Dynamic programming and dimensionality in convex stochastic control

Teemu Pennanen* Ari-Pekka Perkkiö†

January 24, 2025

Abstract

This paper studies convex stochastic control problems in the "decision-hazard-decision" form where at each stage, the system state is controlled both by predictable as well as adapted controls. Such an information structure may result in a lower dimensional system state than what is required in more traditional "decision-hazard" or "hazard-decision" formulations. We allow for general randomness and characterize optimal solutions and optimum values in terms of solutions to generalized Bellman equations. Existence of solutions to the Bellman equations is established under general conditions that do not require compactness. We also extend the problem format beyond linear system equations without increasing the dimensionality of the Bellman equations and we describe a version of the Stochastic Dual Dynamic Programming algorithm applicable to the extended format.

Keywords. Stochastic optimal control, dynamic programming, convexity

AMS subject classification codes. 93E20, 90C39, 26B25

1 Introduction

Let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $(\mathcal{F}_t)_{t=0}^T$ and consider the optimal control problem

$$\begin{aligned} &\text{minimize} \quad && E\left[\sum_{t=0}^{T-1} L_t(X_t, \check{U}_t, \hat{U}_{t+1}) + J(X_T)\right] \quad \text{over} \quad (X, \check{U}, \hat{U}) \in \mathcal{N}, \quad (1) \\ &\text{subject to} \quad && X_t = A_t X_{t-1} + \check{B}_t \check{U}_{t-1} + \hat{B}_t \hat{U}_t + W_t \quad t = 1, \dots, T \ a.s., \end{aligned}$$

*Department of Mathematics, King's College London, Strand, London, WC2R 2LS, United Kingdom, teemu.pennanen@kcl.ac.uk

†Mathematics Institute, Ludwig-Maximilian University of Munich, Theresienstr. 39, 80333 Munich, Germany, a.perkkioe@lmu.de.

where the state X_t and the controls $\check{U}_t$ and $\hat{U}_t$ take values in $\mathbb{R}^N$, $\mathbb{R}^{\check{M}}$ and $\mathbb{R}^{\hat{M}}$, respectively, A_t, $\check{B}_t$ and $\hat{B}_t$ are $\mathcal{F}_t$-measurable random matrices of appropriate dimensions, W_t are $\mathcal{F}_t$-measurable random vectors and L_t are extended real-valued functions on $\mathbb{R}^N \times \mathbb{R}^{\check{M}} \times \mathbb{R}^{\hat{M}} \times \Omega$ such that the expectation above is well-defined. The set $\mathcal{N}$ denotes the space of adapted stochastic processes, i.e. those where $(X_t, \check{U}_t, \hat{U}_t)$ is $\mathcal{F}_t$-measurable for each $t = 0, \ldots, T$. The linear constrains in (1) are called the *system equations*.

Problems of the form (1) were introduced in [2] as a unification of more common formulations of stochastic control problems where one of the control variables $\check{U}_t$ or $\hat{U}_t$ is absent; see also [7] and its references on more specific instances of control problems with similar information structures. While $\check{U}_t$ represents control decisions taken under the information available at time t only, the control $\hat{U}_{t+1}$ may depend on information available at time $t+1$ after e.g. the values of the matrices A_{t+1} and B_{t+1} in the system equations are observed. Problems with controls $\check{U}_t$ only are sometimes said to be in the "decision-hazard" format while problems with $\hat{U}_t$ only are said to be in "hazard-decision" format. Accordingly, in [2] problems of the form (1) are said to be in the "decision-hazard-decision" format.

It is possible to reformulate problem (1) in the hazard-decision format by augmenting the system state by the here-and-now control variable $\check{U}_t$; see Remark 4 below. This, however, increases the dimensionality of the dynamic programming equations by the dimension of $\check{U}_t$. Dynamic programming algorithms are often sensitive to the dimension of the system state so the state augmentation may be problematic from the computational point of view. The original formulation in (1) allows for dynamic programming equations in the original state space. Indeed, [2, Theorem 13] expresses the optimum value of (1) as $J_0(X_0)$ where the initial state X_0 is fixed and the function J_0 is given recursively by the *dynamic programming equations* (aka Bellman equations)

$$J_t(X_t) = \inf_{\check{U}_t} E_t \left[\inf_{\hat{U}_{t+1}} \{ L_t(X_t, \check{U}_t, \hat{U}_{t+1}) + J_{t+1}(A_{t+1}X_t + \check{B}_{t+1}\check{U}_t + \hat{B}_{t+1}\hat{U}_{t+1} + W_{t+1}) \} \right]. \tag{2}$$

Here and in what follows, we use the short hand notation $E_t = E^{\mathcal{F}_t}$ for the $\mathcal{F}_t$-conditional expectation of a random variable. As opposed to the T stage problem in (1), the optimization problems in the definition of $J_t(X_t)$ above are two-stage stochastic optimization problems. Theorem 13 of [2] assumed that the randomness is driven by a sequence of stagewise independent random variables and the cost functions L_t were assumed to be "lower semianalytic" in order to establish sufficient measurability so that the expectations in (2) are well-defined. When the underlying probability space is finite, all the measurability questions disappear. In practice, however, one is often faced with continuously distributed random variables.

