a non-empty compact subset of $\mathbb{R}^{n_t}$.

(b) The set $\mathcal{X}_t$ is bounded for all $t \in [T]$.

Remark 2.3. Note that the linearity assumption (see Assumption 6), immediately implies that Assumption 9 (a) is not only satisfied for all $x_{t-1} \in \mathcal{X}_{t-1}$, but for all $x_{t-1} \in \operatorname{conv}(\mathcal{X}_{t-1})$, where $\operatorname{conv}(S)$ denotes the convex hull of a set S.

The set $\mathcal{X}_t \in \mathbb{R}^{n_t}$ is called *reachable set* in [71] and *effective feasible region* in [116]. It may as well be referred to as the state space sometimes, because in our setting x_t also takes the role of a state variable. However, in other cases the larger polyhedral set X_t may be called state space.

The boundedness of $\mathcal{X}_t$ in (b) is required for some of the convergence results on SDDP presented in Section 4. It follows naturally if X_t is bounded, since $\mathcal{X}_t \subseteq X_t$. Property (a) is convenient, but not necessarily required. We discuss possible ways to relax it in Section 17.

With all the ingredients defined, we can now model the decision problem in a form that can be tackled by SDDP. Based on its properties, in the following we refer to this problem as a multistage stochastic linear programming problem (MSLP). If not specified otherwise, throughout this paper, we assume that (MSLP) satisfies Assumptions 1 to 9. We first discuss two different modeling approaches which are common in the literature.

2.3. Single-problem Formulation. One way to model the decision problem (MSLP) is to formulate it as a single optimization problem. This modeling approach is common in the stochastic programming community. The optimization problem can be obtained by combining (2.1) with the constraints in (2.2) for all $t \in [T]$.

Then, under Assumptions 1 to 9, (MSLP) can be written as

$$v^* := \begin{cases} \min\limits_{x_1, \boldsymbol{x_2}, \ldots, \boldsymbol{x_T}} & \mathbb{E}\left[\sum\limits_{t\in[T]} \big(\boldsymbol{c}_t(\xi_t)\big)^\top \boldsymbol{x}_t(\xi_{[t]})\right] \\ \text{s.t.} & x_1 \in \mathcal{X}_1 \\ & \boldsymbol{x}_t \in \mathcal{X}_t(\boldsymbol{x}_{t-1}(\xi_{[t-1]}), \xi_t) \quad \forall \xi_t \in \Xi_t \ \forall t = 2, \ldots, T. \end{cases} \tag{2.3}$$

Importantly, the decision variables $x_t \in \mathbb{R}^{n_t}$ depend on $\boldsymbol{\xi_t}$ (and on x_{t-1}), so in this representation we optimize over policies. A policy $(\boldsymbol{x}_t(\xi_{[t]}))_{t\in[T]}$ is called *feasible* (or *admissible*) if it satisfies the constraints in (MSLP) for almost every realization of the random data [200].

Assumption 9 (a) implies that the feasible set of (MSLP) is compact and non-empty, and by linearity of the objective (Assumption 6) it follows that v^* is finite. Even more, (MSLP) has *relatively complete recourse*, as we discuss in detail in the next subsection.

Due to optimizing over policies, without Assumption 5, (MSLP) is an infinite-dimensional optimization problem. With Assumption 5, however, it can be reformulated to a more accessible form. More precisely, it can be reformulated to a large-scale deterministic problem, the so-called *deterministic equivalent* of (MSLP) in *extensive form* (see [200]). To this end, let $\mathcal{S}$ denote the set of all (stage-T) scenarios. Then, for each scenario $s \in \mathcal{S}$ a separate copy x_t^s of variables x_t can be introduced, so that the optimization over implementable policies translates to an optimization over a finite number of decision variables. However, the problem size grows exponentially in the number of stages T. Therefore, even for a finite number of scenarios, this large-scale LP is too large to be solved by off-the-shelf solvers for all but very small instances.

A preferable solution approach is therefore to use tailored solution techniques which decompose (MSLP) into smaller subproblems. Note that from Assumption 7 and the definition of $\mathcal{X}_t(\cdot)$ in (2.2), it is evident that the constraints of (MSLP) are block-diagonal, as only consecutive stages are coupled in the constraints. This is visualized in Figure 3.

This sequential and block-diagonal structure can be exploited to achieve the required decomposition. This is crucial for the derivation of SDDP. Interestingly, this decomposition idea directly leads to the second common modeling approach for our decision problem.

2.4. Dynamic Programming Equations. An alternative, but equivalent way to model (MSLP) is to exploit the well-known optimality principle by Bellman [13] and to formulate a recursion of so-called *dynamic programming equations* (DPE), where a multistage decision process with stagewise independent (or Markovian) uncertainty is modeled as a coupled sequence of optimization problems.

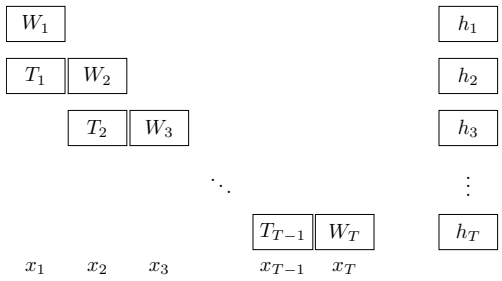


Fig. 3: Block-diagonal structure of constraints in (MSLP).

Whereas this modeling approach is often used in stochastic programming as a way to reformulate and decompose the single problem (2.3) into a computationally tractable form, in dynamic programming it often serves as the starting point of modeling decision problems. However, in contrast to many approaches in dynamic programming we do not discretize x_t, see also Section 5.1.

Under Assumptions 1 to 9, for $t = T, \dots, 2$, the DPE are given by

$$Q_t(x_{t-1}, \xi_t) := \begin{cases} \min\limits_{x_t} & \big(c_t(\xi_t)\big)^\top x_t + \mathcal{Q}_{t+1}(x_t) \\ \text{s.t.} & x_t \in \mathcal{X}_t(x_{t-1}, \xi_t), \end{cases} \tag{2.4}$$

where

$$\mathcal{Q}_{t+1}(x_t) := \mathbb{E}_{\boldsymbol{\xi}_{t+1}}\left[Q_{t+1}(x_t, \boldsymbol{\xi}_{t+1})\right] \tag{2.5}$$

and $\mathcal{Q}_{T+1}(x_T) \equiv 0$. $Q_t(\cdot,\cdot)$ is called *value function* and $\mathcal{Q}_t(\cdot)$ is called *expected value function*, *(expected) cost-to-go function*, *future cost function* or *recourse function*. For the first stage, we obtain

$$v^* = \begin{cases} \min\limits_{x_1} & c_1^\top x_1 + Q_2(x_1) \\ \text{s.t.} & x_1 \in \mathcal{X}_1. \end{cases} \tag{2.6}$$

For a formal proof of the equivalence of (2.3) and its DPE, we refer to [200] and Section 12. Importantly, in subproblem (2.4) x_t is a deterministic variable and not a function, since a fixed realization of ξ_t is considered.

We should emphasize that the equivalence of (2.3) and its DPE does not require Assumption 5. This implies that also the DPE (2.4)-(2.6) are computationally intractable in case of general continuous random variables. While the subproblems are deterministic and finite-dimensional, there exist infinitely many value functions $Q_t(\cdot,\cdot)$ and the evaluation of $\mathcal{Q}_t(\cdot)$ requires the evaluation of (multidimensional) integrals. Therefore, also from this perspective Assumption 5 is crucial.

Remark 2.4 (Dynamic programming control perspective). Recall Remark 2.2. Using a distinction between state variables r_t and controls u_t, the DPE to (MSLP) can be formulated as

$$Q_t(r_t, \xi_t) = \min_{u_t \in U_t(r_t, \xi_t)} f_t(u_t, \xi_t) + \mathcal{Q}_{t+1}(\mathfrak{T}_t(r_t, u_t, \xi_t)). \tag{2.7}$$

Bellman Operator. In the French literature on SDDP, in addition to taking the optimal control perspective discussed in Remarks 2.2 and 2.4, a more formal way to

define the DPE is prevalent, see [71, 118] for instance. To this end, a linear Bellman operator $\widehat{\mathfrak{B}}_t$ is introduced, which applied to some lower semicontinuous function $V : \mathbb{R}^{n_t} \to \mathbb{R} \cup \{+\infty\}$ is defined as [71]

$$\widehat{\mathfrak{B}}_t(V)(x_{t-1}, \xi_t) := \min_{x_t \in \mathcal{X}_t(x_{t-1}, \xi_t)} \big(c_t(\xi_t)\big)^\top x_t + V(x_t), \tag{2.8}$$

i.e., it maps (x_{t-1}, ξ_t) to the optimal value of an optimization problem containing function $V(\cdot)$. We can then further define the operator

$$\mathfrak{B}_t(V)(x_{t-1}) := \mathbb{E}\Big[\widehat{\mathfrak{B}}_t(V)(x_{t-1}, \boldsymbol{\xi}_t)\Big]. \tag{2.9}$$

Setting V to $\mathcal{Q}_t(\cdot)$ for $t = 2, \dots, T$, the (expected) value functions can then be recursively defined in a very compact form. We summarize the different notations for a better overview:

$$\boxed{\begin{aligned} \widehat{\mathfrak{B}}_t(\mathcal{Q}_{t+1})(x_{t-1}, \xi_t) &= Q_t(x_{t-1}, \xi_t) \\ \mathfrak{B}_t(\mathcal{Q}_{t+1})(x_{t-1}) &= \mathcal{Q}_t(x_{t-1}) \end{aligned}}$$

In the remainder of this work, we stick to notation (2.4), as it is most common in the literature on SDDP which we reference in this paper.

We obtain the following properties for the DPE which are standard for SDDP:

LEMMA 2.5. *Under Assumptions 1 to 9, for the DPE defined by (2.4)-(2.6) the following properties hold:*

(a) We have relatively complete recourse, i.e.*, for any $x_{t-1} \in \mathcal{X}_{t-1}$, the stage-$t$ subproblem (2.4) is feasible for all $\xi_t \in \Xi_t$.*

(b) The value functions $Q_t(\cdot, \cdot)$ and expected value functions $\mathcal{Q}_t(\cdot)$ are finite-valued on $\mathrm{conv}(\mathcal{X}_{t-1})$ for all $t = 2, \dots, T$ and all $\xi_t \in \Xi_t$.

(c) Problem (2.6) is feasible and bounded.

Remark 2.6. In addition to Remark 2.2, we should highlight that (MSLP) (both, in single-problem formulation (2.3) and DPE (2.4)-(2.6)) can be straightforwardly enhanced with local decision variables $y_t \in Y_t$ and local constraints, not appearing in different stages. In principle, they can even be incorporated without changes to our models by extending the dimension of the (state) variables x_t and adapting the matrices T_t and W_t accordingly. However, as we explain in Section 4, the complexity of SDDP grows exponentially in the dimension of the state space, so this is computationally detrimental and should be avoided. Instead, purely local variables and constraints should be handled separately from the ones we introduced above. This approach is referred to as *generalized dual dynamic programming* (GDDP) in [18].

While almost every practical application will require the introduction of these additional elements, in this work, for the most part we restrict to coupling variables and constraints which are required to illustrate the mechanics of SDDP.

Remark 2.7. Further note that the local objective functions may also include the states x_{t-1} instead of just depending on x_t and ξ_t. For notational simplicity, we consider a less general form of the objective function in this review.

2.5. Approximations of the Value Functions. The main challenge in exploiting the DPE to solve (MSLP) is that the (expected) value functions are not known in analytical form in advance. The key idea in SDDP is to iteratively approximate them from below using linear functions, which are called *cutting-planes*, or short

cuts. Together, these linear functions build polyhedral outer approximations $\mathfrak{Q}_t(\cdot)$ of $\mathcal{Q}_t(\cdot)$ for all $t = 2, \dots, T$, which we refer to as *cut approximations*. In that regard, SDDP can be considered as a special variant of Kelley's cutting-plane method [110] and closely related to Benders decomposition [17], see also Section 5.2. Note that in contrast to SDP this avoids a state discretization, as $Q_t(\cdot, \cdot)$ and $\mathcal{Q}_t(\cdot)$ do not have to be evaluated at all possible states, but only at well-chosen trial points where new cuts are constructed, cf. Section 5.1.

For this approximation by cuts, the following properties are crucial.

THEOREM 2.8 ([26]). *Let h_t, T_{t-1}, c_t be elements of some convex sets and $x_{t-1} \in conv(\mathcal{X}_{t-1})$. Then, under Assumptions 1 to 9, for all $t = 2, \dots, T$ and a given noise realization ξ_t, the value function $Q_t(\cdot, \xi_t)$*

(a) is piecewise linear and convex in (h_t, T_{t-1}),
(b) is piecewise linear and concave in c_t,
(c) is piecewise linear and convex in x_{t-1} on $conv(\mathcal{X}_{t-1})$.

The main idea here is that given the definition of $\mathcal{X}_{t-1}(\cdot)$ in (2.2), h_t, T_{t-1} and x_{t-1} do only appear in the right-hand side (RHS) of problem (2.4). Therefore, the dual feasible set is independent of those elements. It possesses finitely many extreme points. This assures piecewise linearity of $Q_t(\cdot, \cdot)$, as known from parametric optimization. The convexity follows with the linearity (Assumption 6) and all vectors and matrices being part of convex sets.

Theorem 2.8 directly implies the piecewise linearity and convexity of $\mathcal{Q}_t(\cdot)$.

COROLLARY 2.9 ([26]). *Under Assumption 5 and the premises of Theorem 2.8, for all $t = 2, \dots, T$, $\mathcal{Q}_t(\cdot)$ is piecewise linear and convex in x_{t-1} on $conv(\mathcal{X}_{t-1})$.*

Theorem 2.8 and Corollary 2.9 also directly imply the Lipschitz continuity of the (expected) value functions.

COROLLARY 2.10. *Under Assumptions 1 to 9, for all $t = 2, \dots, T$ and all $\xi_t \in \Xi_t$, $Q_t(\cdot, \xi_t)$ and $\mathcal{Q}_t(\cdot)$ are Lipschitz continuous on $conv(\mathcal{X}_{t-1})$.*

Replacing the true expected value functions with cut approximations in (2.4), we can define *approximate value functions*

$$\underline{Q}_t(x_{t-1}, \xi_t) := \begin{cases} \min\limits_{x_t} & \big(c_t(\xi_t)\big)^\top x_t + \mathfrak{Q}_{t+1}(x_t) \\ \text{s.t.} & x_t \in \mathcal{X}_t(x_{t-1}, \xi_t). \end{cases} \tag{2.10}$$

Trivially, for $\mathcal{Q}_{T+1}(\cdot) \equiv 0$, we have $\mathfrak{Q}_{T+1}(\cdot) \equiv 0$. In SDDP, it is important that $\mathfrak{Q}_t(\cdot)$ should be initialized with a finite and valid lower bound for all $t = 2, \dots, T$ to prevent termination due to the occurrence of unbounded subproblems.

Note that apart from x_{t-1} and ξ_t, $\underline{Q}_t(\cdot, \cdot)$ is also a function of the cut approximation $\mathfrak{Q}_{t+1}(\cdot)$. This is especially relevant when these approximations are iteratively updated in SDDP, leading to different approximate value functions. Using the Bellman operators defined in (2.8)-(2.9) this can be expressed in a very concise way:

$$\underline{Q}_t(\cdot, \cdot) = \mathfrak{B}_t(\mathfrak{Q}_{t+1})(\cdot, \cdot).$$

Similarly, we could express this by adding an argument to $\underline{Q}_t(\cdot, \cdot)$, *i.e.*, by writing $\underline{Q}_t(x_{t-1}, \xi_t \,|\, \mathfrak{Q}_{t+1})$ or $\underline{Q}_t(\mathfrak{Q}_{t+1})(x_{t-1}, \xi_t)$. However, for notational simplicity, we do not state this explicitly, but when dealing with SDDP use the iteration index i for distinction. This means that $\underline{Q}_t^i(\cdot, \cdot)$ indicates that $\underline{Q}_t(\cdot, \cdot)$ is considered with cut approximation $\mathfrak{Q}_{t+1}^i$.

We summarize the different notations for a better overview:

$$\boxed{\begin{aligned}\widehat{\mathfrak{B}}_t(\mathfrak{Q}_{t+1}^i)(x_{t-1},\xi_t) &= \underline{Q}_t^i(x_{t-1},\xi_t)\\ \mathfrak{B}_t(\mathfrak{Q}_{t+1}^i)(x_{t-1}) &= \underline{\mathcal{Q}}_t^i(x_{t-1}) := \mathbb{E}_{\boldsymbol{\xi_t}}\left[\underline{Q}_t^i(x_{t-1},\boldsymbol{\xi_t})\right]\end{aligned}} \tag{2.11}$$

Finally, we can observe that given that the cut approximations $\mathfrak{Q}_{t+1}(\cdot)$ are polyhedral, the approximate value functions $\underline{Q}_t(\cdot,\cdot)$ inherit the previously stated properties from $Q_t(\cdot,\cdot)$. In particular:

LEMMA 2.11. *Let x_{t-1} be element of some convex set and $\mathfrak{Q}_{t+1}(\cdot)$ a polyhedral function. Then, under Assumptions 1 to 9, for all $t = 2,\dots,T$ and a given noise realization ξ_t, $\underline{Q}_t(\cdot,\xi_t)$ is piecewise linear and convex in x_{t-1} on conv$(\mathcal{X}_{t-1})$.*

On the other hand, as they are polyhedral, the cut approximations $\mathfrak{Q}_t(\cdot)$ for $t = 2,\dots,T$ are *nonlinear* functions. Therefore, the subproblems defined by (2.10) are no LPs, even if (MSLP) is a stochastic *linear* problem. Importantly, for computations, subproblems (2.10) can be reformulated as equivalent LPs by using a partial epigraph reformulation and the fact that $\mathfrak{Q}_t(\cdot)$ is polyhedral, that is, defined as the maximum of finitely many affine functions modeled by some set $\mathcal{K}$ with $|\mathcal{K}| \in \mathbb{N}$:

$$\underline{Q}_t(x_{t-1},\xi_t) = \begin{cases} \min\limits_{x_t,\theta_{t+1}} & \left(c_t(\xi_t)\right)^\top x_t + \theta_{t+1} \\ \text{s.t.} & x_t \in \mathcal{X}_t(x_{t-1},\xi_t) \\ & -\left(\beta_{t+1,k}^i\right)^\top x_t + \theta_{t+1} \geq \alpha_{t+1,k}^i, \quad \forall i\ \forall k \in \mathcal{K}. \end{cases} \tag{2.12}$$

This problem is an LP, but contains an additional decision variable θ_{t+1} and finitely many additional linear constraints indexed by i and k. The structure and indexing of these constraints become clear in the next section when we present the cut generation process for SDDP.

3. Standard SDDP.

We are now able to introduce SDDP in its standard form.

3.1. Main Principle.

SDDP consists of two main steps in each iteration i, a *forward pass* and a *backward pass* through the stages $t \in [T]$.

In each forward pass, using the approximate value functions $\underline{Q}_t^i(\cdot,\cdot)$ (recall that this means that we use cut approximation $\mathfrak{Q}_{t+1}^i(\cdot)$ in (2.10)), a sequence of *trial points* $(x_t)_{t\in[T]}$ is generated, at which then new cuts are constructed in the following backward pass to improve the approximation. These trial points are also called incumbents or candidate solutions, and their sequence is called a *state trajectory* (especially in optimal control). The idea behind this approach is that the approximate value functions implicitly define a feasible (suboptimal) policy for problem (MSLP). The trial points are generated by evaluating this policy for one or several scenarios which are sampled from $\mathcal{S}$, *i.e.*, by solving the respective subproblems. This has the advantage that cuts are constructed at points which (at least for some scenario) are optimal given the current cut approximation. This step can also be interpreted as a Monte Carlo *simulation* of the current policy.

In the backward pass, dual information of the subproblems at the trial points is used to construct cuts, passing them back to the previous stage and updating $\mathfrak{Q}_t^i(\cdot)$ to $\mathfrak{Q}_t^{i+1}(\cdot)$ for all $t = 2,\dots,T$. This way, if not optimal, the current policy is amended (at least if the *right* scenario is sampled). In this step, also a *true* lower bound $\underline{v}$ for v^* is determined.

Remark 3.1 (Statistical learning perspective). The basic principle of SDDP can also be interpreted from a perspective of supervized learning as *learning* a policy (or

expected value functions $\mathcal{Q}_t(\cdot)$ for all $t = 2, \ldots, T$) or *training* a model of this policy (or cut approximations $\mathfrak{Q}_t(\cdot)$ for all $t = 2, \ldots, T$) using backpropagation. In the forward pass the inputs are propagated through the stages using the current model, and in the backward pass cuts (representing the error of the current approximation) are propagated back through the stages to update the model.

Algorithm 3.1 provides a pseudo-code for SDDP. We now provide a more detailed and technical look at the algorithmic steps.

Algorithm 3.1 SDDP

Input: Problem (MSLP) satisfying Assumptions 1 to 9. Bounds $\underline{\theta}_t, t = 2, \ldots, T$. Stopping criterion.

`Initialization`

Initialize cut approximations with $\theta_t \geq \underline{\theta}_t$ for all $t = 2, \ldots, T$.
Initialize lower bound with $\underline{v}^0 = -\infty$.
Set iteration counter to $i \leftarrow 0$.

`SDDP Loop`

while Stopping criterion not satisfied **do**
  Set $i \leftarrow i + 1$.

  `Forward Pass`

  Sample a subset $\mathcal{K} \subseteq S$ of scenarios.
  Solve the approximate first-stage problem (3.1) to obtain trial point $x_1^i = x_1^{ik}$ for all $k \in \mathcal{K}$.
  for stages $t = 2, \ldots, T$ **do**
    for samples $k \in \mathcal{K}$ **do**
      Solve the approximate stage-t subproblem (2.10) associated with $\underline{Q}_t^i(x_{t-1}^{ik}, \xi_t^k)$ to obtain trial point x_t^{ik}.
    end for
  end for

  `Backward Pass`

  for stages $t = T, \ldots, 2$ **do**
    for samples $k \in \mathcal{K}$ **do**
      for noise terms $j = 1, \ldots, q_t$ **do**
        Solve the updated approximate stage-t subproblem (2.10) associated with $\underline{Q}_t^{i+1}(x_{t-1}^{ik}, \xi_{tj})$. Store the optimal value and dual vector π_t^{ikj}.
      end for
      Use relations (3.4)-(3.5) and (3.7) to create an optimality cut for $\mathcal{Q}_t(\cdot)$.
      Update the cut approximation $\mathfrak{Q}_t^i(\cdot)$ to $\mathfrak{Q}_t^{i+1}(\cdot)$ using relation (3.6).
    end for
  end for
  Solve the approximate first-stage problem (3.8) to obtain a lower bound $\underline{v}^i$.
end while

Output: (Approximately) optimal feasible policy for (MSLP) defined by x_1^i and cut approximations $\mathfrak{Q}_t^i(\cdot), t = 2, \ldots, T$. x_1^i defines an (approximately) optimal solution to problem (2.6) with $\overline{v}_{\mathcal{K}}^i \approx v^*$.

3.2. Forward Pass. At the start of each iteration i, at first a subset $\mathcal{K} \subseteq \mathcal{S}$ of scenarios is sampled with $|\mathcal{K}| \ll |\mathcal{S}|$ (note that we may equivalently sample stage by stage during the forward pass). The number of samples $|\mathcal{K}|$ may vary by iteration, but we do not state this possible dependence explicitly. Traditionally, and most commonly, in SDDP some random sampling is used, but also a deterministic sampling is possible. We further discuss sampling techniques in Section 6.

Then, at the first-stage, the approximate subproblem

$$\min_{x_1 \in \mathcal{X}_1(x_0)} c_1^\top x_1 + \mathfrak{Q}_2^i(x_1). \tag{3.1}$$

is solved, which yields the trial point $x_1^i = x_1^{ik}$ for all $k \in \mathcal{K}$. Afterwards, for each stage $t = 2, \dots, T$ and each sample $k \in \mathcal{K}$, recursively the approximate value functions $\underline{Q}_t^i(x_{t-1}^{ik}, \xi_t^k)$ are evaluated (this means that the subproblems (2.10) are solved for x_{t-1}^{ik}, ξ_t^k and the current cut approximation $\mathfrak{Q}_{t+1}^i(\cdot)$). This way, for each sample $k \in \mathcal{K}$, a sequence of trial points $(x_t^{ik})_{t \in [T]}$ is obtained.

The forward pass of SDDP is illustrated in Figure 4 for the recombining scenario tree from Figure 1 and $\mathcal{K} = \{1, 3, 9\}$, *i.e.*, $|\mathcal{K}| = 3$. The three sampled scenario paths are highlighted in green. The figure shows that for sample paths ξ^3 and ξ^9 the same node is reached at stage 3.

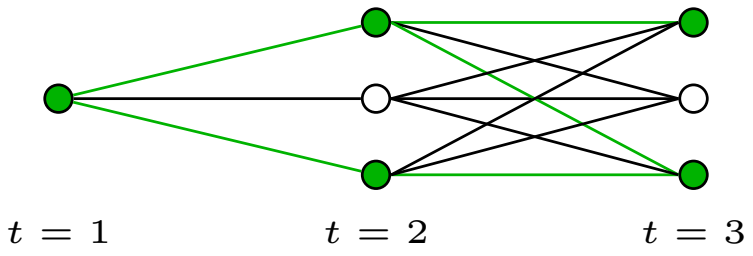


Fig. 4: Illustration of SDDP forward pass for $|\mathcal{K}| = 3$.

3.3. Backward Pass. Main Principle. The backward pass starts at stage T. Here, for all samples $k \in \mathcal{K}$, we consider subproblems (2.10) for the trial point x_{T-1}^{ik} computed in the forward pass, all noise realizations $\xi_{Tj}, j = 1, \dots, q_T$, and $\mathfrak{Q}_{T+1}^{i+1}(\cdot) \equiv 0$. That is, we consider functions $\underline{Q}_T^{i+1}(x_{T-1}^{ik}, \xi_{tj})$ for $j = 1, \dots, q_T$.

As $\underline{Q}_T^{i+1}(\cdot, \xi_{Tj})$ is convex in x_{T-1} by Lemma 2.11, it can be underestimated by a linear function using some subgradient $\beta_{Tkj}^i \in \partial \underline{Q}_T^{i+1}(\cdot, \xi_{Tj})$ for any $j = 1, \dots, q_T$ and any $k \in \mathcal{K}$:

$$\underline{Q}_T^{i+1}(x_{T-1}, \xi_{Tj}) \geq \underline{Q}_T^{i+1}(x_{T-1}^{ik}, \xi_{Tj}) + (\beta_{Tkj}^i)^\top (x_{T-1} - x_{T-1}^{ik}).$$

Since $\underline{Q}_T^{i+1}(\cdot, \xi_{Tj})$ is a lower approximation of the true value function $Q_T(\cdot, \xi_{Tj})$, this directly implies

$$Q_T(x_{T-1}, \xi_{Tj}) \geq \underline{Q}_T^{i+1}(x_{T-1}^{ik}, \xi_{Tj}) + (\beta_{Tkj}^i)^\top (x_{T-1} - x_{T-1}^{ik}).$$

Taking expectations with respect to $\boldsymbol{\xi}_T$ on both sides, we obtain

$$
\begin{aligned}
&\mathcal{Q}_T(x_{T-1}) \\
&\geq \mathbb{E}_{\boldsymbol{\xi}_T}\left[\underline{Q}_T^{i+1}(x_{T-1}^{ik}, \boldsymbol{\xi}_T)\right] + \mathbb{E}_{\boldsymbol{\xi}_T}\left[(\boldsymbol{\beta}_{Tj}^i)^\top (x_{T-1} - x_{T-1}^{ik})\right] \\
&= \mathbb{E}_{\boldsymbol{\xi}_T}\left[\underline{Q}_T^{i+1}(x_{T-1}^{ik}, \boldsymbol{\xi}_T) - (\boldsymbol{\beta}_{Tk}^i)^\top x_{T-1}^{ik}\right] + \left(\mathbb{E}_{\boldsymbol{\xi}_T}\left[\boldsymbol{\beta}_{Tk}^i\right]\right)^\top x_{T-1} \\
&= \underbrace{\sum_{j=1}^{q_T} p_{Tj}\left(\underline{Q}_T^{i+1}(x_{T-1}^{ik}, \xi_{Tj}) - (\beta_{Tkj}^i)^\top x_{T-1}^{ik}\right)}_{=:\alpha_{Tk}^i} + \underbrace{\left(\sum_{j=1}^{q_T} p_{Tj}\beta_{Tkj}^i\right)}_{=:\beta_{Tk}^i}^\top x_{T-1},
\end{aligned}
\tag{3.2}
$$

where we exploit the finiteness of $\boldsymbol{\xi}_T$ (Assumption 5). α_{Tk}^i is called *cut intercept* and β_{Tk}^i is called *cut gradient*. Defining

$$
\phi_{Tk}^i(x_{T-1}) := \alpha_{Tk}^i + (\beta_{Tk}^i)^\top x_{T-1},
$$

we can express (3.2) as

$$
\mathcal{Q}_T(x_{T-1}) \geq \phi_{Tk}^i(x_{T-1}). \tag{3.3}
$$

Inequality (3.3) defines a cut for the expected value function $\mathcal{Q}_T(\cdot)$. Such a cut is constructed for each $k \in \mathcal{K}$. With these new cuts, the cut approximation $\mathfrak{Q}_T^i(\cdot)$ is updated to

$$
\mathfrak{Q}_T^{i+1}(x_{T-1}) := \max\left\{\mathfrak{Q}_T^i(x_{T-1}),\ \phi_{T1}^i(x_{T-1}), \ldots, \phi_{T|\mathcal{K}|}^i(x_{T-1})\right\}.
$$

Thus, assuming that $|\mathcal{K}|$ does not change over the iterations, $\mathfrak{Q}_T^{i+1}(\cdot)$ consists of $i|\mathcal{K}|$ affine functions $\phi_{Tk}^i(\cdot)$, cf. formulation (2.12).

In the same way, for stages $t = T-1, \ldots, 2$, cuts for $\mathcal{Q}_t(\cdot)$ can be constructed by solving subproblems (2.10) for the trial points x_{t-1}^{ik} computed in the forward pass and all noise realizations $\xi_{tj}, j = 1, \ldots, q_t$. Importantly, by going backwards through the stages, at stage t we can already factor in the cuts that have been constructed at the following stage $t+1$, thus using a better approximation as the basis to construct a new cut. This means that we consider $\mathfrak{Q}_{t+1}^{i+1}(\cdot)$ and by that $\underline{Q}_t^{i+1}(\cdot,\cdot)$ with index $i+1$ in the backward pass of iteration i.

As for stage T, we obtain

$$
\mathcal{Q}_t(x_{t-1}) \geq \underbrace{\sum_{j=1}^{q_t} p_{tj}\left(\underline{Q}_t^{i+1}(x_{t-1}^{ik}, \xi_t) - (\beta_{tkj}^i)^\top x_{t-1}^{ik}\right)}_{=:\alpha_{tk}^i} + \underbrace{\left(\sum_{j=1}^{q_t} p_{tj}\beta_{tkj}^i\right)}_{=:\beta_{tk}^i}^\top x_{t-1}, \tag{3.4}
$$

where β_{tkj}^i denotes a subgradient of $\underline{Q}_t^{i+1}(\cdot, \xi_{tj})$ at x_{t-1}^{ik} for $k \in \mathcal{K}, j = 1, \ldots, q_t$. Again, by defining

$$
\phi_{tk}^i(x_{t-1}) := \alpha_{tk}^i + (\beta_{tk}^i)^\top x_{t-1},
$$

we can obtain a cut

$$
\mathcal{Q}_t(x_{t-1}) \geq \phi_{tk}^i(x_{t-1}) \tag{3.5}
$$

for each $k \in \mathcal{K}$ and can update the cut approximation to

$$\mathfrak{Q}_t^{i+1}(x_{t-1}) := \max\left\{\mathfrak{Q}_t^i(x_{t-1}),\ \phi_{t1}^i(x_{t-1}), \ldots, \phi_{t|\mathcal{K}|}^i(x_{t-1})\right\}. \tag{3.6}$$

Computing Subgradients. So far, we have discussed the main idea of the cut generation process in the backward pass of SDDP, which is based on evaluating approximate value functions $\underline{Q}_t^{i+1}(\cdot,\cdot)$ and using subgradients for them at trial points x_{t-1}^{ik}. For the interested reader, we now address in more detail how to compute those subgradients. This step uses *dual information*, *i.e.*, it is based on duality theory of linear programs. For simplicity, we assume $X_t = \{x_t \in \mathbb{R}^{n_t} \ : \ x_t \geq 0\}$ for all $t \in [T]$.

Consider stage T, some $k \in \mathcal{K}$ and some $j \in \{1, \ldots, q_T\}$. Then, the dual problem to the linear stage-T subproblem (2.10) is

$$\begin{cases} \max\limits_{\pi_T} & \left(h_{Tj} - T_{T-1,j} x_{T-1}^{ik}\right)^\top \pi_T \\ \text{s.t.} & W_{Tj}^\top \pi_T \leq c_{Tj}. \end{cases}$$

Let π_T^{ikj} be an optimal dual basic solution. Such solution does always exist by relatively complete recourse and boundedness (see Assumption 9 and Lemma 2.5). By strong duality of linear programs, it follows

$$\begin{aligned} \underline{Q}_T^{i+1}(x_{T-1}^{ik}, \xi_{Tj}) &= \left(h_{Tj} - T_{T-1,j} x_{T-1}^{ik}\right)^\top \pi_T^{ikj} \\ &= -(\pi_T^{ikj})^\top T_{T-1,j} x_{T-1}^{ik} + (\pi_T^{ikj})^\top h_{Tj}. \end{aligned}$$

Importantly, the dual feasible set does not depend on x_{T-1}, but remains unchanged for all trial points. In particular, π_T^{ikj} is always dual feasible, but not necessarily dual optimal for all x_{T-1}. Therefore, and because of minimization, it follows

$$\begin{aligned} \underline{Q}_T^{i+1}(x_{T-1}, \xi_{Tj}) &\geq -(\pi_T^{ikj})^\top T_{T-1,j} x_{T-1} + (\pi_T^{ikj})^\top h_{Tj} \\ &= -(\pi_T^{ikj})^\top T_{T-1,j}(x_{T-1} + x_{T-1}^{ik} - x_{T-1}^{ik}) + (\pi_T^{ikj})^\top h_{Tj} \\ &= \underline{Q}_T^{i+1}(x_{T-1}^{ik}, \xi_{Tj}) - (\pi_T^{ikj})^\top T_{T-1,j}(x_{T-1} - x_{T-1}^{ik}). \end{aligned}$$

Hence,

$$\beta_{Tkj}^i = -(\pi_T^{ikj})^\top T_{T-1,j}$$

is a subgradient of $\underline{Q}_T^{i+1}(\cdot, \xi_{Tj})$ at x_{T-1}^{ik}.

The previous derivation provides some additional insight. Since the dual feasible set is polyhedral and does not depend on x_{T-1}, for each noise term $\xi_{Tj}, j = 1, \ldots, q_T$, there exist only finitely many dual extreme points (dual basic solutions) that can be attained. Therefore, only finitely many different cut coefficients can be generated. This is crucial for some convergence proofs of SDDP, as we discuss later.

For earlier stages $t = T-1, \ldots, 2$, the dual problem to subproblem (2.10) looks a bit more sophisticated, as the cut approximations $\mathfrak{Q}_{t+1}^{i+1}(\cdot)$ have to be taken into account, which requires additional dual multipliers ρ_t^r for all cuts $r \in \Gamma_{t+1}$, where Γ_{t+1} denotes the index set of cuts generated for the following stage. However, the derivation is completely analogous and, again, we arrive at

$$\underline{Q}_t^{i+1}(x_{t-1}, \xi_{tj}) \geq \underline{Q}_t^{i+1}(x_{t-1}^{ik}, \xi_{tj}) - (\pi_t^{ikj})^\top T_{t-1,j}(x_{t-1} - x_{t-1}^{ik}),$$

so that

$$\beta^i_{tkj} = -(\pi^{ikj}_t)^\top T_{t-1,j} \tag{3.7}$$

is a subgradient of $\underline{Q}^{i+1}_t(\cdot, \xi_{tj})$ at x^{ik}_{t-1}. Interestingly, the optimal dual multipliers ρ^{rikj}_t are not explicitly required in this formula.

3.4. Bounds and Stopping. At the first stage, the subproblem

$$\underline{v}^i := \min_{x_1 \in \mathcal{X}_1(x_0)} c_1^\top x_1 + \mathfrak{Q}^{i+1}_2(x_1). \tag{3.8}$$

is solved. As $\mathfrak{Q}^{i+1}_2(\cdot)$ is a lower approximation of $\mathcal{Q}_2(\cdot)$, $\underline{v}^i$ is a valid lower bound to the optimal value v^* of (MSLP). This bound can be initialized with $\underline{v}^0 = -\infty$ or any a priori known lower bound for v^*.

In contrast, we are not guaranteed to obtain a valid upper bound for v^* during iterations of standard SDDP, as we only consider a small subset $\mathcal{K} \subseteq \mathcal{S}$ of all scenarios. This means that in the forward pass, the feasible policy for (MSLP), which is implicitly defined by the current cut approximations $\mathfrak{Q}^i_t(\cdot), t = 2, \dots, T$, is only evaluated for a subset of all scenarios. By evaluating these scenarios in the objective of (MSLP) and taking the sample average

$$\overline{v}^i_{\mathcal{K}} := \frac{1}{|\mathcal{K}|} \sum_{k \in \mathcal{K}} \underbrace{\sum_{t=1}^{T} \big(c_t(\xi^k_t)\big)^\top x^{ik}_t}_{=: v^i(\xi^k)} \tag{3.9}$$

we only obtain an unbiased estimator of the true upper bound $\overline{v}^i$ (a *statistical upper bound*) associated with the current policy, see Section 7 for more details.

After each iteration of SDDP, one or several stopping criteria are checked, which may or may not be based on $\overline{v}^i_{\mathcal{K}}$. We discuss different stopping criteria in detail in Section 7. If SDDP does not stop, a new iteration $i+1$ is started with a forward pass.

It is worth mentioning that the first-stage subproblems (3.1) and (3.8) are the same for consecutive backward and forward passes, and in principle only have to be solved once. The same is true for consecutive forward and backward pass problems at the final stage T.

3.5. Cut Properties. We discuss convergence of SDDP in Section 4. It relies on three key properties of the derived cuts:

LEMMA 3.2. *For any stage $t = 2, \dots, T$ and any $k \in \mathcal{K}$, the functions $\phi^i_{tk}(\cdot)$ are*
(a) valid *lower approximations of $\mathcal{Q}_t(\cdot)$,*
(b) tight *for $\underline{\mathcal{Q}}^{i+1}_t(\cdot)$ (as defined in (2.11)) at x^{ik}_{t-1},*
(c) finite*, i.e., only finitely many different cuts can be generated, if we restrict to dual basic solutions to generate cuts.*

Proof. Property (a) follows immediately from (3.3) and (3.5). (b) holds because of strong duality for linear programs and taking expected values over the obtained optimal values. Alternatively, we can rearrange the RHS of inequality (3.4) to obtain

$$\phi^i_{tk}(x_{t-1}) = \underline{\mathcal{Q}}^{i+1}_t(x^{ik}_{t-1}) + \sum_{j=1}^{q_t} p_{tj} (\beta^i_{tkj})^\top (x_{t-1} - x^{ik}_{t-1}). \tag{3.10}$$

Inserting x^{ik}_{t-1} yields $\phi^i_{tk}(x^{ik}_{t-1}) = \underline{\mathcal{Q}}^{i+1}_t(x^{ik}_{t-1})$.

Property (c) follows by induction using the arguments on the dual feasible region previously discussed for stage T. ∎

Note that $\phi^i_{tk}(\cdot)$ is not necessarily tight for the true expected value function $\mathcal{Q}_t(\cdot)$ in early iterations for $t \neq T$, but rather might provide a loose cut only. However, by the finiteness and tightness properties it can be shown recursively, that eventually the derived cuts become tight for $\mathcal{Q}_t(\cdot)$ as well. In fact, after finitely many steps, the polyhedral function $\mathcal{Q}_t(\cdot)$ is represented exactly for all $t = 2, \ldots, T$. This is a key property for the convergence of SDDP.

3.6. Illustrative Example. To illustrate the key steps of SDDP, we present a simple example.

EXAMPLE 3.3. *Consider the 3-stage (MSLP)*

$$\begin{aligned} \min \quad & x_1 + x_2 + x_{31} + x_{32} \\ \text{s.t.} \quad & x_1 \leq 6 \\ & x_2 \geq \xi_2 - x_1 \\ & x_{31} - x_{32} = \xi_3 - x_2 \\ & x_1, x_2, x_{31}, x_{32} \geq 0, \end{aligned} \tag{3.11}$$

which is inspired by Example 2 in Chapter 5 of [26]. The uncertain data in the RHS is stagewise independent and uniformly distributed with $\xi_2 \in \{4, 5, 6\}$ *and* $\xi_3 \in \{1, 2, 4\}$.

Problem (3.11) *has not entirely the same structure as problem (MSLP), but can be easily converted to it by introducing slack variables. However, for illustrative purposes, we abstain from this. The problem can be expressed by means of the value functions*

$$Q_3(x_2, \xi_3) = \begin{cases} \min\limits_{x_3} & x_{31} + x_{32} \\ \text{s.t.} & x_{31} - x_{32} = \xi_3 - x_2 \\ & x_{31}, x_{32} \geq 0 \end{cases} \tag{3.12}$$

and

$$Q_2(x_1, \xi_2) = \begin{cases} \min\limits_{x_2} & x_2 + \mathcal{Q}_3(x_2, \xi_2) \\ \text{s.t.} & x_2 \geq \xi_2 - x_1 \\ & x_2 \geq 0. \end{cases}$$

The first-stage problem then is

$$v^* = \begin{cases} \min\limits_{x_1} & x_1 + \mathcal{Q}_2(x_1) \\ s.t. & x_1 \in [0, 6]. \end{cases}$$

The optimal solution is given by $x_1^* = 3$ *with* $v^* = \frac{53}{9}$.

As shown in [26], the stage-3 value functions can be written in closed-form as $Q_3(x_2, \xi_2, \xi_3) = |\xi_3 - x_2|$ *for all scenarios. Taking expectations, a closed-form expres-*

sion for $\mathcal{Q}_3(\cdot)$ can be derived, and by recursion we obtain

$$\mathcal{Q}_2(x_1) = \begin{cases} \frac{23}{3} - \frac{16}{9}x_1, & x_1 \in [0,1] \\ \frac{67}{9} - \frac{10}{9}x_1, & x_1 \in [1,2] \\ \frac{59}{9} - \frac{10}{9}x_1, & x_1 \in [2,3] \\ \frac{47}{9} - \frac{2}{3}x_1, & x_1 \in [3,4] \\ \frac{31}{9} - \frac{2}{9}x_1, & x_1 \in [4,5] \\ \frac{7}{3}, & x_1 \in [5,6]. \end{cases}$$

The optimal value is $v^* = \frac{56}{9}$.

We apply SDDP for illustration. We assume loose initial bounds $\theta_2, \theta_3 \geq -10$ for simplicity. In the forward pass, we sample one scenario path per iteration, i.e., $|\mathcal{K}| = 1$. In iteration 1, let $(\xi_2, \xi_3) = (5, 4)$ define this path. Solving the approximate subproblems (2.10) for all stages $t = 1, 2, 3$ and $(\xi_2, \xi_3) = (5, 4)$, we obtain $\overline{v}^i_{\mathcal{K}} = 6$. In fact, this is no valid upper bound for v^*.

In the backward pass, cuts for $\mathcal{Q}_t(\cdot), t = 2, 3$, are derived at the trial points. For stage 3, the cut gradient is $\beta_3(4) = 1$. Moreover, $\underline{\mathcal{Q}}^2_3(4) = \frac{5}{3}$. With formulas (3.5) and (3.10) this yields the cut $\mathcal{Q}_3(x_2) \geq -\frac{7}{3} + x_2$, which is incorporated into the stage-2 subproblems. Solving these problems yields the cut $\mathcal{Q}_t(x_1) \geq \frac{23}{3} - 2x_1$. At the first stage, the lower bound computes to $\underline{v}^1 = \frac{5}{3}$.

The expected value functions and the obtained cuts for three iterations are depicted in Figure 5. In the second and the third iteration, the same scenario path $(\xi_2, \xi_3) = (6, 1)$ is sampled in the forward pass.

Figure 6 displays the bounds $\underline{v}^i$ and $\overline{v}^i_{\mathcal{K}}$ for ten iterations of SDDP. It shows that the lower bounds stabilize quickly at v^*, whereas the values of $\overline{v}^i_{\mathcal{K}}$ oscillate around v^*.

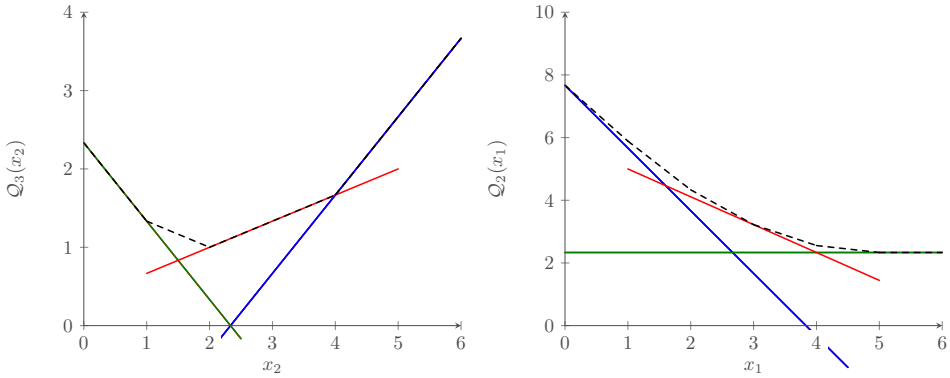


Fig. 5: Expected value functions for Example 3.3 with cuts obtained in first three iterations depicted in blue, green and red.

3.7. Policy Assessment. As mentioned before, in standard SDDP no valid upper bound $\overline{v}$ for v^* is determined. While in each iteration a statistical upper

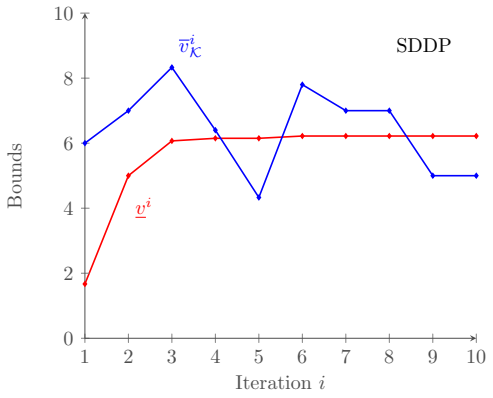


Fig. 6: Bounds for 10 iterations of SDDP applied to Example 3.3.

bound (3.9) can be computed, the number of samples $|\mathcal{K}|$ may often be too small to appropriately assess the quality of the current policy. In particular, $|\mathcal{K}|$ is often chosen to be 1 in practice, and thus $\overline{v}^i_{\mathcal{K}}$ is not a meaningful estimate for $\overline{v}$.

Therefore, to assess the obtained policy, usually an additional forward simulation is conducted once SDDP has terminated. For this simulation a much higher number of sample paths through the scenario tree is used, *e.g.* $|\mathcal{K}| \in \{1000, 10000\}$, leading to a reasonable estimator $\overline{v}_{\mathcal{K}}$. In this step, the simulation can be either performed *in-sample* (using sample paths through the recombining scenario tree) or *out-of-sample* (using the true underlying distribution, *e.g.*, if $\boldsymbol{\xi}_t$ is a continuous random variable that is discretized to satisfy Assumption 5, see Section 11).

Remark 3.4. In the light of Remark 3.1 this policy assessment step can also be interpreted from a statistical learning perspective. After the model has been trained, a model validation (using in-sample data) or a model test (using out-of-sample data) are performed.

4. Convergence and Complexity. The convergence behavior of SDDP has been thoroughly analyzed over the years. We discuss the main convergence results in this section. We first focus on *finite* convergence of SDDP, and then afterwards discuss the actual convergence *rate*, *i.e.*, the computational complexity of SDDP. Our overview is loosely based on the review chapter in [71].

4.1. Finite Convergence. The first convergence analyses related to SDDP have been conducted in [38] and [124], however implicitly assuming independence of sampled random variables and convergent subsequences of algorithm iterates. A first complete convergence proof is given by Philpott and Guan in [163] for the case where uncertainty only enters the RHS of (MSLP) (in fact, they consider a more general algorithm than SDDP, including sampling in the backward pass). The same reasoning is used by Shapiro [197] for the case where also W_t, c_t and T_{t-1} are uncertain.

The convergence behavior of SDDP can be explained using two main arguments: First, as stated in Lemma 3.2, only finitely many different cuts, and by that only finitely many different cut approximations $\mathfrak{Q}_t(\cdot)$ can be constructed for all $t = 2, \dots, T$. This result requires linearity (Assumption 6) and finite random variables (Assumption 5). Moreover, these finitely many cuts also satisfy some tightness property, which implies that they are sufficient to exactly represent the polyhedral (expected) value

functions (see Theorem 2.8). For a deterministic algorithm, this would result in finite convergence to the true optimal point and value (see the convergence properties of Benders decomposition [17] and Kelley's cutting-plane method [110]).

For SDDP, it has to be taken into account that scenarios are sampled in the forward pass. This means that the cut approximations might not further improve for some iterations if the *wrong* scenarios are sampled. Therefore, the second key argument for many proofs of finite convergence of SDDP is that each scenario is visited infinitely many times with probability 1 given that the algorithm does not terminate. Intuitively, this means that after finitely many iterations the *right* scenarios will be sampled with probability 1, leading to the construction of a new cut. This requirement is satisfied under *independent sampling*, that is, if the sampling in the forward pass of Algorithm 3.1 is random and independent of previous iterations. It is also satisfied for an exhaustive enumeration of all scenarios in the sampling process.

Using these two arguments, the following main convergence result can be obtained

THEOREM 4.1 (Almost sure finite convergence of SDDP). *Under Assumptions 1 to 9 and using an independent random sampling procedure in the forward pass, SDDP converges with probability 1 to an optimal policy of (MSLP) in a finite number of iterations.*

Importantly, almost sure finite convergence to an optimal policy of (MSLP) does not imply that the trajectories $(x_t^{ik})_{t\in[T]}, k \in \mathcal{K}$, and the corresponding sample averages $\overline{v}_{\mathcal{K}}^i$ obtained in SDDP converge, as both are random and depend on the current sample $\mathcal{K}$. However, the lower bounds $\underline{v}^i$ obtained in SDDP converge to v^*.

Deterministic Sampling. Recently, convergence analyses of SDDP and related algorithms have often made use of deterministic sampling techniques instead of random sampling [10, 11]. Here, the idea is that the approximation error in SDDP can be controlled and guided to zero in a deterministic way if in each iteration scenarios are sampled for which the current approximation gap is maximized. This requires, however, that the approximation gap itself can be bounded rigorously. Therefore, in addition to the lower cut approximation $\mathfrak{Q}_t(\cdot)$ also an upper approximation $\overline{\mathfrak{Q}}_t(\cdot)$ is constructed and iteratively refined [10, 229], so that deterministic lower bounds $\underline{v}^i$ and upper bounds $\overline{v}^i$ are computed in each iteration. For more details on deterministic sampling and deterministic upper bounds we refer to Sections 6 and 8.

