Periodical Multistage Stochastic Programs

Alexander Shapiro*and **Lingquan Ding**

School of Industrial and Systems Engineering,
Georgia Institute of Technology,
Atlanta, Georgia 30332-0205.

Abstract. In some applications the considered multistage stochastic programs have a periodical behavior. We show that in such cases it is possible to drastically reduce the number of stages by introducing a periodical analog of the so-called Bellman equations for discounted infinite horizon problems, used in Markov Decision Processes and Stochastic Optimal Control. Furthermore, we describe a variant of the Stochastic Dual Dynamic Programming algorithm, applied to the constructed periodical Bellman equations, and provide numerical experiments for the Brazilian interconnected power system problem.

Key Words: multistage programs, decision rules, dynamic programming, Bellman equations, SDDP algorithm, fixed point theorem

*This research was partly supported by NSF grant 1633196.

1 Introduction

There are many practical problems where one has to make decisions sequentially based on data (observations) available at time of the decision. Trying to make such decisions under uncertainty in some optimal way, looking forward in time, leads to the area of multistage optimization. Traditionally the uncertainty in such problems was modeled as randomness using tools of probability theory. We can mention the following areas of research dealing with such optimization problems: Markov Decision Processes (MDP) (e.g., [10]), Stochastic Optimal Control (SOC) (e.g., [3]), and Stochastic Programming (SP). The main tools, used both in MDP and SOC, are based on utilising dynamic programming equations. Until recently the dominant approach in SP was based on generating scenarios. Since the number of scenarios needed to approximate the considered data process grows exponentially with increase of the number of stages, the scenario approach could be computationally infeasible even for a moderate number of stages (cf., [14]).

On the other hand the dynamic programming approach suffers from the so-called "curse of dimensionality", the term coined by Bellman already in the early work on dynamic programming. This motivated development of approximate approaches to dynamic programming. One type of algorithms, which can be classified as approximate dynamic programming, is the so-called Stochastic Dual Dynamic Programming (SDDP) method; it was introduced in [8] and became quite popular for solving multistage convex stochastic programs. The SDDP method is based on approximation of value (cost-to-go) functions of the dynamic programming equations by piecewise linear functions and can be viewed as a dynamic variant of the cutting plane method. It was demonstrated numerically that in some cases the SDDP method could produce an acceptably accurate solution even when the number of stages is large.

In some applications the considered multistage stochastic programs have a periodical behavior. *The main result of this paper is to demonstrate that in such cases it is possible to drastically reduce the number of stages by introducing a periodical analogue of the so-called Bellman equations, used in MDP and SOC, and consequently applying a variant of the SDDP algorithm.* The paper is organized as follows. In the next section we give a general formulation of infinite horizon risk averse stochastic programming problems. The main theoretical developments are presented in section 3 where we introduce Bellman equations adjusted to the periodical behavior of the considered infinite horizon problems. In section 4 we briefly discuss statistical inference of the Sample Average Approximation (SAA) approach to discretization of possibly continuous distributions of the data process. An SDDP type algorithm for solving the obtained periodical dynamical equations is described in section 5. In section 6 we provide numerical experiments with the proposed variant of SDDP algorithm applied to the Brazilian interconnected power system (cf., [15]). Section 7 gives a brief discussion of a possible extension to a Markovian setting, and section 8 is devoted to conclusion remarks.

2 Multistage programming

Consider the following risk averse multistage stochastic programming problem, given in the nested form,

$$\min_{\substack{A_1x_1=b_1\\ x_1\in\mathcal{X}_1}} f_1(x_1) \;+\; \gamma\rho_{|\xi_1}\Big(\min_{\substack{B_2x_1+A_2x_2=b_2\\ x_2\in\mathcal{X}_2}} f_2(x_2,\xi_2)+\cdots \\ +\gamma\rho_{|\xi_{[T-1]}}\big(\min_{\substack{B_Tx_{T-1}+A_Tx_T=b_T\\ x_T\in\mathcal{X}_T}} f_T(x_T,\xi_T)\big)\Big). \tag{2.1}$$

Here $\gamma\in(0,1)$ is the discount coefficient, $\xi_t\in\mathbb{R}^d$, $t=1,...,T$, is a random data process, $f_t:\mathbb{R}^n\times\mathbb{R}^d\to\mathbb{R}$ are cost functions, $b_t=b_t(\xi_t)$ are right side vectors, $B_t=B_t(\xi_t)$ and $A_t=A_t(\xi_t)$ are matrices of appropriate dimensions and $\mathcal{X}_t\subset\mathbb{R}^n$ are closed sets. The number of stages T can be finite or it can be that $T=\infty$. In order to make formulation (2.1) precise, certain regularity conditions should be introduced, we will discuss this later. We denote by $\xi_{[t]}:=(\xi_1,...,\xi_t)$ the history of the data process up to time t. The conditional functionals $\rho_{|\xi_{[t]}}$ are supposed to be conditional counterparts of a law invariant coherent risk measure $\varrho:\mathcal{Z}\to\mathbb{R}$, defined on a linear space $\mathcal{Z}$ of random variables. With some abuse of the notation, sometimes we view ξ_t as a variable in $\mathbb{R}^d$ rather than a random vector, the particular meaning will be clear from the context.

It is said that $\varrho:\mathcal{Z}\to\mathbb{R}$ is a coherent risk measure if it is subadditive (i.e., if $Z,Z'\in\mathcal{Z}$, then $\varrho(Z+Z')\le\varrho(Z)+\varrho(Z')$), monotone (i.e., if $Z,Z'\in\mathcal{Z}$ are such that $Z\ge Z'$ a.s., then $\varrho(Z)\ge\rho(Z')$), translation equivariant (i.e., if $Z\in\mathcal{Z}$ and $a\in\mathbb{R}$, then $\varrho(Z+a)=\varrho(Z)+a$), and positively homogeneous (i.e., if $Z\in\mathcal{Z}$ and $t>0$, then $\varrho(tZ)=t\varrho(Z)$). It is said that ϱ is law invariant if $\varrho(Z)=\varrho(Z')$ for any distributionally equivalent (with respect to the reference probability measure) random variables $Z,Z'\in\mathcal{Z}$. Axioms of coherent risk measures were introduced in [1], we refer to [5] and [13] for a general discussion of these concepts. When $\varrho=\mathbb{E}$ is the expectation operator, conditional risk measures $\rho_{|\xi_{[t]}}$ become the respective conditional expectations $\mathbb{E}_{|\xi_{[t]}}$. In that case we refer to the problem (2.1) as *risk neutral*.

It is assumed that the first stage data vector ξ_1 is known (deterministic), and hence $\rho_{|\xi_1}=\varrho$. Optimization in (2.1) is performed over feasible policies (also called decision rules). A policy is a sequence of (measurable) functions $x_t(\xi_{[t]})$, $t=1,...,T$. Each $x_t(\xi_{[t]})$ is a function of the data process up to time t, this ensures the *nonanticipative* property of a considered policy. The first stage decision x_1 is deterministic, i.e., does not depend on realizations of random data. The constraints should be satisfied for almost every realization of the random data process. We denote by Π the set of policies $\pi=\big(x_1,x_2(\xi_{[2]}),...,x_T(\xi_{[T]})\big)$ satisfying the feasibility constraints and such that the corresponding conditional risk measures are well defined. Sometimes we write x_t^π for $x_t=x_t(\xi_{[t]})$ to emphasize dependence on the considered policy. It is assumed that the probability distribution of the data process $\xi_1,...,\xi_T$ does *not* depend on considered policies $\pi\in\Pi$. The constraints $B_tx_{t-1}+A_tx_t=b_t$ define dynamics of the system and often represent balance equations.

Assuming that the risk measure ϱ is law invariant and coherent, for finite time horizon T, problem (2.1) can be written as

$$\min_{\pi\in\Pi} f_1(x_1^\pi)+\mathcal{R}_T\left(\textstyle\sum_{t=2}^T\gamma^{t-1}f_t(x_t^\pi,\xi_t)\right), \tag{2.2}$$

where $B_1 = 0$ (and hence the term $B_1 x_0$ at the first stage is omitted), $f_1(x_1, \xi_1) = f_1(x_1)$, and $\mathcal{R}_T$ is the respective nested risk measure $\mathcal{R}_T(\cdot) := \rho_{|\xi_1}\big(\cdots \rho_{|\xi_{[T-1]}}(\cdot)\big)$. That is, for a given policy $\pi \in \Pi$, $Z_t := f_t(x_t(\xi_{[t]}), \xi_t)$ and $S_T := \sum_{t=2}^{T} \gamma^{t-1} Z_t$ we have that

$$\mathcal{R}_T(S_T) = \gamma \rho_{|\xi_1}\big(Z_2 + \cdots + \gamma \rho_{|\xi_{[T-1]}}(Z_T)\big).$$

We are going to deal with limit risk measures as $T \to \infty$. In order to proceed we need to impose certain boundedness conditions. Suppose that random variables Z_t are bounded, say $|Z_t| \leq \kappa$ a.s. for all t. Then

$$\begin{aligned} \mathcal{R}_{T+1}(S_{T+1}) &= \gamma \rho_{|\xi_1}\big(Z_2 + \cdots + \gamma \rho_{|\xi_{[T-1]}}(Z_T + \gamma \rho_{|\xi_{[T]}}(Z_{T+1}))\big) \\ &\leq \gamma \rho_{|\xi_1}\big(Z_2 + \cdots + \gamma \rho_{|\xi_{[T-1]}}(Z_T + \gamma \kappa)\big) \leq \mathcal{R}_T(S_T) + \gamma^T \kappa. \end{aligned}$$

Also

$$\begin{aligned} \mathcal{R}_T(S_T) &= \gamma \rho_{|\xi_1}\big(Z_2 + \cdots + \gamma \rho_{|\xi_{[T-1]}}(Z_T + \gamma \rho_{|\xi_{[T]}}(Z_{T+1}) - \gamma \rho_{|\xi_{[T]}}(Z_{T+1}))\big) \\ &\geq \gamma \rho_{|\xi_1}\big(Z_2 + \cdots + \gamma \rho_{|\xi_{[T-1]}}(Z_T + \gamma \rho_{|\xi_{[T]}}(Z_{T+1}))\big) - \gamma^T \kappa, \end{aligned}$$

and hence $|\mathcal{R}_{T+1}(S_{T+1}) - \mathcal{R}_T(S_T)| \leq \gamma^T \kappa$. And so on for $T < T'$ we have that

$$\big|\mathcal{R}_{T'}(S_{T'}) - \mathcal{R}_T(S_T)\big| \leq \kappa \gamma^T \sum_{k=0}^{\infty} \gamma^k = \frac{\kappa \gamma^T}{1-\gamma}, \tag{2.3}$$

and hence $\mathcal{R}_2(S_2), ...,$ is a Cauchy sequence. Letting $T \to \infty$ we have then that the following limit does exist

$$\mathcal{R}_\infty(S_\infty) := \lim_{T \to \infty} \rho_{|\xi_1}\big(\cdots \rho_{|\xi_{[T-1]}}(S_T)\big). \tag{2.4}$$

In particular if $\varrho := \mathbb{E}$ is given by the expectation operator, with $\rho_{|\xi_{[t]}} = \mathbb{E}_{|\xi_{[t]}}$ being the corresponding conditional expectations, then $\mathcal{R}_T$ is also the respective expectation operator. In that case we refer to (2.2) as the risk neutral formulation of the multistage program.

- Unless stated otherwise we assume that the process ξ_t is stagewise independent, i.e., for $t = 1, ...,$ random vector ξ_{t+1} is independent of $\xi_{[t]}$.