This paper extends [2, Theorem 13] in the convex case by allowing for general randomness and providing optimality conditions and sufficient conditions for existence of solutions $(J_t)_{t=0}^T$ to the Bellman equations (2) as well as solutions

$(X,\check U,\hat U)$ to problem (1). Moreover, we will give conditions that reduce the dimensionality of the cost-to-go functions J_t with respect to the underlying randomness. The last section describes an extension of the Stochastic Dual Dynamic Programming for (1) alluded to at the end of [2, Section 6].

Much like [12] and [4], our approach is based on the theory of *normal integrands* developed for optimization problems involving integral functionals; see e.g. [11], [1], [14, Chapter 14] and the references there. Recall that a function $h:\mathbb{R}^n\times\Omega\to\overline{\mathbb{R}}$ is a *normal integrand* if the epigraphical mapping $\omega\mapsto\operatorname{epi}h(\cdot,\omega)$ is closed-valued and measurable; see [14, Chapter 14]. Given a sub-σ-algebra $\mathcal{G}$ of $\mathcal{F}$, the *$\mathcal{G}$-conditional expectation* of h, when it exists, is a $\mathcal{G}$-measurable normal integrand $E^{\mathcal{G}}h$ such that

$$(E^{\mathcal{G}}h)(x)=E[h(x)|\mathcal{G}]\ a.s.$$

for every $\mathcal{G}$-measurable $\mathbb{R}^n$-valued random variable x. Here, $E[h(x)|\mathcal{G}]$ denotes the $\mathcal{G}$-conditional expectation of the random variable $h(x)$. If h is *lower bounded* in the sense that there exists $m\in L^1$ such that

$$h(x,\omega)\geq m(\omega)\quad\forall x\in\mathbb{R}^n$$

for almost every ω, then $E^{\mathcal{G}}h$ exists and is unique; see [8, Theorem 2.13]. When $\mathcal{G}=\mathcal{F}_t$, we will use the short hand notation $E_th=E^{\mathcal{F}_t}h$.

We say that two sequences $(J_t)_{t=0}^T$ and $(\tilde V_t)_{t=0}^T$ of normal integrands solve the *Bellman equations* for problem (1) if

$$\begin{aligned}
J_T&=J,\\
\tilde V_t(X_t,\check U_t)&=\inf_{\hat U_{t+1}}\{L_t(X_t,\check U_t,\hat U_{t+1})+J_{t+1}(A_{t+1}X_t+\check B_{t+1}\check U_t+\hat B_{t+1}\hat U_{t+1}+W_{t+1})\},\\
J_t(X_t)&=\inf_{\check U_t}(E_t\tilde V_t)(X_t,\check U_t)
\end{aligned}\tag{3}$$

almost surely for every $t=0,\dots,T$ and $(X,\check U,\hat U)\in(\mathbb{R}^{N+\check M+\hat M})^{T+1}$. If the conditional expectations E_t can be expressed as "regular conditional expectations", then the last equation in (3) can be interpreted pointwise; see [8, Example 2.102]. Substituting out the normal integrand $\tilde V_t$, then shows that the normal integrands J_t in (3) satisfy (2) and, in particular, that the infimums in (2) are measurable functions. Indeed, the pointwise infimum of a normal integrand is automatically measurable; see e.g. [8, Corollary 1.23].

The three theorems below will be proved in Section 4. The first one characterizes the optimum value and optimal solutions of (1) in terms of the solutions of the Bellman equations (3). In continuous time control theory, such theorems are sometimes called "verification theorems".

Theorem 1 (Optimality principle)**.** *Let $(J_t,\tilde V_t)_{t=0}^T$ be a lower bounded convex solution of* (3)*. Then the optimum value of* (1) *coincides with that of*

$$\begin{aligned}
&\text{minimize}\quad && E\left[\sum_{s=0}^{t-1}L_s(X_s,\check U_s,\hat U_{s+1})+J_t(X_t)\right]\quad\text{over}\quad(X^t,\check U^t,\hat U^t)\in\mathcal{N}^t,\\
&\text{subject to}\quad && X_s=A_sX_{s-1}+\check B_s\check U_{s-1}+\hat B_s\hat U_s+W_s\quad s=1,\dots,t\ a.s
\end{aligned}$$

for all $t = 0, \dots, T$ *and, moreover, an* $(\bar{X}, \bar{\check{U}}, \bar{\hat{U}}) \in \mathcal{N}$ *solves* (1) *if and only if it satisfies the system equations and*

$$\begin{aligned}
\bar{X}_0 &\in \operatorname*{argmin}_{X_0} J_0(X_0) \ a.s.,\\
\bar{\check{U}}_t &\in \operatorname*{argmin}_{\check{U}_t} (E_t \tilde{V}_t)(X_t, \check{U}_t) \ a.s.\\
\bar{\hat{U}}_{t+1} &\in \operatorname*{argmin}_{\hat{U}_{t+1}} \{L_t(X_t, \check{U}_t, \hat{U}_{t+1}) + J_{t+1}(A_{t+1}X_t + \check{B}_{t+1}\check{U}_t + \hat{B}_{t+1}\hat{U}_{t+1} + W_{t+1})\} \ a.s.
\end{aligned}$$

for all $t = 0, \dots, T$.