Generalizations. It has been shown that some of the basic assumptions (Assumptions 1 to 9) can be relaxed without compromising convergence of SDDP. Girardeau et al. [78] analyze the case where SDDP is applied to multistage problems with nonlinear convex subproblems, *i.e.*, Assumption 6 is relaxed. In this case, the value functions $Q_t(\cdot)$ are no longer polyhedral, but still convex. The authors show that almost sure convergence is still satisfied as long as some convexity and compactness assumptions and some tightened recourse assumption are satisfied. We discuss this result in detail in Section 15 when we formally introduce convex multistage stochastic nonlinear problems. The main idea is that even without polyhedrality, $Q_t(\cdot)$ can be guaranteed to be Lipschitz continuous, so that the approximations of $\mathcal{Q}_t(\cdot)$ get better in a whole neighborhood of the trajectories $(x_t^{ik})_{t\in[T]}, k \in \mathcal{K}$.

Guigues generalizes this convergence result to the risk-averse case where Assumption 8 is relaxed [85]. Forcier and Leclère provide convergence results for a generalized framework of SDDP-related algorithms, including SDDP with inexact cuts or regularization (see also Section 21), risk-averse SDDP (see also Section 12) and extensions to convex nonlinear or non-convex mixed-integer (nonlinear) problems (see also Sections 15 and 16). Moreover, they prove convergence for (MSLP) without finite ran-

domness, *i.e.*, dropping Assumption 5. Further convergence proofs are provided for multi-cut SDDP [8], SDDP with cut selection [8, 87], adaptive partition-based SDDP [205] (see also Section 21), using SDDP with saddle cuts [55] (see also Section 14) and variants of distributionally robust SDDP [65, 161] (see also Section 13), Another proof of almost sure finite convergence for extensions to non-convex problems is provided in [229].

4.2. Complexity. Theorem 4.1 guarantees almost sure finite convergence of SDDP. While this result is of theoretical interest, it may not be very relevant in practical applications, as it provides no result on the rate of convergence. As pointed out in [71], especially the argument of scenarios being sampled repeatedly (infinitely many times) is almost never applicable to SDDP in practice due to the sheer amount of scenarios in $\mathcal{S}$. Important for the rate of convergence are the computational cost per iteration and the required number of iterations.

Cost per Iteration. For the computational cost per iteration, the number of LPs to be solved in the backward pass is crucial. Per sample $k \in \mathcal{K}$ in the forward pass, q_t subproblems are solved for each stage except for $t = 1$ in the backward pass. Therefore, the total number of LPs solved is $1 + |\mathcal{K}| \sum_{t=2}^{T} q_t$. Hence, the number of problems to be solved grows linearly in the number of stages T, in the number of samples $|\mathcal{K}|$ and in the number of noise terms q_t [178].

Expected Number of Iterations. The computational bottleneck for SDDP is the expected required number of iterations to achieve convergence. Recently, there has been active research on computing theoretical bounds on this number, with Lan [116] as well as Zhang and Sun [229] publishing similar results using slightly different approaches. In both cases, the authors start by considering some case of deterministic sampling (in [116] the associated algorithm is referred to as *explorative dual dynamic programming* (EDDP)) before enhancing their results to the random sampling variant of SDDP. We discuss deterministic sampling in more detail in Section 6. The main idea to derive iteration bounds is the following: By exploiting Lipschitz continuity of $\mathfrak{Q}_t(\cdot)$ and $\mathcal{Q}_t(\cdot)$, it is possible to control the approximation error also at points where no cuts are constructed, as long as they lie in a neighborhood of some trial point x_t^{ik}. As long as the state space is bounded for all $t \in [T]$ (cf. Assumption 9), it can be completely covered by finitely many such neighborhoods [229]. A similar reasoning is applied in [71].

More formally, Lan [116] introduces the notion of *saturated points* $\bar{x}_{t-1}$, in which the approximation of $\mathcal{Q}_t(\cdot)$ is already ε-close for some predefined tolerance $\varepsilon > 0$, *i.e.*,

$$\mathcal{Q}_t(\bar{x}_{t-1}) - \mathfrak{Q}_t^i(\bar{x}_{t-1}) \leq \varepsilon,$$

and *distinguishable points* $\bar{x}_{t-1}$, which have at least a δ-distance to the set X_{t-1}^{sat} of already saturated points for some $\delta > 0$, that is

$$\|\bar{x}_{t-1} - x_{t-1}\| > \delta, \quad \forall x_{t-1} \in X_{t-1}^{sat}.$$

If some trial point x_t^{ik} is saturated and distinguishable, the iteration i can be called *effective* [71]. Using deterministic sampling, all iterations in SDDP can be shown to be effective, and thus the number of iterations can be bounded in the aforementioned way. For random sampling, this is not true, but the probability for an effective iteration is at least $\frac{1}{N}$ with $N := \Pi_{t=2}^{T-1} n_t$.

In the light of Assumption 9 (b), for any $t \in [T]$, we call the bound D_t satisfying

$$\|x_t - x_t'\| \leq D_t, \ \forall x_t, x_t' \in \mathcal{X}_t$$

the *diameter* of the state space. Additionally, let L denote a Lipschitz constant for the objective function of (MSLP), which exists due to Corollary 2.10.

Then, the following complexity results are satisfied by SDDP.

THEOREM 4.2 (Complexity of SDDP [116, 229]). *Let $D_t \leq D$ for all $t \in [T]$. For some arbitrary $\varepsilon > 0$, the (expected) number of required iterations of SDDP (Algorithm 3.1) to obtain*

- *an ε-optimal solution using deterministic sampling is*
  - *polynomial in $T, (\frac{1}{\varepsilon}), L$ and D,*
  - *exponential in n_t,*
- *a $(T\varepsilon)$-optimal solution using deterministic sampling is*
  - *linear in T,*
  - *linear in T and $\overline{X}$ if $\mathcal{X}_t$ is finite with cardinality $|\mathcal{X}_t| \leq \overline{X}$,*
- *an ε-optimal solution using random sampling is*
  - *polynomial in $q_t, T, (\frac{1}{\varepsilon}), L$ and D,*
  - *exponential in T and n_t.*

This means that for standard SDDP (using random sampling) the expected number of iterations grows exponentially in the horizon T and the dimension n_t of the state space. This is computationally important. The exponential complexity with respect to the state dimension is not that surprising, as it is well-known for cutting-plane methods [145] and inherited by SDDP. Similarly, the exponential complexity with respect to the number of stages directly follows from the exponential number of scenarios that may have to be sampled in the worst-case. Interestingly, under deterministic sampling, the complexity is independent of the number q_t of noise terms per stage, as this number only affects the computational cost per iteration.

We see that using some deterministic sampling scheme a polynomial or even linear iteration complexity in T can be achieved, whereas the iteration complexity in the state space cannot be alleviated [229].

5. Comparison with Related Methods. We briefly compare SDDP to solution methods that it is (historically) related to, as discussed in Section 1.

5.1. Relation to SDP. SDDP is closely related to stochastic dynamic programming (SDP). SDP usually is applied in a setting where not only state variables, but additional local variables are considered, see Remarks 2.2 and 2.4. Therefore, the DPE and value functions are considered in the form of (2.7), which we repeat here for convenience:

$$Q_t(x_{t-1}, \xi_t) = \min_{u_t \in U_t(x_{t-1}, \xi_t)} f_t(u_t, \xi_t) + \mathcal{Q}_{t+1}(\mathfrak{T}_t(x_{t-1}, u_t, \xi_t)).$$

The main idea of SDP is to explicitly evaluate the (expected) value functions for all possible cases during a forward or backward iteration through the stages $t \in [T]$. This is only possible if the support Ξ_t of $\boldsymbol{\xi}_t$ and the state space $\mathcal{X}_t \subset X_t$ are finite for all $t \in [T]$. Otherwise, infinitely many evaluations would be required. Additionally, it is required that also the action space $U_t(x_{t-1}, \xi_t)$ is finite for all $x_{t-1} \in X_{t-1}, \xi_t \in \Xi_t$, so that the minimum in (2.7) can be computed by finitely many evaluations. For this reason, all these sets may have to be discretized first [168].

The computational effort of SDP scales linearly in T and in the cardinalities $|X_t|, |U_t(x_{t-1}, \xi_t)|$ and $|\Xi_t|$. The three sets might be multidimensional, and thus require to be discretized in each dimension d_{t-1}, κ_t and $\tilde{d}_t$. Hence, their cardinality grows exponentially in these dimensions, which is computationally prohibitive for

high-dimensional problems. This is known as the curse of dimensionality of SDP, see also Section 1.

SDDP avoids the requirements of state space and action space discretization by not evaluating $\mathcal{Q}_t(\cdot), t \in [T]$, exactly for all (finitely many) possible actions and states, but approximating them by an iteratively refined polyhedral outer approximation $\mathfrak{Q}_t(\cdot)$, constructed by linear cuts. It can thus be considered an *approximate dynamic programming* (ADP) method.

5.2. Relation to NBD. In stochastic programming, it is common practice to consider problems (MSLP) with finite randomness (Assumption 5), but without the requirement of stagewise independence of $\boldsymbol{\xi}_t$ (Assumption 2). In that case the uncertainty can be modeled by a finite scenario tree, which compared to the recombining tree from Section 2 exhibits some path dependence and satisfies the usual *tree* property that each node n has a finite set of child nodes $\mathcal{C}(n)$, but a unique parent node $a(n)$. An example of a scenario tree with $T = 3$ and $|\mathcal{S}| = 9$ is illustrated in Figure 7. This scenario tree represents the same number of scenarios $|\mathcal{S}|$ as the recombining one in Figure 1, but requires $\sum_{t=2}^{T} q_t^{t-1} + 1$ instead of $\sum_{t=2}^{T} q_t + 1$ nodes.

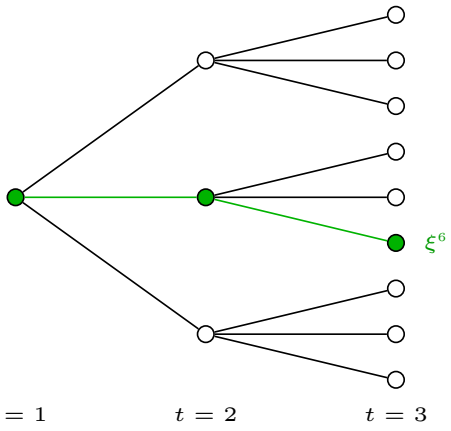


Fig. 7: Scenario tree with 9 scenarios and ξ^6 highlighted.

To solve (MSLP) associated with a general scenario tree, in principle the same approach as in SDDP can be used to approximate $\mathcal{Q}_t(\cdot)$ with linear cuts. However, due to the path dependence, the value functions $Q_t(\cdot)$ and expected value function $\mathcal{Q}_t(\cdot)$ depend on the history $\xi_{[t-1]}$ of the data process $(\boldsymbol{\xi}_t)_{t\in[T]}$. In other words, each node n has its own value function $Q_n(\cdot)$, and with each node (except for leaf nodes) is associated an expected value function $\mathcal{Q}_{\mathcal{C}(n)}(\cdot)$. Therefore, to update the approximations $\mathfrak{Q}^i_{\mathcal{C}(n)}(\cdot)$ of all $\mathcal{Q}_{\mathcal{C}(n)}(\cdot)$ in each iteration, all nodal subproblems have to be solved in the backward pass, which in turn requires to compute trial points $x^i_{a(n)}$ for all nodes, *i.e.*, solving all nodal subproblems in the forward pass as well.

Because of its close relation to the L-shaped method for solving two-stage stochastic linear programs [226] and to Benders decomposition [17] this solution method is called *nested Benders decomposition* (NBD) or just *nested decomposition*. It was first proposed by Birge in 1985 [25] and can be interpreted as a decomposition method for the extensive form of the deterministic equivalent of (MSLP). Contrary to SDDP, NBD guarantees that valid lower bounds $\underline{v}$ and upper bounds $\overline{v}$ of v^* are determined

in each iteration and by that allows for a deterministic stopping criterion in a straightforward way. The upper bounds can be computed as

$$\overline{v}^i := \mathbb{E}\left[\sum_{n\in\mathcal{T}} c_n x_n^i\right],$$

where $\mathcal{T}$ is the set of all nodes in the scenario tree.

On the other hand, due to the sheer amount of subproblems to be solved in each iteration, which grows exponentially in T, it is computationally tractable for problems of moderate size. By moderate we mean instances with some hundreds or a few thousand scenarios, and 4 or 5 stages at maximum [225].

Note that for general scenario trees also sampling scenarios from $\mathcal{S}$ in the forward pass does not necessarily help to reduce the computational burden and to speed-up the solution process, as it reduces the computational effort per iteration, but at the same time implicates that the cut approximations $\mathfrak{Q}^i_{\mathcal{C}(n)}(\cdot)$ are only improved for some $\mathcal{Q}_{\mathcal{C}(n)}(\cdot)$ in each iteration.

Under stagewise independence (Assumption 2) this is different. The scenario tree collapses to a recombining tree, and there exists only one expected value function $\mathcal{Q}_t(\cdot)$ for each $t = 2, \dots, T$. If now only a sample $\mathcal{K} \subset \mathcal{S}$ of scenarios is considered in each iteration i, as in SDDP, still the cut approximations $\mathfrak{Q}^i_t(\cdot)$ for all $\mathcal{Q}_t(\cdot)$ are updated with new cuts. The key difference is that in the stagewise independent case, for some stage t differing scenarios still share the same nodes, and thus value functions in the recombining scenario tree.

From this perspective, SDDP can be interpreted as a sampling variant of NBD which reduces the computational effort per iteration significantly [178], but heavily relies on stagewise independence of $(\boldsymbol{\xi}_t)_{t\in[T]}$ in order to leverage the sampling with respect to value function approximations.

Remark 5.1 (Cut-sharing). In the literature, the previous property of SDDP is often referred to as *cut-sharing*. The idea behind this phrase is the following. Consider a stagewise independent data process $(\boldsymbol{\xi}_t)_{t\in[T]}$. Even if not being the best possible representation, this process can be modeled using a classical finite scenario tree instead of a recombining one – hiding the fact that there exists no path dependence. In each iteration of SDDP now only a sample $\mathcal{K}$ of scenarios is considered, *i.e.*, only a subset of nodes from $\mathcal{T}$ is visited. Nonetheless, the constructed cuts are valid for all scenarios or nodes, respectively, no matter if they have been visited or not. This leads to the impression that despite being generated using a specific scenario, cuts can be *shared* with other scenarios.

Arguably, the phrase can be considered misleading, though, as it is based on a classical scenario tree perspective and may evoke the connotation that cuts are actually shared between different approximations $\mathfrak{Q}^i_{\mathcal{C}(n)}(\cdot)$. However, the crucial point is that due to stagewise independence, there is only *one* function $\mathcal{Q}_t(\cdot)$ to be approximated for each stage, with no need to share cuts. Cuts are shared between scenarios because these scenarios *share nodes* in the recombining scenario tree.

5.3. Complexity Comparison. We summarize the main complexity results for SDDP and the related methods in Table 3.

In contrast to SDP, SDDP does not require a state space and action space discretization. Especially, the later is computationally important in practice, while the former at least does not translate into an improvement of the worst-case complexity class. On the other hand, SDDP does not have linear complexity in T.

Table 3: Complexity of SDDP and related solution methods.

	Det. Equiv.	NBD	SDDP	SDP
Requirements				
stagewise independence	no	no	yes	yes*
state discretization	no	no	no	yes
action discretization	no	no	no	yes
Complexity				
in T	exponential	exponential	exponential	linear
in n_t	linear	linear	exponential	exponential
in q_t	polynomial	polynomial	polynomial	polynomial

* Markovian uncertainty is possible as well.

Compared to NBD, SDDP mainly reduces the computational effort per iteration significantly, but does not get rid of the exponential growth of the computational cost with respect to T. In return, it heavily relies on stagewise independence (Assumption 2) and has worse complexity with respect to the state dimension n_t.

We can conclude that SDDP, while mitigating some of the weaknesses of SDP and NBD (sometimes advertized as "breaking the curse of dimensionality"), does not manage to leave the respective worst-case complexity classes. On the contrary, it inherits some of the complexity drawbacks of both methods. Still, in many applications (where not worst-case complexity is decisive) it shows considerable performance improvements compared to SDP and NBD, especially for problems with continuous action space, a medium number of stages T and a moderate state dimension n_t. While Theorem 4.2 indicates that convergence may take extremely long in large-scale applications, and too long to be computationally tractable, SDDP has shown good performance for large-scale instances of (MSLP) in many applications, as we discuss in Section 9. This is also due to various improvements, which we address in the following sections.

6. Sampling. Sampling is a central element of SDDP, see Section 3. In the forward pass, a finite number $|\mathcal{K}|$ of scenarios is sampled to simulate the current policy and compute a trajectory of trial points $(x_t^{ik})_{t\in[T]}$ for all $k \in \mathcal{K}$. Often, this sampling is done from a finite set of scenarios $\mathcal{S}$ (see Assumption 5), with $|\mathcal{K}| \ll |\mathcal{S}|$. Alternatively, it is possible to directly sample from a given (continuous) distribution.

In this section, we discuss different sampling techniques which can be used in SDDP. As indicated in Sections 3 and 4, we can distinguish between *random sampling* and *deterministic sampling* methods. In standard SDDP, as originally proposed in [151], random sampling is used. For random sampling, the main requirement is that the samples should be independent and identically distributed (i.i.d.). This is important for two reasons:

(1) This way, almost sure finite convergence of SDDP can be ensured, as any scenario is sampled infinitely many times with probability 1, assuming that the algorithm does not terminate, see Section 4.

(2) In the originally proposed stopping criterion of SDDP a confidence interval is used, which is built using the sample mean $\overline{v}_{\mathcal{K}}^i$ (3.9), see Section 7. However, by the Central Limit Theorem, even an approximate confidence interval can

only be obtained for a sequence of i.i.d. random variables.

6.1. Monte Carlo Sampling. The simplest sampling method satisfying the above requirement is Monte Carlo (MC) sampling. Here, samples are drawn randomly from the probability distribution of $\boldsymbol{\xi}_t$ in each iteration, by first sampling from a uniform distribution and then using appropriate transforms. Under stagewise independence (Assumption 2), this is done independently for each stage $t \in [T]$.

As the quantities $v^i(\xi^k)$ are i.i.d., the value $\overline{v}^i_{\mathcal{K}}$ (3.9) that can be computed in the SDDP forward pass is an unbiased estimator of $\overline{v}^i$ and according to the Strong Law of Large Numbers converges to $\overline{v}^i$ for $|\mathcal{K}|$ approaching infinity. Still, the sampling error can be significant. The variance of $\overline{v}^i_{\mathcal{K}}$ can be estimated by $\frac{1}{|\mathcal{K}|}\left(\sigma^i_{\overline{v},\mathcal{K}}\right)^2$. This means that the variance can be reduced either by increasing the number of samples $|\mathcal{K}|$ or by reducing the sample variance $\left(\sigma^i_{\overline{v},\mathcal{K}}\right)^2$. Increasing the sample size may look promising at first glance, but may become computationally intractable in practice [149]. Recall that for every sample $k \in \mathcal{K}$ a number of $1 + \sum_{t=2}^{T} q_t$ subproblems has to be solved in the backward pass of each iteration. Therefore, the more promising approach is combining MC sampling with variance reduction techniques [149].

6.2. Variance Reduction Techniques. Incorporating variance reduction techniques into sampling in SDDP is studied extensively in [104, 149]. For a review on sampling techniques in stochastic programming in general, we refer to [103].

Randomized QMC Sampling. In [104], it is proposed to use Quasi-Monte Carlo (QMC) sampling within SDDP. In this case, instead of randomly sampling from the uniform distribution, a deterministic sequence of points $u^1, \dots, u^N$ from $(0,1)^{\kappa_t}$ is chosen. This is done in such a way that the sampled points fill $(0,1)^{\kappa_t}$ as homogeneously as possible (so the empirical distribution is as close to a uniform distribution as possible). Then, after an appropriate transformation, they provide a better representation of $\boldsymbol{\xi}_t$ than randomly sampled points.

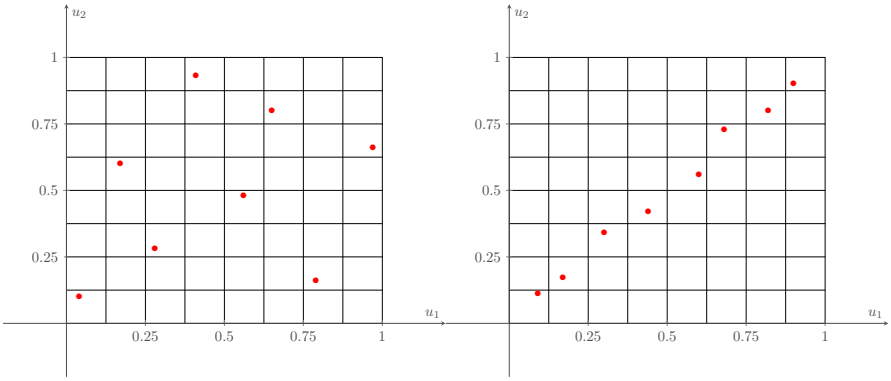
A drawback of QMC methods is that the sample points are not random, the obtained estimator is biased and no confidence interval can be established. *Randomized* QMC (RQMC) methods, where the choice of QMC points is combined with some kind of randomness, avoid this drawback and allow for standard error estimation [104].

Compared to MC sampling, RQMC methods achieve better convergence rates of $\mathcal{O}\big(|\mathcal{K}|^{-1}(\log|\mathcal{K}|)^{\kappa_t}\big)$, and thus are considered more efficient. However, the convergence rate depends on the dimension κ_t of $\boldsymbol{\xi}_t$ [104].

Latin Hypercube Sampling. In Latin Hypercube Sampling (LHS) [140], the space $(0,1)^{\kappa_t}$ is divided into equidistant subintervals and then scenarios are sampled from each subinterval in such a way that in each row and column of the grid only one point is sampled. This is illustrated in Figure 8 (a).

In this way, again, a more homogeneous distribution of the sample points can be obtained, and compared to MC sampling, the variance can be reduced. On the flipside, poor space-filling or correlation between the sample points has to be ruled out, see Figure 8 (b), which requires significant additional effort.

Incorporation into SDDP. It is important to notice that while reducing the variance compared to the classical MC estimators, scenarios sampled by RQMC and LHS are no longer i.i.d. Therefore, both sampling techniques cannot be incorporated into SDDP without modification, if convergence properties (and the stopping criterion) should not be compromised. Homem-de-Mello et al. therefore suggest to build sampling blocks [104]. This means that the total number of samples $|\mathcal{K}|$ is divided into M blocks $\ell = 1, \dots, M$ with $M \geq 5$ a divisor of $|\mathcal{K}|$. Then, for each block



(a) Good space-filling. (b) Correlated sample points.

Fig. 8: Latin Hypercube Sampling for two dimensions.

ℓ, $|\mathcal{K}'| := |\mathcal{K}|/M$ scenarios are obtained using conditional sampling with RQMC or LHS, which are not independent. For each $k' \in \mathcal{K}'$, values $v^i(\xi^{k'})$ are determined and averaged to $\overline{v}^{i,\ell}$.

This is repeated for each block ℓ. Then, the mean $\overline{v}^i_{\mathcal{K}}$ of all values $\overline{v}^{i,\ell}, \ell = 1, \ldots, M$, and the sample variance are determined. As the scenarios of different blocks are independent, this still yields a useful confidence interval to stop the algorithm.

Another challenge reported in [104] is that it is computationally expensive to generate samples using RQMC for high dimensions. To reduce the computational effort, it may be reasonable to apply RQMC only to important components, *e.g.*, to early stages in $[T]$, and standard MC or LHS to the other ones. This strategy is called *padding* and applied after 6 or 12 stages for numerical tests in [104].

Experiments in [104] imply that RQMC and LHS both lead to upper bounds $\overline{v}_{\mathcal{K}}$ oscillating around the lower bound $\underline{v}$ more quickly compared to MC sampling.

6.3. Importance Sampling. In [149], Parpas et al. propose incorporating importance sampling into SDDP. Importantly, in difference to the previously described techniques, it can be used to obtain i.i.d. samples in the forward pass.

The main idea of importance sampling in general is to attach different importance to subregions of the sample space and to sample more often from subregions of higher importance. In context of SDDP, this means that it is sampled with priority from scenarios that contribute more to the value of the expected value functions $\mathcal{Q}_t(\cdot)$.

This is achieved by sampling from a different distribution than the original one, the so-called *importance sampling distribution*, but correcting the bias introduced by this difference. Then, an importance sampling estimate of $\overline{v}$ can be calculated as

$$\overline{v}^{IS,i}_{\mathcal{K}} := \frac{1}{|\mathcal{K}|} \sum_{k \in \mathcal{K}} v^i(\xi^k) \Lambda(\xi^k)$$

with $\Lambda(\boldsymbol{\xi}) := \frac{f(\boldsymbol{\xi})}{g(\boldsymbol{\xi})}$, where f denotes the original distribution and g the importance sampling distribution. The likelihood function $\Lambda(\cdot)$ is used to correct for sampling from the wrong distribution. It can be shown that importance sampling can reduce the variance of sampling estimators significantly. In the SDDP case, as shown in [149],

the variance is minimized for choice

$$g_t^*(\xi_t) := \frac{|Q_t(x_{t-1}^{ik}, \xi_t)|}{\mathbb{E}_f|Q_t(x_{t-1}^{ik}, \boldsymbol{\xi}_t)|} f_t(\xi_t).$$

However, clearly, this *zero-variance distribution* is a theoretical construct and not known, which is referred to as the *curse of circularity*. Therefore, it is proposed to first approximate g^* using a framework including Kernel density estimation [149].

In numerical experiments, SDDP with importance sampling is shown to outperform MC and QMC sampling based methods, in case that it is difficult to sample from the original probability distribution and that the original problem has moderate or high variance [149].

6.4. Deterministic Sampling. As already discussed in Section 4, in step 6 of SDDP (Algorithm 3.1) also some deterministic sampling can be used. In this case, $|\mathcal{K}| = 1$. In the literature, two different approaches are considered.

Worst Approximation Sampling. The first one requires that in addition to the (lower) cut approximation $\mathfrak{Q}_t(\cdot)$ of $\mathcal{Q}_t(\cdot)$ also an upper approximation $\overline{\mathfrak{Q}}_t(\cdot)$ is constructed and iteratively refined in SDDP. Assume that in the forward pass on stage $t-1$ the trial point x_{t-1}^i has been computed. Then, for stage t the approximate subproblem (2.10) is solved for x_{t-1}^i and for all noise terms $\xi_{tj}, j = 1, \ldots, q_t$, yielding optimal states x_{tj}. For the next stage, the trial point $x_t^i = x_{tj'}$ is chosen such that

$$j' \in \underset{j=1,\ldots,q_t}{\arg\max} \left\{ \overline{\mathfrak{Q}}_t^i(x_{tj}) - \mathfrak{Q}_t^i(x_{tj}) \right\},$$

i.e., that the gap between the current upper and lower approximations is maximized. This corresponds to sampling noise term $\xi_{tj'}$ on stage t.

This form of deterministic sampling is used for SDDP in [229]. Its computational drawback is that at each stage q_t subproblems have to be solved instead of only $|\mathcal{K}| \ll q_t$. A similar approach was first proposed by Baucke et al. in [10, 11] and called *problem-child node selection*. However, their setting differs a bit from original SDDP, as each subproblem contains specific variables $x_{tj}, j = 1, \ldots, q_t$, for all random outcomes, and therefore in their case only one subproblem has to be solved in the sampling step. Another related sampling scheme is used in *robust dual dynamic programming* (RDDP) [76]. In that case, $\xi_{tj'}$ is determined by solving a special upper bounding problem containing $\overline{\mathfrak{Q}}_t^i(\cdot)$

Explorative Sampling. Explorative deterministic sampling is proposed in [116] as part of EDDP. It is based on the concepts of saturated and distinguishable points, which we introduced in Section 4.2. As for the previous sampling scheme, the idea is to solve the forward pass subproblems for all $\xi_{tj}, j = 1, \ldots, q_t$. Instead of maximizing an approximation gap, however, the trial point $x_t^i = x_{tj'}$ is chosen such that

$$j' \in \underset{j=1,\ldots,q_t}{\arg\max} \min_{x_t \in X_t^{sat}} \|x_{tj} - x_t\|,$$

i.e., the minimum distance to already saturated points is maximized. In other words, a maximum distinguishable point is chosen.

As shown in [71], worst approximation sampling and explorative sampling are equivalent in the sense that both approaches are guaranteed to lead to effective iterations, see Section 4.2.

7. Stopping Criteria. In each iteration i of SDDP, a valid lower bound $\underline{v}^i$ for the optimal value v^* is determined. Additionally, a statistical upper bound $\overline{v}^i_{\mathcal{K}}$ can be computed. Since the latter is not necessarily valid, an important question is when to consider an obtained policy $(\boldsymbol{x}_t(\xi_{[t]}))_{t\in[T]}$ as (approximately) *optimal* and to stop the SDDP method. If the stopping criterion is too conservative, the algorithm may iterate much longer than required, if it is too optimistic, then SDDP may stop prematurely.

Confidence Stopping Criteria. In their seminal work on SDDP, Pereira and Pinto propose to use a confidence interval based stopping criterion [152]. An approximate confidence interval for a true valid upper bound $\overline{v}^i$ is determined as follows using the estimates $v^i(\xi^k)$ from (3.9).

Under random independent sampling, the values $v^i(\xi^k)$ are i.i.d. random variables with expected value $\overline{v}^i$ and variance $(\sigma^i)^2$. Moreover, knowing the sample mean $\overline{v}^i_{\mathcal{K}}$ (3.9), we can define a standardized random variable

$$Z^i_{\mathcal{K}} := \frac{\overline{v}^i_{\mathcal{K}} - \overline{v}^i}{\frac{\sigma^i}{\sqrt{K}}}. \tag{7.1}$$

According to the Central Limit Theorem, this random variable asymptotically, that is, for $|\mathcal{K}| \to \infty$, follows a standard normal distribution $\mathcal{N}(0,1)$. This implies that for sufficiently large $|\mathcal{K}|$, $Z^i_{\mathcal{K}}$ is approximately standard normal distributed.

Due to symmetry of the standard normal distribution, it follows

$$\mathbb{P}(-z_{1-\alpha/2} \leq Z^i_{\mathcal{K}} \leq z_{1-\alpha/2}) \approx 1-\alpha,$$

where $z_{1-\alpha/2}$ denotes $(1-\frac{\alpha}{2})$-quantiles of $\mathcal{N}(0,1)$ for some level $\alpha \in (0,1)$.

Inserting (7.1) and rearranging yields an approximate $(1-\alpha)$-confidence interval for the true upper bound $\overline{v}^i$:

$$\left[\overline{v}^i_{\mathcal{K}} - z_{1-\frac{\alpha}{2}}\frac{\sigma^i}{\sqrt{|\mathcal{K}|}}, \overline{v}^i_{\mathcal{K}} + z_{1-\frac{\alpha}{2}}\frac{\sigma^i}{\sqrt{|\mathcal{K}|}}\right].$$

As σ^i is unknown, it can be replaced by the sample standard distribution $\sigma^i_{\overline{v},K}$ which is defined by the sample variance

$$(\sigma^i_{\overline{v},\mathcal{K}})^2 := \frac{1}{|\mathcal{K}|-1}\sum_{k\in\mathcal{K}}(v^i(\xi^k) - \overline{v}^i_{\mathcal{K}})^2.$$

In that case, the standardized variable approximately follows a Student's t-distribution with degree of freedom $|\mathcal{K}|-1$. In the literature on SDDP, even in this case, the $(1-\alpha)$-confidence interval for the true upper bound $\overline{v}^i$ is usually approximated using a standard Normal distribution [200], though, which yields:

$$\left[\overline{v}^i_{\mathcal{K}} - z_{1-\frac{\alpha}{2}}\frac{\sigma^i_{\overline{v},\mathcal{K}}}{\sqrt{|\mathcal{K}|}}, \overline{v}^i_{\mathcal{K}} + z_{1-\frac{\alpha}{2}}\frac{\sigma^i_{\overline{v},\mathcal{K}}}{\sqrt{|\mathcal{K}|}}\right]. \tag{7.2}$$

Pereira and Pinto propose choosing $\alpha = 0.05$, which implies $z_{1-\alpha/2} = 1.96$, and stopping SDDP if the lower bound $\underline{v}^i$ is included in this confidence interval [152].

As pointed out by Shapiro [197], this stopping criterion has several flaws. For instance, the higher the sample variance $(\sigma^i_{\overline{v},\mathcal{K}})^2$, the earlier $\underline{v}^i$ exceeds the lower end of the confidence interval, which provides a misguided incentive to increase $(\sigma^i_{\overline{v},\mathcal{K}})^2$.

The same is true for increasing the confidence $1-\alpha$, which contradicts the intuition behind α. Additionally, faster stopping can be achieved by reducing the sample size $|\mathcal{K}|$. Finally, the above stopping criterion may favor premature stopping, as it is rather unlikely that $\overline{v}^i$ is located exactly at the lower bound of the confidence interval.

For this reason, Shapiro proposes a more conservative stopping criterion where SDDP terminates if the difference between the upper bound of the confidence interval (7.2) and $\underline{v}^i$ is sufficiently small.

Sometimes it is also suggested to include values $v^j(\xi^k)$ from previous iterations $j < i$ in (3.9), for instance if $|\mathcal{K}|$ is too small to obtain a reasonable bound. However, this destroys the independence between the different samples. Thus, the Central Limit Theorem can no longer be applied and the confidence-based stopping criteria are not applicable. [49].

Hypothesis Test Criteria. Considering that hypothesis tests and confidence intervals are closely related, the above stopping criterion can also be interpreted in terms of a hypothesis test with hypotheses [104]:

$$H_0 : \overline{v}^i = \underline{v}^i, \quad \text{against} \quad H_1 : \overline{v}^i \neq \underline{v}^i.$$

The null hypothesis H_0 is tested using the test statistic $\overline{v}^i_{\mathcal{K}}$, which is assumed to be approximately normal distributed. This can be reasoned using the Central Limit Theorem for sufficiently large $|\mathcal{K}|$. Then, the region of acceptance for H_0 in iteration i is given by the interval (7.2). By choosing α, the type I error (rejecting optimality although SDDP has converged) can be controlled. However, this comes at the cost of a possibly high type II error (stopping the algorithm prematurely) [104].

To avoid stopping prematurely, Homem-de-Mello et al. propose a modified hypothesis test controlling type I and type II errors simultaneously [104]. First, it is checked whether

$$\rho^i_{\mathcal{K}} := \frac{\overline{v}^i_{\mathcal{K}} - z_{1-\alpha}\frac{\sigma^i_{\overline{v},\mathcal{K}}}{\sqrt{|\mathcal{K}|}}}{\underline{v}^i},$$

is larger than 1. $\rho^i_{\mathcal{K}}$ describes the ratio between the lower bound of the region of acceptance related to a one-sided hypotheses test with $H_0 : \overline{v}^i \leq \underline{v}^i$, and the lower bound $\underline{v}^i$. If it is larger than 1, then optimality is rejected. This is completely in line with the original SDDP hypothesis test.

However, if $\rho^i_{\mathcal{K}} \leq 1$, optimality is not directly retained. Instead, the idea is to predefine a bound $\gamma > 0$ on the probability of a type II error given that the true upper bound $\overline{v}^i$ exceeds the lower bound $\underline{v}^i$ by more than a percentage δ. This means that at least for large deviations, the type II error can be controlled. For given γ and α, and given sample estimates, the value δ can be computed for which γ satisfies this bound criterion:

$$\delta^i = (z_{1-\alpha} + z_{1-\gamma})\frac{\sigma^i_{\overline{v},\mathcal{K}}}{\underline{v}^i\sqrt{|\mathcal{K}|}}.$$

If δ^i is below some predefined threshold $\overline{\delta}$, the sample estimates guarantee that for deviations larger than $\overline{\delta}$, the type II error is under control. Therefore, SDDP stops. Otherwise, the control of the type II error is not considered sufficient, and the algorithm proceeds.

Computational experiments with $\overline{\delta} = 0.1$ and $\gamma = 0.05$ indicate that this stopping criterion is effective in preventing SDDP from premature stopping [104]. Still, it is

a heuristic, and so far, no proposed statistical testing procedure guarantees that the probability of stopping prematurely is bounded by some $\gamma > 0$ in general.

Predefined Criteria. The previous statistical stopping criteria are computationally demanding and require $|\mathcal{K}|$ to be sufficiently large to yield reasonable approximate confidence intervals. Furthermore, in practical applications (MSLP) is often too large to achieve convergence in reasonable time. Finally, the statistical stopping criteria do not necessarily generalize to extensions of SDDP, such as risk-averse variants, see Section 12. Therefore, in practice often more convenient stopping criteria are used for SDDP. For instance, it is common to stop SDDP after a fixed number of iterations $I \in \mathbb{N}$, after a fixed number of cuts $|\mathcal{K}|I$, after a predefined time or if the lower bounds $\underline{v}^i$ have stalled. Neither guarantees that an optimal policy is determined, though.

Deterministic Stopping. Finally, SDDP can be stopped deterministically as long as valid upper bounds $\overline{v}^i$ for v^* are computed in addition to lower bounds $\underline{v}^i$. In that case, for some predefined optimality tolerance $\varepsilon > 0$, SDDP stops with an (approximately) optimal policy if $\overline{v}^i - \underline{v}^i \leq \varepsilon$.

This stopping criterion requires significant additional computational effort to determine true upper bounds $\overline{v}^i$. Hence, there is a trade-off between achieving a more reasonable stopping criterion and spending computational resources on computations offside of the core elements of SDDP. We address how such exact upper bounds can be computed in the next section.

Summarizing, despite various attempts at developing reasonable termination criteria for SDDP, optimally stopping SDDP remains an open challenge.

8. Exact Upper Bounds and Upper Approximations. The idea of computing deterministic upper bounds $\overline{v}$ for v^* and deterministic upper approximations $\overline{\mathfrak{Q}}_t(\cdot)$ of $\mathcal{Q}_t(\cdot)$ has drawn a lot of interest in the research community recently, both in analyzing the convergence behavior of SDDP, see Section 4, and in developing deterministic stopping criteria, see Section 7.

An intuitive way to determine upper approximations $\overline{\mathfrak{Q}}_t(\cdot)$ of $\mathcal{Q}_t(\cdot)$ is based on the observation that due to convexity all secants of $\mathcal{Q}_t(\cdot)$ lie above or on its graph. Therefore, an upper approximation is possible by a convex combination of points $(x_{t-1}, \mathcal{Q}_t(x_{t-1}))$. From another perspective, the convex epigraph $\operatorname{epi}(\mathcal{Q}_t)$ of $\mathcal{Q}_t(\cdot)$ can be approximated by the convex hull $\operatorname{conv}(w_{t-1}^1, \dots, w_{t-1}^{M_t})$ of finitely many points $w_{t-1} := (x_{t-1}, \mathcal{Q}_t(x_{t-1}))$ in $\operatorname{epi}(\mathcal{Q}_t)$. Outside of this convex hull, this inner approximation can be extended using a Lipschitz constant L_t of $\mathcal{Q}_t(\cdot)$ to obtain an approximation on the whole state space. Such constant exists according to Corollary 2.10. In this light, $\overline{\mathfrak{Q}}_t(\cdot)$ can also be constructed as [229]

$$\overline{\mathfrak{Q}}_t(\cdot) := \operatorname{conv}\Big(\mathcal{Q}_t(x_{t-1}^m) + L_t\|x_{t-1} - x_{t-1}^m\|, m = 1, \dots, M_t\Big).$$

This idea is illustrated in Figure 9.

In principle, there are two different approaches to realize this idea. One uses the above perspective, which we refer to as *primal*, and one is related to some *dual* perspective on SDDP and its value functions [97, 118].

8.1. Primal Inner Approximation. Similar to subproblems (2.10), based on upper approximations $\overline{\mathfrak{Q}}_t(\cdot)$ of $\mathcal{Q}_t(\cdot)$, approximating subproblems can be defined by replacing $\mathcal{Q}_t(\cdot)$ with $\overline{\mathfrak{Q}}_t(\cdot)$ in the DPE for all $t \in [T]$. This idea is first introduced by Philpott et al. [160]. As they consider only the RHS of (MSLP) to be uncertain, we adopt this assumption, although it is not required.

For stages $t = T-1, \dots, 2$, each element m in a given set of points $x_t^1, \dots, x_t^{M_{t-1}}$

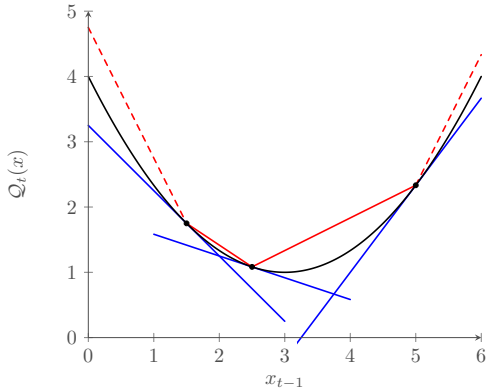


Fig. 9: Inner and outer approximation of $\mathcal{Q}_t(\cdot)$.

and each $\xi_{tj}, j = 1, \ldots, q_t$, the following subproblem can be solved by backward recursion:

$$\overline{Q}_t(x_{t-1}^m, \xi_{tj}) := \begin{cases} \min\limits_{x_t} & c_t^\top x_t + \overline{\Omega}_{t+1}(x_t) \\ \text{s.t.} & x_t \in \mathcal{X}_t(x_{t-1}^m, \xi_{tj}). \end{cases} \tag{8.1}$$

Here, as indicated above, the upper approximation $\overline{\Omega}_{t+1}(\cdot)$ is defined as a convex combination of points $\left(x_t^m, \overline{\mathcal{Q}}_{t+1}(x_t^m)\right), m = 1, \ldots, M_t$. The only difference is that instead of $\mathcal{Q}_{t+1}(x_t^m)$ here $\overline{\mathcal{Q}}_{t+1}(x_t^m) := \mathbb{E}\left[\, \overline{Q}_{t+1}(x_t^m, \boldsymbol{\xi}_{t+1})\right]$ is used, as $\mathcal{Q}_{t+1}(\cdot)$ is not known:

$$\overline{\Omega}_{t+1}(x_t) := \begin{cases} \min\limits_{w} & \sum\limits_{m=1}^{M_t} w_m \overline{\mathcal{Q}}_{t+1}(x_t^m) \\ \text{s.t.} & \sum\limits_{m=1}^{M_t} w_m x_t^m = x_t \\ & \sum\limits_{m=1}^{M_t} w_m = 1 \\ & w_m \geq 0, \quad m = 1, \ldots, M_t. \end{cases} \tag{8.2}$$

By recursion, it can be shown that

$$\overline{Q}_t(x_{t-1}^m, \xi_{tj}) \geq Q_t(x_{t-1}^m, \xi_{tj})$$

for all $m = 1, \ldots, M_{t-1}$ and $j = 1, \ldots, q_t$. This implies

$$\overline{\mathcal{Q}}_t(x_{t-1}^m) \geq \mathcal{Q}_t(x_{t-1}^m).$$

The first-stage problem then yields

$$\overline{v}^{IA} := \begin{cases} \min\limits_{x_t} & c_1^\top x_1 + \overline{\Omega}_2(x_1) \\ \text{s.t.} & x_1 \in \mathcal{X}_1, \end{cases}$$

with $\overline{v}^{IA}$ an exact valid upper bound to v^*.

The main challenge with this approach is to appropriately choose the set of points $x_{t-1}^m, m = 1, \dots, M_{t-1}$. On the one hand, they should be chosen such that as much of $\mathcal{X}_{t-1}$ is spanned as possible. On the other hand, choosing (at least some of) those points as extreme points leads to $M_t \geq 2^{n_t}$ points, *i.e.*, the number of required points grows exponentially in the dimension of the state space.

An alternative is to use the trial points from the SDDP forward pass [160]. Even using these points, the computational effort may become excessive, though. Similarly to the SDDP backward pass, subproblems (8.1) have to be solved for each stage $t \in [T]$, each point $x_{t-1}^m, m = 1, \dots, M_{t-1}$, and each noise term $\xi_{tj}, j = 1, \dots, q_t$. However, in contrast to the backward pass, the number M_{t-1} of points to be considered grows with each iteration, as it contains all previous trial solutions. It is therefore suggested to only use the upper bound computation every few hundred iterations, and not to permanently incorporate it into the backward pass [160]. This hinders using the upper bounds $\overline{v}^{IA}$ in the stopping criterion of SDDP in each iteration, though.

Moreover, the obtained bounds $\overline{v}^{IA}$ may be very loose, especially in problems (MSLP) with a high number of stages. Computational tests are required to assess whether the information gain justifies the additional computational effort and, possibly, higher number of iterations.

Baucke et al. provide a different perspective on the previous inner approximation idea [10]. Instead of (8.2), they use its dual representation

$$\overline{\mathfrak{Q}}_{t+1}(x_t) = \begin{cases} \max\limits_{\mu,\lambda} & x_t^\top \lambda + \mu \\ \text{s.t.} & (x_t^m)^\top \lambda + \mu \leq \overline{\mathcal{Q}}_{t+1}(x_t^m), \quad m = 1, ..., M_t. \end{cases} \tag{8.3}$$

This shows that $\overline{\mathfrak{Q}}_{t+1}(\cdot)$ can be equivalently described by maximizing over the coefficients of all supporting hyperplanes for points $\left(x_t^m, \overline{\mathcal{Q}}_{t+1}(x_t^m)\right), m = 1, \dots, M_t$.

In [10], the dual problem is additionally regularized, *i.e.*, enhanced by constraint

$$\|\lambda\| \leq L_t,$$

with L_t denoting a Lipschitz constant of $\overline{Q}_t(\cdot, \cdot)$. This way, a reasonable approximation is also achieved for points outside of the convex hull of the set defined by the points $x_t^m, m = 1, \dots, M_t$.

Using this expression for the inner approximation functions, Baucke et al. propose a deterministic algorithm for multistage stochastic convex programs. In their case, subproblems (8.1) are solved in each backward pass iteration, and $\overline{\mathfrak{Q}}_{t+1}^i(\cdot)$ is updated by adding constraint $(x_t^{\widetilde{m}})^\top \lambda + \mu \leq \overline{\mathcal{Q}}_{t+1}^i(x_t^{\widetilde{m}})$ for the current iterate $x_t^{\widetilde{m}}$. The proposed algorithm differs in further regards from standard SDDP, for instance it requires a multi-cut approach, see Section 21. Moreover, choosing a reasonable and valid value for L_t can be very challenging, but is crucial for the proposed method to work as intended.

8.2. Dual SDDP. To compute deterministic upper bounds $\overline{v}$ for v^* recently a dual perspective on SDDP and the DPE (2.4) has gained attention.

Using Convex Conjugates of Value Functions. The first proposal in this context, by Leclère et al. [118], exploits convex conjugates and the related duality concepts to derive *dual value functions* for (MSLP) where uncertainty only appears in the RHS $h_t(\xi_t)$.

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty, \infty\}$. Then its *convex conjugate* $f^*(\cdot)$ is defined as [185]

$$f^*(\lambda) := \sup_{x \in \mathbb{R}^n} \lambda^\top x - f(x).$$

For (MSLP), the convex conjugates $D_t(\cdot) := Q_t^*(\cdot)$ of the value functions $Q_t(\cdot)$ can be considered as dual value functions for $t = 2, \dots, T$. It can be shown that these functions also satisfy some DPE with linear subproblems on each stage. Whereas Leclère et al. consider a more general setting including control variables u_t (see Remark 2.2), for (MSLP) as defined in Section 2 (and especially under Assumption 5), for $t = 2, \dots, T$, these subproblems can be expressed by

$$D_t(\lambda_{t-1}) := \begin{cases} \min\limits_{\lambda_t, \mu_t, \gamma_t} & \sum\limits_{j=1}^{q_t} p_{tj} \Big(- h_{tj}^\top \mu_{tj} + D_{t+1}(\lambda_{tj}) \Big) \\ \text{s.t.} & T_{t-1}^\top \left(\sum\limits_{j=1}^{q_t} p_{tj} \mu_{tj} \right) - \sum\limits_{j=1}^{q_t} p_{tj} \gamma_{tj} + \lambda_{t-1} = 0 \\ & W_t^\top \mu_{tj} = \lambda_{tj} + c_t, \qquad j = 1, \dots, q_t \\ & \gamma_{tj} \le 0, \qquad j = 1, \dots, q_t. \end{cases} \tag{8.4}$$

For the first stage, we obtain a deterministic problem, which by $T_0 \equiv 0$ simplifies to

$$D_1(\lambda_0) = \min_{\mu_1} \; h_1^\top \mu_1 + D_t\big(W_1^\top \mu_1 - c_1\big)$$

for some arbitrary initial $\lambda_0 \le 0$ (note that more general formulations of (MSLP) may lead to a dependence on λ_0).

Using this dynamic recursion, it is possible to apply an SDDP-type algorithm, called *dual SDDP*, to $D_t(\cdot)$, using iteratively improving outer approximations $\mathfrak{D}_t^i(\cdot)$ for $D_t(\cdot)$. Analogously to SDDP, this iterative method yields a converging deterministic lower bound for the first-stage optimal value, *i.e.*, $\mathfrak{D}_1^i(\lambda_0) \le D_1(\lambda_0)$. Applying conjugacy theory again, we obtain

$$\overline{v}^i = \big(\mathfrak{D}_1^i\big)^*(x_0) \ge D_1^*(x_0) = Q_1^{**}(x_0) = Q_1(x_0) = v^*.$$

Hence, deterministic upper bounds for v^* can be obtained as conjugates of the first-stage approximations $\mathfrak{D}_t^i(\cdot)$ evaluated at $x_0 = 0$, and $(\overline{v}^i)_i$ defines a sequence converging to v^* [118].

Using the Dual of (MSLP). Guigues et al. propose an alternative way to define dual value functions and DPE that can be exploited in a dual SDDP algorithm [97]. Instead of working with conjugates of the primal value functions $Q_t(\cdot)$, they first derive the dual to (MSLP) formulated as a single problem (2.3), and then show that this dual problem can be decomposed using DPE and dual value functions

$$\widetilde{D}_t(\pi_{t-1}) := \begin{cases} \max\limits_{\pi_t} & \sum\limits_{j=1}^{q_t} p_{tj} \Big(- h_{tj}^\top \pi_{tj} + \widetilde{D}_{t+1}(\pi_{tj}) \Big) \\ \text{s.t.} & \sum\limits_{j=1}^{q_t} p_{tj} \Big(T_{t-1,j}^\top \pi_{tj} \Big) + W_{t-1}^\top \pi_{t-1} \le c_{t-1}. \end{cases} \tag{8.5}$$

It can be argued that these dual DPE are simpler and more intuitive, as they do not require conjugacy theory. Moreover, we immediately obtain that the first-stage

optimal value $\widetilde{D}_1(\pi_0)$ equals v^* by strong duality for linear programs. Therefore, using outer approximations $\widetilde{\mathfrak{D}}_t^i(\cdot)$ of these value functions in dual SDDP, again a sequence $(\overline{v}^i)_i$ of deterministic and valid upper bounds converging to v^* can be computed [97], but without requiring to consider conjugates. On the other hand, the dual value functions $\widetilde{D}_t(\cdot)$ cannot be directly related to the original value functions $Q_t(\cdot)$.

Remark 8.1. Even if the dual DPE (8.4) and (8.5) are derived using different tools and perspectives, they are still closely related. Note that subproblem (8.4) can be reformulated as

$$D_t(\lambda_{t-1}) = \begin{cases} \min\limits_{\pi_t, \mu_t} & \sum\limits_{j=1}^{q_t} p_{tj}\Big(- h_{tj}^\top \mu_{tj} + D_{t+1}(\lambda_{tj})\Big) \\ \text{s.t.} & T_{t-1}^\top \left(\sum\limits_{j=1}^{q_t} p_{tj} \mu_{tj} \right) + \lambda_{t-1} \leq 0 \\ & W_t^\top \mu_{tj} = \lambda_{tj} + c_t, \qquad j = 1, \dots, q_t. \end{cases}$$

Using the last constraint, the state λ_{t-1} can be expressed through the dual variables μ_{t-1} from the previous stage: $\lambda_{t-1} = W_{t-1}^\top \mu_{t-1} - c_{t-1}$. Replacing this, the subproblems only contain dual variables μ_t, which have to be considered as state variables. By adapting the optimization sense in the objective, we get exactly the structure of (8.5).