When the number T of stages is finite, the nested formulation (2.1) leads to the following dynamic programming equations (cf., [13]). At stage $t = T$ the value (cost-to-go) function $V_T(x_{T-1}, \xi_T)$ is given by the optimal value of the problem

$$\begin{array}{ll} \min\limits_{x_T \in \mathcal{X}_T} & f_T(x_T, \xi_T) \\ \text{s.t.} & B_T x_{T-1} + A_T x_T = b_T. \end{array} \tag{2.5}$$

Note that variable x_T in (2.5) is viewed now as a deterministic vector $x_T \in \mathbb{R}^{n_T}$ and should satisfy the corresponding feasibility constraints. At stages $t = T-1, ..., 2$, the value function $V_t(x_{t-1}, \xi_t)$ is given by the optimal value of the problem

$$\begin{array}{ll} \min\limits_{x_t \in \mathcal{X}_t} & f_t(x_t, \xi_t) + \gamma \mathcal{V}_{t+1}(x_t) \\ \text{s.t.} & B_t x_{t-1} + A_t x_t = b_t, \end{array} \tag{2.6}$$

with

$$\mathcal{V}_{t+1}(x_t) = \varrho\big(V_{t+1}(x_t, \xi_{t+1})\big). \tag{2.7}$$

At the first stage the following problem should be solved

$$\begin{array}{ll} \min\limits_{x_1 \in \mathcal{X}_1} & f_1(x_1) + \gamma \mathcal{V}_2(x_1) \\ \text{s.t.} & A_1 x_1 = b_1. \end{array} \tag{2.8}$$

3 Infinite horizon periodical setting

Consider the infinite horizon setting (i.e., $T = \infty$) where problem (2.2) has the following periodical behavior with period $\mathsf{m} \in \mathbb{N}$. That is, let us make the following assumptions leading to stationary solutions.

(A1) The random vectors ξ_t and $\xi_{t+\mathsf{m}}$ have the same distribution, with support $\Xi \subset \mathbb{R}^d$, for $t \geq 2$ (recall that ξ_1 is deterministic).

(A2) The functions $b_t(\cdot)$, $B_t(\cdot)$, $A_t(\cdot)$ and $f_t(\cdot, \cdot)$ have period m, i.e., are the same for $t = \tau$ and $t = \tau + \mathsf{m}$, $t = 2, ...$, and the sets $\mathcal{X}_t$ are nonempty and $\mathcal{X}_t = \mathcal{X}_{t+\mathsf{m}}$ for all t.

(A3) The functional $\varrho : \mathcal{Z} \to \mathbb{R}$ is a law invariant coherent risk measure, defined on space $\mathcal{Z} = L_p(\Xi, \mathcal{B}, P)$ for some $p \in [1, \infty]$.

(A4) The functions $f_t(\cdot, \cdot)$ are bounded, i.e., there is $\kappa > 0$ such that $|f_t(x_t, \xi_t)| \leq \kappa$ for all $(x_t, \xi_t) \in \mathcal{X}_t \times \Xi$ and all t.

(A5) For every $x_{t-1} \in \mathcal{X}_{t-1}$ the set $\{x_t \in \mathcal{X}_t : B_t(\xi_t)x_{t-1} + A_t(\xi_t)x_t = b_t(\xi_t)\}$ is nonempty for a.e. $\xi_t \in \Xi$ and $t \geq 2$.

In assumption (A3), $\mathcal{B}$ is the Borel sigma algebra of subsets of Ξ and P is a probability measure on $(\Xi, \mathcal{B})$, referred to as the reference measure. Unless stated otherwise all probabilistic statements are made with respect to the reference measure P. In particular, we sometimes write $\mathbb{E}_P$ to emphasize that the expectation is taken with respect to the distribution P. Assumption (A5) usually is referred to as the relatively complete recourse.

Under these assumptions the value functions $\mathcal{V}_t(\cdot)$ and $\mathcal{V}_{t+\mathsf{m}}(\cdot)$ of dynamic equations (2.6) - (2.7) are the same for all $t \geq 2$. This leads to the following periodical variant of Wald-Bellman (WB) equations (often simply referred to as Bellman equations in Optimal Control and MDP) for the value functions $\mathcal{V}_2(\cdot), ..., \mathcal{V}_{\mathsf{m}+1}(\cdot)$,

$$\mathcal{V}_\tau(x_{\tau-1}) = \varrho(V_\tau(x_{\tau-1}, \xi_\tau)), \tag{3.1}$$

with

$$V_\tau(x_{\tau-1}, \xi_\tau) = \inf_{x_\tau \in \mathcal{X}_\tau} \left\{ f_\tau(x_\tau, \xi_\tau) + \gamma \mathcal{V}_{\tau+1}(x_\tau) : B_\tau x_{\tau-1} + A_\tau x_\tau = b_\tau \right\}, \tag{3.2}$$

for $\tau = 2, ..., \mathsf{m}+1$, and $\mathcal{V}_{\mathsf{m}+2}$ replaced by $\mathcal{V}_2$ for $\tau = \mathsf{m}+1$ (note that ξ_τ in (3.1) is viewed as a *random* vector, while in (3.2) it is treated as a vector valued variable). Consequently for $t \geq \mathsf{m}+2$ the corresponding value functions are defined recursively as $V_t(\cdot, \xi_t) = V_{t-\mathsf{m}}(\cdot, \xi_t)$, and hence $\mathcal{V}_t(\cdot) = \mathcal{V}_{t-\mathsf{m}}(\cdot)$. At the first stage problem (2.8) should be solved.

In order to show that equations (3.1) - (3.2) have a solution we proceed as follows (cf., [3]). Let $\mathbb{B}(\mathcal{X}_t)$ be the space of bounded functions $g : \mathcal{X}_t \to \mathbb{R}$. Note that $\mathbb{B}(\mathcal{X}_t)$, equipped with the sup-norm $\|g\| = \sup_{x\in\mathcal{X}_t} |g(x)|$, is a Banach space. Consider the space $\mathfrak{B} := \mathbb{B}(\mathcal{X}_1) \times \cdots \times \mathbb{B}(\mathcal{X}_{\mathsf{m}})$ equipped with the max-norm $\|\mathfrak{g}\|_{\mathfrak{B}} := \max\{\|g_2\|, ..., \|g_{\mathsf{m}+1}\|\}$, for $\mathfrak{g} = (g_2, ..., g_{\mathsf{m}+1}) \in \mathfrak{B}$ composed from functions $g_t : \mathcal{X}_{t-1} \to \mathbb{R}$. Moreover consider mapping $\mathfrak{T} : \mathfrak{B} \to \mathfrak{B}$ defined for $\mathsf{m} \geq 2$ as

$$\mathfrak{T}(\mathfrak{g})(\mathfrak{x}) := \left(\varrho(\Psi_{g_3}(x_1, \xi_2)), ..., \varrho(\Psi_{g_{\mathsf{m}+1}}(x_{\mathsf{m}-1}, \xi_{\mathsf{m}})), \varrho(\Psi_{g_2}(x_{\mathsf{m}}, \xi_{\mathsf{m}+1}))\right), \tag{3.3}$$

where $\mathfrak{x} = (x_1, ..., x_{\mathsf{m}}) \in \mathfrak{X}$, with $\mathfrak{X} := \mathcal{X}_1 \times \cdots \times \mathcal{X}_{\mathsf{m}}$, and

$$\Psi_{g_{\tau+1}}(x_{\tau-1}, \xi_\tau) := \inf_{x_\tau \in \mathcal{X}_\tau} \{f_\tau(x_\tau, \xi_\tau) + \gamma g_{\tau+1}(x_\tau) : B_\tau x_{\tau-1} + A_\tau x_\tau = b_\tau\}, \tag{3.4}$$

for $\tau = 2, ..., \mathsf{m}$, and $g_{\mathsf{m}+2}(\cdot) = g_2(\cdot)$ for $\tau = \mathsf{m}+1$. The case of $\mathsf{m} = 1$ is discussed in Remark 3.1 below.

Assumptions A(4) and A(5) imply that for any $g_{\tau+1} \in \mathbb{B}(\mathcal{X}_\tau)$ the functions $\Psi_{g_{\tau+1}}(x_{\tau-1}, \xi_\tau)$ are bounded. In order to verify that $\varrho\left(\Psi_{g_{\tau+1}}(x_{\tau-1}, \xi_\tau)\right)$ are well defined we still need to verify that $\Psi_{g_{\tau+1}}(x_{\tau-1}, \cdot)$ are measurable. We implicitly assume this. When the set Ξ is finite this holds automatically. In general we refer to [3],[11] for a discussion of such measurability issues. Then the mapping $\mathfrak{T}$ is well defined, i.e., for any $\mathfrak{g} \in \mathfrak{B}$ its image $\mathfrak{T}(\mathfrak{g})$ belongs to the space $\mathfrak{B}$. We have that $\mathfrak{V} = (\mathcal{V}_2, ..., \mathcal{V}_{\mathsf{m}+1})$ is a solution of equations (3.1)–(3.2) iff it is a fixed point of the mapping $\mathfrak{T}$. Let us show that $\mathfrak{T}$ is a contraction mapping with respect to the norm $\|\cdot\|_{\mathfrak{B}}$ (in that we follow [3, Section 4.2]). Note that the space $\mathfrak{B}$, with the norm $\|\cdot\|_{\mathfrak{B}}$, is a Banach space.

Proposition 3.1 *For every $\mathfrak{g}, \mathfrak{g}' \in \mathfrak{B}$ the following inequality holds*

$$\|\mathfrak{T}(\mathfrak{g}) - \mathfrak{T}(\mathfrak{g}')\|_{\mathfrak{B}} \leq \gamma \|\mathfrak{g} - \mathfrak{g}'\|_{\mathfrak{B}}. \tag{3.5}$$

Proof. It follows from monotonicity of ϱ that $\mathfrak{T}$ has the following monotonicity property: if $\mathfrak{g}, \mathfrak{g}' \in \mathfrak{B}$ are such that $\mathfrak{g} \succeq \mathfrak{g}'$, then $\mathfrak{T}(\mathfrak{g}) \succeq \mathfrak{T}(\mathfrak{g}')$ (by writing $\mathfrak{g} \succeq \mathfrak{g}'$ we mean that $g_t(\cdot) \geq g'_t(\cdot)$ for $t = 2, ..., \mathsf{m}+1$). It follows from translation equivariance of ϱ that $\mathfrak{T}$ has the following property of constant shift: for any $\mathfrak{g} \in \mathfrak{B}$ and $c \in \mathbb{R}$ it follows that $\mathfrak{T}(\mathfrak{g} + c\mathfrak{e}) = \mathfrak{T}(\mathfrak{g}) + \gamma c\mathfrak{e}$, where $\mathfrak{e} \in \mathfrak{B}$ is such that $\mathfrak{e}(\mathfrak{x}) = (1, ..., 1)$ for any $\mathfrak{x} \in \mathfrak{X}$. These two properties imply that $\mathfrak{T}$ is a contraction mapping. Indeed, consider $\mathfrak{g}, \mathfrak{g}' \in \mathfrak{B}$ and let $\delta := \|\mathfrak{g} - \mathfrak{g}'\|_{\mathfrak{B}}$. Then we have that $\mathfrak{g} + \delta\mathfrak{e} \succeq \mathfrak{g}'$, and hence by monotonicity of $\mathfrak{T}$ it follows that $\mathfrak{T}(\mathfrak{g} + \delta\mathfrak{e}) \succeq \mathfrak{T}(\mathfrak{g}')$. Applying the constant shift property we obtain that $\mathfrak{T}(\mathfrak{g}) + \gamma\delta\mathfrak{e} \succeq \mathfrak{T}(\mathfrak{g}')$. That is $\gamma\delta\mathfrak{e} \succeq \mathfrak{T}(\mathfrak{g}') - \mathfrak{T}(\mathfrak{g})$. The opposite inequality can be shown in the similar way, and hence (3.5) follows. ∎

The contraction property of $\mathfrak{T}$ implies the following.