The optimality condition in Theorem 1 characterizes the optimal solutions of (1), for each time t and state ω as solutions of a two-stage convex optimization problems. This is a major simplification over (1) where one optimizes over adapted processes $(X, \check{U}, \hat{U})$ over T stages. The above, of course, requires knowledge of the Bellman functions J_t which are normal integrands on $\mathbb{R}^N \times \Omega$. One should note, however, that while, at each stage in (1), one optimizes over $(X_t, \check{U}_t, \hat{U}_t)$, the Bellman functions J_t only depend on the system state X_t. When the problem does not depend on $\hat{U}$, Theorem 1 reduces to [8, Theorem 2.91], where the optimality condition concerning X_0 was omitted by mistake.

The following gives sufficient conditions for the existence of solutions to the Bellman equations (3). Given a normal integrand $h : \mathbb{R}^n \times \Omega \to \overline{\mathbb{R}}$, the function h^∞ obtained from h by defining $h^\infty(\cdot, \omega)$, for each $\omega \in \Omega$, as the *recession function* $h(\cdot, \omega)$ is a positively homogeneous convex normal integrand; see e.g. [8, Example 14.54a] or [8, Theorem 1.39].

Theorem 2 (Existence of solutions)**.** *Assume that the set*

$$\begin{aligned}
\{(X, \check{U}, \hat{U}) \in \mathcal{N} \mid \sum_{t=0}^{T-1} L_t^\infty(X_t, \check{U}_t, \hat{U}_{t+1}) + J^\infty(X_T) \le 0,\\
X_t = A_t X_{t-1} + \check{B}_t \check{U}_{t-1} + \hat{B}_t \hat{U}_t \ t = 1, \dots, T \ a.s.\}
\end{aligned}$$

is linear. Then (1) *has a solution* $(\bar{X}, \bar{\check{U}}, \bar{\hat{U}}) \in \mathcal{N}$ *and the Bellman equations* (3) *have a unique solution* $(J_t, \tilde{V}_t)_{t=0}^T$ *of lower bounded convex normal integrands.*

So far, we have not assumed anything about the underlying randomness in (1). Consequently, the cost-to-go functions J_t are merely $\mathcal{F}_t$-measurable normal integrands. The independence assumption made in [2], on the other hand, resulted in nonrandom functions cost-to-go functions. We can reduce the dimensionality of the Bellman equations with respect to randomness more generally as follows.

Given an $\mathbb{R}^d$-valued random variable ξ, we say that another random variable C depends on ω only through ξ if there is a measurable function $\hat{C}$ on $\mathbb{R}^d$ such that $C(\omega) = \hat{C}(\xi(\omega))$ almost surely. By the Doob-Dynkin lemma, this happens if and only if C is $\sigma(\xi)$-measurable. Similarly, we say that a normal integrand $h : \mathbb{R}^n \times \Omega \to \overline{\mathbb{R}}$ depends on ω only through ξ if there is a normal integrand $H : \mathbb{R}^n \times \mathbb{R}^d \to \overline{\mathbb{R}}$ such that $h(x,\omega) = H(x,\xi(\omega))$ for all $x \in \mathbb{R}^n$ almost surely. Here, we endow $\mathbb{R}^d$ with the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$. By [8, Corollary 1.34] such a representation exists if and only if h is a $\sigma(\xi)$-measurable normal integrand.

Theorem 3 (Dimensionality with respect to scenarios)**.** *Let $(J_t, \tilde{V}_t)_{t=0}^T$ be a lower bounded solution of* (3) *and assume that L_{t-1}, A_t, B_t and W_t depend on ω only through random variables $(\theta_{t-1}, \theta_t, \eta_t)$, where $\theta = (\theta)_{t=0}^T$ is a Markov process and $\eta = (\eta_t)_{t=0}^T$ is a sequence of random variables such that each η_t is independent of θ and of η_s for $s \neq t$. Then J_t depends on ω only through θ_t. In particular, if L_{t-1}, A_t, B_t and W_t are independent of θ, then J_t is deterministic.*

