



HAL
open science

Time Blocks Decomposition of Multistage Stochastic Optimization Problems

Pierre Carpentier, Jean-Philippe Chancelier, Michel de Lara, Thomas Martin,
Tristan Rigaut

► **To cite this version:**

Pierre Carpentier, Jean-Philippe Chancelier, Michel de Lara, Thomas Martin, Tristan Rigaut. Time Blocks Decomposition of Multistage Stochastic Optimization Problems. 2022. hal-01757113v3

HAL Id: hal-01757113

<https://hal.science/hal-01757113v3>

Preprint submitted on 11 May 2022 (v3), last revised 25 Apr 2023 (v6)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Time Blocks Decomposition of Multistage Stochastic Optimization Problems

Pierre Carpentier*, Jean-Philippe Chancelier†, Michel De Lara‡,
Thomas Martin†, Tristan Rigaut‡

May 11, 2022

Abstract

Multistage stochastic optimization problems are, by essence, complex as their solutions are indexed both by stages and by uncertainties. Their large scale nature makes decomposition methods appealing, like dynamic programming which is a sequential decomposition using a state variable defined at all stages. In this paper, we introduce the notion of state reduction by time blocks, that is, at stages that are not necessarily all the original stages. Then, we prove a reduced dynamic programming equation. We position our result with respect to the most well-known mathematical frameworks for dynamic programming. We illustrate our contribution by showing its potential for applied problems with two time scales.

Keywords: multistage stochastic optimization, time blocks decomposition, dynamic programming

1 Introduction

Solutions of multistage stochastic optimization problems are indexed both by time and by uncertainties. This makes such problems complex. However, their structure makes decomposition methods appealing to solve them. The most common approaches are time decomposition (state-based resolution methods), like stochastic dynamic programming, in stochastic optimal control, and scenario decomposition, like progressive hedging, in stochastic programming. On the one hand, stochastic programming deals with an underlying random process taking a finite number of values, called scenarios [12]. Solutions are indexed by a scenario tree, the size of which increases exponentially with the number of stages (hence generally a

*UMA, ENSTA Paris, Institut Polytechnique de Paris, Palaiseau, France

†CERMICS, Ecole des Ponts ParisTech, Marne-la-Vallée, France

‡Efficacity, Marne-la-Vallée, France

few stages in practice). However, to overcome this obstacle, stochastic programming takes advantage of scenario decomposition methods (progressive hedging [11]). On the other hand, stochastic control deals with a state model driven by a white noise, that is, the noise is made of a sequence of independent random variables. Under such assumptions, stochastic dynamic programming is able to handle many stages, as it offers reduction of the search for a solution among state feedbacks (instead of functions of the past noise) [2, 9].

In a word, dynamic programming is good at handling multiple stages — but at the price of assuming that noises are stagewise independent — whereas stochastic programming does not require such assumption, but can only handle a few stages. Could we take advantage of both methods? Is there a way to apply stochastic dynamic programming at a slow time scale — a scale at which noise would be statistically independent — crossing over fast time scale optimization problems where independence would not hold? This question is one of the motivations of this paper, and we indeed provide a method to decompose multistage stochastic optimization problems by time blocks. This decomposition method and the main result are, mathematically speaking, quite natural, but the main difficulty is notational. Indeed, the rigorous formulation of multistage stochastic optimization problems on so-called history spaces requires heavy notation.

Although specialists in stochastic optimal control and dynamic programming will find the results as natural and non surprising, or as part of folklore, the fact is that we have not been able to find references that treat the case of a state defined only at a subset of stages. This is why we set out to write this paper, without any real theoretical ambition, but with the objective that this result be established and can be used for applications using several forms of decomposition¹

The methodology developed in this paper has been successfully applied to a multistage stochastic optimization problem involving several millions time steps, namely a battery management problem over 20 years involving both the battery operating (with a fast time step of 1 minute) and the battery replacement (with a slow time step of one day) [10]. It is assumed that the vectors of noises (energy demand minus renewable energy production) are independent day by day, so that we are able to write the dynamic programming equations at the slow time scale for this two time scales optimization problem. Then we use decomposition techniques to obtain lower and upper bounds for the Bellman value functions: the corresponding approximated value functions are also computed by backward recursion, involving intraday costs (fast time scale) which are computable offline. Finally, taking into account some periodicity properties in the computation of intraday costs allows to solve the problem using a reasonable CPU time.

The paper is organized as follows. In Sect. 2, we present the standard approaches to solve, by dynamic programming, a stochastic optimal control problem formulated in discrete time. In Sect. 3, we revisit the notion of “state” by defining state reduction by time blocks — that is, at stages that are not necessarily all the original stages — and then we prove a reduced

¹The starting point of our reflections on this subject were conversations that three of us held with Roger Wets in Bogota in 2013. We discussed the interest and the way of mixing the techniques of scenario trees (to be able to take into account correlated noises) with the techniques of dynamic programming (to have a vision of the optimal future costs).

dynamic programming equation. In Sect. 4, we illustrate our contribution by showing its potential for applied problems with two time scales, as the crude oil procurement problem. We relegate technical results in Appendix A.

2 Stochastic dynamic programming with histories

In §2.1, we recall standard approaches to solve, by dynamic programming, a stochastic optimal control problem formulated in discrete time. We emphasize that, in all of these approaches, either a state is given for all times or no state is given. We highlight that our approach is intermediate, in that a state will possibly be obtained, but only at certain times. In §2.2, we formulate multistage stochastic optimization problems over the so-called history space, with history feedbacks, and we obtain a general dynamic programming equation.

2.1 Background on stochastic dynamic programming

We first recall the notion of stochastic kernel, used in the modeling of stochastic control problems. Let $(\mathbb{X}, \mathcal{X})$ and $(\mathbb{Y}, \mathcal{Y})$ be two measurable spaces. A *stochastic kernel* from $(\mathbb{X}, \mathcal{X})$ to $(\mathbb{Y}, \mathcal{Y})$ is a function $\rho : \mathbb{X} \times \mathcal{Y} \rightarrow [0, 1]$ such that, for any $Y \in \mathcal{Y}$, the function $\rho(\cdot, Y) : \mathbb{X} \rightarrow [0, 1]$ is \mathcal{X} -measurable and, for any $x \in \mathbb{X}$, the function $\rho(x, \cdot) : \mathcal{Y} \rightarrow [0, 1]$ is a probability measure. By a slight abuse of notation, a stochastic kernel is also denoted as a mapping $\rho : \mathbb{X} \rightarrow \Delta(\mathbb{Y})$ from the measurable space $(\mathbb{X}, \mathcal{X})$ towards the space $\Delta(\mathbb{Y})$ of probability measures over $(\mathbb{Y}, \mathcal{Y})$, with the property that the function $x \in \mathbb{X} \mapsto \int_Y \rho(dy | x)$ is measurable for any $Y \in \mathcal{Y}$.

We now sketch the most classical frameworks for stochastic dynamic programming in discrete time. We use the notation $\llbracket r, s \rrbracket = \{r, r + 1, \dots, s - 1, s\}$ for any two integers r, s such that $r \leq s$. We will also use the shorter notation $r:s = \llbracket r, s \rrbracket$, for example in subscripts as in $h_{r:s}$. In what follows, $t_0 \in \mathbb{N}$ and $T \in \mathbb{N}^*$ are two integers such that $t_0 < T$.

Witsenhausen approach The most general stochastic dynamic programming principle is sketched by Witsenhausen at the end of [17]. However, we do not detail it as its formalism is too far from the following ones, though we will touch the subject when we discuss Yüksel's approach below. We present here what Witsenhausen calls an optimal stochastic control problem in *standard form* (see [15]). The ingredients are the following:

1. time $t \in \llbracket t_0, T \rrbracket$ is discrete and runs among a finite set of consecutive integers;
2. $(\mathbb{X}_{t_0}, \mathcal{X}_{t_0})$ (Nature), $(\mathbb{X}_{t_0+1}, \mathcal{X}_{t_0+1}), \dots, (\mathbb{X}_T, \mathcal{X}_T)$ (state spaces) are measurable spaces;
3. $(\mathbb{U}_{t_0}, \mathcal{U}_{t_0}), \dots, (\mathbb{U}_{T-1}, \mathcal{U}_{T-1})$ are measurable spaces (control spaces);
4. \mathcal{J}_t is a subfield of \mathcal{X}_t , for $t \in \llbracket t_0, T - 1 \rrbracket$ (information);
5. $f_t : (\mathbb{X}_t \times \mathbb{U}_t, \mathcal{X}_t \otimes \mathcal{U}_t) \rightarrow (\mathbb{X}_{t+1}, \mathcal{X}_{t+1})$ is measurable, for $t \in \llbracket t_0, T - 1 \rrbracket$ (dynamics);

6. π_{t_0} is a probability on $(\mathbb{X}_{t_0}, \mathcal{X}_{t_0})$;
7. $j : (\mathbb{X}_T, \mathcal{X}_T) \rightarrow \mathbb{R}$ is a measurable function (criterion).

With these ingredients, Witsenhausen formulates a stochastic optimization problem, whose solutions are to be searched among adapted feedbacks, namely $\lambda_t : (\mathbb{X}_t, \mathcal{X}_t) \rightarrow (\mathbb{U}_t, \mathcal{U}_t)$ with the property that $\lambda_t^{-1}(\mathcal{U}_t) \subset \mathcal{J}_t$ for all $t \in \llbracket t_0, T-1 \rrbracket$. Then, he establishes a dynamic programming equation, where the Bellman functions are function of the (unconditional) distribution of the original state $x_t \in \mathbb{X}_t$, and where the minimization is done over adapted feedbacks. The main objective of Witsenhausen is to establish a dynamic programming equation for nonclassical information patterns.