Theorem 3.1 *Under assumptions* (A1)–(A5) *the following holds:* (i) *There exists unique set of functions $\mathcal{V}_2, ..., \mathcal{V}_{\mathsf{m}+1}$ satisfying the WB equations* (3.1)-(3.2). (ii) *For any $\mathfrak{g}_0 \in \mathfrak{B}$, the sequence of functions defined iteratively $\mathfrak{g}_{k+1} = \mathfrak{T}(\mathfrak{g}_k)$, $k = 0, 1, ...$, converges (in the norm $\|\cdot\|_{\mathfrak{B}}$) to $\mathfrak{V} = (\mathcal{V}_2, ..., \mathcal{V}_{\mathsf{m}+1})$ as $k \to \infty$.* (iii) *If, moreover, for all t the sets $\mathcal{X}_t$ are convex and $f_t(x_t, \xi_t)$ are convex in $x_t \in \mathcal{X}_t$, then the functions $\mathcal{V}_t : \mathcal{X}_{t-1} \to \mathbb{R}$, $t \geq 2$, are convex.*

Proof. Since the mapping $\mathfrak{T}$ is a contraction mapping, the assertions (i) and (ii) follow by the Banach Fixed Point Theorem. Suppose now that $\mathcal{X}_t$ and $f_t(\cdot, \xi_t)$ are convex, and consider $\mathfrak{g} \in \mathfrak{B}$ with convex component functions $g_{\tau+1}$. It follows then that the functions $\Psi_{g_{\tau+1}}(\cdot, \xi_\tau)$ are convex. Consequently by convexity and monotonicity of ϱ, we have that $\varrho\left(\Psi_{g_{\tau+1}}(\cdot, \xi_\tau)\right)$ are convex. Consequently the components of $\mathfrak{T}^k(\mathfrak{g})(\cdot)$ are convex for all k and hence the fixed point of $\mathfrak{T}$ has convex components. This proves (iii). ∎

Let $\mathcal{V}_2, ..., \mathcal{V}_{\mathsf{m}+1}$ be the value functions satisfying the WB equations (3.1)-(3.2), and for $t \geq \mathsf{m}+2$ the corresponding value functions be defined recursively as $\mathcal{V}_t(\cdot) = \mathcal{V}_{t-\mathsf{m}}(\cdot)$. The minimizers

$$\begin{aligned} &\bar{x}_1 \in \arg\min_{x_1 \in \mathcal{X}_1} \left\{ f(x_1) + \gamma \mathcal{V}_2(x_1) : A_1 x_1 = b_1 \right\}, \\ &\bar{x}_t \in \arg\min_{x_t \in \mathcal{X}_t} \left\{ f_t(x_t, \xi_t) + \gamma \mathcal{V}_{t+1}(x_t) : B_t(\xi_t)\bar{x}_{t-1} + A_t(\xi_t) x_t = b_t(\xi_t) \right\}, \ t \geq 2, \end{aligned} \tag{3.6}$$

give an optimal policy of the corresponding infinite horizon problem. Note that for $t = 2, ...,$ the minimizer $\bar{x}_t$ is a function of $\bar{x}_{t-1}$ and ξ_t. The respective optimal value is given by the optimal value of the first stage problem.

Remark 3.1 For $\mathsf{m} = 1$ the data is the same for all t, therefore in that case we remove the subscript t from the data functions. The WB equations for the value function $\mathcal{V} = \mathcal{V}_2$ become

$$\begin{aligned} \mathcal{V}(x) &= \varrho(V(x, \xi)), \\ V(x, \xi) &= \inf_{x' \in \mathcal{X}} \left\{ f(x', \xi) + \gamma \mathcal{V}(x') : B(\xi)x + A(\xi)x' = b(\xi) \right\}, \end{aligned} \tag{3.7}$$

where ξ denotes random vector having distribution of ξ_t. Solution of these equations are given by the fixed point of the mapping $\mathfrak{T} : \mathbb{B}(\mathcal{X}) \to \mathbb{B}(\mathcal{X})$ defined as

$$\mathfrak{T}(g)(x) := \varrho\left(\Psi_g(x, \xi)\right) \tag{3.8}$$

with

$$\Psi_g(x, \xi) := \inf_{x' \in \mathcal{X}} \left\{ f(x', \xi) + \gamma g(x') : B(\xi)x + A(\xi)x' = b(\xi) \right\}. \tag{3.9}$$

Under the specified assumptions the mapping $\mathfrak{T}$ is a contraction mapping and the assertions of Theorem 3.1 do apply.

4 Statistical inference

For the sake of simplicity let us consider first the case of the period length $\mathsf{m} = 1$, with the corresponding WB equations given in (3.7). In order to evaluate the value function $\mathcal{V}(\cdot)$ numerically we need to discretize the (possibly continuous) distribution of the random vector ξ. We discuss this in the framework of the (infinite horizon) problem (2.2) with $T = \infty$.

Let $\hat{P}$ be an approximation of the reference distribution P and $\hat{\varrho}$ be the corresponding risk measure. In the risk neutral case we have that $\varrho = \mathbb{E}_P$ and then $\hat{\varrho} = \mathbb{E}_{\hat{P}}$. Consider mapping $\mathfrak{T}$, defined in (3.8), corresponding to the original (true) problem and its approximation counterpart $\hat{\mathfrak{T}} : \mathbb{B}(\mathcal{X}) \to \mathbb{B}(\mathcal{X})$ defined in the similar way

$$\hat{\mathfrak{T}}(g)(x) := \hat{\varrho}\left(\Psi_g(x, \xi)\right). \tag{4.1}$$

Under the specified regularity conditions, both mappings $\mathfrak{T}$ and $\hat{\mathfrak{T}}$ are contraction mappings. Let $\mathcal{V}$ and $\hat{\mathcal{V}}$ be their fixed points, i.e., $\mathfrak{T}(\mathcal{V}) = \mathcal{V}$ and $\hat{\mathfrak{T}}(\hat{\mathcal{V}}) = \hat{\mathcal{V}}$. Then by using the contraction property of $\hat{\mathfrak{T}}$ we have

$$\begin{aligned} \|\mathcal{V} - \hat{\mathcal{V}}\| &= \|\mathfrak{T}(\mathcal{V}) - \hat{\mathfrak{T}}(\hat{\mathcal{V}})\| = \|\mathfrak{T}(\mathcal{V}) - \hat{\mathfrak{T}}(\mathcal{V}) + \hat{\mathfrak{T}}(\mathcal{V}) - \hat{\mathfrak{T}}(\hat{\mathcal{V}})\| \\ &\leq \|\mathfrak{T}(\mathcal{V}) - \hat{\mathfrak{T}}(\mathcal{V})\| + \gamma\|\mathcal{V} - \hat{\mathcal{V}}\|. \end{aligned}$$

It follows that

$$\|\mathcal{V} - \hat{\mathcal{V}}\| \leq (1-\gamma)^{-1}\|\mathfrak{T}(\mathcal{V}) - \hat{\mathfrak{T}}(\mathcal{V})\| = (1-\gamma)^{-1}\left\|\varrho\left(\Psi_{\mathcal{V}}(x,\xi)\right) - \hat{\varrho}\left(\Psi_{\mathcal{V}}(x,\xi)\right)\right\|. \tag{4.2}$$

Consider now the Sample Average Approximation (SAA) approach to discretization of the true problem. That is, a random sample $\xi^1, ..., \xi^N$, of iid realizations of random vector ξ, is generated by Monte Carlo sampling techniques and distribution P is approximated by the corresponding empirical distribution $\hat{P}_N = N^{-1}\sum_{j=1}^N \delta_{\xi^j}$. We can view the associated risk measure $\hat{\varrho}_N$ and value function $\hat{\mathcal{V}}_N$ as the SAA estimates of ϱ and $\mathcal{V}$, respectively. In particular, in the risk neutral case, when $\varrho = \mathbb{E}_P$, we have that

$$\hat{\varrho}_N\left(\Psi_{\mathcal{V}}(x,\xi)\right) = \mathbb{E}_{\hat{P}_N}\left[\Psi_{\mathcal{V}}(x,\xi)\right] = \frac{1}{N}\sum_{j=1}^N \Psi_{\mathcal{V}}(x,\xi^j). \tag{4.3}$$

Assuming that the set $\mathcal{X}$ is compact, $\Psi_{\mathcal{V}}(x,\xi)$ is continuous in $x \in \mathcal{X}$ and is dominated by an integrable function, we have by the uniform Law of Large Numbers (LLN) (e.g., [13, Theorem 7.53]) that

$$\lim_{N\to\infty}\left\|\mathbb{E}_P\left[\Psi_{\mathcal{V}}(x,\xi)\right] - \mathbb{E}_{\hat{P}_N}\left[\Psi_{\mathcal{V}}(x,\xi)\right]\right\| = 0 \text{ w.p.1.} \tag{4.4}$$

There is also a uniform LLN for general law invariant coherent risk measures ϱ (cf., [12]). It follows then by (4.2) that $\hat{\mathcal{V}}_N(\cdot)$ converges w.p.1 to $\mathcal{V}(\cdot)$ uniformly on $\mathcal{X}$. This in turn implies that the optimal value and first stage optimal solutions of the SAA problem converge w.p.1 to their counterparts of the (true) problem (e.g., [13, Theorem 5.3]).

In the risk neutral setting, when $\varrho = \mathbb{E}_P$, there are also available results for uniform in $x \in \mathcal{X}$ rates of convergence of $\mathbb{E}_{\hat{P}_N}\left[\Psi_{\mathcal{V}}(x,\xi)\right]$ to $\mathbb{E}_P\left[\Psi_{\mathcal{V}}(x,\xi)\right]$ (e.g., [13, Section 7.2.10]). By (4.2) this implies that for $\varepsilon > 0$, under appropriate regularity conditions, probability of the event $\{\|\mathcal{V} - \hat{\mathcal{V}}_N\| \geq \varepsilon\}$ converges to zero exponentially fast with increase of the sample size N. It also follows from those results that in order to solve the true problem with accuracy $\varepsilon > 0$ by solving the SAA problem with accuracy $\delta \in [0, \varepsilon)$, for a given confidence one needs a sample size N of order $O\left((\varepsilon - \delta)^{-2}\right)$ (cf., [13, Section 5.3.2]).

In the general setting of $\mathsf{m} \geq 1$ we can proceed in the similar way. The true (marginal) distribution of every ξ_t is approximated by generating an iid sample $\xi_t^1, ..., \xi_t^N$, say of the same size N for every $t = 2, ..., \mathsf{m}+1$. (Recall that the process ξ_t is assumed to be stagewise independent.) Consequently the value functions $\mathcal{V}_t(x_{t-1})$ are approximated by their sample average approximations

$$\hat{\mathcal{V}}_{t,N}(x_{t-1}) = \hat{\varrho}_N\left(\Psi_{\mathcal{V}_{t+1}}(x_{t-1},\xi_t)\right),\ t = 2, ..., \mathsf{m}+1,$$

by replacing the true (original) distribution of vectors ξ_t in WB equations (3.1) - (3.2) with their respective empirical distributions based on the generated samples. Similar to (4.2) by using the max-norm $\|\cdot\|_{\mathfrak{B}}$, we have that for $t = 2, ..., \mathsf{m}+1$,

$$\|\mathcal{V}_t - \hat{\mathcal{V}}_{t,N}\| \leq (1-\gamma)^{-1} \max_{2\leq\tau\leq\mathsf{m}+1} \left\|\varrho\left(\Psi_{\mathcal{V}_{\tau+1}}(x_{\tau-1},\xi_\tau)\right) - \hat{\varrho}_N\left(\Psi_{\mathcal{V}_{\tau+1}}(x_{\tau-1},\xi_\tau)\right)\right\|, \tag{4.5}$$

with $\mathcal{V}_{\mathsf{m}+2} = \mathcal{V}_2$.