Remark 4. *Defining $V_t := E_t\tilde{V}_t$, we can write* (3) *in hazard-decision format as*

$$\begin{aligned}
V_T(X_T, \check{U}_T) &= K(X_T),\\
\tilde{V}_t(X_t, \check{U}_t) &= \inf_{\hat{U}_{t+1}, \check{U}_{t+1}} \{L_t(X_t, \check{U}_t, \hat{U}_{t+1}) + V_{t+1}(A_{t+1}X_t + \check{B}_{t+1}\check{U}_t + \hat{B}_{t+1}\hat{U}_{t+1} + W_{t+1}, \check{U}_{t+1})\},\\
V_t &= E_t\tilde{V}_t,
\end{aligned}$$

where we used the assumption that L_t is independent of $\check{U}_{t+1}$. This can be interpreted as the Bellman equations for the hazard-decision problem with state $(X_t, \check{U}_t)$, control $(\hat{U}_t, \tilde{U}_t)$ and system equations

$$\begin{bmatrix} X_{t+1} \\ \check{U}_{t+1} \end{bmatrix} = \begin{bmatrix} A_{t+1}X_t + \check{B}_{t+1}\check{U}_t + \hat{B}_{t+1}\hat{U}_{t+1} + W_{t+1} \\ \tilde{U}_t \end{bmatrix}.$$

Such a reformulation of (1) *corresponds to the hazard-decision reformulation suggested in [15] and [3] in the context of the Stochastic Dual Dynamic Programming (SDDP) algorithm. One should note that the dimensionality of the Bellman equations in the hazard-decision reformulation is $N + \check{M}$ while in the decision-hazard-decision formulation it is N. The dimension of the state may have a significant effect on the performance of e.g. the SDDP algorithm.*

2 Problems of Lagrange

It turns out that Theorems 1, 2 and 3 from the introduction extend to more general problem formats. This section studies an abstract stochastic optimization model inspired by calculus of variations; see [13] or [8] and their references. This format does not split the decision variables to states and controls as in (1) so the cost-to-go function at time t will be a function of all the decision variables chosen at time t. The format covers many more specific formats studied in stochastic programming and, in particular, in the context of the stochastic dual dynamic programming algorithm; see e.g. [9].

Let $K_t : \mathbb{R}^d \times \mathbb{R}^d \times \Omega \to \overline{\mathbb{R}}$ be convex normal integrands and consider the problem

$$\text{minimize} \quad E\left[\sum_{t=1}^{T} K_t(x_{t-1}, x_t)\right] \quad \text{over } x \in \mathcal{N} \tag{4}$$

and the corresponding Bellman equations

$$\begin{aligned} \tilde{V}_T &= 0, \\ V_t &= E_t \tilde{V}_t, \\ \tilde{V}_{t-1}(x_{t-1}) &= \inf_{x_t \in \mathbb{R}^d} \{K_t(x_{t-1}, x_t) + V_t(x_t)\}. \end{aligned} \tag{5}$$

We assume throughout that K_t are lower-bounded and proper.

Theorem 5 (Optimality principle)**.** *Let $(V_t)_{t=0}^T$ be a lower bounded convex solution of* (5)*. Then the optimum value of* (4) *is given by*

$$\inf (4) = \inf_{x^t \in \mathcal{N}^t} E\left[\sum_{s=0}^{t} (E_t K_s)(x_{s-1}, x_s) + V_t(x_t)\right]$$

for all $t = 0, \ldots, T$ and, moreover, an $\bar{x} \in \mathcal{N}$ solves (4) *if and only if*

$$\bar{x}_t \in \operatorname*{argmin}_{x_t \in \mathbb{R}^d} \{(E_t K_t)(\bar{x}_{t-1}, x_t) + V_t(x_t)\} \quad a.s.$$

for all $t = 0, \ldots, T$.

Proof. Apply [8, Theorem 2.106] to the normal integrands $\hat{K}_t(x_t, \Delta x_t, \omega) := K_t(x_t - \Delta x_t, x_t, \omega)$. □

Theorem 6 (Existence of solutions)**.** *Assume that the set*

$$\{x \in \mathcal{N} \mid \sum_{t=0}^{T} K_t^\infty(x_{t-1}, x_t) \le 0 \ a.s.\}$$

is linear. Then (4) *has a solution $\bar{x} \in \mathcal{N}$ and the Bellman equations* (5) *have a unique solution $(V_t)_{t=0}^T$ of lower bounded convex normal integrands.*

Proof. Applying [8, Theorem 2.108] to the normal integrands $\hat{K}_t(x_t, \Delta x_t, \omega) := K_t(x_t - \Delta x_t, x_t, \omega)$ proves the second claim. Combining the last statements of [8, Theorem 2.108] and [8, Theorem 2.106] proves the first claim. □

Theorem 7 (Dimensionality with respect to scenarios)**.** *Let $(V_t)_{t=0}^T$ be a lower bounded solution of the Bellman equations* (5) *and assume that K_t depends on ω only through random variables $(\theta_{t-1}, \theta_t, \eta_t)$, where $\theta = (\theta)_{t=0}^T$ is a Markov process and $\eta = (\eta_t)_{t=0}^T$ is a sequence of random variables such that each η_t is independent of θ and of η_s for $s \neq t$. Then V_t depends on ω only through θ_t. In particular, if K_t is independent of θ, then V_t is deterministic.*