Evstigneev approach The ingredients of the approach developed in [5] are the following:

1. time $t \in \llbracket t_0, T \rrbracket$ is discrete and runs among a finite set of consecutive integers;
2. $(\mathbb{U}_{t_0}, \mathcal{U}_{t_0}), \dots, (\mathbb{U}_{T-1}, \mathcal{U}_{T-1})$ are measurable spaces (control spaces);
3. (Ω, \mathcal{F}) is a measurable space (Nature);
4. $\{\mathcal{F}_t\}_{t \in \llbracket t_0, T-1 \rrbracket}$ is a filtration of \mathcal{F} (information);
5. \mathbb{P} is a probability on (Ω, \mathcal{F}) ;
6. $j : (\Omega \times \prod_{t \in \llbracket t_0, T-1 \rrbracket} \mathbb{U}_t, \mathcal{F} \otimes \bigotimes_{t \in \llbracket t_0, T-1 \rrbracket} \mathcal{U}_t) \rightarrow \mathbb{R}$ is a measurable function (criterion).

With these ingredients, Evstigneev formulates a stochastic optimization problem, whose solutions are to be searched among adapted processes, namely random processes with values in $\prod_{t \in \llbracket t_0, T-1 \rrbracket} \mathbb{U}_t$ and adapted to the filtration $\{\mathcal{F}_t\}_{t \in \llbracket t_0, T-1 \rrbracket}$. Then, he establishes a dynamic programming equation, where the Bellman function at time t is an \mathcal{F}_t -integrand depending on controls up to time t (random variables) and where the minimization is done over \mathcal{F}_t -measurable random variables at time t . The main objective of Evstigneev is to establish an existence theorem for an optimal adapted process (under proper technical assumptions, especially on the objective function j , that we do not detail here). Notice that there is no notion of state variable.

Puterman approach The ingredients of the approach developed in [9, Sect. 2.1] are the following:

1. time $t \in \llbracket t_0, T \rrbracket$ is discrete and runs among a finite set of consecutive integers;
2. $(\mathbb{X}_{t_0}, \mathcal{X}_{t_0}), \dots, (\mathbb{X}_T, \mathcal{X}_T)$ are measurable spaces (state spaces);
3. $(\mathbb{U}_{t_0}, \mathcal{U}_{t_0}), \dots, (\mathbb{U}_{T-1}, \mathcal{U}_{T-1})$ are measurable spaces (control spaces);
4. $\nu_{t:t+1} : \mathbb{X}_t \times \mathbb{U}_t \rightarrow \Delta(\mathbb{X}_{t+1})$ is a stochastic kernel, for $t \in \llbracket t_0, T-1 \rrbracket$ (transitions);

5. $L_t : \mathbb{X}_t \times \mathbb{U}_t \rightarrow \mathbb{R}$, for $t \in \llbracket t_0, T - 1 \rrbracket$, and $K : \mathbb{X}_T \rightarrow \mathbb{R}$, are measurable functions (instantaneous and final costs).

With these ingredients, Puterman formulates a stochastic optimization problem with a time additive cost function over given state and control spaces, whose solutions are to be searched among history feedbacks, namely sequences of mappings $\mathbb{X}_{t_0} \times \prod_{s=t_0}^{t-1} (\mathbb{U}_s \times \mathbb{X}_{s+1}) \rightarrow \mathbb{U}_t$. Then, he establishes a dynamic programming equation, where the Bellman functions are function of the history $h_t \in \mathbb{X}_{t_0} \times \prod_{s=t_0}^{t-1} (\mathbb{U}_s \times \mathbb{X}_{s+1})$. He identifies cases where no loss of optimality results from reducing the search to Markovian feedbacks $\mathbb{X}_t \rightarrow \mathbb{U}_t$. In such cases, the Bellman functions are function of the state $x_t \in \mathbb{X}_t$, and the minimization in the dynamic programming equation is done over controls $u_t \in \mathbb{U}_t$. The main objective of Puterman is to explore infinite horizon criteria, average reward criteria, the continuous time case, and to present many examples.

Hernández-Lerma and Lasserre approach The ingredients of the approach developed in [6, §2.2, §3.2, §3.3] are the following:

1. time $t \in \llbracket t_0, T \rrbracket$ is discrete and runs among a finite set of consecutive integers;
2. $(\mathbb{X}_{t_0}, \mathcal{X}_{t_0}), \dots, (\mathbb{X}_T, \mathcal{X}_T)$ are Borel spaces (state spaces);
3. $(\mathbb{U}_{t_0}, \mathcal{U}_{t_0}), \dots, (\mathbb{U}_{T-1}, \mathcal{U}_{T-1})$ are Borel spaces (control spaces); there are also feasible state-dependent control constraints that we do not present here;
4. $\nu_{t:t+1} : \mathbb{X}_t \times \mathbb{U}_t \rightarrow \Delta(\mathbb{X}_{t+1})$, for $t \in \llbracket t_0, T - 1 \rrbracket$, are Borel-measurable stochastic kernels (transitions);
5. $L_t : \mathbb{X}_t \times \mathbb{U}_t \rightarrow \mathbb{R}$, for $t \in \llbracket t_0, T - 1 \rrbracket$, and $K : \mathbb{X}_T \rightarrow \mathbb{R}$ are Borel-measurable functions (instantaneous and final costs).

With these ingredients, Hernández-Lerma and Lasserre formulate a stochastic optimization problem with a time additive cost function over given state and control spaces. They introduce the “canonical construction” where the history at time t consists in the states and the controls prior to t . Then, they study optimization problems whose solutions (policies) are to be searched among history feedbacks (or randomized history feedbacks), namely sequences of mappings $\mathbb{X}_{t_0} \times \prod_{s=t_0}^{t-1} (\mathbb{U}_s \times \mathbb{X}_{s+1}) \rightarrow \mathbb{U}_t$. They identify cases where no loss of optimality results from reducing the search to (relaxed) Markovian feedbacks $\mathbb{X}_t \rightarrow \mathbb{U}_t$. Then, they establish a dynamic programming equation, where the Bellman functions are function of the state $x_t \in \mathbb{X}_t$, and where the minimization is done over controls $u_t \in \mathbb{U}_t$. For finite horizon problems, the mathematical challenge is to set up a mathematical framework — the Borel assumptions plus additional topological ones presented in [6, §3.3] — for which optimal policies exists. The main objective of [6] is to offer a unified and comprehensive treatment of discrete-time Markov control processes, with emphasis on the case of Borel state and control spaces, and possibly unbounded costs and noncompact control constraint sets.

Bertsekas and Shreve approach The ingredients of the approach developed in [3] (more precisely in [3, Definition 10.1]) are the following:

1. time $t \in \llbracket t_0, T \rrbracket$ is discrete and runs among a finite set of consecutive integers;
2. $(\mathbb{X}_{t_0}, \mathcal{X}_{t_0}), \dots, (\mathbb{X}_T, \mathcal{X}_T)$ are Borel spaces (state spaces);
3. $(\mathbb{U}_{t_0}, \mathcal{U}_{t_0}), \dots, (\mathbb{U}_{T-1}, \mathcal{U}_{T-1})$ are Borel spaces (control spaces); there are also feasible state-dependent control constraints that we do not present here;
4. $(\mathbb{W}_{t_0}, \mathcal{W}_{t_0}), \dots, (\mathbb{W}_T, \mathcal{W}_T)$ are Borel spaces (noise);
5. $f_t : (\mathbb{X}_t \times \mathbb{U}_t \times \mathbb{W}_t, \mathcal{X}_t \otimes \mathcal{U}_t \otimes \mathcal{W}_t) \rightarrow (\mathbb{X}_{t+1}, \mathcal{X}_{t+1})$, for $t \in \llbracket t_0, T-1 \rrbracket$, are Borel-measurable mappings (dynamics);
6. $\rho_{t:t+1} : \mathbb{X}_t \times \mathbb{U}_t \rightarrow \Delta(\mathbb{W}_{t+1})$, for $t \in \llbracket t_0, T-1 \rrbracket$, are Borel-measurable stochastic kernels (noise distributions);
7. $L_t : \mathbb{X}_t \times \mathbb{U}_t \rightarrow \mathbb{R}$, for $t \in \llbracket t_0, T-1 \rrbracket$, and $K : \mathbb{X}_T \rightarrow \mathbb{R}$ are lower semianalytic functions (instantaneous and final costs).

With these ingredients, Bertsekas and Shreve formulate a stochastic optimization problem with a time additive cost function over given state spaces, control spaces and uncertainty spaces. They introduce the notion of history at time t which consists in the states and the controls prior to t and study optimization problems whose solutions (policies) are to be searched among history feedbacks (or relaxed history feedbacks), namely sequences of mappings from history space $\mathbb{X}_{t_0} \times \prod_{s=t_0}^{t-1} (\mathbb{U}_s \times \mathbb{X}_{s+1}) \rightarrow \mathbb{U}_t$. They identify cases where no loss of optimality results from reducing the search to (relaxed) Markovian feedbacks $\mathbb{X}_t \rightarrow \mathbb{U}_t$. Then, they establish a dynamic programming equation, where the Bellman functions are function of the state $x_t \in \mathbb{X}_t$, and where the minimization is done over controls $u_t \in \mathbb{U}_t$. For finite horizon problems, the mathematical challenge is to set up a mathematical framework (the Borel assumptions) for which optimal policies exists. The main objective of Bertsekas and Shreve is to state conditions under which the dynamic programming equation is mathematically sound, namely with universally measurable Bellman functions and with universally measurable relaxed control strategies in the context of Borel spaces. The interested reader will find all the subtleties about Borel spaces and universally measurable concepts in [3, Chapter 7].

Yüksel approach As said at the beginning, the most general stochastic dynamic programming principle is sketched by Witsenhausen at the end of [17]. This approach builds upon the so-called Witsenhausen intrinsic model [16] which does not consider state, but information under the form of σ -fields (see [18] for the functional form). In [15], Witsenhausen provides conditions to express stochastic control optimization problems — with information constraints, but without state — in standard form with a state (the first approach that we have considered above).