Under mild regularity conditions (in particular assuming compactness of the sets $\mathcal{X}_t$) a uniform LLN can be applied to the empirical risk measures $\hat{\varrho}_N$ (cf., [12],[13, Section 7.2.6]). It follows then that the optimal value and first stage optimal solutions of the SAA problem converge w.p.1 to their counterparts of the (true) problem. In the risk neutral case, when $\varrho = \mathbb{E}$, it is also possible to apply a general analysis of sample complexity of the SAA approach in the multistage setting (we refer to [13, Section 5.8.2] for the derivations and discussion of required regularity conditions). Again for a given accuracy $\varepsilon > 0$ and $\delta \in [0, \varepsilon)$ one needs the sample size N (per stage) of order $O\left((\varepsilon-\delta)^{-2}\right)$. Note however that the total number of scenarios (sample paths) of the constructed SAA problem is $\mathcal{N} = N^{\mathsf{m}}$. Therefore the corresponding number of *scenarios* $\mathcal{N}$ is of order $O\left((\varepsilon-\delta)^{-2\mathsf{m}}\right)$.

5 Cutting plane algorithm

There are different approaches to computing value function, satisfying the WB equations, which were suggested in the literature (see, e.g., [3]). For a recent discussion of limitations of classical approaches we can refer to [16]. In this section we discuss a cutting planes approach to approximation of the value functions. This is somewhat similar to the Stochastic Dual Dynamic Programming (SDDP) method (introduced in [8]), which became popular for solving multistage linear (convex) stochastic programs. Such type of algorithms for solving the WB equations in Optimal Control, with $\mathsf{m} = 1$, were recently considered in [2] and [16].

Let us consider first the case of $\mathsf{m} = 1$ (see Remark 3.1); the general periodic case will be discussed in sections below. We assume that the distribution of the random vector ξ is discretized, say by the SAA method, so that the set $\Xi = \{\xi^1, ..., \xi^N\}$ is finite, equipped with probabilities $p_j > 0$, $j = 1, ..., N$. We assume that $f(x,\xi)$ is convex in x, in particular it can be linear $f(x,\xi) = c(\xi)^\top x$, and the set $\mathcal{X}$ is convex, and hence functions $V(\cdot,\xi)$ and $\mathcal{V}(\cdot)$ are convex. For now we consider the risk neutral case, when $\varrho = \mathbb{E}$; later we will comment how this can be extended to general coherent risk measures. The WB equations can be written here as

$$\begin{aligned} \mathcal{V}(x) &= \sum_{j=1}^N p_j V^j(x), \\ V^j(x) &= \min_{x'\in\mathcal{X}} \left\{f^j(x') + \gamma\mathcal{V}(x') : B^j x + A^j x' = b^j\right\},\ j = 1, ..., N, \end{aligned} \tag{5.1}$$

where $f^j(x) = f(x,\xi^j)$, $B^j = B(\xi^j), A^j = A(\xi^j), b^j = b(\xi^j)$, and p_j is the probability of scenario j. Given a current piecewise linear (under) approximation $\underline{\mathcal{V}}(\cdot)$ of function $\mathcal{V}(\cdot)$ and a trial point $x \in \mathcal{X}$, the next cutting plane is constructed by computing a subgradient g^j, at x, of the current estimate of value function

$$\underline{V}^j(x) = \min_{x'\in\mathcal{X}} \left\{f^j(x') + \gamma\underline{\mathcal{V}}(x') : B^j x + A^j x' = b^j\right\} \tag{5.2}$$

for $j = 1, ..., N$. This subgradient is obtained by solving the respective dual problem. If $f(x, \xi^j) = (c^j)^\top x$ is linear and $\mathcal{X}$ is a polyhedral set given by a finite number of affine constraints, then problem (5.2) is a linear programming problem. Then the respective subgradient of $\mathcal{V}(\cdot)$ at the point x is estimated as $\sum_{j=1}^N p_j g^j$. There are different strategies for generating trial point(s) at each iteration, we will discuss this in section 5.1.1 below.

For an estimate $\underline{\mathcal{V}}(\cdot)$ of the value function, consider policy

$$\bar{x}_t \in \arg\min_{x_t \in \mathcal{X}} \left\{ f(x_t, \xi_t) + \gamma \underline{\mathcal{V}}(x_t) : B(\xi_t)\bar{x}_{t-1} + A(\xi_t)x_t = b(\xi_t) \right\}, \ t = 1, ..., \tag{5.3}$$

with $\bar{x}_0 = 0$ and $f(x_1, \xi_1) = f_1(x_1)$. This is a feasible policy for the respective infinite horizon problem. For a sample path (realization) $\xi_1, ...,$ of the random process the point estimate of corresponding value of this policy is $\sum_{t=1}^\infty \gamma^{t-1} f_t(\bar{x}_t, \xi_t)$. Hence the value of this policy can be estimated by averaging these point estimates for M randomly generated sample paths (this can be applied to the discretized or the original distribution of the random process). This will give an upper estimate of the optimal value of problem. This is similar to the forward step of the SDDP algorithm.

Note that for finite sum $\sum_{t=1}^T \gamma^{t-1} f_t(\bar{x}_t, \xi_t)$ we can estimate the respective error as (compare with (2.3))

$$\left| \sum_{t=1}^\infty \gamma^{t-1} f_t(\bar{x}_t, \xi_t) - \sum_{t=1}^T \gamma^{t-1} f_t(\bar{x}_t, \xi_t) \right| = \left| \sum_{t=T+1}^\infty \gamma^{t-1} f_t(\bar{x}_t, \xi_t) \right| \leq \frac{\kappa \gamma^T}{1-\gamma}, \tag{5.4}$$

where κ is such that $|f_t(\bar{x}_t, \xi_t)| \leq \kappa$.

Similar procedure can be applied to a law invariant coherent risk measure ϱ. The only difference is how a subgradient of $\underline{V}^j(x)$ is computed (cf., [15]). Unfortunately for general coherent risk measures it is not clear how to compute an upper bound for the optimal value of the considered risk averse problem by the randomization method, discussed above for the risk neutral problem. Construction of the upper bound by an outer approximation was suggested in [9]. That procedure is based on a certain discretization of the set $\mathcal{X}$; as a consequence the number of discretization points needed to achieve a given accuracy typically grows exponentially with increase of the dimension of $x \in \mathcal{X}$.

In the next section we discuss an implementation of SDDP type algorithm, referred to as the periodical SDDP algorithm, applied to the periodical WB equations, for general period $\mathsf{m} \geq 1$ and risk measures.

5.1 SDDP type algorithm

In this section we discuss the proposed periodical variant SDDP algorithm applied to periodical infinite horizon problems. We follow the formalism, and apply software, described in [4]. That is, consider risk functional $\mathcal{R}_\infty$, defined in (2.4), and the following problem

$$\min_{\pi \in \Pi} \mathcal{R}_\infty \left(\sum_{t=1}^\infty \gamma^{t-1} (u_t^\top x_t + v_t^\top y_t) \right), \tag{5.5}$$

with decision variables (x_t, y_t) and respective policies $\pi \in \Pi$ satisfying the constraints

$$B_t x_{t-1} + A_t x_t + C_t y_t \geq b_t. \tag{5.6}$$

In that approach x_t can be viewed as state variables and y_t as local (control) variables similar to the model used in the Optimal Control. The difference from the Optimal Control approach is that the state variables x_t are not determined completely by the local variables y_t and the minimization in (5.5) is performed jointly in x_t and y_t. Anyway in the above formulation the corresponding value functions are functions of state variables and do not depend on y_t.

Suppose that the above problem has a periodical behavior with period $\mathsf{m} \geq 1$. The stage-wise independent stochastic processes $u_t, v_t, A_t, B_t, C_t, b_t$ are discretized into finite set Ω_t for each stage $t = 2, \ldots, \mathsf{m}+1$ (recall that the first stage data u_1 and v_1 are deterministic). The discretized problem is solved by a periodical version of the SDDP algorithm (see Algorithm 1 below).

5.1.1 Construction of trial points

There is a delicate issue of choosing trial points where the value functions are updated by construction of the respective cutting planes. This is related to the question of eventual convergence of the algorithm. Recall that the aim here is solving the corresponding infinite horizon problem (with the discount factor $\gamma < 1$). The WB equations describe the value functions for this infinite horizon problem. Available proofs of convergence of standard SDDP algorithms, for finite horizon problems, are quite technical (see [6],[7], and references therein). Careful derivation of such proofs for the considered framework is outside the scope of this paper and will be left for a future research.

When the number of scenarios (sample paths) is finite, such proofs of convergence are based on the property that the considered algorithm eventually goes through *every* scenario. Of course when the number of scenarios is astronomically large, this can not happen in a reasonable time. Therefore although such proofs have a theoretical interest, they do not guarantee convergence to an arbitrary small precision for larger problems. Therefore although such theoretical developments give general guidelines, in practice various heuristics are used. We used the following heuristics trying to mimic the standard SDDP approach, this worked reasonably well.

In the forward step of SDDP algorithm, by iteratively solving the first T stage problems, we obtain solutions of state variables, denoted by $\tilde{x}_t$, for each stage $t = 1, \ldots, T$. Those solutions are considered as candidates of trial points for stage 2 to stage $\mathsf{m}+1$ in the backward step. Various trial points selection methods can be considered. The method we adopt here is as follows. Consider all consecutive m-stage periods for the T stages, i.e., $\big\{(s, s+1, \ldots, s+\mathsf{m}-1) : s = 1, \mathsf{m}+1, 2\mathsf{m}+1 \ldots, s+\mathsf{m}-1 \leq T\big\}$. Randomly pick one of the cycles $(\hat{s}, \hat{s}+1, \ldots, \hat{s}+\mathsf{m}-1)$. Trial points are then selected as $\bar{x}_k = \tilde{x}_{s_0+k-1}, k = 1, \ldots, \mathsf{m}$.

The backward step then solves each stage problem at those trial points $\bar{x}_k, k = \mathsf{m}, \ldots, 1$, from stage $\mathsf{m}+1$ to stage 2 backwardly and iteratively adds a corresponding cutting plane to the previous stage. The cutting plane obtained from stage 2 is also added to stage $\mathsf{m}+1$ since $\hat{\mathcal{V}}_{\mathsf{m}+2}(\cdot) = \hat{\mathcal{V}}_2(\cdot)$. In the end, by solving the first stage problem, the backward step is able to produce a lower bound for the infinite horizon problem. Following the paper [4], we add local copy variables $z_t, t = 1, \ldots, T$, to simplify dual computation and modeling process.