Proof. The claim clearly holds for $t = T$. Assume that it holds for $t+1$. The normal integrand $\tilde{V}_{t-1}$ then depends on ω only through $(\theta_{t-1}, \theta_t, \eta_t)$. Applying [8, Theorem 2.21] with $\mathcal{G} := \mathcal{F}_t$ and $\mathcal{H} := \sigma(\theta_t)$ says that $E[\tilde{V}_t \mid \mathcal{F}_t] = E[\tilde{V}_t \mid \sigma(\theta_t)]$. In particular, V_t is $\sigma(\xi_t)$-measurable so the claim follows from [8, Corollary 1.34]. □

3 Dimensionality of the Bellman equations

An essential feature of the optimal control problem (1) is that consecutive stages are only linked through the state variable X_t. Accordingly, the Bellman functions J_t in (3) only depend on the state instead of all the decision variables of stage t as in (5). This structure generalizes as follows.

Consider again problem (4) and assume that the decision variables decompose as $x_t = (X_t, \check{U}_t, \hat{U}_t)$ and that each K_t is independent of $(\hat{U}_{t-1}, \check{U}_t)$. With a slight misuse of notation, we can then write problem (4) as

$$\text{minimize} \quad \sum_{t=1}^{T} K_t(X_{t-1}, \check{U}_{t-1}, X_t, \hat{U}_t) \quad \text{over } (X, \check{U}, \hat{U}) \in \mathcal{N} \tag{6}$$

and the corresponding Bellman equations (5) as

$$\begin{aligned} \tilde{V}_T &= 0, \\ V_t &= E_t \tilde{V}_t, \\ \tilde{V}_{t-1}(X_{t-1}, \check{U}_{t-1}, \hat{U}_{t-1}) &= \inf_{X_t, \check{U}_t, \hat{U}_t} \{K_t(X_{t-1}, \check{U}_{t-1}, X_t, \hat{U}_t) + V_t(X_t, \check{U}_t, \hat{U}_t)\}. \end{aligned} \tag{7}$$

By induction over $t = T, \dots, 0$, the functions $\tilde{V}_t$ and V_t are independent of $\hat{U}_t$. If $(V_t)_{t=0}^T$ solve the above Bellman equations, and the functions

$$J_t(X_t) := \inf_{\check{U}_t} (E_t \tilde{V}_t)(X_t, \check{U}_t)$$

are normal integrands, then they solve the *reduced Bellman equations*

$$\begin{aligned} \tilde{V}_T &= 0, \\ J_t(X_t) &= \inf_{\check{U}_t} (E_t \tilde{V}_t)(X_t, \check{U}_t), \\ \tilde{V}_{t-1}(X_{t-1}, \check{U}_{t-1}) &= \inf_{X_t, \hat{U}_t} \{K_t(X_{t-1}, \check{U}_{t-1}, X_t, \hat{U}_t) + J_t(X_t)\} \end{aligned}$$

or, equivalently,

$$\begin{aligned} J_T &= 0, \\ \tilde{V}_t(X_t, \check{U}_t) &= \inf_{X_{t+1}, \hat{U}_{t+1}} \{K_{t+1}(X_t, \check{U}_t, X_{t+1}, \hat{U}_{t+1}) + J_{t+1}(X_{t+1})\}, \\ J_t(X_t) &= \inf_{\check{U}_t} (E_t \tilde{V}_t)(X_t, \check{U}_t). \end{aligned} \tag{8}$$

Theorem 8 (Optimality principle)**.** *Let $(J_t, \tilde{V}_t)_{t=0}^T$ be a lower bounded convex solution of* (8)*. Then the optimum value of* (6) *is given by*

$$\inf (6) = \inf_{(X,\check{U},\hat{U})^t \in \mathcal{N}^t} E\left[\sum_{s=1}^t K_s(X_{s-1}, \check{U}_{s-1}, X_s, \hat{U}_s) + J_t(X_t)\right]$$

for all $t = 0, \ldots, T$ and, moreover, an $(\bar{X}, \bar{\check{U}}, \bar{\hat{U}}) \in \mathcal{N}$ solves (6) *if and only if*

$$\begin{aligned} \bar{X}_0 &\in \operatorname*{argmin}_{X_0} J_0(X_0) \ a.s., \\ \bar{\check{U}}_t &\in \operatorname*{argmin}_{\check{U}_t} (E_t \tilde{V}_t)(\bar{X}_t, \check{U}_t) \ a.s., \\ (\bar{X}_{t+1}, \bar{\hat{U}}_{t+1}) &\in \operatorname*{argmin}_{X_{t+1}, \hat{U}_{t+1}} \{K_{t+1}(\bar{X}_t, \bar{\check{U}}_t, X_{t+1}, \hat{U}_{t+1}) + J_{t+1}(X_{t+1})\} \ a.s. \end{aligned}$$

for all $t = 0, \ldots, T$.