Although Witsenhausen established a dynamic programming equation in [15], Yüksel notes in [19] that “Witsenhausen’s construction [...] does not address the well-posedness of such a dynamic program” and that “the existence problem was not considered”. In the spirit of [15], Yüksel entails in [19] “a general approach establishing that any sequential team optimization may admit a formulation appropriate for a dynamic programming analysis”. One of the contributions of [19] is to propose a construction of standard Borel controlled state and action spaces and to establish a universal dynamic program for stochastic control optimization problems — with information constraints, but without state — thus addressing some of the issues raised and left open by Witsenhausen. The ingredients are the following:

1. time $t \in \llbracket t_0, T \rrbracket$ is discrete and runs among a finite set of consecutive integers;
2. (Ω, \mathcal{F}) is a measurable space (Nature);
3. $(\mathbb{U}_{t_0}, \mathcal{U}_{t_0}), \dots, (\mathbb{U}_{T-1}, \mathcal{U}_{T-1})$ are measurable spaces (control spaces);
4. $(\mathbb{Y}_{t_0}, \mathcal{Y}_{t_0}), \dots, (\mathbb{Y}_{T-1}, \mathcal{Y}_{T-1})$ are measurable spaces (“observation” spaces);
5. $\{\eta_t : (\Omega \times \prod_{s \in \llbracket t_0, t \rrbracket} \mathbb{U}_s, \mathcal{F} \otimes \bigotimes_{s \in \llbracket t_0, t \rrbracket} \mathcal{U}_s) \rightarrow (\mathbb{U}_t, \mathcal{U}_t)\}_{t \in \llbracket t_0, T-1 \rrbracket}$ are measurable mappings (“measurement constraints”);
6. \mathbb{P} is a probability on (Ω, \mathcal{F}) ;
7. $j : (\Omega \times \prod_{t \in \llbracket t_0, T-1 \rrbracket} \mathbb{U}_t, \mathcal{F} \otimes \bigotimes_{t \in \llbracket t_0, T-1 \rrbracket} \mathcal{U}_t) \rightarrow \mathbb{R}_+$ is a measurable function (criterion).

With these ingredients, Yüksel formulates a stochastic team optimization problem whose solutions (policies) are to be searched among sequences of measurable mappings (“design constraints”) $\mathbb{Y}_{t-1} \rightarrow \mathbb{U}_t$, and their “randomized” versions (so-called strategic measures). He establishes a dynamic programming equation, where the Bellman functions are function of probability distributions and where the minimization is done over proper design mappings. One objective of Yüksel is to set up a mathematical framework under which the dynamic programming equation is mathematically sound [19, Theorem 3.6].

Our approach The ingredients that we use are the following:

1. time $t \in \llbracket t_0, T \rrbracket$ is discrete and runs among a finite set of consecutive integers;
2. $(\mathbb{U}_{t_0}, \mathcal{U}_{t_0}), \dots, (\mathbb{U}_{T-1}, \mathcal{U}_{T-1})$ are measurable spaces (control spaces);
3. $(\mathbb{W}_{t_0}, \mathcal{W}_{t_0}), \dots, (\mathbb{W}_T, \mathcal{W}_T)$ are measurable spaces (noise);
4. $\rho_{t:t+1} : \mathbb{W}_{t_0} \times \prod_{s=t_0}^{t-1} (\mathbb{U}_s \times \mathbb{W}_{s+1}) \rightarrow \Delta(\mathbb{W}_{t+1})$, for $t \in \llbracket t_0, T-1 \rrbracket$, are stochastic kernels (noise distributions);
5. $j : (\mathbb{W}_{t_0} \times \prod_{s=t_0}^{T-1} (\mathbb{U}_s \times \mathbb{W}_{s+1}), \mathcal{W}_{t_0} \otimes \bigotimes_{s=t_0}^{T-1} (\mathcal{U}_s \otimes \mathcal{W}_{s+1})) \rightarrow [0, +\infty]$ is a measurable function (criterion);

6. $t_0 < \dots < t_N = T$ are the indices of multiple consecutive time blocks $\llbracket t_0, t_1 \rrbracket, \dots, \llbracket t_{N-1}, t_N \rrbracket$, with $N \geq 1$ an integer;
7. $\{(\mathbb{X}_{t_j}, \mathcal{X}_{t_j})\}_{j \in \llbracket 0, N \rrbracket}$ are measurable spaces (time block state spaces);
8. $\left\{ \theta_{t_j} : \mathbb{W}_{t_0} \times \prod_{s=t_0}^{t_j-1} (\mathbb{U}_s \times \mathbb{W}_{s+1}) \rightarrow \mathbb{X}_{t_j} \right\}_{j \in \llbracket 1, N \rrbracket}$ and $\theta_{t_0} : \mathbb{W}_{t_0} \rightarrow \mathbb{X}_{t_0}$ are measurable mappings (time block reduction of history towards state);
9. $\left\{ f_{t_j:t_{j+1}} : \mathbb{X}_{t_j} \times \prod_{s=t_j}^{t_{j+1}-1} (\mathbb{U}_s \times \mathbb{W}_{s+1}) \rightarrow \mathbb{X}_{t_{j+1}} \right\}_{j \in \llbracket 0, N-1 \rrbracket}$ are measurable mappings (time block dynamics).

The framework developed in this paper is intermediate between the ones of Evstigneev in [5] and of Yüksel in [19] — notable by the absence of a state space — and the ones of Witsenhausen [15], Hernández-Lerma and Lasserre [6], Bertsekas and Shreve [3] and Puterman [9] — where the state spaces are given for all times.

This said, our preoccupation could be adapted to any of the above frameworks. Indeed, our objective is to establish a dynamic programming equation with a state, not at any time $t \in \llbracket t_0, T \rrbracket$, but at some specified instants $t_0 < t_1 < \dots < t_N = T$. The state spaces are introduced as image sets (codomains) of what we call (*time block*) *history reduction mappings* (where history at time t consists of all uncertainties and controls prior to time t).

2.2 Stochastic dynamic programming with history feedbacks

To prepare the main result in Sect. 3, we establish a dynamic programming equation when the state is the history, that is, the uncertainties and the controls prior to the current stage (see the “canonical construction” in [6, p. 15]). Although quite natural, this equation is generally not written in the literature, as most frameworks in dynamic programming assume the *a priori* existence of a state (see §2.1).

From now on, time is discrete and runs among the integers $t \in \llbracket 0, T \rrbracket$, where $T \in \mathbb{N}^*$ is a positive integer (and where, for the sake of simplicity, we have taken $t_0 = 0$ regarding the notation in §2.1). We first define the basic and the composite spaces that we need to formulate multistage stochastic optimization problems. Then, we introduce a class of solutions called history feedbacks.

Histories and history spaces For each time $t \in \llbracket 0, T - 1 \rrbracket$, the control u_t takes its values in a measurable set \mathbb{U}_t equipped with a σ -field \mathcal{U}_t . For each time $t \in \llbracket 0, T \rrbracket$, the uncertainty w_t takes its values in a measurable set \mathbb{W}_t equipped with a σ -field \mathcal{W}_t . For

$t \in \llbracket 0, T \rrbracket$, we define the *history space* \mathbb{H}_t equipped with the *history field* \mathcal{H}_t

$$\begin{aligned}\mathbb{H}_t &= \mathbb{W}_0 \times \prod_{s=1}^t (\mathbb{U}_{s-1} \times \mathbb{W}_s), \\ \mathcal{H}_t &= \mathbb{W}_0 \otimes \bigotimes_{s=1}^t (\mathcal{U}_{s-1} \otimes \mathcal{W}_s), \quad \forall t \in \llbracket 0, T \rrbracket,\end{aligned}$$

with the particular case $\mathbb{H}_0 = \mathbb{W}_0$, $\mathcal{H}_0 = \mathbb{W}_0$. A generic element $h_t = (w_0, (u_{s-1}, w_s)_{s=1, \dots, t}) = (w_0, u_0, w_1, u_1, w_2, \dots, u_{t-2}, w_{t-1}, u_{t-1}, w_t) \in \mathbb{H}_t$ is called a *history* at time t . For $1 \leq r \leq s \leq t$, we introduce the $(r:s)$ -*history subpart* $h_{r:s} = (u_{r-1}, w_r, \dots, u_{s-1}, w_s) \in \mathbb{H}_{r:s} = \prod_{\tau=r}^s (\mathbb{U}_{\tau-1} \times \mathbb{W}_\tau)$, so that we have $h_t = (h_{r-1}, h_{r:t})$.

History feedbacks For $0 \leq r \leq t \leq T-1$, we define a $(r:t)$ -*history feedback* as a sequence $\{\gamma_s\}_{s=r, \dots, t}$ of measurable mappings $\gamma_s : (\mathbb{H}_s, \mathcal{H}_s) \rightarrow (\mathbb{U}_s, \mathcal{U}_s)$. We call $\Gamma_{r:t}$ the set of $(r:t)$ -history feedbacks. The history feedbacks reflect the following information structure. At the end of the time interval $[t-1, t]$, an uncertainty variable w_t is produced. Then, at the beginning of the time interval $[t, t+1[$, a decision-maker chooses a control u_t contingent on no more than the past, giving the chronology $w_0 \rightsquigarrow u_0 \rightsquigarrow w_1 \rightsquigarrow u_1 \rightsquigarrow \dots \rightsquigarrow w_t \rightsquigarrow u_t \rightsquigarrow \dots \rightsquigarrow w_{T-1} \rightsquigarrow u_{T-1} \rightsquigarrow w_T$.

Family of optimization problems with stochastic kernels We introduce a family of optimization problems with stochastic kernels. Then, we show how such problems can be solved by stochastic dynamic programming. In what follows, we say that a function is *numerical* if it takes its values in $\overline{\mathbb{R}} = [-\infty, +\infty]$ (also called *extended* or *extended real-valued* function). To build a family of optimization problems over the time span $\llbracket 0, T-1 \rrbracket$, we require two ingredients:

- a family $\{\rho_{s-1:s}\}_{s \in \llbracket 1, T \rrbracket}$ of stochastic kernels

$$\rho_{s-1:s} : (\mathbb{H}_{s-1}, \mathcal{H}_{s-1}) \rightarrow \Delta(\mathbb{W}_s), \quad \forall s \in \llbracket 1, T \rrbracket, \quad (1)$$

that represents the distribution of the next uncertainty w_s parameterized by past history h_{s-1} ,

- a numerical function, playing the role of a cost to be minimized,

$$j : (\mathbb{H}_T, \mathcal{H}_T) \rightarrow [0, +\infty], \quad (2)$$

assumed to be nonnegative² and measurable with respect to the field \mathcal{H}_T .