Algorithm 1 Periodical SDDP algorithm for stagewise independent infinite horizon problem with a discount factor $\gamma \in (0,1)$ and a period $\mathsf{m} \geq 1$

Given sample size N_t and discretization $\Omega_t = \{u_t^j, v_t^j, A_t^j, B_t^j, C_t^j, b_t^j\}_{1 \leq j \leq N_t}$, for $t = 2, \ldots, \mathsf{m}+1$
Initial approximation of value functions: $\hat{\mathcal{V}}_2^{(0)}, \ldots, \hat{\mathcal{V}}_{\mathsf{m}+1}^{(0)}, \hat{\mathcal{V}}_{\mathsf{m}+2}^{(0)} (= \hat{\mathcal{V}}_2^{(0)})$
Initialize: $i = 1$, LB = -∞
while no stopping criterion is met **do**
  (Forward Step)
  for $t = 1, \ldots, T (T \geq \mathsf{m}+1)$ **do**
    if $t = 1$ **then** $t' = 1$
    else
      $t' \equiv t \pmod{\mathsf{m}}$, where $t' \in \{2, \ldots, \mathsf{m}+1\}$
      Draw a sample $(u_{t'}, v_{t'}, A_{t'}, B_{t'}, C_{t'}, b_{t'})$ from $\Omega_{t'}$
    end if
    $\tilde{x}_t, \tilde{y}_t = \operatorname{argmin} \{u_{t'}^\top x_t + v_{t'}^\top y_t + \gamma \hat{\mathcal{V}}_{t'+1}^{(i-1)}(x_t) : A_{t'} x_t + B_{t'} z_t + C_{t'} y_t \geq b_{t'}, z_t = \tilde{x}_{t-1}\}$
  end for
  Trial point selection: $(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{\mathsf{m}})$ selected from candidates $(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_T)$
  (Backward Step)
  for $t = \mathsf{m}+1, \ldots, 2$ **do**
    for $j = 1, \ldots, N_t$ **do**
      if $t = \mathsf{m}+1$ **then** $i' = i-1$
      else $i' = i$
      end if
      Solve min $\{u_{tj}^\top x_t + v_{tj}^\top y_t + \gamma \hat{\mathcal{V}}_{t+1}^{(i')}(x_t) : A_t^j x_t + B_t^j z_t + C_t^j y_t \geq b_t^j, z_t = \bar{x}_{t-1}\}$ and get optimal value $\mathscr{V}_{tj}$ and dual solution $\mathscr{G}_{tj}$ corresponding to constraint $z_t = \bar{x}_{t-1}$
    end for
    Compute coefficients of a new cutting plane, $\mathscr{V}_t$ and $\mathscr{G}_t$, from $\mathscr{V}_{tj}$ and $\mathscr{G}_{tj}$. In particular, for risk neutral measure: $\mathscr{V}_t := \frac{1}{N_t} \sum_{j=1}^{N_t} \mathscr{V}_{tj}, \mathscr{G}_t := \frac{1}{N_t} \sum_{j=1}^{N_t} \mathscr{G}_{tj}$
    $\hat{\mathcal{V}}_t^{(i)} \leftarrow \{\theta \in \hat{\mathcal{V}}_t^{(i-1)}, \theta \geq \mathscr{G}_t (x_{t-1} - \bar{x}_{t-1}) + \mathscr{V}_t\}$
  end for
  LB = $\min\{u_1^\top x_1 + v_1^\top y_1 + \gamma \hat{\mathcal{V}}_2^{(i)}(x_1) : A_1 x_1 + B_1 z_1 + C_1 y_1 \geq b_1, z_1 = x_0\}$
  $i = i + 1$
end while

5.1.2 Policy evaluation

After solving the discretized problem, the obtained policy is feasible and implementable for infinite number of stages. The classical SDDP on the other hand, produces a policy that is implementable up to a finite number of stages. Nevertheless, we can still compare the two approaches by evaluating the obtain policies up to some finite number of stages both for the discretized problem and the true problem (by "true" we mean the original problem with possibly continuous distributions of the random variables). The evaluation of the periodical SDDP can be done by running the forward step in Algorithm 1 multiple times.

In the risk neutral setting, the finite summation of individual discounted stage costs $\sum_{t=1}^{T} \gamma^{t-1} \left[u_{t'}^\top \bar{x}_t + v_{t'}^\top \bar{y}_t \right]$, for sufficiently large T, can be regarded as an approximation of the policy value for $T = \infty$ (the error of such approximation is given by (5.4)). Therefore, we are able to construct confidence interval of the (approximated) policy value. If we replace the discretized set Ω_t by the true distribution of the stochastic processes, we are also able to construct confidence interval of the policy value for the true problem. In the end, the optimality gap of the discretization problem is given by the percentage difference between the upper end of the confidence interval and the lower bound. It is also possible to compute the optimality gap of the true problem by randomizing the discretized problem. Stopping criteria and parallelization technique for the classical SDDP are also valid here. All of the evaluation techniques are quite similar to the classical SDDP. For more details, we direct readers to the paper [4].

6 Numerical experiments

We implemented the periodical SDDP algorithm for the Brazilian interconnected power system problem (cf., [15]). The system has four regions. In each region i, demand is satisfied by energy generation q_i from an integrated reservoir, energy generation g_k from local thermal plants $k \in \Omega_i$ and energy inflow $\mathrm{ex}_{j\to i}$ from other regions ($j \in \{1, 2, 3, 4\}$) or a transshipment station ($j = 5$). If demand d_i can not be satisfied, system i borrows df_{ij} units of energy from the deficit account in system j and a cost of e_{ij} will be incurred. The deficit accounts are set up in a way that ensures the complete recourse condition is held. There are tiny costs for each unit of energy spillage and energy exchanges, denoted by b_i, $c_{j\to i}$ (energy flowed from region j to region i) respectively. We assume a cost of u_k for each unit of energy thermal plant k produces. The dynamics of the system is that the stored energy v_{it} of each reservoir i in stage t is equal to the previous stored energy $v_{i,t-1}$ plus the current water inflow a_{it} minus the energy generation q_{it} and the water spillage s_{it}. The transshipment station is solely for water exchanges between regions. The inflow energy is modeled as a periodical first order autoregressive process with multiplicative errors. The model has a period of 12 in the sense that the deterministic demand d_{it}, the stochastic process $\{a_{it}\}$ are both with period 12 from stage 2. The objective is to obtain a sequential scheme to minimize the total operation cost over a design period of T stages with a discount factor γ while meeting energy requirements

and feasibility constraints.

$$\min \mathcal{R}_T\left(\sum_{t=1}^{T}\gamma^{t-1}\left\{\sum_{i=1}^{4}\left[b_{it}s_{it}+\sum_{j=1}^{5}\mathrm{c}_{j\to i,t}\mathrm{ex}_{j\to i,t}+\sum_{k\in\Omega_i}u_k g_{kt}+\sum_{j=1}^{4}e_{ij}\mathrm{df}_{ijt}\right]\right.\right.$$
$$\left.\left.+\sum_{j=1}^{4}c_{j\to 5,t}\mathrm{ex}_{j\to 5,t}\right\}\right)$$
$$\text{s.t. } \forall t=1,\ldots,T$$
$$v_{it}+s_{it}+q_{it}-v_{i,t-1}=a_{it},\forall i,$$
$$q_{it}+\sum_{k\in\Omega_i}\mathrm{g}_{kt}+\sum_{j=1}^{4}\mathrm{df}_{ijt}-\sum_{j=1}^{5}\mathrm{ex}_{i\to j,t}+\sum_{j=1}^{5}\mathrm{ex}_{j\to i,t}=d_{it},\forall i,$$
$$\sum_{j=1}^{4}\mathrm{ex}_{j\to 5,t}=\sum_{j=1}^{4}\mathrm{ex}_{5\to j,t},$$
$$0\le v_{it}\le\bar{v}_i,0\le q_{it}\le\bar{q}_i,\forall i,$$
$$\underline{g}_k\le g_{kt}\le\bar{g}_k,\forall k,$$
$$0\le df_{ijt}\le\bar{df}_{ij},0\le ex_{i\to j,t}\le\bar{ex}_{i\to j},\forall i,\forall j,$$
$$v_{i,0}\text{ are given, }a_{i,1}\text{ are deterministic, }\forall i.$$

This problem was originally a 60-stage (60-month) multistage problem with a monthly discount factor $\gamma = 0.9906$ (this discount factor corresponds to the annual discount rate of 12%, i.e., $1/\gamma^{12} = 1.12$ for $\gamma = 0.9906$, and was used in the original project). In order to deal with the so-called end-of-horizon effect, 60 more stages are added to the problem, which results in $T = 120$ stage problem. However, implementing the classical SDDP for the resulting 120-stage problem requires huge computational effort and resources. By contrast, the periodical SDDP leverages the periodical structure of the problem, considers $T = \infty$ and only needs to solve a 13-stage problem (with deterministic first stage problem and $\mathsf{m} = 12$).

For the discount factor $\gamma = 0.9906$ in risk neutral setting, the error estimate given in (5.4), is $\gamma^T/(1-\gamma) \approx 34$. This estimate is still large, indicating that the cost-to-go function at the 120th stage may be far from zero and the end-of-horizon effect may still be significant. If setting a smaller discount factor, for example $\gamma = 0.8$, the coefficient in (5.4) becomes $\gamma^T/(1-\gamma) \approx 1.17e-11$, while the upper bound κ of the objective is less than $1e10$ for this problem. It then makes sense to regard the cost-to-go function at the 120th stage as zero. In such case, the 120-stage problem is approximately the same as the infinite horizon problem. It can be noted that the monthly discount factor $\gamma = 0.8$ may not have an economic sense, but it is still useful for algorithmic comparisons.

In this section, we numerically compare the classical SDDP and the proposed (periodical) variant of SDDP algorithm on this Brazilian power problem with discount factors 0.8 and 0.9906 in risk neutral, risk averse, and integer settings. The true problem is discretized by SAA with sample size $N_t = 100$ at every stage $t \ge 2$. Both algorithms are implemented on SAA discretized problems for 300 iterations and are parallelized by 10 processes (i.e., 10 forward and backward steps are simultaneously done for each iteration, and hence in total

effectively around 3000 iterations are implemented). For every 100 iterations, we evaluate the obtained policy from both approaches and compare the bound, policy value and solution. Since the periodical SDDP solves a much smaller problem per iteration (ten times smaller in this case since there are around 10 periods for $T = 120$ and $\mathsf{m} = 12$), if the comparison shows similar results, the periodical SDDP will be significantly more efficient for periodical multistage problems than the classical SDDP. All experiments are done using MSPPy [4], Gurobi 8.1.0, Python 3.7.1 in Red Hat Linux. More details can be found at github.com/lingquant/msppy.

6.1 Risk neutral problem

We start with the risk neutral version of the problem, with $\varrho = \mathbb{E}$, and discount factor $\gamma = 0.8$. Table 1 compares the speed of convergence for the two approaches. The lower bound, confidence interval of policy values, and optimality gap evolve in a similar pattern for the classical SDDP and periodical SDDP. The periodical SDDP has similar rate of convergence for the same number of iterations. Figure 1 illustrates the evolution of individual stage costs both for the discretized problem and the true problem. For both approaches, the individual stage costs stabilize after 300 iterations. Figure 2 exhibits the evolution of stored energy in one of the regions for the discretized problem and the true problem. The stored energy stabilizes very quickly and it is slightly more cyclical for the periodical SDDP than for the classical SDDP. Both plots also reflect that the discretized problem is a good approximation to the true problem.

