Data-Driven Stochastic Dual Dynamic Programming: Performance Guarantees and Regularization Schemes

Hyuk Park, Zhuangzhuang Jia, and Grani A. Hanasusanto
Department of Industrial and Enterprise Systems Engineering
University of Illinois Urbana-Champaign, United States

Abstract

We propose a data-driven extension of the stochastic dual dynamic programming (SDDP) algorithm for multistage stochastic linear programs under a continuous-state, non-stationary Markov data process. Unlike traditional SDDP methods—which often assume a known probability distribution, stagewise independent data process, or uncertainty restricted to the right-hand side of constraints—our approach overcomes these limitations, making it more applicable to various real-world applications. Our scheme avoids the construction of an exponentially growing scenario tree while providing theoretical out-of-sample performance guarantees for the proposed SDDP variant. However, sparse training data may induce an optimistic bias, degrading out-of-sample performance. To address this, we incorporate distributionally robust optimization based on the modified χ^2 distance and show its equivalence to the variance regularization. We validate our approach through real-world applications in finance and energy.

Keywords: stochastic dual dynamic programming; multistage stochastic programming; Markov dependence; distributionally robust optimization

1 Introduction

Multistage stochastic programming is an optimization framework for modeling sequential decision-making under uncertainty, in which uncertain data is revealed over time and decisions are adjusted accordingly. Specifically, the planning horizon consists of T stages, with decisions made at each stage that adapt to realizations of the underlying stochastic data process. This framework intersects with fields such as Markov decision process, stochastic optimal control, and reinforcement learning, as noted in [21], although each community adopts distinct modeling assumptions and solution methods.

Multistage stochastic linear programming (MSLP) is the case where both the objective function and constraints at each stage are linear. MSLP has been used to model various real-world applications, such as hydrothermal scheduling [16, 39, 38, 46], unit commitment [5, 11, 26, 43], portfolio optimization [9, 10, 15,

23, 36], and manufacturing and capacity planning [1, 2, 20, 49]. Despite its various applications, there is a consensus that solving general multistage stochastic programs is challenging [32, 45].

One of the difficulties arises from representing the underlying data process: if it has a continuous state space, numerically representing the data process is impossible. Even with a discrete state space, the number of possible realizations can be overwhelmingly large. Therefore, an approximate version of the problem is typically constructed and solved, with the goal of obtaining a solution that serves as a good suboptimal proxy for the true MSLP problem. In the literature, it is common to assume that the probability distribution governing the data process is known. Therefore, one can easily construct the approximate problem by Monte Carlo sampling.

Stochastic dual dynamic programming (SDDP) is a popular approach for solving such approximate MSLP problems and can be viewed as a multistage extension of the Benders decomposition method [6]. Originally proposed in the seminal work [39] to address a large-scale hydrothermal scheduling problem, SDDP evaluates cost-to-go (or value) functions by the dynamic programming framework. Because these functions are convex on the feasible regions, SDDP iteratively constructs piecewise linear outer approximations, known as cutting planes or *cuts*. These cuts allow the cost-to-go function to be evaluated at carefully chosen feasible solutions, thereby alleviating the intractability resulting from the discretization of a high-dimensional feasible region—commonly referred to as the *curse of dimensionality*. Asymptotic convergence has been established in several studies [12, 22, 40, 44] and computational complexity has recently been investigated in [33, 55]. Extensions of the standard SDDP framework include risk-averse formulations [13, 24, 46] as well as adaptations for nonlinear convex [53, 54] and nonconvex cost-to-go functions [3, 18, 56].

Standard SDDP requires certain restrictions on the data process to be solved efficiently. In the literature, it is often assumed that the data process is stagewise independent, i.e., the uncertainty at the next stage does not depend on the history of the data process. Without this assumption, the scenario tree used to approximate the data process would grow exponentially: specifically, the number of cost-to-go functions that must be evaluated becomes $\sum_{t=1}^{T-1} \prod_{s=1}^{t} N_s$, where N_s represents the number of realizations at stage $s+1$ for $s = 1, \ldots, T-1$. In principle, SDDP can still be used to solve this MSLP problem even though it no longer has significant computational advantages over other solution methods. This limitation is a reason many works on SDDP often rely on the stagewise independence assumption.

However, stagewise independence is a strong assumption that often does not hold in practice. For instance, future stock prices or renewable energy outputs (e.g., rainfall or wind speed) are typically correlated with their historical values—and may also be influenced by some observable side information, such as economic indices for stock prices or atmospheric conditions for wind predictions. To relax this assumption, several extensions have been proposed that assume the data process is Markovian (i.e., exhibits inter-stage dependence). The most popular approaches for incorporating Markov dependence are the time-series SDDP and the Markov chain SDDP methods.

In the time-series approach, the data process is fitted to a first-order time series model, and the scenario

tree is constructed from the residuals of the model (i.e., stagewise independent errors) [29, 46]. However, this method assumes a *linear* time series model and requires the introduction of auxiliary variables, which aggravates the curse of dimensionality. Moreover, from a modeling standpoint, this approach is limited to cases where the uncertain data with Markov dependence appear only on the right-hand side of the constraints, since any other configuration may destroy the convexity of the cost-to-go functions. Alternatively, the Markov chain discretization method partitions the data process at each stage into a pre-defined number of segments and determines a center point for each segment, which is known as the optimal quantization problem [8]. Unlike the time-series approach, this method can accommodate more general Markov data processes, i.e., uncertainty can appear anywhere in the problem. However, the optimal quantization problem is NP-hard and thus typically requires approximation methods [35].

The existing SDDP algorithms are developed under the strong assumption that the underlying probability distribution is known. However, in many real-world applications, only historical trajectories are at our disposal, emphasizing the need for a data-driven approach with performance guarantees. While one can principally extend the time-series and the Markov chain discretization methods to the data-driven setting, neither approach currently offers theoretical out-of-sample performance guarantees for the general MSLP problems. Furthermore, the Markov chain discretization method cannot even guarantee convergence to the true data process, while the time-series approach is too restrictive because it requires that the true data process be linear.

Motivated by these challenges, we pose the following question:

> *"Can we develop a data-driven SDDP variant for general Markov data processes that remains computationally tractable and sample efficient, with theoretical guarantees?"*

In this paper, we answer this question by developing a novel SDDP framework that provides out-of-sample performance guarantees. Notably, our theoretical results show that incorporating Markov dependence does not worsen the sample complexity compared to the stagewise independent setting: the complexity scales only polynomially with the planning horizon T. This result highlights that the proposed approach remains both computationally scalable and sample-efficient—ensuring that even for MSLP problems with large T, one can expect good out-of-sample performance with a sufficiently large number of samples.

In many practical settings, however, the available data may be scarce, and collecting additional samples can be impractical. In such small-data regimes, our data-driven SDDP approach may overfit, leading to significant suboptimality when solving the approximate MSLP. To address this issue, we employ distributionally robust optimization (DRO), an emerging framework that relaxes the stringent assumption of a known distribution by constructing an (ambiguity) set of plausible distributions consistent with the available data. We are not the first to combine SDDP with DRO. Similar to our approach, Philpott et al. [41] use the modified χ^2 ambiguity set to derive a closed-form solution for the inner maximization problem, generating cuts. However, their work assumes stagewise independence with randomness restricted to the right-hand side. In contrast, Silva et al. [48] address a general Markov data process using a hidden Markov model

to capture unobservable states. They then formulate a DRO version of the problem with a total variation ambiguity set. In Table 1, we compare our proposed method with existing DRO-based SDDP variants.

The main contributions of this paper can be summarized as follows:

1. We propose a novel data-driven approximation scheme for MSLP problems under general Markov data processes. When solved using dynamic programming, our approximation scheme achieves polynomial computational complexity with respect to the planning horizon T, avoiding the exponential complexity of the sample average approximation method. Our scheme further provides out-of-sample performance guarantees with a suboptimality bound of $\tilde{O}(T^{\frac{3}{2}}/N^{\frac{2}{p+4}})$, where N denotes the number of sample trajectories and p is the dimension of the data process. To the best of our knowledge, no existing work achieves such a mild dependence on T for MSLP problems under general Markov dependence. To solve the data-driven approximation scheme, we propose an SDDP algorithm that systematically addresses the evaluations over continuous decision spaces.

2. Our theoretical analysis motivates the use of a variance-based regularization scheme to enhance out-of-sample performance. Although this scheme yields a nonconvex problem, we establish an equivalence between the variance-regularized formulation and the DRO formulation using the modified χ^2 ambiguity set, which is amenable to our SDDP solution scheme. To our knowledge, these formulations have not been proposed in the literature on SDDP.

3. We validate our approach through numerical experiments on real-life portfolio optimization and wind energy commitment problems, demonstrating that our data-driven schemes can significantly outperform the stagewise independent SDDP and other existing DRO-based SDDP variants in out-of-sample performance tests.

Notation

Scalars are denoted by non-bold letters n or N. For any positive integer N, $[N] = \{1, \ldots, N\}$. Bold lowercase letter $\boldsymbol{w} \in \mathbb{R}^N$ denotes a vector while bold upper-case letter $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ represents a matrix. The vector of all ones is denoted as $\mathbf{e}$. A realization of the random data process up to stage t is denoted as

Table 1. Comparison of distributionally robust SDDP algorithms

Algorithms	**Huang et al., 2017 [28]**	**Philpott et al., 2018 [41]**	**Duque and Morton, 2020 [19]**	**Silva et al., 2021 [48]**	**This Paper**
Ambiguity set	∞-norm	modified χ^2	Wasserstein	total variation	modified χ^2
Data Process	stagewise independent	stagewise independent	stagewise independent	Markov dependent	Markov dependent
Randomness	right-hand side	right-hand side	right-hand side	anywhere	anywhere
Out-of-sample guarantee	-	-	-	-	✓

$\boldsymbol{\xi}_{[t]} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_t)$. The tilde symbol denotes randomness (e.g., $\tilde{\boldsymbol{\xi}}_t$) to differentiate from its realization (e.g., $\boldsymbol{\xi}_t$). In asymptotic analysis, we use the standard little-o and big-O notations: $o(\cdot)$ and $O(\cdot)$. The notation $\tilde{O}(\cdot)$ is used to suppress multiplicative terms with logarithmic dependence. The set $\Delta^N = \{\boldsymbol{w} \in \mathbb{R}^N_+ : \mathbf{e}^\top \boldsymbol{w} = 1\}$ is the probability simplex, while $\mathcal{C}^N = \left\{(\boldsymbol{u}, v) \in \mathbb{R}^N \times \mathbb{R} : \|\boldsymbol{u}\|_2 \leq v\right\}$ is the second-order cone.

2 Problem Statement

Consider the following MSLP problem under *Markov dependence* with the planning horizon T:

$$\min_{\boldsymbol{x}_1 \in \mathcal{X}_1(\boldsymbol{x}_0, \boldsymbol{\xi}_1)} \boldsymbol{c}_1^\top \boldsymbol{x}_1 + \mathbb{E}\left[\min_{\boldsymbol{x}_2 \in \mathcal{X}_2(\boldsymbol{x}_1, \tilde{\boldsymbol{\xi}}_2)} \tilde{\boldsymbol{c}}_2^\top \boldsymbol{x}_2 + \mathbb{E}\left[\cdots + \mathbb{E}\left[\min_{\boldsymbol{x}_T \in \mathcal{X}_T(\boldsymbol{x}_{T-1}, \tilde{\boldsymbol{\xi}}_T)} \tilde{\boldsymbol{c}}_T^\top \boldsymbol{x}_T \;\middle|\; \tilde{\boldsymbol{\xi}}_{T-1}\right] \cdots \;\middle|\; \tilde{\boldsymbol{\xi}}_2\right] \;\middle|\; \boldsymbol{\xi}_1\right]. \quad (1)$$

Here, $\boldsymbol{x}_0 \in \mathbb{R}^{d_0}$ is given as a deterministic vector, and the parameters are defined as follows: for the first stage ($t = 1$), $\boldsymbol{\xi}_1 = (\boldsymbol{c}_1, \boldsymbol{b}_1, \boldsymbol{A}_1, \boldsymbol{B}_1) \in \mathbb{R}^{p_1}$ are a deterministic initial state of the underlying data process, while for stages $t \geq 2$, the random vector $\tilde{\boldsymbol{\xi}}_t = (\tilde{\boldsymbol{c}}_t, \tilde{\boldsymbol{b}}_t, \tilde{\boldsymbol{A}}_t, \tilde{\boldsymbol{B}}_t) \in \mathbb{R}^{p_t}$, supported on a set $\Xi_t \subseteq \mathbb{R}^{p_t}$, follows an *unknown* Markov process (i.e., $\tilde{\boldsymbol{\xi}}_t$ depends only on $\tilde{\boldsymbol{\xi}}_{t-1}$). The feasible region at stage t is a polytope defined as $\mathcal{X}_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t) = \{\boldsymbol{x}_t \in \mathbb{R}^{d_t}_+ : \boldsymbol{A}_t \boldsymbol{x}_t + \boldsymbol{B}_t \boldsymbol{x}_{t-1} = \boldsymbol{b}_t\}$, where $\boldsymbol{x}_{t-1}$ and $\boldsymbol{\xi}_t$ denote the previous-stage decision and the realized values of the uncertain data $\tilde{\boldsymbol{\xi}}_t$, respectively. Note that the first-stage feasible region $\mathcal{X}_1(\boldsymbol{x}_0, \boldsymbol{\xi}_1)$ is deterministic, as both $\boldsymbol{x}_0$ and $\boldsymbol{\xi}_1$ are given as deterministic vectors.

At stage 1, the decision $\boldsymbol{x}_1$ is selected from $\mathcal{X}_1(\boldsymbol{x}_0, \boldsymbol{\xi}_1)$, incurring cost $\boldsymbol{c}_1^\top \boldsymbol{x}_1$. At each subsequent stage $t \geq 2$, the realization $\boldsymbol{\xi}_t = (\boldsymbol{c}_t, \boldsymbol{b}_t, \boldsymbol{A}_t, \boldsymbol{B}_t)$ of $\tilde{\boldsymbol{\xi}}_t$ is observed, and the decision $\boldsymbol{x}_t$ is chosen from $\mathcal{X}_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$ with cost $\boldsymbol{c}_t^\top \boldsymbol{x}_t$. The goal is to find a policy $\{\boldsymbol{x}_t(\boldsymbol{\xi}_{[t]})\}_{t=1}^T$ that minimizes the expected total cost, where each decision $\boldsymbol{x}_t : \mathbb{R}^{p_1} \times \cdots \times \mathbb{R}^{p_t} \to \mathbb{R}^{d_t}$ is a function of the history $\boldsymbol{\xi}_{[t]} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_t)$ of the data process up to stage t. Also, note that the conditional expectations $\mathbb{E}[\,\cdot \mid \boldsymbol{\xi}_t]$ in (1) depend only on the most recent realization $\boldsymbol{\xi}_t$ rather than the full history $\boldsymbol{\xi}_{[t]}$ due to Markov dependence.

In this paper, we refer to (1) as the *true* (MSLP) problem, as it is defined under the true data process. Using the Bellman optimality principle, the true problem can be rewritten as a sequence of cost-to-go functions $Q_t : \mathbb{R}^{d_{t-1}} \times \mathbb{R}^{p_t} \to \mathbb{R}$: for each stage $t \in [T]$, $\boldsymbol{x}_{t-1} \in \mathcal{X}_{t-1}(\cdot)$, and $\boldsymbol{\xi}_t \in \Xi_t$,

$$\begin{aligned} Q_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t) = \quad & \min \quad \boldsymbol{c}_t^\top \boldsymbol{x}_t + \mathcal{Q}_{t+1}(\boldsymbol{x}_t, \boldsymbol{\xi}_t) \\ & \text{s.t.} \quad \boldsymbol{x}_t \in \mathbb{R}^{d_t}_+, \\ & \qquad \boldsymbol{A}_t \boldsymbol{x}_t + \boldsymbol{B}_t \boldsymbol{x}_{t-1} = \boldsymbol{b}_t. \end{aligned} \quad (2)$$

Here,

$$\mathcal{Q}_{t+1}(\boldsymbol{x}_t, \boldsymbol{\xi}_t) = \mathbb{E}\left[Q_{t+1}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1}) \;\middle|\; \boldsymbol{\xi}_t\right]$$

represents the conditional expectation of the cost-to-go function at stage $t+1$, given the most recent history $\boldsymbol{\xi}_t$. We assume $Q_{T+1}(\cdot) = 0$, which implies that no additional costs beyond the terminal stage T.

2.1 Assumptions

In the literature, it is commonly assumed that (i) the data process is stagewise independent and (ii) the probability distribution governing the data process is known. In contrast, we relax these assumptions to develop a data-driven approach for (2). We formally state the following assumptions.

(**A1**) **Markov Process**. The underlying data process $\tilde{\boldsymbol{\xi}}_{[T]} = (\boldsymbol{\xi}_1, \tilde{\boldsymbol{\xi}}_2, \ldots, \tilde{\boldsymbol{\xi}}_T)$ is a discrete-time, continuous-state, non-stationary Markov process, where each vector $\boldsymbol{\xi}_t$ summarizes the data appearing in the objective function and constraints. The initial state $\boldsymbol{\xi}_1 = (\boldsymbol{c}_1, \boldsymbol{b}_1, \boldsymbol{A}_1, \boldsymbol{B}_1) \in \mathbb{R}^{p_1}$ is deterministic, while for each $t \in [T] \setminus \{1\}$, the uncertain data is denoted by $\tilde{\boldsymbol{\xi}}_t = (\tilde{\boldsymbol{c}}_t, \tilde{\boldsymbol{b}}_t, \tilde{\boldsymbol{A}}_t, \tilde{\boldsymbol{B}}_t) \in \mathbb{R}^{p_t}$.

(**A2**) **Unknown Distribution**. For each $t \in [T] \setminus \{1\}$, the conditional density function $f_{t+1}(\boldsymbol{\xi}_{t+1} \mid \boldsymbol{\xi}_t)$ is unknown. Instead, we have access to N i.i.d. sample trajectories $\boldsymbol{\xi}^i_{[T]} = (\boldsymbol{\xi}^i_1, \ldots, \boldsymbol{\xi}^i_T)$ for all $i \in [N]$, where each $\boldsymbol{\xi}^i_t = (\boldsymbol{c}^i_t, \boldsymbol{b}^i_t, \boldsymbol{A}^i_t, \boldsymbol{B}^i_t)$ represents the observed data at stage t.

2.2 Approximate MSLP Problem

If the underlying data process has a continuous state space, solving the true MSLP problem becomes intractable. This is because even evaluating the expected cost-to-go functions $\mathcal{Q}_{t+1}(\boldsymbol{x}_t, \boldsymbol{\xi}_t)$ requires computing multivariate (conditional) expectations, which is challenging [27].