²We could also consider any $j : \mathbb{H}_t \rightarrow \mathbb{R}$, measurable bounded function, or measurable and uniformly bounded below function. However, for the sake of simplicity, we will deal in the sequel with measurable non-negative numerical functions. When $j(h_T) = +\infty$, this materializes joint constraints between uncertainties and controls.

We define, for any feedback $\{\gamma_s\}_{s=t,\dots,T-1} \in \Gamma_{t:T-1}$, a new family of stochastic kernels $\rho_{t:T}^\gamma : (\mathbb{H}_t, \mathcal{H}_t) \rightarrow \Delta(\mathbb{H}_T)$, that capture the transitions between histories when the dynamics $h_{s+1} = (h_s, u_s, w_{s+1})$ is driven by $u_s = \gamma_s(h_s)$ for all s in $\llbracket t, T-1 \rrbracket$ (see Definition 10 in Appendix A for the detailed construction of $\rho_{t:T}^\gamma$; note that $\rho_{t:T}^\gamma$ generates a probability distribution on the space \mathbb{H}_T of histories over the whole timespan $\llbracket 0, T \rrbracket$). We consider the following family of optimization problems, indexed by t in $\llbracket 0, T-1 \rrbracket$ and parameterized by the history $h_t \in \mathbb{H}_t$: for all t in $\llbracket 0, T-1 \rrbracket$, we define the minimum value

$$V_t(h_t) = \inf_{\gamma_{t:T-1} \in \Gamma_{t:T-1}} \int_{\mathbb{H}_T} j(h'_T) \rho_{t:T}^\gamma(dh'_T | h_t), \quad \forall h_t \in \mathbb{H}_t, \quad (3a)$$

$$\text{and we also define } V_T(h_T) = j(h_T), \quad \forall h_T \in \mathbb{H}_T. \quad (3b)$$

The numerical function $V_t : \mathbb{H}_t \rightarrow [0, +\infty]$ is called the *value function* at time t .

In the next paragraph, we show how the family $\{V_t\}_{t \in \llbracket 0, T \rrbracket}$ of value functions can be used to solve, via dynamic programming, the optimization problem of interest whose value is

$$\begin{aligned} V_0(w_0) &= \inf_{\gamma_{0:T-1} \in \Gamma_{0:T-1}} \int_{\mathbb{H}_T} j(h'_T) \rho_{0:T}^\gamma(dh'_T | w_0) \\ &= \inf_{\gamma_{0:T-1} \in \Gamma_{0:T-1}} \int_{\mathbb{W}_{1:T}} j(\Phi_{0:T}^\gamma(w_{0:T})) \\ &\quad \prod_{s=1}^T \rho_{s-1:s}(dw_s | \Phi_{0:s-1}^\gamma(w_{0:s-1})), \end{aligned} \quad (4)$$

by (40), where the flows $\Phi_{0,s}^\gamma$ for $s \in \llbracket 0, T-1 \rrbracket$ are defined by Equation (38a) in Appendix A.

Bellman operators and dynamic programming We show that the value functions in (3) are *Bellman functions*, in that they are solution of a Bellman or dynamic programming equation. For t in $\llbracket 0, T \rrbracket$, let $\mathbb{L}_+^0(\mathbb{H}_t, \mathcal{H}_t)$ be the space of universally measurable nonnegative numerical functions over \mathbb{H}_t (see [3, § 7.7] for further details). For t in $\llbracket 0, T-1 \rrbracket$, we define the *Bellman operator* $\mathcal{B}_{t+1:t}$ by, for all $\varphi \in \mathbb{L}_+^0(\mathbb{H}_{t+1}, \mathcal{H}_{t+1})$, and for all $h_t \in \mathbb{H}_t$,

$$(\mathcal{B}_{t+1:t}\varphi)(h_t) = \inf_{u_t \in \mathbb{U}_t} \int_{\mathbb{W}_{t+1}} \varphi(h_t, u_t, w_{t+1}) \rho_{t:t+1}(dw_{t+1} | h_t). \quad (5)$$

Since $\varphi \in \mathbb{L}_+^0(\mathbb{H}_{t+1}, \mathcal{H}_{t+1})$, we have that $\mathcal{B}_{t+1:t}\varphi$ is a well defined nonnegative numerical function. The proof of the following theorem is given in Appendix A.

Theorem 1 *Assume that all the spaces introduced in §2.2 are Borel spaces, the stochastic kernels in (1) are Borel-measurable, and that the criterion j in (2) is a nonnegative lower semianalytic numerical function. Then, the Bellman operators in (5) are such that $\mathcal{B}_{t+1:t} : \mathbb{L}_+^0(\mathbb{H}_{t+1}, \mathcal{H}_{t+1}) \rightarrow \mathbb{L}_+^0(\mathbb{H}_t, \mathcal{H}_t)$, and the value functions V_t defined in (3) are universally measurable and satisfy the Bellman equation, or (stochastic) dynamic programming equation,*

$$V_T = j, \quad V_t = \mathcal{B}_{t+1:t}V_{t+1}, \quad \text{for } t = T-1, \dots, 1, 0. \quad (6)$$

This theorem is mainly inspired by [3, Chap. 8], with the feature that the state x_t is, in our case, the canonical history h_t , with the canonical dynamics $h_{t+1} = (h_t, u_t, w_{t+1})$. This very general dynamic programming result will be the basis of all future developments in this paper. In the sequel, we assume that all the assumptions of Theorem 1 are fulfilled; all the spaces (like the ones introduced in §2.2) are Borel spaces; all the stochastic kernels (like the ones introduced in (1)) are Borel-measurable; all the criteria (like the one introduced in (2)) are nonnegative lower semianalytic functions.

This Sect. 2 is mostly made of recalls and of statements that are straightforward consequences of results already established in the literature. However, the developments in §2.2 are indispensable to tackle time blocks decomposition in the coming Sect. 3.

3 State reduction by time blocks and dynamic programming

In this section, we consider the question of reducing the history using a compressed “state” variable. Differing with traditional practice, such a variable may not be available at any time $t \in \llbracket 0, T \rrbracket$, but at some specified stages $0 = t_0 < \dots < t_N = T$. We have seen in §2.2 that the history h_t is itself a state variable with associated canonical dynamics $h_{t+1} = (h_t, u_t, w_{t+1})$. However the size of this canonical state increases with t , which is an unpleasant feature for dynamic programming, hence the practical need to introduce a (ideally low dimensional) state space, at least at some specified stages, as done in this paper. As said in the introduction, the main difficulty is notational.

3.1 State reduction on a single time block

We first present the case where the reduction only occurs at two instants denoted by r and t , and such that $0 \leq r < t \leq T$.

Definition 2 *Let $(\mathbb{X}_r, \mathcal{X}_r)$ and $(\mathbb{X}_t, \mathcal{X}_t)$ be two measurable state spaces, θ_r and θ_t be two measurable reduction mappings*

$$\theta_r : \mathbb{H}_r \rightarrow \mathbb{X}_r, \quad \theta_t : \mathbb{H}_t \rightarrow \mathbb{X}_t, \quad (7a)$$

and $f_{r:t}$ be a measurable dynamics

$$f_{r:t} : \mathbb{X}_r \times \mathbb{H}_{r+1:t} \rightarrow \mathbb{X}_t. \quad (7b)$$

The triplet $(\theta_r, \theta_t, f_{r:t})$ is called a state reduction across $(r:t)$ if we have³

$$\theta_t((h_r, h_{r+1:t})) = f_{r:t}(\theta_r(h_r), h_{r+1:t}), \quad \forall h_t \in \mathbb{H}_t. \quad (7c)$$

The state reduction $(\theta_r, \theta_t, f_{r:t})$ is said to be compatible with the family $\{\rho_{s-1:s}\}_{r+1 \leq s \leq t}$ of stochastic kernels (1) if

³Notice that, if only the couple $(\theta_r, f_{r:t})$ is given, we can define θ_t by (7c), and thus obtain a triplet $(\theta_r, \theta_t, f_{r:t})$ which is a state reduction across $(r:t)$.

- there exists a reduced stochastic kernel $\tilde{\rho}_{r:r+1} : \mathbb{X}_r \rightarrow \Delta(\mathbb{W}_{r+1})$, such that the stochastic kernel $\rho_{r:r+1}$ in (1) can be factored, for all $h_r \in \mathbb{H}_r$, as $\rho_{r:r+1}(dw_{r+1} | h_r) = \tilde{\rho}_{r:r+1}(dw_{r+1} | \theta_r(h_r))$,
- for all s in $\llbracket r+2, t \rrbracket$, there exists a reduced stochastic kernel $\tilde{\rho}_{s-1:s} : \mathbb{X}_r \times \mathbb{H}_{r+1:s-1} \rightarrow \Delta(\mathbb{W}_s)$, such that the stochastic kernel $\rho_{s-1:s}$ can be factored, for all $h_{s-1} \in \mathbb{H}_{s-1}$, as $\rho_{s-1:s}(dw_s | (h_r, h_{r+1:s-1})) = \tilde{\rho}_{s-1:s}(dw_s | (\theta_r(h_r), h_{r+1:s-1}))$.

The above definition is quite similar to the sufficient statistics idea in stochastic control: the state variable, which summarizes the history, is sufficient for the controller to design its control policy ([14, p. 19], [3, Definition 10.6], [13]). However, sufficient statistics in the stochastic control literature are defined at the original time stages. By contrast, Definition 2 — and the coming Definition 4 — consider a notion of sufficient statistics only for a subset of stages, which is the main focus of this paper.