Table 1: Comparison of classical SDDP and periodical SDDP, with the lower bound, the 95% confidence interval of policy values for the SAA discretized problem and the true problem, and the optimality gap for the SAA discretized problem

	periodical SDDP			classical SDDP		
#iter	LB ($m)	CI approx.($m) CI true($m)	gap	LB ($m)	CI approx.($m) CI true($m)	gap
100	7.09	7.57, 8.03 7.32, 7.81	13.27%	7.14	7.70, 8.19 7.42, 7.93	14.71%
200	7.31	7.55, 8.01 7.32, 7.81	9.56%	7.30	7.67, 8.15 7.40, 7.91	11.6%
300	7.39	7.51, 7.97 7.27, 7.76	7.84%	7.37	7.65, 8.13 7.38, 7.89	10.3%

Figure 1: Individual stage costs (in average value and 0.9 quantile) by periodical SDDP (on the left) and classical SDDP (on the right) for the SAA discretized problem (on the above) and the true problem (on the bottom) for the risk neutral case with discount factor 0.8, after 100, 200 and 300 iterations

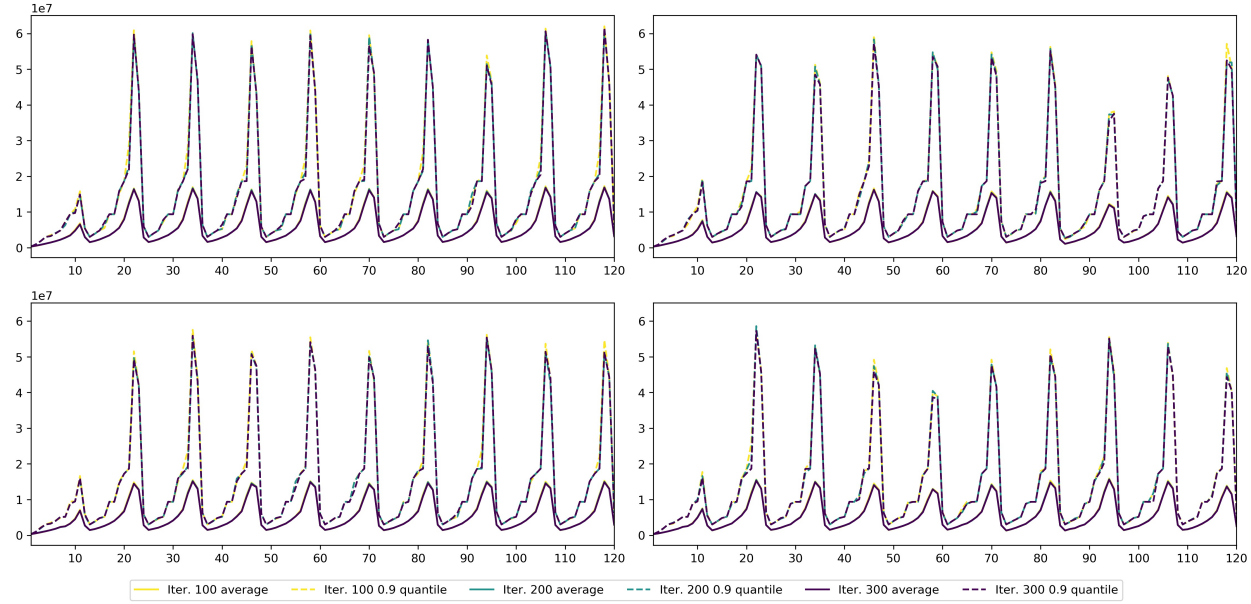
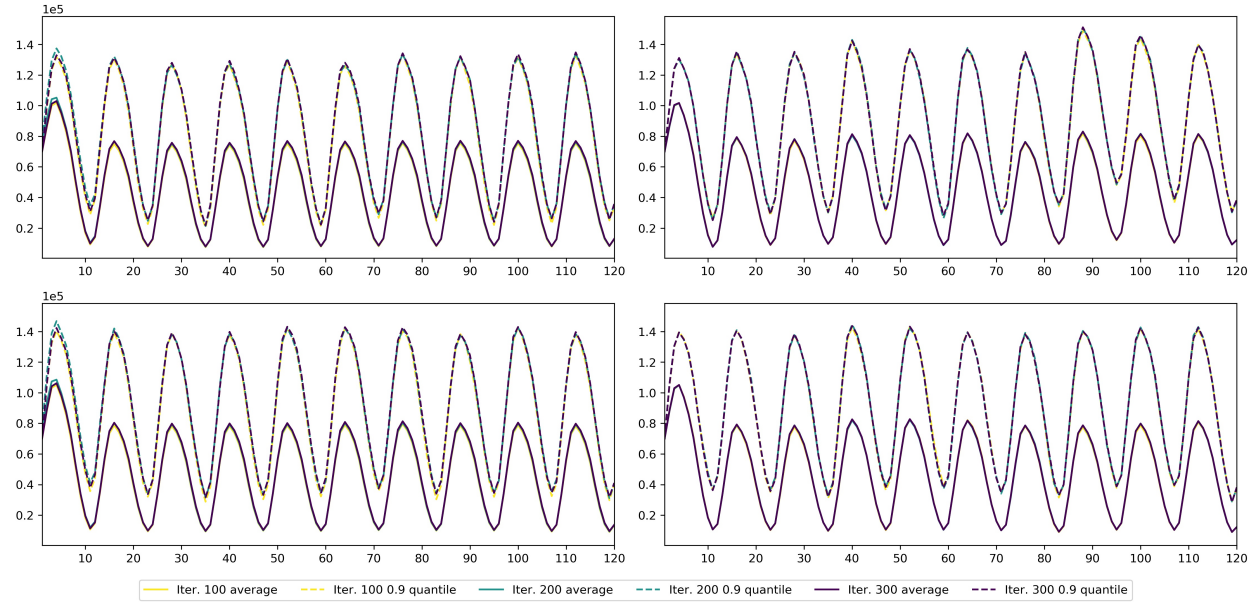


Figure 2: Stored energy (in average value and 0.9 quantile) by periodical SDDP (on the left) and classical SDDP (on the right) for the SAA discretized problem (on the above) and the true problem (on the bottom) for the risk neutral case with discount factor 0.8, after 100, 200 and 300 iterations



6.2 Changing discount factor from 0.8 to 0.9906

The original problem has a discount factor of $\gamma = 0.9906$. As shown above, the 120 stage problem may not be a good approximation of the infinite horizon problem. Indeed as shown in Figure 3, the implementation results demonstrate significant discrepancies between the two approaches. The stage costs for the periodical SDDP are similar to those for the classical SDDP for about the first 10 stages, while in the classical SDDP they decay faster in later stages. It appears that the reason for that is the end-of-horizon effect: when approaching the end of the horizon, the classical SDDP tends to use up all available stored energy and thus reduces the stage costs. This argument is verified by Figure 4. That being said, in practice, people typically only consider the costs for the first few stages and run a rolling horizon based SDDP algorithm. From that point of view these two approaches behave in a similar way.

Figure 3: Individual stage costs (in average value and 0.9 quantile) by periodical SDDP (on the left) and classical SDDP (on the right) for the SAA discretized problem (on the above) and the true problem (on the bottom) for the risk neutral case with discount factor 0.9906, after 100, 200 and 300 iterations

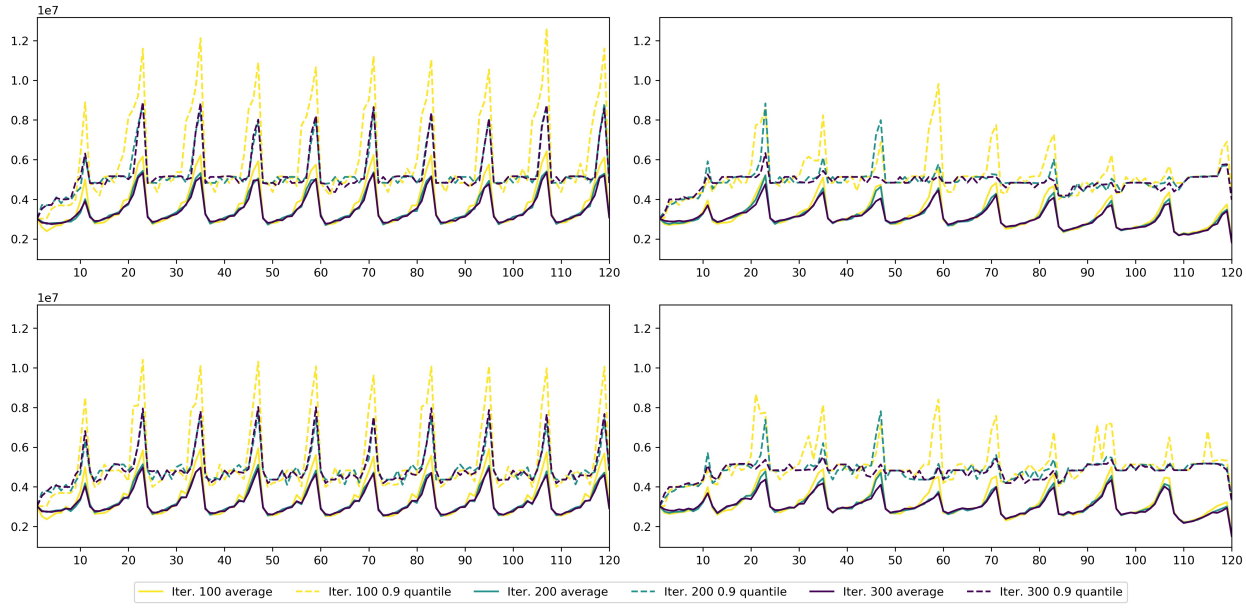
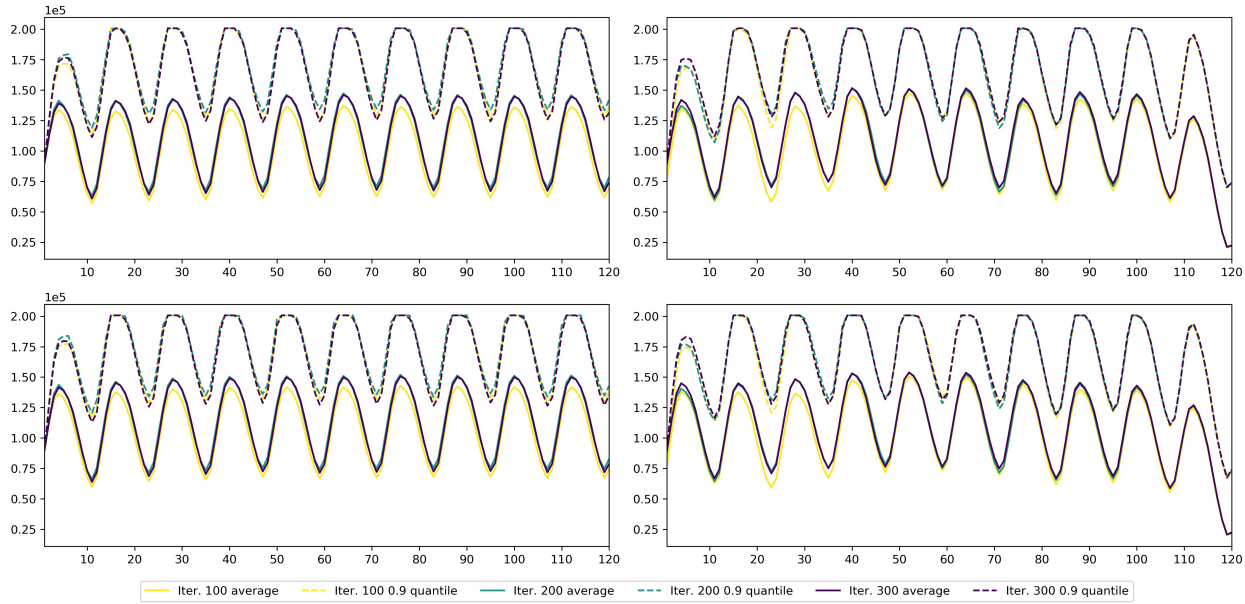


Figure 4: Stored energy (in average value and 0.9 quantile) by periodical SDDP (on the left) and classical SDDP (on the right) for the SAA discretized problem (on the above) and the true problem (on the bottom) for the risk neutral case with discount factor 0.9906, after 100, 200 and 300 iterations



6.3 Incorporating risk

We compare now the two approaches in a risk averse setting. Specifically, we consider a risk measure which is a convex combination of the expectation and the Average Value-at-Risk,

$$\mathrm{AV@R}_\alpha(Z) := \inf_{x\in\mathbb{R}} \{x + \alpha^{-1}\mathbb{E}[Z - x]_+\},$$

risk measure (also called Conditional Value-at-Risk, Expected Shortfall, Expected Tail Loss),

$$\varrho(\cdot) := (1-\lambda)\mathbb{E}(\cdot) + \lambda \mathrm{AV@R}_\alpha(\cdot),$$

with $\lambda = 0.2$ and $\alpha = 0.1$.

Again, as shown in Figures 5 - 8, both approaches stabilize after 300 iterations. For discount factor 0.8, periodical SDDP shows similar but slightly more cyclical result then the classical SDDP. For discount factor 0.9906, the classical SDDP suffers more from the end-of-horizon effect than in the risk neutral setting and the periodical SDDP produces more cyclical and regular solutions.