In our data-driven setting, the state space of the true data process is discretized by N sample trajectories as depicted in Figure 1, which we refer to as a scenario tree. Each node in the tree directly takes values from the sample trajectories $\boldsymbol{\xi}^i_{[T]}$ for all $i \in [N]$, and the transition probabilities between any successive nodes, i.e., conditional probabilities, are estimated via Nadaraya-Watson kernel regression [37, 52]. Specifically, transition probabilities between $\boldsymbol{\xi}^i_t$ and $\boldsymbol{\xi}^j_{t+1}$ are estimated as

$$\widehat{w}_{t+1}(\boldsymbol{\xi}^i_t, \boldsymbol{\xi}^j_t) = \mathcal{K}\Big(\frac{\boldsymbol{\xi}^i_t - \boldsymbol{\xi}^j_t}{h}\Big) \Big/ \sum_{k \in [N]} \mathcal{K}\Big(\frac{\boldsymbol{\xi}^i_t - \boldsymbol{\xi}^k_t}{h}\Big) \quad \forall i \in [N]\ \forall j \in [N], \tag{3}$$

where $\mathcal{K} : \mathbb{R}^{p_t} \times \mathbb{R}_{++} \to \mathbb{R}_{++}$ is a kernel function of choice and $h > 0$ is a parameter known as bandwidth, which controls the smoothness of the estimator. In our work, we adopt the exponential kernel $\mathcal{K}((\boldsymbol{\xi}^i_t - \boldsymbol{\xi}^j_t)/h) = \exp(-\|\boldsymbol{\xi}^i_t - \boldsymbol{\xi}^j_t\|_2/h)$. This leads to the approximate MSLP problem expressed in the following cost-to-go functions: for the first stage,

$$\begin{aligned} V_1(\boldsymbol{x}_0, \boldsymbol{\xi}_1) = \quad & \min \quad \boldsymbol{c}_1^\top \boldsymbol{x}_1 + \mathcal{V}_2(\boldsymbol{x}_1, \boldsymbol{\xi}_1) \\ & \text{s.t.} \quad \boldsymbol{x}_1 \in \mathbb{R}^{d_1}_+, \\ & \qquad \boldsymbol{A}_1 \boldsymbol{x}_1 + \boldsymbol{B}_1 \boldsymbol{x}_0 = \boldsymbol{b}_1, \end{aligned} \tag{4}$$

and, for each subsequent stage $t \in [T] \setminus \{1\}$, $\boldsymbol{x}_{t-1} \in \mathcal{X}_{t-1}(\cdot)$ and $i \in [N]$,

$$\begin{aligned} V_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}^i_t) = \quad & \min \quad {\boldsymbol{c}^i_t}^\top \boldsymbol{x}_t + \mathcal{V}_{t+1}(\boldsymbol{x}_t, \boldsymbol{\xi}^i_t) \\ & \text{s.t.} \quad \boldsymbol{x}_t \in \mathbb{R}^{d_t}_+, \\ & \qquad \boldsymbol{A}^i_t \boldsymbol{x}_t + \boldsymbol{B}^i_t \boldsymbol{x}_{t-1} = \boldsymbol{b}^i_t, \end{aligned} \tag{5}$$

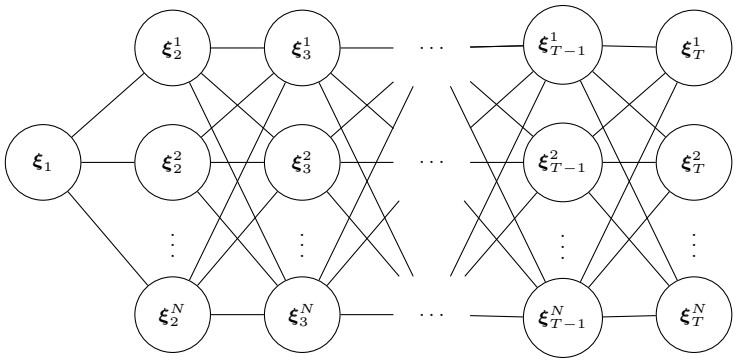

Figure 1. Scenario tree using N sample trajectories.

where $V_{T+1}(\cdot) = 0$ similar to $Q_{T+1}(\cdot) = 0$ in the true problem. Here,

$$\mathcal{V}_{t+1}(\boldsymbol{x}_t, \boldsymbol{\xi}_t^i) = \widehat{\mathbb{E}}\left[V_{t+1}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1}) \;\middle|\; \boldsymbol{\xi}_t^i\right] = \sum_{j\in[N]} \widehat{w}_{t+1}(\boldsymbol{\xi}_t^i, \boldsymbol{\xi}_t^j) \cdot V_{t+1}(\boldsymbol{x}_t, \boldsymbol{\xi}_{t+1}^j) \tag{6}$$

represents the *approximate* expected cost-to-go function at stage $t+1$, given the most recent sample $\boldsymbol{\xi}_t^i$ at stage t. Note that we define $\widehat{\mathbb{E}}[V_{t+1}(\cdot)|\boldsymbol{\xi}_t^i]$ as the approximate conditional expectation computed via the kernel estimation (3).

We aim to solve the approximate problem as a proxy for the true problem (2). Although the scenario tree discretizes the state space so that the approximate cost-to-go functions (5) are evaluated on a finite number of samples, it still needs to be evaluated for every $\boldsymbol{x}_{t-1}$ in the polytope $\mathcal{X}_{t-1}(\cdot)$, which contains infinitely many points. This challenge necessitates an efficient strategy for selecting the evaluation points $\boldsymbol{x}_{t-1}$, which is a central idea of SDDP, as we discuss in more detail later. We first analyze the quality of the solution to this data-driven approximation scheme.

Remark 1. *Standard SDDP typically assumes stagewise independent data process, which is a special case of the above procedure where we simply replace the kernel estimate to equal probability, i.e., $w_{t+1}(\boldsymbol{\xi}_t^i, \boldsymbol{\xi}_t^j) = 1/N$ for all $i \in [N]$ and $j \in [N]$. Another method for solving MSLP is the nested Benders decomposition (NBD) by Birge [7]. Unlike SDDP, NBD can incorporate a general data process beyond Markov dependence. However, the number of cost-to-go functions grows exponentially with a planning horizon T. Hence, as noted in [51], NBD is only computationally tractable for moderate-sized MSLP, say, the number of realizations at each stage being several hundreds and T being 4 or 5 at maximum. As noted in [21], SDDP is a sampling-based variant of NBD, reducing the computational burden experienced with NBD at the expense of restrictions on the underlying data process.*

2.3 Out-of-sample Performance

Let $\boldsymbol{x}_1^\star$ and $z^\star$ denote the optimal first-stage solution and objective value of the *true* problem (1). Similarly, let $\widehat{\boldsymbol{x}}_1^N$ be the optimal first-stage solution of the *approximate* problem (4) based on N sample trajectories.

Since $\widehat{\boldsymbol{x}}_1^N$ is feasible in $\mathcal{X}_1(\boldsymbol{x}_0, \boldsymbol{\xi}_1)$, it serves as a valid suboptimal solution for the true problem. In particular, we have

$$\underbrace{z^\star = \boldsymbol{c}_1^\top \boldsymbol{x}_1^\star + \mathcal{Q}_2(\boldsymbol{x}_1^\star, \boldsymbol{\xi}_1)}_{\text{True optimal objective value}} \leq \underbrace{\boldsymbol{c}_1^\top \widehat{\boldsymbol{x}}_1^N + \mathcal{Q}_2(\widehat{\boldsymbol{x}}_1^N, \boldsymbol{\xi}_1)}_{\text{Out-of-sample performance}} \iff 0 \leq \underbrace{\boldsymbol{c}_1^\top \widehat{\boldsymbol{x}}_1^N + \mathcal{Q}_2(\widehat{\boldsymbol{x}}_1^N, \boldsymbol{\xi}_1) - z^\star}_{\text{Suboptimality}},$$

which indicates that solving the approximate problem introduces suboptimality. In addition, $\widehat{\boldsymbol{x}}_1^N$ is a *random* vector, as it depends on the realizations of the N sample trajectories, and consequently, the suboptimality is also a random variable. Therefore, it is essential to derive a high-probability bound on the suboptimality.

In this section, we present the out-of-sample performance guarantee and further provide the suboptimality bound. We begin by stating the following assumptions.

(**A3**) **Relatively Complete Recourse**. The feasible region is nonempty and compact: there exists a constant $D > 0$ such that $\sup_{\boldsymbol{x}_t, \boldsymbol{x}_t' \in \mathcal{X}_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)} \|\boldsymbol{x}_t - \boldsymbol{x}_t'\| \leq D$ for any $t \in [T]$, $\boldsymbol{x}_{t-1} \in \mathcal{X}_{t-1}(\cdot)$, and $\boldsymbol{\xi}_t \in \Xi_t$. For simplicity of exposition, we assume that the dimension of the feasible region $\mathcal{X}_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$ is the same at all stages, i.e., $d = d_1 = \cdots = d_T$.

(**A4**) **Cost-to-go Function between 0 and 1**. For $t \in [T]$, the value of $Q_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$ falls in the interval $[0, 1]$ for all $\boldsymbol{x}_{t-1} \in \mathcal{X}_{t-1}(\cdot)$ and $\boldsymbol{\xi}_t \in \Xi_t$.

(**A5**) **Compact Uncertainty Set**. For each stage $t \in [T] \setminus \{1\}$, the random data $\tilde{\boldsymbol{\xi}}_t \in \mathbb{R}^{p_t}$ is supported on a *compact* set Ξ_t. For simplicity of exposition, the dimension of random data at all stages is equal to p, i.e., $p = p_2 = \cdots = p_T$.

(**A6**) **Differentiability**. For each $t \in [T] \setminus \{1\}$, the density function $f_t(\boldsymbol{\xi}_t | \boldsymbol{\xi}_{t-1})$ given any $\boldsymbol{\xi}_{t-1} \in \Xi_{t-1}$ is non-zero and twice differentiable with continuous and bounded partial derivatives.

(**A7**) **Bandwidth**. The bandwidth parameter h for the kernel function $\mathcal{K}(\cdot)$ is scaled such that $\lim_{N\to\infty} h_N = 0$ and $\lim_{N\to\infty} N h_N^p = \infty$.

These are standard in the literature. Note that the relatively complete recourse condition in (**A3**) ensures that the cost-to-go function always attains a finite value. Thus, (**A4**) is not restrictive and is assumed solely for notational simplicity in our analysis, since one can always scale and translate the objective function to restrict its finite values between 0 and 1 without altering the optimal solution. As shown in [25], (**A7**) ensures that (3) asymptotically converges to the true conditional distribution as N tends to infinity.

We begin by establishing preliminary results about Lipschitz continuity.

Lemma 1 (Lipschitz Cost-to-go Function)**.** *For each stage $t \in [T]$, the true cost-to-go function $Q_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$ and the approximate cost-to-go function $V_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$ are L_t-Lipschitz continuous in $\boldsymbol{x}_{t-1} \in \mathcal{X}_{t-1}(\cdot)$. That is, for each $t \in [T]$, there exists a constant $L_t > 0$ such that*

$$\begin{aligned}
&\left|Q_t(\boldsymbol{x}, \boldsymbol{\xi}_t) - Q_t(\boldsymbol{x}', \boldsymbol{\xi}_t)\right| \leq L_t \|\boldsymbol{x} - \boldsymbol{x}'\| \quad \forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}_{t-1}\left(\boldsymbol{x}_{t-2}, \boldsymbol{\xi}_{t-1}\right)\ \forall \boldsymbol{\xi}_t \in \Xi_t, \\
&\left|V_t(\boldsymbol{x}, \boldsymbol{\xi}_t) - V_t(\boldsymbol{x}', \boldsymbol{\xi}_t)\right| \leq L_t \|\boldsymbol{x} - \boldsymbol{x}'\| \quad \forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}_{t-1}\left(\boldsymbol{x}_{t-2}, \boldsymbol{\xi}_{t-1}\right)\ \forall \boldsymbol{\xi}_t \in \Xi_t.
\end{aligned}$$

The proof of Lemma 1 can be found in Appendix A.

Based on the result, we obtain the following corollary on the Lipschitz continuity of the expected cost-to-go functions.

Corollary 1. *At any stage $t \in [T] \setminus \{1\}$, for any $\boldsymbol{\xi}_{t-1} \in \Xi_{t-1}$, we have*

$$\begin{aligned}
\left|\mathbb{E}\left[Q_t(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}_t) \,\middle|\, \boldsymbol{\xi}_{t-1}\right] - \mathbb{E}\left[Q_t(\boldsymbol{x}', \tilde{\boldsymbol{\xi}}_t) \,\middle|\, \boldsymbol{\xi}_{t-1}\right]\right| &\leq L_t\|\boldsymbol{x} - \boldsymbol{x}'\| \qquad \forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}_{t-1}(\boldsymbol{x}_{t-2}, \boldsymbol{\xi}_{t-1}),\\
\left|\widehat{\mathbb{E}}\left[Q_t(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}_t) \,\middle|\, \boldsymbol{\xi}_{t-1}\right] - \widehat{\mathbb{E}}\left[Q_t(\boldsymbol{x}', \tilde{\boldsymbol{\xi}}_t) \,\middle|\, \boldsymbol{\xi}_{t-1}\right]\right| &\leq L_t\|\boldsymbol{x} - \boldsymbol{x}'\| \qquad \forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}_{t-1}(\boldsymbol{x}_{t-2}, \boldsymbol{\xi}_{t-1}),\\
\left|\mathbb{E}\left[V_t(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}_t) \,\middle|\, \boldsymbol{\xi}_{t-1}\right] - \mathbb{E}\left[V_t(\boldsymbol{x}', \tilde{\boldsymbol{\xi}}_t) \,\middle|\, \boldsymbol{\xi}_{t-1}\right]\right| &\leq L_t\|\boldsymbol{x} - \boldsymbol{x}'\| \qquad \forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}_{t-1}(\boldsymbol{x}_{t-2}, \boldsymbol{\xi}_{t-1}),\\
\left|\widehat{\mathbb{E}}\left[V_t(\boldsymbol{x}, \tilde{\boldsymbol{\xi}}_t) \,\middle|\, \boldsymbol{\xi}_{t-1}\right] - \widehat{\mathbb{E}}\left[V_t(\boldsymbol{x}', \tilde{\boldsymbol{\xi}}_t) \,\middle|\, \boldsymbol{\xi}_{t-1}\right]\right| &\leq L_t\|\boldsymbol{x} - \boldsymbol{x}'\| \qquad \forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}_{t-1}(\boldsymbol{x}_{t-2}, \boldsymbol{\xi}_{t-1}).
\end{aligned}$$

Corollary 1 can be verified by directly applying Lemma 1.

Next, we present the following result on the conditional variance of the cost-to-go functions.

Lemma 2 (Bounded Conditional Variance)**.** *For each $t \in [T-1]$, there exists a constant $\sigma_{t+1}^2 > 0$ such that*

$$\mathbb{V}\left[Q_{t+1}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1}) \,\middle|\, \boldsymbol{\xi}_t\right] \leq \sigma_{t+1}^2 \quad \forall \boldsymbol{x}_t \in \mathcal{X}_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)\ \forall \boldsymbol{\xi}_t \in \Xi_t.$$

Proof. The compactness of Ξ_t in (**A5**) and the square-integrability in (**A6**) guarantee that $\mathbb{V}[Q_{t+1}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1})|\boldsymbol{\xi}_t]$ is finite for every $\boldsymbol{x}_t \in \mathcal{X}_t(\cdot)$ and $\boldsymbol{\xi}_t \in \Xi_t$. □

With the preparatory results, we introduce the generalization error bound in [50].

Proposition 1. *At each $t \in [T-1]$, for any fixed $\boldsymbol{x}_t \in \mathcal{X}_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$, $\boldsymbol{\xi}_t \in \Xi_t$, and $\delta \in [0,1]$, we have*

$$\left|\mathbb{E}\left[Q_{t+1}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1}) \,\middle|\, \boldsymbol{\xi}_t\right] - \widehat{\mathbb{E}}\left[Q_{t+1}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1}) \,\middle|\, \boldsymbol{\xi}_t\right]\right| \leq \sqrt{\frac{\mathbb{V}\left[Q_{t+1}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1}) \,\middle|\, \boldsymbol{\xi}_t\right]}{O(N^{\frac{4}{p+4}})(1+o(1))g_{t+1}(\boldsymbol{\xi}_t)} \log\left(\frac{1}{\delta}\right)} \tag{7}$$

with probability at least $1-\delta$. Here,

$$g_{t+1}(\boldsymbol{\xi}_t) = \frac{f_{t+1}(\boldsymbol{\xi}_{t+1} \mid \boldsymbol{\xi}_t)}{2\int_{\mathbb{R}^p} \mathcal{K}^2(\boldsymbol{\omega})\mathrm{d}\boldsymbol{\omega}}$$

is the scaled conditional density function given $\boldsymbol{\xi}_t$.

The proof of Proposition 1 can be found in Corollary 1 in [50].

We now extend this result for a uniform generalization bound over $\mathcal{X}_t(\cdot)$.

Corollary 2. *At each $t \in [T-1]$, for any fixed $\boldsymbol{\xi}_t \in \Xi_t$, $\delta \in [0,1]$, and $\eta > 0$, we have*

$$\left|\mathbb{E}\left[Q_{t+1}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1}) \,\middle|\, \boldsymbol{\xi}_t\right] - \widehat{\mathbb{E}}\left[Q_{t+1}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1}) \,\middle|\, \boldsymbol{\xi}_t\right]\right| \leq \sqrt{\sigma_{t+1}^2 \frac{\log\left(\frac{O(1)(D/\eta)^d}{\delta}\right)}{O(N^{\frac{4}{p+4}})(1+o(1))g_{t+1}(\boldsymbol{\xi}_t)}} + 2L_{t+1}\eta \tag{8}$$

*for all $\boldsymbol{x}_t \in \mathcal{X}_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$ with probability at least $1-\delta$. Here, $g_{t+1}(\boldsymbol{\xi}_t)$ is defined in Proposition 1, L_{t+1} is a Lipschitz constant in Lemma 1, σ_{t+1}^2 is a constant in Lemma 2, and D is a constant by (**A3**).*

We defer the proof of Corollary 2 to Appendix B.

In contrast to our work, [50] considers a static setting where the data process consists of a single-stage random vector $\tilde{\boldsymbol{\xi}}$ correlated with some observable side information $\tilde{\boldsymbol{\gamma}}$. In that context, given a realization $\boldsymbol{\gamma}$, they seek a decision $\boldsymbol{x}$ that minimizes the expected cost $\mathbb{E}[\ell(\boldsymbol{x}, \tilde{\boldsymbol{\xi}})|\boldsymbol{\gamma}]$, where $\ell(\cdot)$ is a *known* cost function. Extending their result to the multistage setting, however, introduces additional challenges. Unlike the static case, the cost-to-go functions $Q_{t+1}(\cdot)$ are not directly available. Indeed, $\widehat{\mathbb{E}}[Q_{t+1}(\cdot)|\boldsymbol{\xi}_t]$ in (8) represents an *approximate* conditional expectation of the *true* cost-to-go function, which differs from $\widehat{\mathbb{E}}[V_{t+1}(\cdot)|\boldsymbol{\xi}_t^i]$ in (6). This illustrates the difficulty of solving MSLP problems with high accuracy: since each approximate cost-to-go function $V_{t+1}(\cdot)$ depends on subsequent ones, errors propagate over time, potentially leading to poor out-of-sample performance, especially when the planning horizon T is large. The main result of this section is presented in the following theorem.

Theorem 1 (Suboptimality Bound)**.** *For any fixed $\delta_t \in (0, 1]\ \forall t \in [T] \setminus \{1\}$, and $\eta > 0$, we have*

$$\boldsymbol{c}_1^\top \widehat{\boldsymbol{x}}_1^N + \mathcal{Q}_2(\widehat{\boldsymbol{x}}_1^N, \boldsymbol{\xi}_1) - z^\star \leq 2 \sum_{t=2}^{T} \left(\sqrt{\sigma_t^2 \frac{\log\left(\frac{O(1)N^{t-2}(D/\eta)^{d(t-1)}}{\delta_t}\right)}{O(N^{\frac{4}{p+4}})(1+o(1))g_t}} + 2L_t\eta \right) \tag{9}$$

with probability at least $1 - \sum_{t=2}^{T} \delta_t$. Here, $g_t = \min_{i\in[N]} g_t(\boldsymbol{\xi}_{t-1}^i)$, and all other parameters are defined as in Proposition 1 and Corollary 2.

The proof of Theorem 1 is deferred to Appendix C.

Theorem 1 shows that the suboptimality is at most $\tilde{O}(T^{\frac{3}{2}}/N^{\frac{2}{p+4}})$. This result is significant both theoretically and practically. From the theoretical point of view, it provides the first statistical analysis of MSLP problems under general Markov dependence in a data-driven setting. A key insight is that the suboptimality bound scales only *polynomially* with the planning horizon T. From the practical perspective, this result implies that, to achieve a given suboptimality, the number of sample trajectories N needs to grow only polynomially with T, suggesting that our method is sample-efficient even for MSLP problems with long planning horizons.

In contrast, if the true distribution is known, in principle, the classical sample average approximation (SAA) method could be applied. Under Markov dependence, SAA employs conditional Monte Carlo sampling: one draws N_1 samples $\{\boldsymbol{\xi}_2^i\}_{i=1}^{N_1}$ conditional on $\boldsymbol{\xi}_1$, then N_2 samples $\{\boldsymbol{\xi}_3^{i,j}\}_{j=1}^{N_2}$ conditional on each $\boldsymbol{\xi}_2^i$, and so on. A previous work [45] analyzes SAA for MSLP problems and derives a suboptimality bound of $\tilde{O}(T^{\frac{1}{2}}/N^{\frac{1}{2T}})$, which exhibits exponential dependence on T. Their result indicates that SAA becomes impractical for MSLP problems with a large T. In contrast, our approach suggests that incorporating Markov dependence does not increase the problem's complexity, offering an efficient data-driven alternative to SAA.

While our suboptimality bound has a mild dependence on T, it exhibits exponential dependence on p, the dimension of the random vector $\tilde{\boldsymbol{\xi}}_t$. Unfortunately, this is a common challenge in high-dimensional regression problems [34]. To mitigate this, one can employ dimensionality reduction techniques. Such methods can

improve the performance of our scheme when the effective dimension of the data process, say p', is smaller than p. In the numerical experiment in Section 5.2, we illustrate how dimensionality reduction can be employed in our framework.