We can make the following additional observations with respect to the dual DPE (8.4) and (8.5). Firstly, in both cases, the subproblems are not necessarily bounded. Therefore, in both cases, artificial bounds are introduced. In [97] they are chosen as $\pi_t \in [\underline{\pi}_t, \overline{\pi}_t]$, whereas in [118] Lipschitz continuity of $\mathcal{Q}_t(\cdot)$ is exploited to impose the bounds $\|\lambda_t\|_\infty \leq L_t$ for Lipschitz constants $L_t, t \in 2, \dots, T$. It is assumed that these bounds are chosen sufficiently large to not affect the optimal solutions.

Secondly, even if the primal DPE (2.4) are assumed to have relatively complete recourse (see Assumption 9 and Lemma 2.5), this does not necessarily translate to the dual subproblems. To ensure feasibility, Guigues et al. propose to either use feasibility cuts (also see Section 17) or a penalization approach [97].

Thirdly, in contrast to the primal perspective, the subproblems do not decompose by realizations of $\boldsymbol{\xi}_t$, but contain separate dual variables π_{tj} (or $\lambda_{tj}, \mu_{tj}, \gamma_{tj}$, respectively) for all $j = 1, \dots, q_t$. In the forward pass of dual SDDP the trial point π_t^i (or λ_t^i) that is used as a parameter in the following stage is sampled from these variables.

Finally, if W_t and c_t become uncertain as well, then the value functions and subproblems additional depend on ξ_t. In fact, in formulation (8.5) the state space has to be extended to include the history ξ_{t-1} of the stochastic process, as the problem contains W_{t-1} and c_{t-1} [97].

Again, an SDDP-type algorithm, also referred to as *dual SDDP* in [97], can be applied to the DPE (8.5). This algorithm is presented in Algorithm 8.1. The two variants of dual SDDP have been extended to the risk-averse case [40] (see also Section 12) and to problems with infinite horizon (see also Section 19) [199].

Dual Inner Approximation. First and foremost, dual SDDP is an alternative to (primal) SDDP to approximate v^* by converging deterministic upper bounds $\overline{v}$. However, as shown in [118], if the dual DPE (8.4) are used, then the obtained approximations $\mathfrak{D}_t^i(\cdot)$ may be translated to inner approximations $\overline{\mathfrak{Q}}_t^i(\cdot)$ of the *primal* value functions $Q_t(\cdot)$. This way, policies $(\boldsymbol{x}_t(\xi_{[t]}))_{t \in [T]}$ for (MSLP) can be computed. The inner approximations can be computed as Lipschitz regularizations (see Sect. 17) of

Algorithm 8.1 Dual SDDP from [97]

Input: Dual to problem (MSLP) satisfying Assumptions 1 to 9. Appropriate multiplier bounds. Stopping criterion.

`Initialization`

Initialize cut approximations with bounded $\widetilde{\mathfrak{D}}_t^0(\cdot)$ for all $t = 2, \dots, T$.
Initialize upper bound with $\overline{v}^0 = +\infty$.
Set iteration counter to $i \leftarrow 0$.

`Dual SDDP Loop`

while Stopping criterion not satisfied **do**
  Set $i \leftarrow i + 1$.

  `Forward Pass`

  Solve the first-stage problem (defined by replacing $\widetilde{D}_2(\cdot)$ with $\widetilde{\mathfrak{D}}_2^i(\cdot)$ and adding multiplier bounds in (8.5)). Store the trial point π_1^i.
  for stages $t = 2, \dots, T$ **do**
    Solve the stage-t subproblem (defined by replacing $\widetilde{D}_{t+1}(\cdot)$ with $\widetilde{\mathfrak{D}}_{t+1}^i(\cdot)$ and adding multiplier bounds in (8.5)) for π_{t-1}^i to obtain $\pi_{tj}^i, j = 1, \dots, q_t$.
    Sample $\widetilde{j}$ from $j = 1, \dots, q_t$ and set $\pi_t^i = \pi_{t\widetilde{j}}$.
  end for

  `Backward Pass`

  for stages $t = T, \dots, 2$ **do**
    Solve the updated stage-t subproblem (2.10) (defined by replacing $\widetilde{D}_{t+1}(\cdot)$ with $\widetilde{\mathfrak{D}}_{t+1}^{i+1}(\cdot)$ and adding multiplier bounds in (8.5)) for π_{t-1}^i. Store the optimal value $\overline{D}_t(\pi_{t-1}^i)$ and the optimal dual vector x_{t-1}^i.
    Compute

$$\alpha_t^{D,i} := \overline{D}_t(\pi_{t-1}^i) - \left(\beta_t^{D,i}\right)^\top \pi_{t-1}^i$$

    and

$$\beta_t^{D,i} := -W_{t-1} x_{t-1}^i.$$

    Update the cut approximation of $\widetilde{D}_t(\cdot)$ to

$$\widetilde{\mathfrak{D}}_t^{i+1}(x_{t-1}) := \min\left\{\widetilde{\mathfrak{D}}_t^i(x_{t-1}),\ \alpha_t^{D,i} + \left(\beta_t^{D,i}\right)^\top \pi_{t-1}\right\}.$$

  end for
  Solve the first-stage problem (defined by replacing $\widetilde{D}_2(\cdot)$ with $\widetilde{\mathfrak{D}}_2^{i+1}(\cdot)$ and adding multiplier bounds in (3.8)) to obtain an upper bound $\overline{v}^i$.
end while

Output: Upper bound $\overline{v}^i$ for v^*.

the convex conjugate of the outer approximations $\mathfrak{D}_t^i(\cdot)$, which is shown to be equivalent to solving problem (8.3) with regularization $\|\lambda\|_\infty \leq L_t$. The key difference to the approach in [10] is the way the primal supporting points x_t^m are determined, that is, by the slopes of the dual outer approximation [118].

Incorporation into SDDP. While dual SDDP can be applied on its own to

approximate v^*, and even compute policies $(\boldsymbol{x}_t(\xi_{[t]}))_{t\in[T]}$, it seems reasonable to incorporate it into (primal) SDDP in order to compute deterministic upper *and* lower bounds for v^*. Guigues et al. suggest to use both variants of SDDP in parallel [97]. In contrast, Leclère et al. propose a framework where primal and dual SDDP are intertwined [118]:

1. Run a forward pass of (primal) SDDP, yielding trial solutions $(x_t^i)_{t\in[T]}$ for the sampled scenario path (the authors choose $|\mathcal{K}| = 1$).
2. Run a backward pass of (primal) SDDP using the trial solutions x_{t-1}^i, obtaining new slopes π_t^i from the cuts.
3. Run a backward pass of dual SDDP using the slopes $\lambda_t^i = \pi_t^i$, obtaining new cuts for the dual problem.
4. Run a forward pass of dual SDDP, to obtain a new dual trajectory $(\lambda_t^i)_{t\in[T]}$ and update the cuts along this trajectory.

One computational drawback of this framework, and of dual SDDP in general, is that each iteration of dual SDDP is much more computational expensive than for standard (primal) SDDP. This hampers the application of a solely deterministic stopping criterion for very large problems [97, 118].

9. Applications. In this section, we present different application areas of SDDP. We also point out applications in which some of the Assumptions 1 to 9 are not satisfied, and therefore either modifications of (MSLP) or algorithmic extensions are required in order to apply SDDP. These use cases can be regarded as a motivation for the enhancements of SDDP that we cover in the following sections.

9.1. Power System Optimization. By far the dominating application field of SDDP is power system optimization, in particular, the operational planning of energy systems including hydro storages by a central planner. This is due to its adequacy for such problems, but also due to its origins in optimizing the operational planning of the Brazilian hydrothermal system [151, 152].

In general, solving power system optimization problems is a very complex task, as it allows for incorporation of various technical and economical details and uncertainties [109, 148, 181, 182, 183, 207, 208, 230]. Including all these details in one single problem is computationally intractable. Therefore, usually a hierarchy of problems is considered, dealing with different time-scales and perspectives [47], such as short-term dispatch (a few days or weeks), mid-term operational planning (1-2 years) and long-term operational planning (3-5 years) [72, 81]. Results from a long-term model can then be incorporated into one with a shorter horizon, but more detail in other modeling aspects.

9.1.1. Long-term Operational Planning. SDDP is most prominently used for long-term operational planning (LTOP) of hydrothermal power systems, also called long-term hydrothermal scheduling (LTHS). In the research literature, SDDP has been applied to LTOP of various hydrothermal systems, with the most prominent ones being the hydro power dominated systems in Brazil [15, 30, 31, 32, 42, 47, 48, 49, 52, 84, 97, 104, 125, 126, 128, 132, 135, 160, 165, 201, 202, 203, 206, 214, 224], other Central or South American countries [6, 70, 178, 211], Norway [80, 186] and New Zealand [159, 161, 227]. Additionally, to this day, SDDP is applied by the Brazilian system operator ONS in practice [133, 134].

The aim in LTOP is to determine an optimal policy for the amount of power to be generated by thermal and hydroelectrical utilities over some planning horizon of several years (usually with monthly resolution) such that demand is satisfied, technical

constraints are fulfilled and the expected cost is minimized [159]. The main focus is on managing hydro reservoirs, and thus the water resource efficiently. This is not trivial. While there is an incentive to use all the water in a reservoir immediately, as no fuel costs occur, also the potential value of storing water for later stages has to be considered, with taking into account the uncertainty of future inflows. For this reason, it can be beneficial to retain water in wet periods for following dryer periods. The ability to store water in reservoirs leads to a temporal coupling of the stages. The number of inflow realizations q_t per stage is typically chosen in a range between 20 and 100. For $T = 60$, this yields a scenario tree with about $1.15e78$ or $1e120$ scenarios. Per forward pass, either a single scenario [48] or 100 to 200 scenarios are sampled.

LTOP can be used to illustrate some of the challenges and limits of standard SDDP, and thus motivate the necessity of extensions.

Autoregressive Uncertainty. In LTOP, the main source of uncertainty are future (usually monthly) inflows into the reservoirs. These inflows often show seasonality and a temporal or spatial coupling which has to be considered in modeling. Therefore, usually *autoregressive* (AR) processes are used to model and forecast them, in particular *periodic autoregressive* (PAR) [132, 133] and related models [138]. This means that for each reservoir and each month a different AR model is fitted, or in other words, that the parameters in the AR model are allowed to differ between months.

Additionally, often hydro reservoirs are organized in cascade systems. Then, the generation of one turbine may affect the inflow of downstream reservoirs, such that they cannot be managed separately. For this reason, inflows often do not only show temporal correlation and seasonality, but also spatial correlation. To address this, instead of PAR, *spatial periodic autoregressive* (SPAR) models can be used [126]. Such model is still linear, but instead of only autoregressive components, *i.e.*, lags of ξ_{it} for some reservoir i, also lags of the inflows of neighboring reservoirs i' are used to explain ξ_{it}. Apart from inflow lags, also different exogenous variables, such as climate indices, precipitation or sea temperature can be used to explain inflows [123, 164].

Whenever an AR process is used for the uncertain data, the assumption of stagewise independence (Assumption 2) is not satisfied. This motivates an extension of SDDP able to handle stagewise dependent uncertainty. We discuss this in Section 14.

Nonlinear Uncertainty. When modeling hydro inflows, the error terms in the AR process are usually assumed to be i.i.d. with normal or log-normal distribution [47, 126]. In the latter case, the model is also referred to as a geometric PAR (GPAR) model [128]:

$$\ln(\xi_t) = \gamma_t + \Phi_t \ln(\xi_{t-1}) + \eta_t. \tag{9.1}$$

GPAR models are usually more accurate in modeling inflows, as these often tend to positive skewness and are thus not normally distributed. Moreover, they have the advantage that the requirement of non-negative inflows is naturally satisfied.

On the other hand, solving (9.1) for ξ_t yields an AR process with multiplicative instead of additive error terms [202], which is a nonlinear model. Incorporating this into the DPE destroys the convexity of $\mathcal{Q}_t(\cdot)$, making a direct application of SDDP impossible. Instead, the nonlinear model has to be approximated linearly [202]. Another idea is to normalize the inflows first using a Box-cox transformation. As such a transformation is nonlinear, still a linear approximation is required afterwards, though [167]. Further strategies to avoid non-negative inflows and nonlinearities are discussed in [47, 176]. In [45] it is suggested to apply bootstrapping to resample directly from the historical residuals instead of applying a nonlinear transformation.

Continuous Uncertainty and Distributional Uncertainty. As stated before, usually a normal or log-normal distribution is assumed for the error terms in the inflow models, both being continuous distributions (an exception is [170] where inflows are modeled as a continuous process with discrete random errors). For this reason, the assumption of finite discrete random variables (Assumption 5) is not satisfied. Additionally, the chosen distribution for the model may not coincide with the *true distribution* of the uncertain data. This raises the questions of how to handle continuous uncertainty and distributional uncertainty in SDDP. We address these questions in Section 11 and Section 13.

Computational Performance. Despite the amenities of SDDP, its performance may suffer for problems with a large number of state variables, due to its exponential complexity in the state dimension d_t, see Section 4.2. For instance, SDDP is computationally prohibitive for a complete model of the Brazilian energy system consisting of about 150 thermal plants and more than 150 hydro storages [48]. This is aggravated if the state dimension is artificially increased, *e.g.*, in order to deal with stagewise dependent uncertainty, see Section 14. As a relief, it is common practice to aggregate reservoirs based on their region and hydrological properties in so-called *energy equivalent reservoirs* (EER) [4], thus reducing the state dimension [134]. However, this comes with an increased abstraction, and may lead to suboptimal policies. Moreover, as outlined in [47], the EER modeling may introduce some nonlinearities into the system, which have to be mitigated by linearization.

The computational complexity with respect to the state space also makes general performance improvements for SDDP indispensable, which we discuss in Section 21.

End-of-horizon Effect. Another challenge when applying SDDP to LTOP in practice is the so-called *end-of-horizon effect*. It relates to the effect that obtained policies do not guarantee a continuous and reliable energy supply *after* the planning period, because in an optimal policy, all energy remaining in the reservoirs will be used at the end of the planning period. A typical planning horizon for LTOP are 5 years with a monthly resolution, leading to 60 stages. A common practice to mitigate the end-of-horizon effect is to add 60 more stages to the problem, *i.e.*, to consider a problem with 120 stages [202], even if only decisions of the first half are about to be implemented. Alternatively, it seems natural to analyze how SDDP can be applied for problems with an infinite horizon or with a random horizon, where Assumption 1 is not satisfied. We address this in Sections 19 and 20.

Risk-aversion. Due to the high importance of system reliability and stability to prevent outages and electricity shortages, system planners may favor more risk-averse policies compared to the risk-neutral ones obtained by standard SDDP. Therefore, there has been an increased interest recently to take risk aversion into account when applying SDDP to LTOP [107, 206]. However, as Assumption 8 is no longer satisfied, this requires to extend standard SDDP to a risk-averse variant. We discuss different approaches to achieve this in Section 12.

9.1.2. Medium-term Operational Planning. Structurally, medium-term operational planning problems (MTOP) do not differ much from LTOP. The main difference is that a shorter, one- or two-year time horizon is considered [47, 159, 160, 178].

Price Uncertainty in the Objective. Especially on a medium-term time horizon, SDDP has also been adopted from the traditional setting with a single system operator to more market-driven systems, in which several electricity suppliers are active. In such systems, besides inflows also spot prices can be considered uncertain. This imposes an additional challenge to SDDP, as it leads to stagewise dependent

uncertainty in the objective. We discuss this in detail in Section 14. To deal with this challenge, for instance, for the operational planning of the Norwegian hydro-storage system, usually a combined SDP/SDDP approach is used [79, 80, 81, 99, 100].

Water Head Effect. In LTOP the so-called *water head effect* of hydro storages is often disregarded, but it may become decision-relevant in (MTOP). This effect describes that the production of a hydro plant increases with the net head of the reservoir. As this production function is multiplied with the water discharge, it introduces non-convexities to the problem. Therefore, if this nonlinear effect is explicitly considered, suitable extensions of SDDP to non-convex problem are required [36, 102, 162]. We cover such extensions in Section 16.

9.1.3. More Energy Applications. We briefly summarize further applications of SDDP in power system optimization.

Short-term Dispatch. SDDP is particularly suited for long-term planning, but it can also be applied to short-term economic dispatch problems [37, 51, 117, 147]. For shorter time horizons, it may be reasonable to include additional system aspects, for instance power flow and security constraints, reserve energy or different ancillary services [131, 214]. If security constraints are considered, usually linear DC power flow models are used, but recently also AC power flow has gained interest [111].

Another research stream considers CO_2 emissions, which can be covered by imposing an emission quota system [14, 179, 177] or by introducing emission trading [180]. The first approach leads to an (MSLP) which has no block-diagonal structure (Assumption 7). We discuss how SDDP can be applied in this case in Section 18.

Using a reasonable extension to mixed-integer programs, see Section 16, also unit commitment problems are accessible by the SDDP idea [232].

Different Storage Systems. As different types of storage systems can be modeled similar to hydro storages, SDDP is also applicable to such systems, for instance, to optimize gas storage facilities [225] or energy storages in microgrids [22].

Optimal Bidding. Instead of minimizing expected system cost from the perspective of a central system operator, in strategic bidding problems power plant operators attempt to determine an optimal bidding policy in order to maximize their expected revenue, while taking into account information uncertainty, for example with respect to inflows or the market-clearing price; see [210, 212] for an overview.

Since the future revenue functions of the price-maker have a sawtooth shape, the resulting problem is non-convex [211]. Therefore, to apply the SDDP idea, tailor-made extensions are required, *e.g.*, convexifications, approximations by saddle cuts [55] or by step functions [162, 227]. For methodological details, we refer to Section 16.

Recently, also applying SDDP to optimize trading in continuous intraday markets has gained attention [204].

Investment Planning. An important long-term optimization problem in power systems is to make optimal (risk-averse) investment decisions, either with respect to the expansion of renewables [33, 122, 213] or to conventional projects.

For conventional power systems, common investment problems address the questions of generation expansion or transmission expansion. The main challenge with such problems is that they naturally impose the introduction of integer decision variables. Therefore, in such a case relaxations [146] or appropriate extensions of SDDP, *e.g.*, SDDiP [232], have to be used (see Section 16). Alternatively, SDDP can be incorporated into a larger Benders decomposition framework, where at the first stage binary investment decisions are taken and at the second stage a multistage stochastic linear program is solved by SDDP [177]. Similar applications are considered in [52]

and [41] with a special focus on risk and reliability constraints.

Coping with Renewable Uncertainty. An increasing share of renewable energy sources introduces more variability to an energy system, which has to be taken into account and balanced by appropriate mechanisms. The usage of distributed grid-level storage, such as batteries or electric vehicles, for smoothing out the variable generation of renewables is examined using SDDP in [66, 233].

9.2. Water Resource Management. In many energy applications of SDDP, managing water resources plays a key role, as it couples subsequent stages. Apart from energy optimization, SDDP is also applied to more general water resource management problems, where not only energy production, but also water usage for irrigation in agriculture [154, 219], flow requirements for navigation [219], groundwater [136] or ecological constraints [218] are taken into account in the operational planning of reservoirs. Also related is the problem of river basin management [187].

Additionally, SDDP is used for assessing various quantities in hydrological systems, *e.g.*, the value of water [222], risk for dam projects [2, 221], resource vulnerabilities [188] or benefits and costs of cooperation or non-cooperation [137, 220].

9.3. Portfolio Management. The optimal management of a portfolio of investments, also referred to as *asset allocation*, can be modeled as an (MSLP) [43]. The aim is to distribute a fixed investment sum among a finite number of assets with uncertain returns, in such a way that the expected return at the end of the considered horizon is maximized. By selling or buying certain amounts of assets, the investor can restructure his portfolio in each time period. Usually, both operations are associated with transactions costs, which leads to a very complex problem [223].

In the literature on SDDP, asset allocation problems are quite popular to test proposed improvements and enhancements of SDDP, such as regularization [90], cut-sharing [84] or inexact cuts [8]. Since most investors are risk-averse, asset allocation problems are a popular application [60, 63, 64, 105, 112, 113], but also one of the main drivers for the development of risk-averse SDDP, which we introduce in Section 12.

For applications of practical interest, asset allocation becomes very challenging, as pointed out in [223]. Firstly, risk aversion parameters such as λ_t or α_t, see Section 12, are not intuitive to choose in such a way that the true preferences of an investor are appropriately represented. For this reason, the authors propose to solve a risk-constrained model with one-period conditional CVaR constraints instead of a usual risk-averse SDDP approach. Secondly, assuming stagewise independence of asset returns may prove unrealistic, requiring a more sophisticated approach such as incorporating a Markov chain, see Section 14. Moreover, the large supply of potential assets leads to a high-dimensional state space.

9.4. Further Applications. Although the focus is on the previous applications, occasionally also other types of applications are investigated using SDDP. Among those applications are dairy farming [61, 83], newsvendor problems [5, 149], inventory management [8, 59, 87, 97], lot-sizing [216] and routing problems [60]. In [50] and [232] airline revenue management is explored, which is an established problem in dynamic programming, but requires integer variables.

10. Software. Until recently, SDDP implementations have been solely restricted to closed research projects or commercial products. For commercial products, most established is the SDDP implementation by PSR, a Brazilian energy consultancy [171]. A newer stochastic programming software, which also includes SDDP ideas, is provided by Quantego and can be accessed using MATLAB, Python and Java [172].

For research projects, various different implementations exist, covering programming languages like AMPL, C++, Fortran, GAMS, Java or MATLAB, see [57].

In the last few years, open-source implementations have gained more and more interest, with the aim to increase research transparency, enhance research exchange and benchmarking, and facilitate access to SDDP in industry and science [57]. The most prominent programming language in this regard is Julia [21], which provides its own algebraic modeling language JUMP [62] and is increasingly used in operations research and especially stochastic progamming. By now, with `StochDynamicProgram.jl` [119], `StructDualDynProg.jl` [120] and `SDDP.jl` [57] there exist three SDDP implementations in Julia. Similarly, SDDP packages are available in MATLAB (`FAST` [34]), C++ (`StOpt` [77]) and Python (`msppy` [50]).

Currently, `SDDP.jl`, which is based on Dowson's concept of policy graphs [56], can be considered the most comprehensive package. It provides many of the features described in this paper, such as cut selection, parallelism, Markov chain SDDP, objective states, belief states, SDDiP, as well as different stopping criteria and sampling approaches. Moreover, it includes some of the approaches discussed for distributionally robust and risk-averse SDDP. However, as most other packages, it requires the underlying stochastic process to be finite. Thus, if Assumption 5 is not satisfied, some discretization has to be applied a priori. Then, the results obtained by SDDP are valid for the discretized problem, but not put into perspective with respect to the *true* problem. `msppy`, on the other hand, integrates both, the discretization by SAA and the solution by SDDP in one package and, thus, can naturally be applied to problems with continuous uncertainty [50].

A more detailed comparison of currently available libraries is presented in [57].

11. SDDP for Continuous Uncertainty [relaxing Assumption 5]. So far, we assumed the uncertainty in (MSLP) to be modeled by some discrete and finite random process, see Assumption 5, in order for SDDP to be applicable. Until the recent work by Forcier and Leclère [71], also all convergence proofs for SDDP leveraged Assumption 5. However, in many practical applications, this assumption is not justified. For example, if the stochastic process governing the uncertain data is modeled by a time series model, the random error terms are usually assumed to follow a continuous distribution [197], see Section 9. In the remainder of this section, we denote a problem with such a continuous data process by $(\widetilde{P})$.

As pointed out in Section 2.3, for problems with sizes of practical interest, problem $(\widetilde{P})$ is computationally intractable. Therefore, if the true distribution $F_{\boldsymbol{\xi}}$ of the stochastic process $(\boldsymbol{\xi}_t)_{t\in[T]}$ is continuous, usually an approximation with finitely many scenarios is used. In the literature on multistage stochastic programming, a variety of techniques are proposed to generate (and reduce) scenario tree approximations of continuous stochastic processes. For an overview we refer to [127].

11.1. Sample Average Approximation (SAA). The most common approximation approach is to use random sampling. That means that the distribution $F_{\boldsymbol{\xi}}$ is *approximated* using an empirical distribution $\widetilde{F}_N$ with a finite number N of scenarios, which is obtained by sampling from $F_{\boldsymbol{\xi}}$ [197]. This yields an approximating problem $(\widetilde{P}_N)$, which then can be handled by SDDP. Often, this technique is referred to as *sample average approximation* (SAA), especially, if classical Monte Carlo sampling is used. We discuss SAA and the application of SDDP to solve an SAA problem in more detail now. For a general analysis of SAA, we refer the interested reader to [200].

SAA and SDDP. Under stagewise independence of $(\boldsymbol{\xi}_t)_{t\in[T]}$ (Assumption 2),

it is desirable to preserve this property in the SAA problem, especially if the latter should be solved by SDDP. To achieve this, random sampling can be applied to each stage $t = 2, \dots, T$ independently with sample size $\widetilde{q}_t$ [197]. The obtained SAA has a total number of $N = \prod_{t=2}^{T} \widetilde{q}_t$ scenarios, *i.e.*, the number of scenarios is exponentially growing in the number of stages [197].

For the SAA problem $(\widetilde{P}_N)$, for each stage $t = 2, \dots, T$ and each sample $j = 1, ..., \widetilde{q}_t$, the DPE can be written as

$$\widetilde{Q}_t\big(x_{t-1}, \widetilde{\xi}_{tj}\big) := \begin{cases} \min\limits_{x_t} & \big(c_t\big(\widetilde{\xi}_{tj}\big)\big)^\top x_t + \widetilde{\mathcal{Q}}_{t+1}(x_t) \\ \text{s.t.} & x_t \in \mathcal{X}_t(x_{t-1}, \widetilde{\xi}_{tj}) \end{cases} \tag{11.1}$$

where

$$\widetilde{\mathcal{Q}}_{t+1}(x_t) := \frac{1}{N_{t+1}} \sum_{j=1}^{\widetilde{q}_{t+1}} \widetilde{Q}_{t+1}(x_t, \widetilde{\xi}_{t+1,j}) \tag{11.2}$$

and $\widetilde{\mathcal{Q}}_{T+1} \equiv 0$. For the first stage, we obtain

$$\widetilde{v}_N := \begin{cases} \min\limits_{x_1} & c_1^\top x_1 + \widetilde{\mathcal{Q}}_2(x_1) \\ \text{s.t.} & x_1 \in \mathcal{X}_1. \end{cases} \tag{11.3}$$

The DPE (11.1)-(11.3) can be approached by SDDP as described in Section 3. However, in contrast to the problems considered there, the SAA problems are random, as they depend on a sample from the true data process $(\boldsymbol{\xi}_t)_{t\in[T]}$.

SAA Properties. Since the aim is to solve the original problem $(\widetilde{P})$, the central question is how the solution and the bounds obtained by applying SDDP to the SAA problem $(\widetilde{P}_N)$ relate to the solution of $(\widetilde{P})$. We denote the optimal value of $(\widetilde{P})$ by $\widetilde{v}^*$ and the bounds obtained by SDDP in iteration i with $\underline{\widetilde{v}}^i$ and $\overline{\widetilde{v}}^i_{\mathcal{K}}$. We summarize important properties of SAA.

(P.11.1) *Consistency.* It can be shown that the optimal value $\widetilde{\boldsymbol{v}}_N$ provides a consistent estimator of the true optimal value $\widetilde{v}^*$, *i.e.*, $\lim_{\widetilde{q}_2,\dots,\widetilde{q}_T\to\infty} \mathbb{E}[\widetilde{\boldsymbol{v}}_N] = \widetilde{v}^*$ with probability 1 [197, 200]. The intuition behind this is that asymptotically, the structure of the true process $(\boldsymbol{\xi}_t)_{t\in[T]}$ is recovered. In practical applications, increasing $\widetilde{q}_t$ to infinity is computationally intractable, though.

(P.11.2) *Bias.* $\widetilde{\boldsymbol{v}}_N$ is a biased estimator of $\widetilde{v}^*$, more precisely, $\mathbb{E}[\widetilde{\boldsymbol{v}}_N] \leq \widetilde{v}^*$ for all N [200], since only a subset of all scenarios is considered and the decisions are optimized with respect to these scenarios [48]. This means that solving the SAA problem provides a (converging) estimator of a lower bound for $\widetilde{v}^*$ [194].

(P.11.3) *Lower Bounds.* In each iteration i of SDDP, we have $\underline{\widetilde{v}}^i \leq \widetilde{v}_N$. Therefore, $\mathbb{E}[\underline{\widetilde{\boldsymbol{v}}}^i] \leq \widetilde{v}^*$ [197], and the SDDP lower bound is a statistical lower bound for $\widetilde{v}^*$. Note, however, that both, $\widetilde{\boldsymbol{v}}_N$ and $\underline{\widetilde{\boldsymbol{v}}}^i$, are lower bounds in expectation only, whereas this is not clear for one specific SAA problem $(\widetilde{P}_N)$.

(P.11.4) *Upper Bounds.* Applying SDDP to the DPE (11.1)-(11.3) yields a policy. Under relatively complete recourse (see Assumption 9) with respect to the true data process $(\boldsymbol{\xi}_t)_{t\in[T]}$, this policy also yields feasible decisions if applied to any realization $(\xi_t)_{t\in[T]}$ of this true process. By computing

$$\mathbb{E}\left[\sum_{t=1}^{T} \big(\boldsymbol{c}_t(\xi_t)\big)^\top \boldsymbol{x}_t^i\big(\xi_{[t]}\big)\right] \tag{11.4}$$

with the expectation taken with respect to the true process, a valid upper bound for $\widetilde{v}^*$ can be obtained [197].

(P.11.5) The sample mean $\overline{\widetilde{v}}^i_{\mathcal{K}}$ determined in iteration i in SDDP is an unbiased and consistent estimator of (11.4). Hence, $\mathbb{E}\left[\overline{\widetilde{\boldsymbol{v}}}^i_{\mathcal{K}}\right] \geq \widetilde{v}^*$.

Even with these theoretical properties, solving $(\widetilde{P})$ using SAA may be computationally intractable. Shapiro shows that even under relatively complete recourse (see Assumption 9) and stagewise independence (Assumption 2) of the true data process $(\boldsymbol{\xi}_t)_{t\in[T]}$, the total number of scenarios required in SAA problem $(\widetilde{P}_N)$ to solve $(\widetilde{P})$ with a reasonable accuracy $\varepsilon > 0$ grows exponentially in the number of stages [195]. Therefore, he proposes to use smaller sample sizes $\widetilde{q}_t$ for later stages, although then the accuracy of the solution cannot be guaranteed anymore [196].

Clearly, there exists a trade-off between the quality of the obtained bounds for $\widetilde{v}^*$ and the computational tractability of the SAA problem. Approximating $F_{\boldsymbol{\xi}}$ with F_N using very large sample sizes $\widetilde{q}_t$ for all $t = 2, \dots, T$, a much better representation of the original process $(\boldsymbol{\xi}_t)_{t\in[T]}$ is obtained, leading to a better approximation of $\widetilde{v}^*$. However, in this case, it may be even impossible to solve the SAA problem to optimality in reasonable time, as it may take too long until all scenarios are *eventually* sampled [197]. On the other hand, a very rough approximation yields a problem $(\widetilde{P}_N)$, which can be solved efficiently by SDDP, but does not provide reasonable information about the solution to the true problem $(\widetilde{P})$ [113].

11.2. Assessing Policy Quality. As it is computationally intractable to solve an SAA problem of $(\widetilde{P})$ with a sample size that guarantees a predetermined accuracy, in practice, usually moderate sample sizes are used. For example, in [48], sample sizes with branching numbers $\tilde{q}_t$ between 5 and 200 are tested.

The bounds $\underline{\widetilde{v}}^i$ and $\overline{\widetilde{v}}^i_{\mathcal{K}}$ in SDDP are determined using one specific sample of $(\boldsymbol{\xi}_t)_{t\in[T]}$. Therefore, they only measure the *in-sample* performance of the determined feasible policy $\big(\boldsymbol{x}_t(\xi_{[t]})\big)_{t\in[T]}$. To assess its quality for the original problem $(\widetilde{P})$, *i.e.*, its *out-of-sample* performance, it is required to evaluate it with respect to the original process $(\boldsymbol{\xi}_t)_{t\in[T]}$. Such an evaluation also allows one to compare policies obtained for different SAA problems, which can be helpful in designing appropriate sampling techniques and sample sizes [48].

Various techniques have been proposed in stochastic programming to measure the performance of feasible policies, such as analyzing optimality conditions, assessing solution stability or estimating the optimality gap [48]. Specifically for SDDP, Morton et al. have made substantial contributions [39, 48, 113], which are based on estimating the optimality gap ([113] analyzes a risk-averse variant of SDDP, see Section 12). We discuss their ideas for the risk-neutral case thoroughly in the remainder of this subsection. In accordance with [48], we only consider uncertainty in the RHS of $(\widetilde{P})$.

Estimating the Optimality Gap. For some feasible policy $\big(\boldsymbol{x}_t(\xi_{[t]})\big)_{t\in[T]}$, let $\widetilde{v}(\xi) = \sum_{t=1}^{T} c_t x_t\big(\xi_{[t]}\big)$ denote the random cost for some arbitrary scenario path $\xi = (\xi_1, \dots, \xi_T)$. From (P.11.4) we have $\mathbb{E}[\widetilde{\boldsymbol{v}}(\xi)] \geq \widetilde{v}^*$. Therefore, the optimality gap induced by policy $\big(\boldsymbol{x}_t(\xi_{[t]})\big)_{t\in[T]}$ can be expressed as

$$\Delta := \mathbb{E}[\widetilde{\boldsymbol{v}}(\xi)] - \widetilde{v}^* \geq 0.$$

This gap cannot be directly evaluated because the optimal value $\widetilde{v}^*$ is not known. Using some lower bound for $\widetilde{v}^*$, Δ can be overestimated though. Such lower bound

is given by $\mathbb{E}[\widetilde{\underline{\boldsymbol{v}}}]$, see (P.11.3). This yields

$$\mathbb{E}[\widetilde{\boldsymbol{v}}(\xi)] - \mathbb{E}[\widetilde{\underline{\boldsymbol{v}}}] \geq \Delta \geq 0. \tag{11.5}$$

Still, the left-hand side of (11.5) is computationally infeasible to evaluate. It requires excessive computational effort to evaluate policy $\big(\boldsymbol{x}_t(\xi_{[t]})\big)_{t\in[T]}$ for all possible scenarios to obtain $\mathbb{E}[\widetilde{\boldsymbol{v}}(\xi)]$. Furthermore, from SDDP only one specific realization of $\widetilde{\underline{v}}$ is known. Therefore, in [48] it is proposed to use estimators for both terms to derive an approximate one-sided confidence interval bounding Δ from above.

Upper Bound Estimation. The SDDP policy $\big(\boldsymbol{x}_t(\xi_{[t]})\big)_{t\in[T]}$ is feasible for the original problem $(\widetilde{P})$, see (P.11.4). Hence, it can be evaluated for any realization of $(\boldsymbol{\xi}_t)_{t\in[T]}$ to assess its out-of-sample performance. Let us sample M_u i.i.d. scenario paths from $(\boldsymbol{\xi}_t)_{t\in[T]}$. For each of those sampled scenarios $\xi^\ell, \ell = 1,\dots,M_u$, the SDDP subproblems (2.10) are solved in forward direction, yielding $x_t(\xi^\ell_{[t]})$ and $\widetilde{v}(\xi^\ell)$ [48]. An upper bound estimator is then defined by the sample mean

$$U_{M_u} := \frac{1}{M_u}\sum_{\ell=1}^{M_u} \widetilde{v}(\xi^\ell). \tag{11.6}$$

Similarly to the in-sample estimator, this estimator is an unbiased and consistent estimator of $\mathbb{E}[\widetilde{\boldsymbol{v}}(\xi)]$. Its sample variance is given by [48]

$$\sigma_U^2 := \frac{1}{M_u - 1}\sum_{\ell=1}^{M_u} (\widetilde{v}(\xi^\ell) - U_{M_u})^2. \tag{11.7}$$

Alternatively, an upper bound estimator can be obtained by sampling a finite number of different SAA problems, and applying the SDDP policy $\big(\boldsymbol{x}_t(\xi_{[t]})\big)_{t\in[T]}$ to each of them [39]. This comes at the cost of increased computational effort.

Lower Bound Estimation with Several SAA Problems. From SDDP, only one single realization of $\widetilde{\underline{v}}$ is known. Hence, it is not directly possible to determine a sampling error for this point estimate and to derive a confidence interval for $\mathbb{E}[\widetilde{\underline{\boldsymbol{v}}}]$.

One approach to derive a lower bound estimator is to solve a finite number of different SAA problems with SDDP and to determine the mean of the lower bounds $\widetilde{\underline{v}}$. To be precise, M_l different SAA problems are constructed, each by sampling $\widehat{q}_t$ realizations per stage from $(\boldsymbol{\xi}_t)_{t\in[T]}$. Then SDDP is run, yielding the lower bounds $\widetilde{\underline{v}}^\ell, \ell = 1,\dots,M_l$ [48]. The sample mean

$$L_{M_l} := \frac{1}{M_l}\sum_{\ell=1}^{M_l} \widetilde{\underline{v}}^\ell \tag{11.8}$$

then defines an estimator for $\mathbb{E}[\widetilde{\underline{\boldsymbol{v}}}]$ with sample variance

$$\sigma_l^2 := \frac{1}{M_l - 1}\sum_{\ell=1}^{M_l} (\widetilde{\underline{v}}^\ell - L_{M_l})^2.$$

Note that instead of lower bounds $\widetilde{\underline{v}}^\ell$, also the optimal values $\widetilde{v}_N^\ell$ could be used in estimator (11.8) [48]. We already discussed in Section 11.1 that it may be computationally intractable to solve one single SAA problem to optimality, though. Thus, using $\widetilde{\underline{v}}^\ell$ may be computationally preferable.

In principle, applying SDDP to not only one, but several SAA problems and building the mean of the obtained bounds seems very reasonable from a statistical perspective, as the outcome of one SAA problem is random. This also has another possible benefit: If SDDP is run for M_l different SAA problems ($\widetilde{P}_N^l$), each of these problems yields a different feasible policy. By calculating the upper bound estimator U_{M_u} (11.6) for each of them, directly M_l different policies could be compared.

However, for problems with multiple stages and for sufficiently high $\widehat{N}_t$, this becomes computationally intractable, even without solving ($\widetilde{P}_N^l$) exactly. Therefore, de Matos et al. [48] follow the strategy to run SDDP once for some SAA problem with larger branch size $\widetilde{q}_t$ to determine a high quality policy and then, afterwards, to run SDDP for M_l SAA problems with smaller branch size $\widehat{q}_t$ only to produce the lower bound estimate L_{M_l} and assess the quality of that policy. In their numerical tests, they choose values between 5 and 200 for $\widetilde{q}_t$ and 5 for $\widehat{q}_t$. In general, it is not clear though, how to choose $\widehat{q}_t$ to reach a reasonable trade-off between computational tractability and an appropriate quality of the lower bound estimator.

Lower Bound Estimation with One SAA Problem. An alternative and less costly lower bound estimator is derived by only using the existing SAA problem, which has been applied to determine the policy that is to be assessed [48].

The idea is then to use the SDDP outcome $\underline{\widetilde{v}}$ as the point estimate L_{M_l} for the lower bound. To estimate the unknown sampling error of $\underline{\widetilde{v}}$, the sampling error of the in-sample upper bound estimator is used. This means that M_l scenarios are sampled from F_N (the SAA problem distribution) and formulas (11.6) and (11.7) with M_l in the role of M_u are used to compute an upper bound estimate $\overline{\widetilde{v}}_{M_l}$ and sample error σ_l^2. The idea behind applying this sampling error is that $\underline{\widetilde{v}}$ and $\mathbb{E}[\overline{\widetilde{\boldsymbol{v}}}_{M_l}]$ are equal if SDDP has been run to optimality. However, this also implies that if SDDP has not converged (or if $\widetilde{q}_t$ is not sufficiently large) the sampling error may be underestimated, and thus the confidence intervals drawn from this become overly optimistic [48].

Confidence Intervals. Using the bound estimators and their sample variances, asymptotically valid confidence intervals can be derived [48].

$$\left(-\infty, U_{M_u} + t_{M_u-1,\alpha}\frac{\sigma_U}{\sqrt{M_u}}\right]$$

is an asymptotically valid, and for finite M_u approximate, $(1-\alpha)\%$ confidence interval for $\mathbb{E}[\widetilde{\boldsymbol{v}}(\xi)]$. Here, $t_{M_u-1,\alpha}$ denotes the $(1-\alpha)$-level quantile of a student's t distribution with $M_u - 1$ degrees of freedom. Similarly,

$$\left[L_{M_l} - t_{M_l-1,\alpha}\frac{\sigma_l}{\sqrt{M_l}}, \infty\right)$$

is an asymptotically valid, and for finite M_l approximate, $(1-\alpha)\%$ confidence interval for $\widetilde{v}^*$. Using only one SAA problem, this confidence interval is only valid if SDDP has converged and if $\widetilde{q}_t$ is sufficiently large. Combining both intervals yields

$$\left[0, [U_{M_u} - L_{M_l}]_+ + t_{M_l-1,\alpha}\frac{\sigma_l}{\sqrt{M_l}} + t_{M_u-1,\alpha}\frac{\sigma_U}{\sqrt{M_u}}\right]$$

as a one-sided approximate confidence interval for the optimality gap Δ [48]. Here, $[x]_+ := \max\{x, 0\}$.

11.3. Variance Reduction Techniques. Instead of MC sampling, also importance sampling [149] and variance reduction techniques (see Section 6.2) can be applied to obtain SAA estimators with reduced bias and variance.

In [104], numerical tests comparing MC, LHS and RQMC indicate that RQMC yields the most promising results when it comes to determining representative SAA problems. In [48] also MC, LHS and RMC are compared for different branch sizes and policy evaluation strategies. The results indicate that with both LHS and RQMC, a reduction of bias and sampling error, a higher policy quality and tighter confidence intervals can be achieved in comparison with MC sampling, especially for smaller branch sizes $\widetilde{q}_t$. For smaller branch sizes LHS appears to be superior, while RQMC yields better results for larger branch sizes. While showing higher variability for MC sampling, if combined with RQMC and LHS sampling, the computationally preferable lower bound estimator using only in-sample scenarios from the existing SAA yields comparable results to the approach solving several SAA problems [48].

12. Risk-averse SDDP [relaxing Assumption 8]. In SDDP, as described in Section 3, a risk-neutral optimal policy is determined for (MSLP) (see Assumption 8). More precisely, (MSLP) minimizes the expectation of the total objective value over all stages $t \in [T]$ over feasible policies $(\boldsymbol{x}_t(\xi_{[t]}))_{t\in[T]}$, which satisfy non-anticipativity and all constraints. Hence, it can be formulated as the single problem (2.3) with objective

$$\min_{x_1, \boldsymbol{x_2}, \ldots, \boldsymbol{x_T}} \mathbb{E}\left[\sum_{t\in[T]} \big(\boldsymbol{c}_t(\xi_t)\big)^\top \boldsymbol{x}_t(\xi_{[t]})\right]. \tag{12.1}$$

As discussed in Section 2.4, this problem can be expressed equivalently using the DPE (2.4)-(2.6). This equivalence is based on two important properties of expected values, first the so-called *tower property*

$$\mathbb{E}_{\boldsymbol{\xi}_t}[\boldsymbol{Z}_t(\xi_t)] = \mathbb{E}_{\xi_{[t-1]}}\big[\mathbb{E}_{\boldsymbol{\xi}_t|\xi_{[t-1]}}[\boldsymbol{Z}_t(\xi_t)]\big] \tag{12.2}$$

for some random variable $\boldsymbol{Z}_t$, and second its strict monotonicity (see property (R2') below for a formal definition) [198].

Recall that the objective value $\sum_{t\in[T]} \big(\boldsymbol{c}_t(\xi_t)\big)^\top \boldsymbol{x}_t(\xi_{[t]})$ is random, and its realizations depend on realizations of $(\boldsymbol{\xi}_t)_{t\in[T]}$. For some specific realization, the SDDP policy may produce an objective value which widely deviates from the expectation in (12.1). In practice, decision makers are often anxious not only to find a policy yielding low costs *on average*, but also to avoid the risk of extremely high cost situations. This motivates to consider *risk-averse* approaches in stochastic programming.

For multistage stochastic programming, incorporating risk-aversion has been a popular research topic in the last decade. This includes theoretical fundamentals on dynamic risk measures [191] as well as algorithmic developments, such as rolling horizon approaches with chance constraints or AVaR constraints, which take risk aversion into account in the constraints of (MSLP) [95, 96]. For SDDP, most focus has been on replacing expectations in the objective (12.1) with some multi-period risk measure $\mathcal{R}[\cdot]$. This yields the risk-averse problem $(P_\mathcal{R})$:

$$\begin{aligned} \min_{x_1, \boldsymbol{x_2}, \ldots, \boldsymbol{x_T}} \quad & \mathcal{R}\left[\sum_{t\in[T]} \big(\boldsymbol{c}_t(\xi_t)\big)^\top \boldsymbol{x}_t(\xi_{[t]})\right] \\ \text{s.t.} \quad & x_1 \in \mathcal{X}_1 \\ & \boldsymbol{x}_t \in \mathcal{X}_t(\boldsymbol{x}_{t-1}(\xi_{[t-1]}), \xi_t) \quad \forall \xi_t \in \Xi_t \ \forall t = 2, \ldots, T. \end{aligned} \tag{12.3}$$

We cover risk-averse SDDP in detail in the remainder of this section, but start with the required foundations of risk measures, especially for multistage problems. As our focus is on algorithmic aspects of SDDP, we refer to the comprehensive coverage of this topic in [198, 200] for technical definitions and derivations.

12.1. Risk Measures. A *static* (or *one-period*) risk measure is a function $\rho : \mathcal{Z} \to \bar{\mathbb{R}}$ from the space $\mathcal{Z}$ of random variables $\boldsymbol{Z}$ to $\bar{R} := \mathbb{R} \cup \{-\infty, +\infty\}$. Often, $\mathcal{Z}$ is assumed to be $\mathcal{L}_1(\Omega, \mathscr{F}, \mathbb{P})$, *i.e.*, the space of all $\mathscr{F}$-measurable functions with finite first moments, as this ensures well-definedness and finiteness of many common risk measures. Importantly, since random variables are functions themselves, risk measures are actually *functionals*. This is sometimes emphasized by calling them *risk functionals* or *risk mapping*.

We summarize some well-known risk measures:

- The expected value $\mathbb{E}[\cdot]$ is the most common risk measure. It is completely risk-neutral.
- The *value-at-risk* $\mathrm{VaR}_\alpha[\cdot]$ to level $\alpha \in (0, 1)$ is defined as the left-side $(1-\alpha)$-quantile of the cumulative distribution of some random variable $\boldsymbol{Z}$:

$$\mathrm{VaR}_\alpha[\boldsymbol{Z}] := \inf\left\{u \in \mathbb{R} \ : \ \mathbb{P}(Z \leq u) \geq 1 - \alpha\right\}. \tag{12.4}$$

  Note that this definition is not used consistently in the literature, and that the RHS of (12.4) may also be defined as $\mathrm{VaR}_{1-\alpha}[\boldsymbol{Z}]$.
- The *average value-at-risk* $\mathrm{AVaR}_\alpha[\cdot]$ to level $\alpha \in (0, 1)$ for some random variable $\boldsymbol{Z}$ is defined by [184]

$$\mathrm{AVaR}_\alpha[\boldsymbol{Z}] := \inf\left\{u \in \mathbb{R} \ : \ u + \frac{1}{\alpha}\mathbb{E}\left[[Z-u]_+\right]\right\}, \tag{12.5}$$

  where $[x]_+$ is defined as $\max\{x, 0\}$. Note that the infimum is always attained in our SDDP setting of finite randomness (Assumption 5) and finite value functions $Q_t(\cdot)$ (see Lemma 2.5).

  Remark 12.1. $\mathrm{AVaR}_\alpha[\cdot]$ is also called *conditional value-at-risk*, *expected shortfall* or *expected tail loss*. In the literature on risk-averse stochastic programming, the first alternative is most frequently used with notation $\mathrm{CVaR}_\alpha[\cdot]$, but to avoid confusion when we introduce conditional risk measures later, we stick to average value-at-risk.

  It can be shown that an equivalent formulation of $\mathrm{AVaR}_\alpha[\boldsymbol{Z}]$ is given by [197]

$$\mathrm{AVaR}_\alpha[\boldsymbol{Z}] = \mathrm{VaR}_\alpha[\boldsymbol{Z}] + \frac{1}{\alpha}\mathbb{E}\Big[\big[\boldsymbol{Z} - \mathrm{VaR}_\alpha[\boldsymbol{Z}]\big]_+\Big], \tag{12.6}$$

  i.e., $u^* = \mathrm{VaR}_\alpha[\boldsymbol{Z}]$ minimizes the RHS in (12.5).
  $\mathrm{AVaR}_\alpha[\cdot]$ has some beneficial properties compared to $\mathrm{VaR}_\alpha[\cdot]$. It does not only consider the probability mass beyond $\mathrm{VaR}_\alpha[\cdot]$, but also its distribution, *e.g.*, if it has fat or long tails. Moreover, it allows to retain convexity of optimization problems, as we discuss later on. $\mathrm{VaR}_\alpha[\cdot]$ and $\mathrm{AVaR}_\alpha[\cdot]$ are illustrated in Figure 10.
- In stochastic programming, often a convex combination of $\mathbb{E}[\cdot]$ and $\mathrm{AVaR}[\cdot]$ is considered, that is

$$\widehat{\rho}_{\alpha,\lambda}[\boldsymbol{Z}] := (1-\lambda)\mathbb{E}[\boldsymbol{Z}] + \lambda\mathrm{AVaR}_\alpha[\boldsymbol{Z}] \tag{12.7}$$

  for some $\lambda \in [0, 1]$. The parameters λ and α control the risk-aversion. Choosing $\lambda = 0$ yields the standard risk-neutral model.

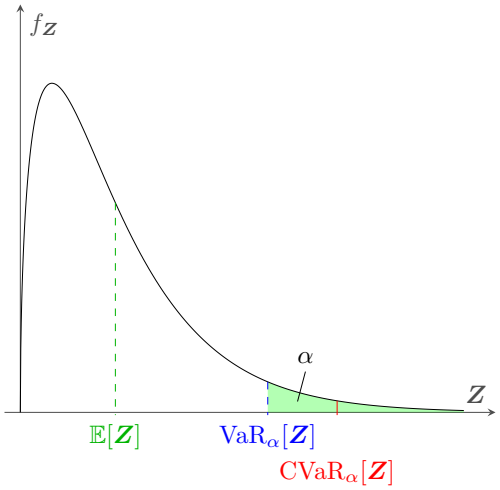


Fig. 10: $\mathrm{VaR}_\alpha[\boldsymbol{Z}]$ and $\mathrm{AVaR}_\alpha[\boldsymbol{Z}]$ for a gamma distributed random variable $\boldsymbol{Z}$.

- For some $\gamma > 0$, the *entropic risk measure* is defined by

$$\mathbb{ENT}_\gamma[\boldsymbol{Z}] := \frac{1}{\gamma} \log\left(\mathbb{E}[e^{\gamma \boldsymbol{Z}}]\right). \tag{12.8}$$

  It generalizes $\mathbb{E}[\cdot]$ (for $\gamma \to 0$) and $\operatorname{ess\,sup}[\cdot]$ (for $\gamma \to \infty$), where $\operatorname{ess\,sup}[\boldsymbol{Z}]$ denotes the essential supremum of a random variable $\boldsymbol{Z}$.