According to Definition 2, the triplet $(\theta_r, \theta_t, f_{r:t})$ is a state reduction across $(r:t)$ if and only if the diagram in the left part of Figure 1 is commutative; it is compatible if and only if the diagram in the middle part of Figure 1 is commutative.

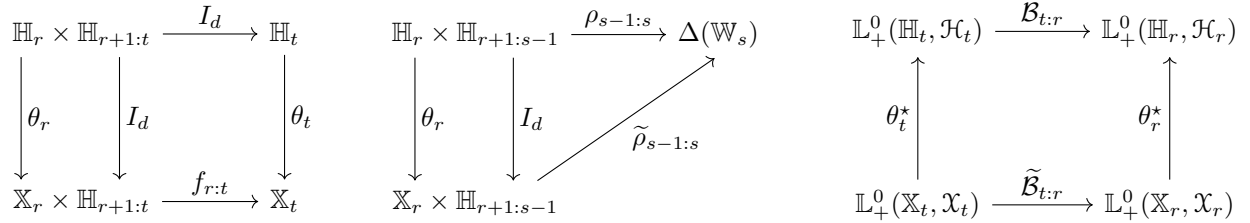


Figure 1: Commutative diagrams in case of state reduction

To prepare a dynamic programming equation, we define the *Bellman operator across* $(t:r)$ $\mathcal{B}_{t:r} : \mathbb{L}_+^0(\mathbb{H}_t, \mathcal{H}_t) \rightarrow \mathbb{L}_+^0(\mathbb{H}_r, \mathcal{H}_r)$ by

$$\mathcal{B}_{t:r} = \mathcal{B}_{r+1:r} \circ \cdots \circ \mathcal{B}_{t:t-1}, \quad (9)$$

where the one time step operators $\mathcal{B}_{s:s-1}$, for s in $\llbracket r+1, t \rrbracket$ are defined in (5). The following proposition is the key ingredient to formulate dynamic programming equations with a reduced state.

Proposition 3 *Suppose that all the assumptions of Theorem 1 are satisfied. Suppose that there exists a state reduction $(\theta_r, \theta_t, f_{r:t})$ that is compatible with the family $\{\rho_{s-1:s}\}_{r+1 \leq s \leq t}$ of stochastic kernels (1) (see Definition 2). Then, there exists a reduced Bellman operator across $(t:r)$ $\tilde{\mathcal{B}}_{t:r} : \mathbb{L}_+^0(\mathbb{X}_t, \mathcal{X}_t) \rightarrow \mathbb{L}_+^0(\mathbb{X}_r, \mathcal{X}_r)$, such that*

$$(\tilde{\mathcal{B}}_{t:r} \tilde{\varphi}_t) \circ \theta_r = \mathcal{B}_{t:r}(\tilde{\varphi}_t \circ \theta_t), \quad \forall \tilde{\varphi}_t \in \mathbb{L}_+^0(\mathbb{X}_t, \mathcal{X}_t). \quad (10)$$

For any $\tilde{\varphi}_t \in \mathbb{L}_+^0(\mathbb{X}_t, \mathcal{X}_t)$ and for any $x_r \in \mathbb{X}_r$, we have that

$$\begin{aligned}
& (\tilde{\mathcal{B}}_{t:r} \tilde{\varphi}_t)(x_r) \\
&= \inf_{u_r \in \mathbb{U}_r} \int_{\mathbb{W}_{r+1}} \tilde{\rho}_{r:r+1}(\mathrm{d}w_{r+1} \mid x_r) \\
&\quad \inf_{u_{r+1} \in \mathbb{U}_{r+1}} \int_{\mathbb{W}_{r+2}} \tilde{\rho}_{r+1:r+2}(\mathrm{d}w_{r+2} \mid x_r, u_r, w_{r+1}) \quad \cdots \\
&\quad \inf_{u_{t-1} \in \mathbb{U}_{t-1}} \int_{\mathbb{W}_t} \tilde{\rho}_{t-1:t}(\mathrm{d}w_t \mid x_r, u_r, w_{r+1}, \dots, u_{t-2}, w_{t-1}) \\
&\quad \tilde{\varphi}_t(f_{r:t}(x_r, u_r, w_{r+1}, \dots, u_{t-1}, w_t)) .
\end{aligned} \tag{11}$$

The proof of Proposition 3 is given in Appendix A. Proposition 3 can be interpreted as follows. Denoting by $\theta_t^* : \mathbb{L}_+^0(\mathbb{X}_t, \mathcal{X}_t) \rightarrow \mathbb{L}_+^0(\mathbb{H}_t, \mathcal{H}_t)$ the operator defined by $\theta_t^*(\tilde{\varphi}_t) = \tilde{\varphi}_t \circ \theta_t$ for any $\tilde{\varphi}_t \in \mathbb{L}_+^0(\mathbb{X}_t, \mathcal{X}_t)$, the relation (10) rewrites as $\theta_r^* \circ \tilde{\mathcal{B}}_{t:r} = \mathcal{B}_{t:r} \circ \theta_t^*$, that is, Proposition 3 states that the diagram in the right part of Figure 1 is commutative.

3.2 State reduction on multiple consecutive time blocks and dynamic programming equations

Proposition 3 can easily be extended to the case of multiple consecutive time blocks $\llbracket t_i, t_{i+1} \rrbracket$, with $N \in \mathbb{N}^*$, $i \in \llbracket 0, N-1 \rrbracket$ and $0 = t_0 < \dots < t_N = T$.

Definition 4 Let $\{(\mathbb{X}_{t_i}, \mathcal{X}_{t_i})\}_{i \in \llbracket 0, N \rrbracket}$ be a family of measurable state spaces, $\{\theta_{t_i}\}_{i \in \llbracket 0, N \rrbracket}$ be a family of measurable reduction mappings $\theta_{t_i} : \mathbb{H}_{t_i} \rightarrow \mathbb{X}_{t_i}$, and $\{f_{t_i:t_{i+1}}\}_{i \in \llbracket 0, N-1 \rrbracket}$ be a family of measurable dynamics

$$f_{t_i:t_{i+1}} : \mathbb{X}_{t_i} \times \mathbb{H}_{t_{i+1}:t_{i+1}} \rightarrow \mathbb{X}_{t_{i+1}} .$$

The triplet $(\{\mathbb{X}_{t_i}\}_{i \in \llbracket 0, N \rrbracket}, \{\theta_{t_i}\}_{i \in \llbracket 0, N \rrbracket}, \{f_{t_i:t_{i+1}}\}_{i \in \llbracket 0, N-1 \rrbracket})$ is called a state reduction across the consecutive time blocks $\llbracket t_i, t_{i+1} \rrbracket$, $i \in \llbracket 0, N-1 \rrbracket$ if every triplet $(\theta_{t_i}, \theta_{t_{i+1}}, f_{t_i:t_{i+1}})$ is a state reduction, for i in $\llbracket 0, N-1 \rrbracket$. The state reduction across the consecutive time blocks $\llbracket t_i, t_{i+1} \rrbracket$ is said to be compatible with the family $\{\rho_{s-1:s}\}_{s \in \llbracket 1, T \rrbracket}$ of stochastic kernels given in (1) if every triplet $(\theta_{t_i}, \theta_{t_{i+1}}, f_{t_i:t_{i+1}})$ is compatible with the family $\{\rho_{s-1:s}\}_{s \in \llbracket t_i+1, t_{i+1} \rrbracket}$, for i in $\llbracket 0, N-1 \rrbracket$.

There is a practical case where state reductions can readily be obtained.

Remark 5 (Composed state dynamics as a straightforward reduction mapping)

We consider here the special case where the model is given by controlled state dynamics driven by noises. That is, we are given a family of measurable state spaces $\{(\mathbb{X}_s, \mathcal{X}_s)\}_{s \in \llbracket 0, T \rrbracket}$ and a family $\{f_{s:s+1}\}_{s \in \llbracket 0, T-1 \rrbracket}$ of measurable dynamics

$$f_{s:s+1} : \mathbb{X}_s \times \mathbb{U}_s \times \mathbb{W}_{s+1} \rightarrow \mathbb{X}_{s+1} . \tag{12}$$

For any time $s \in \llbracket 0, T-1 \rrbracket$, we define the composition $f_{0:s+1} = f_{s:s+1} \circ f_{s-1:s} \circ \dots \circ f_{0:1}$ with the abuse of notation that the composition is performed on the state argument. Setting $\mathbb{W}_0 = \mathbb{X}_0$, we obtain that $f_{0:s+1} : \mathbb{H}_{s+1} \rightarrow \mathbb{X}_{s+1}$ is a mapping from the history space \mathbb{H}_{s+1} taking values in the state space \mathbb{X}_{s+1} .

Now, given an integer $N > 0$ and an increasing sequence $0 = t_0 < \dots < t_N = T$ of times, we define the family $\{\theta_{t_i}\}_{i \in \llbracket 0, N \rrbracket}$ of measurable reduction mappings by $\theta_{t_i} = f_{0:t_i} : \mathbb{H}_{t_i} \rightarrow \mathbb{X}_{t_i}$ for $i > 0$, and by $\theta_0 = I_d$ (the identity mapping on \mathbb{W}_0) for $i = 0$. Moreover, given i and $j \in \llbracket 0, N \rrbracket$, with $i < j$ we obtain, for all $h_{t_j} \in \mathbb{H}_{t_j}$, that

$$\theta_{t_j}(h_{t_j}) = \theta_{t_j}((h_{t_i}, h_{t_{i+1}:t_j})) = f_{t_i:t_j}(\theta_{t_i}(h_{t_i}), h_{t_{i+1}:t_j}), \quad (13)$$

with $f_{t_i:t_j} = f_{t_j-1:t_j} \circ f_{t_j-2:t_j-1} \circ \dots \circ f_{t_i:t_{i+1}}$ which gives the state reduction Equation (7c).

There is a practical case where *compatible* state reductions can readily be obtained.