3 Data-Driven SDDP

Our solution method for the approximate MSLP problem (4) is based on stochastic dual dynamic programming (SDDP), which is introduced by Pereira and Pinto [39]. SDDP exploits the convexity properties of the cost-to-go functions—as we rigorously demonstrate in a later section, for all $t \in [T]$ and $i \in [N]$, the cost-to-go function $V_{t+1}(\boldsymbol{x}_t, \boldsymbol{\xi}_{t+1}^i)$ is *piecewise linear and convex* in $\boldsymbol{x}_t \in \mathcal{X}_t(\cdot)$, with finitely many pieces. This property allows the cost-to-go functions to be equivalently expressed as pointwise maximum of affine functions, known as *cuts*:

$$V_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t^i) = \max\left\{ {\alpha_{t,l}^i}^\top \boldsymbol{x}_{t-1} + \beta_{t,l}^i \ : \ l \in [\mathcal{S}_t^i] \right\} \quad \forall t \in [T] \setminus \{1\} \ \ \forall i \in [N].$$

Here, $\mathcal{S}_t^i$ denotes the total number of cuts associated with the i-th sample $\boldsymbol{\xi}_t^i$ at stage t, and for each $l \in \mathcal{S}_t^i$, $\alpha_{t,l}^i \in \mathbb{R}^{d_{t-1}}$ and $\beta_{t,l}^i \in \mathbb{R}$ represent the gradient and intercept of the l-th cut, respectively. Such cut information can be obtained by solving the dual of the minimization problem associated with each cost-to-go function.

SDDP is an iterative algorithm that approximates $V_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t^i)$ by constructing lower bounds denoted as $\underline{V}_t^k(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t^i)$ where k represents the k-th iteration. The lower bound can be initialized as a poor approximation, e.g., $\underline{V}_t^0(\cdot) = -\infty$. Then, for each iteration k, we compute and add a cut in reverse time order, i.e., $t = T, T-1, \ldots, 2$:

$$\underline{V}_t^k(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t^i) \leftarrow \max\left\{ \underline{V}_t^{k-1}(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t^i), \ {\boldsymbol{\alpha}_{t,k}^i}^\top \boldsymbol{x}_{t-1} + \beta_{t,k}^i \right\} \quad \forall i \in [N]. \tag{10}$$

Here, $\boldsymbol{\alpha}_{t,k}^i$ and $\beta_{t,k}^i$ is cut information evaluated at a previous-stage solution $\boldsymbol{x}_{t-1}^k$ and the i-th sample $\boldsymbol{\xi}_t^i$ and can be computed as follows:

$$\boldsymbol{\alpha}_{t,k}^i = -{\boldsymbol{B}_t^i}^\top \boldsymbol{\pi}_{t,i,k}^\star \quad \text{and} \quad \beta_{t,k}^i = {\boldsymbol{\alpha}_{t,k}^i}^\top \boldsymbol{x}_{t-1}^k - \underline{V}_t^k(\boldsymbol{x}_{t-1}^k, \boldsymbol{\xi}_t^i). \tag{11}$$

Here, $\underline{V}_t^k(\boldsymbol{x}_{t-1}^k, \boldsymbol{\xi}_t^i)$ denotes the value of the lower bound at the k-th iteration evaluated at $\boldsymbol{x}_{t-1}^k$ and $\boldsymbol{\xi}_t^i$, and $\boldsymbol{\pi}_{t,i,k}^\star$ is the corresponding optimal dual extreme point. As is common in dynamic programming approaches, the cost-to-go functions are evaluated starting from the terminal stage and proceeding to the initial stage, hence, the process is called the *backward pass*.

However, this approach alone does not address the issue of evaluating the cost-to-go functions on infinitely many solutions $\boldsymbol{x}_{t-1} \in \mathcal{X}_{t-1}(\cdot)$. Instead of a brute-force discretization of the feasible region—which quickly becomes intractable as the dimension of $\boldsymbol{x}_{t-1}$ increases—SDDP performs a so-called *forward pass*: one (or multiple) trajectories $\{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2^{i_{k+1}}, \ldots, \boldsymbol{\xi}_{T-1}^{i_{k+1}}\}$ are sampled from the scenario tree according to the transition

probabilities (3): i.e., $\boldsymbol{\xi}_2^{i_{k+1}}$ is sampled given $\boldsymbol{\xi}_1$, then $\boldsymbol{\xi}_3^{i_{k+1}}$ is sampled given $\boldsymbol{\xi}_2^{i_{k+1}}$, and so on. Then, we obtain optimal solutions $\{\boldsymbol{x}_t^{k+1}\}_{t\in[T-1]}$ to the corresponding optimization problems using the current lower bounds: for $t \in [T-1]$

$$\boldsymbol{x}_t^{k+1} \in \arg\min_{\boldsymbol{x}_t}\Big\{ {\boldsymbol{c}_t^{i_{k+1}}}^\top \boldsymbol{x}_t + \underline{\mathcal{V}}_{t+1}^k(\boldsymbol{x}_t, \boldsymbol{\xi}_t^{i_{k+1}}) : \boldsymbol{x}_t \in \mathbb{R}_+^{d_t}, \boldsymbol{A}_t^{i_{k+1}}\boldsymbol{x}_t + \boldsymbol{B}_t^{i_{k+1}}\boldsymbol{x}_{t-1}^{k+1} = \boldsymbol{b}_t^{i_{k+1}} \Big\},$$

where $\underline{\mathcal{V}}_{t+1}^k(\boldsymbol{x}_t, \boldsymbol{\xi}_t^{i_{k+1}})$ denotes the approximate conditional expectation of the current (i.e., k-th) lower bounds $\underline{V}_{t+1}^k(\boldsymbol{x}_t, \boldsymbol{\xi}_{t+1}^i)$ for all $i \in [N]$ given $\boldsymbol{\xi}_t^{i_{k+1}}$ (analogous to (6)). Note that in the constraints in (3) the previous-stage solution $\boldsymbol{x}_{t-1}^{k+1}$ is used. Then, in the subsequent backward pass, the cost-to-go functions are evaluated only at these finitely many sampled solutions, also called *trial solutions*. In a later section on convergence, we will rigorously show that this construction of lower bounds ensures the validity of the cuts at every iteration, and that the initial-stage lower bound converges to $V_1(\boldsymbol{x}_0, \boldsymbol{\xi}_1)$.

Algorithm 1: Data-Driven SDDP (DD-SDDP)

Input: N sample trajectories $\boldsymbol{\xi}_{[T]}^i$, for all $i \in [N]$; $\overline{V}_1^0(\boldsymbol{x}_0, \boldsymbol{\xi}_1) = \infty$, $\underline{V}_1^0(\boldsymbol{x}_0, \boldsymbol{\xi}_1) = -\infty$;
$\overline{V}_{t+1}^0(\boldsymbol{x}_t, \boldsymbol{\xi}_{t+1}^i) = \infty$, $\underline{V}_{t+1}^0(\boldsymbol{x}_t, \boldsymbol{\xi}_{t+1}^i) = -\infty \quad \forall \boldsymbol{x}_t\ \forall i \in [N]\ \forall t \in [T-1]$;
$\overline{\mathcal{V}}_{T+1}^k(\boldsymbol{x}_T, \boldsymbol{\xi}_T^i) = 0$, $\underline{\mathcal{V}}_{T+1}^k(\boldsymbol{x}_T, \boldsymbol{\xi}_T^i) = 0 \quad \forall \boldsymbol{x}_T\ \forall i \in [N]\ \forall k \geq 1$

$k = 1$
while $\overline{V}_1^{k-1}(\boldsymbol{x}_0, \boldsymbol{\xi}_t) - \underline{V}_1^{k-1}(\boldsymbol{x}_0, \boldsymbol{\xi}_t) > \epsilon$ **do**
  Sample $(\boldsymbol{\xi}_2^{i_k}, \ldots, \boldsymbol{\xi}_{T-1}^{i_k})$ using transition probabilities (3)
  for $t = 1, \ldots, T-1$ **do**
    if $t = 1$ **then**
      Obtain $\boldsymbol{x}_1^k \in \arg\min_{\boldsymbol{x}_1}\Big\{ \boldsymbol{c}_1^\top \boldsymbol{x}_1 + \underline{\mathcal{V}}_2^{k-1}(\boldsymbol{x}_1, \boldsymbol{\xi}_1) : \boldsymbol{x}_1 \in \mathbb{R}_+^{d_1},\ \boldsymbol{A}_1\boldsymbol{x}_1 + \boldsymbol{B}_1\boldsymbol{x}_0 = \boldsymbol{b}_1 \Big\}$;
    else
      Obtain $\boldsymbol{x}_t^k \in \arg\min_{\boldsymbol{x}_t}\Big\{ {\boldsymbol{c}_t^{i_k}}^\top \boldsymbol{x}_t + \underline{\mathcal{V}}_{t+1}^{k-1}(\boldsymbol{x}_t, \boldsymbol{\xi}_t^{i_k}) : \boldsymbol{x}_t \in \mathbb{R}_+^{d_t},\ \boldsymbol{A}_t^{i_k}\boldsymbol{x}_t + \boldsymbol{B}_t^{i_k}\boldsymbol{x}_{t-1}^k = \boldsymbol{b}_t^{i_k} \Big\}$;
  for $t = T, \ldots, 2$ **do**
    for $i = 1, \ldots, N$ **do**
      Given $\boldsymbol{x}_{t-1}^k$, solve:
      $\min_{\boldsymbol{x}_t}\Big\{ {\boldsymbol{c}_t^i}^\top \boldsymbol{x}_t + \underline{\mathcal{V}}_{t+1}^k(\boldsymbol{x}_t, \boldsymbol{\xi}_t^i) : \boldsymbol{x}_t \in \mathbb{R}_+^{d_t},\ \boldsymbol{A}_t^i\boldsymbol{x}_t + \boldsymbol{B}_t^i\boldsymbol{x}_{t-1}^k = \boldsymbol{b}_t^i \Big\}$;
      Compute $\boldsymbol{\alpha}_{t,i,k}$ and $\beta_{t,i,k}$ in (11), and update the lower bound:
      $$\underline{V}_t^k(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t^i) \leftarrow \max\Big\{ \underline{V}_t^{k-1}(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t^i),\ {\boldsymbol{\alpha}_{t,k}^i}^\top \boldsymbol{x}_{t-1} + \beta_{t,k}^i \Big\}.$$
      Compute $\gamma_{t,k}^i$ in (12), and update the upper bound:
      $$\overline{V}_t^k(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t^i) \leftarrow \mathrm{env}\Big(\min\Big\{ \overline{V}_t^{k-1}(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t^i),\ \gamma_{t,k}^i + \mathbb{I}_{\{\boldsymbol{x}_{t-1}^k\}}(\boldsymbol{x}_{t-1}) \Big\} \Big).$$
  $k \leftarrow k + 1$;

3.1 Deterministic Upper Bound

In standard SDDP, the algorithm terminates when the gap between the initial-stage lower bound (i.e., $\underline{V}_1^k(\boldsymbol{x}_0, \boldsymbol{\xi}_1)$) and the *stochastic* upper bound falls within a predetermined tolerance level (see [44]). Although a small gap may indicate that the current approximations yield a good suboptimal policy, the randomness

in constructing the upper bound might trigger an early termination that does not accurately reflect the true quality of the approximation. To overcome this, we replace the stochastic upper bound with the deterministic upper bound proposed in [42, 48].

The initial upper bounds, denoted by $\overline{V}_t^0(\cdot)$, can be initialized with a sufficiently large constant. Then, analogous to the lower bound update, we update the upper bound in reverse time order, $t = T, T-1, \ldots, 1$, as follows:

$$\overline{V}_t^k(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t^i) \leftarrow \mathrm{env}\Big(\min\Big\{ \overline{V}_t^{k-1}(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t^i),\ \gamma_{t,k}^i + \mathbb{I}_{\{\boldsymbol{x}_{t-1}^k\}}(\boldsymbol{x}_{t-1}) \Big\} \Big) \quad \forall i \in [N]. \tag{12}$$

Here, $\gamma_{t,k}^i = \overline{V}_t^{k-1}(\boldsymbol{x}_{t-1}^k, \boldsymbol{\xi}_t^i)$ denotes the value of the $(k-1)$-th upper bound evaluated at $(\boldsymbol{x}_{t-1}^k, \boldsymbol{\xi}_t^i)$. Specifically, this is defined as

$$\begin{aligned} \overline{V}_t^{k-1}(\boldsymbol{x}_{t-1}^k, \boldsymbol{\xi}_t^i) = \quad \min \quad & {\boldsymbol{c}_t^i}^\top \boldsymbol{x}_t + \overline{\mathcal{V}}_{t+1}^{k-1}(\boldsymbol{x}_t, \boldsymbol{\xi}_t^i) + M_t \|\boldsymbol{y}_t\|_1 \\ \text{s.t.} \quad & \boldsymbol{x}_t \in \mathbb{R}_+^{d_t}, \quad \boldsymbol{\theta} \in \Delta^{k-1}, \quad \boldsymbol{y}_t \in \mathbb{R}^{d_t}, \\ & \boldsymbol{A}_t^i \boldsymbol{x}_t + \boldsymbol{B}_t^i \boldsymbol{x}_{t-1}^j = \boldsymbol{b}_t^i, \\ & \boldsymbol{y}_t = \boldsymbol{x}_t - \sum_{l \in [k-1]} \theta_l \boldsymbol{x}_t^l, \end{aligned} \tag{13}$$

where M_t is a sufficiently large constant and $\overline{\mathcal{V}}_{t+1}^{k-1}(\boldsymbol{x}_t, \boldsymbol{\xi}_t^i)$ denotes the approximate conditional expectation of the upper bounds $\overline{V}_{t+1}^{k-1}(\boldsymbol{x}_t, \boldsymbol{\xi}_{t+1}^j)$ for all $j \in [N]$ given $\boldsymbol{\xi}_t^i$. In (12), $\mathrm{env}(\cdot)$ denotes the lower convex envelope, and the indicator function $\mathbb{I} : \mathbb{R}^{d_{t-1}} \times \mathbb{R}^{d_{t-1}} \to \{0, \infty\}$ is defined as $\mathbb{I}_{\{\boldsymbol{x}_{t-1}^k\}}(\boldsymbol{x}_{t-1}) = 0$ if $\boldsymbol{x}_{t-1} = \boldsymbol{x}_{t-1}^k$, and ∞ otherwise. This update can be performed during the backward pass using the same trial solutions as those used in the lower bound update. Moreover, since the construction of the upper bounds is independent of that for the lower bounds, these processes can be parallelized to reduce runtime for each iteration.

Similar to the lower bound, the construction of the upper bound leverages the convexity of the cost-to-go function, ensuring that the lower convex envelope (12) provides a valid upper bound—the validity will be discussed in a later section. Since an earlier-stage upper bound depends on later-stage bounds, the envelope functions are constructed starting from the terminal stage. As the approximations $\overline{V}_t^k(\cdot)$ for all $t \in [T] \setminus \{1\}$ are refined, the initial-stage upper bound $\overline{V}_1^k(\boldsymbol{x}_0, \boldsymbol{\xi}_1)$ becomes tighter. Eventually, a small initial-stage gap, i.e., $\overline{V}_1^k(\boldsymbol{x}_0, \boldsymbol{\xi}_1) - \underline{V}_1^k(\boldsymbol{x}_0, \boldsymbol{\xi}_1)$, indicates that the resulting policy is a good suboptimal solution for the approximate MSLP problem.

Our data-driven SDDP (DD-SDDP) algorithm, which utilizes the deterministic upper bound, is provided in Algorithm 1.

3.2 Convergence

We defer the convergence proof to Section 4.4 since the distributionally robust version of DD-SDDP in the next section generalizes DD-SDDP.

4 Regularization Schemes

The theoretical results in Section 2 provide insight into the out-of-sample performance of our approximation scheme. Specifically, in Theorem 1, the dependence of the generalization bound on the conditional variance suggests that a large variance leads to a looser generalization bound, which in turn results in poor out-of-sample performance.

This observation motivates the development of a regularization scheme based on the conditional variance to enhance out-of-sample performance. However, directly incorporating the conditional variance is intractable due to its nonconvexity. Interestingly, we can establish equivalence between the variance-regularized formulation and a distributionally robust formulation based on the modified χ^2 distance. Leveraging this equivalence, we propose the data-driven distributionally robust SDDP (DDR-SDDP) scheme as a tractable alternative.

4.1 Variance-regularized Formulation

In Corollary 2, a constant upper bound σ_{t+1}^2 on the conditional variance $\mathbb{V}[Q_{t+1}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1})|\boldsymbol{\xi}_t]$ is used solely for the theoretical analysis. Since $\sigma_{t+1}^2 \geq \mathbb{V}[Q_{t+1}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1})|\boldsymbol{\xi}_t]$ for all feasible $\boldsymbol{x}_t$ and $\boldsymbol{\xi}_t \in \Xi_t$, replacing σ_{t+1}^2 with the true conditional variance leads to a tighter generalization bound. This observation suggests that incorporating the conditional variance into the objective function could yield policies with improved out-of-sample performance. However, under (**A2**), the *true* conditional variance is unknown. Instead, we could use its empirical counterpart defined by

$$\widehat{\mathbb{V}}\left[V_{t+1}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1}) \,\middle|\, \boldsymbol{\xi}_t\right] = \widehat{\mathbb{E}}\left[V_{t+1}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1})^2 \,\middle|\, \boldsymbol{\xi}_t\right] - \widehat{\mathbb{E}}\left[V_{t+1}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1}) \,\middle|\, \boldsymbol{\xi}_t\right]^2. \tag{14}$$

Using this empirical variance, we obtain a variance-regularized formulation of the approximate MSLP problem, expressed in terms of the stage-t cost-to-go function:

$$\begin{aligned} V_t^{\mathrm{VR}}(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t^i) = \quad & \min \quad \boldsymbol{c}_t^{i\top} \boldsymbol{x}_t + \mathcal{V}_{t+1}^{\mathrm{VR}}(\boldsymbol{x}_t, \boldsymbol{\xi}_t^i) + \lambda \sqrt{\widehat{\mathbb{V}}\left[V_{t+1}^{\mathrm{VR}}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1}) \,\middle|\, \boldsymbol{\xi}_t^i\right]} \\ & \text{s.t.} \quad \boldsymbol{x}_t \in \mathbb{R}_+^{d_t}, \\ & \quad\;\; \boldsymbol{A}_t^i \boldsymbol{x}_t + \boldsymbol{B}_t^i \boldsymbol{x}_{t-1} = \boldsymbol{b}_t^i. \end{aligned} \tag{15}$$

Here, $\mathcal{V}_{t+1}^{\mathrm{VR}}(\boldsymbol{x}_t, \boldsymbol{\xi}_t^i)$ and $\widehat{\mathbb{V}}[V_{t+1}^{\mathrm{VR}}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1})|\boldsymbol{\xi}_t^i]$ denote the approximate conditional expectation and variance, respectively (analogous to (6) and (14)). The parameter $\lambda \geq 0$ controls the contribution of the regularization term.

4.2 Distributionally Robust Formulation

Unfortunately, solving the variance regularized version of the MSLP problem in (15) is intractable since the conditional variance term (14) is nonconvex in $\boldsymbol{x}_t$. Instead, we employ distributionally robust optimization (DRO). Specifically, for $t \in [T-1]$ and $\boldsymbol{\xi}_t \in \Xi_t$, we consider the ambiguity set of distributions using the

modified χ^2 distance as follows:

$$\mathcal{P}^{\lambda}_{t+1}(\boldsymbol{\xi}_t) = \left\{ \sum_{i\in[N]} w^i \cdot \delta_{\boldsymbol{\xi}^i_{t+1}} \colon \boldsymbol{w} \in \Delta^N, \left(\left(\frac{w^i - \widehat{w}_{t+1}(\boldsymbol{\xi}_t, \boldsymbol{\xi}^i_t)}{\sqrt{\widehat{w}_{t+1}(\boldsymbol{\xi}_t, \boldsymbol{\xi}^i_t)}} \right)_{i\in[N]}, \lambda \right) \in \mathcal{C}^N \right\}. \tag{16}$$

Here, $\lambda \geq 0$ is a user-defined parameter, and $\delta_{\boldsymbol{\xi}^i_{t+1}}$ represents the Dirac distribution that assigns a unit mass at the sample point $\boldsymbol{\xi}^i_{t+1}$. Therefore, $\mathcal{P}^{\lambda}_{t+1}(\boldsymbol{\xi}_t)$ contains all distributions $\mathbb{P}_{t+1}$ supported on the N samples $\{\boldsymbol{\xi}^i_{t+1}\}_{i\in[N]}$ whose modified χ^2 distance from the kernel estimates $(\widehat{w}_{t+1}(\boldsymbol{\xi}_t, \boldsymbol{\xi}^i_t))_{i\in[N]}$ is within λ. With the ambiguity set (16), we formulate the DRO version of the approximate MSLP problem (4) as

$$\begin{array}{rll} V^{\mathrm{DR}}_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}^i_t) = & \min & \boldsymbol{c}^{i\top}_t \boldsymbol{x}_t + \max\limits_{\mathbb{P}_{t+1}\in\mathcal{P}^{\lambda}_{t+1}(\boldsymbol{\xi}^i_t)} \mathbb{E}_{\mathbb{P}_{t+1}}\left[V^{\mathrm{DR}}_{t+1}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1}) \,\middle|\, \boldsymbol{\xi}^i_t \right] \\ & \text{s.t.} & \boldsymbol{x}_t \in \mathbb{R}^{d_t}_+, \\ & & \boldsymbol{A}^i_t \boldsymbol{x}_t + \boldsymbol{B}^i_t \boldsymbol{x}_{t-1} = \boldsymbol{b}^i_t. \end{array} \tag{17}$$

Notably, there is a connection between the DRO formulation and the variance regularized formulation, as stated in the following proposition.