Figure 5: Individual stage costs (in average value and 0.9 quantile) by periodical SDDP (on the left) and classical SDDP (on the right) for the SAA discretized problem (on the above) and the true problem (on the bottom) for the risk averse case with discount factor 0.8, after 100, 200 and 300 iterations

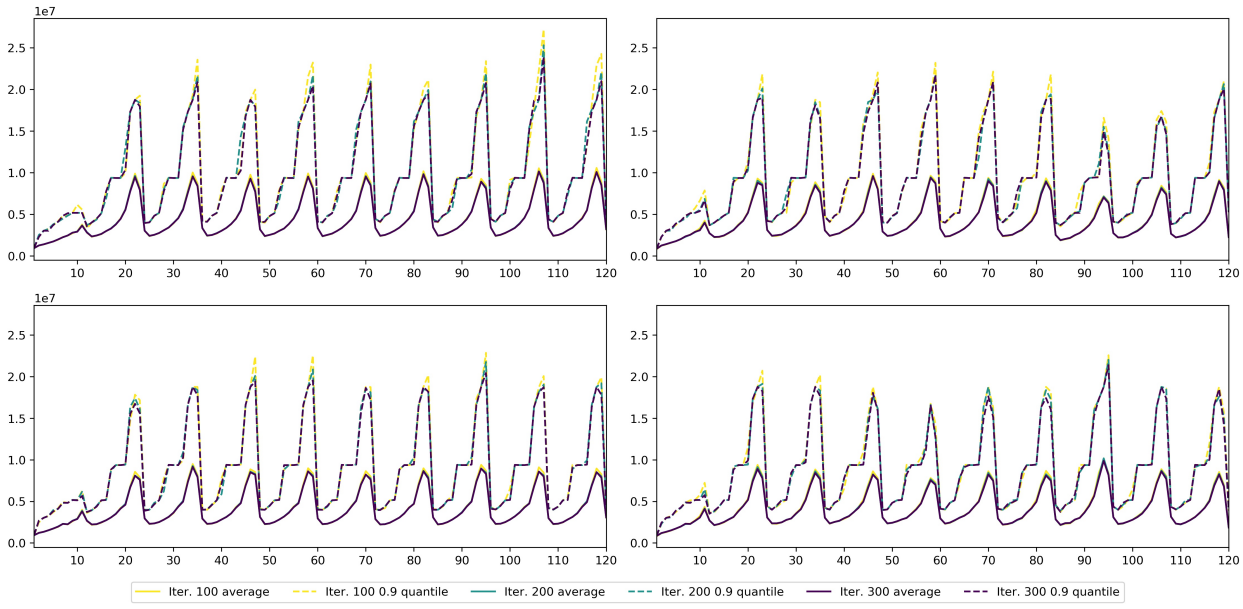


Figure 6: Stored energy (in average value and 0.9 quantile) by periodical SDDP (on the left) and classical SDDP (on the right) for the SAA discretized problem (on the above) and the true problem (on the bottom) for the risk averse case with discount factor 0.8, after 100, 200 and 300 iterations

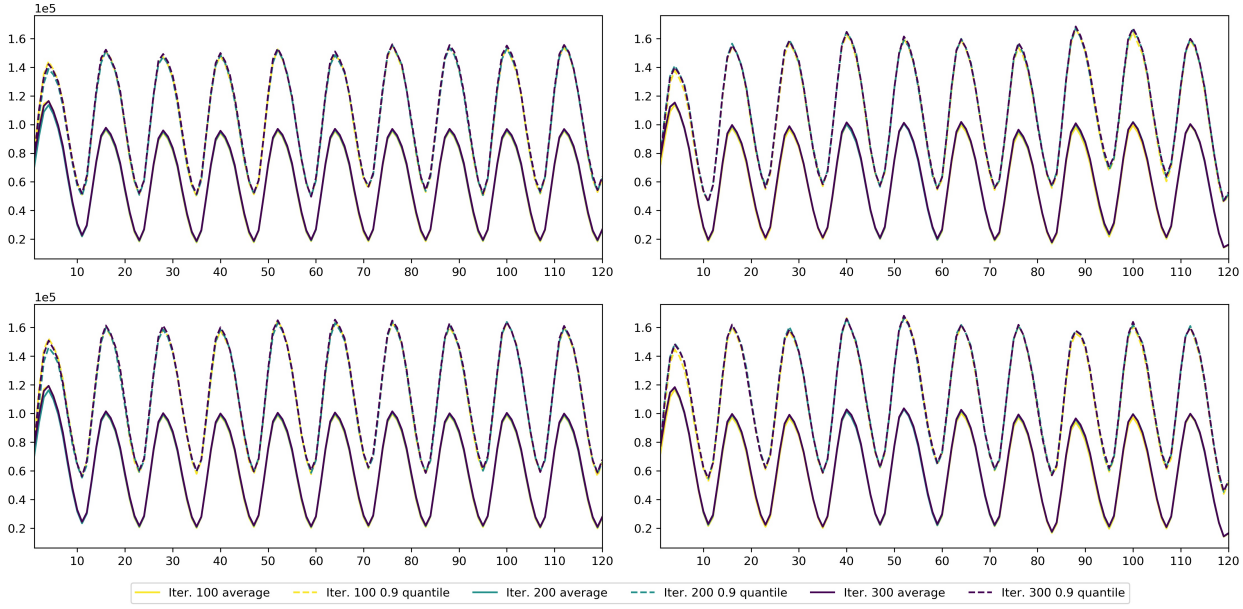


Figure 7: Individual stage costs (in average value and 0.9 quantile) by periodical SDDP (on the left) and classical SDDP (on the right) for the SAA discretized problem (on the above) and the true problem (on the bottom) for the risk averse case with discount factor 0.9906, after 100, 200 and 300 iterations

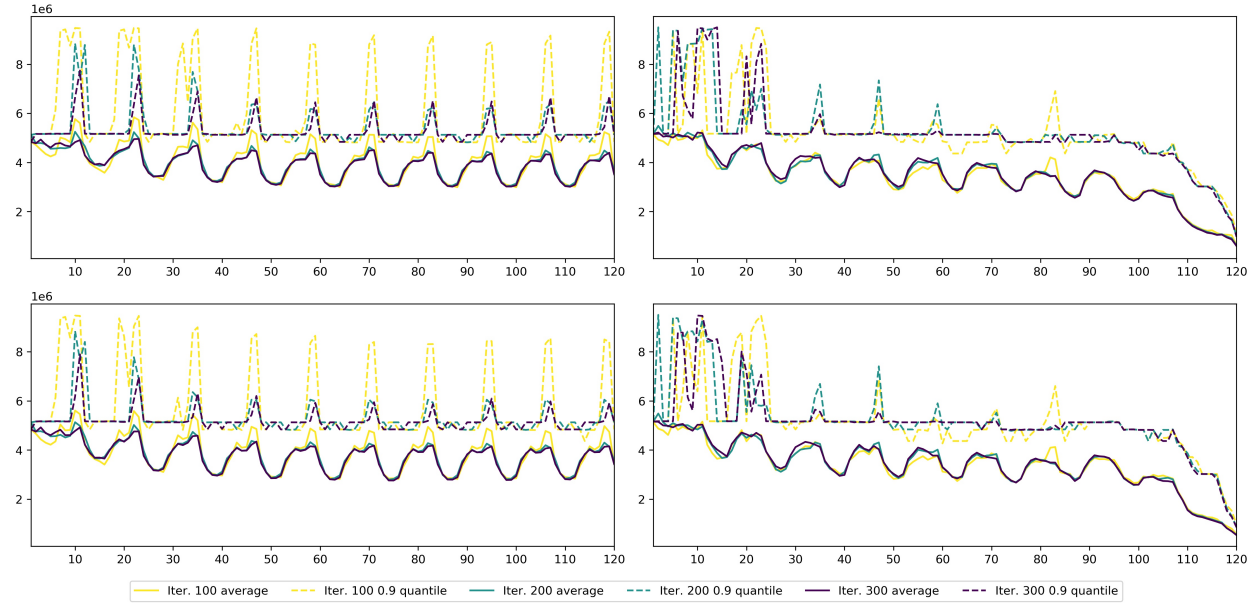
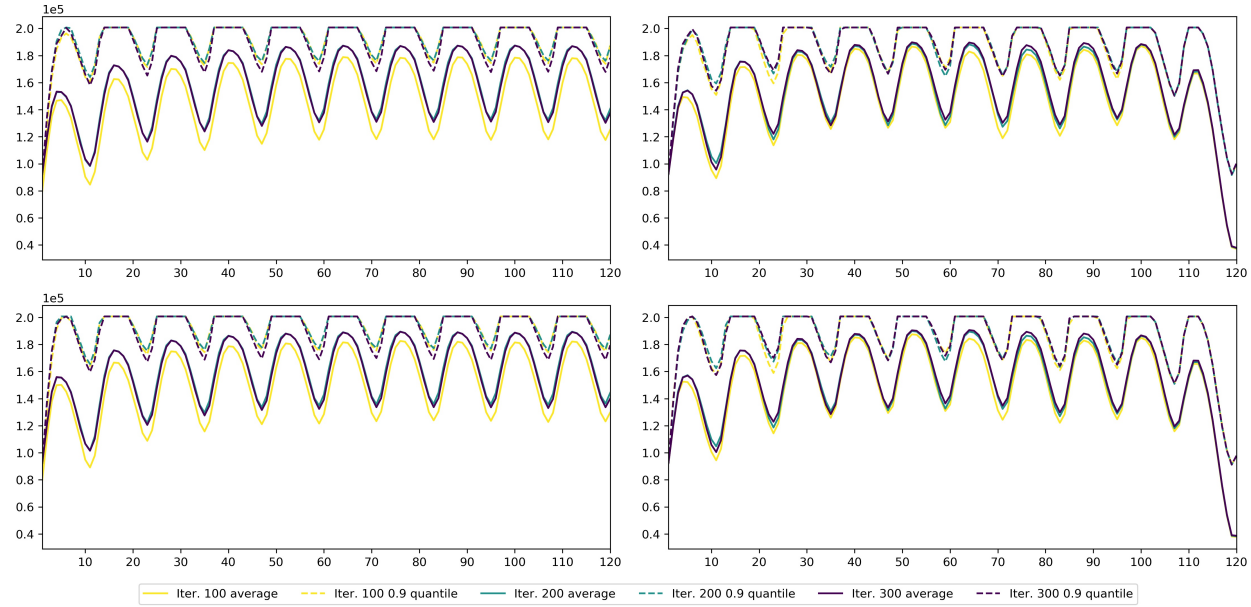


Figure 8: Stored energy (in average value and 0.9 quantile) by periodical SDDP (on the left) and classical SDDP (on the right) for the SAA discretized problem (on the above) and the true problem (on the bottom) for the risk averse case with discount factor 0.9906, after 100, 200 and 300 iterations



6.4 Including binary variables

One of the concerns about the Brazilian power system is to include a thermal security constraint. That is, whenever the stored energy is below certain level, thermal energy must be produced to ensure enough operational maneuver options for hydro plants and for electrical network reliability. This can be achieved by considering risk neutral measure $\varrho = \mathbb{E}$ and including binary variables as follows,

$$v_{it} \geq z_i v_i^*, \sum_{k \in \omega_i} g_{kt} \geq (1 - z_i) g_i^*, g_i^* \text{ and } v_i^* \text{ given}, z_i \in \{0, 1\}, \forall i.$$

Adding the above additional binary variables and constraints makes the problem a multistage stochastic mixed-integer program. Such type of problems are solved by the so-called SDDiP algorithm proposed in [17], i.e., SDDP with strengthen Benders' cuts or Lagrangian cuts. Following this, the paper [4] makes some preliminary attempts to solve the above problem by SDDiP with strengthen Benders' cuts for only up to 6 stages. As far as we know, the above problem has not been solved in the literature because of the heavy computational cost.