Theorem 9 (Existence of solutions)**.** *Assume that the set*

$$\{(X, \check{U}, \hat{U}) \in \mathcal{N} \mid \sum_{t=0}^T K_t^\infty(X_{t-1}, \check{U}_{t-1}, X_t, \hat{U}_t) \leq 0 \ a.s.\}$$

is linear. Then (6) *has a solution $(\bar{X}, \bar{\check{U}}, \bar{\hat{U}}) \in \mathcal{N}$ and the Bellman equations* (8) *have a unique solution $(\tilde{V}_t, J_t)_{t=0}^T$ of lower bounded convex normal integrands.*

Proof. By Theorem 6, (6) has a solution and the Bellman equations (7) has a solution $(V_t)_{t=0}^T$ of lower bounded convex normal integrands. Applying [8, Theorem 2.108] to the normal integrands

$$\hat{K}_t(x_t, \Delta x_t, \omega) := K_t(X_t - \Delta X_t, \check{U}_t - \Delta \check{U}_t, X_t, \hat{U}_t, \omega)$$

we get that the measurable mappings

$$\omega \mapsto \{(X_t, \check{U}_t, \hat{U}_t) \in \mathbb{R}^{N+\check{M}+\hat{M}} \mid K_t^\infty(0, 0, X_t, \hat{U}_t, \omega) + V_t^\infty(X_t, \check{U}_t, \omega) \leq 0\}$$

are almost surely linear-valued. Setting $X_t = 0$ and $\hat{U}_t = 0$ shows that the mapping

$$\omega \mapsto \{\check{U}_t \in \mathbb{R}^{\check{M}} \mid V_t^\infty(0, \check{U}_t, \omega) \leq 0\}$$

is linear-valued as well. By [8, Theorem 1.40], the function J_t is thus a convex normal integrand. Lower boundedness of V_t implies that of J_t. □

The following is a direct consequence of Theorem 7.

Theorem 10 (Dimensionality with respect to scenarios)**.** *Let $(J_t, \tilde{V}_t)_{t=0}^T$ be a lower bounded solution of the Bellman equations* (8) *and assume that K_t depends on ω only through random variables $(\theta_{t-1}, \theta_t, \eta_t)$, where $\theta = (\theta)_{t=0}^T$ is a Markov process and $\eta = (\eta_t)_{t=0}^T$ is a sequence of random variables such that each η_t is independent of θ and of η_s for $s \neq t$. Then J_t depends on ω only through θ_t. In particular, if K_t is independent of θ, then J_t is deterministic.*

4 Proofs of Theorems 1, 2 and 3

The optimal control problem in the introduction of this paper is an instance of (6) with the normal integrands K_t defined, for $t = 1, \dots T-1$, by

$$K_t(X_{t-1}, \check{U}_{t-1}, X_t, \hat{U}_t, \omega) = L_{t-1}(X_{t-1}, \check{U}_{t-1}, \hat{U}_t, \omega)$$

if $X_t = A_t(\omega)X_{t-1} + \check{B}_t(\omega)\check{U}_{t-1} + \hat{B}_t(\omega)\hat{U}_t + W_t(\omega)$ and

$$K_t(X_{t-1}, \check{U}_{t-1}, X_t, \hat{U}_t, \omega) = +\infty$$

otherwise, while

$$K_T(X_{T-1}, \check{U}_{T-1}, X_T, \hat{U}_T, \omega) = L_{T-1}(X_{T-1}, \check{U}_{T-1}, \hat{U}_T, \omega) + K(X_T, \omega)$$

if $X_T = A_T(\omega)X_{T-1} + \check{B}_T(\omega)\check{U}_{T-1} + \hat{B}_T(\omega)\hat{U}_T + W_T(\omega)$ and

$$K_T(X_{T-1}, \check{U}_{T-1}, X_T, \hat{U}_T, \omega) = +\infty$$

otherwise. In this case, the Bellman equations (8) become

$$\begin{aligned}
J_T &= K, \\
\tilde{V}_t(X_t, \check{U}_t) &= \inf_{\hat{U}_{t+1}} \{L_t(X_t, \check{U}_t, \hat{U}_{t+1}) + J_{t+1}(A_{t+1}X_t + \check{B}_{t+1}\check{U}_t + \hat{B}_{t+1}\hat{U}_{t+1} + W_{t+1})\}, \\
J_t(X_t) &= \inf_{\check{U}_t} (E_t\tilde{V}_t)(X_t, \check{U}_t),
\end{aligned}$$

which are the Bellman equations (3) for the optimal control problem (1). Theorem 1, Theorem 2 and Theorem 3 thus follow directly from Theorem 8, Theorem 9 and Theorem 10, respectively.

5 Stochastic dual dynamic programming

Stochastic dual dynamic programming (SDDP) is an iterative algorithm for constructing approximate solutions to Bellman equations. SDDP was proposed in [9] under the assumption that the randomness is driven by independent noises (no Markov process θ) so that the Bellman functions are deterministic; see Theorem 3. Moreover, [9] did not split the decision variables into states and controls so the cost-to-go function at time t was a function of all the decision variables chosen at time t. The SDDP algorithm was extended to linear problems with non-independent noises in [5]; see also [10, 6] and their references.