It is often required that risk measures satisfy some special properties, especially in an optimization context. First, we assume that all considered risk measures are proper. Another desired property is *coherence*, a concept introduced by Artzner et al. [3]. We employ a slightly different definition from [200] and state it for the general case of continuous random variables:

DEFINITION 12.2. *A risk measure $\rho : \mathcal{Z} \to \bar{\mathbb{R}}$ is called* coherent, *if it satisfies*
(R1) Convexity: for any $\boldsymbol{Z}_1, \boldsymbol{Z}_2 \in \mathcal{Z}$ and all $\lambda \in [0,1]$ it holds

$$\rho(\lambda \boldsymbol{Z}_1 + (1-\lambda)\boldsymbol{Z}_2) \leq \lambda \rho(\boldsymbol{Z}_1) + (1-\lambda)\rho(\boldsymbol{Z}_2),$$

(R2) Monotonicity: If $\boldsymbol{Z}_1 \leq \boldsymbol{Z}_2$ almost surely, then $\rho(\boldsymbol{Z}_1) \leq \rho(\boldsymbol{Z}_2)$,
(R3) Translation Equivariance: If $a \in \mathbb{R}$ and $\boldsymbol{Z} \in \mathcal{Z}$, then $\rho(\boldsymbol{Z} + a) = \rho(\boldsymbol{Z}) + a$,
(R4) Positive Homogeneity: If $\lambda > 0$ and $\boldsymbol{Z} \in \mathcal{Z}$, then $\rho(\lambda \boldsymbol{Z}) = \lambda \rho(\boldsymbol{Z})$.

A risk measure satisfying only properties (R1), (R2) and (R3) is called *convex*. In fact, a key feature of coherent risk measures is that they are convex, and thus convex objective functions as they appear in $(P_{\mathcal{R}})$ and its DPE remain convex if $\rho[\cdot]$ is applied to them. $\mathrm{VaR}_\alpha[\cdot]$ is not a coherent risk measure, but $\mathrm{AVaR}_\alpha[\cdot]$ is [156]. Therefore, in optimization $\mathrm{AVaR}_\alpha[\cdot]$ is usually preferred over $\mathrm{VaR}_\alpha[\cdot]$.

We introduce some additional relevant properties.

DEFINITION 12.3. *Let $\rho : \mathcal{Z} \to \bar{\mathbb{R}}$ be some risk measure. Then, the following properties can be defined.*
(R2') If the inequalities in (R2) in Definition 12.2 are strict, we call this property strict monotonicity.

Table 4: Properties of common risk measures.

	(R1)	(R2)	(R3)	(R4)	(R2')
$\mathbb{E}[\cdot]$	✓	✓	✓	✓	✓
$\mathrm{VaR}_\alpha[\cdot]$	-	✓	✓	✓	-
$\mathrm{AVaR}_\alpha[\cdot]$	✓	✓	✓	✓	-
$\widehat{\rho}_{\alpha,\lambda}[\cdot]$	✓	✓	✓	✓	✓*
$\mathbb{ENT}_\gamma[\cdot]$	✓	✓	✓	-	✓

* only for $\lambda \in [0, 1)$.

(R5) Law Invariance: ρ is called law invariant *with respect to $\mathbb{P}$, if for all $\boldsymbol{Z}, \boldsymbol{Z}' \in \mathcal{Z}$ with the same distribution also $\rho(\boldsymbol{Z}) = \rho(\boldsymbol{Z}')$ holds.*

Property (R5) implies that the risk measure ρ only depends on the distribution of the considered random variable $\boldsymbol{Z}$.

We summarize properties of the previously introduced risk measures in Table 4.

Remark 12.4. A classical approach in economics is to take risk aversion into account by means of non-decreasing and convex disutility (or concave utility) functions $g : \mathbb{R} \to \bar{\mathbb{R}}$ that are applied to some random variable $\boldsymbol{Z}$ before taking expectations. However, the obtained risk measure $\rho[\boldsymbol{Z}] = \mathbb{E}[g(\boldsymbol{Z})]$ does not satisfy property (R3) which is required to equivalently express $(P_\mathcal{R})$ using DPE.

12.2. Multi-period Risk Measures. In a multistage setting, single-period risk measures have to be extended to several periods, more precisely, to a sequence of random variables $\boldsymbol{Z} := \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_T$, which in our case model the stagewise objectives of (MSLP). We define such multi-period risk measures as functionals $\mathcal{R} : \mathcal{Z} \to \mathbb{R}$ with $\mathcal{Z} = \mathcal{Z}_1 \times \mathcal{Z}_2 \times \ldots \times \mathcal{Z}_T$.

Choosing multi-period risk measures in a reasonable way is a challenging task. Firstly, it is not clear how risk should be measured in a multistage setting [105]. Several different options exist [60, 105, 200], such as

$$\mathcal{R}[\boldsymbol{Z}] = \rho[\boldsymbol{Z}_1 + \cdots + \boldsymbol{Z}_T] \qquad \text{(end-of-horizon risk)} \tag{12.9}$$

$$\mathcal{R}[\boldsymbol{Z}] = \rho\Big[\boldsymbol{Z}_1 + \rho_{|Z_1}\big[\boldsymbol{Z}_2 + \ldots + \rho_{|Z_{T-1}}[\boldsymbol{Z}_T]\cdots\big]\Big] \qquad \text{(nested risk)} \tag{12.10}$$

$$\mathcal{R}[\boldsymbol{Z}] = \rho[\boldsymbol{Z}_1] + \ldots \rho[\boldsymbol{Z}_T] \qquad \text{(stage-wise risk)}, \tag{12.11}$$

where ρ is some one-period risk measure and $\rho_{|Z_t}$ is ρ conditioned on $\mathscr{F}_t$ (or $\xi_{[t-1]}$, respectively), a so-called *conditional* risk measure. The idea of nested conditional risk measures goes back to Ruszczyński and Shapiro [192]. If ρ is law-invariant (property (R5) in Definition 12.3), then $\rho_{|Z_t}$ can be obtained by replacing the distribution with the corresponding conditional distribution [200]. Note that coherence of conditional risk measures can be defined completely analogously to unconditional ones.

Remark 12.5. Under stagewise independence (Assumption 2), as we assume it for SDDP, the conditional risk measures in (12.10) become unconditional ones.

Secondly, in an optimization context, multi-period risk measures have to be carefully chosen, in such a way that the resulting problem $(P_\mathcal{R})$ possesses desirable properties. In addition to convexity, especially time-consistency is a crucial property.

12.2.1. Time Consistency. In the literature, various different definitions of time consistency exist, see among others [35, 105, 46, 157, 198] and references within. The term is ambiguous in the sense that it is used for risk measures, policies and optimization problems. We only state some of these concepts that are relevant for SDDP, and for technical definitions and detailed discussions refer to [64, 105, 198, 200].

A common definition is that an optimal policy $\left(\bar{\boldsymbol{x}}_t(\xi_{[t]})\right)_{t\in[T]}$ for $(P_{\mathcal{R}})$ (see (12.3)) is called *time consistent* if for any $\tau \in [T]$, the policy $\left(\bar{\boldsymbol{x}}_t(\xi_{[t]})\right)_{t=\tau,\dots,T}$ is optimal for $(P_{\mathcal{R}})$ restricted to horizon $t = \tau, \dots, T$ conditional on $\mathscr{F}_{t-1}$ and $\bar{x}_{t-1}$ [200]. This means that the optimal policy remains optimal after some of the uncertain data has been revealed. The problem $(P_{\mathcal{R}})$ is then called *weakly time consistent*, if at least one of its optimal policies is time consistent, or *time consistent*, if every optimal policy is time consistent [200] (note that there exist deviating definitions in the literature).

Policies obtained using DPE (such as (2.4)-(2.6)) naturally satisfy time consistency. Therefore, the concept of time consistency is closely related to equivalently reformulating $(P_{\mathcal{R}})$ (see (12.3)) into DPE [200]. For nested risk measures $\mathcal{R}[\cdot]$, see (12.10), this equivalence holds under strict monotonicity (property (R2') in Definition 12.3) of ρ (or $\rho_{|\xi_{[t]}}$, respectively). More precisely, under (R2'), by interchanging risk measures and minimization operators, $(P_{\mathcal{R}})$ with nested risk can be expressed in the nested fashion [200]

$$\begin{aligned}\min_{x_1\in\mathcal{X}_1} c_1^\top x_1 + \rho_2\Bigg[\min_{\boldsymbol{x}_2\in\mathcal{X}_2(x_1)} (\boldsymbol{c}_2(\xi_2))^\top \boldsymbol{x}_2 + \rho_{3|\xi_{[2]}}\Big[\dots \\ \dots + \rho_{T|\xi_{[T-1]}}\Big[\min_{\boldsymbol{x}_T\in\mathcal{X}_T(x_{T-1})} (\boldsymbol{c}_T(\xi_T))^\top \boldsymbol{x}_T\Big]\dots\Big]\Bigg],\end{aligned} \tag{12.12}$$

which naturally allows for a reformulation to DPE. Note that for stage 2 no conditional expectation is used as the first-stage data is deterministic. If ρ (or $\rho_{|\xi_{[t]}}$) only satisfy (R2) instead of (R2'), then only weak consistency of $(P_{\mathcal{R}})$ is guaranteed, as any optimal policy for the DPE is also optimal for problem $(P_{\mathcal{R}})$ with nested risk, but not necessarily vice versa.

As indicated by Table 4, $\mathrm{AVaR}_\alpha[\cdot]$ is not strictly monotone. Therefore, even if applied in a nested conditional way, time consistency is not assured. In contrast, it can be ensured using risk measure $\widehat{\rho}_{\alpha,\lambda}[\cdot]$ defined in (12.7), given that $\lambda \in [0, 1)$. A drawback of nested risk is that it is less amenable to suitable interpretations, although some economic interpretations are possible [189].

For single-period risk measures $\rho[\cdot]$ that are applied as an end-of-horizon risk measure (12.9), it is well known that time consistency is often not satisfied. For instance, some simple examples in [64, 105] show that using a single-period risk measure $\rho[\cdot]$, such as $\mathrm{VaR}_\alpha[\cdot]$ or $\mathrm{AVaR}_\alpha[\cdot]$, in this setting leads to time-inconsistent decisions. Moreover, in [189], an illustrative example is presented in which even under stagewise independence (Assumption 2), the risk measure $\widehat{\rho}_{\alpha,\lambda}[\cdot]$ does not yield time-consistent policies from an end-of-horizon perspective. To achieve time consistency, it is required that problem $(P_{\mathcal{R}})$ (see (12.3)) with end-of-horizon risk measure $\rho[\cdot]$ can be converted to an equivalent problem with nested risk using conditional risk measures $\rho_{|\xi[t]}$. For this reason, Dowson et al. [60] define time consistency (in their case referred to as *conditional consistency*) of a single-period risk measure $\rho[\cdot]$ as an equivalence between the associated end-of-horizon risk and nested risk.

In fact, the only law-invariant coherent single-period risk measures $\rho[\cdot]$ allowing for such a reformulation are $\mathbb{E}[\cdot]$ and $\operatorname{ess\,sup}[\cdot]$ [200]. Hence, using $\mathrm{AVaR}_\alpha[\cdot]$ as an

end-of-horizon risk measure not even weak time consistency can be guaranteed for $(P_{\mathcal{R}})$. It can be shown, though, that the non-coherent, but convex risk measure $\mathbb{ENT}_{\gamma}[\cdot]$ from (12.8) is conditionally consistent, and thus is sufficient to ensure time consistency of $(P_{\mathcal{R}})$. The equivalence of different formulations for problem $(P_{\mathcal{R}})$ is illustrated in Figure 11.

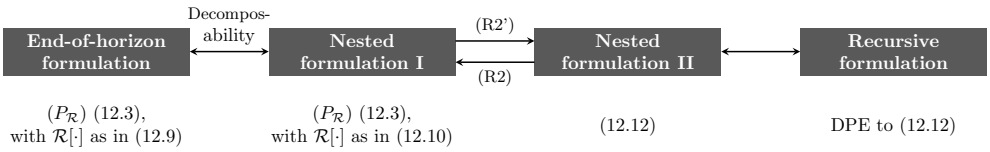


Fig. 11: Different forms of $(P_{\mathcal{R}})$ and conditions for their equivalence.

Remark 12.6. In view of conditional consistency, note that nested risk measures $\mathcal{R}[\cdot]$ from (12.10) can always be expressed equivalently using an associated end-of-horizon risk measure (12.9), the so-called *composite risk measure*. However, as the previous discussion shows, the reverse direction is only true if $\rho[\cdot]$ allows for a decomposition using its conditional analogues; similar to (12.2) [198, 200].

Additionally, some notion of time consistency can be satisfied using *expected conditional risk measures* $\mathcal{R}[\cdot]$, which measure the risk stage by stage (see (12.11)), if the included (conditional) risk measures are coherent [105]. Applying such a risk measure in $(P_{\mathcal{R}})$ (problem (12.3)), we obtain the problem

$$
\begin{aligned}
\min_{x_1, \boldsymbol{x_2}, \ldots, \boldsymbol{x_T}} \quad & c_1^\top x_1 + \rho_2\left[\left(\boldsymbol{c}_2(\xi_2)\right)^\top \boldsymbol{x}_2(\xi_{[2]})\right] + \mathbb{E}_{\xi_{[2]}}\left[\rho_{3|\xi_{[2]}}\left[\left(\boldsymbol{c}_3(\xi_3)\right)^\top \boldsymbol{x}_3(\xi_{[3]})\right]\right] \\
& \qquad + \cdots + \mathbb{E}_{\xi_{[T-1]}}\left[\rho_{T|\xi_{[T-1]}}\left[\left(\boldsymbol{c}_T(\xi_T)\right)^\top \boldsymbol{x}_T(\xi_{[T]})\right]\right] \qquad (12.13) \\
\text{s.t.} \quad & x_1 \in \mathcal{X}_1 \\
& \boldsymbol{x}_t \in \mathcal{X}_t(\boldsymbol{x}_{t-1}(\xi_{[t-1]}), \xi_t) \quad \forall \xi_t \in \Xi_t \ \forall t = 2, \ldots, T.
\end{aligned}
$$

In the remainder of this section, we discuss the incorporation of risk-aversion into SDDP from an algorithmic perspective. The first two methodological studies of risk-averse SDDP are [93] for problems with end-of-horizon risk (12.9), in particular using polyhedral risk measures, and [197] for problems with nested conditional risk mappings (12.10). While some articles on this topic also cover SAA [113, 197, 202], see Section 11, we restrict to finite random variables here.

12.3. SDDP with Polyhedral Risk Measures. Multiperiod polyhedral risk measures are a special type of risk measure, which can be formulated as the optimal value of certain T-stage linear stochastic programs [67]. The arguments of the risk measure, *e.g.*, in our case the objective function of (MSLP), enter these linear programs on the RHS.

In [93], multiperiod extended polyhedral risk measures are introduced, for which the corresponding linear program has a slightly more general form. This class comprises polyhedral risk measures, spectral risk measures and also $\mathrm{AVaR}_{\alpha}[\cdot]$. These risk measures can be shown to be convex and coherent under certain assumptions [93].

The main strength of (extended) polyhedral risk measures is that they can naturally be used in a multistage stochastic programming setting. The LP representation

of $\mathcal{R}[\cdot]$ and the original LP formulation of (MSLP) can be conflated to a single large-scale risk-averse linear programming problem $(P_{\mathcal{R}})$, which allows for a reformulation by means of DPE [93]. These DPE can then be approached by standard risk-neutral SDDP to compute lower bounds and statistical upper bounds for the risk-averse optimal value $v^*_{\mathcal{R}}$. Guigues and Römisch derive the cut formulas and give a convergence proof for some special cases of extended polyhedral risk measures [93] and the special case of spectral risk measures [94].

While polyhedral risk measures allow for a straightforward formulation of linear DPE, they have a significant drawback with respect to SDDP. The stage-t subproblems have to be enhanced with additional state variables z_{t-1} and $y_1, \dots, y_{t-1}$, which are required to store the history of previous decisions. In general, this is unfavorable, see Section 4.2. The specific computational cost depends on the chosen extended polyhedral risk measure. Even if for general extended polyhedral risk measures the augmentation of the state space may yield prohibitive computational cost [160], the approach has been successfully applied with $\mathrm{AVaR}_\alpha[\cdot]$ in [84].

12.4. SDDP with Nested Risk Measures. As mentioned in Section 12.2.1, to obtain a risk-averse problem $(P_{\mathcal{R}})$ with time-consistent solutions, it is often proposed to use (conditional) coherent one-period risk measures $\rho[\cdot]$ (or $\rho_{t|\xi_{[t]}}[\cdot]$) for all $t \in [T]$ in a nested fashion. This yields the nested problem (12.12). We denote its optimal value by $v^*_{\mathcal{R}}$. As indicated before, we can derive an equivalent formulation using DPE [200]. Using Remark 12.5 they become

$$Q_{\mathcal{R},t}(x_{t-1}, \xi_t) := \begin{cases} \min\limits_{x_t} & \big(c_t(\xi_t)\big)^\top x_t + \mathcal{Q}_{\mathcal{R},t+1}(x_t) \\ \text{s.t.} & x_t \in \mathcal{X}_t(x_{t-1}, \xi_t) \end{cases} \tag{12.14}$$

with some *risk-adjusted value function*

$$\mathcal{Q}_{\mathcal{R},t+1}(x_t) := \rho_{t+1}\left[Q_{\mathcal{R},t+1}(x_t, \boldsymbol{\xi}_{t+1})\right] \tag{12.15}$$

and $\mathcal{Q}_{\mathcal{R},T+1}(\cdot) \equiv 0$. The corresponding first-stage problem is

$$v^*_{\mathcal{R}} = \begin{cases} \min\limits_{x_1} & c_1^\top x_1 + \mathcal{Q}_{\mathcal{R},2}(x_1) \\ \text{s.t.} & x_1 \in \mathcal{X}_1. \end{cases} \tag{12.16}$$

Fortunately, for coherent risk measures $\rho_t[\cdot], t \in [T]$, also the nested risk measure $\mathcal{R}[\cdot]$ preserves convexity of $\mathcal{Q}_{\mathcal{R},t+1}(\cdot)$. Therefore, a cutting-plane approximation as in SDDP can be applied.

Nested conditional risk measures are by far the most frequently chosen approach for risk-averse extensions of SDDP [64, 105, 113, 159, 160, 197, 202]. Most typically, the risk measure $\widehat{\rho}_{\alpha,\lambda}[\cdot]$ (see (12.7)) is used, which is coherent according to Table 4.

12.4.1. Reformulating the DPE. The general DPE for $(P_{\mathcal{R}})$ with nested risk measures are formulated in (12.14)-(12.16). To determine $\mathcal{Q}_t(\cdot), t \in [T]$, for $\widehat{\rho}_{\alpha,\lambda}[\cdot]$ specifically, the AVaR of $Q_t(\cdot,\cdot)$ has to be evaluated. Using its definition as the optimal value of an optimization problem with decision variable $u \in \mathbb{R}$ [184], see (12.5), we are able to further reformulate the DPE.

Remark 12.7. For finite random variables $\boldsymbol{Z}$ (under Assumption 5 for SDDP), $\mathrm{AVaR}_\alpha[\cdot]$ may as well be defined as

$$\mathrm{AVaR}_\alpha[\boldsymbol{Z}] = \mathbb{E}\big[\boldsymbol{Z} \big| \boldsymbol{Z} \geq \mathrm{VaR}_\alpha[\boldsymbol{Z}]\big].$$

However, to reformulate the DPE, representation (12.5) is preferrable.

Additional State Variable Approach. Using (12.5), the risk-adjusted value function (12.15) can be expressed as

$$\begin{aligned}\mathcal{Q}_{\mathcal{R},t+1}(x_t) = \min_{u_t\in\mathbb{R}} \; \mathbb{E}_{\boldsymbol{\xi}_{t+1}}\Big[&(1-\lambda_{t+1})Q_{\mathcal{R},t+1}(x_t,\boldsymbol{\xi}_{t+1}) \\ &+ \lambda_{t+1}\Big(u_t + \frac{1}{\alpha_{t+1}}\big[Q_{\mathcal{R},t+1}(x_t,\boldsymbol{\xi}_{t+1}) - u_t\big]_+\Big)\Big].\end{aligned} \tag{12.17}$$

Recall that λ_t and α_t, $t = 2,\dots,T$, are user-controlled parameters.

The minimization over u_t can be incorporated into the stage-t subproblems [197], which yields

$$\widetilde{Q}_{\mathcal{R},t}(x_{t-1},\xi_t) = \begin{cases}\min\limits_{x_t,u_t} & \big(c_t(\xi_t)\big)^\top x_t + \lambda_{t+1}u_t + \widetilde{\mathcal{Q}}_{\mathcal{R},t+1}(x_t,u_t) \\ \text{s.t.} & x_t \in \mathcal{X}_t(x_{t-1},\xi_t)\end{cases} \tag{12.18}$$

with some modified risk-adjusted value function

$$\begin{aligned}\widetilde{\mathcal{Q}}_{\mathcal{R},t+1}(x_t,u_t) = \mathbb{E}_{\boldsymbol{\xi}_{t+1}}\Big[&(1-\lambda_{t+1})\widetilde{Q}_{\mathcal{R},t+1}(x_t,\boldsymbol{\xi}_{t+1}) \\ &+ \frac{\lambda_{t+1}}{\alpha_{t+1}}\big[\widetilde{Q}_{\mathcal{R},t+1}(x_t,\boldsymbol{\xi}_{t+1}) - u_t\big]_+\Big],\end{aligned} \tag{12.19}$$

$\widetilde{\mathcal{Q}}_{\mathcal{R},T+1}(\cdot,\cdot) \equiv 0$ and $\lambda_{T+1} \equiv 0$ [197]. The first stage changes to

$$v^*_{\mathcal{R}} = \begin{cases}\min\limits_{x_1,u_1} & c_1^\top x_1 + \lambda_2 u_1 + \widetilde{\mathcal{Q}}_{\mathcal{R},2}(x_1,u_1) \\ \text{s.t.} & x_1 \in \mathcal{X}_1.\end{cases} \tag{12.20}$$

The risk-adjusted value functions $\widetilde{\mathcal{Q}}_{\mathcal{R},t+1}(\cdot,\cdot)$ differ from the ones defined in (12.17), but can be proven to be convex as well.

With equations (12.18)-(12.20), the risk measures $\rho_t[\cdot]$ are incorporated into the subproblems, such that only expectations have to be evaluated in the DPE. However, as pointed out in [113], in comparison with the DPE (2.4)-(2.6) of the risk-neutral case, we still observe some fundamental differences: Firstly, an additional, albeit one-dimensional, state variable $u_t \in \mathbb{R}$ is introduced at each stage to estimate the VaR-level, augmenting the state space by one. Secondly, the risk-adjusted value functions $\mathcal{Q}_{\mathcal{R},t+1}(\cdot,\cdot)$ do not only depend on x_t, but also on u_t and parameters λ_t, α_t. Thirdly, they contain the nonlinear, *i.e.*, piecewise linear, function $[\cdot]_+$.

Philpott and de Matos provide an alternative reformulation of the DPE, eliminating the nonlinear expression via an epigraph reformulation [159]. To this end, the random term in the brackets in (12.19) is fully incorporated into the value functions. For $t = 2,\dots,T-1$, this yields

$$\begin{aligned}&\widehat{Q}_{\mathcal{R},t}(x_{t-1},u_{t-1},\xi_t) \\ &= \begin{cases}\min\limits_{x_t,u_t,w_t} & (1-\lambda_t)\Big(\big(c_t(\xi_t)\big)^\top x_t + \lambda_{t+1}u_t + \widehat{\mathcal{Q}}_{\mathcal{R},t+1}(x_t,u_t)\Big) + \dfrac{\lambda_t}{\alpha_t}w_t \\ \text{s.t.} & x_t \in \mathcal{X}_t(x_{t-1},\xi_t) \\ & w_t - \big(c_t(\xi_t)\big)^\top x_t - \lambda_{t+1}u_t - \widehat{\mathcal{Q}}_{\mathcal{R},t+1}(x_t,u_t) \geq -u_{t-1}.\end{cases}\end{aligned} \tag{12.21}$$

Using this formulation, the risk value function is defined more naturally as

$$\widehat{\mathcal{Q}}_{\mathcal{R},t+1}(x_t,u_t) = \mathbb{E}_{\boldsymbol{\xi}_{t+1}}\Big[\widehat{Q}_{\mathcal{R},t+1}(x_t,u_t,\boldsymbol{\xi}_{t+1})\Big]. \tag{12.22}$$

Again, $\widehat{\mathcal{Q}}_{\mathcal{R},T+1}(\cdot,\cdot) \equiv 0$ and $\lambda_{T+1} \equiv 0$.

The first-stage problem reads then

$$v^*_{\mathcal{R}} = \begin{cases} \min\limits_{x_1,u_1,w_1} & c_1^\top x_1 + \lambda_2 u_1 + \widehat{\mathcal{Q}}_{\mathcal{R},2}(x_1,u_1) \\ \text{s.t.} & x_1 \in \mathcal{X}_1. \end{cases} \tag{12.23}$$

In comparison to the formulation (12.18)-(12.20) by Shapiro [197], additional variables and constraints have to be introduced. Both formulations allow application of SDDP, but share the drawback of augmenting the state space. Since the computational effort of SDDP grows exponentially in the state space dimension, see Theorem 4.2, such increase should be avoided.

Modifying the Probability Measure. An alternative idea is to exploit that $u^* = \mathrm{VaR}_\alpha[\boldsymbol{Z}]$ in the definition of $\mathrm{AVaR}_\alpha[\boldsymbol{Z}]$ (see (12.5)) and that $\mathrm{VaR}_\alpha[\boldsymbol{Z}]$ is the $(1-\alpha)$-quantile of a random variable $\boldsymbol{Z}$. As we assume finite randomness (Assumption 5) and solve the subproblems for all realizations $\xi_{tj}, j = 1,\dots,q_t$, in the backward pass of SDDP, this quantile can be manually determined for the value functions [202].

Without loss of generality, assume that for all $t = 2,\dots,T$ and any fixed trial solution $\bar{x}_{t-1}$ the values of $Q_{\mathcal{R},t}(\bar{x}_{t-1},\xi_{tj})$ are ordered for all $j = 1,\dots,q_t$. That means, we have $Q_{\mathcal{R},t}(\bar{x}_{t-1},\xi_{t1}) \le \dots \le Q_{\mathcal{R},t}(\bar{x}_{t-1},\xi_{t,q_t})$. Then, in (12.17) the variable u_t can be replaced by the $(1-\alpha)$-quantile $Q_{\mathcal{R},t+1}(\bar{x}_t,\xi_{t+1,j^*})$ with j^* chosen such that $\sum_{j=1}^{j^*} p_{t+1,j} \ge 1-\alpha_{t+1}$:

$$\begin{aligned} \mathcal{Q}_{\mathcal{R},t+1}(x_t) = \mathbb{E}_{\boldsymbol{\xi}_{t+1}}\Big[&(1-\lambda_{t+1})Q_{\mathcal{R},t+1}(x_t,\boldsymbol{\xi}_{t+1}) + \lambda_{t+1}\Big(Q_{\mathcal{R},t+1}(\bar{x}_t,\xi_{t+1,j^*}) \\ &+ \frac{1}{\alpha_t}\big[Q_{\mathcal{R},t+1}(x_t,\boldsymbol{\xi}_{t+1}) - Q_{\mathcal{R},t+1}(\bar{x}_t,\xi_{t+1,j^*})\big]_+\Big)\Big]. \end{aligned} \tag{12.24}$$

In SDDP, relation (12.24) cannot directly be applied, since $Q_{\mathcal{R},t+1}(\cdot,\xi_{t+1,j})$ is not known and also not evaluated for all $j = 1,\dots,q_{t+1}$. However, the same principle can also be applied to the approximate value functions $\underline{Q}_{\mathcal{R},t+1}(\cdot,\xi_{t+1,j})$.

In [160], this idea is considered from a dual perspective and used to reformulate the risk measure (12.7) even before formulating the DPE. The key concept is the so-called *dual representation* of $\mathrm{AVaR}_\alpha[\cdot]$, which is defined as

$$\mathrm{AVaR}_\alpha[\boldsymbol{Z}] = \begin{cases} \sup\limits_{\zeta} & \sum\limits_{j=1}^{q} p_j\zeta_j Z(\xi_j) \\ \text{s.t.} & \sum\limits_{j=1}^{q} p_j\zeta_j = 1 \\ & \zeta_j \ge 0, \quad j = 1,\dots,q \\ & \zeta_j \le \dfrac{1}{\alpha}, \quad j = 1,\dots,q. \end{cases} \tag{12.25}$$

It shows that $\mathrm{AVaR}_\alpha[\cdot]$ can be interpreted as some worst-case probability measure $\widetilde{\mathbb{P}}$ with $\widetilde{p}_j := p_j\zeta_j$ for all $j = 1,\dots,q$.

As shown in [160], using this definition and explicitly computing the supremum, risk measure (12.7) can be written as

$$\widehat{\rho}_{t,\alpha_t,\lambda_t}[\boldsymbol{Z}] = \sum_{j=1}^{q_t} p_{tj}\zeta_{tj} Z(\xi_{tj}) \tag{12.26}$$

with

$$\zeta_{tj} = \begin{cases} (1-\lambda_t), & j < j^*, \\ (1-\lambda_t) + \dfrac{1}{p_{tj^*}} \Big(\lambda_t - \dfrac{\lambda_t}{\alpha_t} \displaystyle\sum_{n=j^*+1}^{q_t} p_{tn} \Big), & j = j^*, \\ (1-\lambda_t) + \dfrac{\lambda}{\alpha_t}, & j > j^*. \end{cases} \tag{12.27}$$

Again, note that the true value functions $Q_t(\cdot)$ are not known explicitly in advance, and therefore the worst-case probability measure $\widetilde{\mathbb{P}}$ stemming from (12.25) is not known either. However, it can be approximated in SDDP. In particular, the DPE (12.14)-(12.16) and their approximations can be used with expectations as in standard SDDP, but with a modified probability measure that is iteratively updated. More precisely, as ζ_{tj} changes with $\bar{x}_{t-1}$, the modified probabilities have to be recomputed for each stage t, iteration i and sample k in SDDP. This principle is also extended to general coherent risk measures in [160].

Recently, this kind of change of the probability measure has also been discussed in [125]. Instead of determining the ordering and j^* based on $\underline{Q}_{t+1}^{i+1}(\cdot)$ for one specific iteration i, also all previous iterations are taken into account there. More precisely, the number of iterations in which an index j exceeds $\mathrm{VaR}_\alpha[\underline{Q}_{\mathcal{R},t+1}(\bar{x}_t, \boldsymbol{\xi}_t)]$ are counted. This is considered as a good proxy for the ordering of the actual value functions. The ordering, and thus the probability measure $\widetilde{\mathbb{P}}$, can either be updated dynamically within SDDP or be determined by running risk-averse SDDP once in advance to identify the outcomes contributing to $\mathrm{AVaR}_\alpha[\cdot]$. The latter approach has the advantage that the changed probability measure $\widetilde{\mathbb{P}}$ can be fixed for the following run, which yields a *risk-neutral* problem and allows for application of standard SDDP.

Additionally, as pointed out in [125], the approximation of $\widetilde{\mathbb{P}}$ may also be used in the forward pass to sample scenarios with "bad" outcomes with higher probability. This *biased sampling* can be considered similar to the importance sampling techniques presented in Section 6.

For the third-stage of Example 3.3, the expected risk value function $\mathcal{Q}_{\mathcal{R},3}(\cdot)$ obtained by applying (12.26) and (12.27) to (12.15) is illustrated in Figure 12 for $\alpha = 0.05$ and different values of λ. It can be seen that with choosing larger values for λ, representing a higher risk-aversion, the stage-3 cost increases compared to the risk-neutral case ($\lambda = 0$).

As an overview, the different forms of DPE for $(P_\mathcal{R})$ using a nested (conditional) risk measure based on $\widehat{\rho}_{\alpha,\lambda}[\cdot]$ are summarized in Table 5.

12.4.2. Forward and Backward Pass. All approaches in Table 5 to formulate the DPE allow for a solution of a risk-averse problem $(P_\mathcal{R})$ using SDDP. Some approaches are more efficient, since the state space, the decision space or the number of constraints are not augmented. Others are advantageous in the sense that $\mathcal{Q}_{\mathcal{R},t}(\cdot)$ is expressed by a neat formula, and thus cut formulas can be derived more easily. With some epigraph reformulation, for all the approaches all subproblems can be formulated as LPs.

The forward pass of SDDP basically remains the same as for risk-neutral SDDP from Section 3. That is, $k \in \mathcal{K}$ scenarios are sampled and considered, with $\mathcal{K} \subset \mathcal{S}$ and $|\mathcal{K}| \ll |\mathcal{S}|$. However, the subproblems and the associated approximate value functions $\underline{Q}_{\mathcal{R},t}^{i}(x_{t-1}^{ik}, \xi_t^k)$ differ from the risk-neutral case. Instead of subproblems (2.10), one

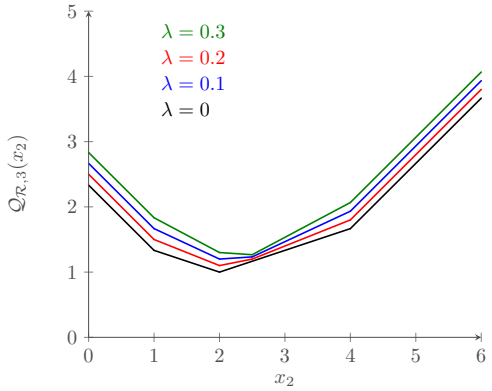


Fig. 12: $\mathcal{Q}_{\mathcal{R},3}(\cdot)$ from Example 3.3 for $\alpha = 0.05$ and different values of λ.

Description	Source	DPE
- general		(12.14)-(12.16)
- augmented state, sophisticated formula for $\mathcal{Q}_{\mathcal{R},t}(\cdot)$	[197]	(12.18)-(12.20)
- augmented state, additional constraints and variables	[159]	(12.21)-(12.23)
- $\text{VaR}_{\alpha_t}[Q_t(\cdot)]$ explicitly determined, sophisticated formula for $\mathcal{Q}_{\mathcal{R},t}(\cdot)$	[202]	(12.18), (12.20), (12.24)
- modified probability measure	[160]	(12.14)-(12.16), (12.26)-(12.27)
- modified probability measure	[125]	(12.18), (12.20), (12.24)

Table 5: DPE formulations for $(P_{\mathcal{R}})$ using a nested (conditional) risk measure based on $\widehat{\rho}_{\alpha,\lambda}[\cdot]$.

of the DPE from Table 5 are chosen and the occurring risk-adjusted value functions $\mathcal{Q}_{\mathcal{R},t+1}(\cdot)$ are replaced by cut approximations $\mathfrak{Q}^i_{\mathcal{R},t+1}(\cdot)$.

In the backward pass, as in risk-neutral SDDP, at each stage $t = T, \ldots, 2$, those subproblems are solved for each trial solution $x^{ik}_{t-1}, k \in \mathcal{K}$, and possible stage-$t$ realization $\xi^k_{tj} \equiv \xi_{tj}, j = 1, \ldots, q_t$, using an updated cut approximation $\mathfrak{Q}^{i+1}_{\mathcal{R},t+1}(\cdot)$. On stage t, a new cut for $\mathcal{Q}_{\mathcal{R},t}(\cdot)$ is derived and handed back to stage $t-1$. The main difference to risk-neutral SDDP is again the definition of $\mathcal{Q}_{\mathcal{R},t}(\cdot)$. Therefore, the cut formulas have to be adapted to the individual approach chosen. For the technical derivation of subgradients in such cases, we refer to the references in Table 5.

12.4.3. Upper Bound Determination and Stopping. The main challenge of applying SDDP is to determine upper bounds for $v^*_{\mathcal{R}}$, and allowing for a reasonable stopping criterion. The reason is that most upper bound construction methods from the risk-neutral case, see Sections 7 and 8, cannot be efficiently extended to the risk-averse case.

Recall that in the risk-neutral case, a feasible policy $(\boldsymbol{x}_t(\xi_{[t]}))_{t\in[T]}$ is determined in the backward pass and evaluated in the forward pass for different scenarios $k \in \mathcal{K}$, yielding a sequence of trial points $(x^{ik}_t)_{t\in[T]}$. Then, a statistical upper bound $\overline{v}_{\mathcal{K}}$ for

v^* is determined as the sample average of the objective values of all these sample paths ξ^k, see (3.9). Analogously, a true upper bound $\overline{v}$ can be obtained by taking the expectation of such objective value for all scenarios $\xi^s, s \in \mathcal{S}$.

However, this is possible only due to the tower property (12.2) of expected values, which is required for the equivalence of the end-of-horizon formulation (12.1) and the nested formulation (12.12), see the discussion in Section 12.2.1. While such an equivalence can be established for nested risk measures $\mathcal{R}[\cdot]$ based on conditional one-period risk measures $\rho_{\xi_{[t]}}[\cdot]$, if their composite risk measure is considered as end-of-horizon risk measure, see Remark 12.6, it usually does not hold for using $\rho[\cdot]$ itself as an end-of-horizon risk measure. The composite risk measure, on the other hand, is usually not known explicitly [200]. Therefore, an analogue to the sample average cannot directly be applied to obtain a (statistical) upper bound.

As determining reasonable upper bounds is a crucial ingredient of SDDP, developing appropriate upper bound estimators has been an active research field in the last decade. In the following, we discuss different approaches that have been proposed. In reviewing them, we follow the presentation of Kozmík and Morton [113], who provide a comprehensive study within their own work on upper bound estimators.

A Sample Average Estimator. In Section 12.4.1, we managed to formulate each $\rho_t[\cdot]$ only by means of expectations in (12.19). Still, this does not assure the tower property, since the risk-adjusted value functions $\mathcal{Q}_{\mathcal{R},t}(\cdot)$ contain a nested nonlinearity due to the $[\cdot]_+$-function. However, we can derive an estimator similar to (3.9) [113]. To this end, we remove the expectation in (12.19) to obtain

$$\begin{aligned}\hat{v}_t(\xi_t^k) := (1-\lambda_t)\Big(\big(c_t(\xi_t^k)\big)^\top x_t^k + \hat{v}_{t+1}(\xi_t^k)\Big) \\ + \lambda_t u_{t-1}^k + \frac{\lambda_t}{\alpha_t}\Big[\big(c_t(\xi_t^k)\big)^\top x_t^k + \hat{v}_{t+1}(\xi_t^k) - u_{t-1}^k\Big]_+,\end{aligned} \tag{12.28}$$

where we replace the value functions $Q_{\mathcal{R},t+1}(\cdot)$ by the estimator of the following stage. For stage T it follows $\hat{v}_{T+1}(\xi_T^k) \equiv 0$ and for the first stage

$$\hat{v}(\xi^k) := c_1^\top x_1 + \hat{v}_2(\xi_1^k). \tag{12.29}$$

Equation (12.29) provides a recursive estimator for the cost associated with sample path ξ^k. This estimator has to be evaluated by backward recursion starting with stage T. Importantly, formula (12.28) is only used for upper bound estimation, whereas the forward and backward problems in SDDP are still based on the original DPE (12.18)-(12.20). Determining estimator (12.29) for all scenarios $\xi^k, k \in \mathcal{K}$, sampled in the forward pass of SDDP, we can form an upper bound estimator

$$U^n := \frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}} \hat{v}(\xi^k), \tag{12.30}$$

which resembles the sample average estimator (3.9).

Estimator (12.30) is an unbiased and consistent estimator of $\overline{v}_{\mathcal{R}}$, but has a large variance. Kozmík and Morton point out several reasons for this behaviour [113]. Only a small portion of the sampled scenarios contributes to estimating $\mathrm{AVaR}_\alpha[\cdot]$, while most solely contribute to the expectation. Therefore, a very large number of scenarios would be required for an appropriate estimate. Additionally, since expectations are not taken conditionally on each stage as in (12.19), and due to to division by $\alpha_t \in (0,1)$, small or large values are very likely to propagate from late to earlier stages in the recursion to determine $\hat{v}(\xi^k)$.

Conditional Sampling Estimator. For the above reasons, estimator U^n in (12.30) is rarely considered in the literature on risk-averse SDDP. Instead, Shapiro discusses a conditional sampling estimator [197]. Here, the idea is to estimate the expectations (12.19) in the nested structure conditionally by sampling on each stage. Since in principle, the upper bound estimator can be determined independently of the scenarios sampled in the forward pass, we denote the set of samples by $\mathcal{M}$ instead of $\mathcal{K}$. $\mathcal{M}_t$ denotes the corresponding scenario set for stage t.

For each stage, $t = 2, \dots, T$, this yields [113]

$$\hat{v}_t^c(\xi_t^k) := \frac{1}{|\mathcal{M}_t|} \sum_{m_t \in \mathcal{M}_t} \Bigg[(1-\lambda_t)\Big(\big(c_t(\xi_t^{m_t})\big)^\top x_t^{m_t} + \hat{v}_{t+1}^c(\xi_t^{m_t}) \Big) + \lambda_t u_{t-1}^{m_t} + \frac{\lambda_t}{\alpha_t} \Big[\big(c_t(\xi_t^{m_t})\big)^\top x_t^{m_t} + \hat{v}_{t+1}^c(\xi_t^{m_t}) - u_{t-1}^{m_t} \Big]_+ \Bigg],$$

and for the first stage the estimator

$$U^c := c_1^\top x_1 + \hat{v}_2^c(\xi_1).$$

As Shapiro himself points out, this estimator has two significant drawbacks. It requires $\prod_{t=2}^{T} |\mathcal{M}_t| + 1$ subproblems to be solved, which is exponentially growing in the number of stages. Moreover, the obtained upper bounds are typically not very tight. Therefore, estimator U^c should not be useful for large-scale problems [113].

Importance Sampling Estimators. Applying conditional sampling appears computationally intractable, but the drawbacks of estimator U^n may also be addressed by importance sampling [112, 113], see Section 6 for an introduction. By sampling scenarios associated with $\text{AVaR}_\alpha[\cdot]$ with higher importance, it is possible to better represent it. Based on this idea, Kozmík and Morton put forward different importance sampling upper bound estimators [113], which are further enhanced in [112].

Using importance sampling with respect to $\text{AVaR}_\alpha[\cdot]$, yields a considerable challenge, though. In order to determine the importance sampling distribution for some stage t, it has to be identified which scenarios are associated with $\text{AVaR}_\alpha[\cdot]$ on that stage, *i.e.*, which of them provide a value $Q_{\mathcal{R},t}(x_{t-1}^k, \xi_{tj}^k)$ beyond the $(1-\alpha)$-quantile. If we estimate this by solving subproblems for several ξ_{tj}^k and determining $Q_{\mathcal{R},t}(x_{t-1}^k, \xi_{tj}^k)$, we face a similar computational burden as for conditional sampling.

Kozmík and Morton propose the following approach: They use an *approximation function* $d_t(x_{t-1}, \xi_t)$, which estimates the recourse value of the decisions x_{t-1} after ξ_t has been observed [113]. Instead of solving the subproblems for several ξ_{tj}^k, they simply evaluate $d_t(x_{t-1}^k, \xi_{tj}^k)$ and sort these values. Based on the obtained order, it can be decided then which scenarios are used to estimate $\text{AVaR}_\alpha[\cdot]$, *i.e.*, $u_d := \text{VaR}_{\alpha_t}[d_t(x_{t-1}, \boldsymbol{\xi}_t)]$ is determined.

This allows defining an importance sampling distribution depending on x_{t-1} [113]. For simplicity, we assume that all scenarios are equally likely in the original distribution, that is, $f_t(\xi_{tj}) = \frac{1}{q_t}$ for all $j = 1, \dots, q_t$. Then, it follows:

$$g_t(\boldsymbol{\xi}_t | x_{t-1}) := \begin{cases} \dfrac{1}{2\lfloor \alpha_t q_t \rfloor}, & d_t(x_{t-1}, \boldsymbol{\xi}_t) \geq u_t, \\ \dfrac{1}{2(q_t - \lfloor \alpha_t q_t \rfloor)}, & d_t(x_{t-1}, \boldsymbol{\xi}_t) < u_t. \end{cases}$$

This distribution ensures that it is equally likely to draw sample observations above

and below u_t. Note that the formula presented in [113] looks a bit different, since it is presented in the context of SAA.

Defining weights

$$\Lambda_t(\boldsymbol{\xi}_t|x_{t-1}) := \frac{f_t(\boldsymbol{\xi}_t)}{g_t(\boldsymbol{\xi}_t|x_{t-1})}$$

and multiplying them along the sample paths

$$\Lambda(\xi^k) := \prod_{t=2}^{T} \Lambda_t(\xi_t^k|x_{t-1})$$

we can derive the estimator

$$U^i := \frac{1}{\sum_{k\in\mathcal{K}} \Lambda(\xi^k)} \sum_{k\in\mathcal{K}} \Lambda(\xi^k)\hat{v}(\xi^k). \tag{12.31}$$

This estimator is similar to (12.30), as the same recursive term $\hat{v}(\xi^k)$ is used, but combined with importance instead of standard MC sampling.

With the assumptions of relatively complete recourse (based on Assumption 9) and stagewise independence (Assumption 2), estimator (12.31) is asymptotically valid, *i.e.*, for $|\mathcal{K}| \to \infty$, U^i converges to $\mathbb{E}_f[\hat{\boldsymbol{v}}(\xi)]$ with probability 1 and $\mathbb{E}_f[\hat{\boldsymbol{v}}(\xi)] \geq v^*_{\mathcal{R}}$. Moreover, for sufficiently good choice of $d_t(\cdot)$, it can be expected that the variance is lower than for U^n [113].

Based on this idea, even better estimators can be developed [112, 113], for example by sampling with higher importance scenarios associated with $\text{AVaR}_\alpha[\cdot]$ as before, but also using only such scenarios to estimate AVaR, which contribute to the $[\cdot]_+$-term [113]:

$$\begin{aligned}\hat{v}_t^d(\xi_t^k) := (1-\lambda_t)\Big(\big(c_t(\xi_t^k)\big)^\top x_t^k + \hat{v}_{t+1}^d(\xi_t^k)\Big) \\ + \lambda_t u_{t-1}^k + \mathcal{I}[d_t(x_{t-1},\xi_t) \geq u_d]\frac{\lambda_t}{\alpha_{t-1}}\Big[\big(c_t(\xi_t^k)\big)^\top x_t^k + \hat{v}_{t+1}^d(\xi_t^k) - u_{t-1}^k\Big]_+.\end{aligned}$$

Here $\mathcal{I}[\cdot]$ denotes an indicator function. For the first stage it follows

$$\hat{v}^d(\xi^k) := c_1^\top x_1 + \hat{v}_2^d(\xi_1^k).$$

Combining this with (11.6), we obtain

$$U^d := \frac{1}{\sum_{k\in\mathcal{K}} \Lambda(\xi^k)} \sum_{k\in\mathcal{K}} \Lambda(\xi^k)\hat{v}^d(\xi^k).$$

The practical applicability of this estimator relies heavily on satisfaction of the following goodness assumption with respect to $d_t(\cdot)$:

$$Q_{\mathcal{R},t}(x_{t-1},\xi_t) \geq \text{VaR}_{\alpha_t}[Q_{\mathcal{R},t}(x_{t-1},\xi_t)] \;\Leftrightarrow\; d_t(x_{t-1},\xi_t) \geq \text{VaR}_{\alpha_t}[d_t(x_{t-1},\xi_t)],$$

which means that $d_t(\cdot)$ correctly classifies whether a realization is in the upper α-tail of the recourse value distribution.

It is proven that this estimator is asymptotically valid as well, but also provides tighter upper bounds than U^i in expectation, as long as the above goodness assumption is satisfied. Moreover, a smaller variance should be expected [113]. Numerical

results in [113] illustrate that even for a medium number of stages, estimator U^d provides significantly better upper bounds than U^n, U^c and U^i and that also the variance of the estimators is reduced significantly.

Apart from the above sampling estimators, some completely different strategies may be used to obtain upper bounds for $v^*_{\mathcal{R},t}$ or to define some stopping criteria for SDDP in the risk-averse case.

Using Deterministic Upper Bounds. As already discussed in Section 8, we may circumvent the determination of sampling-based upper bound estimators completely if we resort to deterministic upper bounding procedures.

To this end, Philpott et al. [160] extend their inner approximation based upper bounding procedure from Section 8 to the risk-averse case with nested (conditional) coherent risk measures. The main downside of this procedure, to require prohibitively large computational effort for a large number of state variables and an increasing number of cuts, also holds in this case, though.

The alternative deterministic upper bounding procedure based on dual SDDP [97, 118] has been extended to a risk-averse setting as well [40].

Determining Bad Outcomes in Advance. As discussed in Section 12.4.1, following the approach of a change of probability measure, see (12.14)-(12.16) and (12.26), it is also possible to run (risk-averse) SDDP once in advance to approximate the probability measure $\widetilde{\mathbb{P}}$, and then a second time, this time fixing the probability measure to the approximation of $\widetilde{\mathbb{P}}$. This is referred to as solving the *change-of-measure risk-neutral problem* in [125]. Whereas this approach has a lot of computational overhead, the advantage is that a risk-neutral problem can be solved by SDDP and therefore, also the standard stopping, upper bounding and policy assessment techniques can be applied. Clearly, solving the change-of-measure risk-neutral problem is not guaranteed to yield optimal policies for $(P_{\mathcal{R}})$, however Liu and Shapiro report that the quality of the policies is similar to those obtained by risk-averse SDDP [125].

Fixing the Number of Iterations. This approach is proposed by Philpott and de Matos [159]. They run a risk-neutral variant of SDDP first and then fix the number of iterations required until termination. The same number of iterations is then used in the risk-averse case, avoiding the challenge of upper bound evaluation.

In some practical applications, in which it is computationally intractable to determine a sophisticated upper bound estimator, this approach may be useful. Promising results are reported in [159]. However, there is no theoretical guarantee to find a sufficiently good solution for a risk-averse version of $(P_{\mathcal{R}})$ in the same number of iterations as for a risk-neutral version. Additionally, for large problems it may already take considerably long to run SDDP one time. Running it a second time for risk-averse problem $(P_{\mathcal{R}})$ may partially annihilate the computational advantage of avoiding upper bound estimation.

Lower Bound Stabilization. As for risk-neutral SDDP, instead of using upper bounds at all, the algorithm can be terminated, once the lower bounds $\underline{v}^i_{\mathcal{R}}$ stabilize. This provides no convergence guarantee but may be worthwhile in large-scale practical applications where other approaches become computationally prohibitive.

Using Benefit Factors. Instead of the lower bounds $\underline{v}^i_{\mathcal{R}}$, it is also possible to condition termination of SDDP on the improvements of the cut approximations $\mathfrak{Q}^i_{\mathcal{R},t}(\cdot), t = 2, \ldots, T$. For that purpose, Brandi et al. define a benefit factor

$$\mathcal{B}^i_{t,k} = \min\left\{1, \frac{\delta(x^{ik}_{t-1})}{\delta^i_{t,\max}}\right\},$$

which determines how much a new cut improves the current cut approximation $\mathfrak{Q}^i_{\mathcal{R},t}(\cdot)$ at x^{ik}_{t-1} [30]. $\delta(x^{ik}_{t-1})$ is the absolute increase, while $\delta^i_{t,\max}$ is a proxy for the maximum improvement possible. For each sample path $k \in \mathcal{K}$, a total benefit factor can be determined by

$$\mathcal{B}^i_k = \max\left\{\mathcal{B}^i_{2,k}, \mathcal{B}^i_{3,k}, \ldots, \mathcal{B}^i_{T,k}\right\}.$$

The risk-averse SDDP method is then stopped if the values $\mathcal{B}^i_k$ for all $k \in \mathcal{K}$ are below a predefined tolerance, either for one iteration or, alternatively and more robustly, for a predefined larger number of iterations.