Remark 6 (Block independent exogenous noises and stochastic kernels)

Assume that the family $\{\rho_{s-1:s}\}_{s \in \llbracket 1, T \rrbracket}$ of stochastic kernels in §2.2 are mappings whose arguments do not include the control part (that is, depend at most on the history uncertainty part (see (37a)). If we interpret stochastic kernels as (conditional) distributions of noises (random process), this means that the system dynamics are driven by an exogenous noise process, say $(\mathbf{W}_t)_{t \in \llbracket 1, T \rrbracket}$. Moreover, assume that the stochastic kernels give rise to noises that are independent block by block, in the sense that the family of noises $\{(\mathbf{W}_t)_{t \in \llbracket t_i+1, t_{i+1} \rrbracket}\}_{i \in \llbracket 0, N-1 \rrbracket}$ is made of independent random vectors, i by i . Then, from Definitions 2 and 4, we deduce that any state reduction across the same time blocks is compatible with the stochastic kernels.

Assuming the existence of a state reduction across the consecutive time blocks $\llbracket t_i, t_{i+1} \rrbracket$ compatible with the family of stochastic kernels (1), we obtain the existence of a family of reduced Bellman operators across the consecutive $\llbracket t_i, t_{i+1} \rrbracket$ as an immediate consequence of multiple applications of Proposition 3, that is, $\tilde{\mathcal{B}}_{t_{i+1}:t_i} : \mathbb{L}_+^0(\mathbb{X}_{t_{i+1}}, \mathcal{X}_{t_{i+1}}) \rightarrow \mathbb{L}_+^0(\mathbb{X}_{t_i}, \mathcal{X}_{t_i})$, $i \in \llbracket 0, N-1 \rrbracket$, such that, for any function $\tilde{\varphi}_{t_{i+1}} \in \mathbb{L}_+^0(\mathbb{X}_{t_{i+1}}, \mathcal{X}_{t_{i+1}})$, we have that $(\tilde{\mathcal{B}}_{t_{i+1}:t_i} \tilde{\varphi}_{t_{i+1}}) \circ \theta_{t_i} = \mathcal{B}_{t_{i+1}:t_i}(\tilde{\varphi}_{t_{i+1}} \circ \theta_{t_{i+1}})$. We now consider the family of optimization problems defined by the associated value functions (3). Thanks to the state reductions, we can enounce the following theorem which establishes dynamic programming equations *across* consecutive time blocks.

Theorem 7 (Time block decomposition) *Suppose that all the assumptions of Theorem 1 are satisfied. Suppose that a state reduction $(\{\mathbb{X}_{t_i}\}_{i \in \llbracket 0, N \rrbracket}, \{\theta_{t_i}\}_{i \in \llbracket 0, N \rrbracket}, \{f_{t_i:t_{i+1}}\}_{i \in \llbracket 0, N-1 \rrbracket})$ exists across the consecutive time blocks $\{\llbracket t_i, t_{i+1} \rrbracket\}_{i \in \llbracket 0, N-1 \rrbracket}$, satisfying $0 = t_0 < \dots < t_N = T$, and which is compatible with the family $\{\rho_{s-1:s}\}_{s \in \llbracket 1, T \rrbracket}$ of stochastic kernels given in (1). Assume that there exists a reduced criterion $\tilde{j} : \mathbb{X}_T \rightarrow [0, +\infty]$ such that the cost function j in (2) can be factored as $j = \tilde{j} \circ \theta_T$. We define the family of reduced value functions $\{\tilde{V}_{t_i}\}_{i \in \llbracket 0, N \rrbracket}$ by*

$$\tilde{V}_{t_N} = \tilde{j} \text{ and } \tilde{V}_{t_i} = \tilde{\mathcal{B}}_{t_{i+1}:t_i} \tilde{V}_{t_{i+1}}, \quad \forall i \in \llbracket 0, N-1 \rrbracket. \quad (15)$$

Then, the family $\{V_{t_i}\}_{i \in \llbracket 0, N \rrbracket}$ in (3) satisfies $V_{t_i} = \tilde{V}_{t_i} \circ \theta_{t_i}$, for all $i \in \llbracket 0, N \rrbracket$.

The proof is an immediate consequence of multiple applications of Theorem 1 and Proposition 3. Then, it is easy, and left to the reader, to prove that the following Corollary holds true.

Corollary 8 (Taking care of instantaneous costs in addition to final cost)

Assume that a state reduction on multiple consecutive time blocks compatible with the family of stochastic kernels (as in Definition 4) exists, and assume moreover that the criterion $j : \mathbb{H}_T \rightarrow [0, +\infty]$ can be factored as

$$j(h_T) = \sum_{i=0}^{N-1} \ell_{t_i}(\theta_{t_i}(h_{t_i}), h_{t_{i+1}:t_{i+1}}) + \ell_{t_N}(\theta_{t_N}(h_{t_N})) . \quad (16)$$

Theorem 7 remains valid with the reduced Bellman value functions given by

$$\tilde{V}_{t_N} = \ell_{t_N} \quad \text{and} \quad \tilde{V}_{t_i} = \bar{\mathcal{B}}_{t_{i+1}:t_i} \tilde{V}_{t_{i+1}} , \quad \forall i \in \llbracket 0, N-1 \rrbracket ,$$

and the reduced Bellman operator across $(t_i:t_{i+1})$ given, for any $i \in \llbracket 0, N-1 \rrbracket$, for any $\tilde{\varphi}_{t_{i+1}} \in \mathbb{L}_+^0(\mathbb{X}_{t_{i+1}}, \mathbb{X}_{t_{i+1}})$ and for any $x_{t_i} \in \mathbb{X}_{t_i}$, by

$$\begin{aligned} (\bar{\mathcal{B}}_{t_{i+1}:t_i} \tilde{\varphi}_{t_{i+1}})(x_{t_i}) &= \inf_{u_{t_i} \in \mathbb{U}_{t_i}} \int_{\mathbb{W}_{t_i+1}} \tilde{\rho}_{t_i:t_{i+1}}(dw_{t_i+1} | x_{t_i}) \\ &\quad \inf_{u_{t_{i+1}} \in \mathbb{U}_{t_{i+1}}} \int_{\mathbb{W}_{t_i+2}} \tilde{\rho}_{t_{i+1}:t_{i+2}}(dw_{t_i+2} | x_{t_i}, u_{t_i}, w_{t_i+1}) \quad \cdots \\ &\quad \inf_{u_{t_{i+1}-1} \in \mathbb{U}_{t_{i+1}-1}} \int_{\mathbb{W}_{t_{i+1}}} \tilde{\rho}_{t_{i+1}-1:t_{i+1}} \\ &\quad (dw_{t_{i+1}} | x_{t_i}, u_{t_i}, w_{t_i+1}, \dots, u_{t_{i+1}-2}, w_{t_{i+1}-1}) \\ &\quad \left(\ell_{t_i}(x_{t_i}, u_{t_i}, w_{t_i+1}, \dots, u_{t_{i+1}-1}, w_{t_{i+1}}) \right. \\ &\quad \left. + \tilde{\varphi}_{t_{i+1}}(f_{t_i:t_{i+1}}(x_{t_i}, u_{t_i}, w_{t_i+1}, \dots, u_{t_{i+1}-1}, w_{t_{i+1}})) \right) . \end{aligned} \quad (17)$$

Of course, solving Equation (15) or Equation (17) can be as difficult as solving the original Bellman equation. However, the interest of such time block decomposition will be illustrated on the two time scale optimization problems, object of the next Sect. 4, as detailed at the end of §4.3.

4 Two time scale optimization problems

Some decisions problems naturally involve two different time scales, because of the timing of decisions — as for example long term investment decision and short term monitoring of physical devices. In this section, we introduce abstract mathematical notations to describe multistage decision problems with two time scales. Then, we show how they can be reformulated on a unique *product timeline* in order to obtain a block decomposition by Theorem 7.

In §4.1 and §4.2 we detail the structure and we formulate the two time scale optimization problems that we consider. In §4.3, we show how to decompose such problems by time blocks. In §4.4, we make the link with the classical framework of stochastic optimal control, and we illustrate the approach on a crude oil procurement problem in §4.5.

4.1 Structure of a two time scale optimization problem

We provide the data for a two time scale optimization problem.

Two time scales. We consider a multistage decision problem, with two time scales. The slow time scale is represented by a finite totally ordered set (\mathbf{S}, \preceq) as follows — where s^+ denotes the successor of $s \in \mathbf{S}$ and s^- its predecessor, and where we use the notation $t \prec t'$ for $t \preceq t'$ and $t \neq t'$ —

$$\min \mathbf{S} = \underline{s} \prec \dots \prec s^- \prec s \prec s^+ \prec \dots \prec \bar{s} = \max \mathbf{S}, \quad (18a)$$

and the fast time scale by a finite totally ordered set (\mathbf{F}, \preceq) :

$$\min \mathbf{F} = \underline{f} \prec \dots \prec f^- \prec f \prec f^+ \prec \dots \prec \bar{f} = \max \mathbf{F}. \quad (18b)$$

In a sense to be made more rigorous later (once a unified timeline will have been defined), each slow time interval $[s, s^+[$ is made up of $|\mathbf{F}|$ (cardinality of \mathbf{F}) fast time steps, hence the denomination “two time scale”. For instance, $\mathbf{S} = \{Mo, Tu, We, Th, Fr, Sa, Su\}$ may represent days, whereas $\mathbf{F} = \llbracket 1, 24 \rrbracket$ may represent hours within a day. In some problems, we might even take $\mathbf{F} = \llbracket 0, 24 \rrbracket$ to handle the fact that two decisions (one slow and one fast) are taken at midnight, hence an additional fast time step 0.