This section gives a further extension of the SDDP algorithm that applies to the general decision-hazard-decision format of Section 3 that allows for nonlinear constraints and objectives, Markovian randomness and cost-to-go functions that only depend on the linking variables X_t.

Assume that K_t are as in Section 3 so that the Bellman functions J_t depend on the decision variables only throught the system state X_t. We will also assume, as in Theorem 10, that K_t depends on ω only through random variables

$(\theta_{t-1},\theta_t,\eta_t)$, where $\theta=(\theta)_{t=0}^T$ is a Markov process and $\eta=(\eta_t)_{t=0}^T$ is a sequence of random variables such that each η_t is independent of ξ and of η_s for $s\neq t$. By Theorem 10, each Bellman function J_t then depends on ω only through θ_t. Furthermore, we will assume that $\xi:=(\theta,\eta)$ is *finitely supported* so that the conditional expectations below be expressed as the finite sums.

0. **Initialization:** Choose convex (polyhedral) lower-approximations J_t^0 of the cost-to-go functions J_t and set $k=0$.
1. **Forward pass:** Sample a path ξ^k of ξ and define X_t^k for $t=0,\dots,T$ by

$$
\begin{aligned}
X_0^k &\in \operatorname{argmin} J_k^0,\\
\check U_t^k &\in \operatorname*{argmin}_{\check U_t\in\mathbb{R}^{\check M}} E\left[\inf_{(X_{t+1},\hat U_{t+1})\in\mathbb{R}^N\times\mathbb{R}^{\hat M}}\{K_{t+1}(X_t^k,\check U_t,X_{t+1},\hat U_{t+1},\xi)+J_{t+1}^k(X_{t+1},\theta_{t+1})\}\middle|\theta_t^k\right],\\
(X_{t+1}^k,\hat U_{t+1}^k) &\in \operatorname*{argmin}_{(X_{t+1},\hat U_{t+1})\in\mathbb{R}^N\times\mathbb{R}^{\hat M}}\{K_{t+1}(X_t^k,\check U_t^k,X_{t+1},\hat U_{t+1},\xi^k)+J_{t+1}^k(X_{t+1},\theta_{t+1}^k)\}.
\end{aligned}
$$

2. **Backward pass:** Let $J_T^{k+1}:=0$ and, for $t=T-1,\dots,0$, compute

$$
\begin{aligned}
\tilde J_t^{k+1}(X_t^k,\theta_t^k) &:= \inf_{\check U_t\in\mathbb{R}^{\check M}} E\left[\inf_{(X_{t+1},\hat U_{t+1})\in\mathbb{R}^N\times\mathbb{R}^{\hat M}}\{K_{t+1}(X_t^k,\check U_t,X_{t+1},\hat U_{t+1},\xi)+J_{t+1}^{k+1}(X_{t+1},\theta_{t+1})\}\middle|\theta_t^k\right],\\
V_t^{k+1} &\in \partial\tilde J_t^{k+1}(X_t^k,\theta_t^k),
\end{aligned}
$$

   and let

$$
J_t^{k+1}(X_t,\theta_t^k):=\max\{J_t^k(X_t,\theta_t^k),\tilde J_t^{k+1}(X_t^k,\theta_t^k)+V_t^{k+1}\cdot(X_t-X_t^k)\}\quad\forall X_t\in\mathbb{R}^N.
$$

   Set $k:=k+1$ and go to 1.

The SDDP algorithm above, breaks the original T-stage optimization problem into a series of two-stage convex stochastic optimization problems. If the θ_t^k-conditional distribution of ξ_{t+1} is supported by M_{t+1} scenarios, then the second stage has M_{t+1} copies of the variables $(X_{t+1},\hat U_{t+1})$. The optimal solution $(X_{t+1}^k,\hat U_{t+1}^k)$ in the forward pass is the optimal second stage solution corresponding to the scenario ξ^k. In the hazard-decision format where the "here-and-now" variables $\check U_t$ are absent, the second stage optimization can be done separately for each scenario and in the forward pass, only the scenario k needs to be treated.

The subgradients V_t^{k+1} of $\tilde J_t^{k+1}$ at X_t^k in the backward pass are provided by standard optimization packages provided they can accommodate the functions K_{t+1}. This of course depends on what the functions are in practice. Even in the abstract setting above, one can always introduce a dummy variable $\tilde X_t$ and add a constraint $\tilde X_t=X_t^k$ and then, the Lagrange multipliers of this constraint are subgradients of $\tilde J_t^{k+1}$ at X_t^k.

In the optimal control format of the introduction and Section 4, one can use the system equations to substitute out the state variables X_{t+1} so the SDDP algorithm above can be written as follows.

0. **Initialization:** Choose convex (polyhedral) lower-approximations J_t^0 of the cost-to-go functions J_t and set $k=0$.