12.5. SDDP with Entropic Risk Measure. As discussed before, nested risk measures come with some drawbacks. Computation-wise, upper bound determination is very challenging. Additionally, applying a standard single-period risk measure $\rho[\cdot]$, *e.g.*, $\mathrm{AVaR}_\alpha[\cdot]$, as an end-of-horizon risk measure (12.9) and (possibly conditionally) in a nested risk measure (12.10) does not yield equivalent policies [60] (this is only the case if we take the composite risk measure associated with the nested risk measure as end-of-horizon risk; however, this risk measure is usually not known explicitly, see Remark 12.6). This makes nested risk measures difficult to interpret from an end-of-horizon perspective.

For this reason, Dowson et al. [60] propose to apply single-period *conditionally consistent* risk measures in the context of SDDP [60], see also [11, 157]. It can be proven that under some technical assumptions, the class of entropic risk measures $\mathbb{ENT}_\gamma[\cdot]$ (see (12.8)) is the only class of risk measures that is conditionally consistent.

As $\mathbb{ENT}_\gamma[\cdot]$ can be applied in a nested fashion, the DPE (12.14)-(12.16) are valid in this case. Moreover, since $\mathbb{ENT}_\gamma[\cdot]$ is a convex risk measure, the (risk-adjusted) value functions are convex. Therefore, SDDP can be applied to derive polyhedral outer approximations.

As for standard SDDP, first, for each scenario $k \in \mathcal{K}$ and all possible stage-t realizations $\xi^k_{tj} \equiv \xi_{tj}, j = 1, \ldots, q_t$, approximate versions of subproblems (12.14) are solved to obtain $\underline{Q}^i_{\mathcal{R},t}(x^k_{t-1}, \xi_{tj})$. Then, based on the dual form of $\mathbb{ENT}_\gamma[\cdot]$, the following auxiliary problem can be solved to evaluate the risk-adjusted value function:

$$\begin{aligned}
&\mathbb{ENT}_\gamma\left[\underline{Q}^i_{\mathcal{R},t}(x^k_{t-1}, \xi_t)\right] \\
&= \begin{cases}
\max\limits_{\tilde{p}_t} & \displaystyle\sum_{j=1}^{q_t} \tilde{p}_{tj} \underline{Q}^i_{\mathcal{R},t}(x^k_{t-1}, \xi_{tj}) - \frac{1}{\gamma_t} \sum_{j=1}^{q_t} \tilde{p}_{tj} \cdot \log\left(\frac{\tilde{p}_{tj}}{p_{tj}}\right) \\
\text{s.t.} & \displaystyle\sum_{j=1}^{q_t} \tilde{p}_{tj} = 1 \\
& \tilde{p}_{tj} \geq 0, \quad j = 1, \ldots, q_t.
\end{cases}
\end{aligned} \tag{12.32}$$

Here, parameter p_{tj} denotes the nominal probabilities of realizations ξ_{tj}, which usually equal $\frac{1}{q_t}$, and the decision variable $\tilde{p}_{tj}$ denotes an alternative probability based on the entropic risk measure. In this way, problem (12.32) can be regarded as building the expectation based on some modified probability measure and with some additional penalty term. Problem (12.32) can be solved algorithmically, but as stated in [60], also a closed form for $\tilde{p}^*_{tj}$ can be derived. Using $\tilde{p}^*_{tj}$ and $\mathbb{ENT}_\gamma\left[\underline{Q}^i_{\mathcal{R},t}(x^k_{t-1}, \xi_t)\right]$, cuts can then be constructed and handed back to the previous stage.

The entropic risk measure does not only ensure conditional consistency of the obtained policies, but it also allows for upper bound computation as in standard SDDP, because the tower property can be employed for $\mathbb{ENT}[\cdot]$. However, these advantages

come at the cost of an aggravated interpretation of the risk measure compared to AVaR-based ones. In this context, it is particularly difficult to make a reasonable choice for the parameter $\gamma_t > 0$ [60].

12.6. SDDP with Expected Conditional CVaR. Another class of multi-period risk measures that can be used as an alternative to nested risk measures are *expected conditional risk measures*, which we briefly introduced in Section 12.2.1 [64, 105]. Here, conditional expectations are used to avoid the risk measure nesting, which proves beneficial in determining upper bounds in SDDP, as it avoids the aforementioned computational difficulties, while still time consistency is ensured.

Recall the risk-averse problem $(P_{\mathcal{R}})$ using expected conditional risk measures stated in (12.13). Using $\rho_t[\cdot] = \mathrm{CVaR}_{\alpha_t}[\cdot]$ yields the so called $\mathbb{E} - \mathrm{CVaR}$ or *multi-period average value-at-risk* [105], which goes back to Pflug and Ruszczyński [158].

As stated in [105], by some lengthy reformulations, the objective function of problem (12.13) can be expressed in a nested way. Therefore, equivalent DPE can be derived and time consistency is assured. Moreover, the $[\cdot]_+$-function can be reformulated by an epigraph approach. Then, for $t = 2, \dots, T$, the DPE read

$$\check{Q}_{\mathcal{R},t}(x_{t-1}, u_t, \xi_t) = \begin{cases} \min\limits_{x_t, u_{t+1}, w_t} & \dfrac{1}{\alpha_t} w_t + u_{t+1} + \check{\mathcal{Q}}_{\mathcal{R},t+1}(x_t, u_{t+1}) \\ \text{s.t.} & x_t \in \mathcal{X}_t(x_{t-1}, \xi_t) \\ & w_t - \big(c_t(\xi_t)\big)^\top x_t \geq -u_t \\ & w_t \geq 0 \end{cases} \tag{12.33}$$

with

$$\check{\mathcal{Q}}_{\mathcal{R},t+1}(x_t, u_{t+1}) = \mathbb{E}_{\boldsymbol{\xi}_{t+1}}\Big[\check{Q}_{\mathcal{R},t}(x_{t-1}, u_t, \boldsymbol{\xi}_t)\Big], \tag{12.34}$$

$\check{\mathcal{Q}}_{\mathcal{R},T+1}(\cdot,\cdot) \equiv 0$ and first stage

$$v^*_{\mathcal{R}} = \begin{cases} \min\limits_{x_1, u_2} & c_1^\top x_1 + u_2 + \check{Q}_{\mathcal{R},2}(x_1, u_2, \xi_t) \\ \text{s.t.} & x_1 \in \mathcal{X}_1. \end{cases} \tag{12.35}$$

In contrast to using nested conditional risk measures, the DPE here only depend on nested sums of (conditional) expectations, *i.e.*, have the same structure as in the risk-neutral case. Hence, standard SDDP can be applied. This has the advantage to allow on to use upper bounding techniques developed for risk-neutral SDDP.

12.7. Bi-objective SDDP. An alternative to risk-averse formulations that allows one to achieve a trade-off between obtaining the best policy in expectation (*e.g.*, the policy with the lowest expected costs) and avoiding bad extreme outcomes (*e.g.*, power outages or load shedding in an electricity network) is to formulate a multistage problem (MSLP) with multiple competing objectives that are optimized simultaneously. Recently, a variant of SDDP for bi-objective problems has been put forward by Dowson et al. [58].

Let $\widetilde{c}_t(\xi_t)$ and $\widehat{c}_t(\xi_t)$ denote the objective coefficients for stage $t \in [T]$ and the two competing objectives. For all but trivial cases, there exists no policy which yields the best objective value with respect to both objectives

$$\widetilde{v}^* := \min_{x_1, \boldsymbol{x_2}, \dots, \boldsymbol{x_T}} \underbrace{\mathbb{E}\left[\sum_{t \in [T]} \big(\widetilde{\boldsymbol{c}}_t(\xi_t)\big)^\top \boldsymbol{x}_t(\xi_{[t]})\right]}_{=:\widetilde{v}(\boldsymbol{x})}$$

and

$$\widehat{v}^* := \min_{x_1, \boldsymbol{x_2}, \ldots, \boldsymbol{x_T}} \underbrace{\mathbb{E}\left[\sum_{t \in [T]} \left(\widehat{\boldsymbol{c}}_t(\xi_t)\right)^\top \boldsymbol{x}_t(\xi_{[t]}) \right]}_{=:\widehat{v}(\boldsymbol{x})}$$

meaning that the two objectives are truly conflicting.

For this reason, if there is no clear preference for one of the objectives, usually the aim is to compute *Pareto-optimal* policies. A policy $\left(\bar{\boldsymbol{x}}_t(\xi_{[t]})\right)_{t \in [T]}$ is Pareto-optimal if it cannot be improved in one objective without getting worse in the other one, *i.e.*, if there exists no other policy $\left(\boldsymbol{x}_t(\xi_{[t]})\right)_{t \in [T]}$ such that $\widetilde{v}(\boldsymbol{x}) \geq \widetilde{v}(\bar{\boldsymbol{x}})$ and $\widehat{v}(\boldsymbol{x}) > \widehat{v}(\bar{\boldsymbol{x}})$ (or non-strict or strict inequality switched). Pareto-optimal solutions are also called *non-dominated*, and the set of non-dominated objective vectors is called the *Pareto front* [58].

A standard approach to compute Pareto-optimal solutions in optimization is to use some *scalarization* approach in which both conflicting objectives are combined to a weighted sum, which is then optimized in a deterministic single-objective problem. In our case, the DPE (2.4)-(2.6) can be adapted to

$$Q_t(x_{t-1}, \xi_t, \lambda) := \begin{cases} \min\limits_{x_t} & \left(\lambda \widetilde{c}_t(\xi_t) + (1-\lambda)\widehat{c}_t(\xi_t)\right)^\top x_t + \mathcal{Q}_{t+1}(x_t) \\ \text{s.t.} & x_t \in \mathcal{X}_t(x_{t-1}, \xi_t) \end{cases} \tag{12.36}$$

where

$$\mathcal{Q}_{t+1}(x_t, \lambda) := \mathbb{E}_{\boldsymbol{\xi}_{t+1}}\left[Q_{t+1}(x_t, \boldsymbol{\xi}_{t+1}, \lambda)\right] \tag{12.37}$$

and $\mathcal{Q}_{T+1}(x_T) \equiv 0$. For the first stage, we obtain

$$v^*(\lambda) = \begin{cases} \min\limits_{x_1} & \left(\lambda \widetilde{c}_1 + (1-\lambda)\widehat{c}_1\right)^\top x_1 + Q_2(x_1) \\ \text{s.t.} & x_1 \in \mathcal{X}_1. \end{cases} \tag{12.38}$$

SDDP can then be applied to these DPE. In the proposed variant, λ is adapted dynamically. To this end, in each iteration i, after the backward pass, one stage $t \in [T]$ is randomly and independently sampled and the corresponding subproblem is solved again for x_{t-1}^k, ξ_t^k and λ^i. Then, λ^i is updated to λ^{i+1}, where the latter is determined as the closest λ to λ^i such that the optimal basis of the constraint equation system changes.

It is proven that this variant of SDDP converges almost surely to a set of Pareto-optimal policies corresponding to the Pareto front of bi-objective (MSLP) in finitely many iterations [58].

13. SDDP with Unknown Distribution [relaxing Assumption 3]. In Section 3 we introduced SDDP assuming that the probability distribution $F_{\boldsymbol{\xi}}$ of the data process $(\boldsymbol{\xi}_t)_{t \in [T]}$ governing the uncertainty in problem (MSLP) is known, see Assumption 3. This allowed us to sample from this specific distribution in the forward pass of SDDP or, in case of continuous random vectors, to obtain a finite sample average approximation, as described in Section 11.

In practical applications, usually, the true distribution $F_{\boldsymbol{\xi}}$ is not known, though. Often, only historical data is available, *i.e.*, some realization of an unknown true distribution. This data is then used to determine a reasonable estimate for the true distribution, from which the required samples are taken. However, using such an estimation imposes the risk of *overfitting* the SDDP policies to this specific distribution,

and thus the available data [161]. Philpott et al. identify this problem as particularly noteworthy if the number of possible outcomes q_t per stage is small. For this reason, it may be reasonable to take a more robust approach and factor in the distributional uncertainty. Considering this type of uncertainty in SDDP is a young research area.

13.1. Distributionally Robust SDDP. One way to consider distributional uncertainty in SDDP is by integrating ideas from robust optimization [16, 20] into (multistage) stochastic programming. More precisely, a set of potential distributions is considered, which is called *distributional uncertainty set* or *ambiguity set* and denoted by $\mathcal{P}$. The expected cost is then minimized over the worst-case probability distribution from this set. This is called *Distributionally Robust Optimization* (DRO).

Usually, the outcomes of the random variables $\boldsymbol{\xi}_t$ are fixed to a finite number of realizations observed in the historical data. The ambiguity set $\mathcal{P}_t$ then models a variety of potential probability measures $\mathbb{P}_t \in \mathcal{P}_t$ supported on this finite set Ξ_t.

In the following, we restrict to DRO specifically in the SDDP context. For a general introduction to DRO, we refer to the review [175] and the tutorial [198]. We assume all assumptions from Section 3 to hold, except for Assumption 3. Furthermore, we only consider uncertainty in the RHS.

Then, the distributionally robust version of (MSLP) can be written as

$$\begin{aligned} \min_{x_1, \boldsymbol{x_2}, \ldots, \boldsymbol{x_T}} \max_{\mathbb{P} \in \mathcal{P}} \quad & \mathbb{E}\left[\sum_{t \in [T]} \big(\boldsymbol{c}_t(\xi_t)\big)^\top \boldsymbol{x}_t(\xi_{[t]})\right] \\ \text{s.t.} \quad & x_1 \in \mathcal{X}_1 \\ & \boldsymbol{x}_t \in \mathcal{X}_t(\boldsymbol{x}_{t-1}(\xi_{[t-1]}), \xi_t) \quad \forall \xi_t \in \Xi_t \ \forall t = 2, \ldots, T. \end{aligned} \tag{13.1}$$

Remark 13.1. Distributionally robust stochastic programming is closely related to risk-averse stochastic programming. In particular, the operator $\max_{\mathbb{P} \in \mathcal{P}} \mathbb{E}[\cdot]$ can be interpreted as a multi-period risk measure $\mathcal{R}[\cdot]$. This risk measure is coherent [198].

For SDDP it is required to reformulate problem (13.1) by means of DPE. This requires that each distribution $\mathbb{P}$ in the ambiguity set $\mathcal{P}$ can be expressed as the cross product of the respective marginal distributions of random vectors $\boldsymbol{\xi}_t$ [198]. Formally,

$$\mathcal{P} := \big\{\mathbb{P} = \mathbb{P}_1 \times \ldots \times \mathbb{P}_T \ : \ \mathbb{P}_t \in \mathcal{P}_t, t \in [T]\big\}.$$

The ambiguity sets $\mathbb{P}_t$ are assumed to be independent of each other. This property is called *rectangularity* of $\mathcal{P}$ and is reminiscent of the stagewise independence assumption for vectors $\boldsymbol{\xi}_t$. Note that $\mathcal{P}_1$ is a singleton containing one distribution with one possible realization.

With the ambiguity sets $\mathcal{P}_t$, then the DPE can be written as

$$Q_{DR,t}(x_{t-1}, \xi_t) := \begin{cases} \min\limits_{x_t} & c_t^\top x_t + \mathcal{Q}_{DR,t+1}(x_t) \\ \text{s.t.} & x_t \in \mathcal{X}_t(x_{t-1}, \xi_t) \end{cases} \tag{13.2}$$

with

$$\mathcal{Q}_{DR,t+1}(x_t) := \max_{\mathbb{P}_{t+1} \in \mathcal{P}_{t+1}} \mathbb{E}_{\mathbb{P}_{t+1}}\left[Q_{DR,t+1}(x_t, \boldsymbol{\xi}_{t+1})\right], \tag{13.3}$$

and $\mathcal{Q}_{DR,T+1}(x_T) \equiv 0$. Compared to Section 3, here, an inner maximization problem is introduced when defining $\mathcal{Q}_{DR,t+1}(\cdot)$ to obtain the expected cost over the worst-case

probability measure in $\mathcal{P}_{t+1}$. The first-stage problem reads

$$v^*_{DR} = \begin{cases} \min\limits_{x_1} & c_1^\top x_1 + \mathcal{Q}_{DR,2}(x_1) \\ \text{s.t.} & x_1 \in \mathcal{X}_1. \end{cases} \tag{13.4}$$

How v^*_{DR} and a corresponding optimal policy can be computed algorithmically, heavily depends on the specific choice of the ambiguity sets $\mathcal{P}_t, t = 2, \dots, T$. Various ambiguity sets are proposed in the literature. Usually, these sets are defined in such a way that they contain all distributions, which are *in some sense* within a given range of some nominal distribution. This nominal distribution, denoted by $\bar{\mathbb{P}}_t$, in turn, is defined by probabilities $\bar{p}_{tj} = \frac{1}{q_t}$ for all $j = 1, \dots, q_t$, where q_t denotes the number of historical data samples. Based on the measure employed to evaluate the distance between two distributions or probability measures, respectively, different classes of ambiguity sets can be defined.

For SDDP, the following three distance measures have been used so far. In [106], the ℓ_∞ metric is used to define the ambiguity set

$$\mathcal{P}_t = \left\{ \mathbb{P}_t \ : \ \sum_{i=1}^{q_t} p_{ti} = 1, \ p_{ti} \geq 0, \ \|p_t - \bar{p}_t\|_\infty \leq r \right\}. \tag{13.5}$$

A similar metric, but with the ℓ_2-norm, is used in [161] to define the ambiguity set

$$\mathcal{P}_t = \left\{ \mathbb{P}_t \ : \ \sum_{i=1}^{q_t} p_{ti} = 1, \ p_{ti} \geq 0, \ \|p_t - \bar{p}_t\|_2 \leq r \right\}. \tag{13.6}$$

This is a special case of the class of ϕ-divergence distances, see [12]. Both these distance measures are only applicable to discrete distributions supported on the observed historical data points.

On the contrary, the Wasserstein distance allows to compare general distributions (see for instance [215]). In our case with finite distributions $\mathbb{P}_t$ and $\bar{\mathbb{P}}_t$, the Wasserstein distance can be defined by the minimization problem

$$\begin{aligned} d_W(\bar{\mathbb{P}}_t, \mathbb{P}_t) := \min_z \quad & \sum_{i=1}^{q_t} \sum_{j=1}^{q_t} \|\xi_t^i - \xi_t^j\| z_{ij} \\ \text{s.t.} \quad & \sum_{j=1}^{q_t} z_{ij} = \bar{p}_{ti} \quad \forall i = 1, \dots, q_t \\ & \sum_{i=1}^{q_t} z_{ij} = p_{tj} \quad \forall j = 1, \dots, q_t \\ & z_{ij} \geq 0 \quad \forall i, j = 1, \dots, q_t, \end{aligned}$$

where for the norm different choices are possible. It can be interpreted as the amount of probability mass that has to be moved between the distributions. This distance is used in [65] to define the Wasserstein ambiguity set

$$\mathcal{P}_t = \left\{ \mathbb{P}_t \ : \ \sum_{i=1}^{q} p_{ti} = 1, \ p_{ti} \geq 0, \ d_W(\bar{\mathbb{P}}_t, \mathbb{P}_t) \leq r \right\}. \tag{13.7}$$

In all three cases, very different strategies are chosen to apply SDDP to the nested min-max structure defined by the DPE (13.2)-(13.4).

13.1.1. Reformulation as a Risk-averse Problem. As shown in [106], using the ambiguity set (13.5), the DPE (13.2)-(13.4) can be reformulated to those of a risk-averse multistage problem with nested conditional $\text{AVaR}_\alpha[\cdot]$, that is equations (12.18)-(12.20) with

$$\lambda_{t+1} = 1 - p^\ell_{t+1}, \quad \alpha_{t+1} = \frac{\lambda_{t+1}}{p^u_{t+1} - p^\ell_{t+1}},$$

where p^ℓ_{t+1} and p^u_{t+1} denote the probabilities associated with the probability measures at the lower and upper bound of ambiguity set (13.5). Therefore, SDDP can be applied as in this risk-averse setting.

13.1.2. Solving the Inner Maximization Problem Separately. Using ambiguity set (13.6) in the DPE (13.2)-(13.4) yields value functions, which can be proven to remain convex, and thus can be approximated by affine cuts [161].

To derive such cuts, Philpott et al. propose to solve the inner maximization problem identifying the worst-case distribution separately. In the backward pass, for some stage t, first the subproblems are solved for all $j = 1, \ldots, q_t$ as usual. Then, using the obtained values of $\underline{Q}^i_t(x^{ik}_{t-1}, \xi_{tj})$, the inner maximization problem is solved. This can be done algorithmically and in some cases even analytically, as shown in [161]. The obtained worst-case probability measure $\mathbb{P}^*$ can then be used to compute subgradients and cut coefficients. Even though these coefficients are determined based on cut approximation $\mathfrak{Q}^{i+1}_t(\cdot)$ and on $\mathbb{P}^*$, which does not necessarily coincide with the worst-case probability measure in the true DPE, valid cuts are constructed and convergence is ensured [161].

13.1.3. Using a Dual Representation. If we use the Wasserstein ambiguity set (13.7) in SDDP, we obtain the inner maximization problem.

$$\begin{aligned}
\min_{z_t, p_{t+1}} \quad & \sum_{j=1}^{q_t} p_{t+1,j} Q_{t+1}(x_t, \xi_{t+1,j}) \\
\text{s.t.} \quad & \sum_{i=1}^{q_t}\sum_{j=1}^{q_t} d_{t+1,ij} z_{tij} \le 1 \\
& \sum_{j=1}^{q_t} z_{tij} = \bar{p}_{ti} \quad \forall i = 1, \ldots, q_t \\
& \sum_{i=1}^{q_t} z_{tij} = p_{tj} \quad \forall j = 1, \ldots, q_t \\
& z_{tij} \ge 0 \quad \forall i, j = 1, \ldots, q_t
\end{aligned}$$

with $d_{t+1,ij} = \|\xi^i_{t+1} - \xi^j_{t+1}\|$. Duque and Morton [65] suggest to replace this problem using its dual problem. This way, the value functions can be evaluated by solving the single-level minimization problem

$$Q_{DR,t}(x_{t-1}, \xi_t) := \begin{cases} \min\limits_{x_t, \gamma_t, \nu_t} & c_t^\top x_t + r\gamma_t + \sum\limits_{i=1}^{q_{t+1}} q^i_{t+1} \nu^i_t \\ \text{s.t.} & x_t \in \mathcal{X}_t(x_{t-1}, \xi_t) \\ & d_{t+1,ij}\gamma_t + \nu_{ti} \ge Q_{DR,t+1}(x_t, \xi_{t+1,j}) \quad \forall i, j = 1, \ldots, q_{t+1} \\ & \gamma_t \ge 0 \end{cases}$$

with dual variables γ_t and ν_t.

As proven in [65], these value functions are piecewise linear and convex on $\mathcal{X}_{t-1}$, and therefore can be represented by finitely many linear cuts. Note, however, that this approach requires to use multi-cut SDDP, see Section 21.2, since otherwise bilinear terms occur.

With all these strategies, the forward pass remains basically the same as in standard SDDP. The sampling can be done from the nominal distribution associated with $\bar{\mathbb{P}}_t, t = 2, \dots, T$, or alternatively the current worst-case distribution associated with $\mathbb{P}_t^*$ [65]. If independent sampling is conducted, convergence follows as for standard SDDP. However, challenges to determine valid upper bounds are prevalent for distributionally robust SDDP similarly to the risk-averse case.

Computational results indicate that taking the dual reformulation approach, better approximations are achieved for multi-cut SDDP than solving the inner maximization in a side computation [65]. Furthermore, out-of-sample tests by Philpott et al. imply that distributionally robust SDDP yields policies which are better suited, *e.g.*, induce lower costs, in periods with a substantial risk of high costs [161].

13.2. Partially Observable Distributions. A different approach to deal with distributionally uncertainty is introduced by Dowson et al. in [59], and is referred to as *partially observable* multistage stochastic programming. The idea is to consider a finite number of potential distributions by combining problem (MSLP) with a hidden Markov model. More precisely, in each stage $t \in [T]$, different nodes can be reached, with each node representing one Markov state. Each node reflects a different candidate distribution, possibly with identical realizations $\xi_j, j = 1, \dots, q$, but different associated probabilities.

To model the uncertainty with respect to the distributions, the nodes are partitioned by a partition $\mathcal{A}$ into ambiguity sets $A \in \mathcal{A}$, with $\bigcup_{A \in \mathcal{A}} = \mathcal{N}$ and $\mathcal{N}$ the number of all nodes besides the root node. For example, the partition can be chosen such that there is one ambiguity set A for each stage.

It is now assumed that at any point, only the current ambiguity set is known, while the specific node within it cannot be observed. However, for each node i, a probability b_i of being in that node can be computed. In other words, for each candidate distribution there exists a probability with which this distribution is considered to be the most accurate representation of the true underlying distribution. These probabilities are stored in a so called *belief state* b.

Each time an ambiguity set A is entered and a particular realization $\widetilde{\xi}$ of the random data is observed, the belief state is updated componentwise by applying Bayes' theorem [59].

In contrast to (MSLP) with perfect distribution information (see Assumption 3), the value functions $Q_t(\cdot)$ have to incorporate this belief state. To this end, let $p_{i\ell}$ be the probability of observing $\xi_{i\ell}$ conditional on being in node i with $\ell = 1, \dots, q^i$. Let $\bar{\mathcal{N}}$ describe all nodes including the root node, ω_{jk} the transition probability from node j to k and $B_k(b, \xi)$ the update rule for the belief state being in (unobservable) node k. Furthermore, let x' denote the current trial solution. Then, the expected value function can be written as

$$\mathcal{Q}_B(x', b) := \sum_{j \in \bar{\mathcal{N}}} b_j \sum_{k \in \mathcal{N}} \omega_{jk} \sum_{\ell=1}^{q^k} p_{k\ell} \, Q_k\big(x', B_k(b, \xi_{k\ell}), \xi_{k\ell}\big). \tag{13.8}$$

This means that the value functions $Q_k(\cdot, \cdot)$ depend on a node and an updated belief

state, and in (13.8) it is looped over all nodes, weighing the corresponding expected value with the current belief and the transition probabilities between the nodes.

As proven in Theorem 1 in [59], the expected value functions $\mathcal{Q}_B(\cdot)$ are saddle functions, as they are convex in x for fixed b, but concave in b for fixed x. Therefore, to apply SDDP, the cut generation has to be adapted to this property. This can be achieved by using an outer approximation for x and an inner approximation for b [59]. The main difference for the cut computation is that apart from taking expectations over the realizations of ξ, it is looped over all nodes in the current ambiguity set A and the cut components are weighed with the current belief [59].

In the forward pass, for each stage $t = 2, \dots, T$, first, a new node is sampled conditionally on the (unobserved) current node. Then, a realization of ξ is sampled conditionally on the obtained node and the associated candidate distribution. For a more detailed description, see [59].

A different method of combining SDDP with a hidden Markov model is given in [66]. One general drawback of such hidden Markov approaches is that transition probabilities between the nodes have to be properly defined a priori.

14. Stagewise Dependent Uncertainty [relaxing Assumption 2]. As explained in Sections 2 and 3, stagewise independence (Assumption 2) is a standard assumption in dynamic programming, and thus also for SDDP. It is also crucial for the computational tractability of SDDP compared to NBD because it ensures that there exists only one expected value function $\mathcal{Q}_t(\cdot)$ per stage and that cuts can be shared between scenarios, see Section 5.2. However, in many applications, the uncertain data in (MSLP) (*e.g.*, demand, fuel prices, electricity prices, inflows) shows correlations over time and assuming stagewise independence is not appropriate.

If the uncertainty in problem (MSLP) is stagewise dependent, the expected value functions $\mathcal{Q}_t(\cdot)$ for $t = 2, \dots, T$ do not only depend on x_{t-1}, but implicitly also depend on the history $\xi_{[t-1]}$ of the process $(\boldsymbol{\xi}_t)_{t\in[T]}$. In order to apply SDDP, this dependence has to be taken into account, for instance by reformulating the model or adapting the algorithmic steps in SDDP. In this section, we consider different cases of stagewise dependent uncertainty and ways of how SDDP can be applied in these cases.

14.1. Expanding the State Space. As a first case of stagewise dependent uncertainty, let us assume that the data process $(\boldsymbol{\xi}_t)_{t\in[T]}$ is a simple linear *autoregressive* (AR) process with lag one, defined by appropriately chosen coefficient vectors γ_t, matrices Φ_t and stagewise independent and i.i.d. error terms η_t:

$$\boldsymbol{\xi}_t = \gamma_t + \Phi_t \boldsymbol{\xi}_{t-1} + \boldsymbol{\eta}_t. \tag{14.1}$$

Remark 14.1. If we still assume finite randomness (Assumption 5), now for $\boldsymbol{\eta}_t$, then $\boldsymbol{\xi}_t$ can be modeled by a classical scenario tree, see Section 5.2.

The most natural approach to deal with this case, is to reformulate (MSLP) in such a way that it exhibits stagewise independent uncertainty [153]. This can be achieved by including ξ_{t-1} as an additional state variable. Then, as shown in [128],

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\xi}_t|\xi_{t-1}}\left[Q_t(x_{t-1}, \boldsymbol{\xi}_t)\right] &= \mathbb{E}_{\boldsymbol{\eta}_t|\xi_{t-1}}\left[Q_t(x_{t-1}, \gamma_t + \Phi_t\xi_{t-1} + \boldsymbol{\eta}_t)\right] \\ &= \mathbb{E}_{\boldsymbol{\eta}_t}\left[Q_t(x_{t-1}, \gamma_t + \Phi_t\xi_{t-1} + \boldsymbol{\eta}_t)\right],\end{aligned}$$

where the second equality holds because η_t and ξ_{t-1} are statistically independent.

By introducing equation (14.1) as a constraint and defining a new value function

$$\widehat{Q}_t(x_{t-1}, \xi_{t-1}, \eta_t) := Q_t(x_{t-1}, \gamma_t + \Phi_t\xi_{t-1} + \eta_t), \tag{14.2}$$

and the corresponding expected value function

$$\widehat{\mathcal{Q}}_t(x_{t-1}, \xi_{t-1}) := \mathbb{E}_{\boldsymbol{\eta}_t}\left[\widehat{Q}_t(x_{t-1}, \xi_{t-1}, \boldsymbol{\eta}_t)\right] \tag{14.3}$$

for all $t = 2, \dots, T$, it follows

$$\mathbb{E}_{\boldsymbol{\xi}_t|\xi_{t-1}}\left[Q_t(x_{t-1}, \boldsymbol{\xi}_t)\right] = \widehat{\mathcal{Q}}_t(x_{t-1}, \xi_{t-1}).$$

The state variables then consist of the resource state x_{t-1} and the information state ξ_{t-1}, while the stagewise independent uncertainty is modeled by $\boldsymbol{\eta}_t$. Importantly, ξ_t is regarded as a decision variable in the reformulated problem, augmenting the dimension of the decision space.

Remark 14.2. It is worth emphasizing that this approach is presented in various different ways in the literature. In some cases, as outlined, equation (14.1) is explicitly incorporated into the DPE as an additional constraint [173, 202]. In some cases, each occurrence of ξ_t in the subproblems is simply replaced by the RHS of (14.1). And in other cases, the dependence on ξ_{t-1} is only expressed by writing $\widehat{Q}_t(\cdot,\cdot,\cdot)$ and $\widehat{\mathcal{Q}}_t(\cdot,\cdot)$ as functions of ξ_{t-1}, whereas the explicit relation (14.1) is only considered in the cut generation process [84, 128, 178]. We revisit this observation in the next subsection.

By the presented procedure, stagewise independence (Assumption 2) is recovered for (MSLP). However, in order to apply SDDP, it also has to be ensured that valid cuts for $\widehat{\mathcal{Q}}_t(\cdot,\cdot)$ can be derived as functions in both types of state variables. This requires that $\widehat{\mathcal{Q}}_t(\cdot,\cdot)$ is convex in both x_{t-1} and ξ_{t-1}. Similarly to Theorem 2.8, it can be shown that under certain assumptions, this property is satisfied.

THEOREM 14.3 ([178]). *Let $\boldsymbol{\xi}_t$ be described by (14.1) and let ξ_{t-1} be contained in some convex set. Then, under Assumptions 1 and 3 to 9, the expected value function $\widehat{\mathcal{Q}}_t(\cdot,\cdot)$ is piecewise linear and*

a) convex in x_{t-1} on $\mathcal{X}_{t-1}$ for fixed ξ_{t-1},
b) convex in $\xi_{t-1} = (T_{t-2}, h_{t-1})$ for fixed $x_{t-1}, W_{t-1}, c_{t-1}$,
c) concave in $\xi_{t-1} = c_{t-1}$ for fixed $x_{t-1}, W_{t-1}, T_{t-2}, h_{t-1}$,
d) convex jointly in x_{t-1} and in $\xi_{t-1} = h_{t-1}$ for fixed $W_{t-1}, T_{t-2}, c_{t-1}$.

Theorem 14.3 shows that convexity in both types of state variables is only guaranteed if the stagewise dependent part of the uncertainty only enters the RHS $\boldsymbol{h}_t(\xi_t)$ of problem (MSLP). Note that this still allows for additional stagewise independent uncertainty in c_t, W_t and T_{t-1}. The result also requires linearity of (MSLP) (Assumption 6) and of the AR process (14.1) defining the random variable $\boldsymbol{\xi}_t$.

Under certain assumptions, Theorem 14.3 can be generalized to convex problems (MSLP) and stagewise dependence in the RHS defined by a convex function [84]. Moreover, the result is not limited to lag-one processes, but can be enhanced to AR processes with higher lag order [84]. This is important for practical applications, as often several lags are required to explain a time series appropriately. In contrast, for general nonlinear stochastic processes or for uncertainty in W_t, c_t or T_{t-1}, such a generalization seems not possible. In order to cover such cases, different approaches are required. We discuss those in later parts of this section.

For simplicity, assume that $X_t = \{x_t \in \mathbb{R}^{n_t} \ : \ x_t \geq 0\}$ for all $t \in [T]$ and recall

the definition of the approximate subproblem (2.10):

$$\underline{Q}_t(x_{t-1},\xi_t) = \begin{cases} \min\limits_{x_t,\theta_{t+1}} & \big(c_t(\xi_t)\big)^\top x_t + \theta_{t+1} \\ \text{s.t.} & W_t(\xi_t)x_t = h_t(\xi_t) - T_{t-1}(\xi_t)x_{t-1} \\ & x_t \geq 0 \\ & -(\beta^r_{t+1})^\top x_t + \theta_{t+1} \geq \alpha^r_{t+1}, \quad \forall r \in \Gamma_{t+1}, \end{cases} \tag{14.4}$$

where Γ_{t+1} is the index set of previously generated cuts. Then, the result in Theorem 14.3 can be illustrated by means of the feasible region of the LP dual to (14.4), which can be written as

$$\begin{aligned} \max_{\pi_t,\rho_t} \quad & \big(h_t(\xi_t) - T_{t-1}(\xi_t)x_{t-1}\big)^\top \pi_t + a_{t+1}^\top \rho_t \\ \text{s.t.} \quad & \big(W_t(\xi_t)\big)^\top \pi_t - B_{t+1}^\top \rho_t \leq c_t(\xi_t) \\ & e^\top \rho_t = 1 \\ & \rho_t \geq 0. \end{aligned} \tag{14.5}$$

Here, we collect all cut gradients β^r_{t+1} in a matrix B_{t+1} and all cut intercepts α^r_{t+1} in a vector a_t for compact representation. π_t denotes the dual variable to the original constraints, and ρ_t denotes the dual variable to the previously generated cuts.

In the case of linear AR processes in the RHS $\boldsymbol{h}_t(\xi_t)$, the dual feasible region is not affected by the new state variable ξ_{t-1} (and also remains polyhedral). This means that the extreme solutions obtained for one state $\bar{\xi}_{t-1}$ remain valid, although not necessarily optimal, for all other states ξ_{t-1} as well. In contrast, in other cases of stagewise dependence, the dual feasible region and its extreme solutions may change for different states, affecting the properties of $\widehat{\mathcal{Q}}_t(\cdot,\cdot)$ [178].

In sum, for affine and convex AR processes occurring in the RHS, expanding the state recovers stagewise independence (Assumption 2), but at the same time convexity of $\widehat{\mathcal{Q}}_t(\cdot,\cdot)$ in all state variables is preserved. Therefore, SDDP can be used as introduced in Section 3. In this case, the obtained cuts are functions of both state variables and can be formulated with a cut gradient for each of them (compare to (3.5)), *i.e.*,

$$\phi_t(x_{t-1},\xi_{t-1}) = \alpha_t + \big(\beta_t^x\big)^\top x_{t-1} + \big(\beta_t^\xi\big)^\top \xi_{t-1}.$$

Unfortunately, depending on the dimension κ_{t-1} of ξ_{t-1}, the state space dimension can increase significantly. This effect is amplified for higher lag orders. As the computational complexity of SDDP grows exponentially in this dimension, see Section 4.2, augmenting the state space is detrimental and should be avoided if possible.

14.2. Scenario-Adaptable Cut Formulas. The previously described adverse effect can be alleviated to some degree by a special cut generation approach that was first proposed by Infanger and Morton [108] and later enhanced by de Queiroz and Morton [173] and Guigues [84]. In all these cases, the process model, such as (14.1), is not explicitly incorporated into the subproblems, see Remark 14.2. Instead, it is merely considered within the cut generation process. The main idea is to derive scenario-adaptable closed-form cut formulas, given AR processes with a specific structure, which allow one to adapt the cut generated for one specific history $\bar{\xi}_{[t-1]}$ to different histories $\xi_{[t-1]}$ of the stochastic process, and thus to different scenarios. This way, the cuts can be *shared* between scenarios (see Section 5.2) without the need to incorporate (14.1) into (MSLP) as a constraint. Importantly, these cut formulas lead to the exact same cuts as the previously described approach.

To illustrate this idea, consider a cut derived using dual problem (14.5) without paying any particular attention to the stagewise dependence. For convenience, but without loss of generality, we assume T_{t-1} to be deterministic and the RHS uncertainty to be defined by

$$\boldsymbol{h}_t(\xi_t) = \Phi_t h_{t-1}(\xi_{t-1}) + \boldsymbol{\eta}_t \tag{14.6}$$

with stagewise independent error terms $\boldsymbol{\eta}_t$, similarly to (14.1). We obtain

$$\begin{aligned}\widehat{\mathcal{Q}}_t(x_{t-1}, \xi_{t-1}) &\geq \mathbb{E}_{\boldsymbol{\xi}_t|\xi_{t-1}}\left[-\boldsymbol{\pi}_t^\top T_{t-1} x_{t-1} + \boldsymbol{\pi}_t^\top \boldsymbol{h}_t(\xi_t) + \boldsymbol{\rho}_t^\top \boldsymbol{a}_{t+1}\right] \\ &= \mathbb{E}_{\boldsymbol{\xi}_t|\xi_{t-1}}\left[-\boldsymbol{\pi}_t^\top T_{t-1}\right] x_{t-1} + \mathbb{E}_{\boldsymbol{\xi}_t|\xi_{t-1}}\left[\boldsymbol{\pi}_t^\top \boldsymbol{h}_t(\xi_t) + \boldsymbol{\rho}_t^\top \boldsymbol{a}_{t+1}\right]\end{aligned} \tag{14.7}$$

We can make the following observations:

(i) Since the probabilities in $\mathbb{E}_{\boldsymbol{\xi}_t|\xi_{t-1}}[\cdot]$ are assumed to not depend on ξ_{t-1} (recall that $\boldsymbol{\eta}_t$ is stagewise independent) and since all scenarios share the same dual feasible region, the cut gradient

$$\beta_t = \mathbb{E}_{\boldsymbol{\xi}_t|\xi_{t-1}}\left[-\boldsymbol{\pi}_t^\top T_{t-1}\right] \tag{14.8}$$

derived for one specific scenario ξ_{t-1}^s, is valid for all other scenarios as well.

(ii) According to (14.6), the RHS $\boldsymbol{h}_t(\xi_t)$ depends on ξ_{t-1}. Therefore, to evaluate the cut for a specific scenario, this term has to be adapted to this scenario. Otherwise, the cut may become invalid. By (14.6), this term can be split up into a scenario-dependent part depending on ξ_{t-1} and a scenario-independent part depending on η_t only.

(iii) The last term $\boldsymbol{a}_{t+1}$ in (14.7) is the cut intercept of the following stage. As we face stagewise dependence, this intercept is not scenario-independent anymore, but should denote $\boldsymbol{a}_{t+1}(\xi_t)$. Moreover, it is defined recursively: The stage-t intercept includes the stage-$(t+1)$ intercept, which includes the stage-$(t+2)$ intercept and so on. This implies that to evaluate $\boldsymbol{a}_{t+1}(\xi_t)$ for a specific scenario, it is basically required to recursively traverse the whole scenario tree starting form stage t. This is computationally intractable.

To address these observations, the main idea by Infanger and Morton [108] is to express the cut intercept $\alpha_t(\xi_{t-1})$ as the sum of a stagewise independent term α_t^{ind} and a stagewise dependent term $\alpha_t^{\text{dep}}(\xi_{t-1})$:

$$\alpha_t(\xi_{t-1}) = \alpha_t^{\text{ind}} + \alpha_t^{\text{dep}}(\xi_{t-1}). \tag{14.9}$$

Let $\bar{\pi}_t = \mathbb{E}_{\boldsymbol{\eta}_t}[\boldsymbol{\pi}_t]$ and $\bar{\rho}_t = \mathbb{E}_{\boldsymbol{\eta}_t}[\boldsymbol{\rho}_t]$ denote the expected value of the dual variables obtained for realizations of $\boldsymbol{\eta}_t$. As explained, these dual values are valid for any history of the stochastic process due to the structure of the dual feasible set. Let $\bar{\mathscr{P}}_t$ define the $(|\Gamma_t| \times m_t)$-matrix containing the values of $\bar{\pi}_t$ and $\bar{\mathscr{R}}_t$ the $(|\Gamma_t| \times |\Gamma_{t-1}|)$-matrix containing the values of $\bar{\rho}_t$ for the previously determined cuts. Furthermore, let the matrix D_t be defined recursively by

$$D_t = \left[\bar{\mathscr{P}}_{t+1} + \bar{\mathscr{R}}_{t+1} D_{t+1}\right] \Phi_t, \quad D_T = 0. \tag{14.10}$$

Then, as shown in [108], the stagewise dependent cut intercept is given by

$$\alpha_t^{\text{dep}}(\xi_{t-1}) = \left[\bar{\pi}_t + \bar{\rho}_t D_t\right] \Phi_t h_{t-1}(\xi_{t-1}). \tag{14.11}$$

This means that a cut can be constructed by using formula (14.8) for the gradient and formulas (14.9), (14.10) and (14.11) for the intercept. The stagewise independent

RHS $h_t(\xi_t)$	Autoregressive model for ξ_t Model	Type	Lag	Formula	Source
const.	AR	L	1	$\xi_t = \Phi_t \xi_{t-1} + \eta_t$	[108]
L	AR	L	1	$\xi_t = \Phi_t \xi_{t-1} + \eta_t$	[173]
const.	PAR	L	1	$\xi_t = \varphi_t(\xi_{t-1} - \mu_{t-1}) + \mu_t + \sigma_t \eta_t$	[209]
const.	AR	L	≥ 1	$\xi_t = \sum_{k=1}^{t-1}(\Phi_k^t \xi_k + \Psi_k^t \eta_k) + \eta_t$	[108]
L/C*	AR	L	≥ 1	$\xi_t = \Phi_t \xi_{[t-1]} + \eta_t$	[84]
L/C*	AR	L	≥ 1	$\xi_t = \Phi_t \xi_{[t-1]} + \Psi_t \eta_t + \Theta_t$	[84]
const.	SPAR	L	≥ 1	$\xi_{ti} = \sum_{i'} \sum_{k=1}^{t-1} \Phi_{ii'k}^t \xi_{ti'} + \eta_{ti}$	[126]
const.	AR	NL	1	$\xi_t = \Phi_t(f_t(v_{t-1}) + \xi_{t-1}) + \eta_t$	[108]
const.	AR	NL	≥ 1	$\xi_t = \sum_{k=1}^{t-1}(\Phi_k^t \xi_k + f_k^t(v_k)) + \eta_t$	[108]
C	AR	C	≥ 1	$\xi_t = f_t(\xi_{[t-1]}, \eta_t)$	[84]

L = affine/linear function, C = convex function, NL = general nonlinear function
* only in case of inequality constraints

Table 6: RHS and uncertainty models considered in the literature on SDDP with stagewise dependence to derive scenario-adaptable closed-form cut formulas.

term can be either determined by an additional formula or by subtracting (14.11) from $\alpha_t(\xi_{t-1})$ [108]. In order for a cut to be shared with a different scenario at stage $t-1$, it is only required to adapt the stagewise dependent intercept (14.11) to this specific scenario. In other words, a given cut can be *corrected* to be valid for a different history of the stochastic process. In particular, it is not required to add (14.6) as a constraint to the stage-t subproblem or to traverse the whole scenario tree (see Remark 14.1). Instead, only the cut gradient, the stagewise independent part of the intercept and the *cumulative expected dual vector* $\left[\bar{\mathscr{P}}_{t+1} + \bar{\mathscr{R}}_{t+1} D_{t+1}\right] \Phi_t$ have to be stored [108].

Whereas we limited our explanations to a very simple AR process so far, similar cut formulas can be derived for more complex processes [84, 108, 173, 178]. We give an overview on different cases covered in the literature in Table 6. Some of the process formulas in Table 6 are presented in a simplified form for reasons of clarity, *e.g.*, by omitting standardization and the incorporation of seasonal or periodical effects. For example, this is true for the SPAR processes considered in [126] (also see Section 9), where spatial dependencies between locations i and i' are taken into account.

Importantly, all processes for which scenario-adaptable closed-form cut formulas can be derived require a specific structure, such as linearity, convexity or separability. As shown by Guigues [84], a generalization to convex AR processes and more complex structures in the RHS is possible. For instance, the RHS h_t does not have to be directly described by the stochastic process (constant $\boldsymbol{h}_t \equiv \boldsymbol{\xi}_t$), but may also be defined as some function $\boldsymbol{h}_t(\cdot)$ of $\boldsymbol{\xi}_t$. Moreover, for the affine case, alternative formulas to the ones provided by Infanger and Morton are presented by Guigues [84]. The main difference is that only a minimal subset of coefficients is used, due to defining the process $(\boldsymbol{\xi}_t)_{t\in[T]}$ componentwise and not in vectorial form compared to (14.1) or (14.6). On the other hand, no recursive formula as in (14.10) is provided to compute the cut coefficients. Finally, Guigues shows that also for feasibility cuts (Section 17) scenario-adaptable cut formulas can be derived.

It is important to emphasize that the presented approach only partially mitigates the drawbacks of augmenting the state space. First of all, the history of the stochastic

process has to be stored to compute ξ_t, even if such computation is possible outside of the subproblems. Guigues provides a detailed discussion on how state vectors of minimal size can be defined in order to keep the stored information as small as possible [84]. Additionally, due to their dependence on ξ_{t-1}, or $\xi_{[t-1]}$ in general, the expected value functions $\widehat{\mathcal{Q}}_t(\cdot,\cdot)$ live in a higher-dimensional space. Therefore, more iterations and cuts may be required to achieve convergence compared to the stagewise independent case, as discussed in Section 4.2.

14.3. Sensitivity of SDDP with AR Processes. Let the uncertainty in (MSLP) be modeled by an AR process. Consider the approach of expanding the state, leading to two types of state variables: x_t and $\xi_{[t]}$. Both contain information on future resource availability (*e.g.*, hydro storage volume and hydro inflow history affecting future inflows), but they differ in several aspects [206]. First, whereas the information provided by the state x_{t-1} is certain, the information provided by $\xi_{[t-1]}$ enters an AR model predicting future realizations, which still involves uncertainty. Second, the parameters of this AR model are estimated from data, and thus can be subject to estimation errors. Third, in practice it can often be observed that the values in $(\xi_t)_{t\in[T]}$ show higher variability over short time than the values of $(x_t)_{t\in[T]}$. This uncertainty and variability raises the question on how much the solutions obtained in SDDP react to changes in $\xi_{[t-1]}$. This can be examined in a *sensitivity analysis*.

A general approach for sensitivity analysis in SDDP is presented in [97] and applied to an inventory problem with AR demand. Also the sensitivity with respect to AR model parameters Φ_t or γ_t is discussed.

For a hydrothermal problem, in [206], it is shown that the solutions obtained in SDDP are more sensitive to changes in the initial information state ξ_1 than to changes in the initial resource state x_0. Based on the previous observations this leads to the unfavorable side effect of expanding the state space that solutions of SDDP exhibit larger variability. This may have severe consequences in economic applications, such as increasing risk, unpredictability of prices or distorted investment signals.

To address this issue, Soares et al. present different mitigation heuristics [206], such as regularizing changes in x_t over time, or using the accurate AR model in the forward pass of SDDP, but predefined unconditional samples in the backward pass in order to avoid the dependence of cuts on $\xi_{[t-1]}$. While they report positive computational results, the authors provide no theoretical results on reasonable parameter choice, cut validity and convergence for their heuristics.

14.4. Markov Chain SDDP. Assume the data process $(\boldsymbol{\xi}_t)_{t\in[T]}$ is Markovian, *i.e.*, as in (14.1), $\boldsymbol{\xi}_t$ only depends on ξ_{t-1} for all $t = 2,\dots,T$ instead of the whole history $\xi_{[t-1]}$. Then, instead of expanding the state space also an alternative approach can be used to apply SDDP.

In this case, the data process can be represented, or at least approximated (if the random variables $\boldsymbol{\xi}_t$ are continuous), by a discrete Markov chain. This approximation can be obtained by lattice quantization techniques [29, 128]. As it contains only finitely many states per stage $t = 2,\dots,T$, this Markov chain can be illustrated as a recombining scenario tree or *scenario lattice* [128], just as in the case of stagewise independence Assumption 2, see Section 2. The difference is that in the Markov chain case the probabilities of transitions to stage-t nodes may differ between different stage-$(t-1)$ nodes. This also includes the possibility that some stage-t nodes may not be reached from certain stage-$(t-1)$ nodes.

Due to this difference, the (expected) value functions $\mathcal{Q}_{t\ell}(\cdot)$ depend on the states $\zeta_\ell, \ell = 1,\dots,L$, of the Markov chain. In other words, for each such state (*i.e.*, each

node in the recombining tree), a different expected value function and a different set of value functions exist. In SDDP, then cuts are derived for each of these functions separately. This idea is called Markov chain SDDP (MC-SDDP) [128] or *approximate dual dynamic programming* (ADDP) [129, 130], whereas for distinction the approach of expanding the state space is referred to as time series SDDP (TS-SDDP).

For problems with moderate state space dimension, expanding the state may be computationally favorable as only one expected value function has to be approximated per stage. On the other hand, a computational advantage of MC-SDDP is that the computational effort grows linearly with the number of Markov states only [200]. In contrast, expanding the state leads to a state space dimension increase in which the complexity of SDDP grows exponentially. Moreover, MC-SDDP requires no linearity and is not limited to stagewise dependent uncertainty only appearing in the RHS of (MSLP). As long as the Markov property is satisfied, it allows for stagewise dependent uncertainty in all data c_t, T_{t-1}, W_t and h_t of (MSLP).

The main drawback of MC-SDDP lies in the relation to the true problem $(\tilde{P})$ in case of a continuous data process $(\boldsymbol{\xi}_t)_{t\in[T]}$, see also Section 11. For SDDP with AR processes and expanding the state space, many results exist that allow for inference of the SAA solution with respect to the true problem, see Section 11. One key property in this regard is that ξ_{t-1} is treated as a possibly continuous state variable in SDDP, such that the derived cuts are also valid at states which are not reached by the scenarios $\xi^s \in \mathcal{S}$ that are considered in SDDP. Similar results are not available for MC-SDDP. In particular, the obtained policy and lower bounds are not necessarily valid for the true problem [128].

In spite of this theoretical downside, Löhndorf and Shapiro report tighter lower bounds and better policies even for the true process based on computational experiments [128]. They conjecture that this is due to a differing exploration of the state space. Expanding the state space introduces additional state variables, which are not under control of the optimal policy (their trajectory is not chosen based on solving the approximate subproblems in the forward pass), but selected randomly in the forward pass). This may lead to selection of states, which do not provide the highest information gain. With MC-SDDP this is partially mitigated by choosing sufficiently different states in advance when constructing the Markov chain.