Proposition 2. *Consider the ambiguity set* (16) *with an arbitrary value of the parameter* $\lambda \geq 0$. *Then, for all* $t \in [T-1]$, $\boldsymbol{x}_t \in \mathcal{X}_t(\cdot)$, *and* $i \in [N]$, *we have*

$$\max_{\mathbb{P}_{t+1}\in\mathcal{P}^{\lambda}_{t+1}(\boldsymbol{\xi}^i_t)} \mathbb{E}_{\mathbb{P}_{t+1}}\left[V^{\mathrm{DR}}_{t+1}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1}) \,\middle|\, \boldsymbol{\xi}^i_t \right] \leq \mathcal{V}^{\mathrm{VR}}_{t+1}(\boldsymbol{x}_t, \boldsymbol{\xi}^i_t) + \lambda \sqrt{\widehat{\mathbb{V}}\left[V^{\mathrm{VR}}_{t+1}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1}) \,\middle|\, \boldsymbol{\xi}^i_t \right]}.$$

Moreover, equality is achieved if $\lambda^2 \leq \widehat{\mathbb{V}}[V^{\mathrm{VR}}_{t+1}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1}) | \boldsymbol{\xi}^i_t]$.

The proof of Proposition 2 can be found in Appendix D.

Proposition 2 establishes the condition under which the DRO formulation (17) is equivalent to the variance-regularized formulation (15). Even if this condition is not satisfied, the DRO formulation asymptotically converges to the variance-regularized formulation as N increases, since—as suggested by the bound (9)—λ should be scaled as $O(1/N^{\frac{2}{p+4}})$ to ensure good out-of-sample performance.

4.3 Data-Driven Distributionally Robust SDDP

Using duality, the DRO problems (17) with the modified χ^2 ambiguity set defined in (16) can be reformulated as a convex second-order cone program (SOCP). Although the SOCP provides a tractable formulation compared to the variance-regularized approach, the resulting reformulation yields a multistage *nonlinear* convex program. From a computational standpoint, this is not ideal, as the original problem we wish to solve involves only a linear objective function and constraints. Therefore, rather than using (16), we consider its polyhedral outer approximation $\mathscr{P}^{\lambda}_{t+1}(\boldsymbol{\xi}_t)$, which is obtained by replacing the second-order cone with the outer approximation

$$\mathscr{C}^N = \left\{ (\boldsymbol{u}, v) \in \mathbb{R}^N \times \mathbb{R} : \|\boldsymbol{u}\|_1 \leq \sqrt{N} v,\ \|\boldsymbol{u}\|_\infty \leq v \right\} \supseteq \mathcal{C}^N.$$

This polyhedral cone is defined through the intersection of 1-norm and ∞-norm balls of appropriate radii such that $\mathscr{C}^N$ tightly contains $\mathcal{C}^N$. The following result provides a single-level formulation of the DRO problems (17) using the ambiguity set $\mathscr{P}^{\lambda}_{t+1}(\boldsymbol{\xi}_t)$.

Proposition 3 (A Tractable Reformulation)**.** *Consider the DRO problems in* (17) *with the ambiguity set* $\mathscr{P}^{\lambda}_{t+1}(\boldsymbol{\xi}_t)$*. Let* $(\mathscr{C}^N)^* = \{(\boldsymbol{\psi}+\boldsymbol{\phi}, \sqrt{N}\eta+\mu) \in \mathbb{R}^N \times \mathbb{R} : \|\boldsymbol{\psi}\|_\infty \leq \eta,\ \|\boldsymbol{\phi}\|_\infty \leq \mu\}$ *be the dual cone of* $\mathscr{C}^N$*. Then, for each stage* $t \in [T]$*, the corresponding min-max optimization problem can be equivalently reformulated as the following polyhedral conic program:*

$$
\begin{aligned}
V_t^{\mathrm{DR}}(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t^i) = \min \quad & \boldsymbol{c}_t^{i\top}\boldsymbol{x}_t + \gamma + \lambda\rho - \sum_{j\in[N]} \sqrt{\widehat{w}_{t+1}(\boldsymbol{\xi}_t^i, \boldsymbol{\xi}_t^j)}\zeta_j \\
s.t. \quad & \boldsymbol{x}_t \in \mathbb{R}^{d_t}_+, \quad \gamma \in \mathbb{R}, \quad (\boldsymbol{\zeta}, \rho) \in (\mathscr{C}^N)^*, \\
& \boldsymbol{A}_t^i\boldsymbol{x}_t + \boldsymbol{B}_t^i\boldsymbol{x}_{t-1} = \boldsymbol{b}_t^i, \\
& \gamma - \frac{\zeta_j}{\sqrt{\widehat{w}_{t+1}(\boldsymbol{\xi}_t^i, \boldsymbol{\xi}_t^j)}} \geq V_{t+1}^{\mathrm{DR}}(\boldsymbol{x}_t, \boldsymbol{\xi}_{t+1}^j) \quad \forall j \in [N].
\end{aligned}
\tag{18}
$$

The proof of Proposition 3 is deferred to Appendix E.

As shown in (18), the reformulation is essentially a linear program. Hence, analogous to solving the approximate MSLP problem, the proposed SDDP method in Algorithm 1 can be applied for solving the DRO problem. We refer to this approach as *data-driven distributionally robust SDDP* (DDR-SDDP) for future reference. Specifically, we formulate optimization problems that provide lower and upper bounds on (18). For each $j \in [N]$, let $\bar{\mathcal{S}}_{t+1}^{j,k}$ denote the number of distinct cuts generated over k iterations for $V_{t+1}^{\mathcal{DR}}(\boldsymbol{x}_t, \boldsymbol{\xi}_{t+1}^j)$. Then, the lower-bound problem of (18) is written as

$$
\begin{aligned}
\underline{V}_{t,k}^{\mathrm{DR}}(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t^i) = \min \ & \boldsymbol{c}_t^{i\top}\boldsymbol{x}_t + \gamma + \lambda\rho - \sum_{j\in[N]} \sqrt{\widehat{w}_{t+1}(\boldsymbol{\xi}_t^i, \boldsymbol{\xi}_t^j)}\zeta_j \\
\text{s.t.} \quad & \boldsymbol{x}_t \in \mathbb{R}^{d_t}_+, \quad \gamma \in \mathbb{R}, \quad (\boldsymbol{\zeta}, \rho) \in (\mathscr{C}^N)^*, \\
& \boldsymbol{A}_t^i\boldsymbol{x}_t + \boldsymbol{B}_t^i\boldsymbol{x}_{t-1} = \boldsymbol{b}_t^i, \\
& \gamma - \frac{\zeta_j}{\sqrt{\widehat{w}_{t+1}(\boldsymbol{\xi}_t^i, \boldsymbol{\xi}_t^j)}} \geq \boldsymbol{\alpha}_{t+1,l}^{j\top}\boldsymbol{x}_t + \beta_{t+1,l}^j \ \ \forall j \in [N]\ \forall l \in \bar{\mathcal{S}}_{t+1}^{j,k}.
\end{aligned}
\tag{19}
$$

Here, $\boldsymbol{\alpha}_{t+1,l}^j \in \mathbb{R}^{d_t}$ and $\beta_{t+1,l}^j \in \mathbb{R}$ are defined as

$$
\boldsymbol{\alpha}_{t+1,l}^j = -\boldsymbol{B}_{t+1}^{j\top}\boldsymbol{\pi}_{t+1,j,l}^{\star} \quad \text{and} \quad \beta_{t+1,l}^j = \boldsymbol{\alpha}_{t+1,l}^{j\top}\boldsymbol{x}_t^l - \underline{V}_{t+1,l}^{\mathcal{DR}}(\boldsymbol{x}_t^l, \boldsymbol{\xi}_{t+1}^j), \tag{20}
$$

where $\boldsymbol{\pi}_{t+1,j,l}^{\star}$ denotes the subvector of an optimal extreme point of the dual problem associated with

$\underline{V}^{\mathrm{DR}}_{t+1,l}(\boldsymbol{x}^l_t, \boldsymbol{\xi}^j_{t+1})$—the full dual problem presented in the next section. The upper-bound problem of (18) is

$$
\begin{aligned}
\overline{V}^{\mathrm{DR}}_{t,k}(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}^i_t) = \min\ & \boldsymbol{c}_t^{i\top}\boldsymbol{x}_t + \gamma + \lambda\rho - \sum_{j\in[N]} \sqrt{\widehat{w}_{t+1}(\boldsymbol{\xi}^i_t, \boldsymbol{\xi}^j_t)}\zeta_j \\
\text{s.t.}\quad & \boldsymbol{x}_t, \boldsymbol{y}^j_t \in \mathbb{R}^{d_t}_+, \quad \gamma \in \mathbb{R}, \quad (\boldsymbol{\zeta}, \rho) \in (\mathscr{C}^N)^*, \quad \boldsymbol{\theta}^j \in \Delta^k \quad \forall j \in [N], \\
& \boldsymbol{A}^i_t\boldsymbol{x}_t + \boldsymbol{B}^i_t\boldsymbol{x}_{t-1} = \boldsymbol{b}^i_t, \\
& \gamma - \frac{\zeta_j}{\sqrt{\widehat{w}_{t+1}(\boldsymbol{\xi}^i_t, \boldsymbol{\xi}^j_t)}} \geq \overline{V}^{\mathrm{DR}}_{t+1,k}(\boldsymbol{x}_t, \boldsymbol{\xi}^j_{t+1}) \qquad \forall j \in [N], \\
& \sum_{l\in[k]} \theta^j_l \cdot \gamma^j_{t+1,l} + M_{t+1}\|\boldsymbol{y}^j_t\|_1 = \overline{V}^{\mathrm{DR}}_{t+1,k}(\boldsymbol{x}_t, \boldsymbol{\xi}^j_{t+1}) \qquad \forall j \in [N], \\
& \boldsymbol{y}^j_t = \boldsymbol{x}_t - \sum_{l\in[k]} \theta^j_l \boldsymbol{x}^l_t \qquad \forall j \in [N],
\end{aligned}
\tag{21}
$$

where $\gamma^j_{t+1,l} = \overline{V}^{\mathrm{DR}}_{t+1,l}(\boldsymbol{x}^l_t, \boldsymbol{\xi}^j_{t+1})$ is the value of the l-th upper bound evaluated at $(\boldsymbol{x}^l_t, \boldsymbol{\xi}^j_{t+1})$ and M_{t+1} is a sufficiently large constant.

Remark 2. *We are not the first to adopt the modified χ^2 ambiguity set in the literature on DRO-based SDDP variants. In [41], the authors use the same ambiguity set for the original MSLP problem, although they assume stagewise independence. However, they take a different approach to handling the inner maximization over the ambiguity set: rather than using duality to obtain the reformulation like* (18)*, they employ the KKT conditions to derive a closed-form solution for the worst-case distribution that maximizes the expected cost-to-go function, and then use this distribution to generate cuts during the backward pass, hence, the problem remains MSLP. A drawback of their approach is that the KKT conditions impose restrictions necessary for obtaining the closed-form solution—i.e., the vector representing the worst-case distribution may take on negative values, forcing an increase in the value of the parameter λ until the conditions hold. Unlike our method, which is applicable for any value of λ, their approach may restrict the tuning process and may lead to overly conservative solutions. It is also worth emphasizing that our specific choice of the modified χ^2 ambiguity set is motivated by our theoretical results: the suboptimality bound and its equivalence to the variance-regularized formulation.*

4.4 Convergence

When $\lambda = 0$ in the ambiguity set (16), the DRO problem (17) reduces to the approximate MSLP problem (5) as the only feasible vector $(w^i)_{i\in[N]}$ is $(\widehat{w}_{t+1}(\boldsymbol{\xi}_t, \boldsymbol{\xi}^i_t))_{i\in[N]}$. In this section, we establish the finite-iteration convergence of DDR-SDDP (and hence DD-SDDP). Specifically, we show that the gap between the lower and upper bounds eventually vanishes with probability one. To facilitate this, we make the following assumption.

(**A7**) **Extreme Points**. For each $t \in [T-1]$, the trial solution $\boldsymbol{x}^k_t$ obtained during the forward pass and the dual solution $\boldsymbol{\pi}^i_{t,k}$ for all $i \in [N]$ obtained during the backward pass are extreme points for all iteration $k \geq 1$.

This is a mild assumption that can be ensured by solving the linear programs during the forward and backward passes using the simplex method.

Before presenting the main convergence result, we formally show that the cost-to-go function $V_t^{\mathrm{DR}}(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$ is piecewise linear and convex in $\boldsymbol{x}_{t-1}$.

Lemma 3. *For any $t \in [T]\setminus\{1\}$ and $i \in [N]$, $V_t^{\mathrm{DR}}(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t^i)$ is piecewise linear and convex in $\boldsymbol{x}_{t-1} \in \mathcal{X}_{t-1}(\cdot)$ with a finite number of pieces.*

The proof of Lemma 3 is provided in Appendix F.

The proof of Lemma 3 illustrates the key idea behind cut generation. As shown in (F.2), an optimal dual extreme point depends on $\boldsymbol{x}_{t-1}$, indicating that for any trial solution obtained during the forward pass, there exists at least one corresponding optimal dual extreme point. That is, the k-th lower bound evaluated at $\{\boldsymbol{x}_{t-1}^l\}_{l\in[k]}$ can be expressed as $\underline{V}_{t,k}^{\mathrm{DR}}(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t^i) = \max\{\boldsymbol{\alpha}_{t,l}^{i\top}\boldsymbol{x}_{t-1} + \beta_{t,l}^i : l \in [\bar{S}_t^{i,k}]\}$, as in (20).
Next, we show that, at each iteration, the upper and lower bounds are valid.

Lemma 4. *Suppose that, for all $t \in [T]\setminus\{1\}$, L_t^{DR} is a Lipschitz constant of $V_t^{\mathrm{DR}}(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$ and the constant M_t in (21) is set sufficiently large such that $M_t \geq L_t^{\mathrm{DR}}$. Then, for any iteration $k \geq 1$, $t \in [T]\setminus\{1\}$, $\boldsymbol{x}_{t-1} \in \mathcal{X}_{t-1}(\cdot)$, and $i \in [N]$, the following inequalities hold:*

$$\underline{V}_{t,k}^{\mathrm{DR}}(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t^i) \leq V_t^{\mathrm{DR}}(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t^i) \leq \overline{V}_{t,k}^{\mathrm{DR}}(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t^i). \tag{22}$$

The proof of Lemma 4 is provided in Appendix G.

With these preliminary results, we now state the convergence guarantee for the proposed DDR-SDDP algorithm.

Theorem 2. *Suppose the lower bound* (19) *and upper bound* (21) *are iteratively constructed using the DDR-SDDP algorithm. Then, the initial-stage gap converges to zero in a finite number of iterations with probability one, i.e.,*

$$\mathrm{Prob}\left(\exists \tilde{k} \in \mathbb{Z}_+ \ : \ \underline{V}_{1,\tilde{k}}^{\mathrm{DR}}(\boldsymbol{x}_0, \boldsymbol{\xi}_1) = V_1^{\mathrm{DR}}(\boldsymbol{x}_0, \boldsymbol{\xi}_1) = \overline{V}_{1,\tilde{k}}^{\mathrm{DR}}(\boldsymbol{x}_0, \boldsymbol{\xi}_1)\right) = 1.$$

The proof of Theorem 2 is provided in Appendix H.

Remark 3. *As noted in [19], DRO-based SDDP variants (as well as SDDP methods incorporating certain risk measures) encounter difficulties in computing a stochastic upper bound. As a result, these approaches typically set a maximum number of iterations, k^{max}, and terminate the algorithm at the k^{max}-th iteration. This strategy can lead to either premature termination or unnecessary additional iterations. In contrast, our discussion demonstrates that convergence guarantees hold for DRO-based SDDP, and adopting a deterministic upper bound is more appealing than relying on setting a maximum number of iterations.*

5 Numerical Experiments

In this section, we present numerical experiments to evaluate the performance of the proposed methods. All optimization problems are implemented in Python 3.7 and solved using Gurobi 9.5.2 via the Gurobipy interface. The experiments are conducted on a laptop equipped with a 2.3GHz 6-core Intel Core i7 processor and 16GB of RAM.

5.1 Portfolio Optimization

We consider a multistage portfolio optimization problem where an investor reallocates assets over time to maximize terminal-stage utility. The portfolio includes K risky assets with random returns and a risk-free asset with a fixed return r_f, and the investor starts with \$1 in the risk-free asset. At each stage $t \in [T-1]$, before observing returns $\boldsymbol{\xi}_{t+1} \in \mathbb{R}^K$, the investor can buy ($\boldsymbol{u}_t^+ \in \mathbb{R}_+^K$) or sell ($\boldsymbol{u}_t^- \in \mathbb{R}_+^K$) risky assets. The value of asset i at stage t, $s_{t,i}$, updates based on its previous value, realized return $\xi_{t,i}$, purchases $u_{t,i}^+$, and sales $u_{t,i}^-$. Transaction costs f_b and f_s apply per unit bought or sold. The problem is formulated recursively via cost-to-go functions for $t \in [T-1]$.

$$\begin{aligned} Q_t(\boldsymbol{s}_{t-1}, \boldsymbol{\xi}_t) = \max \;\; & \mathbb{E}\left[Q_{t+1}(\boldsymbol{s}_t, \tilde{\boldsymbol{\xi}}_{t+1}) \;\middle|\; \boldsymbol{\xi}_t \right] \\ \text{s.t.} \;\; & \boldsymbol{s}_t \in \mathbb{R}_+^{K+1}, \quad \boldsymbol{u}_t^+, \boldsymbol{u}_t^- \in \mathbb{R}_+^K, \\ & s_{t,i} = \xi_{t,i} s_{t-1,i} + u_{t,i}^+ - u_{t,i}^- \quad \forall i \in [K], \\ & s_{t,K+1} = r_f s_{t-1,K+1} - (1+f_b)\mathbf{e}^\top \boldsymbol{u}_t^+ + (1-f_s)\mathbf{e}^\top \boldsymbol{u}_t^-, \end{aligned} \tag{23}$$

where $s_{0,i} = 0 \; \forall i \in [K]$, $s_{0,K+1} = 1$, and the terminal-stage cost-to-go function is

$$\begin{aligned} Q_T(\boldsymbol{s}_{T-1}, \boldsymbol{\xi}_T) = \;\; \max \;\; & u(\boldsymbol{\xi}_T^\top \boldsymbol{x}_T) \\ \text{s.t.} \;\; & \boldsymbol{s}_T \in \mathbb{R}_+^{K+1}, \\ & s_{T,i} = \xi_{T,i} s_{T-1,i} \quad \forall i \in [K], \\ & s_{T,K+1} = r_f s_{T-1,K+1}. \end{aligned} \tag{24}$$

Here, the function $u : \mathbb{R}_+ \to \mathbb{R}_+$ in (24) is a piecewise linear approximation of the log utility function. Also, note that, in (23), no immediate cost is imposed since the goal is solely to maximize the utility of cumulative wealth at the terminal stage T.

We compare our DD-SDDP and the distributionally robust version (DDR-SDDP) with several benchmark approaches. These include the stagewise independent SDDP scheme (Ind-SDDP), the hidden Markov model-based SDDP scheme (HMM-SDDP) proposed in [48], and the equally weighted portfolio (Equal) studied in [17].

We test the schemes with the historical weekly returns of the following data sets from December 2003 to January 2023: the 10 Industry Portfolios and 12 Industry Portfolios from the Fama-French online data library[‡], which include US stock portfolios categorized by industries; and the iShares Exchange-Trades Funds

[‡] `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

(iShares) data set downloaded from *yfinance*[‡‡]. We collect 120 weekly price trajectories per data set with a time horizon of $T = 8$ weeks. The first 50 are used for training, and the remaining 70 for out-of-sample evaluations. To tune the ambiguity parameter λ for DDR-SDDP and HMM-SDDP, we use cross-validation.