1. **Forward pass:** Sample a path ξ^k of ξ and define X_t^k for $t = 0, \dots, T$ by

$$X_0^k \in \operatorname{argmin} J_k^0,$$

$$\check{U}_t^k \in \operatorname*{argmin}_{\check{U}_t \in \mathbb{R}^{\check{M}}} E\left[\inf_{\hat{U}_{t+1} \in \mathbb{R}^{\hat{M}}} \{L_t(X_t^k, \check{U}_t, \hat{U}_{t+1}) + J_{t+1}(A_{t+1}X_t^k + \check{B}_{t+1}\check{U}_t + \hat{B}_{t+1}\hat{U}_{t+1} + W_{t+1})\} \middle| \theta_t^k\right],$$

$$\hat{U}_{t+1}^k \in \operatorname*{argmin}_{\hat{U}_{t+1} \in \mathbb{R}^{\hat{M}}} \{L_t(X_t^k, \check{U}_t^k, \hat{U}_{t+1}) + J_{t+1}(A_{t+1}X_t + \check{B}_{t+1}\check{U}_t^k + \hat{B}_{t+1}\hat{U}_{t+1} + W_{t+1})\}.$$

2. **Backward pass:** Let $J_T^{k+1} := 0$ and, for $t = T-1, \dots, 0$, compute

$$\tilde{J}_t^{k+1}(X_t^k, \theta_t^k) := \inf_{\check{U}_t \in \mathbb{R}^{\check{M}}} E\left[\inf_{\hat{U}_{t+1} \in \mathbb{R}^{\hat{M}}} \{L_t(X_t^k, \check{U}_t, \hat{U}_{t+1}) + J_{t+1}(A_{t+1}X_t^k + \check{B}_{t+1}\check{U}_t + \hat{B}_{t+1}\hat{U}_{t+1} + W_{t+1})\} \middle| \theta_t^k\right],$$

$$V_t^{k+1} \in \partial \tilde{J}_t^{k+1}(X_t^k, \theta_t^k)$$

and let

$$J_t^{k+1}(X_t, \theta_t^k) := \max\{J_t^k(X_t, \theta_t^k), \tilde{J}_t^{k+1}(X_t^k, \theta_t^k) + V_t^{k+1} \cdot (X_t - X_t^k)\} \quad \forall X_t \in \mathbb{R}^N.$$

Set $k := k+1$ and go to 1.

Again, in the hazard-hecision format, the second stage optimizations in the above two-stage problems can be done scenariowise. Moreover, in that case, the gradient in the backward pass is given by averaging the gradients of the infimum values with respect to $\hat{U}_{t+1}$. Note also that in the decision-hazard format where the "wait-and-see" variables $\hat{U}_t$ are absent, the optimization problems in the above control format of the SDDP algorithm become static problems of minimizing the expectations.

References

[1] J.-M. Bismut. Intégrales convexes et probabilités. *J. Math. Anal. Appl.*, 42:639–673, 1973.

[2] P Carpentier, J.-P. Chancelier, M. De Lara, T. Martin, and T. Rigaut. Time block decomposition of multistage stochastic optimization problems. *J. Convex Anal.*, 30(2):627–658, 2023.

[3] O. Dowson. The policy graph decomposition of multistage stochastic programming problems. *Networks*, 76(1):3–23, 2020.

[4] I. V. Evstigneev. Measurable selection and dynamic programming. *Math. Oper. Res.*, 1(3):267–272, 1976.

[5] G. Infanger and D. P. Morton. Cut sharing for multistage stochastic linear programs with interstage dependency. *Math. Programming*, 75(2):241–256, 1996.

[6] N. Löhndorf and A. Shapiro. Modeling time-dependent randomness in stochastic dual dynamic programming. *European Journal of Operational Research*, 273(2):650–661, 2019.

[7] C. Martinez Parra, M. De Lara, J.-P. Chancelier, P. Carpentier, J.-M. Janin, and M. Ruiz. A two-timescale decision-hazard-decision formulation for storage usage values calculation. *arXiv preprint arXiv:2408.17113*, 2024.

[8] T. Pennanen and A.-P. Perkkiö. *Convex Stochastic Optimization*. Springer Cham, 2024.

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

[10] A.B. Philpott and V.L. de Matos. Dynamic sampling algorithms for multi-stage stochastic programs with risk aversion. *European Journal of Operational Research*, 218(2):470–483, 2012.

[11] R. T. Rockafellar. Integrals which are convex functionals. *Pacific J. Math.*, 24:525–539, 1968.

[12] R. T. Rockafellar and R. J.-B. Wets. Nonanticipativity and L^1-martingales in stochastic optimization problems. *Math. Programming Stud.*, (6):170–187, 1976. Stochastic systems: modeling, identification and optimization, II (Proc. Sympos., Univ Kentucky, Lexington, Ky., 1975).

[13] R. T. Rockafellar and R. J.-B. Wets. Deterministic and stochastic optimization problems of Bolza type in discrete time. *Stochastics*, 10(3-4):273–312, 1983.

[14] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1998.

[15] A. Street, D. Valladão, A. Lawson, and A. Velloso. Assessing the cost of the hazard-decision simplification in multistage stochastic hydrothermal scheduling. *Applied Energy*, 280:115939, 2020.