14.5. SDDP with Integrated Markov Chain. By Theorem 14.3, a natural extension of SDDP to stagewise dependent uncertainty using expanding the state space is only possible for linear (or at least convex) AR processes appearing in the RHS of problem (MSLP). In all other cases, expanding the state space destroys the convexity of the expected value functions $\widehat{\mathcal{Q}}_t(\cdot,\cdot)$. Therefore, in such cases, different approaches are required. One such approach is to integrate a discrete Markov chain into the uncertainty modeling. This approach is quite established in the literature on and in practical application of SDDP. Importantly, this approach does not necessarily coincide with the previous case where the process $(\boldsymbol{\xi}_t)_{t\in[T]}$ is assumed to be Markovian and approximated by a Markov chain.

Modeling. Consider a Markov chain with finitely many possible states $\zeta_\ell, \ell = 1,\dots,L$, with $L \in \mathbb{N}$. At each stage $t \in [T]$, we denote the current state of the Markov chain as ψ_t (again, we assume that ψ_1 is deterministic). The transition probabilities between state $\psi_{t-1} = \zeta_\ell$ at stage $t-1$ and $\psi_t = \zeta_{\ell'}$ at stage t are then denoted by $\omega_{\ell\ell'}$ for $\ell, \ell' \in \{1,\dots,L\}$. For simplicity, we assume the Markov chain to be time-homogeneous, such that $\omega_{\ell\ell'}$ does not depend on t, even though this is not required.

We now assume that the distribution of random variable $\boldsymbol{\xi}_t$ at stage $t \in [T]$ may

depend on the state ψ_t of the Markov chain. In other words, for each possible state $\zeta_\ell, \ell = 1, \dots, L$, the distribution of $\boldsymbol{\xi}_t$ may differ. We emphasize this by writing $\boldsymbol{\xi}_t^\ell$.

The value functions $Q_t(\cdot, \cdot)$ for (MSLP) then do not only depend on x_{t-1} and the realization ξ_t of $\boldsymbol{\xi}_t$, but also on the current Markov state ψ_t. As this state can only take finitely many values, we denote this by $Q_{t\ell}(x_{t-1}, \xi_t)$, where index ℓ indicates conditioning on $\psi_t = \zeta_\ell$. Based on this definition, the expected value functions can be expressed as

$$\mathcal{Q}_{t\ell}(x_{t-1}) := \sum_{\ell'=1}^{L} \omega_{\ell\ell'} \mathbb{E}_{\boldsymbol{\xi}_t|\ell'} \left[Q_{t\ell'}(x_{t-1}, \boldsymbol{\xi}_t^{\ell'}) \right]. \tag{14.12}$$

The index ℓ of the expected value function refers to the previous Markov state $\psi_{t-1} = \zeta_\ell$. Compared to standard SDDP, the expectation is not only taken over the realizations of $\boldsymbol{\xi}_t^{\ell'}$, but also the state transitions from ψ_{t-1} to ψ_t are taken into account. Using this definition, the DPE for stages $t = 2, \dots, T$ can be written as

$$Q_{t\ell}(x_{t-1}, \xi_t^\ell) := \begin{cases} \min\limits_{x_t} & \zeta_\ell^\top x_t + \mathcal{Q}_{t+1,\ell}(x_t) \\ \text{s.t.} & x_t \in \mathcal{X}_t(\xi_t^\ell). \end{cases} \tag{14.13}$$

Note that the dependence on ζ_ℓ in (14.12) resembles the expanding-the-state approach from Section 14.1. However, there are important differences. ψ_{t-1} does not enter the subproblems and it can only take a finite number of different values, whereas $\xi_{[t-1]}$, even if discrete, is treated like a continuous state variable when expanding the state. Furthermore, as the transition probabilities $\omega_{\ell\ell'}$ may differ for each ζ_ℓ, the cut components are weighted differently and cuts cannot be shared between different Markov states. Consequently, it is required to store separate expected value functions $\mathcal{Q}_{t\ell}(\cdot)$ for each $\ell = 1, \dots, L$. In return, the non-convexity of these functions is circumvented, since each $\mathcal{Q}_{t\ell}(\cdot)$ remains convex and is approximated on its own, see also the discussion in Section 14.4.

As an example, consider a problem with $L = 2$ Markov states and $q^\ell = 2$ realizations for $\boldsymbol{\xi}_t^\ell$ for each of them, which is borrowed from [159]. The corresponding scenario tree with underlying Markov chain is illustrated in Figure 13. For the transition probabilities let $\omega_{11} = q, \omega_{12} = 1 - q, \omega_{21} = 1 - p$ and $\omega_{22} = p$. For all t and $\ell \in \{1, 2\}$, the distribution of ξ_t^ℓ is given by $p_{tj} = \frac{1}{2}$ for $j \in \{1, 2\}$.

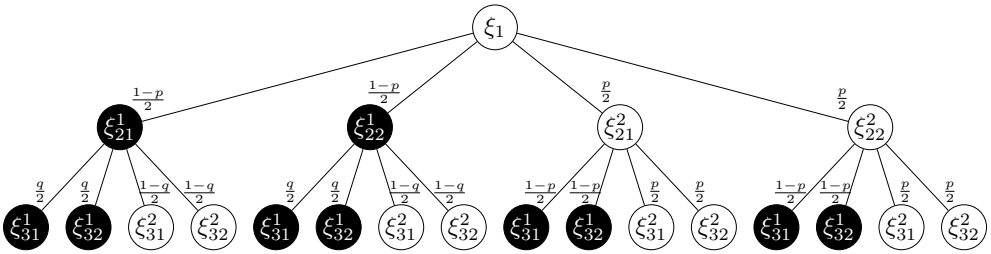


Fig. 13: Scenario tree with underlying Markov chain (state 1 printed in black, state 2 printed in white). Replication from [159].

As an alternative to the scenario tree in Figure 13, the stochastic process with underlying Markov chain can be represented by a Markovian policy graph with finitely many nodes per stage [56]

SDDP. Let us now address how SDDP works in this case. In the forward pass, different approaches are used in the literature. The most natural one is for each stage t and each sample path $k \in \mathcal{K}$, to sample first from the Markov states and then conditionally from $\boldsymbol{\xi}_t^\ell$ [160]. Sometimes it is also proposed to use historical values here, *e.g.*, true inflow spot-price combinations [79]. In such a case, it is possible that a spot price is drawn which is not a valid state of the Markov chain. Then, a strategy is to use the *in some sense* closest state from the Markov chain [79]. Another one is to use a linear interpolation between the hyperplanes of neighbouring states [81, 227].

For stages $t = 2, \ldots, T$, states $\ell = 1, \ldots, L$ and samples $k \in \mathcal{K}$, based on (14.13), the approximate subproblems solved in the forward pass of SDDP have the form

$$\underline{Q}^i_{t\ell}(x^{ik}_{t-1}, \xi^{\ell k}_t) := \begin{cases} \min\limits_{x_t} & \left(c_t(\xi^\ell_t)\right)^\top x_t + \mathfrak{Q}^i_{t+1\ell}(x_t) \\ \text{s.t.} & x_t \in \mathcal{X}_t(x^{ik}_{t-1}, \xi^\ell_t). \end{cases} \tag{14.14}$$

Importantly, each function $\mathcal{Q}_{t\ell}(\cdot), \ell = 1, \ldots, L$, is approximated by an individual cut approximation $\mathfrak{Q}_{t\ell}(\cdot)$.

In the backward pass of some iteration i, the stages are traversed in backward direction as usual to improve the cut approximations. At each stage t, the subproblems (14.14) updated with $\mathfrak{Q}^{i+1}_{t\ell}(\cdot)$ are solved for each trial state x^{ik}_{t-1}, $k \in \mathcal{K}$, each stage-t Markov state $\psi_t = \zeta_\ell, \ell = 1, \ldots, L$, and all realizations $\xi^\ell_{tj}, j = 1, \ldots, q^\ell_t$.

Then, for each x^{ik}_{t-1} and $\psi_{t-1} = \zeta_\ell, \ell = 1, \ldots, L$, a valid cut can be derived for $\mathcal{Q}_{t\ell}(\cdot)$. Let $\beta^i_{t\ell kj}$ denote a subgradient for $\underline{Q}^i_{t\ell}(\cdot, \cdot)$ at x^{ik}_{t-1}. In accordance with (3.4), but also taking into account the Markov chain transition probabilities, we can then define cut coefficients

$$\beta^i_{t\ell k} := \sum_{\ell'=1}^{L} \omega_{\ell\ell'} \left(\sum_{j=1}^{q_{t\ell}} p_{t\ell j} \left(\underline{Q}^{i+1}_{t\ell}(x^{ik}_{t-1}, \xi^{\ell k}_t) - (\beta^i_{t\ell kj})^\top x^{ik}_{t-1} \right) \right),$$

$$\alpha_{t\ell} := \sum_{\ell'=1}^{L} \omega_{\ell\ell'} \left(\sum_{j=1}^{q_{t\ell}} p_{t\ell j} \beta^i_{t\ell kj} \right),$$

where $q_{t\ell}$ and $p_{t\ell j}$ denote the number of realizations and probabilities of $\boldsymbol{\xi}^\ell_t$.

A cut (3.5) for $\mathcal{Q}_{t\ell}(\cdot)$ is then given by function

$$\phi^i_{t\ell k}(x_{t-1}) := \alpha^i_{t\ell k} + (\beta^i_{t\ell k})^\top x_{t-1}$$

and can be used to update $\mathfrak{Q}^i_{t\ell}(\cdot)$. Philpott et al. derive similar formulas for the multi-cut and risk-averse case [160].

Use Cases. There exist different use cases for modeling the uncertainty in (MSLP) with an integrated Markov chain.

- The data process $(\boldsymbol{\xi}_t)_{t\in[T]}$ can be modeled as a nonlinear AR process or a nonlinear transformation of a linear AR process (see Section 9), which, if handled by expanding the state space, destroys the convexity of $\widehat{\mathcal{Q}}_t(\cdot, \cdot)$. Sometimes such a nonlinear process can be approximated by assuming that the realizations ξ_t depend on an underlying system state which follows a Markov process [160], thus not capturing the nonlinearity explicitly in a formula. As the value functions are also not convex in this, possibly continuous, Markov state, the Markov process is approximated using a discrete Markov chain.

- Instead of a single AR process, sometimes the data process $(\boldsymbol{\xi}_t)_{t\in[T]}$ may be best modeled by a finite set of different AR processes, which are valid representations, and thus active, under different circumstances (*e.g.*, macroeconomic, political or ecological situations). A discrete Markov chain can then be used to model these overall system states, and AR models can be used to describe realizations of the uncertain data conditioned on these states. Such *regime-switching* models are very common in wind forecasting [231].
- *Hybrid SDP/SDDP.* Different parts of the data in (MSLP) exhibit stagewise dependent uncertainty. While some of them, namely uncertainty in the RHS h_t, can be treated by expanding the state space, for others, *e.g.*, stagewise dependent uncertainty in the objective coefficients c_t, it would destroy the convexity of $\widehat{\mathcal{Q}}_t(\cdot,\cdot)$. Therefore, this part of the uncertainty may be modeled by a discrete Markov chain instead. Since one part of the uncertainty is treated as in standard SDDP (allows for cut-sharing between scenarios), while another one is treated by enumerating separate expected value functions for each $\ell = 1,\dots,L$ (cuts cannot be shared between Markov states), this is often referred to as a hybrid SDP/SDDP method [79].
  For instance, this setting often occurs in medium-term hydrothermal scheduling problems (see Section 9) when inflow uncertainty in the RHS as well as spot-price uncertainty in the objective function are taken into account. The idea to address this by using a Markov chain goes back to Gjelsvik et al. who modeled this kind of scheduling problem for the Norwegian power system [79, 81, 82]. Since then, this approach has been employed in several applications, for example, hydrothermal scheduling including balancing market bids [99, 100], risk management [107, 115, 141] and fuel contracts [37]. It is also applied to model fuel price uncertainty [150].
  In contrast to the presented general approach, in this case it is usually assumed that the uncertainty in the RHS and in the objective are independent of each other. Therefore, for each state $\zeta_\ell, \ell = 1,\dots,L$, the distribution of $\boldsymbol{\xi}_t$ is the same, and marginal distributions can be used in the expectation in (14.12). Moreover, note that in this specific case the Markov chain states are not underlying the distribution of $\boldsymbol{\xi}_t$, but instead entering the subproblems explicitly, *e.g.*, as objective coefficients. Still SDDP can be applied using the same ideas as above.

The described approach allows for the incorporation of even nonlinear stagewise dependent uncertainty into SDDP, but also gives rise to some challenges. Among those is the assumption of the Markov property, which may not always be appropriate. Moreover, it is required to define useful values $\zeta_\ell, \ell = 1,\dots,L$, and transition probabilities $\omega_{\ell\ell'}$ for the Markov states [81, 142]. Most importantly, cuts cannot be shared between, but only within Markov states, so that separate expected value functions have to be considered for each $\ell = 1,\dots,L$. Therefore, the number of Markov states should be rather small to preserve computational tractability.

14.6. Hybrid NBD/SDDP. In the previous section, we presented a hybrid SDP/SDDP method as a tool to model different stagewise dependent uncertain data in (MSLP) by different approaches. Instead of modeling the "complicating" part of the uncertainty by a discrete Markov chain, also a scenario tree can be used. Instead of a hybrid SDP/SDDP method, this yields a hybrid NBD/SDDP method [178], see also Section 5.2.

Assume that the random vector $\boldsymbol{\xi}_t$ modeling the uncertainty in c_t, W_t, T_{t-1} and

h_t can be separated into two separate and independent parts, $\boldsymbol{\xi}_t^S$ and $\boldsymbol{\xi}_t^T$. The first vector $\boldsymbol{\xi}_t^S$ can either be stagewise independent or exhibit some linear dependency if it occurs in the RHS. In the latter case, it can be handled by expanding the state space. Within SDDP, in each iteration samples of $\boldsymbol{\xi}_t^S$ are considered. The second vector $\boldsymbol{\xi}_t^T$, on the other hand, may lead to non-convexities in the value functions if it is approached by expanding the state space. Therefore, it is modeled by a scenario tree, which is treated exactly in SDDP. This means that for this particular part of the uncertainty, no samples are drawn, but all scenarios are considered in each iteration of SDDP, as in NBD, see Section 5.2. This approach is similar to hybrid SDP/SDDP in the sense that the expected value functions $\mathcal{Q}_t(\cdot)$ depend on the scenarios from $\boldsymbol{\xi}_t^S$ and that cuts can only be shared within, but not between such scenarios.

By only treating the crucial part $\boldsymbol{\xi}^T$ of $\boldsymbol{\xi}$ as a scenario tree and the remainder $\boldsymbol{\xi}^S$ still by sampling, complex uncertainty processes can be considered, while at the same time the increase of computational complexity is kept as small as possible [178]. To take advantage of this, the scenario tree associated with $\boldsymbol{\xi}^S$ should not be too large.

Compared to hybrid SDP/SDDP, in specific applications the one or the other approach may be favorable. For instance, the Markov chain approaches allow for dependencies between different uncertainty processes. Moreover, in the case that each realization of $\boldsymbol{\xi}_t$ is assigned to one specific Markov state $\zeta_\ell, \ell = 1, \dots, L$, the number of LPs to be solved per iteration can be kept equal to standard SDDP. The scenario tree approach, by contrast, requires independence of $\boldsymbol{\xi}^S$ and $\boldsymbol{\xi}^T$. By design, it considers all combinations of scenarios of $\boldsymbol{\xi}^T$ and $\boldsymbol{\xi}^S$, so no assignment of realizations of $\boldsymbol{\xi}^S$ to scenarios of $\boldsymbol{\xi}^T$ is required. However, the number of LPs to be solved grow exponentially in the number of stages [178]. On the other hand, a scenario tree may be more appropriate to model very complex processes, *e.g.*, referring to macroeconomical, political or structural decisions [178], for which the Markov property is not appropriate.

14.7. Saddle Cuts. We consider the special case of stagewise dependent objective coefficients $\boldsymbol{c}_t(\xi_t)$ in (MSLP), as they appear for uncertain prices models by AR processes. So far, we introduced SDDP with integrated Markov chain as a suitable solution approach in this case. Now, we discuss as second one.

As discussed in Section 14.1, by expanding the state space, stagewise independence (Assumption 2) can be recovered, but in return the expected value functions $\widehat{\mathcal{Q}}_t(\cdot,\cdot)$ are no longer convex. In Theorem 14.3 it is shown that $\widehat{\mathcal{Q}}_t(\cdot,\cdot)$ is in fact convex in x_{t-1}, but concave in c_{t-1}, which yields a saddle shape. Therefore, linear cuts are not sufficient to approximate them. As a resort, exploiting the saddle shape, special *saddle cuts* can be used.

To derive this formally, in the vein of [55], we assume the objective coefficients to be described by $\big(y_t(\xi_t)\big)^\top C_t$ instead of $c_t(\xi_t)$. While the matrix C_t is considered deterministic, $y_t(\xi_t)$ is defined by the following AR process

$$y_t(\xi_t) = B_t(\xi_t) y_{t-1}(\xi_{t-1}) + b_t(\xi_t) \tag{14.15}$$

for all stages $t = 2, \dots, T$. Here, the matrix B_t and the vector b_t are uncertain and depend on the realization of $\boldsymbol{\xi}_t$. Thus, the sequence $\big(y_t(\xi_t)\big)_{t=1}^T$ is scenario-dependent.

Inserting relation (14.15) into the objective function and considering y_{t-1} as an

additional state variable, for $t = 2, \dots, T$, we obtain the subproblems

$$\widehat{Q}_t(x_{t-1}, y_{t-1}, \xi_t) = \begin{cases} \min\limits_{x_t} & \big(B_t(\xi_t)y_{t-1} + b_t(\xi_t)\big)^\top C_t x_t + \widehat{\mathcal{Q}}_{t+1}\big(x_t, B_t(\xi_t)y_{t-1} + b_t(\xi_t)\big) \\ \text{s.t.} & x_t \in \mathcal{X}_t(x_{t-1}, \xi_t) \end{cases}$$

where

$$\widehat{\mathcal{Q}}_{t+1}(x_t, y_t) = \mathbb{E}_{\boldsymbol{\xi}_{t+1}}\left[\widehat{Q}_{t+1}(x_t, y_t, \boldsymbol{\xi}_{t+1}))\right]$$

and $\widehat{\mathcal{Q}}_{T+1}(x_T, y_T) \equiv 0$. For the first stage, we obtain

$$v^* = \begin{cases} \min\limits_{x_1} & b_1 C_1 x_1 + \widehat{\mathcal{Q}}_2(x_1, y_1) \\ \text{s.t.} & x_1 \in \mathcal{X}_1. \end{cases}$$

The additional state y_{t-1} is referred to as an *objective state*. This state is not allowed to appear in the constraints [55]. As stated before, $\widehat{\mathcal{Q}}_t(\cdot, \cdot)$ is piecewise linear and convex in x_{t-1}, but piecewise linear concave in y_{t-1} and as such, a piecewise bilinear saddle function.

The concept of approximating saddle functions with saddle cuts goes back to Baucke et al., who propose a deterministic algorithm to solve stochastic minimax dynamic programs [11]. A related approach is used in *robust dual dynamic programming* (RDDP), which uses an SDDP-like framework to solve multistage robust programs [76]. The main idea is to compute lower and upper bounding saddle functions, which combine the ideas of an outer approximation by cutting-planes and an inner approximation by convex combinations of function values, the latter of which we discuss thoroughly in Section 8. For stagewise dependent objective coefficients, it is sufficient to only use the lower bounding saddle functions, so-called *saddle cuts*, from [11] to approximate the expected value functions in SDDP.

Let (3.4) define β_t and α_t as in standard SDDP. Then, the r-th saddle cut for $\widehat{\mathcal{Q}}_{t+1}(\cdot, \cdot)$ is defined as the solution to the optimization problem

$$\begin{aligned} \min_{\mu_t,\ \theta_{t+1}} \quad & y_t^\top \mu_t + \theta_{t+1} \\ \text{s.t.} \quad & (y_t^r)^\top \mu_t + \theta_{t+1} \geq \alpha_{t+1}^r + (\beta_{t+1}^r)^\top x_t \\ & \|\mu_t\|_\infty \leq \nu \end{aligned} \tag{14.16}$$

where $y_t^r = y_t^{ik}$ denotes the current objective state in iteration i and for scenario $k \in \mathcal{K}$. Importantly, this problem has x_t and y_t as parameters. Hence, a saddle cut gives a valid lower approximation for $\widehat{\mathcal{Q}}_{t+1}(\cdot, \cdot)$ for all x_t and y_t and can be shared between scenarios. Moreover, the saddle cuts are tight at the trial state given by x_t^{ik} and y_t^{ik}, at which they are created.

A crucial part of applying this approach is to bound the decision variable μ_t in (14.16) by an appropriate constant ν. To this end, the expected value functions $\widehat{\mathcal{Q}}_t(\cdot, \cdot)$ are required to be Lipschitz continuous with respect to y_{t-1}. As shown in [11], to ensure validity of the saddle cuts, the parameter ν has to be chosen at least as large as the Lipschitz constant of $\widehat{\mathcal{Q}}_t(\cdot, \cdot)$ with respect to y_{t-1} under the dual norm $\|\cdot\|_1$ of $\|\cdot\|_\infty$. If it is chosen smaller, this may result in invalid cuts and suboptimal solutions. If it is chosen too large, the cuts may become very weak [55].

Incorporating the saddle cuts, for each stage $t = 2, \dots, T$, iteration i and scenario $k \in \mathcal{K}$, the SDDP subproblems can be formulated as

$$\underline{\widehat{Q}}_t^i(x_{t-1}^{ik}, y_{t-1}^{ik}, \xi_{tj}) = \begin{cases} \min\limits_{x_t, \mu_t, \theta_{t+1}} & (y_t^{ik})^\top C_t x_t + (y_t^{ik})^\top \mu_t + \theta_{t+1} \\ \text{s.t.} & x_t \in \mathcal{X}_t(x_{t-1}^{ik}, \xi_{tj}) \\ & (y_t^r)^\top \mu_t + \theta_{t+1} - (\beta_{t+1}^r)^\top x_t \geq \alpha_{t+1}^r, \quad r \in \Gamma_{t+1} \\ & \|\mu_t\|_\infty \leq \nu, \end{cases}$$

where $y_t^{ik} = B_t(\xi_t^k) y_{t-1}^{ik} + b_t(\xi_t^k)$.

It can be shown that only finitely many different saddle cuts can be constructed. As a consequence, the convergence results are the same as for standard SDDP [55].

14.8. Applying Dual SDDP. A third alternative that is tailored to stagewise dependent objective coefficients $\boldsymbol{c}_t(\xi_t)$ in (MSLP) is to apply dual SDDP [97], as presented in Section 8. Recall the value functions derived from the dual problem of (MSLP):

$$\widetilde{D}_t(\pi_{t-1}) := \begin{cases} \max\limits_{\pi_t} & \sum\limits_{j=1}^{q_t} p_{tj} \Big(- h_{tj}^\top \pi_{tj} + \widetilde{D}_{t+1}(\pi_{tj}) \Big) \\ \text{s.t.} & \sum\limits_{j=1}^{q_t} p_{tj} \Big(T_{t-1,j}^\top \pi_{tj} \Big) + W_{t-1}^\top \pi_{t-1} \leq c_{t-1}. \end{cases} \tag{14.17}$$

These value functions are concave in π_{t-1}. Crucially, here the objective coefficients c_{t-1} appear in the RHS. If $(\boldsymbol{c}_t)_{t \in [T]}$ is described as a linear AR process, we can expand the state space as for the primal subproblems in Section 14.1, and the new state variable $c_{[t-2]}$ appears in the RHS. Therefore, the obtained value functions are also concave in $c_{[t-2]}$ and can be approximated from above by linear cuts. This can be done by applying dual SDDP [97], see Section 8.

14.9. Conditional Cuts. The previously discussed approaches all have in common that they require to expand the state space or to set up a scenario tree or a discrete Markov chain from the true (continuous) data process (or from existing historical data). van Ackooij and Warin propose an alternative approach that works without these requirements [225]. The approach is based on established methods in mathematical finance and optimal stopping theory. A crucial assumption is that the data process $(\boldsymbol{\xi}_t)_{t \in [T]}$ is Markovian.

Assume that a finite set $\mathcal{S}$ of scenarios $\xi^s, s \in \mathcal{S}$, is given, *e.g.*, historical observations of the data. This set is chosen in advance and not changed within SDDP. The first key ingredient of the proposed variant of SDDP is to partition the set of possible values of $\boldsymbol{\xi}_t$ for each stage $t \in [T]$ into a finite number $|L_t|$ of hypercubes $D_{t\ell}$, $\ell = 1, \dots, |L_t|$, also called *meshes*. This partitioning is done in such a way that approximately a uniform distribution of the samples is achieved [225].

In the forward pass of SDDP, a subset $L_t \subseteq \mathcal{S}_t$ of scenarios are sampled for each stage. This is done with the aim to obtain a trial solution x_t^ℓ for each mesh in expectation for all $t = 2, \dots, T$. Each of these trial solutions is then used in the backward pass to derive cuts.

In the backward pass, in principle, the whole set of scenarios $\mathcal{S}$ is considered as candidates for cut derivation. For any sequence $(x_t^{i\ell})_{t \in [T]}$ of trial solutions, let

$\big(d(t)^{i\ell}\big)_{t=1}^{T}$ denote the sequence of corresponding meshes, *i.e.*, $x_t^{i\ell}$ has been determined in the forward pass for $\xi_t^\ell \in D_{t,d(t)^{i\ell}}$. At each stage $t = T, \dots, 2$, the SDDP subproblems are now solved for all scenarios ξ_t^s for which $\xi_{t-1}^s \in D_{t-1,d(t-1)^{i\ell}}$. This means that for each trial solution, all scenarios are considered which share the same mesh with the scenario used to obtain the trial solution.

After solving these subproblems, the obtained solutions are used to construct cuts. However, the cut derivation process differs from standard SDDP. The cut coefficients are determined as estimates of the corresponding conditional expectations [225]:

$$\alpha_{t\ell}^i(\xi_{t-1}) = \hat{\mathbb{E}}^S_{|\xi_{t-1}}\Big[(\boldsymbol{\pi}_t^{i\ell s})^\top \boldsymbol{h}_t(\xi_t) + \sum_{r\in\Gamma_{t+1}} \boldsymbol{\rho}_t^{i\ell sr}\alpha_{t+1}^r\Big]$$

and

$$\beta_{t\ell}^i(\xi_{t-1}) = -\hat{\mathbb{E}}^S_{|\xi_{t-1}}\Big[(\boldsymbol{\pi}_t^{i\ell s})^\top T_{t-1}\Big].$$

These estimates are computed by linearly regressing the terms for each considered scenario ξ_t^s on a finite number of local base functions, *e.g.*, monomials in $\mathbb{R}^{p_t}$, with support on the considered mesh. This yields a cut

$$\mathcal{Q}_t(x_{t-1},\xi_{t-1}) \geq \phi_{t\ell}^i(x_{t-1},\xi_{t-1}) = \big(\beta_{t\ell}^i(\xi_{t-1})\big)^\top x_{t-1} + \alpha_{t\ell}^i(\xi_{t-1}),$$

which provides a local update of the cut approximation in the current mesh $D_{t-1,d(t-1)}$ and is zero otherwise. For this reason, the cut is associated with this specific mesh and stored in a corresponding index set.

For each subproblem solved in the forward or backward pass, only the set of cuts is taken into account which is associated with the currently explored mesh then [225]. Therefore, these cuts are called *conditional cuts*.

For problems with a low-dimensional vector $\boldsymbol{\xi}_t$ and Markovian dependency, the policies obtained using conditional cuts are reported to be competitive with those obtained by expanding the state space, but without an increase of the state dimension and without the need to set up a scenario tree [225].

15. Extension to Convex Programs [relaxing Assumption 6]. A natural extension of SDDP can be achieved by relaxing the assumption of linearity, *i.e.*, Assumption 6, but assuming a multistage stochastic convex problem (MSCP). We take the following assumptions [78, 85].

ASSUMPTION 10. *Let $f_t(\cdot)$ and $g_t(\cdot,\cdot)$ (componentwise) be convex lower semicontinuous proper and differentiable functions and X_t nonempty convex compact sets for all $t \in [T]$.*

Under stagewise independence (Assumption 2) and Assumption 10, (MSCP) can be expressed using its DPE in the following form. For $t = 2, \dots, T$ they read

$$Q_{t,C}(x_{t-1},\xi_t) := \begin{cases} \min\limits_{x_t} & f_t(x_t,\xi_t) + \mathcal{Q}_{t+1,C}(x_t) \\ \text{s.t.} & g_t(x_{t-1},x_t,\xi_t) \leq 0 \\ & x_t \in X_t, \end{cases} \tag{15.1}$$

with expected value functions defined as usual by

$$\mathcal{Q}_{t+1,C}(x_t) := \mathbb{E}_{\boldsymbol{\xi}_{t+1}}\left[Q_{t+1,C}(x_t,\boldsymbol{\xi}_{t+1})\right] \tag{15.2}$$

and $\mathcal{Q}_{T+1,C}(x_T) \equiv 0$. For the first stage, this yields

$$v_C^* = \begin{cases} \min\limits_{x_1} & f_1(x_1) + \mathcal{Q}_{2,C}(x_1) \\ \text{s.t.} & g_1(x_1) = 0 \\ & x_1 \in X_1. \end{cases} \tag{15.3}$$

Applying SDDP to (MSCP) with convergence guarantees requires a more strict recourse assumption compared to Assumption 9, which we present under finite randomness (Assumption 5).

ASSUMPTION 11. *(Extended relatively complete recourse [78]) Let aff($\mathcal{X}_t$) be the affine hull of the reachable set $\mathcal{X}_t$ and $B_t(\delta_t) = \{y \in \text{aff}(\mathcal{X}_t) : \|y\| < \delta_t\}$ for some $\delta_t > 0$ and some norm $\|\cdot\|$.*

For all $t \in t = 2, \dots, T$, all $x_{t-1} \in \mathcal{X}_{t-1} + B_t(\delta_t)$ and all $\xi_{tj}, j = 1, \dots, q_t$, the feasible set of subproblems (15.1) is non-empty.

Intuitively, Assumption 11 demands that feasibility of the subproblems is also ensured for x_{t-1} slightly outside of $\mathcal{X}_t$. This is required in order to guarantee Lipschitz continuity of all value functions $Q_{t,C}(\cdot,\cdot)$ and expected value functions $\mathcal{Q}_{t,C}(\cdot)$ [78]. Additionally, all value functions are convex, and thus can be approximated by linear cuts. Such cuts can be generated using Lagrangian duality. More precisely, for all $t = 2, \dots, T$ and $x_{t-1} \in \mathcal{X}_{t-1}$, we introduce the Lagrangian function

$$L_{t,C}(\pi_t; x_{t-1}, x_t, \xi_t) = f_t(x_t, \xi_t) + \pi_t^\top g_t(x_{t-1}, x_t, \xi_t), \tag{15.4}$$

the corresponding dual function

$$\mathcal{L}_{t,C}(\pi_t; x_{t-1}, \xi_t) = \min_{x_t \in X_t} L_t(\pi_t; x_{t-1}, x_t, \xi_t) \tag{15.5}$$

and the corresponding Lagrangian dual problem

$$\max_{\pi_t \geq 0} \mathcal{L}_t(\pi_t; x_{t-1}, \xi_t). \tag{15.6}$$

Further, we make the following assumption which ensures no duality gap between the primal subproblems (15.1) and their dual problems (15.6) [85]. Here, ri(S) denotes the relative interior of some set S.

ASSUMPTION 12. *(Slater condition [85]) For all $x_{t-1} \in \mathcal{X}_{t-1}$ and all $\xi_{tj}, j = 1, \dots, q_t$, there exists $x_t \in ri(X_t)$ such that $g_t(x_{t-1}, x_t, \xi_{tj}) < 0$.*

Then, exploiting differentiability, a subgradient of $\mathcal{Q}_{t,C}(\cdot)$ at $\bar{x}_{t-1}$ is given by

$$\bar{\beta}_t = \partial \mathcal{Q}_t(\bar{x}_{t-1}) = \sum_{j=1}^{q_t} p_{tj} \nabla_{x_{t-1}} L_{t,C}(\bar{\pi}_{tj}; \bar{x}_{t-1}, \bar{x}_{tj}, \xi_{tj}),$$

where $\bar{x}_{tj}$ is an optimal solution to the primal problem (15.1) and $\bar{\pi}_{tj}$ is an optimal solution to the dual problem (15.6) given ξ_{tj}. Moreover, $\nabla_x h(\cdot)$ denotes the gradient of some function $h(\cdot)$ with respect to x. Using this subgradient, a cut for $\mathcal{Q}_t(\cdot)$ is given by [85]

$$\mathcal{Q}_t(x_{t-1}) \geq \mathcal{Q}_t(\bar{x}_{t-1}) + \bar{\beta}_t^\top (x_{t-1} - \bar{x}_{t-1}).$$

Under Assumption 11, the norm of the obtained subgradients can be shown to be bounded [85].

This cut derivation can be generalized to DPE including $\mathfrak{Q}_t(\cdot)$ instead of $\mathcal{Q}_t(\cdot)$. The results can also be generalized to cost functions $f_t(x_{t-1}, x_t, \xi_t)$ depending on the state x_{t-1}, see [85] for details.

Contrary to the linear case, however, the expected value functions $\mathcal{Q}_{t,C}(\cdot)$ are no longer polyhedral. As a consequence, they cannot be represented exactly by a finite number of cuts. However, it can be shown that given the above assumptions and Assumptions 1 to 8 almost sure asymptotic convergence of SDDP is ensured. In [78] this is proven for the case that x_{t-1} only enters the subproblems (15.1) in linear constraints, that is, $g_t(\cdot)$ being a linear function. In [85] the convergence proof is extended to the more general setting presented above. For both convergence proofs also the differentiability requirement can be dropped. As shown in [71], almost sure *finite* convergence can be achieved for ε-optimal policies, for some predefined $\varepsilon > 0$.

In [92], Guigues and Monteiro propose a slightly different algorithmic approach, called StoDCuP (Stochastic Dynamic Cutting Plane), in which not only $\mathcal{Q}_t(\cdot), t = 2, \dots, T$, but also some or all nonlinear functions $f_t(\cdot)$ and $g_t(\cdot)$ are iteratively approximated by affine functions at the trial points visited in the forward pass.

Another variant of SDDP is DASC (decomposition algorithm for multistage stochastic programs with strongly convex cost functions), which is introduced in [86]. It can be applied when the (expected) value functions in (MSCP) are *strongly convex*. For this type of problems, it is proposed to approximate them using functions $\mathfrak{Q}_t(\cdot)$ which are defined as the pointwise maximum of quadratic cuts instead of affine cuts. In contrast to standard SDDP, this means that the subproblems to be solved in SDDP become nonlinear, but in return good approximations of the expected value functions are obtained much quicker, and thus less iterations are expected [86]

While most research on SDDP deals with problems (MSLP), some of the extensions presented previously and in the following sections have also been enhanced to the convex case, *e.g.*, risk-aversion [85], inexact cuts [88], regularization [90] or exact upper bounding procedures [10, 118]. [85] contains an extension of the convergence proof from [78] to the risk-averse case. Furthermore, the idea to use inexact cuts is generalized to convex non-differentiable problems [91], see Section 21.

16. Extensions to Mixed-integer and Non-convex Problems [relaxing Assumption 6]. In many practical applications, multistage stochastic problems do involve integer decision variables or nonlinear, but non-convex terms in the objective function or constraints, see Section 9. In general, such programs can be formulated in the same way as in the convex case, but with the functions $f_t(\cdot)$ and $g_t(\cdot)$ possibly being non-convex. Moreover, in this case, X_t is the intersection of a convex compact set, *e.g.*, representing box constraints, with possible integer constraints, *i.e.*, $X_t \subset \mathbb{R}_t^{n_{t1}} \times \mathbb{Z}_+^{n_{t2}}$ with $n_t = n_{t1} + n_{t2}$. We denote the optimal value by v_{NC}^*.

Under stagewise independence (Assumption 2), the DPE can be written as (15.1)-(15.3), but for distinction we denote the value functions by $Q_{t,NC}(x_{t-1}, \xi_t)$ and the expected value functions by $\mathcal{Q}_{t,NC}(x_{t-1})$ for all $t = 2, \dots, T$. Both, integer variables and non-convex functions make this a non-convex multistage stochastic programming problem (MSNCP). Importantly, $Q_{t,NC}(\cdot, \cdot)$ and $\mathcal{Q}_{t,NC}(\cdot)$ are no longer ensured to be convex, but become non-convex functions in x_{t-1}. They are also not guaranteed to be (Lipschitz) continuous. This poses significant challenges on approximation algorithms such as SDDP, as linear cuts are not sufficient to approximate $\mathcal{Q}_{t,NC}(\cdot)$.

To approach (MSNCP) by SDDP, different strategies can be used. As nonlinear or mixed-integer stochastic programming are large research areas on their own, we give a brief overview here and for methodological details refer to the cited literature.

16.1. Convexification. A standard approach in practice is to solve a static convex relaxation ($\widehat{P}_{NC}$) of (MSNCP), which is associated with convex expected value functions $\widehat{\mathcal{Q}}_t(\cdot)$ for all $t \in [T]$. Such relaxation can be achieved by relaxing the integrality constraints and replacing non-convex functions with convex relaxations, such as McCormick envelopes [139]. In this case, the Benders cuts determined by SDDP can be very loose, though. Therefore, only some rough under-approximation $\widehat{v}^*_{NC}$ of the optimal value v^*_{NC} may be determined. However, sometimes this is considered sufficient to obtain reasonable policies for practical implementation. Also note that even if convex relaxations are considered when running SDDP to compute a policy, the simulation of this policy afterwards can be executed including integrality constraints and non-convex functions.

A second strategy is to keep the subproblems in SDDP non-convex, but to convexify the expected value functions $\mathcal{Q}_{t,NC}(\cdot)$ *in some sense*. Often, in this case, the nonlinearities in (MSNCP) are first relaxed by piecewise linear approximations, such that all subproblems are MILPs [36, 217]. In the backward pass, given some incumbent x^{ik}_{t-1}, for all $t = T, \dots, 2$ and all $\xi_{tj}, j = 1, \dots, q_t$, instead of solving an LP relaxation of the subproblems (2.10) (or its LP dual), a Lagrangian relaxation is solved where the coupling constraints $g_t(x_{t-1}, x_t, \xi_{tj}) \leq 0$ are relaxed. This relaxation can be written as

$$\begin{aligned} \mathfrak{L}^{i+1}_t(\pi_t; x^{ik}_{t-1}, \xi_{tj}) := \min_{x_t} \quad & f_t(x_t, \xi_{tj}) + \mathfrak{Q}_{t+1}(x_t) + \pi_t^\top g_t(x^{ik}_{t-1}, x_t, \xi_{tj}) \\ \text{s.t.} \quad & x_t \in \mathcal{X}_t. \end{aligned}$$

In the Lagrangian dual, this dual function is maximized over all multipliers π_t:

$$v^{i+1}_{t,LD}(x^{ik}_{t-1}, \xi_{tj}) := \max_{\pi_t \geq 0} \mathfrak{L}^{i+1}_t(\pi_t; x^{ik}_{t-1}, \xi_{tj}). \tag{16.1}$$

It is known from the theory on Lagrangian relaxation that the optimal value $v^{i+1}_{t,LD}(x^{ik}_{t-1}, \xi_{tj})$ coincides with the lower convex envelope of $\underline{Q}^{i+1}_{t,NC}(\cdot, \xi_{tj})$ at x^{ik}_{t-1} [75]. Therefore, cuts obtained based on (16.1) are associated with a convexification of the value function. In order to derive utilizable cut formulas from (16.1) some specific conditions have to be satisfied by the constraints. Suppose the constraints $g_t(x_{t-1}, x_t, \xi_t) \leq 0$ can be rewritten as

$$\widehat{g}_t(x_{t-1}) - \bar{g}_t(x_t, \xi_t) \leq 0, \quad \tilde{g}_t(x_t, \xi_t) \leq 0,$$

i.e., the nonlinear function being separable with respect to x_{t-1}, and let π^{ikj}_t denote optimal multipliers in (16.1). Then, in line with Sect. 3.3, *Lagrangian cuts* can be derived as [211]

$$\mathcal{Q}_{t,NC}(x_{t-1}) \geq \alpha^i_{tk} + (\beta^i_{tk})^\top \widehat{g}_t(x_{t-1}),$$

with

$$\begin{aligned} \alpha^i_{tk} &= \sum_{j=1}^{q_t} p_{tj} \Big(\mathfrak{L}_t(\pi^{ikj}_t; x^{ik}_{t-1}, \xi_{tj}) - (\pi^{ikj}_t)^\top \widehat{g}_t(x^{ik}_{t-1}) \Big), \\ \beta^i_{tk} &= \sum_{j=1}^{q_t} p_{tj} \pi^{ikj}_t. \end{aligned}$$

For linear functions $\widehat{g}_t(\cdot)$ and $\bar{g}_t(\cdot, \cdot)$, a similar result is derived in [36].

The obtained Lagrangian cuts provably dominate standard Benders cuts, which can be obtained by solving LP relaxations [211]. However, convergence of SDDP is not guaranteed, since there may still be some duality gap between $v_{t,LD}^{i+1}(x_{t-1}^{ik}, \xi_{tj})$ and $\underline{Q}_{t,NC}^{i+1}(x_{t-1}^{ik}, \xi_{tj})$.

Moreover, generating Lagrangian cuts can be computationally costly. Various methods have been proposed to solve the Lagrangian dual (16.1), such as cutting-plane methods [110], subgradient methods [69, 166] or bundle methods [121], but all of them may take considerable time compared to solving an LP relaxation. Advantageously, even suboptimal Lagrangian multipliers π_t yield valid cuts for $\mathcal{Q}_{t,NC}(\cdot)$.

Instead of a static convexification approach [36], Steeger and Rebennack [209, 211], also apply the above principle in a dynamic fashion by considering DPE for the Lagrangian relaxations in the backward pass.

16.2. Exact Methods. Recently, there has been more research on directly applying the SDDP idea to problems (MSNCP) to avoid the requirement of convexification and to close the optimality gap.

Step Functions. Given that the value functions $Q_{t,NC}(\cdot)$ are monotonically increasing or decreasing, they can be approximated by special step functions instead of affine functions. This idea is incorporated into the SDDP framework in the so-called *mixed-integer dynamic approximation scheme* (MIDAS) [162]. To determine the step functions, mixed-integer linear subproblems have to be solved exactly at each stage and in each iteration. In contrast to the previous approaches, convergence of MIDAS to an approximately optimal policy for (MSNCP) is guaranteed.

SDDiP. For the mixed-integer *linear* case, the *stochastic dual dynamic integer programming* (SDDiP) approach by Zou, Ahmed and Sun [232] allows for the computation of optimal policies for (MSNCP) as long as all state variables x_t are binary (or bounded integer).

Consider the subproblems (2.10), but with binary state variables $x_t \in \{0,1\}^{n_t}$. Similarly to the approaches in [36, 211, 217], Lagrangian dual problems are solved in the backward pass to derive valid cuts. However, in SDDiP a new class of Lagrangian cuts is proposed. The crucial idea is to introduce local copies z_t of the state variables x_{t-1} and to relax the corresponding copy constraints in the Lagrangian relaxation:

$$\begin{aligned} \mathcal{L}_t^{i+1}(\pi_t; x_{t-1}^{ik}, \xi_{tj}) := \min_{x_t} \quad & \big(c_t(\xi_{tj})\big)^\top x_t + \mathfrak{Q}_{t+1}(x_t) + \pi_t^\top (x_{t-1}^{ik} - z_t) \\ \text{s.t.} \quad & x_t \in \mathcal{X}_t(z_t, \xi_t) \\ & z_t \in [0,1]^{d_{a(n)}}. \end{aligned}$$

In the Lagrangian dual, this dual function is maximized over all multipliers π_t:

$$\tilde{v}_{t,LD}^{i+1}(x_{t-1}^{ik}, \xi_{tj}) := \max_{\pi_t} \ \mathcal{L}_t^{i+1}(\pi_t; x_{t-1}^{ik}, \xi_{tj}).$$

Then, Lagrangian cuts can be determined as

$$\mathcal{Q}_{t,NC}(x_{t-1}) \geq \alpha_{tk}^i + (\beta_{tk}^i)^\top x_{t-1}, \tag{16.2}$$

with

$$\alpha_{tk}^{i} = \sum_{j=1}^{q_t} p_{tj} \Big(\mathcal{L}_t(\pi_t^{ikj}; x_{t-1}^{ik}, \xi_{tj}) - (\pi_t^{ikj})^\top x_{t-1}^{ik} \Big),$$

$$\beta_{tk}^{i} = \sum_{j=1}^{q_t} p_{tj} \pi_t^{ikj}.$$

These cuts can be proven to be valid and, in particular, tight, as defined in Lemma 3.2. The key aspect behind this tightness property is that for $x_{t-1} \in \{0,1\}^{n_t}$ the value functions $Q_{t,NC}(\cdot)$ coincide with their lower convex envelopes at all x_{t-1}. Therefore, Lagrangian cuts recovering the latter are also tight for the former.

Moreover, if only dual basic solutions are considered, the cuts (16.2) are also finite in the sense of Lemma 3.2. Therefore, almost sure finite convergence of SDDiP to an optimal policy of (MSNCP) is guaranteed [232].

If the state variables x_t are bounded and general integer or even continuous, they can be componentwise approximated by a (weighted) sum of binary variables in order to apply SDDiP [232]:

$$x_{tj} \approx \sum_{k=1}^{K_{tj}} 2^{k-1} \beta_{tj} \lambda_{tkj},$$

with discretization precision β_{ti} (for integer x_t, $\beta_t = 1$), binary variables $\lambda_{tkj}, k = 1, \ldots, K_{tj}$, and $K_{tj} \in \mathbb{N}$ for all $j = 1, \ldots, n_t$. Under some recourse assumptions, it can be proven that for a sufficiently fine binary expansion, an approximately optimal policy for (MSNCP) is computed. However, it may be challenging to choose an appropriate precision in advance in practice.

SDDiP is applied in the case studies [102], [174] and [232]. In the latter, additional non-convex functions occur in (MSNCP), which are linearized using a Big-M reformulation.

Non-convex Lipschitz cuts. As long as the value functions are assured to be Lipschitz continuous (*e.g.* because the *complete continuous recourse* [232] property is satisfied), the requirement of binary state variables can be dropped. This is exploited by the *stochastic Lipschitz dynamic programming* (SLDP) method proposed by Ahmed et al. in [1], which enhances SDDiP to general MILPs. In contrast to the Lagrangian cuts (16.2), here, two types of non-convex, but Lipschitz continuous cuts are derived to approximate $\mathcal{Q}_{t,NC}(\cdot)$: Reverse-norm cuts, which are constructed by using Lipschitz constants, and augmented Lagrangian cuts, which are based on (16.2), but contain an additional penalization term $-\mu\|x_{t-1} - x_{t-1}^{i}\|$, where μ denotes some user-controlled parameter and $\|\cdot\|$ some arbitrary norm.

This idea is further refined by Zhang and Sun in [229] who propose a new framework to solve multistage non-convex stochastic MINLPs as part of their complexity analysis of SDDP-like algorithms, see Sect. 4. The first key ingredient of their framework is to consider Lipschitz regularizations of the value functions, see Sect. 17.2. This ensures that the considered value functions are Lipschitz continuous without the requirement of restricting recourse assumptions for (MSNCP). The second idea is to construct nonlinear *generalized conjugacy cuts* by solving conjugate dual problems, similar to the approach in SLDP. Whereas of theoretical interest, this method has not been applied in computational experiments yet. In particular, it is not clear how to solve the conjugate dual problems efficiently in general. Moreover, the framework requires the costly solution of MINLP subproblems in each iteration.

Based on concepts from SDDiP and [229], Füllner and Rebennack present a new framework to solve multistage (stochastic) non-convex MINLPs [73]. Here, the original MINLP is outer approximated by MILPs using piecewise linear relaxations, which are iteratively improved in an outer loop. In an inner loop, those MILPs are solved by an SDDP- and NBD-like decomposition scheme, which combines the Lipschitz regularization approach from [229] with binary approximation to generate non-convex cuts. In contrast to SDDiP, the binary approximation is applied only temporarily to derive linear cuts in the lifted binary space, which are then projected back to the original state space. The pointwise maximum of this projection yields a Lipschitz continuous non-convex cut for the value functions. The projection is computationally important, as it allows to construct cuts which are guaranteed to be valid also for the outer loop MINLPs. The binary approximation is dynamically refined within the algorithm, instead of a static choice in advance. Another key difference compared to the approach from [229] is that it is not required to solve MINLPs in each iteration to derive cuts. The cut projection closure for a non-convex and discontinuous value function is illustrated in Figure 14.

Similar to SLDP [1], however, it is required to introduce a potentially large number of auxiliary variables and constraints to express the non-convex approximations by mixed-integer linear constraints. While the framework in [73] is presented for deterministic problems, the inner loop decomposition method can be enhanced to the stochastic case. Therefore, by appropriate modifications of the refinement and stopping criteria, also the larger framework may be enhanced to stochastic problems.

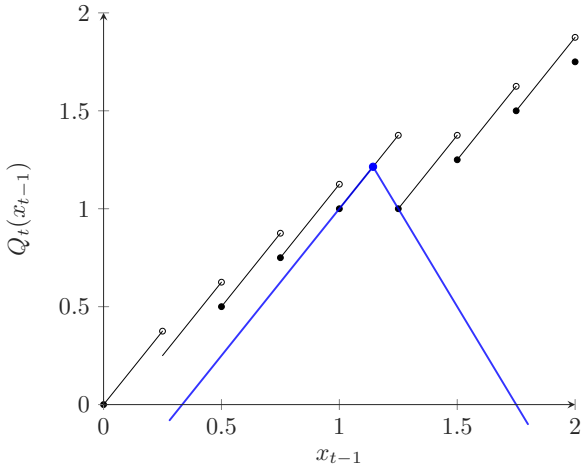


Fig. 14: Non-convex and discontinuous value function with tight non-convex cut.

17. Infeasible Subproblems [relaxing Assumption 9]. Under relatively complete recourse (see Assumption 9), it is guaranteed that any subproblem occurring in the DPE (2.4)-(2.6) and their approximations (2.10) has a feasible solution. As we also assume boundedness, for each of these subproblems there exists some optimal point with finite optimal value. Moreover, all value functions are finite-valued.

In some practical applications, Assumption 9 may not be satisfied. For instance, variable bounds may prevent equality constraints from being satisfied for all x_{t-1} and all realizations of ξ_t, as is illustrated by a toy example in [84]. In such a case, the

primal subproblems become infeasible and the corresponding dual problems become unbounded. Different measures can be taken to cope with infeasibilities.

17.1. Feasibility Cuts. One approach is to approximate the effective domains $\mathrm{dom}(\mathcal{Q}_t)$ of $\mathcal{Q}_t(\cdot)$ by cutting away states $x_{t-1}^{ik} \in \mathcal{X}_t$ leading to infeasible subproblems on stage t. This can be achieved by generating so called *feasibility cuts* in addition to the *optimality cuts* derived in Section 3. These cuts have the form $(\beta_t^f)^\top x_{t-1} \leq \alpha_t^f$, with cut gradient β_t^f, cut intercept α_t^f and the superscript f signifying the cut as a feasibility cut. They can be derived as follows [84].