First, we show the convergence of our schemes. As shown in Figure 2, we observe that both DD-SDDP and DDR-SDDP are able to close the gap to less than 3% after 800 iterations, and the gap for DDR-SDDP is about three times smaller than one for DD-SDDP. Table 2 reports the out-of-sample performance for five different schemes. The results indicate that the DDR-SDDP scheme performs favorably relative to other benchmarks: it achieves the largest utility, the largest mean return, and the largest Sharpe ratio over all data sets. In addition, compared to DD-SDDP, DDR-SDDP provides a less risky policy in terms of standard deviation for all data sets, illustrating the connection with the variance regularization discussed in Section 4. Meanwhile, we observe that DD-SDDP also performs significantly better than the stagewise independent scheme, demonstrating the benefits of incorporating the time dependence present in real-world data.

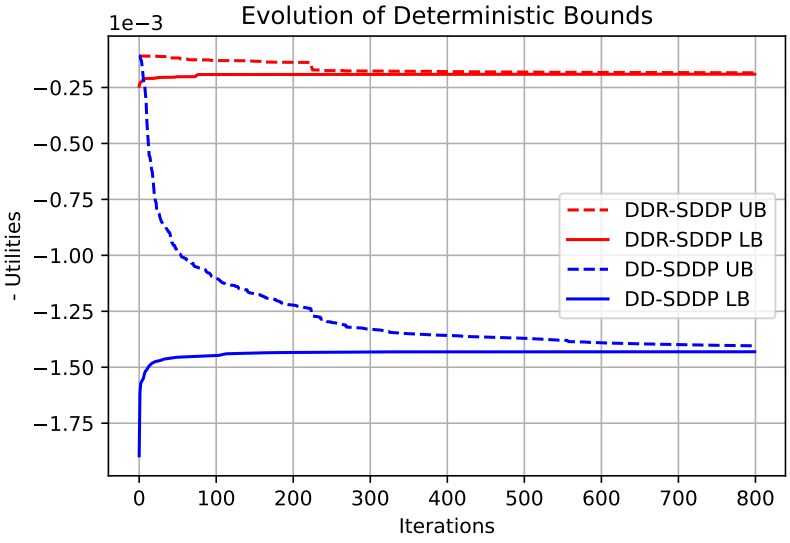

Figure 2. Evolution of both upper and lower bounds for DDR-SDDP and DD-SDDP over 800 iterations using the 10 Industry Portfolios data set.

5.2 Day-Ahead Wind Energy Commitment

We consider a wind energy commitment problem in the day-ahead market, considered in [31]. At the start of day t, the producer observes day-ahead prices $\boldsymbol{p}_t \in \mathbb{R}^{24}_+$ and selects 24-hour commitment levels $\boldsymbol{u}_t \in \mathbb{R}^{24}_+$, before actual production $\boldsymbol{w}_{t+1} \in \mathbb{R}^{24}_+$ is realized. On day $t+1$, commitments are met using wind generation or by discharging one of three storage devices. Surplus energy may charge storage or be discarded if capacity is full. Unmet commitments incur a penalty of twice the day-ahead price.

[‡‡] `https://pypi.org/project/yfinance/`

Table 2. Out-of-sample statistics of different schemes

Data set	Model	Utility	Mean return	Std. dev.	Sharpe ratio
10 Industry Fama-French	DDR-SDDP	**0.03190**	**0.03411**	0.06348	**0.47656**
	DD-SDDP	0.02654	0.02904	0.06957	0.36176
	Ind-SDDP	0.01582	0.01762	0.06228	0.22088
	HMM-SDDP	0.02428	0.02540	0.04697	0.45878
	Equal	-0.00215	-0.00197	**0.02743**	-0.03518
12 Industry Fama-French	DDR-SDDP	**0.02486**	**0.02669**	0.05974	**0.38238**
	DD-SDDP	0.02348	0.02591	0.06899	0.31962
	Ind-SDDP	0.02014	0.02203	0.06316	0.28780
	HMM-SDDP	0.01968	0.02086	0.04970	0.34211
	Equal	-0.00152	-0.00136	**0.02715**	-0.03166
iShares	DDR-SDDP	**0.01380**	**0.01425**	0.03605	**0.28860**
	DD-SDDP	0.00734	0.00945	0.06855	0.08160
	Ind-SDDP	0.00343	0.00534	0.06511	0.02283
	HMM-SDDP	0.00568	0.00609	0.03433	0.06526
	Equal	-0.05000	-0.04865	**0.02729**	-0.31754

Each storage device $l \in [3]$ has capacity $\bar{s}^l$, leakage γ^l, and (dis)charging efficiencies γ_c^l, γ_d^l. Let $s_t^l \in \mathbb{R}_+^{24}$ denote storage profiles, and $\boldsymbol{s}_t = (s_{t,24}^1, s_{t,24}^2, s_{t,24}^3)$ the end-of-day state. Random parameters $\boldsymbol{\xi}_t = (\boldsymbol{p}_t, \boldsymbol{w}_t)$ include prices and production. The producer aims to maximize expected profit over $T = 7$ days by solving a multistage dynamic program.

$$
\begin{aligned}
Q_t(\boldsymbol{s}_t, \boldsymbol{\xi}_t) = \max\ & \boldsymbol{p}_t^\top \boldsymbol{u}_t - 2\boldsymbol{p}_t^\top \mathbb{E}\left[\boldsymbol{e}_{t+1}^u \mid \boldsymbol{\xi}_t\right] + \mathbb{E}[Q_{t+1}(\boldsymbol{s}_{t+1}, \tilde{\boldsymbol{\xi}}_{t+1}) \mid \boldsymbol{\xi}_t] \\
\text{s.t.}\quad & \boldsymbol{u}_t, \boldsymbol{e}_{t+1}^{\{s,u,d\}} \in \mathbb{R}_+^{24}, \quad \boldsymbol{e}_{t+1}^{\{+,-\},l}, \boldsymbol{s}_{t+1}^l \in \mathbb{R}_+^{24} && \forall l \in [3], \\
& w_{t+1,h} = e_{t+1,h}^s + e_{t+1,h}^{+,1} + e_{t+1,h}^{+,2} + e_{t+1,h}^{+,3} + e_{t+1,h}^d && \forall h \in [24], \\
& u_{t,h} = e_{t+1,h}^s + e_{t+1,h}^{-,1} + e_{t+1,h}^{-,2} + e_{t+1,h}^{-,3} + e_{t+1,h}^u && \forall h \in [24], \\
& s_{t+1,h}^l = \gamma^l s_{t+1,h-1}^l + \gamma_c^l e_{t+1,h}^{+,l} - \frac{1}{\gamma_d^l} e_{t+1,h}^{-,l} \quad s_{t+1,h}^l \le \bar{s}^l && \forall h \in [24]\ \forall l \in [3],
\end{aligned}
$$

Here, $\boldsymbol{e}_{t+1}^s$ and $\boldsymbol{e}_{t+1}^u$ represent the amounts of satisfied and unsatisfied energy commitments, respectively, while $\boldsymbol{e}_{t+1}^d$ represents the amounts of dumped wind energy. In addition, $\boldsymbol{e}_{t+1}^{+,l}$ represent the amounts of wind energy used to charge storage l and $\boldsymbol{e}_{t+1}^{-,l}$ the amounts of energy discharged from storage l to meet the commitments.

We obtain the hourly day-ahead prices in the PJM market and the hourly wind energy from 2002 to 2011 at the following locations: **Ohio** (41.8125N, 81.5625W) and **North Carolina** (33.9375N, 77.9375W). By setting $T = 7$ days, we obtain 520 historical trajectories for each location. We apply principal component analysis to the high-dimensional data $\boldsymbol{\xi}_t = (\boldsymbol{p}_t, \boldsymbol{w}_t) \in \mathbb{R}_+^{48}$, reducing it to a 6-dimensional subspace that captures over 90% of the variance and mitigates the exponential dependence on p in the suboptimality bound (9).

Table 3 presents the out-of-sample performance for DDR-SDDP, DD-SDDP, and the stagewise independent scheme (Independent). Similar to the previous example, our data-driven schemes outperform the stagewise independent scheme in all criteria. We observe that DDR-SDDP wins in all the categories. Particularly, it is more robust in terms of the 10th percentile compared to other schemes, while the stagewise independent scheme has a significant risk of incurring a loss (negative profits).

Table 3. Out-of-sample statistics of profit (in $100,000)

Data set	Model	Mean	Variance	10th pct.
Ohio	DDR-SDDP	**7.12**	**18.61**	**2.72**
	DD-SDDP	6.71	22.57	2.07
	Ind-SDDP	5.61	21.58	0.78
North Carolina	DDR-SDDP	**8.62**	**42.82**	**1.47**
	DD-SDDP	8.27	53.47	1.07
	Ind-SDDP	7.55	57.12	-0.30

6 Conclusion

In this paper, we introduced a novel data-driven stochastic dual dynamic programming (SDDP) approach for multistage stochastic linear programming (MSLP) under a Markov data process. In contrast to existing SDDP variants that incorporate Markov dependence, we established out-of-sample performance guarantees for the data-driven solutions. These theoretical results motivated the development of a data-driven distributionally robust SDDP (DDR-SDDP) scheme as a tractable regularization method. Our numerical experiments demonstrate that DDR-SDDP outperforms all other benchmarks in real-world applications.

As demonstrated in several works, SDDP intersects with other sequential decision-making frameworks, such as reinforcement learning (RL). For instance, recent work [4] uses the concept of batch learning—commonly used in RL algorithms—to enhance the convergence of SDDP. Hence, there is potential to integrate ideas from the relevant fields. A key drawback of SDDP is its lack of an efficient online learning extension: as new samples become available, there is no straightforward method to update the current policy without re-solving the entire MSLP problem from scratch. While some SDDP variants using deep neural networks have been proposed, they lack any theoretical guarantees [14, 30]. In contrast, many RL algorithms are well-suited for online settings. Extending SDDP to incorporate an efficient online update would improve the practicality of the approach and present an interesting direction for future research.

Acknowledgements

This research was supported by the National Science Foundation grants no. 2343869 and 2404413.

References

[1] Shabbir Ahmed and Nikolaos V Sahinidis. An approximation scheme for stochastic integer programs arising in capacity expansion. *Operations Research*, 51(3):461–471, 2003.

[2] Shabbir Ahmed, Alan J King, and Gyana Parija. A multi-stage stochastic integer programming approach for capacity expansion under uncertainty. *Journal of Global Optimization*, 26(1):3–24, 2003.

[3] Shabbir Ahmed, Filipe Goulart Cabral, and Bernardo Freitas Paulo da Costa. Stochastic lipschitz dynamic programming. *Mathematical Programming*, 191(2):755–793, 2022.

[4] Daniel Ávila, Anthony Papavasiliou, and Nils Löhndorf. Batch learning sddp for long-term hydrothermal planning. *IEEE Transactions on Power Systems*, 39(1):614–627, 2023.

[5] Rüdiger Barth, Heike Brand, Peter Meibom, and Christoph Weber. A stochastic unit-commitment model for the evaluation of the impacts of integration of large amounts of intermittent wind power. In *International Conference on Probabilistic Methods Applied to Power Systems*, pages 1–8. IEEE, 2006.

[6] J Benders. Partitioning procedures for solving mixed-variables programming problems. *Computational Management Science*, 2(1), 2005.

[7] John R Birge. Decomposition and partitioning methods for multistage stochastic linear programs. *Operations Research*, 33(5):989–1007, 1985.

[8] J Frédéric Bonnans, Zhihao Cen, and Thibault Christel. Energy contracts management by stochastic programming techniques. *Annals of Operations Research*, 200(1):199–222, 2012.

[9] Stephen P Bradley and Dwight B Crane. A dynamic model for bond portfolio management. *Management Science*, 19(2):139–151, 1972.

[10] David R Carino, Terry Kent, David H Myers, Celine Stacy, Mike Sylvanus, Andrew L Turner, Kouji Watanabe, and William T Ziemba. The russell-yasuda kasai model: An asset/liability model for a Japanese insurance company using multistage stochastic programming. *Interfaces*, 24(1):29–49, 1994.

[11] Santiago Cerisola, Álvaro Baíllo, José M Fernández-López, Andrés Ramos, and Ralf Gollmer. Stochastic power generation unit commitment in electricity markets: A novel formulation and a comparison of solution methods. *Operations Research*, 57(1):32–46, 2009.

[12] Zhi-Long Chen and Warren B Powell. 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.

[13] Bernardo Freitas Paulo da Costa and Vincent Leclere. Dual sddp for risk-averse multistage stochastic programs. *Operations Research Letters*, 51(3):332–337, 2023.

[14] Hanjun Dai, Yuan Xue, Zia Syed, Dale Schuurmans, and Bo Dai. Neural stochastic dual dynamic programming. *arXiv preprint arXiv:2112.00874*, 2021.

[15] George B Dantzig and Gerd Infanger. Multi-stage stochastic linear programs for portfolio optimization. *Annals of Operations Research*, 45(1):59–76, 1993.

[16] Vitor L De Matos, Andy B Philpott, and Erlon C Finardi. Improving the performance of stochastic dual dynamic programming. *Journal of Computational and Applied Mathematics*, 290:196–208, 2015.

[17] Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The review of Financial studies*, 22(5):1915–1953, 2009.

[18] Anthony Downward, Oscar Dowson, and Regan Baucke. Stochastic dual dynamic programming with stagewise-dependent objective uncertainty. *Operations Research Letters*, 48(1):33–39, 2020.

[19] Daniel Duque and David P Morton. Distributionally robust stochastic dual dynamic programming. *SIAM Journal on Optimization*, 30(4):2841–2865, 2020.

[20] Laureano F Escudero, Pasumarti V Kamesam, Alan J King, and Roger JB Wets. Production planning via scenario modelling. *Annals of Operations Research*, 43(6):309–335, 1993.

[21] CHRISTIAN Füllner and STEFFEN Rebennack. Stochastic dual dynamic programming and its variants—a review. *Preprint, available at http://www. optimization-online. org/DB_FILE/2021/01/8217. pdf*, 2023.

[22] Pierre Girardeau, Vincent Leclere, and Andrew B Philpott. On the convergence of decomposition methods for multistage stochastic convex programs. *Mathematics of Operations Research*, 40(1):130–145, 2015.

[23] Bennett Golub, Martin Holmer, Raymond McKendall, Lawrence Pohlman, and Stavros A Zenios. A stochastic programming model for money management. *European Journal of Operational Research*, 85 (2):282–296, 1995.

[24] Vincent Guigues and Claudia Sagastizábal. Risk-averse feasible policies for large-scale multistage stochastic linear programs. *Mathematical Programming*, 138(1):167–198, 2013.

[25] László Györfi, Michael Kohler, Adam Krzyzak, Harro Walk, et al. *A Distribution-Free Theory of Nonparametric Regression*, volume 1. Springer, 2002.

[26] Grani A Hanasusanto and Daniel Kuhn. Robust data-driven dynamic programming. *In Advances in Neural Information Processing Systems*, 26, 2013.

[27] Grani A Hanasusanto, Daniel Kuhn, and Wolfram Wiesemann. A comment on "computational complexity of stochastic programming problems". *Mathematical Programming*, 159:557–569, 2016.

[28] Jianqiu Huang, Kezhuo Zhou, and Yongpei Guan. A study of distributionally robust multistage stochastic optimization. *arXiv preprint arXiv:1708.07930*, 2017.

[29] Gerd Infanger and David P Morton. Cut sharing for multistage stochastic linear programs with interstage dependency. *Mathematical Programming*, 75(2):241–256, 1996.

[30] Chanyeong Kim, Jongwoong Park, Hyunglip Bae, and Woo Chang Kim. Transformer-based stagewise decomposition for large-scale multistage stochastic optimization. *arXiv preprint arXiv:2404.02583*, 2024.

[31] Jae Ho Kim and Warren B Powell. Optimal energy commitments with storage and intermittent supply. *Operations Research*, 59(6):1347–1360, 2011.

[32] Daniel Kuhn, Wolfram Wiesemann, and Angelos Georghiou. Primal and dual linear decision rules in stochastic and robust optimization. *Mathematical Programming*, 130(1):177–209, 2011.

[33] Guanghui Lan. Complexity of stochastic dual dynamic programming. *Mathematical Programming*, 191 (2):717–754, 2022.

[34] Pascal Lavergne and Valentin Patilea. Breaking the curse of dimensionality in nonparametric testing. *Journal of Econometrics*, 143(1):103–122, 2008.

[35] Nils Löhndorf, David Wozabal, and Stefan Minner. Optimizing trading decisions for hydro storage systems using approximate dual dynamic programming. *Operations Research*, 61(4):810–823, 2013.

[36] John M Mulvey and Hercules Vladimirou. Stochastic network programming for financial planning problems. *Management Science*, 38(11):1642–1664, 1992.

[37] Elizbar A Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142, 1964.

[38] Mario VF Pereira and Leontina MVG Pinto. Stochastic optimization of a multireservoir hydroelectric system: A decomposition approach. *Water Resources Research*, 21(6):779–792, 1985.

[39] Mario VF Pereira and Leontina MVG Pinto. Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming*, 52(1):359–375, 1991.

[40] Andrew B Philpott and Ziming Guan. On the convergence of stochastic dual dynamic programming and related methods. *Operations Research Letters*, 36(4):450–455, 2008.

[41] Andrew B Philpott, Vitor L de Matos, and Lea Kapelevich. Distributionally robust SDDP. *Computational Management Science*, 15(3):431–454, 2018.

[42] Andy Philpott, Vitor de Matos, and Erlon Finardi. On solving multistage stochastic programs with coherent risk measures. *Operations Research*, 61(4):957–970, 2013.

[43] Suvrajeet Sen, Lihua Yu, and Talat Genc. A stochastic programming approach to power portfolio optimization. *Operations Research*, 54(1):55–72, 2006.

[44] Alexander Shapiro. Analysis of stochastic dual dynamic programming method. *European Journal of Operational Research*, 209(1):63–72, 2011.

[45] Alexander Shapiro and Arkadi Nemirovski. On complexity of stochastic programming problems. In Vaithilingam Jeyakumar and Alexander Rubinov, editors, *Continuous Optimization: Current Trends and Modern Applications*, pages 111–146. Springer, 2005.

[46] Alexander Shapiro, Wajdi Tekaya, Joari Paulo da Costa, and Murilo Pereira Soares. Risk neutral and risk averse stochastic dual dynamic programming method. *European Journal of Operational Research*, 224(2):375–391, 2013.

[47] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczynski. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, 2021.

[48] Thuener Silva, Davi Valladão, and Tito Homem-de Mello. A data-driven approach for a class of stochastic dynamic optimization problems. *Computational Optimization and Applications*, 80(3):687–729, 2021.

[49] Kavinesh J Singh, Andy B Philpott, and R Kevin Wood. Dantzig-wolfe decomposition for solving multistage stochastic capacity-planning problems. *Operations Research*, 57(5):1271–1286, 2009.

[50] Prateek R Srivastava, Yijie Wang, Grani A Hanasusanto, and Chin Pang Ho. On data-driven prescriptive analytics with side information: A regularized Nadaraya-Watson approach. *arXiv preprint arXiv:2110.04855*, 2021.

[51] Wim van Ackooij and Xavier Warin. On conditional cuts for stochastic dual dynamic programming. *EURO Journal on Computational Optimization*, 8(2):173–199, 2020.

[52] Geoffrey S Watson. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 359–372, 1964.

[53] Shixuan Zhang and Xu Andy Sun. On distributionally robust multistage convex optimization: new algorithms and complexity analysis. *arXiv preprint arXiv:2010.06759*, 2020.

[54] Shixuan Zhang and Xu Andy Sun. On distributionally robust multistage convex optimization: Data-driven models and performance. *arXiv preprint arXiv:2210.08433*, 2022.

[55] Shixuan Zhang and Xu Andy Sun. Stochastic dual dynamic programming for multistage stochastic mixed-integer nonlinear optimization. *Mathematical Programming*, pages 1–51, 2022.

[56] Jikai Zou, Shabbir Ahmed, and Xu Andy Sun. Stochastic dual dynamic integer programming. *Mathematical Programming*, 175(1):461–502, 2019.

Appendices

A Proof of Lemma 1

The proof of Lemma 1 relies on the Hoffman's Lemma shown below.