By virtue of the proposed techniques, we only need to solve the corresponding 13-stage problem, which makes the SDDiP approach computationally much more tractable. As shown in Table 2, such large-scale mixed integer problem is able to be solved in a reasonable accuracy. Figures 9 and 10 compare the results with and without thermal security constraint. For the discount factor 0.9906, the imposed thermal security does not have much impact on the solutions and costs. While for the discount factor 0.8, the imposed thermal security constraint leads to retain more stored energy over the entire horizon.

Table 2: Result of SDDiP with Benders' cuts for the mixed integer problem with discount factor 0.8.

	periodical SDDP			
#iter	LB ($m)	CI ($m)	CI for true ($m)	gap
100	9.50	9.94, 10.29	9.85, 10.23	8.22%
200	9.65	9.92, 10.26	9.83, 10.21	6.36%
300	9.71	9.90, 10.26	9.84, 10.22	5.67%

Figure 9: Stored energy (on the left) and discounted stage cost (on the right) in average value and 0.9 quantile for the problem with/without thermal security constraint for the SAA discretized problem (on the above) and the true problem (on the bottom) for the risk neutral with discount factor 0.8

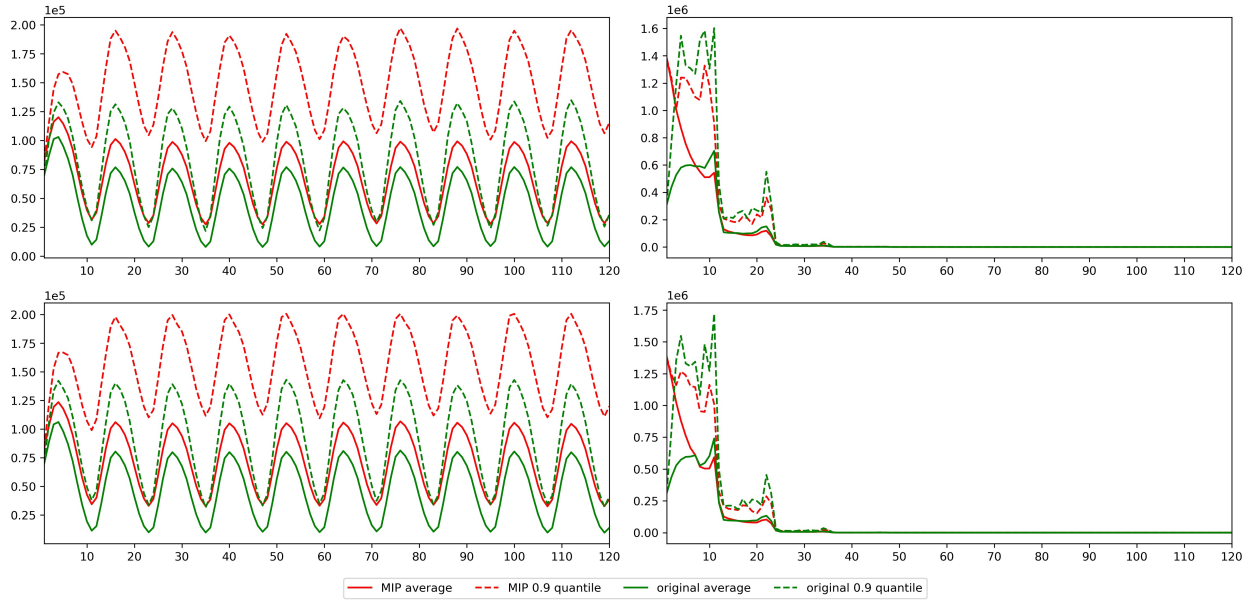
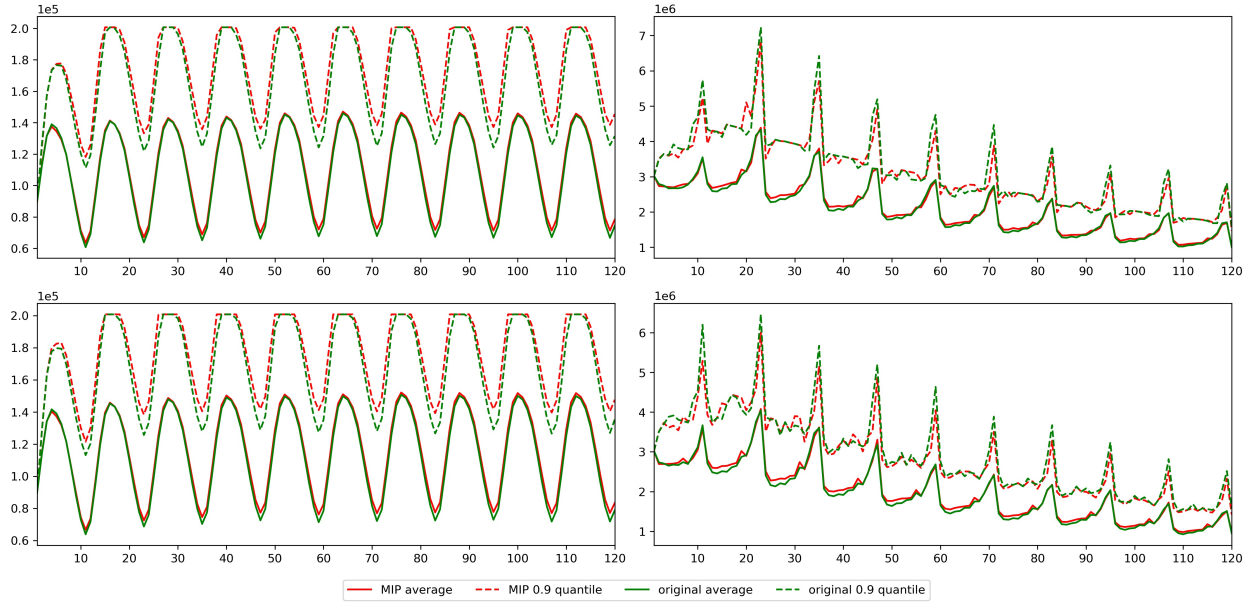


Figure 10: Stored energy (on the left) and discounted stage cost (on the right) in average value and 0.9 quantile for the problem with/without thermal security constraint for the SAA discretized problem (on the above) and the true problem (on the bottom) for the risk neutral case with discount factor 0.9906



7 Markovian setting

Let us finally mention the following possible extension to a Markovian setting, which could be a topic for future research. For the sake of simplicity consider the case of $\mathsf{m} = 1$ (see Remark 3.1). Suppose that the process ξ_t is Markovian and each ξ_t has the same marginal distribution. For example ξ_t can be stationary first order autoregressive process $\xi_t = \mu + \Phi\xi_{t-1} + \varepsilon_t$, or (after discretization) ξ_t can be a stationary homogeneous Markov chain. Let ξ and ξ' be random vectors with ξ having distribution of ξ_t, and ξ' having conditional distribution of ξ_{t+1} given ξ_t. For the Markovian process, the WB equations take the form (compare with equations (3.7))

$$\begin{gathered} \mathcal{V}(x,\xi) = \varrho_{|\xi}[V(x,\xi')], \\ V(x,\xi) = \min_{x' \in \mathcal{X}} \big\{ f(x',\xi) + \gamma \mathcal{V}(x',\xi) : B(\xi)x + A(\xi)x' = b(\xi) \big\}. \end{gathered} \tag{7.1}$$

The corresponding mapping

$$\mathfrak{T}_\xi(g)(x) := \rho_{|\xi}\left[\Psi_g(x,\xi')\right]$$

depends on ξ.

For example, consider the risk neutral case of stationary homogeneous Markov chain with the state space $\{(B^1, A^1, b^1), ..., (B^k, A^k, b^k)\}$ and probability of moving from state (B^i, A^i, b^i) to state (B^j, A^j, b^j) is p_{ij}. Then the WB equations take the form

$$\begin{gathered} V^i(x) = \min_{x' \in \mathcal{X}} \big\{ f^i(x') + \gamma \mathcal{V}^i(x') : B^i x + A^i x' = b^i \big\}, \\ \mathcal{V}^i(x) = \textstyle\sum_{j=1}^k p_{ij} V^j(x), \ i = 1, ..., k. \end{gathered} \tag{7.2}$$

Here the value function $\mathcal{V}^i(\cdot)$ should be computed for every $i = 1, ..., k$.

8 Conclusion

In conclusion, in all cases above, the proposed periodical SDDP algorithm provides similar, and sometimes more reasonable solutions as opposed to the classical SDDP for the same number of iterations, for *periodical* problems. Giving the benefit of significantly reduced computational time it brings about, the periodical SDDP algorithm can be an important substitute of the classical one. Having said that, we notice that for the discount factor that is close to 1, it is computationally expensive to evaluate policy value and thus hard to compute the upper bound. We leave it to a future research for construction of a more efficient way to compute the upper bound.

Of course the periodical SDDP algorithm applies only to problems with the periodical structure. It could be mentioned, however, that the user can choose different period lengths m for the considered problem. For instance, in the example of the Brazilian interconnected power system problem (discussed in section 6), one can use $\mathsf{m} = 24$ (two years) or even $\mathsf{m} = 60$ (five years). Still it will reduce the computational effort and will take care of the end-of-horizon effect.

Acknowledgments

The authors are indebted to Joari Paulo da Costa for constructive comments which helped to improve presentation of the manuscript.

References

[1] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

[2] R. Baucke. An algorithm for solving infinite horizon Markov dynamic programmes. *Optimization online*, 2018.

[3] D.P. Bertsekas and S.E. Shreve. *Stochastic Optimal Control, The Discrete Time Case*. Academic Press, New York, 1978.

[4] L. Ding, S. Ahmed, and A. Shapiro. A python package for multi-stage stochastic programming. *Optimization online*, 2019.

[5] H. Föllmer and A. Schied. *Stochastic Finance: An Introduction in Discrete Time*. Walter de Gruyter, Berlin, 2nd edition, 2004.

[6] P. Girardeau, V. Leclere, and A. B. Philpott. On the convergence of decomposition methods for multistage stochastic convex programs. *Mathematics of Operations Research*, 40:130–145, 2016.

[7] V. Guigues. Dual dynamic programing with cut selection: Convergence proof and numerical experiments. *European Journal of Operational Research*, 258:47–57, 2017.

[8] M.V.F. Pereira and L.M.V.G. Pinto. Multi-stage stochastic optimization applied to energy planning. *Mathematical programming*, 52(1-3):359–375, 1991.

[9] A.B. Philpott, V.L. de Matos, and E. Finardi. On solving multistage stochastic programs with coherent risk measures. *Operations Research*, 61(4):957–970, 2013.

[10] M. L. Puterman. *Markov Decision Processes.: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, 2014.

[11] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*. Springer, New York, 1998.

[12] A. Shapiro. Consistency of sample estimates of risk averse stochastic programs. *Journal of Applied Probability*, 50:533–541, 2013.

[13] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, Philadelphia, 2nd edition, 2014.

[14] A. Shapiro and A. Nemirovski. On complexity of stochastic programming problems. In V. Jeyakumar and A.M. Rubinov, editors, *Continuous Optimization: Current Trends and Applications*, pages 111–144. Springer-Verlag, 2005.

[15] A. Shapiro, W. Tekaya, J.P. da Costa, and M. Pereira Soares. Risk neutral and risk averse stochastic dual dynamic programming method. *European Journal of Operational Research*, 224:375–391, 2013.

[16] J. Warrington, N. Beuchat, and J. Lygeros. Generalized dual dynamic programming for infinite horizon problems in continuous state and action spaces. *IEEE Transactions on Automatic Control*, 2019.

[17] Jikai Zou, Shabbir Ahmed, and Xu Andy Sun. Stochastic dual dynamic integer programming. *Mathematical Programming*, 175(1-2):461–502, 2019.