Consider some stage-t subproblem

$$\underline{Q}_t^i(x_{t-1}^{ik}, \xi_t^k) = \begin{cases} \min\limits_{x_t} & \left(c_t(\xi_t^k)\right)^\top x_t + \mathfrak{Q}_{t+1}^i(x_t) \\ \text{s.t.} & x_t \in \mathcal{X}_t(x_{t-1}^{ik}, \xi_t^k) \\ & (\beta_{t+1}^{fr})^\top x_t \leq \alpha_{t+1}^{fr}, \quad r \in \Gamma_{t+1}^f \end{cases} \tag{17.1}$$

in the forward pass of SDDP. This problem may already contain some feasibility cuts, which are indexed by $r \in \Gamma_{t+1}^f$. To assess feasiblity of problem (17.1) and construct a feasibility cut if required, we consider the auxiliary feasibility problem

$$v_t^f(x_{t-1}^{ik}, \xi_t^k) := \begin{cases} \min\limits_{x_t, y_t^+, y_t^-, z_t} & e^\top y_t^+ + e^\top y_t^- + e^\top z_t \\ \text{s.t.} & W_t(\xi_t^k)x_t + I y_t^+ - I y_t^- = h_t(\xi_t^k) - T_{t-1}(\xi_t^k)x_{t-1}^{ik} \\ & (\beta_{t+1}^{fr})^\top x_t + I z_t \leq \alpha_{t+1}^{fr}, \quad r \in \Gamma_{t+1}^f \\ & x_t \geq 0 \\ & y_t^+, y_t^-, z_t \geq 0. \end{cases}$$

Here, slack variables y_t^+, y_t^- and z_t are introduced to (17.1) to ensure feasibility. The symbol I denotes the identity matrix and e denotes a vector of ones. If we have $v_t^f(x_{t-1}^{ik}, \xi_t^k) = 0$, the subproblem (17.1) is feasible, otherwise, it is infeasible.

By strong duality of linear programs, $v_t^f(x_{t-1}^{ik}, \xi_t^k)$ can be expressed as

$$v_t^f(x_{t-1}^{ik}, \xi_t^k) = \left(h_t(\xi_t^k) - T_{t-1}(\xi_t^k)x_{t-1}^{ik}\right)^\top \sigma_t + \sum_{r \in R_{t+1}^f} (\alpha_{t+1}^{fr})^\top \omega_t^r \tag{17.2}$$

with optimal dual vectors σ_t^{ik} and $\omega_t^{ikr}, r \in R_{t+1}^f$. Then, in case of infeasibility it follows that the term in (17.2) is larger than 0.

To avoid the observed infeasibility on stage t in future iterations, the stage-$(t-1)$ trial point x_{t-1}^{ik} should removed from the feasible set on stage $t-1$. This can be achieved by adding the feasibility cut

$$-(\sigma_t^{ik})^\top T_{t-1}(\xi_t^k)x_{t-1} + (\sigma_t^{ik})^\top h_t(\xi_t^k) + \sum_{r \in R_{t+1}^f} (\omega_t^{ikr})^\top \alpha_{t+1}^{fr} \leq 0 \tag{17.3}$$

to stage $t-1$. By defining

$$\alpha_{t-1}^f := -(\sigma_t^{ik})^\top h_t(\xi_t^k) - \sum_{r \in R_{t+1}^f} (\omega_t^{ikr})^\top \alpha_{t+1}^{fr}$$

and

$$\beta_{t-1}^f := -(\sigma_t^{ik})^\top T_{t-1}(\xi_t^k),$$

the cut (17.3) can be expressed in the previously stated form.

An important question when using feasibility cuts in SDDP is how to proceed, once an infeasible subproblem has been detected and a new feasibility cut (17.3) has been generated. For example, it is possible to stop the forward pass and traverse the stages in backward direction until the root node of the scenario tree is reached. Alternatively, the current subproblem can be resolved to obtain a new trial point x_{t-1}^{ik} and the forward pass can be continued. For SDDP, no assessment and comparison of these strategies has been conducted so far.

Another drawback is that feasibility cuts do not necessarily prevent infeasibilities when the obtained policy is simulated outside of SDDP [84]. For this reason, most commonly, the construction of feasibility cuts is circumvented in SDDP.

17.2. Penalization. Another common approach is to artificially enforce relatively complete recourse for a problem at hand, even if it is not satisfied initially. This can be achieved by using *soft-constraints*, that is, introducing slack variables to relax certain constraints and then penalizing their violation in the objective function. In some applications, this may even be practically justifiable, *e.g.*, in load balance equations in power optimization slack variables can be used to model load shedding or curtailment. However, a reasonable choice of the penalty parameters is not trivial and may distort the expected value functions [84].

Lipschitz Regularization. A specific penalization approach is to consider *Lipschitz regularizations*, also called *Pasch-Hausdorff envelopes* of the value functions. More precisely, let $\|\cdot\|$ denote some norm, $\sigma_t > 0$ some constant and z_t a local stage-t copy of x_{t-1}. Then, by allowing z_t to deviate from the incumbent x_{t-1}^{ik} and penalizing such deviations in the objective, for all $t = 2, \dots, T$ and the approximate value functions (2.10) we obtain the approximate Lipschitz-regularized value functions

$$\underline{Q}_t^{R;i+1}(x_{t-1}^{ik}, \xi_t; \|\cdot\|) := \min_{z_t \geq 0} \left\{ \underline{Q}_t^{i+1}(x_{t-1}^{ik}, \xi_t) + \sigma_t \|z_t - x_{t-1}^{ik}\| \right\}.$$

These functions are proven to be Lipschitz continuous on $\mathbb{R}^{d_{a(n)}}$ with Lipschitz constant σ_t. Moreover, for sufficiently large σ_t for all $t \in [T]$, it can be shown that by considering the regularized problems still the original (MSLP) is solved to optimality [68, 229]. However, choosing σ_t in a sufficient way is an open challenge in practice.

18. No Block-diagonal Structure [relaxing Assumption 7]. A key element of dynamic programming methods is that in the multistage decision process only subsequent stages are linked in the constraints, as it allows one to express (MSLP) using the DPE (2.4)-(2.6). In the single-problem formulation (2.3) of (MSLP), this coincides with a block-diagonal structure, see Assumption 7.

In some cases, it may be relevant to include constraints spanning multiple stages instead. One example is the incorporation of emission quotas that are not allowed to be exceeded for a given time horizon in energy optimization problems [14, 177, 179].

In order to apply SDDP, the considered (MSLP) has to be reformulated to a problem satisfying Assumption 7. This can be achieved by aggregating stages [54], even though this changes the structure, solution and interpretability of (MSLP). An alternative approach is augmenting the state space. For emission quotas, for instance, instead of summing emissions over several stages and comparing them with the upper bound, at a given stage the remaining emission allowances can be considered as an additional state variable [177], see Section 9.

19. Infinite Horizon [relaxing Assumption 1]. So far, we considered problems (MSLP) with a finite time horizon $T < \infty$ (Assumption 1). In some practical applications, however, repeated decisions have to be modeled without a clear bound on the horizon. Considering such infinite-horizon problems is for instance common for Markov decision processes [19]. In such a case, to ensure that v^* is finite, a geometric discount factor $\delta < 1$ is introduced for the cost at each stage.

Since SDDP performs a forward and a backward pass through all stages in each iteration, it is not directly applicable to such problems, as no iteration would ever be completed. Therefore, often different solution methods are utilized in such a setting, see for example [9]. Still, recently there has been some focus on enhancing the SDDP idea to problems with infinite time horizon.

One approach, called Benders squared or B^2, is based on limiting each iteration of SDDP to a finite horizon of τ stages, but to dynamically increase τ per iteration, *e.g.*, by 1, until convergence is reached [144]. By presuming that the uncertainty occurs in the RHS and is not only stagewise independent, but also i.i.d. for all stages $t \in [T]$, almost sure convergence to an approximately optimal policy is assured. The reason is that under this special assumption, $\mathcal{Q}_t(\cdot)$ are the same for all stages, so cuts computed at stage t cannot only be incorporated at stage $t-1$, but at *all* stages [144].

A different option to adopt SDDP to infinite horizon problems exists if such problems possess some kind of periodical behavior. This idea is put forward by Shapiro and Ding [201]. Assume that for some period $m \in \mathbb{N}$, the distributions of ξ_t as well as the data c_t, W_t, h_t and T_{t-1} are the same for $t = \tau$ and $t = \tau + m$ for all $\tau = 2, \dots$ Then, under Assumption 9, the functions $\mathcal{Q}_t(\cdot)$ and $\mathcal{Q}_{t+m}(\cdot)$ are equivalent as well. This means that it is sufficient to derive cuts for $\mathcal{Q}_{t+m}(\cdot)$ at stages $t = 2, \dots, m+1$ in order to obtain valid cuts for all stages.

In the forward pass of SDDP, it is proposed to only consider a finite number of T stages starting from stage 1, with $T \geq m+1$ in order to determine at least one trial point for each of the differing expected value functions. In case of $T > m+1$, multiple candidate trial points exist, at which cuts can be constructed in the backward pass. Before starting the backward pass, the used trial points can be chosen from such a candidate set randomly or by some heuristic.

For both approaches, B^2 and periodic SDDP, for discount factors δ close to 1, the influence of late stages on v^* may be substantial, and thus policy evaluation and upper bound determination may become very challenging and computationally costly. Still, Shapiro and Ding propose some proxies based on some finite, but sufficiently large T [201]. However, they do not provide a convergence proof.

A big advantage of SDDP for periodical problems is that it can also be applied to increase solution performance for problems with a finite, but very large number of stages, given that they satisfy some notion of periodicity. The authors present an example where instead of a 60-month horizon, exploiting the periodical structure of the problem, only a 13-stage problem has to be solved [201]. This can make even large problems amenable to SDDP and computationally tractable. It is also considered to mitigate the so-called *end-of-horizon effect*, which we discuss in Section 9.

On a different note, the policy graph approach introduced by Dowson [56] to model (MSLP) provides a natural extension to infinite-horizon problems, as it allows for cyclic graphs. Solving such problems, similarly to [144], relies on a truncation after a finite number of nodes in the graph. Then, approximate convergence can be proven.

20. Random Horizon [relaxing Assumption 1]. Another way to relax Assumption 1 is to assume that the horizon T is random. For simplicity, we discuss this aspect for the linear case only, even though it is presented in [89] for the more general convex case.

Consider (MSLP) from Sect. 2.3, satisfying Assumptions 2 to 8, but with T not being fixed. Instead, we take the following assumption:

ASSUMPTION 13. *The time horizon T is a discrete random variable taking values in $\{2, \dots, \overline{T}\}$ with known $\overline{T} \in \mathbb{N}$.*

Then, the horizon T induces the Bernoulli process $(\boldsymbol{D}_t)_{t \in [\overline{T}]}$ with realizations

$$D_t = \begin{cases} 0, & \text{if the optimization period ended at } t \text{ or before} \\ 1, & \text{otherwise,} \end{cases}$$

and therefore T can be written as

$$T = \min \left\{ t \in [1, \overline{T}] \ : \ D_t = 0 \right\}.$$

The decisions $\boldsymbol{x}_t(\cdot)$ are functions of $\boldsymbol{\xi}_t$ (given Assumption 2), $\boldsymbol{D}_t$ and $\boldsymbol{D}_{t-1}$. In other words, $\boldsymbol{x}_t$ is $\overline{\mathscr{F}}_t$-measurable with $\overline{\mathscr{F}}_t$ the sigma-algebra $\sigma(\boldsymbol{\xi}_t, \boldsymbol{D}_j, j \leq t)$ [89].

As shown in [89], for (MSLP) with this type of random horizon, the following DPE equations can be derived. Importantly, the state space is augmented by D_{t-1}:

$$Q_t(x_{t-1}, D_t, D_{t-1}, \xi_t) = \min_{x_t \in \mathcal{X}_t(x_{t-1}, \xi_t)} D_{t-1} \big(c_t(\xi_t)\big)^\top x_t + \mathcal{Q}_{t+1}(x_t, D_t),$$

where

$$\mathcal{Q}_{t+1}(x_t, D_t) = \mathbb{E}_{\boldsymbol{\xi}_{t+1}, \boldsymbol{D}_t | D_{t-1}} \left[Q_{t+1}(x_t, \boldsymbol{\xi}_{t+1}) \right]$$

and $\mathcal{Q}_{\overline{T}+1}(x_{\overline{T}}, D_{\overline{T}}) \equiv 0$. For the first stage, we obtain

$$v^* = \min_{x_1 \in \mathcal{X}_1(x_0, \xi_1)} c_1^\top x_1 + \mathcal{Q}_2(x_1, D_1).$$

These DPE are the same as those that would be obtained for a problem with a fixed number of stages $\overline{T} \in \mathbb{N}$, but an objective function including the stagewise dependent stochastic process $(\boldsymbol{D}_t)_{t \in [\overline{T}]}$. As $(\boldsymbol{D}_t)_{t \in [\overline{T}]}$ can be modeled by an inhomogeneous Markov chain with two states, SDDP for processes with Markov chains can be applied [89], see Sect. Section 14.5.

21. Performance Improvements. Apart from extensions to different problem classes, a lot of research on SDDP has focused on improving its computational performance, because standard SDDP may suffer from various performance issues.

As shown in Section 4.2, its worst-case iteration complexity is exponential in the number of stages T and the dimension n_t of the state space, the latter being a well-known drawback of cutting-plane methods in general. Whereas SDDP is successfully applied to various large-scale problems in practice, see Section 9, with the optimality gap closed in reasonable time, especially for problems with a large state space it may empirically fail to converge. For instance, Ávila et al. report instances for which the lower bounds $\underline{v}^i$ already start to stall at a gap of about 22% [6].

In addition to the high number of iterations required, also the computational effort in each iteration can become substantial, even if the number of subproblems

solved per iteration has linear complexity, see Section 4.2. The reason is that with each iteration of SDDP, the subproblems (2.10) become larger, as additional cuts are included. This can increase the computational effort per iteration significantly, especially for problems (MSLP) which require many iterations to converge, and thus many cuts to be generated.

In this section, we give an overview on modifications of SDDP to address these issues and improve its performance. In general, those techniques attempt to either speed up the SDDP iterations by reducing the computational effort, or to reduce the required number of iterations by improving the progress achieved in single iterations.

21.1. Cut Elimination and Selection. As mentioned before, with each added cut, the subproblems (2.10) become larger, and thus potentially harder to solve. However, computational results indicate that SDDP tends to generate a large number of similar or redundant cutting planes, which do not contribute much to the approximation quality in later iterations [6, 203]. Therefore, the computational burden of SDDP may be reduced if only a subset of all cuts is taken into account. However, this requires careful elimination of cuts which are dominated and do not contribute to the solution process, or careful selection of decisive cuts, as otherwise the performance of SDDP may even become worse.

Cut Elimination. One way to reduce the number of cuts is to eliminate some cuts permanently. This can be done by solving an auxiliary problem checking feasibility of the system

$$\begin{cases} \theta_{t+1} \leq \alpha_{t+1}^{\tilde{r}} + (\beta_{t+1}^{\tilde{r}})^\top x_t \\ \theta_{t+1} \geq \alpha_{t+1}^{r} + (\beta_{t+1}^{r})^\top x_t, \quad r \in \Gamma_{t+1} \setminus \{\tilde{r}\} \\ x_t \in X_t \end{cases}$$

for each $\tilde{r} \in \Gamma_{t+1}$, where X_t is assumed to be a compact set [203].

If this system is infeasible, then the cut $\theta_{t+1} \geq \alpha_{t+1}^{\tilde{r}} + (\beta_{t+1}^{\tilde{r}})^\top x_t$ is redundant and can be eliminated. The drawback of this method is that the auxiliary problem has to be solved for all cuts in the system.

A different approach is to permanently store all cuts for each stage t, but only select a subset of those cuts to be considered in the backward pass of the current iteration i. Selection techniques based on this approach are introduced in [8, 49].

Selecting Last Cuts. In this naive strategy, only the $\Gamma \in \mathbb{N}$ most recently added cuts are selected. Although on average, late cuts may provide a better approximation of $\mathcal{Q}_t(\cdot)$ than early ones, this strategy does not guarantee that all important cuts are considered.

Level of Dominance. This strategy is a heuristic in order to consider only non-dominated cuts, but avoid the computational effort of the above cut elimination approach. Using the most basic approach, only cuts are selected, which yield the highest function value at one of the trial solutions considered so far within the algorithm. This is called *Level 1 Dominance* [49]. A similar approach is proposed in [155], but there cuts are permanently removed if they are dominated.

Let x_t^ℓ be the trial solution corresponding to the ℓ-th cut, $\ell \in \Gamma_{t+1}$, and $\phi^r(x_t^\ell)$ the corresponding function value of cut r. Then, the values $v(\ell) := \max_{r \in \Gamma_{t+1}} \{\phi^r(x_t^\ell)\}$ and $r(\ell) := \arg\max_{r \in \Gamma_{t+1}} \{\phi^r(x_t^\ell)\}$ can be saved in a list and be updated every time a new cut is constructed. Similarly, a *Level* Γ *Dominance* strategy can be used, selecting the $\Gamma \in \mathbb{N}$ highest cuts for all trial solutions. Using this strategy, only previous trial points are taken into consideration, though. Therefore, cuts may be excluded which

provide a significant benefit at not yet visited feasible states.

Another challenge is that this strategy draws a lot of resources to store all the required cut information – especially, since the number of visited trial points increases significantly in the course of SDDP. Memory requirements can even be relevant for Level 1, especially if the maximum function value at the trial solutions is attained by several cuts. As a resort, in [87], the *Limited Memory Level 1* strategy is introduced, selecting only the oldest of such cuts. In [8] this technique is applied to SDDP and almost sure convergence is proven.

Dynamic Cut Selection. A dynamic, but also computationally more expensive strategy is to select cuts dynamically within the SDDP framework. In [49] it is proposed to remove all cuts at the beginning of each iteration. Then, for each stage t, each scenario k, and each function $\phi^r(\cdot), r \in \Gamma_{t+1}$, the forward pass subproblem (2.10) is solved. If the current cut yields the highest value at the obtained trial solution, it is added to the subproblem, and the next cut is considered.

This way, only cuts are selected that contribute to the optimal solution in the current iteration. On the other hand, the additional loop may slow down the convergence speed. The computational effort can be reduced by inheriting all added cuts from the already considered scenarios.

A similar approach is considered in [31]. Here, cuts are iteratively added as long as they induce a substantial change in the current optimal value and up to some predefined maximum number of cuts. Instead of iterating over all cuts, in each step, the cut with the highest value at the current incumbent is chosen as a candidate for selection.

Numerical results for sampling about 5,000 scenarios and computing 10,000 cuts in SDDP indicate that all cut selection techniques can significantly speed-up the classical SDDP method [49]. For example, the Level 1 strategy is reported to be ten times faster than SDDP without cut selection. For dynamic cut selection, the reported speed-up is much smaller. It is also shown that the cut selection strategies do not have a significant impact on the quality of the determined policies and bounds. In [8], Limited Memory Level 1 is identified as more efficient than pure Level 1.

21.2. Multi-cut SDDP. In the backward pass of SDDP, for any $t \in [T]$ and any $x_{t-1}^{ik}, k \in \mathcal{K}$, subproblems (2.10) are solved for all noise realizations ξ_{tj}, $j = 1, \dots, q_t$. By taking expected values, a cut (3.5) is derived. Such cuts are then incorporated into the stage-$(t-1)$ subproblem using a single variable $\theta_t \in \mathbb{R}$ by

$$\phi_{tk}^i(x_{t-1}) = (\beta_{tk}^i)^\top x_{t-1} + \alpha_{tk}^i \leq \theta_t,$$

see Section 3.3. This is referred to as a *single-cut* approach.

A different approach, that is well-studied for (nested) Benders decomposition [27, 74, 143], is to not aggregate the dual information, but to generate a separate cut for each noise realizations $\xi_{tj}, j = 1, \dots, q_t$. This requires to introduce variables $\theta_{t,\ell}$ and cut approximations $\mathfrak{Q}_{t+1,\ell}^{i+1}(\cdot)$ for all $\ell = 1, \dots, q_t$ in the stage-t subproblem. In this case, we obtain cuts

$$\phi_{tkj}^i(x_{t-1}) := (\beta_{tkj}^i)^\top x_{t-1} + \alpha_{tkj}^i \leq Q_t(x_{t-1}, \xi_{tj}), \qquad j = 1, \dots, q_t,$$

where, analogously to the derivation in Section 3.3, β_{tkj}^i denotes a subgradient of $\underline{Q}_t^{i+1}(\cdot, \xi_{tj})$ at x_{t-1}^{ik} for $k \in \mathcal{K}, j = 1, \dots, q_t$, and α_{tkj}^i is defined by

$$\alpha_{tkj}^i := \underline{Q}_t^{i+1}(x_{t-1}^{ik}, \xi_t) - (\beta_{tkj}^i)^\top x_{t-1}^{ik}.$$

The expectation is then taken in the objective function instead of the cut formula:

$$\underline{Q}_t^{i+1}(x_{t-1}^{ik}, \xi_{tj}) = \min_{x_t \in \mathcal{X}_t(x_{t-1}, \xi_t)} \quad \big(c_t(\xi_{tj})\big)^\top x_t + \sum_{\ell=1}^{q_{t+1}} p_{t+1,\ell} \mathfrak{Q}_{t+1,\ell}^{i+1}(x_t).$$

This way, more specific information about the value functions is incorporated into the subproblems, hopefully leading to fewer iterations. On the downside, the number of decision variables and cuts grows significantly compared to the single-cut approach, especially if q_t is large, which increases the computational effort for each iteration. Therefore, so far multi-cut SDDP has rarely been considered in the literature. It should be most promising when q_t is only of moderate size. For the two-stage case, a rule of thumb is that a single-cut approach should be preferred if the number of realizations is considerably larger than the number of first-stage constraints [26]. Note that in principle also a trade-off between single-cut and multi-cut is possible by partially aggregating cuts [23, 28]. Another approach to reduce the computational burden of multi-cut SDDP is to combine it with cut selection strategies, see Section 21.1, as proposed in [8]. In this paper, also almost sure finite convergence of multi-cut SDDP is proven.

We return to the illustrative problem from Example 3.3 to illustrate the multi-cut approach.

EXAMPLE 21.1. *(Continuation of Example 3.3) Using multi-cut SDDP, at stage 3, instead of $\mathcal{Q}_3(\cdot)$, the functions $Q_3(\cdot, \xi_3)$ are separately approximated by cuts for $\xi_3 \in \{1, 2, 4\}$. These value functions are displayed in Figure 15. Each of them consists of only two linear pieces, so two cuts are required to represent them exactly. In contrast, $\mathcal{Q}_3(\cdot)$ consists of four linear segments. Therefore, multi-cut SDDP should need less iterations than single-cut SDDP to achieve convergence.*

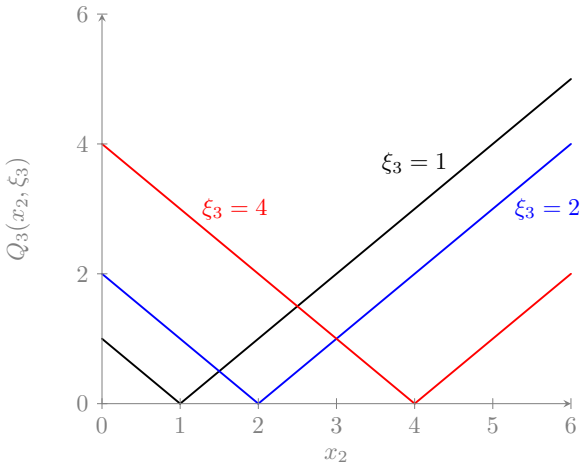


Fig. 15: Stage-3 value functions for Example 3.3.

21.3. Sampling Schemes. SDDP allows to use a variety of different sampling schemes which affect its computational performance.

Number of Forward Samples per Iteration. In standard SDDP, see Section 3, $|\mathcal{K}|$ out of all $|\mathcal{S}|$ scenarios defined by the finite data process $(\boldsymbol{\xi}_t)_{t \in [T]}$ are

sampled in each iteration, with $|\mathcal{K}| \ll |\mathcal{S}|$ and $\mathcal{K} \subset \mathcal{S}$. Philpott and Guan even propose a method with only $|\mathcal{K}| = 1$ for all iterations [163]. This strategy may be particularly efficient in earlier iterations in order to obtain a rough approximation of $\mathcal{Q}_t(\cdot)$ fast without wasting too much effort in regions which are likely to be far from optimal. On the other hand, this strategy faces some drawbacks. Firstly, if the current policy is already reasonably good, but convergence is not achieved yet, it should be beneficial to generate more than one new cut per stage and iteration, and thus choose $\mathcal{K} > 1$ [49]. Secondly, if $|\mathcal{K}| = 1$, then it is not possible to apply a statistical stopping criterion, see Section 7.

Using a *scenario incrementation* strategy in which $|\mathcal{K}|$ is gradually increased, combines the advantages of both approaches – a small number of samples in early iterations and a higher number of samples and a valid statistical stopping criterion in later iterations. This is proposed in [202] and tested in [49].

Subsampling Trial Points. In the *reduced sampling method* (ReSa) [101] the forward pass follows the same principle as in SDDP by sampling scenarios $\xi_t^k, k \in \mathcal{K}$, for $\mathcal{K} \subset \mathcal{S}$. In the backward pass, however, to reduce the number of subproblems to be solved, not all trial points x_{t-1}^{ik} are considered on each stage $t = T, \dots, 2$, but only a subsample $\tilde{\mathcal{K}} \subset \mathcal{K}$ is drawn. This also means that less cuts are derived per iteration than in SDDP (given the same set $\mathcal{K}$). In other words, the number of sample paths through the recombining tree considered in the forward and backward pass may differ. In light of the Central Limit Theorem, this allows to get an accurate statistical upper bound in the forward pass without increasing the computational effort in the backward pass by too much.

Donohue and Birge use a very similar approach in their *abridged nested decomposition* (AND) method [53], an advancement of Birge's NBD method. They claim that SDDP is not well-designed for bushier scenario trees with a high number q_t of noise terms for each stage $t \in [T]$. Their reasoning is that, on the one hand, it is required to choose $|\mathcal{K}|$ reasonably large to get reliable statistical upper bounds, but also to incorporate information on sufficiently many scenarios in the trajectories $(x_t^{ik})_{k \in \mathcal{K}}$. On the other hand, solving the subproblems for all $\xi_{tj}, j = 1, \dots, q_t$, for all $k \in \mathcal{K}$ with large $\mathcal{K}$ may quickly become computational prohibitive. Therefore, the authors propose an alternative sampling scheme.

In the forward pass, on each stage a set $\mathcal{K}_t$ of realizations is sampled. However, the method does not proceed forward from the solutions x_t^{ik} for all $k \in \mathcal{K}_t$ on the next stage, but only from a subset, so-called *branching values*. These branching values $x_t^{i\kappa}$ can either be sampled from $\mathcal{K}_t$ or be a convex combination of all $x_t^{ik}, k \in \mathcal{K}_t$. The latter idea allows to compute trial points which contain information on a large set of scenarios, while not increasing the computational effort in the backward pass tremendously. In the backward pass, the updated subproblems are only solved for all noise terms $\xi_{tj}, j = 1, \dots, q_t$, for all branching values $x_t^{i\kappa}$ on each stage $t = T, \dots, 2$.

The main drawback of AND is that the special structure of the forward pass allows no direct estimate of an upper bound [101]. Therefore, an additional sampling procedure has to be started every few iterations to test for stopping.

Sampling in the Cut Generation Process. The computational effort of generating valid cuts for $\mathcal{Q}_t(\cdot), t = 2, \dots, T$, can be reduced if the subproblems (2.10) are not solved for all noise terms ξ_{tj}, $j = 1, \dots, q_t$, in each iteration, but only for a subsample. The remaining elements that are required to compute a cut for $\mathcal{Q}_t(\cdot)$ can then be used from previous iterations where the corresponding noise ξ_{tj} was sampled.

Even more, if the uncertainty in $(\boldsymbol{\xi}_t)_{t \in [T]}$ is restricted to the RHS h_t of (MSLP),

then the dual feasible set does not depend on ξ_{tj}. Therefore, optimal dual multipliers, which correspond to extreme points of the dual feasible set, and by formula (3.7) also subgradients β_t, can be re-used between different realizations $j = 1, \dots, q_t$.

This allows for the following procedure: Assume that in each iteration i, for each stage $t \in [T]$ only one noise term $\hat{\xi}_t^i$ is sampled and used to compute optimal dual multipliers $\hat{\pi}_t^i$ and (scenario-specific) cut intercepts $\hat{\alpha}_t^i$ as in (21.2). For each stage $t = 2, \dots, T$, all dual multipliers and intercepts obtained up to iteration i are then stored in the set $\mathcal{D}_t^i$ together with ξ_t^i. In other words, in iteration i, this set is updated by $\mathcal{D}_t^i = \mathcal{D}_t^{i-1} \cup \left\{ \left(\hat{\pi}_t^i, \hat{\alpha}_t^i, \hat{\xi}_t^i \right) \right\}$.

For any $\xi_{tj}, j = 1, \dots, q_t$, and a given incumbent x_{t-1}^i, the dual multipliers used to compute a new cut can then be determined based on

$$\left(\hat{\pi}_t^j, \hat{\alpha}_t^j, \hat{\xi}_t^j \right) = \underset{(\hat{\pi}_t, \hat{\alpha}_t, \hat{\xi}_t) \in \mathcal{D}_t^i}{\arg\max} \left\{ \hat{\alpha}_t - \hat{\pi}_t^\top T_{t-1} x_{t-1}^i + \hat{\pi}_t^\top \left(h_t(\xi_{tj}) - h_t(\hat{\xi}_t) \right) \right\}.$$

Hence, not necessarily optimal dual multipliers of the subproblem corresponding to ξ_{tj} are used, but the previously generated ones yielding the best approximation for realization ξ_{tj} at x_{t-1}^i.

Let $\pi_{tj}^i = \hat{\pi}_t^j$ and $\alpha_{tj}^i = \hat{\alpha}_t^j + (\hat{\pi}_t^j)^\top \left(h_t(\xi_{tj}) - h_t(\hat{\xi}_t^j) \right)$ for all $j = 1, \dots, q_t$. Then, a cut can be defined by using subgradient formula (3.7) and taking expectations as in formula (3.4). Note that our description slightly differs from the presentation in the literature, as we adapted it to our cut formulas in Section 3.3.

This idea for the cut generation process is used in two algorithms related to SDDP, which mainly differ by *when* cuts are constructed.

The CUPPS (*convergent cutting-plane and partial-sampling*) method by Chen and Powell [38] does not contain a backward pass, but only a forward pass through the stages, in which both trial points are computed and cuts are generated. This means that the sample $\xi_t^{k'}$ used for the cut generation is the same as drawn for the forward simulation, *i.e.*, $\xi_t^{k'} = \xi_t^k$.

While the computational effort to derive new cuts is reduced, CUPPS has the drawback that the obtained cuts are not necessarily tight. Firstly, the dual multipliers obtained from formula (21.3) are not necessarily optimal for all $j = 1, \dots, q_t$. Secondly, no backward pass is used, and thus new information in form of cuts for stage $t + 1$ are not taken into account when deriving a new cut for stage t.

In the *dynamic outer approximation sampling algorithm* (DOASA) by Philpott and Guan [163], as for SDDP, there exist a forward pass and a backward pass through the stages $t \in [T]$. In the forward pass, a trajectory of trial points $(x_t^{ik})_{k \in \mathcal{K}}$ is computed for all forward samples $k \in \mathcal{K}$ (note that in [163] $|\mathcal{K}| = 1$ is chosen, but this is not mandatory). In the backward pass, cuts are constructed using a backward sample $\xi_t^{k'}$ and formula (21.3). Philpott and Guan prove that this generalization of SDDP also exhibits almost sure finite convergence [163].

21.4. Batch Learning and Experience Replay. While SDDP is used in stochastic programming, dynamic programming or optimal control, its methodology also shares some characteristics with Q-learning algorithms, which are studied in reinforcement learning, see Remark 3.1. This can be exploited by translating established performance enhancing techniques from reinforcement learning to SDDP [6].

As one such technique, Ávila et al. [6] propose to use a *batch learning* technique called *experience replay* in SDDP. The motivation of this is the following: In SDDP, the cut approximations $\mathfrak{Q}_t(\cdot)$ of the expected value functions $\mathcal{Q}_t(\cdot)$ are generated

recursively in a backward pass through the stages $t = T, \dots, 2$. This means that approximation errors at later stages are propagated to earlier stages by means of the cut approximations $\mathfrak{Q}_t(\cdot)$, which then leads to loose cuts at these earlier stages and so on. However, this implies that errors are accumulated at early stages. The authors identify this as a driver for the slow convergence of SDDP, as it favors over-exploring of suboptimal regions and the generation of redundant cuts throughout the iterations.

Experience replay addresses this issue by revisiting previous trial points x_t^i and updating the cut approximations $\mathfrak{Q}_t(\cdot)$ at these points. This seems counterintuitive at first glance because cuts are generated at already visited points instead of improving the approximation of $\mathcal{Q}_t(\cdot)$ at regions of $\mathcal{X}_t$ that have not been visited yet. However, by taking into account all the information currently available to update $\mathfrak{Q}_t(\cdot)$ at x_t^i, it avoids that on earlier stages $\tau < t$ unnecessarily poor approximations of $\mathcal{Q}_t(\cdot)$ at x_t^i are used for several more iterations.

More precisely, the proposed SDDP method works as follows. A predefined number of iterations of standard SDDP are executed and the corresponding trial points x_t^i are stored in a replay memory M_t for all $t \in [T-1]$. When the sizes of the replay memories reach a predefined cardinality Z, then the experience replay step is initiated. This step performs a backward pass through the stages $t = T-1, \dots, 2$. For each stage t, first, a batch $B_t \subseteq M_t$ of trial points is selected from the replay memory (also a full batch $B_t = M_t$ is possible). For each trial point $\tilde{x}_t^\ell$ from this batch, with $\ell \in |B_t|$, the previously generated cut is removed from $\mathfrak{Q}_{t+1}^{i+1}(\cdot)$ and a new cut is constructed by solving the associated subproblems (2.10) (including the experience replay updates from following stages) for $\tilde{x}_t^\ell$. With these cuts, $\mathfrak{Q}_{t+1}^{i}(\cdot)$ is updated and then, the previous stage is explored.

It is shown that experience replay manages to improve the convergence behavior of SDDP, and also the out-of-sample performance of the obtained policies, in computational tests [6]. However, experience replay comes at an increased computational effort, as every Z iterations an additional backward pass solving $q_t|B_t|$ for each stage $t = T, \dots, 2$ has to be performed. For full batches, this adds up to $q_t|\mathcal{K}|Z$ LPs per stage. For this reason, the authors suggest to parallelize both standard SDDP iterations as well as the experience replay. They report computational results which indicate that batch learning is better exploiting parallelism than standard SDDP.

21.5. Regularization. As Kelley's cutting-plane method [110, 145], SDDP exhibits an iteration complexity which is exponential in the dimension n_t of the state variables, see Section 4.2. An unfavorable characteristic of cutting-plane methods, and also of SDDP, in this regard is *zig-zagging* behavior. This means that trial points x_t^i and x_t^{i+1} computed in subsequent iterations can be located far away from each other in different regions of the state space, and that with each new cut the minimum of the subproblems (2.10) is again attained in the respective other region. In particular, this implies that these regions of $\mathcal{X}_t$ experience very tight, but almost redundant approximations $\mathfrak{Q}_t(\cdot)$ of $\mathcal{Q}_t(\cdot)$, while other regions are not properly explored and thus the approximation quality at the true optimum improves very slowly.

In convex and nonsmooth optimization, regularization techniques called *bundle methods* are shown to entail faster convergence than classical cutting-plane methods [121], as they mitigate zig-zagging by stabilizing subsequent trial points around a *stability center* (also called incumbent). Hence, it looks promising to translate these regularization techniques to SDDP.

A common regularization approach, which is predominantly used in two-stage stochastic programming [190, 193], is convex quadratic regularization. Here, some

quadratic deviation of x_t from a stability center $\widehat{x}_t$ is penalized in the objective function for stabilization. An application of quadratic regularization to SDDP is not straightforward, since using a separate stability center for each scenario $s \in \mathcal{S}$ is computationally infeasible due to the exponential growth of $|\mathcal{S}|$ in T [5].

Therefore, Asamov and Powell [5] propose a regularization technique for linear problems, in which stability centers are considered part of the state variable, and thus are the same for all realizations of $\xi_{tj}, j = 1, \dots, q_t$. Then, in the forward pass the objective function is modified to

$$c_t^\top x_t + \mathfrak{Q}_{t+1}^i(x_t) + \frac{\gamma^i}{2}(x_t - \widehat{x}_t^{i-1})^\top H_t (x_t - \widehat{x}_t^{i-1}), \tag{21.1}$$

with a positive semidefinite matrix H_t and some sequence $(\gamma^i)_{i\in\mathbb{N}}$ satisfying $\gamma^i \geq 0$ for all i and $\lim_{i\to\infty} \gamma^i = 0$. The stability centers $\widehat{x}_t^{i-1}$ are chosen as the previous forward pass solution, *i.e.*, the solution is stabilized around a "known" region of the domain of $\mathcal{Q}_t(\cdot)$. This idea is generalized to nonlinear problems and improved in [90] by considering weighted averages of several previous forward pass solutions.

Using objective (21.1), a convex, continuous and linearly constrained quadratic programming problem has to be solved in each forward pass step of SDDP, hopefully, reducing the required number of iterations. Importantly, only the forward pass of SDDP is changed, while the backward pass remains the same. In particular, only LPs have to be solved in the backward pass. As the cuts are still finite (see Lemma 3.2), almost sure finite convergence is assured. In computational tests, it is shown that this method exhibits faster convergence than SDDP, in particular for a high dimension n_t of the state variable x_t [5]. This speed-up is especially important for regularized DDP, see the numerical experiments in [90]. DDP (Dual Dynamic Programming) is the corresponding deterministic counterpart of SDDP (when ξ_t is deterministic for all $t \in [T]$).

While the above approach stabilizes the solution around a "known" region of the domain of $\mathcal{Q}_t(\cdot)$, in a sampling setting, it is not clear whether this is always beneficial. For the current sample ξ_t^k a region may be identified and used for stabilization, which is no appropriate indicator for all $\xi_{tj}, j = 1, \dots, q_t$. Additionally, as pointed out in [224], the condition $\lim_{i\to\infty} \gamma^i = 0$ may evoke that the regularization is diminished and the proposed method in [5] reduces to standard SDDP *before* convergence is obtained, although regularization may be particularly important close to the optimal solution. Therefore, this is claimed to be detrimental to convergence speed [224].

Van Ackooij et al. [224] also address that convergence of proximal bundle methods usually requires the stability centers to be feasible, which is not guaranteed for SDDP subproblems where the feasible set changes with x_{t-1}^i. Therefore, they propose to combine SDDP with a level bundle method, which does not face this requirement.

For stage t and scenario ξ_t^k, trial solutions x_t^{ik} are obtained by solving

$$\begin{cases} \min\limits_{x_t} & \psi_t(x_t) \\ \text{s.t.} & x_t \in \mathbb{X}_t(x_{t-1}^{ik}; \ell_t) \end{cases} \tag{21.2}$$

with $\psi_t(x_t) : \mathbb{R}^{n_t} \to \mathbb{R}$ a given convex function, *e.g.*, $\psi_t(x_t) := x_t^\top x_t$, and

$$\mathbb{X}_t(x_{t-1}^{ik}; \ell_t) := \begin{cases} \arg\min\limits_{x_t \geq 0} & \max\left\{c_t^\top x_t + \mathfrak{Q}_{t+1}^i(x_t), \ell_t\right\} \\ \text{s.t.} & W_t x_t = h_t - T_{t-1} x_{t-1}^{ik}. \end{cases} \tag{21.3}$$

If the maximum in (21.3) is attained by the first term, then x_t^{ik} obtained by solving (21.2) is an ordinary SDDP trial point, referred to as a *normal iterate*. Otherwise, problem (21.2) reduces to a typical level bundle method subproblem, yielding a regularized *level iterate* x_t^{ik}.

The determination of a good level ℓ_t and of an efficient regularization for SDDP are still open questions, and heuristics are proposed in [224] to choose ℓ_t.

An alternative stabilization approach is proposed in [15] based on the concept of Chebyshev centers of polyhedrons. Here, in the forward pass of SDDP, the subproblems (2.10) are modified such that the computed trial states are defined as Chebyshev centers of the polyhedrons given by previously constructed cuts and an appropriate upper bound. It can be shown that this approach is equivalent to modifying the cut formula to

$$-(\beta_{t+1}^r)^\top x_t + \theta_{t+1} \geq \alpha_{t+1}^r + \bar{\sigma}_t \|(1, c_t + \beta_{t+1}^r)\|, \quad r \in \Gamma_{t+1}. \tag{21.4}$$

The authors use the Euclidean norm $\|\cdot\|_2$ in (21.4), however, different choices are possible as well.

Geometrically, the additional term in (21.4) changes the cut intercept, thus lifting the cut. For $\bar{\sigma}_t = 0$, the usual SDDP trial point x_t^i is determined, whereas for $\bar{\sigma}_t > 0$ an offset in the objective compared to the standard SDDP subproblem is considered, yielding a different iterate. To actually improve the performance of SDDP, choosing $\bar{\sigma}_t$ appropriately is crucial, yet not trivial. Adversely, if $\bar{\sigma}_t$ is chosen too large, basically any feasible point can become the new trial solution. Moreover, to ensure convergence, it has to be ensured that $\bar{\sigma}_t$ converges to zero in the course of the algorithm. In [15] heuristics are used to determine $\bar{\sigma}_t$, but it is not clear whether they guarantee performance gains for SDDP.

21.6. Inexact SDDP. Recall Lemma 3.2 (b), stating that the cuts generated in the backward pass of SDDP are *tight* for $\underline{\mathcal{Q}}_t^{i+1}(\cdot)$ at the incumbent x_{t-1}^{ik}. This result is premised on using optimal dual multipliers in the cut formula, *i.e.*, solving the LP subproblem or its dual to global optimality (*exact* solution). Whereas such an exact solution is the standard assumption in the literature on SDDP, computationally it may be more efficient to solve subproblems only approximately, especially early in the solution process when the cut approximations are suboptimal anyway [88].

We first introduce the notion of inexact cuts.

DEFINITION 21.2 (ε-inexact cut). *For any $t = 2, \dots, T$, $\varepsilon > 0$ and a trial point x_{t-1}^{ik}, let $\phi_t : \mathbb{R}^{d_{a(n)}} \to \mathbb{R}$ be an affine function satisfying*

$$\mathcal{Q}_t(x_{t-1}) \geq \underline{\mathcal{Q}}_t^{i+1}(x_{t-1}) \geq \phi_t(x_{t-1}) \quad \textit{(validity)}$$

for all $x_{t-1} \in \mathcal{X}_{t-1}$ and

$$\underline{\mathcal{Q}}_t^{i+1}(x_{t-1}^{ik}) - \phi_t(x_{t-1}^{ik}) \leq \varepsilon \quad \textit{(ε-tightness).}$$

Then, $\phi_t(\cdot)$ defines an ε-inexact cut *at x_{t-1}^{ik} [88].*

Importantly, inexact cuts still yield valid lower approximations of $\mathcal{Q}_t(\cdot)$ for all $t = 2, \dots, T$. We now address how inexact cuts can be determined.

Linear Problems. For any iteration i in SDDP, any $t = 2, \dots, T$ and any trial point x_{t-1}^{ik}, consider the linear subproblem (2.10). In particular, assume that we have relatively complete recourse, *i.e.*, Assumption 9 is satisfied. Also, for simplicity assume that $X_t = \{x_t \in \mathbb{R}^{n_t} \ : \ x_t \geq 0\}$.

For some $\varepsilon > 0$, let π^i_{tjk} be an ε-optimal feasible solution for the dual problem of (2.10) given ξ_{tj} and let θ^i_{tjk} be the corresponding dual objective value for $j = 1, \ldots, q_t$. Then, analogously to Section 3.3, an ε-inexact cut can be defined by [88]

$$\mathcal{Q}_t(x_{t-1}) \geq \phi^i_{tk}(x_{t-1}) := \alpha^i_{tk} + (\beta^i_{tk})^\top x_{t-1},$$

with intercept and subgradient defined by

$$\begin{aligned} \alpha^i_{tk} &= \sum_{j=1}^{q_t} p_{tj}\big(\theta^i_{tjk} - (\beta^i_{tkj})^\top x^{ik}_{t-1}\big), \\ \beta^i_{tk} &= -\sum_{j=1}^{q_t} p_{tj}(\pi^i_{tkj})^\top T_{t-1,j}. \end{aligned}$$

Nonlinear Differentiable Problems. Consider a multistage stochastic convex program (MSCP) as introduced in Section 15, that is, satisfying Assumptions 10 to 12. Moreover, recall the definitions of the Lagrangian function (15.4), the dual function (15.5) and the Lagrangian dual problem (15.6).

Then, an ε-inexact cut can be derived using a pair of approximate primal-dual solutions as follows [88]. Let $\bar{x}_{tj}$ be an ε-optimal feasible primal solution for problem (15.1) given some noise realization ξ_{tj}, $j = 1, \ldots, q_t$, and some trial point $\bar{x}_{t-1}$, and let $\bar{\pi}_{tj}$ be an ε-optimal feasible solution for the corresponding Lagrangian dual (15.6).

We define

$$\eta(\varepsilon) := \ell(\bar{\pi}_{tj}; \bar{x}_{t-1}, \bar{x}_{tj}\xi_{tj}) := \max_{x_t \in X_t} \nabla_{x_t} L_{t,C}(\bar{\pi}_{tj}; \bar{x}_{t-1}, \bar{x}_{tj}, \xi_{tj})^\top (\bar{x}_{tj} - x_t). \tag{21.5}$$

Assume that $f_t(x_t, \xi_{tj})$ takes finite values for all $x_t \in X_t$ and that the term in (21.5) is finite. Then, an ε-inexact cut can be defined by

$$\mathcal{Q}_t(x_{t-1}) \geq \phi^i_{tk}(x_{t-1}) := \alpha^i_{tk} + (\beta^i_{tk})^\top x_{t-1},$$

with intercept and subgradient defined by

$$\begin{aligned} \alpha^i_{tk} &= \sum_{j=1}^{q_t} p_{tj}\big(L_{t,C}(\bar{\pi}_{tj}; \bar{x}_{t-1}, \bar{x}_{tj}, \xi_{tj}) - \eta(\varepsilon) - (\beta^i_{tkj})^\top x^{ik}_{t-1}\big), \\ \beta^i_{tk} &= \sum_{j=1}^{q_t} p_{tj} \nabla_{x_{t-1}} L_{t,C}(\bar{\pi}_{tj}; \bar{x}_{t-1}, \bar{x}_{tj}, \xi_{tj}). \end{aligned}$$

We refer to [88] for a convergence analysis of SDDP using inexact cuts, both for the linear and the nonlinear convex case. In particular, it is shown that the obtained dual solutions are almost surely bounded and that the error terms $\eta(\varepsilon^i_t)$ vanish as i approaches $+\infty$.

Non-differentiable Problems. Using SDDP with inexact cuts is generalized to non-differentiable problems in [91]. In this paper, inexact cuts are derived using two different approaches. In the first approach, it is assumed that the objective and constraint functions have saddle-point representations. The second approach is more general, but requires the introduction of additional variables and constraints.

More precisely, consider a multistage stochastic convex program (MSCP) as introduced in Section 15 and assume that it is satisfying Assumptions 10 and 11 except

for the differentiability properties. Using a local copy z_t of the state variable x_{t-1}, the approximate value functions can be reformulated as

$$Q_{t,C}(x_{t-1},\xi_t) := \begin{cases} \min\limits_{x_t} & f_t(x_t,\xi_t) + \mathcal{Q}_{t+1,C}(x_t) \\ \text{s.t.} & g_t(z_t,x_t,\xi_t) \leq 0 \\ & x_t \in X_t \\ & x_{t-1} = z_t. \end{cases} \tag{21.6}$$

Assume that this modified subproblem satisfies a slater condition analogous to Assumption 12. Additionally, consider the Lagrangian dual problem

$$\max_{\pi_t} \mathcal{L}_t(\pi_t; x_{t-1}, \xi_t). \tag{21.7}$$

with dual function

$$\mathcal{L}_{t,C}(\pi_t; x_{t-1}, \xi_t) = \begin{cases} \min\limits_{x_t \in X_t} & f_t(x_t,\xi_t) + \mathcal{Q}_{t+1,C}(x_t) + \pi_t^\top (x_{t-1} - z_t) \\ \text{s.t.} & g_t(z_t,x_t,\xi_t) \leq 0 \\ & x_t \in X_t, \end{cases}$$

which is obtained by relaxing the copy constraint.

Given a trial point $\bar{x}_{t-1}$ and a noise realization $\xi_{tj}, j = 1,\dots,q_t$, let $\bar{x}_{tj}$ denote an ε_P-optimal feasible solution of problem (21.6) and let $\bar{\pi}_{tj}$ be an ε_D-optimal feasible solution of problem (21.7). Then, an $(\varepsilon_P + \varepsilon_D)$-inexact cut is defined by function

$$\phi^i_{tk}(x_{t-1}) := \sum_{j=1}^{q_t} p_{tj} \Big(f_t(\bar{x}_{tj},\xi_{tj}) - (\varepsilon_P + \varepsilon_D) + \bar{\pi}_{tj}^\top (x_{t-1} - \bar{x}_{t-1}) \Big).$$

For more details and a convergence analysis we refer to [91].

21.7. Parallelization. The performance of SDDP cannot only be improved by modifications of the algorithm itself, but also by its implementation and computational execution. Since several computational steps in SDDP are independent of each other, a performance improvement can be achieved by parallelization.

Different parallelization strategies have been proposed for SDDP. They can be classified with respect to how the workload is distributed among different processors and how the processors are synchronized. Based on this observation, Ávila et al. [6] present a taxonomy of parallelization strategies, which we follow in this section.

Parallelization by Scenario. This is the predominant parallelization strategy for SDDP in the literature. Mostly, a *synchronized* version is proposed. In the forward pass, for all $t \in [T]$, the subproblems (2.10) are solved for $|\mathcal{K}|$ different scenarios, which are sampled independently. The uncertain data ξ_t^k and the trial solutions x_{t-1}^k in each of those problems do only depend on scenario k. Therefore, the different scenarios $\xi^k, k \in \mathcal{K}$, can be assigned to different processors. Assuming P different processors, each processor is assigned $\frac{P}{|\mathcal{K}|}$ scenarios and solves all corresponding subproblems. A master process is then used to aggregate the objective values and compute the upper bound estimate (3.9). This means that there is a synchronization point for all processors at the end of the forward pass.

In the backward pass, a similar approach is followed. The subproblems are again distributed among the processors by scenarios, in such way that for a specific stage $t = T,\dots,2$ and a scenario-based trial point x_{t-1}^k, the subproblems for all noise

realizations $\xi_{tj}, j = 1, \dots, q_t$, are solved by the same processor. Evenly distributing the problems between processors, this way each processor solves $\frac{P}{|\mathcal{K}|} q_t$ subproblems. However, it is also possible to let the master process assign new scenarios to processes once they become idle instead of using a fixed assignment scheme [165].

After solving all associated subproblems, each processor then generates a cut for $\mathcal{Q}_t(\cdot)$ and sends it to the master process. When cut generation is finished for all $k \in \mathcal{K}$, the processors are synchronized so that all of them can proceed with the same set of cuts on stage $t-1$. As stated in [98], this synchronization can be partially relaxed to avoid waiting for single slow processors. Instead, the master process can assign stage-$(t-1)$ subproblems to available processors even if not all cuts have been generated for stage t yet. Numerical results show that such partial relaxation can improve the computational performance of SDDP. However, the number of cuts to wait for to achieve an optimal trade-off between faster iterations and better approximation of $\mathcal{Q}_t(\cdot)$ is problem-dependent.

Even more, an *asynchronous* approach can be used where processors immediately get back to stage $t-1$ after generating their cuts at stage t, using all cuts currently available without waiting for other processes to finish [57].