Lemma 5 (Hoffman Lemma (Theorem 9.14 in [47]))**.** *Let $\mathcal{M}(\boldsymbol{u}) = \{\boldsymbol{x} \in \mathbb{R}^n_+ \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{u}\}$ be a nonempty polyhedron parameterized by the right-hand side vector $\boldsymbol{u} \in \mathbb{R}^m$. Consider $\boldsymbol{u}' \in \mathbb{R}^m$ such that $\mathcal{M}(\boldsymbol{u}') \neq \emptyset$. Then, there exists a positive constant r such that, for any $\boldsymbol{x} \in \mathcal{M}(\boldsymbol{u})$*

$$dist\ (\boldsymbol{x}, \mathcal{M}(\boldsymbol{u}')) \leq r\|\boldsymbol{u} - \boldsymbol{u}'\|,$$

where $r = \max_{\boldsymbol{\lambda}_1 \in S_0} \|\boldsymbol{\lambda}_1\|_1$. Here, S_0 is a bounded polyhedral set satisfying $S = S_0 + C$, where $S = \{(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) \ : \ \|\boldsymbol{A}^\top \boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2\|_1 \leq 1\}$ is a polyhedron set that depends only on $\boldsymbol{A}$, and C is a polyhedral cone.

Proof of Lemma 1. We proceed by backward induction. At $t = T$, we have

$$Q_T(\boldsymbol{x}_{T-1}, \boldsymbol{\xi}_T) = V_T(\boldsymbol{x}_{T-1}, \boldsymbol{\xi}_T) = \min\left\{\boldsymbol{c}_T^\top \boldsymbol{x}_T : \boldsymbol{x}_T \in \mathbb{R}^{d_T}_+, \boldsymbol{A}_T\boldsymbol{x}_T + \boldsymbol{B}_T\boldsymbol{x}_{T-1} = \boldsymbol{b}_T\right\}.$$

Hence, it is sufficient to show the result for $Q_T(\boldsymbol{x}_{T-1}, \boldsymbol{\xi}_T)$. For any $\boldsymbol{\xi}_T \in \Xi_T$, let $\boldsymbol{x}_{T-1}$ and $\boldsymbol{x}'_{T-1}$ be two points in the domain of $Q_T(\boldsymbol{x}_{T-1}, \boldsymbol{\xi}_T)$. Fix any feasible point $\boldsymbol{x}_T \in \mathcal{X}_T(\boldsymbol{x}_{T-1}, \boldsymbol{\xi}_T)$. Applying Lemma 5 with the right-hand side vectors $\boldsymbol{u}_T = \boldsymbol{b}_T - \boldsymbol{B}_T\boldsymbol{x}_{T-1}$ and $\boldsymbol{u}'_T = \boldsymbol{b}_T - \boldsymbol{B}_T\boldsymbol{x}'_{T-1}$, there exists a feasible point $\boldsymbol{x}'_T \in \mathcal{X}_T(\boldsymbol{x}'_{T-1}, \boldsymbol{\xi}_T)$ such that

$$\|\boldsymbol{x}_T - \boldsymbol{x}'_T\| \leq r_T\|\boldsymbol{u}_T - \boldsymbol{u}'_T\| \leq r_T\|\boldsymbol{B}_T\| \cdot \|\boldsymbol{x}_{T-1} - \boldsymbol{x}'_{T-1}\|,$$

where $r_T > 0$ is a constant that depends only on $\boldsymbol{A}_T$. Therefore, the Lipschitz continuity of $\boldsymbol{c}_T^\top \boldsymbol{x}_T$ implies

$$Q_T(\boldsymbol{x}'_{T-1}, \boldsymbol{\xi}_T) \leq \boldsymbol{c}_T^\top \boldsymbol{x}'_T \leq \boldsymbol{c}_T^\top \boldsymbol{x}_T + r_T\|\boldsymbol{c}_T\| \cdot \|\boldsymbol{B}_T\| \cdot \|\boldsymbol{x}_{T-1} - \boldsymbol{x}'_{T-1}\|.$$

Taking minimization over $\boldsymbol{x}_T \in \mathcal{X}_T(\boldsymbol{x}_{T-1}, \boldsymbol{\xi}_T)$, we have

$$Q_T(\boldsymbol{x}'_{T-1}, \boldsymbol{\xi}_T) \leq Q_T(\boldsymbol{x}_{T-1}, \boldsymbol{\xi}_T) + r_T\|\boldsymbol{c}_T\| \cdot \|\boldsymbol{B}_T\| \cdot \|\boldsymbol{x}_{T-1} - \boldsymbol{x}'_{T-1}\|.$$

By symmetry, we have

$$\left|Q_T(\boldsymbol{x}_{T-1}, \boldsymbol{\xi}_T) - Q_T(\boldsymbol{x}'_{T-1}, \boldsymbol{\xi}_T)\right| \leq r_T\|\boldsymbol{c}_T\| \cdot \|\boldsymbol{B}_T\| \cdot \|\boldsymbol{x}_{T-1} - \boldsymbol{x}'_{T-1}\|.$$

By the compactness of the uncertainty set in (**A5**), there exist $\bar{r}_T, \bar{\boldsymbol{c}}_T,$ and $\bar{\boldsymbol{B}}_T$ such that the right-hand side is upper bounded by $\bar{r}_T\|\bar{\boldsymbol{c}}_T\| \cdot \|\bar{\boldsymbol{B}}_T\| \cdot \|\boldsymbol{x}_{T-1} - \boldsymbol{x}'_{T-1}\|$. Setting $L_T = \bar{r}_T\|\bar{\boldsymbol{c}}_T\| \cdot \|\bar{\boldsymbol{B}}_T\|$, we conclude that, for all $\boldsymbol{x}_{T-1}, \boldsymbol{x}'_{T-1} \in \mathcal{X}_{T-1}(\cdot)$, and $\boldsymbol{\xi}_T \in \Xi_T$

$$\left|Q_T(\boldsymbol{x}_{T-1}, \boldsymbol{\xi}_T) - Q_T(\boldsymbol{x}'_{T-1}, \boldsymbol{\xi}_T)\right| \leq L_T\|\boldsymbol{x}_{T-1} - \boldsymbol{x}'_{T-1}\|.$$

Since $Q_{T+1}(\cdot) = V_{T+1}(\cdot) = 0$, this further implies that, for all $\boldsymbol{x}_{T-1}, \boldsymbol{x}'_{T-1} \in \mathcal{X}_{T-1}(\cdot)$, and $\boldsymbol{\xi}_T \in \Xi_T$

$$\left|V_T(\boldsymbol{x}_{T-1}, \boldsymbol{\xi}_T^i) - V_T(\boldsymbol{x}'_{T-1}, \boldsymbol{\xi}_T^i)\right| \leq L_T \|\boldsymbol{x}_{T-1} - \boldsymbol{x}'_{T-1}\|.$$

Hence, the desired result holds for stage T. By induction, suppose the statement in the lemma holds for stage $t+1 \leq T$. That is, there exists a constant $L_{t+1} > 0$ such that, for all $\boldsymbol{x}_t, \boldsymbol{x}'_t \in \mathcal{X}_t(\cdot)$, and $\boldsymbol{\xi}_{t+1} \in \Xi_{t+1}$, we have

$$\left|Q_{t+1}(\boldsymbol{x}_t, \boldsymbol{\xi}_{t+1}) - Q_{t+1}(\boldsymbol{x}'_t, \boldsymbol{\xi}_{t+1})\right| \leq L_{t+1}\|\boldsymbol{x}_t - \boldsymbol{x}'_t\| \text{ and } \left|V_{t+1}(\boldsymbol{x}_t, \boldsymbol{\xi}_{t+1}) - V_{t+1}(\boldsymbol{x}'_t, \boldsymbol{\xi}_{t+1})\right| \leq L_{t+1}\|\boldsymbol{x}_t - \boldsymbol{x}'_t\|.$$

We now show the Lipschitz continuity of $Q_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$. For any feasible $\boldsymbol{x}_t, \boldsymbol{x}'_t \in \mathcal{X}_t(\cdot)$ and any $\boldsymbol{\xi}_t \in \Xi_t$, the following inequalities hold for the expected cost-to-go function:

$$\begin{aligned}
\left|\mathcal{Q}_{t+1}(\boldsymbol{x}_t, \boldsymbol{\xi}_t) - \mathcal{Q}_{t+1}(\boldsymbol{x}'_t, \boldsymbol{\xi}_t)\right| = &\left| \int_{\boldsymbol{\xi}_{t+1} \in \Xi_{t+1}} Q_{t+1}(\boldsymbol{x}_t, \boldsymbol{\xi}_{t+1}) f\left(\boldsymbol{\xi}_{t+1}|\boldsymbol{\xi}_t\right) \mathrm{d}\boldsymbol{\xi}_{t+1} \right. \\
&\left. - \int_{\boldsymbol{\xi}_{t+1} \in \Xi_{t+1}} Q_{t+1}(\boldsymbol{x}'_t, \boldsymbol{\xi}_{t+1}) f\left(\boldsymbol{\xi}_{t+1}|\boldsymbol{\xi}_t\right) \mathrm{d}\boldsymbol{\xi}_{t+1} \right| \\
\leq &\int_{\boldsymbol{\xi}_{t+1} \in \Xi_{t+1}} \left|Q_{t+1}(\boldsymbol{x}_t, \boldsymbol{\xi}_{t+1}) - Q_{t+1}(\boldsymbol{x}'_t, \boldsymbol{\xi}_{t+1})\right| \cdot f(\boldsymbol{\xi}_{t+1}|\boldsymbol{\xi}_t) \mathrm{d}\boldsymbol{\xi}_{t+1} \\
\leq &L_{t+1}\|\boldsymbol{x}_t - \boldsymbol{x}'_t\|.
\end{aligned}$$

Therefore, $\boldsymbol{c}_t^\top \boldsymbol{x}_t + \mathcal{Q}_{t+1}(\boldsymbol{x}_t, \boldsymbol{\xi}_t)$ is a Lipschitz function with constant $L_{t+1} + \|\boldsymbol{c}_t\|$. For any $\boldsymbol{\xi}_t \in \Xi_t$, let $\boldsymbol{x}_{t-1}$ and $\boldsymbol{x}'_{t-1}$ be two points in the domain of $Q_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$. Fix any feasible point $\boldsymbol{x}_t \in \mathcal{X}_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$. Applying Lemma 5 with the right-hand side vectors $\boldsymbol{u}_t = \boldsymbol{b}_t - \boldsymbol{B}_t\boldsymbol{x}_{t-1}$ and $\boldsymbol{u}'_t = \boldsymbol{b}_t - \boldsymbol{B}_t\boldsymbol{x}'_{t-1}$, we have that there exists a feasible point $\boldsymbol{x}'_t \in \mathcal{X}_t(\boldsymbol{x}'_{t-1}, \boldsymbol{\xi}_t)$ such that

$$\|\boldsymbol{x}_t - \boldsymbol{x}'_t\| \leq r_t\|\boldsymbol{u}_t - \boldsymbol{u}'_t\| \leq r_t\|\boldsymbol{B}_t\| \cdot \|\boldsymbol{x}_{t-1} - \boldsymbol{x}'_{t-1}\|$$

where $r_t > 0$ is a constant that depends only on $\boldsymbol{A}_t$. Therefore, the Lipschitz continuity of $\boldsymbol{c}_t^\top \boldsymbol{x}_t + \mathcal{Q}_{t+1}(\boldsymbol{x}_t, \boldsymbol{\xi}_t)$ implies

$$\begin{aligned}
Q_t(\boldsymbol{x}'_{t-1}, \boldsymbol{\xi}_t) &\leq \boldsymbol{c}_t^\top \boldsymbol{x}'_t + \mathcal{Q}_{t+1}(\boldsymbol{x}'_t, \boldsymbol{\xi}_t) \\
&\leq \boldsymbol{c}_t^\top \boldsymbol{x}_t + \mathcal{Q}_{t+1}(\boldsymbol{x}_t, \boldsymbol{\xi}_t) + r_t(L_{t+1} + \|\boldsymbol{c}_t\|) \cdot \|\boldsymbol{B}_t\| \cdot \|\boldsymbol{x}_{t-1} - \boldsymbol{x}'_{t-1}\|.
\end{aligned}$$

Taking minimization over $\boldsymbol{x}_t \in \mathcal{X}_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$, we have

$$Q_t(\boldsymbol{x}'_{t-1}, \boldsymbol{\xi}_t) \leq Q_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t) + r_t(L_{t+1} + \|\boldsymbol{c}_t\|) \cdot \|\boldsymbol{B}_t\| \cdot \|\boldsymbol{x}_{t-1} - \boldsymbol{x}'_{t-1}\|.$$

By symmetry, we have

$$\left|Q_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t) - Q_t(\boldsymbol{x}'_{t-1}, \boldsymbol{\xi}_t)\right| \leq r_t(L_{t+1} + \|\boldsymbol{c}_t\|) \cdot \|\boldsymbol{B}_t\| \cdot \|\boldsymbol{x}_{t-1} - \boldsymbol{x}'_{t-1}\|.$$

By the compactness of the uncertainty set in (**A5**), we know there exist $\bar{r}_t, \bar{\boldsymbol{c}}_t,$ and $\bar{\boldsymbol{B}}_t$ such that the right-hand side is upper bounded by $\bar{r}_t(L_{t+1} + \|\bar{\boldsymbol{c}}_t\|) \cdot \|\bar{\boldsymbol{B}}_t\| \cdot \|\boldsymbol{x}_{t-1} - \boldsymbol{x}'_{t-1}\|$. Setting $L_t = \bar{r}_t(L_{t+1} + \|\bar{\boldsymbol{c}}_t\|) \cdot \|\bar{\boldsymbol{B}}_t\|$, we conclude that, for all $\boldsymbol{x}_{t-1}, \boldsymbol{x}'_{t-1} \in \mathcal{X}_{t-1}(\cdot)$, and $\boldsymbol{\xi}_t \in \Xi_T$, we have

$$\left|Q_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t) - Q_t(\boldsymbol{x}'_{t-1}, \boldsymbol{\xi}_t)\right| \leq L_t\|\boldsymbol{x}_{t-1} - \boldsymbol{x}'_{t-1}\|.$$

Therefore, the result holds for $Q_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$. We omit the proof for $V_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$ since it is analogous to the derivation above. Hence, the desired result holds for stage t, and this completes the induction. □

B Proof of Corollary 2

Proof. For notational simplicity, for any $t \in [T-1]$, $\boldsymbol{\xi}_t \in \Xi_t$, and $\boldsymbol{x}_t \in \mathcal{X}_t(\cdot)$, we define $\epsilon_{t+1}(\boldsymbol{x}_t, \boldsymbol{\xi}_t)$ as the error term appearing on the left-hand side of (8). Recall that, as assumed in (**A3**), the feasible region $\mathcal{X}_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t) \subseteq \mathbb{R}^d$ has a finite diameter D for any $\boldsymbol{x}_{t-1}$ and $\boldsymbol{\xi}_t$. Also, let us define a set of finite number of points $\mathcal{X}_t^\eta(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t) \subseteq \mathcal{X}_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$ such that for any $\boldsymbol{x}_t \in \mathcal{X}_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$, there exists $\boldsymbol{x}'_t \in \mathcal{X}_t^\eta(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$ such that $\|\boldsymbol{x}_t - \boldsymbol{x}'_t\| \leq \eta$. As shown in [45], the cardinality of such a finite set depends on the value of the parameter η, particularly, $|\mathcal{X}_t^\eta(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)| = O(1)(D/\eta)^d$.

Lemma 1 implies that the Lipschitz continuity of the *expected* cost-to-go function, i.e., $\mathcal{Q}_{t+1}(\boldsymbol{x}_t, \boldsymbol{\xi}_t)$ is L_{t+1}-Lipschitz continuous in $\boldsymbol{x}_t$. Therefore, for any $\boldsymbol{x}_t \in \mathcal{X}_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$, there exists $\boldsymbol{x}'_t \in \mathcal{X}_t^\eta(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$, $\|\boldsymbol{x}_t - \boldsymbol{x}'_t\| \leq \eta$, such that $\mathcal{Q}_{t+1}(\boldsymbol{x}_t, \boldsymbol{\xi}_t) - \mathcal{Q}_{t+1}(\boldsymbol{x}'_t, \boldsymbol{\xi}_t) \leq L_{t+1}\eta$. Then, from Proposition 1, for any fixed $\boldsymbol{x}'_t \in \mathcal{X}_t^\eta(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$, $\boldsymbol{\xi}_t \in \Xi_t$, and $\delta \in [0,1]$, we have

$$\epsilon_{t+1}(\boldsymbol{x}'_t, \boldsymbol{\xi}_t) \leq \sqrt{\frac{\mathbb{V}\left[Q_{t+1}(\boldsymbol{x}'_t, \tilde{\boldsymbol{\xi}}_{t+1}) \,\middle|\, \boldsymbol{\xi}_t\right]}{O(N^{\frac{4}{p+4}})(1+o(1))g_{t+1}(\boldsymbol{\xi}_t)} \log\left(\frac{1}{\delta}\right)} \tag{B.1}$$

with probability at least $1-\delta$. Then, applying union bound over $\mathcal{X}_t^\eta(\cdot)$, we have

$$\begin{aligned}
\epsilon_{t+1}(\boldsymbol{x}'_t, \boldsymbol{\xi}_t) &\leq \sqrt{\frac{\mathbb{V}\left[Q_{t+1}(\boldsymbol{x}'_t, \tilde{\boldsymbol{\xi}}_{t+1}) \,\middle|\, \boldsymbol{\xi}_t\right]}{O(N^{\frac{4}{p+4}})(1+o(1))g_{t+1}(\boldsymbol{\xi}_t)} \log\left(\frac{|\mathcal{X}_t^\eta(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)|}{\delta}\right)} \\
&\leq \sqrt{\sigma_{t+1}^2 \frac{\log\left(\frac{O(1)(D/\eta)^d}{\delta}\right)}{O(N^{\frac{4}{p+4}})(1+o(1))g_{t+1}(\boldsymbol{\xi}_t)}} \qquad \forall \boldsymbol{x}'_t \in \mathcal{X}_t^\eta(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)
\end{aligned}$$

with probability at least $1-\delta$. Using the Lipschitz continuity of $\mathcal{Q}_{t+1}(\cdot)$, we obtain (8). This completes the proof. □

C Proof of Theorem 1

Proof of Theorem 1. To simplify notations, we define the following term for the error bound in Corollary 2: for all $t \in [T] \setminus \{1\}$ and $i, j \in \mathbb{Z}_+$

$$\varepsilon_t(i,j) = \sqrt{\sigma_t^2 \frac{\log\left(\frac{O(1)N^i(D/\eta)^{dj}}{\delta_t}\right)}{O(N^{\frac{4}{p+4}})(1+o(1))g_t}}.$$

Using induction, we first show that at stage t we have for any fixed $\boldsymbol{x}_{t-2}$ and $\boldsymbol{\xi}_{t-1}$

$$\left|\mathcal{Q}_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_{t-1}) - \mathcal{V}_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_{t-1})\right| \leq \sum_{s=t}^{T} \left(\varepsilon_s(s-t, s-t+1) + 2L_s\eta\right) \tag{C.1}$$

for all $\boldsymbol{x}_{t-1} \in \mathcal{X}_1(\boldsymbol{x}_{t-2}, \boldsymbol{\xi}_{t-1})$ with probability at least $1 - \sum_{s=t}^{T} \delta_t$.

Stage $t = T$: We have $Q_T(\cdot, \boldsymbol{\xi}_T) = V_T(\cdot, \boldsymbol{\xi}_T)$ for all $\boldsymbol{\xi}_T$ due to the fact that $Q_{T+1}(\cdot) = V_{T+1}(\cdot) = 0$. From Corollary 2, for any fixed $\boldsymbol{x}_{T-2}$ and $\boldsymbol{\xi}_{T-1}$ we have

$$\Big|\mathcal{Q}_T(\boldsymbol{x}_{T-1}, \boldsymbol{\xi}_{T-1}) - \mathcal{V}_T(\boldsymbol{x}_{T-1}, \boldsymbol{\xi}_{T-1})\Big| \leq \varepsilon_T(0,1) + 2L_T\eta \tag{C.2}$$

for all $\boldsymbol{x}_{T-1} \in \mathcal{X}_{T-1}(\boldsymbol{x}_{T-2}, \boldsymbol{\xi}_{T-1})$ with probability at least $1 - \delta_T$. Hence, the base case of our induction holds.