Unified timeline. We define the unified timeline of the decision problem in two steps. First, we equip the product set $\mathbf{S} \times \mathbf{F}$ with the following lexicographic order:

$$\begin{aligned} (\underline{s}, \underline{f}) \prec \dots \prec (s^-, \bar{f}) \prec (s, \underline{f}) \prec (s, \underline{f}^+) \prec \dots \\ \dots \prec (s, \bar{f}^-) \prec (s, \bar{f}) \prec (s^+, \underline{f}) \prec \dots \prec (\bar{s}, \bar{f}). \end{aligned} \quad (19)$$

More formally, we denote by $(s, f)^+$ the successor of (s, f) in $\mathbf{S} \times \mathbf{F} \setminus \{(\bar{s}, \bar{f})\}$, with

$$(s, f)^+ = \begin{cases} (s, f^+) & \text{if } f \neq \bar{f}, \\ (s^+, \underline{f}) & \text{if } f = \bar{f}. \end{cases} \quad (20a)$$

Similarly, we denote by $(s, f)^-$ the predecessor of (s, f) in $\mathbf{S} \times \mathbf{F} \setminus \{(\underline{s}, \underline{f})\}$, with

$$(s, f)^- = \begin{cases} (s, f^-) & \text{if } f \neq \underline{f}, \\ (s^-, \bar{f}) & \text{if } f = \underline{f}. \end{cases} \quad (20b)$$

We adopt the convention that the slow time $s \in \mathbf{S}$ is identified with the two scale time (s, \bar{f}) , as illustrated in Figure 2. For instance Monday is identified with $(Mo, 24)$.

In the product set $\mathbf{S} \times \mathbf{F}$, the first time $(\underline{s}, \underline{f})$ does not coincide with a slow time (the couple $(Mo, 0)$ does not correspond to Monday in our running example). Thus, we add to the product set $\mathbf{S} \times \mathbf{F}$ an extra time denoted by $(\underline{s}^-, \bar{f})$, corresponding to the extra slow time \underline{s}^- , which is such that $(\underline{s}, \underline{f})^- = (\underline{s}^-, \bar{f})$. We denote by $\bar{\mathbf{S}}$ the set $\{\underline{s}^-\} \cup \mathbf{S}$ and by $\bar{\mathbf{S}} \times \bar{\mathbf{F}}$ the set $(\underline{s}^-, \bar{f}) \cup (\mathbf{S} \times \mathbf{F})$, also called the *extended timeline* when equipped with an order \preceq as follows — where we use the notation $(s, f) \prec (s', f')$ for $(s, f) \preceq (s', f')$ and $(s, f) \neq (s', f')$ —

$$\begin{aligned} (\underline{s}^-, \bar{f}) \prec (\underline{s}, \underline{f}) \prec \dots \prec (s^-, \bar{f}) \prec (s, \underline{f}) \prec (s, \underline{f}^+) \prec \dots \\ \dots \prec (s, \bar{f}^-) \prec (s, \bar{f}) \prec (s^+, \underline{f}) \prec \dots \prec (\bar{s}, \bar{f}). \end{aligned} \quad (21)$$

The two time scale optimization problem will be formulated on the extended timeline $\bar{\mathbf{S}} \times \bar{\mathbf{F}}$, which we trivially identify with the time set $\llbracket 0, T \rrbracket$, where $T = |\mathbf{S}| \times |\mathbf{F}|$.

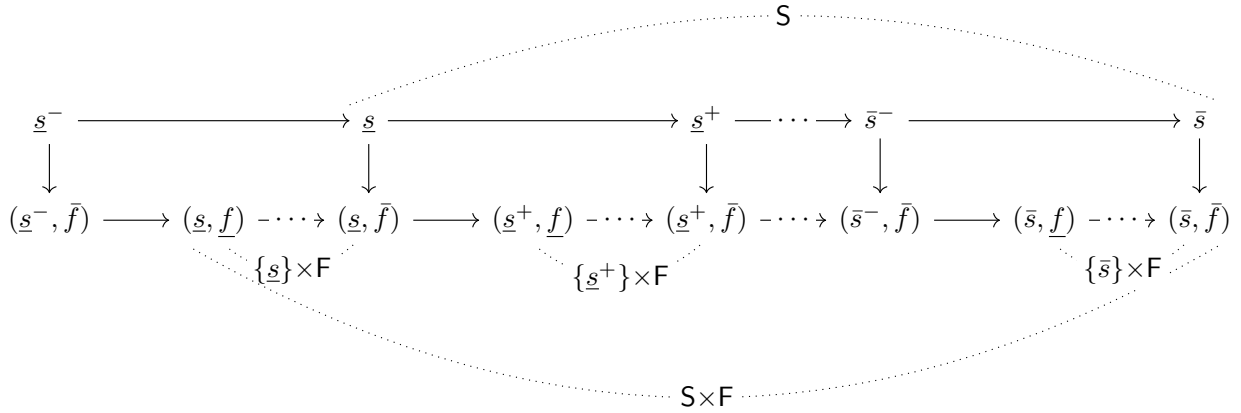


Figure 2: The product timeline with an extra starting point $(\underline{s}^-, \bar{f})$

Decisions. We suppose given

- a family $\{\mathbb{U}_s^s\}_{s \in \bar{\mathbf{S}} \setminus \{\bar{s}\}}$ of *slow time scale decision measurable sets*, and a family $\{\mathbb{W}_s^s\}_{s \in \mathbf{S}}$ of *slow time scale uncertainty measurable sets*,
- a family $\{\mathbb{U}_{(s,f)}^{sf}\}_{(s,f) \in \mathbf{S} \times (\mathbf{F} \setminus \{\bar{f}\})}$ of *fast time scale decision measurable sets*, and a family $\{\mathbb{W}_{(s,f)}^{sf}\}_{(s,f) \in \mathbf{S} \times (\mathbf{F} \setminus \{\bar{f}\})}$ of *fast time scale uncertainty measurable sets*.

Dynamics. We suppose given a family $\{\mathbb{X}_s^s\}_{s \in \bar{\mathbf{S}}}$ and a family $\{\mathbb{X}_{(s,f)}^{sf}\}_{(s,f) \in \mathbf{S} \times (\mathbf{F} \setminus \{\bar{f}\})}$ of *slow time scale* and *fast time scale state measurable sets*. We also suppose given a family

$\{\mathcal{F}_s^s\}_{s \in \bar{S} \setminus \{\bar{s}\}}$ of *slow time scale dynamics measurable mappings*, that represent the evolution “driven at the slow time scale” given, for $s \in \bar{S} \setminus \{\bar{s}\}$, by⁴

$$\begin{aligned} \mathcal{F}_s^s &: \mathbb{X}_s^s \times \mathbb{U}_s^s \times \mathbb{W}_{s^+}^s \rightarrow \mathbb{X}_{(s^+, \underline{f})}^{\text{sf}}, \\ (x_s^s, u_s^s, w_{s^+}^s) &\mapsto x_{(s^+, \underline{f})}^{\text{sf}} = \mathcal{F}_s^s(x_s^s, u_s^s, w_{s^+}^s). \end{aligned} \quad (22a)$$

We suppose given a family $\{\mathcal{F}_{(s,f)}^{\text{sf}}\}_{(s,f) \in \mathbb{S} \times (\mathbb{F} \setminus \{\bar{f}\})}$ of *fast time scale dynamics measurable mappings*, that represent the evolution “driven at the fast time scale” given, for all $s \in \mathbb{S}$ and $f \in \mathbb{F} \setminus \{\bar{f}\}$, by

$$\begin{aligned} \mathcal{F}_{(s,f)}^{\text{sf}} &: \mathbb{X}_{(s,f)}^{\text{sf}} \times \mathbb{U}_{(s,f)}^{\text{sf}} \times \mathbb{W}_{(s,f)^+}^{\text{sf}} \rightarrow \mathbb{X}_{(s,f)^+}^{\text{sf}}, \\ (x_{(s,f)}^{\text{sf}}, u_{(s,f)}^{\text{sf}}, w_{(s,f)^+}^{\text{sf}}) &\mapsto x_{(s,f)^+}^{\text{sf}} \\ &= \mathcal{F}_{(s,f)}^{\text{sf}}(x_{(s,f)}^{\text{sf}}, u_{(s,f)}^{\text{sf}}, w_{(s,f)^+}^{\text{sf}}), \end{aligned} \quad (22b)$$

where, for the sake of simplicity, we use the notation $\mathbb{X}_{(s,\bar{f})}^{\text{sf}} = \mathbb{X}_s^s$ for all $s \in \mathbb{S}$.

Criterion. We suppose given a family $\{\Lambda_s\}_{s \in \bar{S} \setminus \{\bar{s}\}}$ of *slow time scale measurable cost functions*, with

$$\Lambda_{s^-} : \mathbb{X}_{s^-}^s \times \mathbb{U}_{s^-}^s \times \mathbb{W}_s^s \times \underbrace{\prod_{f \in \mathbb{F} \setminus \{\bar{f}\}} (\mathbb{X}_{(s,f)}^{\text{sf}} \times \mathbb{U}_{(s,f)}^{\text{sf}} \times \mathbb{W}_{(s,f)^+}^{\text{sf}})}_{\text{interval } [s^-, s[} \rightarrow \mathbb{R},$$

for $s \in \mathbb{S}$, and we suppose given a function $\Lambda_{\bar{s}}$ representing a final cost, with

$$\Lambda_{\bar{s}} : \mathbb{X}_{\bar{s}}^s \rightarrow \mathbb{R},$$

that make up, by summation, an intertemporal criterion

$$\sum_{s \in \mathbb{S}} \Lambda_{s^-}(x_{s^-}^s, u_{s^-}^s, w_s^s, \{x_{(s,f)}^{\text{sf}}, u_{(s,f)}^{\text{sf}}, w_{(s,f)^+}^{\text{sf}}\}_{f \in \mathbb{F} \setminus \{\bar{f}\}}) + \Lambda_{\bar{s}}(x_{\bar{s}}^s). \quad (24)$$

Stochastic kernels. Finally, we suppose given a family of *constant slow time scale stochastic kernels* $\{\rho_{s:s^+}^s\}_{s \in \bar{S} \setminus \{\bar{s}\}}$

$$\rho_{s:s^+}^s \in \Delta(\mathbb{W}_{s^+}^s), \quad \forall s \in \bar{S} \setminus \{\bar{s}\}, \quad (25a)$$