A major shortcoming of parallelization by scenario is that using more processors becomes more beneficial the more scenarios $|\mathcal{K}|$ are sampled in the forward pass. However, as discussed in Section 21.3, it is often favorable to only consider one or a few scenarios per iteration, especially in earlier iterations. Choosing large $|\mathcal{K}|$ may lead to the accumulation of similar trial points and the generation of redundant cuts [6]. Therefore, exploiting the potential performance gains of additional processors may wrongly incentivize to sample more scenarios than reasonable, thus not accelerating but slowing down the solution process. Additionally, Ávila et al. report computational results indicating that (synchronized) parallelization by scenario scales poorly when increasing the number of samples $|\mathcal{K}|$ due to the combination of long waiting times between processors and low quality cuts [6].

Parallelization by Node. Using parallelization by node, the strategy is to draw only one or a few samples in the forward pass, as this is often computationally preferable. Then, the forward pass is not necessarily parallelized. In the backward pass, the work is distributed among the processors by nodes of the recombining tree (cf. Section 2.1). That means that even for the same $k \in \mathcal{K}$ and the associated trial point x_{t-1}^k, the subproblems (2.10) for different realizations $\xi_{tj}, j = 1, \dots, q_t$, may be solved by different processors. The processors are synchronized at each stage to generate aggregated cuts (given that a single-cut approach is used).

In [6], the authors report clear computational benefits using parallelization by node compared to parallelization by scenario and better scaling properties. However, these results require that the processors can access a shared memory, otherwise the computational overhead is too large. Another drawback is that distributing subproblems for different ξ_{tj}, but the same x_{t-1}^{ik} among different scenarios prevents the exploitation of warm starting techniques.

Parallelization by node can also be used in an asynchronous way, as proposed by Machado et al. [135] in their *asynchronous SDDP* method. In this method, the subproblems of all stages $t = 1, \dots, T$ are solved simultaneously. More precisely, in each *step*, for all stages $t = 1, \dots, T$ and scenarios $k \in \mathcal{K}$, the subproblems for all realizations $\xi_{tj}, j = 1, \dots, q_t$, are solved. Once a processor is finished, it constructs a new cut for $\mathcal{Q}_t(\cdot)$ using all available information. If a required processor has not finished yet, multipliers π_{tkj} from previous steps are re-used. The generated cut can

then be incorporated in stage $t-1$ in the next step. Additionally, each processor generates a new trial point which can be used at stage t in the next step. In contrast to SDDP iterations, this approach requires several steps to propagate information through all stages. Therefore, an ordinary forward pass can only be observed implicitly over several stages. This has to be considered in the computation of upper bounds.

Independent of the applied strategy, parallelizing SDDP in practice comes with considerable challenges, such as communication overhead, problem-dependent performance and lack of reproducibility of results. Therefore, its potential to speed up SDDP in general naturally limited [6].

21.8. Aggregation Techniques. Aggregating information in (MSLP) is another tool with potential to speed up the SDDP solution process.

One approach is to aggregate the variables and constraints of several time periods in a single stage, thus solving a problem with a smaller horizon T. This is straightforward for NBD [54], where each node of the aggregated problem is a subtree of the original scenario tree, even though only few time periods can be aggregated to keep the subproblems tractable. However, it cannot be directly generalized to the sampling and stagewise independent setting in SDDP. The main issue is that it is difficult to model the uncertainty appropriately, without violating non-anticipativity [54].

An alternative approach is to aggregate realizations of $\boldsymbol{\xi}_t$ on each stage (or a subset of stages) [205]. To this end, for some stage t, the support Ξ_t is partitioned into clusters $C_t^\ell, \ell = 1, \dots, L_t$, with $L_t \in \mathbb{N}$. Instead of solving subproblems associated with $\underline{Q}_t^{i+1}(x_{t-1}^{ik}, \xi_{tj})$ for all $j = 1, \dots, q_t$ in the backward pass of SDDP, subproblems associated with $\underline{Q}_t^{i+1}(x_{t-1}^{ik}, \bar{\xi}_t^\ell)$ are solved for clusters $\ell = 1, \dots, L_t$, with $\bar{\xi}_t^\ell := \sum_{j \in C_t^\ell} \frac{p_{tj}}{\bar{p}_t^\ell} \xi_{tj}$, and $\bar{p}_t^\ell$ the probability of cluster C_t^ℓ. This should be beneficial in early iterations where policies are still far away from optimal and a fine information structure unnecessarily slows down the solution process.

Using subgradients and intercepts associated with clusters $C_t^\ell, \ell = 1, \dots, L_t$, *coarse cuts* can be generated for $\mathcal{Q}_t(\cdot)$. Given that W_t and c_t are deterministic, these cuts are valid underestimators for $\mathcal{Q}_t(\cdot)$ by Jensen's inequality [205]. They are not guaranteed to be tight, though.

The authors in [205] discuss several different refinement strategies, such as refinements within the SDDP backward pass (the partition at stage t is refined as soon as a coarse cut does not improve the approximation of $\mathcal{Q}_t(\cdot)$ at the trial point x_{t-1}^{ik}) or refinements outside of SDDP. In the latter case, SDDP is performed on a coarse recombining tree, which is iteratively refined once the algorithm has stopped. Computational results show that this latter approach performs significantly better than the first one due to less computational overhead. However, identifying *when* SDDP should be best stopped to perform a refinement remains a challenging task.

22. Outlook. In this tutorial-type review, we give an overview on the motivation, theory, strengths and weaknesses, extensions and applications of SDDP.

While many proposals have been made in the last 30 years on how to extend SDDP and on how to improve its performance, there still remain open research questions, leaving room for future improvement. Among the most crucial topics are the following.

1. *Stopping.* To this date, in many applications SDDP is stopped heuristically, *e.g.*, based on a fixed number of iterations or stabilization of lower bounds, which leaves the task to define a reasonable stopping criterion to the user. Recently, there has been some pioneering work on developing deterministic upper bounding techniques and stopping criterions, but these are still limited,

as they require significant computational effort.

2. *Upper bounds in risk-averse SDDP.* Developing efficient upper bounding techniques is especially relevant to risk-averse variants of SDDP, where the commonly used nested risk measures do not allow for employment of their pendants from risk-neutral SDDP. Lately, different risk measures have been proposed, which avoid this issue. However, such risk measures usually hamper interpretability. Therefore, it can still be regarded an open question how risk should be optimally measured in SDDP in order to obtain a computationally tractable problem and at the same time to properly reflect the true risk preferences of a decision-maker.
3. *Distributionally robust SDDP.* Recently, the consideration of distributional uncertainty in SDDP has gained more interest. However, while distributionally robust optimization is a flourishing research area, incorporating it into SDDP is still in its early stages, with potential for further improvements.
4. *Non-convex extensions.* In many applications, nonlinear functions or integer variables are required to appropriately model the problem at hand. As the (expected) value functions become non-convex in this case, traditional cutting-plane techniques fail to approximate them correctly. Starting with SDDiP, recently, there has been a trend to extend the NBD and SDDP frameworks to non-convex problems. Lagrangian-type cuts, which are possibly non-convex, show theoretical potential in approximating non-convex functions. However, their construction is computationally costly and subject to rather strong technical assumptions, such that especially large-scale non-convex problems remain computationally intractable. Consequently, in the future, the trade-off between computationally efficient cut generation techniques and best possible approximations of the value functions needs to be further explored.
5. *Regularization.* As a descendant of Kelley's cutting-plane method, SDDP has a computational complexity which grows exponentially in the dimension of the state variables. Therefore, it can become computationally intractable for problems with high-dimensional state space. This is aggravated by common reformulations, *e.g.* in case of stagewise dependent uncertainty, that artificially augment the state space. For Kelley's method, regularization methods have proven helpful in accelerating the solution process. Whereas some first attempts have been made to regularize SDDP, an efficient regularization remains an open challenge.
6. *Reinforcement learning techniques.* As the case of batch learning shows, SDDP can benefit from acceleration techniques that are well-known and established in reinforcement learning, but have not been translated to SDDP setting yet. By exploiting its affinity to Q-learning, there should be a lot of potential to improve the computational performance of SDDP in practice.
7. *Decision-dependent uncertainty.* The only standard assumption for SDDP that has not been relaxed in the literature yet, is to allow for stagewise-dependent stochastic processes modeling the uncertainty in (MSLP). This topic has still to be studied.

Acknowledgments. The authors thank the anonymous referees and the editor for their detailed and insightful feedback which helped to improve earlier versions of this manuscript substantially.

REFERENCES

[1] S. Ahmed, F. G. Cabral, and B. Freitas Paulo da Costa. Stochastic Lipschitz dynamic programming. *Mathematical Programming*, 191:755–793, 2022.
[2] D. Arjoon, Y. Mohamed, Q. Goor, and A. Tilmant. Hydro-economic risk assessment in the eastern Nile River basin. *Water Resources and Economics*, 8:16–31, 2014.
[3] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.
[4] N. V. Arvanitidis and J. Rosing. Composite representation of a multireservoir hydroelectric power systems. *IEEE Transactions on Power Apparatus and Systems*, 89(2):319–326, 1970.
[5] T. Asamov and W. B. Powell. Regularized decomposition of high-dimensional multistage stochastic programs with markov uncertainty. *SIAM Journal on Optimization*, 28(1):575–595, 2018.
[6] D. Ávila, A. Papavasiliou, and N. Löhndorf. Batch learning SDDP for long-term hydrothermal planning. *IEEE Transactions on Power Systems*, 2023.
[7] H. Bakker, F. Dunke, and S. Nickel. A structuring review on multi-stage optimization under uncertainty: Aligning concepts from theory and practice. *Omega*, 96, 2020.
[8] M. Bandarra and V. Guigues. Single cut and multicut stochastic dual dynamic programming with cut selection for multistage stochastic linear programs: convergence proof and numerical experiments. *Computational Management Science*, 18:125–141, 2021.
[9] R. Baucke. An algorithm for solving infinite horizon markov dynamic programmes. Preprint, available online at http://www.optimization-online.org/DB_FILE/2018/04/6565.pdf, 2018.
[10] R. Baucke, A. Downward, and G. Zakeri. A deterministic algorithm for solving multistage stochastic programming problems. Preprint, available online at http://www.optimization-online.org/DB_FILE/2017/07/6138.pdf, 2017.
[11] R. Baucke, A. Downward, and G. Zakeri. A deterministic algorithm for stochastic minimax dynamic programmes. Preprint, available online at http://www.optimization-online.org/DB_FILE/2018/02/6449.pdf, 2018.
[12] G. Bayraksan and D. K. Love. Data-driven stochastic programming using phi-divergences. In *Tutorials in Operations Research: The Operations Research Revolution*, pages 1–19. INFORMS, 2015.
[13] R. E. Bellman. *Dynamic programming*. Princeton University Press, Princeton, New Jersey, 1957.
[14] M.M. Belsnes, A. Haugstad, B. Mo, and P. Markussen. Quota modeling in hydrothermal systems. In *IEEE Bologna Power Tech Conference Proceedings*, 2003.
[15] F. Beltrán, E. C. Finardi, G. M. Fredo, and W. de Oliveira. Improving the performance of the stochastic dual dynamic programming algorithm using Chebyshev centers. *Optimization and Engineering*, 2020.
[16] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton University Press, 2009.
[17] J. F. Benders. Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4(1):238–252, 1962.
[18] J. M. Velásquez Bermúdez. GDDP: generalized dual dynamic programming theory. *Annals of Operations Research*, 117(1-4):21–31, 2002.
[19] D. P. Bertsekas. *Dynamic Programming and Optimal Control, Volume II*. Athena Scientific, 4 edition, 2017.
[20] D. Bertsimas, D. B. Brown, and C. Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501, 2011.
[21] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah. Julia: a fresh approach to numerical computing. *SIAM Review*, 59(1):65–98, 2017.
[22] A. Bhattacharya, J. P. Kharoufeh, and B. Zeng. Managing energy storage in microgrids: a multistage stochastic programming approach. *IEEE Transactions on Smart Grid*, 9(1):483–496, 2018.
[23] M. Biel and M. Johansson. Dynamic cut aggregation in L-shaped algorithms. Preprint, available online at https://arxiv.org/abs/1910.13752, 2019.

[24] J. R. Birge. Solution methods for stochastic dynamic linear programs. Technical report, Stanford University CA Systems Optimization Lab, 1980.
[25] J. R. Birge. Decomposition and partitioning methods for multistage stochastic linear programs. *Operations Research*, 33(5):989–1007, 1985.
[26] J. R. Birge and F. Louveaux. *Introduction to stochastic programming*. Springer Series in Operations Research and Financial Engineering. Springer Science & Business Media, 2nd edition, 2011.
[27] J. R. Birge and F. V. Louveaux. A multicut algorithm for two-stage stochastic linear programs. *European Journal of Operational Research*, 34:384–392, 1988.
[28] X. Blanchot, F. Clautiaux, B. Detienne, A. Froger, and M. Ruiz. The Benders by batch algorithm: Design and stabilization of an enhanced algorithm to solve multicut Benders reformulation of two-stage stochastic programs. *European Journal of Operational Research*, 309(1):202–216, 2023.
[29] J. F. Bonnans, Z. Cen, and T. Christel. Energy contracts management by stochatic programming techniques. *Annals of Operations Research*, 200:199–222, 2012.
[30] R. B. S. Brandi, A. L. M. Marcato, B. H. Dias, T. P. Ramos, and I. C. da Silva Junior. A convergence criterion for stochastic dual dynamic programming: application to the long-term operation planning problem. *IEEE Transactions on Power Systems*, 33(4):3678–3690, 2018.
[31] R. B. S. Brandi, T. P. Ramos, B. H. Dias, A. L. M. Marcato, and I. Chaves da Silva Junior. Improving stochastic dynamic programming on hydrothermal systems through an iterative process. *Electric Power Systems Research*, 123:147–153, 2015.
[32] A. Brigatto, A. Street, and D. M. Valladão. Assessing the cost of time-inconsistent operation policies in hydrothermal power systems. *IEEE Transactions on Power Systems*, 32(6):4541–4550, 2017.
[33] S. Bruno, S. Ahmed, A. Shapiro, and A. Street. Risk neutral and risk averse approaches to multistage renewable investment planning under uncertainty. *European Journal on Operational Research*, 250:979–989, 2016.
[34] L. Cambier and D. Scieur. FAST (Finally An SDDP Toolbox). Accessed April 27, 2021, https://stanford.edu/~lcambier/cgi-bin/fast/index.php.
[35] P. Carpentier, J-.P. Chancelier, G. Cohen, M. De Lara, and P. Girardeau. Dynamic consistency for stochastic optimal control problems. *Annals of Operations Research*, 200:247–263, 2012.
[36] S. Cerisola, J. M. Latorre, and A. Ramos. Stochastic dual dynamic programming applied to nonconvex hydrothermal models. *European Journal of Operational Research*, 218(3):687–697, 2012.
[37] R. M. Chabar, M. V. F. Pereira, S. Granville, L. A. Barroso, and N. A. Iliadis. Optimization of fuel contracts management and maintenance scheduling for thermal plants under price uncertainty. In *IEEE Power Systems Conference and Exposition*, pages 923–930. IEEE, 2006.
[38] Z.-L. Chen and W. B. Powell. A convergent cutting-plane and partial-sampling algorithm for multistage stochastic linear programs with recourse. *Journal of Optimization Theory and Applications*, 102(3):497–524, 1999.
[39] A. Chiralaksanakul and D. P. Morton. Assessing policy quality in multi-stage stochastic programming. *Stochastic Programming E-Print Series*, 12, 2004.
[40] B. F. P. da Costa and V. Leclère. Dual SDDP for risk-averse multistage stochastic programs. *Operations Research Letters*, 51(3):332–337, 2023.
[41] L. C. da Costa Jr., F. S. Thomé, J. Dias Garcia, and M. V. F. Pereira. Reliability-constrained power system expansion planning: a stochastic risk-averse optimization approach. *IEEE Transactions on Power Systems*, 36(1):97–106, 2021.
[42] E. L. da Silva and E. C. Finardi. Parallel processing applied to the planning of hydrothermal systems. *IEEE Transactions on Parallel and Distributed Systems*, 14(8):721–729, 2003.
[43] G. B. Dantzig and G. Infanger. Multi-stage stochastic linear programs for portfolio optimization. *Annals of Operations Research*, 45:59–76, 1993.
[44] G.B. Dantzig. Linear programming under uncertainty. *Management Science*, 1(3/4):197–206, 1955.
[45] C. M. B. de Castro, A. L. M. Marcato, R. Castro Souza, I. Chaves Silva Junior, F. L. Cyrino Oliveira, and T. Pulinho. The generation of synthetic inflows via bootstrap to increase the energy efficiency of long-term hydrothermal dispatches. *Electric Power Systems Research*, 124:33–46, 2015.
[46] M. de Lara and V. Leclère. Building up time-consistency for risk-measures and dynamic optimization. *European Journal of Operational Research*, 249:177–187, 2016.

[47] V. L. de Matos and E. C. Finardi. A computational study of a stochastic optimization model for long term hydrothermal scheduling. *International Journal of Electrical Power & Energy Systems*, 43(1):1443–1452, 2012.

[48] V. L. de Matos, D. P. Morton, and E. C. Finardi. Assessing policy quality in a multistage stochastic program for long-term hydrothermal scheduling. *Annals of Operations Research*, 253:713–731, 2017.

[49] V. L. de Matos, A. B. Philpott, and E. C. Finardi. Improving the performance of Stochastic Dual Dynamic Programming. *Journal of Computational and Applied Mathematics*, 290:196–208, 2015.

[50] L. Ding, S. Ahmed, and A. Shapiro. A Python package for multi-stage stochastic programming. Preprint, available online at http://www.optimization-online.org/DB_FILE/2019/05/7199.pdf, 2019.

[51] T. Ding, X. Zhang, R. Lu, M. Qu, M. Shahidehpour, Y. He, and T. Chen. Multi-stage distributionally robust stochastic dual dynamic programming to multi-period economic dispatch with virtual energy storage. *IEEE Transactions on Sustainable Energy*, 13(1):1–13, 2022.

[52] A. L. Diniz, M. E. P. Maceira, C. L. V. Vasconcellos, and D. D. J. Penna. A combined SDDP/Benders decomposition approach with a risk-averse surface concept for reservoir operation in long term power generation planning. *Annals of Operations Research*, 292:649–681, 2020.

[53] C. J. Donohue and J. R. Birge. The abridged nested decomposition method for multistage stochastic linear programs with relatively complete recourse. *Algorithmic Operations Research*, 1(1):20–30, 2006.

[54] T. N. dos Santos and A. L. Diniz. A new multiperiod stage definition for the multistage Benders decomposition approach applied to hydrothermal scheduling. *IEEE Transactions on Power Systems*, 24(3):1383–1392, 2009.

[55] A. Downward, O. Dowson, and R. Baucke. Stochastic dual dynamic programming with stagewise-dependent objective uncertainty. *Operations Research Letters*, 48:33–39, 2020.

[56] O. Dowson. The policy graph decomposition of multistage stochastic programming problems. *Networks*, 76(1):3–23, 2020.

[57] O. Dowson and L. Kapelevich. SDDP.jl: a Julia package for stochastic dual dynamic programming. *INFORMS Journal on Computing*, 33(1):27–3, 2021.

[58] O. Dowson, D. P. Morton, and A. Downward. Bi-objective multistage stochastic linear programming. *Mathematical Programming*, 196:907–933, 2022.

[59] O. Dowson, D. P. Morton, and B. K. Pagnoncelli. Partially observable multistage stochastic programming. *Operations Research Letters*, 48:505–512, 2020.

[60] O. Dowson, D. P. Morton, and B. K. Pagnoncelli. Incorporating convex risk measures into multistage stochastic programming algorithms. *Annals of Operations Research*, 2022.

[61] O. Dowson, A. Philpott, A. Mason, and A. Downward. A multi-stage stochastic optimization model of a pastoral dairy farm. *European Journal of Operations Research*, 274(3):1077–1089, 2019.

[62] I. Dunning, J. Huchette, and M. Lubin. JuMP: a modeling language for mathematical optimization. *SIAM Review*, 59(2):295–320, 2017.

[63] J. Dupačová and V. Kozmík. SDDP for multistage stochastic programs: preprocessing via scenario reduction. *Computational Management Science*, 14(1):67–80, 2017.

[64] J. Dupačová and V. Kozmík. Structure of risk-averse multistage stochastic programs. *OR Spectrum*, 37:559–582, 2015.

[65] D. Duque and D. P. Morton. Distributionally robust stochastic dual dynamic programming. *SIAM Journal on Optimization*, 30(4):2841–2865, 2020.

[66] J. L. Durante, J. Nascimento, and W. B. Powell. Risk directed importance sampling in stochastic dual dynamic programming with hidden Markov models for grid level energy storage. Preprint, available online at https://arxiv.org/pdf/2001.06026.pdf, 2020.

[67] A. Eichhorn and W. Römisch. Polyhedral risk measures in stochastic programming. *SIAM Journal on Optimization*, 16(1):69–95, 2005.

[68] M.J. Feizollahi, S. Ahmed, and X. A. Sun. Exact augmented Lagrangian duality for mixed integer linear programming. *Mathematical Programming*, 161(1-2):365–387, 2017.

[69] M. L. Fisher. An applications oriented guide to Lagrangian relaxation. *Interfaces*, 15(2):10–21, 1985.

[70] B. C. Flach, L. A. Barroso, and M. V. F. Pereira. Long-term optimal allocation of hydro generation for a price-maker company in a competitive market: latest developments and a stochastic dual dynamic programming approach. *IET Generation, Transmission & Distribution*, 4(2):299–314, 2010.

[71] M. Forcier and V. Leclère. Convergence of trajectory following dynamic programming algorithms for multistage stochastic problems without finite support assumptions. Preprint, available online at https://optimization-online.org/wp-content/uploads/2022/06/8946.pdf, 2023.
[72] S.M. Frank and S. Rebennack. An introduction to optimal power flow: theory, formulation, and examples. *IIE Transactions*, 48(12):1172–1197, 2016.
[73] C. Füllner and S. Rebennack. Non-convex nested Benders decomposition. *Mathematical Programming*, 196:987–1024, 2022.
[74] H. I. Gassmann. MSLiP: A computer code for the multistage stochastic linear programming problem. *Mathematical Programming*, 47:407–423, 1990.
[75] A. M. Geoffrion. Lagrangean relaxation for integer programming. In M.L. Balinski, editor, *Approaches to integer programming*, pages 82–114. Springer Berlin Heidelberg, Berlin, Heidelberg, 1974.
[76] A. Georghiou, A. Tsoukalas, and W. Wiesemann. Robust Dual Dynamic Programming. *Operations Research*, 67(3):813–830, 2019.
[77] H. Gevret, N. Langrené, J. Lelong, R. Lobato, T. Ouillon, X. Warin, and A. Maheshwari. STochastic OPTimization library in C++. Technical report, EDF Lab, 2018.
[78] P. Girardeau, V. Leclère, and A. B. Philpott. On the convergence of decomposition methods for multistage stochastic convex programs. *Mathematics of Operations Research*, 40(1):130–145, 2015.
[79] A. Gjelsvik, M. M. Belsnes, and A. Haugstad. An algorithm for stochastic medium-term hydrothermal scheduling under spot price uncertainty. In *Proceedings of 13th power systems computation conference*, 1999.
[80] A. Gjelsvik, M. M. Belsnes, and M. Håland. A case of hydro scheduling with a stochastic price model. In E. Broch, D.K. Lysne, N. Flatabø, and E. Helland-Hansen, editors, *Procedings of the 3rd international conference on hydropower*, pages 211–218. A.A. Balkema, 1997.
[81] A. Gjelsvik, B. Mo, and A. Haugstad. Long- and medium-term operations planning and stochastic modelling in hydro-dominated power systems based on stochastic dual dynamic programming. In S. Rebennack, P.M. Pardalos, M.V.F. Pereira, and N.A. Iliadis, editors, *Handbook of power systems*, Energy Systems. Springer-Verlag Berlin Heidelberg, 2010.
[82] A. Gjelsvik and S. W. Wallace. Methods for stochastic medium-term scheduling in hydro-dominated power systems. Technical report, Norwegian Electric Power Research Institute, Trondheim, 1996.
[83] Z. Guan and A. B. Philpott. A multistage stochastic programming model for the New Zealand dairy industry. *International Journal of Production Economics*, 134(2):289–299, 2011.
[84] V. Guigues. SDDP for some interstage dependent risk-averse problems and application to hydro-thermal planning. *Computational Optimization and Applications*, 57:167–203, 2014.
[85] V. Guigues. Convergence analysis of sampling-based decomposition methods for risk-averse multistage stochastic convex programs. *SIAM Journal on Optimization*, 26(4):2468–2494, 2016.
[86] V. Guigues. DASC: a Decomposition Algorithm for multistage stochastic programs with Strongly Convex cost functions. Preprint, available online at https://arxiv.org/pdf/1711.04650.pdf, 2017.
[87] V. Guigues. Dual dynamic programing with cut selection: convergence proof and numerical results. *European Journal of Operational Research*, 258:47–57, 2017.
[88] V. Guigues. Inexact cuts in stochastic dual dynamic programming. *SIAM Journal on Optimization*, 30(1):407–438, 2020.
[89] V. Guigues. Multistage stochastic programs with a random number of stages: dynamic programming equations, solution methods, and application to portfolio selection. *Optimization Methods and Software*, 36(1):211–236, 2021.
[90] V. Guigues, M. A. Lejeune, and W. Tekaya. Regularized stochastic dual dynamic programming for convex nonlinear optimization problems. *Optimization and Engineering*, 21:1133–1165, 2020.
[91] V. Guigues, R. Monteiro, and B. Svaiter. Inexact cuts in SDDP applied to multistage stochastic nondifferentiable problems. *SIAM Journal on Optimization*, 31(3):2084–2110, 2021.
[92] V. Guigues and R. D. C. Monteiro. Stochastic dynamic cutting plane for multistage stochastic convex programs. *Journal of Optimization Theory and Applications*, 189:513–559, 2021.
[93] V. Guigues and W. Römisch. Sampling-based decomposition methods for multistage stochastic programs based on extended polyhedral risk measures. *SIAM Journal on Optimization*, 22(2):286–312, 2012.

[94] V. Guigues and W. Römisch. SDDP for multistage stochastic linear programs based on spectral risk measures. *Operations Research Letters*, 40:313–318, 2012.
[95] V. Guigues and C. Sagastizábal. The value of rolling-horizon policies for risk-averse hydrothermal planning. *European Journal of Operational Research*, 217:129–140, 2012.
[96] V. Guigues and C. Sagastizábal. Risk-averse feasible policies for large-scale multistage stochastic linear programs. *Mathematical Programming*, 138(1-2):167–198, 2013.
[97] V. Guigues, A. Shapiro, and Y. Cheng. Duality and sensitivity analysis of multistage linear stochastic programs. *European Journal of Operational Research*, 308(2):752–767, 2023.
[98] A. Helseth and H. Braaten. Efficient parallelization of the stochastic dual dynamic programming algorithm applied to hydropower scheduling. *Energies*, 8(12):14287–14297, 2015.
[99] A. Helseth, M. Fodstad, and B. Mo. Optimal medium-term hydropower scheduling considering energy and reserve capacity markets. *IEEE Transactions on Sustainable Energy*, 7(3):934–942, 2016.
[100] A. Helseth, B. Mo, M. Fodstad, and M. N. Hjelmeland. Co-optimizing sales of energy and capacity in a hydropower scheduling model. In *IEEE Power Tech*, pages 1–6, Eindhoven, The Netherlands, 2015.
[101] M. Hindsberger. ReSa: A method for solving multistage stochastic linear programs. *Journal of Applied Operational Research*, 6(1):2–15, 2014.
[102] M. N. Hjelmeland, J. Zou, A. Helseth, and S. Ahmed. Nonconvex medium-term hydropower scheduling by stochastic dual dynamic integer programming. *IEEE Transactions on Sustainable Energy*, 10(1), 2019.
[103] T. Homem-de-Mello and G. Bayraksan. Monte Carlo sampling-based methods for stochastic optimization. *Surveys in Operations Research and Management Science*, 19:56–85, 2014.
[104] T. Homem-de-Mello, V. L. De Matos, and E. C. Finardi. Sampling strategies and stopping criteria for stochastic dual dynamic programming: a case study in long-term hydrothermal scheduling. *Energy Systems*, 2(1):1–31, 2011.
[105] T. Homem-de-Mello and B. K. Pagnoncelli. Risk aversion in multistage stochastic programming: A modeling and algorithmic perspective. *European Journal of Operational Research*, 249:188–199, 2016.
[106] J. Huang, K. Zhou, and Y. Guan. A study of distributionally robust multistage stochastic optimization. Preprint, available online at https://arxiv.org/abs/1708.07930, 2017.
[107] N.A. Iliadis, V.F. Perira, S. Granville, M. Finger, P.A. Haldi, and L.A. Barroso. Benchmarking of hydroelectric stochastic risk management models using financial indicators. In *2006 IEEE power engineering society general meeting*, 2006.
[108] G. Infanger and D. P. Morton. Cut sharing for multistage stochastic linear programs with interstage dependency. *Mathematical Programming*, 75(2):241–256, 1996.
[109] J. Kallrath, P. M. Pardalos, S. Rebennack, and M. Scheidt, editors. *Optimization in the energy industry*. Springer, 2009.
[110] J. E. Kelley. The cutting-plane method for solving convex programs. *Journal of the Society for Industrial and Applied Mathematics*, 8(4):703–712, 1960.
[111] A. Kiszka and D. Wozabal. Stochastic dual dynamic programming for optimal power flow problems under uncertainty. Preprint, available online at https://optimization-online.org/wp-content/uploads/2021/11/8689.pdf, 2021.
[112] V. Kozmík. On variance reduction of mean-CVaR Monte Carlo estimators. *Computational Management Science*, 12(2):221–242, 2015.
[113] V. Kozmík and D. P. Morton. Evaluating policies in risk-averse multi-stage stochastic programming. *Mathematical Programming*, 152:275–300, 2015.
[114] V. Krasko and S. Rebennack. Two-stage stochastic mixed-integer nonlinear programming model for post-wildfire debris flow hazard management: mitigation and emergency evacuation. *European Journal of Operational Research*, 263(1):265–282, 2017.
[115] T. Kristiansen. Hydropower scheduling and financial risk management. *Optimal Control Applications and Methods*, 27:1–18, 2006.
[116] G. Lan. Complexity of stochastic dual dynamic programming. *Mathematical Programming*, 191:717–754, 2022.
[117] Y. Lan, Q. Zhai, X. Liu, and X. Guan. Fast stochastic dual dynamic programming for economic dispatch in distribution systems. *IEEE Transactions on Power Systems*, pages 1–13, 2022.
[118] V. Leclère, P. Carpentier, J.-P. Chancelier, A. Lenoir, and F. Pacaud. Exact converging bounds for stochastic dual dynamic programming via Fenchel duality. *SIAM Journal on Optimization*, 30(2):1223–1250, 2020.
[119] V. Leclère, F. Pacaud, T. Rigaut, and H. Gerard. StochDynamicProgramming.jl. Code released on GitHub https://github.com/JuliaStochOpt/StochDynamicProgramming.jl.

[120] B. Legat. StructDualDynProg.jl. Code released on GitHub https://github.com/JuliaStochOpt/StructDualDynProg.jl.
[121] C. Lemaréchal, A. Nemirovskii, and Y. Nesterov. New variants of bundle methods. *Mathematical Programming*, 69(1-3):111–147, 1995.
[122] C. M. Leocadio, C. S. G. Richa, M. Z. Fortes, V. H. Ferreira, and B. H. Dias. Economic evaluation of a CHP biomass plant using stochastic dual dynamic programming. *Electrical Engineering*, 102:2605–2615, 2020.
[123] L. M. M. Lima, E. Popova, and P. Damien. Modeling and forecasting of Brazilian reservoir inflows via dynamic linear models. *International Journal of Forecasting*, 30(3):464–476, 2014.
[124] K. Linowsky and A. B. Philpott. On the convergence of sampling-based decomposition algorithms for multistage stochastic programs. *Journal of Optimization Theory and Applications*, 125(2):349–366, 2005.
[125] R. P. Liu and A. Shapiro. Risk neutral reformulation approach to risk averse stochastic programming. *European Journal of Operational Research*, 286(1):21–31, 2020.
[126] T. Lohmann, A. S. Hering, and S. Rebennack. Spatio-temporal hydro forecasting of multireservoir inflows for hydro-thermal scheduling. *European Journal of Operational Research*, 255(1):243–258, 2016.
[127] N. Löhndorf. An empirical analysis of scenario generation methods for stochastic optimization. *European Journal of Operational Research*, 255:121–132, 2016.
[128] N. Löhndorf and A. Shapiro. Modeling time-dependent randomness in stochastic dual dynamic programming. *European Journal of Operational Research*, 273(2):650–661, 2019.
[129] N. Löhndorf and D. Wozabal. Gas storage valuation in incomplete markets. *European Journal of Operational Research*, 2020.
[130] N. Löhndorf, D. Wozabal, and S. Minner. Optimizing trading decisions for hydro storage systems using approximate dual dynamic programming. *Operations Research*, 61(4):810–823, 2013.
[131] R. Lu, T. Ding, B. Qin, J. Ma, X. Fang, and Z. Dong. Multi-stage stochastic programming to joint economic dispatch for energy and reserve with uncertain renewable energy. *IEEE Transactions on Sustainable Energy*, 11(3):1140–1151, 2020.
[132] M. E. P. Maceira and J. M. Damázio. Use of the PAR (p) model in the stochastic dual dynamic programming optimization scheme used in the operation planning of the Brazilian hydropower system. *Probability in the Engineering and Informational Sciences*, 20(1):143–156, 2006.
[133] M. E. P. Maceira, V. S. Duarte, D. D. J. Penna, L. A. M. Moraes, and A. C. G. Melo. Ten years of application of stochastic dual dynamic programming in official and agent studies in Brazil - description of the NEWAVE program. In *16th PSCC, Glasgow, Scotland*, 2008.
[134] M. E. P. Maceira, D. D. J. Penna, A. L. Diniz, R. J. Pinto, A. C. G. Melo, C. V. Vasconcellos, and C. B. Cruz. Twenty years of application of stochastic dual dynamic programming in official and agent studies in Brazil: main features and improvements on the NEWAVE model. In *2018 Power Systems Computation Conference (PSCC)*. IEEE, 2018.
[135] F. D. R. Machado, A. L. Diniz, C. L. T. Borges, and L. C. Brandão. Asynchronous parallel stochastic dual dynamic programming applied to hydrothermal generation planning. *Electric Power Systems Research*, 191:1–14, 2021.
[136] H. Macian-Sorribes, A. Tilmant, and M. Pulibo-Velazquez. Improving operating policies of large-scale surface-groundwater systems through stochastic programming. *Water Resources Research*, 53:1407–1423, 2017.
[137] G. F. Marques and A. Tilmant. The economic value of coordination in large-scale multireservoir systems: The Parana River case. *Water Resources Research*, 49:7546–7557, 2013.
[138] Y. Mbeutcha, M. Gendreau, and G. Emiel. Benefit of PARMA modeling for long-term hydroeletric scheduling using stochastic dual dynamic programming. *Journal of Water Resources Planning and Management*, 147(3):1–12, 2021.
[139] G. P. McCormick. Computability of global solutions to factorable nonconvex programs: part I – convex underestimating problems. *Mathematical Programming*, 10(1):147–175, 1976.
[140] M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.
[141] B. Mo, A. Gjelsvik, and A. Grundt. Integrated risk management of hydro power scheduling and contract management. *IEEE Transactions on Power Systems*, 16(2):216–221, 2001.
[142] B. Mo, A. Gjelsvik, A. Grundt, and K. Karesen. Optimisation of hydropower operation in a liberalised market with focus on price modelling. In *PowerTech Proceedings*, 2001.
[143] D. P. Morton. An enhanced decomposition algorithm for multistage stochastic hydroelectric

scheduling. *Annals of Operations Research*, 64(1):211–235, 1996.
[144] G. Nannicini, E. Traversi, and R. Wolfler Calvo. A benders squared (b^2) framework for infinite-horizon stochastic linear programs. *Mathematical Programming Computation*, 13:645–681, 2021.
[145] Y. Nesterov. *Introductory lectures on convex optimization: a basic course*. Springer, Boston, 2004.
[146] N. Newham. *Power system investment planning using stochastic dual dynamic programming*. PhD thesis, University of Canterbury, 2008.
[147] A. Papavasiliou, Y. Mou, L. Cambier, and D. Scieur. Application of stochastic dual dynamic programming to the real-time dispatch of storage under renewable supply uncertainty. *IEEE Transactions on Sustainable Energy*, 9(2):547–558, 2018.
[148] P. M. Pardalos, S. Rebennack, M. V. F. Pereira, N. A. Iliadis, and V. Pappu, editors. *Handbook of wind power systems*. Springer, 2013.
[149] P. Parpas, B. Ustun, M.Webster, and Q. K. Tran. Importance sampling in stochastic programming: a Markov Chain Monte Carlo approach. *INFORMS Journal on Computing*, 27(2):358–377, 2015.
[150] M. V. F. Pereira, N. M. Campodónico, and R. Kelman. Application of stochastic dual DP and extensions to hydrothermal scheduling. Technical report, PSRI, 1999.
[151] M. V. F. Pereira and L. M. V. G. Pinto. Stochastic optimization of a multireservoir hydroelectric system: a decomposition approach. *Water Resources Research*, 21(6):779–792, 1985.
[152] M. V. F. Pereira and L. M. V. G. Pinto. Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming*, 52(1-3):359–375, 1991.
[153] M.V.F. Pereira. Optimal scheduling of hydrothermal systems - an overview. In *IFAC Symposium on Planning and Operation of Electric Energy Systems., Rio de Janeiro, Brazil, 22-25 July*, pages 1–6, 1985.
[154] S. J. Pereira-Cardenal, B. Mo, A. Gjelsvik, N. D. Riegels, K. Arnbjerg-Nielsen, and P. Bauer-Gottwein. Joint optimization of regional water-power systems. *Advances in Water Resources*, 92:200–207, 2016.
[155] L. Pfeiffer, R. Apparigliato, and S. Auchapt. Two methods of pruning Benders' cuts and their application to the management of a gas portfolio. Preprint, available online at http://www.optimization-online.org/DB_FILE/2012/11/3683.pdf, 2012.
[156] G. C. Pflug. *Probabilistic Constrained Optimization: Methodology and Applications*, chapter Some Remarks on the Value-at-Risk and the Conditional Value-at-Risk., pages 272–281. Kluwer Academic Publishers, 2000.
[157] G. C. Pflug and A. Pichler. Time-consistent decisions and temportal decomposition of coherent risk functionals. *Mathematics of Operations Research*, 41(2):682–699, 2016.
[158] G. C. Pflug and A. Ruszczyński. Measuring risk for income streams. *Computational Optimization and Applications*, 32(1):161–178, 2005.
[159] A. B. Philpott and V. L. de Matos. Dynamic sampling algorithms for multi-stage stochastic programs with risk aversion. *European Journal of Operational Research*, 218(2):470–483, 2012.
[160] A. B. Philpott, V. L. de Matos, and E. Finardi. On solving multistage stochastic programs with coherent risk measures. *Operations Research*, 61(4):957–970, 2013.
[161] A. B. Philpott, V. L. de Matos, and L. Kapelevich. Distributionally robust SDDP. *Computational Management Science*, 15:431–454, 2018.
[162] A. B. Philpott, F. Wahid, and F. Bonnans. MIDAS: A mixed integer dynamic approximation scheme. *Mathematical Programming*, 181:19–50, 2020.
[163] A.B. Philpott and Z. Guan. On the convergence of stochastic dual dynamic programming and related methods. *Operations Research Letters*, 36(4):450–455, 2008.
[164] J. Pina, A. Tilmant, and P. Côté. Optimizing multireservoir system operating policies using exogenous hydrologic variables. *Water Resources Research*, 53:9845–9859, 2017.
[165] R. J. Pinto, C. L. T. Borges, and M. E. P. Maceira. An efficient parallel algorithm for large scale hydrothermal system operation planning. *IEEE Transactions on Power Systems*, 28(4):4888–4896, 2013.
[166] B. T. Polyak. Minimization of unsmooth functionals. *USSR Computational Mathematics and Mathematical Physics*, 9(3):14–29, 1969.
[167] H. Poorsepahy-Samian, V. Espanmanesh, and B. Zahraie. Improved inflow modeling in stochastic dual dynamic programming. *Journal of Water Resources Planning and Management*, 142(12):1–10, 2016.
[168] W. B. Powell. *Approximate dynamic programming: solving the curses of dimensionality*. Wiley-Interscience, 2007.

[169] W. B. Powell. *Reinforcement Learning and Stochastic Optimization – A Unified Framework for Sequential Decisions*. John Wiley & Sons, 2022.
[170] G. Pritchard. Stochastic inflow modeling for hydropower scheduling problems. *European Journal of Operational Research*, 246:496–504, 2015.
[171] PSR. SDDP - Stochastic hydrothermal dispatch with network restrictions. Accessed April 27, 2021, https://www.psr-inc.com/softwares-en/.
[172] Quantego. QUASAR. Accessed April 27, 2021, http://quantego.com/.
[173] A. R. Queiroz and D. P. Morton. Sharing cuts under aggregated forecast when decomposing multi-stage stochastic programs. *Operations Research Letters*, 41(3):311–316, 2013.
[174] F. Quezada, C. Gicquel, and S. Kedad-Sidhoum. Combining polyhedral approaches and stochastic dual dynamic integer programming for solving the uncapacitated lot-sizing problem under uncertainty. *INFORMS Journal on Computing*, 34(2):1024–1041, 2022.
[175] H. Rahimian and S. Mehrotra. Distributionally robust optimization: a review. Preprint, available online at https://arxiv.org/pdf/1908.05659.pdf, 2019.
[176] L. Raso, P.-O. Malaterre, and J.-C. Bader. Effective streamflow process modeling for optimal reservoir operation using stochastic dual dynamic programming. *Journal of Water Resources Planning and Management*, 143(4):1–11, 2017.
[177] S. Rebennack. Generation expansion planning under uncertainty with emissions quotas. *Electric Power Systems Research*, 114:78–85, 2014.
[178] S. Rebennack. Combining sampling-based and scenario-based nested Benders decomposition methods: application to stochastic dual dynamic programming. *Mathematical Programming*, 156(1-2):343–389, 2016.
[179] S. Rebennack, B. Flach, M.V.F. Pereira, and P.M. Pardalos. Stochastic hydro-thermal scheduling under CO_2 emissions constraints. *IEEE Transactions on Power Systems*, 27(1):58–68, 2012.
[180] S. Rebennack, N.A. Iliadis, M.V.F. Pereira, and P.M. Pardalos. Electricity and co_2 emissions system prices modeling and optimization. In *PowerTech, 2009 IEEE Bucharest*, pages 1–6. IEEE, 2009.
[181] S. Rebennack, P.M. Pardalos, M.V.F. Pereira, and N.A. Iliadis, editors. *Handbook of power systems I*. Springer, 2010.
[182] S. Rebennack, P.M. Pardalos, M.V.F. Pereira, and N.A. Iliadis, editors. *Handbook of power systems II*. Springer, 2010.
[183] M. Resener, S. Rebennack, P. M. Pardalos, and S. Haffner, editors. *Handbook of optimization in electric power distribution systems*. Springer, 2020.
[184] R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of risk*, 2(3):21–42, 2000.
[185] R. T. Rockafellar and R. J. B. Wets. *Variational Analysis*. Springer Berlin Heidelberg, 2009.
[186] T. A. Rotting and A. Gjelsvik. Stochastic dual dynamic programming for seasonal scheduling in the Norwegian power system. *IEEE Transactions on Power Systems*, 7(1):273–279, February 1992.
[187] C. Rougé and A. Tilmant. Using stochastic dual dynamic programming in problems with multiple near-optimal solutions. *Water Resources Research*, 52:4151–4163, 2016.
[188] C. Rougé, A. Tilmant, B. Zaitchik, A. Dezfuli, and M. Salman. Identifying key water resource vulnerabilities in data-scarce transboundary river basins. *Water Resources Research*, 54:5264–5281, 2018.
[189] B. Rudloff, A. Street, and D. M. Valladão. Time consistency and risk averse dynamic decision models: definition, interpretation and practical consequences. *European Journal of Operational Research*, 234:743–750, 2014.
[190] A. Ruszczynski. Decomposition methods in stochastic programming. *Mathematical Programming*, 79(1-3):333–353, 1997.
[191] A. Ruszczyński. Risk-averse dynamic programming for Markov decision processes. *Mathematical Programming*, 125:235–261, 2010.
[192] A. Ruszczyński and A. Shapiro. Conditional risk mappings. *Mathematics of Operations Research*, 31:544–561, 2006.
[193] A. Ruszczynski and A. Swietanowski. Accelerating the regularized decomposition method for two stage stochastic linear problems. *European Journal of Operational Research*, 101:328–342, 1997.
[194] A. Shapiro. Inference of statistical bounds for multistage stochastic programming problems. *Mathematical Methods of Operations Research*, 58(1):57–68, 2003.
[195] A. Shapiro. On complexity of multistage stochastic programs. *Operations Research Letters*, 34:1–8, 2005.
[196] A. Shapiro. Stochastic programming approach to optimization under uncertainty. *Mathemat-*

ical Programming, 112(1):183–220, 2008.
[197] A. Shapiro. Analysis of stochastic dual dynamic programming method. *European Journal of Operational Research*, 209:63–72, 2011.
[198] A. Shapiro. Tutorial on risk neutral, distributionally robust and risk averse multistage stochastic programming. *European Journal of Operational Research*, 288(1):1–13, 2021.
[199] A. Shapiro and Y. Cheng. Dual bounds for periodical stochastic programs. *Operations Research*, 71(1):120–128, 2022.
[200] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory, Second Edition*. Society for Industrial and Applied Mathematics, 2014.
[201] A. Shapiro and L. Ding. Periodical multistage stochastic programs. *SIAM Journal on Optimization*, 30(3):2083–2102, 2020.
[202] A. Shapiro, W. Tekaya, J. P. da Costa, and M. P. Soares. Risk neutral and risk averse stochastic dual dynamic programming method. *European Journal of Operational Research*, 224(2):375–391, 2013.
[203] A. Shapiro, W. Tekaya, M. P. Soares, and J. P. da Costa. Worst-case-expectation approach to optimization under uncertainty. *Operations Research*, 61(6):1435–1449, 2013.
[204] P. Shinde, I. Kouveliotis-Lysikatos, and M. Amelin. Multistage stochastic programming for VPP trading in continuous intraday electricity markets. *IEEE Transactions on Sustainable Energy*, 13(2):1–12, 2022.
[205] M. Siddig and Y. Song. Adaptive partition-based SDDP algorithms for multistage stochastic linear programming with fixed recourse. *Computational Optimization and Applications*, 81:201–250, 2022.
[206] M. P. Soares, A. Street, and D. M. Valladão. On the solution variability reduction of stochastic dual dynamic programming applied to energy planning. *European Journal of Operational Research*, 258(2):743–760, 2017.
[207] A. Sorokin, S. Rebennack, P.M. Pardalos, N.A. Iliadis, and M.V.F. Pereira, editors. *Handbook of networks in power systems I*. Springer Science & Business Media, 2012.
[208] A. Sorokin, S. Rebennack, P.M. Pardalos, N.A. Iliadis, and M.V.F. Pereira, editors. *Handbook of networks in power systems II*. Springer Science & Business Media, 2012.
[209] G. Steeger, T. Lohmann, and S. Rebennack. Strategic bidding for a price-maker hydroelectric producer: stochastic dual dynamic programming and Lagrangian relaxation. *IISE Transactions*, 47:1–14, 2018.
[210] G. Steeger and S. Rebennack. Strategic bidding for multiple price-maker hydroelectric producers. *IIE Transactions*, 47(9):1013–1031, 2015.
[211] G. Steeger and S. Rebennack. Dynamic convexification within nested Benders decomposition using Lagrangian relaxation: an application to the strategic bidding problem. *European Journal of Operational Research*, 257(2):669–686, 2017.
[212] G.M. Steeger. *Strategic bidding for price-maker hydroelectric producers*. PhD thesis, Colorado School of Mines, 2014.
[213] A. Street, L. A. Barroso, B. Flach, M. V. Pereira, and S. Granville. Risk constrained portfolio selection of renewable sources in hydrothermal electricity markets. *IEEE Transactions on Power Systems*, 24(3):1136–1144, 2009.
[214] A. Street, A. Brigatto, and D. M. Valladão. Co-optimization of energy and ancillary services for hydrothermal operation planning under a general security criterion. *IEEE Transactions on Power Systems*, 32(6):4914–4923, 2017.
[215] X. A. Sun and A. J. Conejo. *Robust optimization in electric energy systems*. International Series in Operations Research & Management Science. Springer, 2022.
[216] S. Thevenin, Y. Adulyasak, and J.-F. Cordeau. Stochastic dual dynamic programming for multiechelon lot sizing with component substitution. *INFORMS Journal on Computing*, 34(6):3151–3169, 2022.
[217] F.S. Thomé. *Representação de não-convexidade no planejamento da operação hidrotérmica utilizando PDDE*. PhD thesis, COPPE-UFRJ, 2013.
[218] A. Tilmant, L. Beevers, and B. Muyunda. Restoring a flow regime through the coordinated operation of a multireservoir system: The case of the Zambezi River basin. *Water Resources Research*, 46:1–11, 2010.
[219] A. Tilmant and R. Kelman. A stochastic approach to analyze trade-offs and risks associated with large-scale water resources systems. *Water Resources Research*, 43(6), 2007.
[220] A. Tilmant and W. Kinzelbach. The cost of noncooperation in international river basins. *Water Resources Research*, 48:1–12, 2012.
[221] A. Tilmant, J. Lettany, and R. Kelman. Hydrological risk assessment in the Euphrates-Tigris river basin: A stochastic dual dynamic programming approach. *Water International*, 32:294–309, 2009.

[222] A. Tilmant, D. Pinte, and Q. Goor. Assessing marginal water values in multipurpose multireservoir systems via stochastic programming. *Water Resources Research*, 44:1–17, 2008.

[223] D. M. Valladão, T. Silva, and M. Poggi. Time-consistent risk-constrained dynamic portfolio optimization with transactional costs and time-dependent returns. *Annals of Operations Research*, 282:379–405, 2019.

[224] W. van Ackooij, W. de Oliveira, and Y. Song. On level regularization with normal solutions in decomposition methods for multistage stochastic programming problems. *Computational Optimization and Applications*, 74:1–42, 2019.

[225] W. van Ackooij and X. Warin. On conditional cuts for stochastic dual dynamic programming. *EURO Journal on Computational Optimization*, 8:173–199, 2020.

[226] R. M. van Slyke and R. Wets. L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics*, 17(4):638–663, 1969.

[227] F. Wahid. *River optimisation: short-term hydro-bidding under uncertainty*. Optimization and control, Université Paris-Saclay; University of Auckland, New Zealand, 2017.

[228] W.W.-G. Yeh. Reservoir management and operations models: a state-of-the-art review. *Water Resources Research*, 21(12):1797–1818, 1985.

[229] S. Zhang and X. A. Sun. Stochastic dual dynamic programming for multistage stochastic mixed-integer nonlinear optimization. *Mathematical Programming*, 196:935–985, 2022.

[230] Q.P. Zheng, S. Rebennack, P.M. Pardalos, M.V.F. Pereira, and N.A. Iliadis, editors. *Handbook of CO2 in power systems*. Springer Science & Business Media, 2012.

[231] X. Zhu, M.G. Genton, Y. Gu, and L. Xie. Space-time wind speed forecasting for improved power system dispatch. *TEST*, 23(1):1–25, 2014.

[232] J. Zou, S. Ahmed, and X. A. Sun. Stochastic dual dynamic integer programming. *Mathematical Programming*, 175:461–502, 2019.

[233] L. Zéphyr and C. L. Anderson. Stochastic dynamic programming approach to managing power system uncertainty with distributed storage. *Computational Management Science*, 15:87–110, 2018.