Stage $t < T$: By our induction hypothesis, we have that (C.1) holds at stage t. Adding and subtracting the term $\boldsymbol{c}_{t-1}^\top \boldsymbol{x}_{t-1}$ on the left-hand side of (C.1) and then taking the minimum over $\mathcal{X}_{t-1}(\cdot)$, we obtain for any fixed $\boldsymbol{x}_{t-2}$ and $\boldsymbol{\xi}_{t-1}$,

$$\begin{aligned}
&\left|\min_{\boldsymbol{x}_{t-1} \in \mathcal{X}_{t-1}(\boldsymbol{x}_{t-2}, \boldsymbol{\xi}_{t-1})} \Big\{\boldsymbol{c}_{t-1}^\top \boldsymbol{x}_{t-1} + \mathcal{Q}_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_{t-1})\Big\} - \min_{\boldsymbol{x}_{t-1} \in \mathcal{X}_{t-1}(\boldsymbol{x}_{t-2}, \boldsymbol{\xi}_{t-1})} \Big\{\boldsymbol{c}_{t-1}^\top \boldsymbol{x}_{t-1} + \mathcal{V}_T(\boldsymbol{x}_{T-1}, \boldsymbol{\xi}_{T-1})\Big\}\right| \\
&= \big|Q_{t-1}(\boldsymbol{x}_{t-2}, \boldsymbol{\xi}_{t-1}) - V_{t-1}(\boldsymbol{x}_{t-2}, \boldsymbol{\xi}_{t-1})\big| \\
&\leq \sum_{s=t}^{T} \big(\varepsilon_s(s-t, s-t+1) + 2L_s\eta\big)
\end{aligned}$$

with probability at least $1 - \sum_{s=t}^{T} \delta_t$. Applying union bound over the N sample points $\{\boldsymbol{\xi}_{t-1}^i\}_{i=1}^N$, we have for any fixed $\boldsymbol{x}_{t-2}$,

$$\Big|Q_{t-1}(\boldsymbol{x}_{t-2}, \boldsymbol{\xi}_{t-1}^i) - V_{t-1}(\boldsymbol{x}_{t-2}, \boldsymbol{\xi}_{t-1}^i)\Big| \leq \sum_{s=t}^{T} \big(\varepsilon_s(s-t+1, s-t+1) + 2L_s\eta\big) \tag{C.3}$$

for all $i \in [N]$ with probability at least $1 - \sum_{s=t}^{T} \delta_t$.

We now attempt to derive a bound on the expected cost-to-go at time $t-1$. From Proposition 1, for any fixed $\boldsymbol{x}_{t-2}$ and $\boldsymbol{\xi}_{t-2}$, we have

$$\left|\mathbb{E}\Big[Q_{t-1}(\boldsymbol{x}_{t-2}, \tilde{\boldsymbol{\xi}}_{t-1})\Big|\boldsymbol{\xi}_{t-2}\Big] - \widehat{\mathbb{E}}\Big[Q_{t-1}(\boldsymbol{x}_{t-2}, \tilde{\boldsymbol{\xi}}_{t-1})\Big|\boldsymbol{\xi}_{t-2}\Big]\right| \leq \varepsilon_{t-1}(0,0) \tag{C.4}$$

with probability at least $1 - \delta_{t-1}$. Note that $\widehat{\mathbb{E}}[Q_{t-1}(\boldsymbol{x}_{t-2}, \tilde{\boldsymbol{\xi}}_{t-1})|\boldsymbol{\xi}_{t-2}]$ in (C.4) is not the approximate expected cost-to-go function $\mathcal{V}_{t-1}(\cdot) = \widehat{\mathbb{E}}[V_{t-1}(\boldsymbol{x}_{t-2}, \tilde{\boldsymbol{\xi}}_{t-1})|\boldsymbol{\xi}_{t-2}]$ that we evaluate during the backward pass of SDDP. Therefore, we use (C.3) to replace $Q_{t-1}(\cdot)$ with $V_{t-1}(\cdot)$, and obtain for any fixed $\boldsymbol{x}_{t-2}$ and $\boldsymbol{\xi}_{t-2}$,

$$\Big|\mathcal{Q}_{t-1}(\boldsymbol{x}_{t-2}, \boldsymbol{\xi}_{t-2}) - \mathcal{V}_{t-1}(\boldsymbol{x}_{t-2}, \boldsymbol{\xi}_{t-2})\Big| \leq \varepsilon_{t-1}(0,0) + \sum_{s=t}^{T} \big(\varepsilon_s(s-t+1, s-t+1) + 2L_s\eta\big) \tag{C.5}$$

with probability at least $1 - \sum_{s=t-1}^{T} \delta_t$. Analogous to the discretization used in the proof of Corollary 2 (Appendix B), we have $|\mathcal{X}_{t-2}^\eta\left(\boldsymbol{x}_{t-3}, \boldsymbol{\xi}_{t-2}\right)| = O(1)(D/\eta)^d$. Since $Q_{t-1}(\boldsymbol{x}_{t-2}, \boldsymbol{\xi}_{t-1})$ is L_{t-1}-Lipschitz continuous in $\boldsymbol{x}_{t-2}$ as shown in Lemma 1, this implies that for any $\boldsymbol{x}_{t-2} \in \mathcal{X}_{t-2}(\cdot)$, there exists $\boldsymbol{x}'_{t-2} \in \mathcal{X}_{t-2}^\eta(\cdot)$, $\|\boldsymbol{x}_{t-2} - \boldsymbol{x}'_{t-2}\| \leq \eta$, such that

$$\Big|\mathcal{Q}_{t-1}(\boldsymbol{x}_{t-2}, \boldsymbol{\xi}_{t-2}) - \mathcal{Q}_{t-1}(\boldsymbol{x}'_{t-2}, \boldsymbol{\xi}_{t-2})\Big| \leq L_{t-1}\eta \tag{C.6}$$

Furthermore, applying union bound to (C.5) over $\boldsymbol{x}'_{t-2} \in \mathcal{X}^{\eta}_{t-2}(\boldsymbol{x}_{t-3}, \boldsymbol{\xi}_{t-2})$, we get

$$\Big|\mathcal{Q}_{t-1}(\boldsymbol{x}'_{t-2}, \boldsymbol{\xi}_{t-2}) - \mathcal{V}_{t-1}(\boldsymbol{x}'_{t-2}, \boldsymbol{\xi}_{t-2})\Big| \leq \sum_{s=t-1}^{T} \varepsilon_s(s-t+1, s-t+1) + \sum_{u=t}^{T} 2L_u\eta$$

for all $\boldsymbol{x}'_{t-2} \in \mathcal{X}^{\eta}_{t-2}(\boldsymbol{x}_{t-3}, \boldsymbol{\xi}_{t-2})$ with probability at least $1 - \sum_{s=t-1}^{T} \delta_t$. Finally, in view of the Lipschitz continuity of $Q_{T-1}(\cdot)$ and $V_{t-1}(\cdot)$, we therefore prove that the bound (C.1) holds at stage $t-1$, which completes the induction step.

Using this bound at stage $t = 2$ for $\boldsymbol{x}_1 = \widehat{\boldsymbol{x}}_1^N$, we have

$$\begin{aligned}
\boldsymbol{c}_1^\top \widehat{\boldsymbol{x}}_1^N + \mathbb{E}\left[Q_2(\widehat{\boldsymbol{x}}_1^N, \tilde{\boldsymbol{\xi}}_2) \;\middle|\; \boldsymbol{\xi}_1\right] &\leq \left(\boldsymbol{c}_1^\top \widehat{\boldsymbol{x}}_1^N + \widehat{\mathbb{E}}\left[V_2(\widehat{\boldsymbol{x}}_1^N, \tilde{\boldsymbol{\xi}}_2) \;\middle|\; \boldsymbol{\xi}_1\right]\right) + \sum_{t=2}^{T} \big(\varepsilon_t(t-2, t-1) + 2L_t\eta\big) \\
&\leq \left(\boldsymbol{c}_1^\top \boldsymbol{x}_1^\star + \widehat{\mathbb{E}}\left[V_2(\boldsymbol{x}_1^\star, \tilde{\boldsymbol{\xi}}_2) \;\middle|\; \boldsymbol{\xi}_1\right]\right) + \sum_{t=2}^{T} \big(\varepsilon_t(t-2, t-1) + 2L_t\eta\big),
\end{aligned}$$

where the second inequality holds since $\boldsymbol{x}_1^\star$ is suboptimal to the approximate problem. A similar bound can be establish for $\boldsymbol{x}_1 = \boldsymbol{x}_1^\star$. Then, combining the two bounds, we establish (9). This completes the proof. □

D Proof of Proposition 2

Proof. To simplify the notation, we define a vector $\boldsymbol{z} \in \mathbb{R}^N$ whose i-th component is $z_i = V_{t+1}^{\mathcal{VR}}(\boldsymbol{x}_t, \boldsymbol{\xi}_{t+1}^i)$. We further simplify the conditional expectation and variance as follows:

$$\begin{aligned}
\overline{z} &= \widehat{\mathbb{E}}\left[V_{t+1}^{\mathrm{VR}}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1}) \;\middle|\; \boldsymbol{\xi}_t\right] = \sum_{i\in[N]} \widehat{w}_{t+1}(\boldsymbol{\xi}_t, \boldsymbol{\xi}_t^i) \cdot V_{t+1}^{\mathrm{VR}}(\boldsymbol{x}_t, \boldsymbol{\xi}_{t+1}^i), \\
s &= \widehat{\mathbb{V}}\left[V_{t+1}^{\mathrm{VR}}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1}) \;\middle|\; \boldsymbol{\xi}_t\right] = \sum_{i\in[N]} w_{t+1}(\boldsymbol{\xi}_t, \boldsymbol{\xi}_t^i) \cdot \Big(V_{t+1}^{\mathrm{VR}}(\boldsymbol{x}_t, \boldsymbol{\xi}_{t+1}^i) - \overline{z}\Big)^2.
\end{aligned}$$

Using these notations, we can define the value of the variance-regularized formulation in Proposition 2 as follows:

$$z_\lambda^{\mathrm{VR}} = \overline{z} + \lambda\sqrt{s} = \widehat{\mathbb{E}}\left[V_{t+1}^{\mathrm{VR}}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1}) \;\middle|\; \boldsymbol{\xi}_t\right] + \lambda\sqrt{\widehat{\mathbb{V}}\left[V_{t+1}^{\mathrm{VR}}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1}) \;\middle|\; \boldsymbol{\xi}_t\right]}.$$

The DRO formulation in Proposition 2 can be equivalently written as

$$z_\lambda^{\mathrm{DR}} = \max_{\boldsymbol{w}_{t+1}} \Big\{\boldsymbol{w}_{t+1}^\top \boldsymbol{z} : \boldsymbol{w}_{t+1} \in \mathbb{R}_+^N, \mathbf{e}^\top \boldsymbol{w}_{t+1} = 1, \sum_{i\in[N]} \frac{(w_{t+1}^i - \widehat{w}_{t+1}(\boldsymbol{\xi}_t, \boldsymbol{\xi}_t^i))^2}{\widehat{w}_{t+1}(\boldsymbol{\xi}_t, \boldsymbol{\xi}_t^i)} \leq \lambda^2\Big\}. \tag{D.1}$$

Define a vector $\boldsymbol{u}_{t+1} \in \mathbb{R}^N$ whose i-th component is $u_{t+1}^i = w_{t+1}^i - \widehat{w}_{t+1}(\boldsymbol{\xi}_t, \boldsymbol{\xi}_t^i)$ by change of variable. Then, the optimization problem (D.1) is equivalent to

$$\begin{aligned}
z_\lambda^{\mathrm{DR}} = \max_{\boldsymbol{u}_{t+1}} \big\{\overline{z} + \boldsymbol{u}_{t+1}^\top(\boldsymbol{z} - \overline{z}\cdot\mathbf{e}) \;:\; & u_{t+1}^i + \widehat{w}_{t+1}(\boldsymbol{\xi}_t, \boldsymbol{\xi}_t^i) \geq 0 \;\; \forall i \in [N], \\
& \mathbf{e}^\top \boldsymbol{u}_{t+1} = 0, \;\; \|\boldsymbol{u}_{t+1}\|_W \leq \lambda\big\},
\end{aligned} \tag{D.2}$$

where $\|\boldsymbol{u}_{t+1}\|_W = \sqrt{\sum_{i\in[N]} \frac{1}{\widehat{w}_{t+1}(\boldsymbol{\xi}_t, \boldsymbol{\xi}_t^i)} \left(u_{t+1}^i\right)^2}$ is defined to be a weighted norm. We further define its dual norm $\|\boldsymbol{u}_{t+1}\|_{W^{-1}} = \sqrt{\sum_{i\in[N]} \widehat{w}_{t+1}(\boldsymbol{\xi}_t, \boldsymbol{\xi}_t^i)\ (u_{t+1}^i)^2}$. Then, we can establish the upper bound on the

problem (D.2):

$$\begin{aligned}\overline{z} + \boldsymbol{u}_{t+1}^{\top}(\boldsymbol{z} - \overline{z}\cdot\mathbf{e}) &\leq \overline{z} + \|\boldsymbol{u}_{t+1}\|_W \cdot \|\boldsymbol{z} - \overline{z}\cdot\mathbf{e}\|_{W^{-1}} \\ &\leq \overline{z} + \lambda \, \|\boldsymbol{z} - \overline{z}\cdot\mathbf{e}\|_{W^{-1}} = \overline{z} + \lambda\sqrt{s} = z_\lambda^{\mathrm{VR}}.\end{aligned} \tag{D.3}$$

Here, the first inequality follows from the Cauchy-Schwarz inequality, and the second inequality is due to the constraint $\|\boldsymbol{u}_{t+1}\|_W \leq \lambda$. Moreover, the first equality holds because of the following equality:

$$\|\boldsymbol{z} - \overline{z}\cdot\mathbf{e}\|_{W^{-1}} = \sqrt{\sum_{i\in[N]} \widehat{w}_{t+1}(\boldsymbol{\xi}_t, \boldsymbol{\xi}_t^i)\cdot(z_i - \overline{z})^2} = \sqrt{\widehat{\mathbb{V}}\left[V_{t+1}^{\mathrm{VR}}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1}) \;\middle|\; \boldsymbol{\xi}_t\right]}.$$

Therefore, (D.3) implies that $z_\lambda^{\mathrm{DR}} \leq z_\lambda^{\mathrm{VR}}$ for any value of $\lambda \geq 0$.

Now we show the equivalence when $\sqrt{s} \geq \lambda$. Consider the following solution $\boldsymbol{u}'_{t+1}$:

$$u'^{i}_{t+1} = \frac{\lambda\, \widehat{w}_{t+1}(\boldsymbol{\xi}_t, \boldsymbol{\xi}_t^i)\,(z_i - \overline{z})}{\sqrt{s}} \quad \forall i \in [N]. \tag{D.4}$$

This choice of $\boldsymbol{u}'_{t+1}$ yields an objective value that coincides with the upper bound, that is, $\overline{z} + \boldsymbol{u}'^{\top}_{t+1}(\boldsymbol{z} - \overline{z}\cdot\mathbf{e}) = z_\lambda^{\mathrm{VR}}$. This implies that the feasibility of the solution $\boldsymbol{u}'_{t+1}$ is sufficient for the equivalence between the two formulations, i.e., $z_\lambda^{\mathrm{DR}} = z_\lambda^{\mathrm{VR}}$. Regardless of the value of λ, $\boldsymbol{u}'_{t+1}$ satisfies the second and third constraints of (D.2): $\mathbf{e}^{\top}\boldsymbol{u}'_{t+1} = 0$ and $\|\boldsymbol{u}'_{t+1}\|_W^2 \leq \lambda$. Therefore, it remains feasible as long as it satisfies the first constraint:

$$u'^{i}_{t+1} = \frac{\lambda\, \widehat{w}_{t+1}(\boldsymbol{\xi}_t, \boldsymbol{\xi}_t^i)\,(z_i - \overline{z})}{\sqrt{s}} \geq -\widehat{w}_{t+1}(\boldsymbol{\xi}_t, \boldsymbol{\xi}_t^i) \quad \Longleftrightarrow \quad \frac{\lambda(z_i - \overline{z})}{\sqrt{s}} \geq -1. \tag{D.5}$$

Since the difference $|z_i - \overline{z}| = \left|V_{t+1}(\boldsymbol{x}_t, \boldsymbol{\xi}_{t+1}^i) - \sum_{n\in[N]} w_{t+1}^n\, V_{t+1}(\boldsymbol{x}_t, \boldsymbol{\xi}_{t+1}^n)\right| \leq 1$ by (**A4**), a sufficient condition for (D.5) is $s \geq \lambda^2$. Thus, if this condition holds, we have $z_\lambda^{\mathrm{DR}} = z_\lambda^{\mathrm{VR}}$. This completes the proof. □

E Proof of Proposition 3

Proof. We consider the inner maximization problem (17) for any fixed feasible solution $\boldsymbol{x}_t \in \mathcal{X}_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t)$: the worst-case conditional expectation of the cost-to-go function given an arbitrary sample $\boldsymbol{\xi}_t \in \Xi_t$ is

$$\max_{\mathbb{P}_{t+1}\in\mathscr{P}_{t+1}^{\lambda}(\boldsymbol{\xi}_t)} \mathbb{E}_{\mathbb{P}_{t+1}}\left[\widehat{Q}_{t+1}^{\mathrm{DR}}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1})\right] \tag{E.1}$$

using the polyhedral ambiguity set $\mathscr{P}_{t+1}^{\lambda}(\boldsymbol{\xi}_t)$. To simplify the notation, we define the vector $\boldsymbol{z} \in \mathbb{R}^N$ whose i-th component $z_i = V_{t+1}^{\mathrm{DR}}(\boldsymbol{x}_t, \boldsymbol{\xi}_{t+1}^i)$ denotes the cost-to-go function evaluated at the sample $\boldsymbol{\xi}_{t+1}^i$ and feasible solution $\boldsymbol{x}_t$. Then we can rewrite (E.1) as the following primal problem

$$\begin{array}{lll} \max & \boldsymbol{w}_{t+1}^{\top}\boldsymbol{z} & \\ \text{s.t.} & \boldsymbol{w}_{t+1} \in \mathbb{R}_+^N & \\ & \mathbf{e}^{\top}\boldsymbol{w}_{t+1} = 1 & : (\gamma), \\ & \left(\left(\dfrac{w_{t+1}^i - \widehat{w}_{t+1}(\boldsymbol{\xi}_t, \boldsymbol{\xi}_t^i)}{\sqrt{\widehat{w}_{t+1}(\boldsymbol{\xi}_t, \boldsymbol{\xi}_t^i)}}\right)_{i\in[N]}, \lambda\right) \in \mathscr{C}^N & : (\boldsymbol{\phi}, \psi), \end{array} \tag{E.2}$$

where $\gamma \in \mathbb{R}$ and $(\boldsymbol{\zeta}, \rho) \in (\mathscr{C}^N)^*$ are dual variables associated with the constraints. Strong duality holds because $\mathscr{C}^N$ is a polyhedral cone and the conic constraint is feasible for $\lambda \geq 0$. Therefore, the optimal value of the dual problem

$$\begin{aligned} \min \quad & \gamma + \lambda\rho - \sum_{i\in[N]} \sqrt{\widehat{w}_{t+1}(\boldsymbol{\xi}_t, \boldsymbol{\xi}_t^i)}\zeta_i \\ \text{s.t.} \quad & (\boldsymbol{\zeta}, \rho) \in (\mathscr{C}^N)^* \\ & z_i + \frac{\zeta_i}{\sqrt{\widehat{w}_{t+1}(\boldsymbol{\xi}_t, \boldsymbol{\xi}_t^i)}} \leq \gamma \qquad\qquad \forall i \in [N] \end{aligned} \tag{E.3}$$

coincides with that of the primal problem (E.2). Then, given that the inner maximization is rewritten as the minimization problem (E.3), rather than fixing $\boldsymbol{x}_t$ as we did in (E.1), we can combine it with the outer minimization problem as in (18). This completes the proof. □

F Proof of Lemma 3

Proof. Stage $t = T$: For any $i \in [N]$, due to (**A3**) and strong duality, we have

$$\begin{aligned} V_T^{\mathrm{DR}}(\boldsymbol{x}_{T-1}, \boldsymbol{\xi}_T^i) &= \min_{\boldsymbol{x}_T} \left\{ \boldsymbol{c}_T^{i\top} \boldsymbol{x}_T \;:\; \boldsymbol{x}_T \in \mathbb{R}_+^{d_T},\ \boldsymbol{A}_T^i \boldsymbol{x}_T + \boldsymbol{B}_T^i \boldsymbol{x}_{T-1} = \boldsymbol{b}_T^i \right\} \\ &= \max_{\boldsymbol{\pi}_T} \left\{ \boldsymbol{\pi}_T^\top (\boldsymbol{b}_T^i - \boldsymbol{B}_T^i \boldsymbol{x}_{T-1}) \;:\; \boldsymbol{A}_T^{i\top} \boldsymbol{\pi}_T \leq \boldsymbol{c}_T^i \right\} \\ &= \max \left\{ \boldsymbol{\pi}_{T,l}^{i\top} (\boldsymbol{b}_T^i - \boldsymbol{B}_T^i \boldsymbol{x}_{T-1}) \;:\; l \in [\mathcal{S}_T^i] \right\}, \end{aligned} \tag{F.1}$$

where $\mathcal{S}_T^i$ denotes the number of extreme points of the dual feasible region $\{\boldsymbol{\pi}_T : \boldsymbol{A}_T^{i\top} \boldsymbol{\pi}_T \leq \boldsymbol{c}_T^i\}$, and for each $l \in [\mathcal{S}_T^i]$, $\boldsymbol{\pi}_{T,l}^i$ denotes the l-th extreme point. Note that (F.1) represents the maximum of $|\mathcal{S}_T^i|$ affine functions of $\boldsymbol{x}_{T-1}$. Since $\mathcal{S}_T^i$ is finite for all $i \in [N]$, the claim holds for stage T.

Stage $t < T$: By our induction hypothesis, suppose that $V_{t+1}^{\mathrm{DR}}(\boldsymbol{x}_t, \boldsymbol{\xi}_{t+1}^i)$ is piecewise linear convex with $\mathcal{S}_{t+1}^i$ pieces, i.e., $V_{t+1}^{\mathrm{DR}}(\boldsymbol{x}_t, \boldsymbol{\xi}_{t+1}^i) = \max\{\boldsymbol{\alpha}_{t+1,k}^{i\top} \boldsymbol{x}_t + \beta_{t,k}^i : k \in [\mathcal{S}_{t+1}^i]\}$. Then, $V_t^{\mathrm{DR}}(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t^i)$ can be written identical to (19) with the index set $[\bar{\mathcal{S}}_{t+1}^{i,k}]$ in the last constraint being replaced by $[\mathcal{S}_{t+1}^i]$. Then, the corresponding dual problem is derived as follows:

$$\begin{aligned} V_t^{\mathrm{DR}}(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t^i) = \max\ & \boldsymbol{\pi}_t^\top (\boldsymbol{b}_t^i - \boldsymbol{B}_t^i \boldsymbol{x}_{t-1}) + \sum_{j\in[N]} \sum_{k\in[\mathcal{S}_{t+1}^j]} y_{t,k}^j \beta_{t+1,k}^j \\ \text{s.t.} \quad & \boldsymbol{\pi}_t \in \mathbb{R}^{\dim(\boldsymbol{b}_t^i)}, \quad \boldsymbol{r}_t \in \mathbb{R}^N, \quad y_{t,k}^j \in \mathbb{R}_+ \quad \forall j \in [N],\ \forall k \in [\mathcal{S}_{t+1}^j], \\ & \boldsymbol{A}_t^{i\top} \boldsymbol{\pi}_t \leq \boldsymbol{c}_t^i + \sum_{j\in[N]} \sum_{k\in[\mathcal{S}_{t+1}^j]} y_{t,k}^j \boldsymbol{\alpha}_{t+1,k}^j, \\ & \sum_{j\in[N]} \sum_{k\in[\mathcal{S}_{t+1}^j]} y_{t,k}^j = 1, \quad \sum_{j\in[N]} r_t^j \leq \lambda\sqrt{N}, \quad r_t^j \leq \lambda \quad \forall j \in [N], \\ & \sum_{k\in[\mathcal{S}_{t+1}^j]} y_{t,k}^j \leq \widehat{w}_{t+1}(\boldsymbol{\xi}_t^i, \boldsymbol{\xi}_t^j) + \sqrt{\widehat{w}_{t+1}(\boldsymbol{\xi}_t^i, \boldsymbol{\xi}_t^j)} \cdot r_t^j \quad \forall j \in [N], \\ & \sum_{k\in[\mathcal{S}_{t+1}^j]} y_{t,k}^j \geq \widehat{w}_{t+1}(\boldsymbol{\xi}_t^i, \boldsymbol{\xi}_t^j) - \sqrt{\widehat{w}_{t+1}(\boldsymbol{\xi}_t^i, \boldsymbol{\xi}_t^j)} \cdot r_t^j \quad \forall j \in [N]. \end{aligned}$$

Here, $\boldsymbol{\pi}_t$ is a vector of dual variables associated with the primal constraints $\boldsymbol{A}_t^i \boldsymbol{x}_t + \boldsymbol{B}_t^i \boldsymbol{x}_{t-1} = \boldsymbol{b}_t^i$. Then, similar to (F.1), the dual problem is equivalently written as

$$V_t^{\mathrm{DR}}(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t^i) = \max \Big\{ \boldsymbol{\pi}_{t,l}^{i\top} (\boldsymbol{b}_t^i - \boldsymbol{B}_t^i \boldsymbol{x}_{t-1}) + \sum_{j \in [N]} \sum_{k \in [\mathcal{S}_{t+1}^j]} y_{t,k,l}^j \cdot \beta_{t+1,k}^j \ : \ l \in [\mathcal{S}_t^i] \Big\}. \tag{F.2}$$

Thus, $V_t^{\mathrm{DR}}(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t^i)$ is piecwise convex linear with $\mathcal{S}_t^i$ pieces for each $i \in [N]$. □

G Proof of Lemma 4

Proof. We proceed by backward induction. Recall that $\bar{\mathcal{S}}_t^{i,k}$ denotes the number of cuts generated through k iterations for $V_t^{\mathrm{DR}}(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}_t^i)$ as defined in (19).

Stage $t = T$: Since $V_{T+1}^{\mathrm{DR}}(\cdot) = 0$, it follows from (20) that, for any $k \geq 1$, we have

$$\begin{aligned}
\underline{V}_{T,k}^{\mathrm{DR}}(\boldsymbol{x}_{T-1}, \boldsymbol{\xi}_T^i) =& \max \Big\{ \boldsymbol{\alpha}_{T,l}^{i\top} \boldsymbol{x}_{T-1} + \beta_{T,l}^i : l \in [\bar{\mathcal{S}}_T^{i,k}] \Big\} \\
=& \max \Big\{ \boldsymbol{\pi}_{T,i,l}^{\star\top} (\boldsymbol{b}_T^i - \boldsymbol{B}_T^i \boldsymbol{x}_{T-1}) : l \in [\bar{\mathcal{S}}_T^{i,k}] \Big\} \\
\leq& \max \Big\{ \boldsymbol{\pi}_T^\top (\boldsymbol{b}_T^i - \boldsymbol{B}_T^i \boldsymbol{x}_{T-1}) : \boldsymbol{A}_T^{i\top} \boldsymbol{\pi}_T \leq \boldsymbol{c}_T^i \Big\} \\
=& \min \Big\{ \boldsymbol{c}_T^{i\top} \boldsymbol{x}_T : \boldsymbol{x}_T \in \mathbb{R}_+^{d_T}, \boldsymbol{A}_T^i \boldsymbol{x}_T + \boldsymbol{B}_T^i \boldsymbol{x}_{T-1} = \boldsymbol{b}_T^i \Big\} \\
=& V_T^{\mathrm{DR}}(\boldsymbol{x}_{T-1}, \boldsymbol{\xi}_T^i).
\end{aligned} \tag{G.1}$$

The inequality holds since $\{\boldsymbol{\pi}_{T,i,l}^\star\}_{l \in [\bar{\mathcal{S}}_T^{i,k}]}$ is a subset of the dual feasible region $\{\boldsymbol{\pi}_T : \boldsymbol{A}_T^{i\top} \boldsymbol{\pi}_T \leq \boldsymbol{c}_T^i\}$. The third equality holds due to (**A3**) and strong duality.

For the upper bound, recall the last constraint in (21), given by $\boldsymbol{y}_{T-1}^i = \boldsymbol{x}_{T-1} - \sum_{l \in [k]} \theta_l^i \boldsymbol{x}_{T-1}^l$ where $\sum_{l \in [k]} \theta_l^i \boldsymbol{x}_{T-1}^l$ is a convex combination of the previous k trial solutions. Then, due to the convexity of $V_T^{\mathrm{DR}}(\cdot)$ established in Lemma 3, for such $\boldsymbol{y}_{T-1}^i$ and $\boldsymbol{x}_{T-1}$, we have

$$V_T^{\mathrm{DR}}(\boldsymbol{x}_{T-1}, \boldsymbol{\xi}_T^i) - \boldsymbol{\phi}_T^\top \boldsymbol{y}_{T-1}^i \leq \sum_{l \in [k]} \theta_l^i \cdot V_T^{\mathrm{DR}}(\boldsymbol{x}_{T-1}^l, \boldsymbol{\xi}_T^i), \tag{G.2}$$

where $\boldsymbol{\phi}_T$ is any subgradient of $V_T^{\mathrm{DR}}(\boldsymbol{x}_{T-1}, \boldsymbol{\xi}_T^i)$ at $\boldsymbol{x}_{T-1}$. Therefore, we have

$$\begin{aligned}
V_T^{\mathrm{DR}}(\boldsymbol{x}_{T-1}, \boldsymbol{\xi}_T^i) &\leq \sum_{l \in [k]} \theta_l^i \cdot V_T^{\mathrm{DR}}(\boldsymbol{x}_{T-1}^l, \boldsymbol{\xi}_T^i) + \boldsymbol{\phi}_T^\top \boldsymbol{y}_{T-1}^i \\
&\leq \sum_{l \in [k]} \theta_l^i \cdot V_T^{\mathrm{DR}}(\boldsymbol{x}_{T-1}^l, \boldsymbol{\xi}_T^i) + \|\boldsymbol{\phi}_T^\top\|_2 \|\boldsymbol{y}_{T-1}^i\|_2 \\
&\leq \sum_{l \in [k]} \theta_l^i \cdot V_T^{\mathrm{DR}}(\boldsymbol{x}_{T-1}^l, \boldsymbol{\xi}_T^i) + M_T \|\boldsymbol{y}_{T-1}^i\|_1 \\
&= \sum_{l \in [k]} \theta_l^i \cdot \overline{V}_{T,l}^{\mathrm{DR}}(\boldsymbol{x}_{T-1}^l, \boldsymbol{\xi}_T^i) + M_T \|\boldsymbol{y}_{T-1}^i\|_1 = \overline{V}_{T,k}^{\mathrm{DR}}(\boldsymbol{x}_{T-1}, \boldsymbol{\xi}_T^i).
\end{aligned} \tag{G.3}$$

The first inequality comes from (G.2). The second inequality follows due to the Cauchy-Schwarz inequality and the fact that $\|\boldsymbol{y}\|_2 \leq \|\boldsymbol{y}\|_1$ for any vector $\boldsymbol{y}$. The last inequality holds because $M_T \geq L_T^{\mathrm{DR}}$ by our assumption. The first equality holds because $V_{T+1}^{\mathrm{DR}}(\cdot) = 0$, implying that the terminal-stage upper bound

evaluated at any feasible $\boldsymbol{x}_{T-1}$ is tight; that is, $\overline{V}^{\mathrm{DR}}_{T,l}(\boldsymbol{x}^l_{T-1}, \boldsymbol{\xi}^i_T) = V^{\mathrm{DR}}_T(\boldsymbol{x}^l_{T-1}, \boldsymbol{\xi}^i_T)$ for all $l \in [k]$. The last equality follows directly from the second-to-last constraint in (21). Therefore, (22) holds for $t = T$.

Stage $t < T$: By induction, suppose that (22) holds for $t+1 \leq T$, i.e., for any $k \geq 1$

$$\underline{V}^{\mathrm{DR}}_{t+1,k}(\boldsymbol{x}_t, \boldsymbol{\xi}^i_{t+1}) = \max\left\{{\boldsymbol{\alpha}^i_{t+1,l}}^\top \boldsymbol{x}_t + \beta^i_{t+1,l} : l \in [\bar{\mathcal{S}}^{i,k}_{t+1}]\right\} \leq V^{\mathrm{DR}}_{t+1}(\boldsymbol{x}_t, \boldsymbol{\xi}^i_{t+1}). \tag{G.4}$$

Note that the only difference between $V^{\mathrm{DR}}_t(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}^i_t)$ in (18) and $\underline{V}^{\mathrm{DR}}_{t,k}(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}^i_t)$ in (19) is the index set: the former uses $[\mathcal{S}^i_{t+1}]$, while the latter uses $[\bar{\mathcal{S}}^{i,k}_{t+1}]$. Since (G.4) implies that the feasible region of (19) is a subset of that of (18), for any $k \geq 1$, $\underline{V}^{\mathrm{DR}}_{t,k}(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}^i_t)$ provides a valid lower bound:

$$\underline{V}^{\mathrm{DR}}_{t,k}(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}^i_t) \leq V^{\mathrm{DR}}_t\left(\boldsymbol{x}_{t-1}, \boldsymbol{\xi}^i_t\right) \quad \forall \text{ feasible } \boldsymbol{x}_{t-1}\ \forall i \in [N]. \tag{G.5}$$

For the upper bound, by hypothesis, we have $V^{\mathrm{DR}}_{t+1}(\cdot) \leq \overline{V}^{\mathrm{DR}}_{t+1,k}(\cdot)$. Then, analogous to (G.3), due to the convexity of $V^{\mathrm{DR}}_t(\cdot)$ and $M_t \geq L^{\mathrm{DR}}_t$ by our assumption, we have, for any $k \geq 1$

$$V^{\mathrm{DR}}_{t+1}(\boldsymbol{x}_t, \boldsymbol{\xi}^i_{t+1}) \leq \overline{V}^{\mathrm{DR}}_{t+1,k}(\boldsymbol{x}_t, \boldsymbol{\xi}^i_{t+1}) \quad \forall \text{ feasible } \boldsymbol{x}_{t-1}\ \forall i \in [N]. \tag{G.6}$$

Combining (G.5) and (G.6), we show (22) holds for t. This completes the induction. □

H Proof of Theorem 2

Proof. We present the lower bound convergence by following the analysis of the standard SDDP algorithm in [44]. We then proceed to the upper bound convergence.

Define the set of all scenarios in the scenario tree in Figure 1 as

$$\Xi_{[T]} = \left\{(\boldsymbol{\xi}_1, \boldsymbol{\xi}^{i_2}_2, \ldots, \boldsymbol{\xi}^{i_T}_T) : i_t \in [N] \text{ for all } t \in \{2, \ldots, T\}\right\},$$

where each scenario in $\Xi_{[T]}$ is denoted by $\boldsymbol{\xi}_{[T]} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}^{i_2}_2, \ldots, \boldsymbol{\xi}^{i_T}_T)$. Similarly, let $\Xi_{[t]} = \{(\boldsymbol{\xi}_1, \boldsymbol{\xi}^{i_2}_2, \ldots, \boldsymbol{\xi}^{i_t}_t) : i_t \in [N] \text{ for all } t \in \{2, \ldots, t\}\}$ denote the set of *scenario sequences* up to stage t with each sequence denoted by $\boldsymbol{\xi}_{[t]} \in \Xi_{[t]}$.

Note that the total number of scenarios, i.e., $|\Xi_{[T]}|$, is finite. Moreover, since the transition probability $\widehat{w}_{t+1}(\boldsymbol{\xi}^i_t, \boldsymbol{\xi}^j_t)$ (in (3)) computed using the exponential kernel are strictly positive for all $i, j \in [N]$ and all $t \in [T-1]$, every scenario $\boldsymbol{\xi}_{[T]} \in \Xi_{[T]}$ has a nonzero probability of being sampled during the forward pass. The same holds for any sequence $\boldsymbol{\xi}_{[t]} \in \Xi_{[t]}$. Also, as shown in [44], under (**A7**), only a finite number of distinct lower bounds can be generated at stage T. By backward induction, the number of lower bounds that can be generated at any stage is finite, since only finitely many distinct lower bounds can be generated at the subsequent stage.

By the Bellman optimality condition, a given policy $\boldsymbol{x}_t(\boldsymbol{\xi}_{[t]})$ is optimal if the following holds: for all $t \in [T]$, and $\boldsymbol{\xi}_{[t]} \in \Xi_{[t]}$, we have

$$\boldsymbol{x}_t(\boldsymbol{\xi}_{[t]}) \in \underset{\boldsymbol{x}_t \in \mathcal{X}_t(\boldsymbol{x}_{t-1}(\boldsymbol{\xi}_{[t-1]}), \boldsymbol{\xi}_t)}{\arg\min} \left\{\boldsymbol{c}_t^\top \boldsymbol{x}_t + \max_{\mathbb{P}_{t+1} \in \mathcal{P}^\lambda_{t+1}(\boldsymbol{\xi}_t)} \mathbb{E}_{\mathbb{P}_{t+1}}\left[V^{\mathrm{DR}}_{t+1}(\boldsymbol{x}_t, \tilde{\boldsymbol{\xi}}_{t+1}) \middle| \boldsymbol{\xi}_t\right]\right\}. \tag{H.1}$$

Here, $\boldsymbol{x}_{t-1}(\boldsymbol{\xi}_{[t-1]})$ is a policy at the previous stage given $\boldsymbol{\xi}_{[t-1]}$. Recall that the DRO formulation in (H.1) is equivalent to the linear program (18) (here, we use (H.1) to save space).

For all $t \in [T]$ and scenario sequence $\boldsymbol{\xi}_{[t]} \in \Xi_{[t]}$, let $\boldsymbol{x}_t^k(\boldsymbol{\xi}_{[t]})$ denote the current policy obtained from the lower bound $\underline{V}_{t,k}^{\mathrm{DR}}(\cdot)$ at the k-th iteration. Suppose the optimality condition (H.1) does not hold for some sequence $\boldsymbol{\xi}_{[t]}$ at some stage $t \in [T]$. This implies that the current policy is suboptimal. Let $t' \leq T$ be the largest such stage at which the policy $\boldsymbol{x}_{t'}^k(\boldsymbol{\xi}_{[t']})$ violates (H.1), indicating that the lower bound can be improved at the trial solution $\boldsymbol{x}_{t'}^k = \boldsymbol{x}_{t'}^k(\boldsymbol{\xi}_{[t']})$. Therefore, during the subsequent backward pass, a new cut is added at $\boldsymbol{x}_{t'}^k$.

Then, as mentioned earlier, since each sequence $\boldsymbol{\xi}_{[t]} \in \Xi_{[t]}$ can be sampled with nonzero probability (during the forward pass), and only finitely many distinct lower bounds can be generated (during the backward pass), the optimality condition (H.1) will eventually hold for all scenario sequence $\boldsymbol{\xi}_{[t']} \in \Xi_{[t']}$ at stage t'. As the condition is first satisfied at later stages and then propagates backward, it ultimately holds at the initial stage, ensuring convergence in finite iterations with probability one.

We now turn to the convergence of the upper bound. Let $\underline{k}^\star$ denote the iteration after which the lower bound has converged, i.e., optimality of the approximate MSLP problem is achieved. This implies that, for any iteration $k \geq \underline{k}^\star$, no further updates are made to the lower bound during the backward pass. Under (**A7**), it further follows that during the forward pass at any iteration $k \geq \underline{k}^\star$, only finitely many distinct trial solutions $(\boldsymbol{x}_1^k, \ldots, \boldsymbol{x}_{T-1}^k)$ can be generated for any scenario $\boldsymbol{\xi}_{[T-1]} \in \Xi_{[T-1]}$, as each trial solution $\boldsymbol{x}_t^k = \boldsymbol{x}_t^k(\boldsymbol{\xi}_{[t]})$ corresponds to an extreme point for each $t \in [T-1]$.

Without loss of generality, suppose that, after the lower bound convergence, each scenario $\boldsymbol{\xi}_{[T]} \in \Xi_{[T]}$ yields a *unique* trajectory of optimal extreme points—otherwise, there are still only finitely many such trajectories. We now show that the gap between the upper and lower bounds closes in finitely many iterations. Specifically, at each stage $t \in [T]$, there exists an iteration $k \geq \underline{k}^\star$ such that, for any $\boldsymbol{\xi}_{[t]} \in \Xi_{[t]}$, we have

$$\overline{V}_{t,k}^{\mathrm{DR}}(\boldsymbol{x}_{t-1}^k(\boldsymbol{\xi}_{[t-1]}), \boldsymbol{\xi}_t) = \underline{V}_{t,k}^{\mathrm{DR}}(\boldsymbol{x}_{t-1}^k(\boldsymbol{\xi}_{[t-1]}), \boldsymbol{\xi}_t). \tag{H.2}$$

Let $\overline{k}_t$ be the iteration after which (H.2) holds at stage t. At the terminal stage T, (H.2) holds when the upper bound is evaluated at all trial solutions $\boldsymbol{x}_T$ corresponding to each scenario $\boldsymbol{\xi}_{[T]} \in \Xi_{[T]}$, since $\overline{V}_{T+1,k}^{\mathrm{DR}}(\cdot) = \underline{V}_{T+1,k}^{\mathrm{DR}}(\cdot) = 0$ for all $k \geq 0$. Thus, after finitely many additional iterations beyond $\overline{k}_T$, (H.2) also holds at stage $T-1$, once all corresponding trial solutions $\boldsymbol{x}_{T-1}$ are used to evaluate the upper bound. by induction, there exists $k \geq \overline{k}_t$ such that (H.2) holds at stage t since (H.2) holds at the subsequent stage. This completes the proof for the upper bound convergence. □