⁴We stress that the slow time scale dynamics (22a) yields as output the first fast state of the slow period (and not the next slow state). Thus, the slow time scale dynamics (22a) is *not* a dynamics from one slow state to the next slow state.

and, for each $s \in \mathbb{S}$, a family $\{\rho_{(s,f):(s,f)^+}^{\text{sf}}\}_{f \in \mathbb{F} \setminus \{\bar{f}\}}$ of *fast time scale stochastic kernels*

$$\rho_{(s,f):(s,f)^+}^{\text{sf}} : \mathbb{W}_s^{\text{s}} \times \underbrace{\prod_{f'=\underline{f}^+}^f \mathbb{W}_{(s,f')}^{\text{sf}}}_{\text{interval } [s^-, s[} \longrightarrow \Delta(\mathbb{W}_{(s,f)^+}^{\text{sf}}),$$

$$\forall s \in \mathbb{S}, \quad \forall f \in \mathbb{F} \setminus \{\bar{f}\}, \quad (25b)$$

with the convention that the Cartesian products of spaces in Equations (25a) and (25b) reduce to the empty set when the upper index of the Cartesian product is strictly lower than the corresponding lower index. Note that, for a given $s \in \mathbb{S}$, each fast time scale stochastic kernel $\rho_{(s,f):(s,f)^+}^{\text{sf}}$, only depends on the noises of the slow time block s .⁵

4.2 Formulation of a two time scale optimization problem on the product timeline

To apply Theorem 7, we introduce sets associated with the extended timeline (21) by

$$\mathbb{X}_{(s,f)} = \begin{cases} \mathbb{X}_s^{\text{s}} & \text{if } f = \bar{f} \\ \mathbb{X}_{(s,f)}^{\text{sf}} & \text{if } f \neq \bar{f} \end{cases}, \quad \forall (s,f) \in \overline{\mathbb{S} \times \mathbb{F}}, \quad (26a)$$

$$\mathbb{U}_{(s,f)} = \begin{cases} \mathbb{U}_s^{\text{s}} & \text{if } f = \bar{f} \\ \mathbb{U}_{(s,f)}^{\text{sf}} & \text{if } f \neq \bar{f} \end{cases}, \quad \forall (s,f) \in \overline{\mathbb{S} \times \mathbb{F}} \setminus \{(\bar{s}, \bar{f})\}, \quad (26b)$$

$$\mathbb{W}_{(s,f)} = \begin{cases} \mathbb{W}_s^{\text{s}} & \text{if } f = \underline{f} \\ \mathbb{W}_{(s,f)}^{\text{sf}} & \text{if } f \neq \underline{f} \end{cases}, \quad \forall (s,f) \in \mathbb{S} \times \mathbb{F}, \quad (26c)$$

with the particular case of the extra initial slow time

$$\mathbb{W}_{(\underline{s}^-, \bar{f})} = \mathbb{X}_{\underline{s}^-}^{\text{s}}, \quad (26d)$$

and a family of state dynamics $\mathcal{F}_{(s,f)} : \mathbb{X}_{(s,f)} \times \mathbb{U}_{(s,f)} \times \mathbb{W}_{(s,f)^+} \rightarrow \mathbb{X}_{(s,f)^+}$ defined by

$$\mathcal{F}_{(s,f)} = \begin{cases} \mathcal{F}_s^{\text{s}} & \text{if } f = \bar{f} \\ \mathcal{F}_{(s,f)}^{\text{sf}} & \text{if } f \neq \bar{f} \end{cases}, \quad \forall (s,f) \in \overline{\mathbb{S} \times \mathbb{F}} \setminus \{(\bar{s}, \bar{f})\}. \quad (27)$$

From these sets, we deduce the history sets and the histories for all $(s,f) \in \overline{\mathbb{S} \times \mathbb{F}}$

$$\mathbb{H}_{(s,f)} = \mathbb{W}_{(\underline{s}^-, \bar{f})} \times \prod_{(\underline{s}', \underline{f}') \preceq (s', f') \preceq (s,f)} \left(\mathbb{U}_{(s', f')^-} \times \mathbb{W}_{(s', f')} \right), \quad (28a)$$

$$h_{(s,f)} = \left(w_{(\underline{s}^-, \bar{f})}, \left(u_{(s', f')^-}, w_{(s', f')} \right)_{(\underline{s}', \underline{f}') \preceq (s', f') \preceq (s,f)} \right), \quad (28b)$$

⁵The (constant) assumption (25a) and the (single block) assumption (25b) correspond to stochastic independence between time blocks, and will be useful in the proof of Proposition 9.

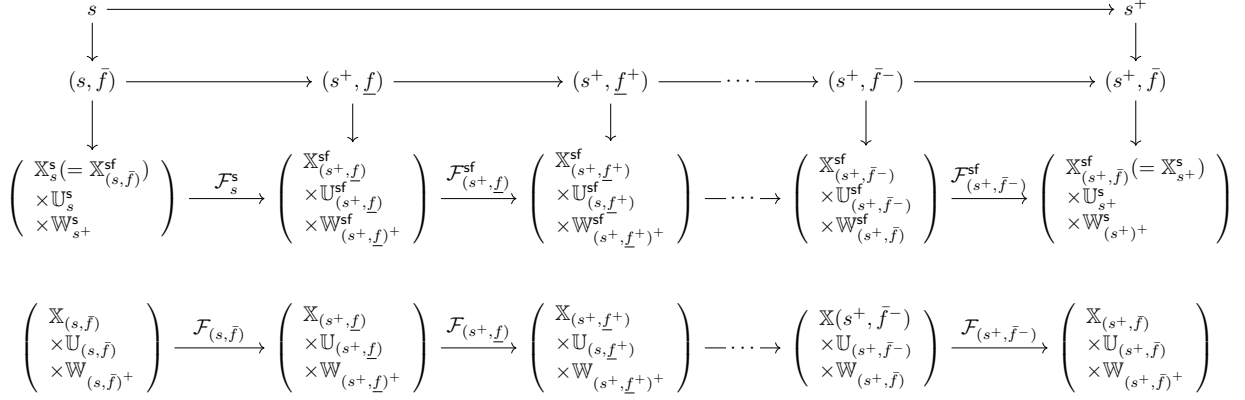


Figure 3: Original dynamics and their reformulation on the product timeline on the slow time interval $[s, s^+]$

and, for suitable indices, the partial history sets and the partial histories

$$\mathbb{H}_{(s, f):(s', f')} = \prod_{(s, f) \preceq (s'', f'') \preceq (s', f')} (\mathbb{U}_{(s'', f'')^-} \times \mathbb{W}_{(s'', f'')}) , \quad (29a)$$

$$h_{(s, f):(s', f')} = \left((u_{(s'', f'')^-}, w_{(s'', f'')})_{(s, f) \preceq (s'', f'') \preceq (s', f')} \right) . \quad (29b)$$

The criterion formulated in Equation (24) combined with state dynamics leads to a criterion $j : \mathbb{H}_{(\bar{s}, \bar{f})} \rightarrow \mathbb{R}$.

Based on the stochastic kernels (25a) and (25b), we introduce stochastic kernels $\rho_{(s, f):(s, f)^+}$ associated with the extended timeline (21), for each $(s, f) \in \overline{\mathbb{S} \times \mathbb{F}} \setminus \{\bar{s}, \bar{f}\}$, by $\rho_{(s, f):(s, f)^+} : \mathbb{H}_{(s, f)} \rightarrow \Delta(\mathbb{W}_{(s, f)^+})$ with

$$\begin{aligned} & \rho_{(s, f):(s, f)^+} (dw_{(s, f)^+} \mid h_{(s, f)}) \\ &= \begin{cases} \rho_{s: s^+}^s (dw_{s^+}^s) & \text{if } f = \bar{f} , \\ \rho_{(s, f):(s, f)^+}^{sf} (dw_{(s, f)^+}^{sf} \mid w_s^s, w_{(s, f^+)}^{sf}, \dots, w_{(s, f)}^{sf}) & \text{if } f \neq \bar{f} . \end{cases} \end{aligned}$$

Note that, for $f \neq \bar{f}$, the kernels $\rho_{(s, f):(s, f)^+} : \mathbb{H}_{(s, f):(s, f)} \rightarrow \Delta(\mathbb{W}_{(s, f)^+})$, only depend on the partial history uncertainty part from (s, \underline{f}) to (s, f) , and not on the (past) controls.

The components of the problem are now formulated on the extended timeline $\overline{\mathbb{S} \times \mathbb{F}}$, already identified with the time set $\llbracket 0, T \rrbracket$. Thus, we are in the framework of §2.2 and we aim at solving an optimization problem as formulated in Equation (4).

4.3 Two time scale decomposition

The existence of Bellman equations for a two time scale optimization problem is given by the following proposition.

Proposition 9 Consider a two time scale optimization problem as formulated in §4.1 and §4.2. The optimization problem (4) has a solution given by a dynamic programming equation at the slow scale. More precisely, let $(V_s)_{s \in \bar{\mathcal{S}}}$ be given by $V_{\bar{s}} = \Lambda_{\bar{s}}$ and, for $s \in \bar{\mathcal{S}} \setminus \{\bar{s}\}$, by the backward induction

$$\begin{aligned}
V_s(x_s^s) &= \inf_{u_s^s \in \mathbb{U}_s^s} \int_{\mathbb{W}_{s^+}^s} \rho_{s:s^+}^s(dw_{s^+}^s) \\
&\quad \inf_{u_{(s^+, \underline{f})}^{\text{sf}} \in \mathbb{U}_{(s^+, \underline{f})}^{\text{sf}}} \int_{\mathbb{W}_{(s^+, \underline{f}^+)}^{\text{sf}}} \rho_{(s^+, \underline{f}):(s^+, \underline{f}^+)}^{\text{sf}}(dw_{(s^+, \underline{f}^+)}^{\text{sf}} | w_{s^+}^s) \cdots \\
&\quad \inf_{u_{(s^+, \bar{f}^-)}^{\text{sf}} \in \mathbb{U}_{(s^+, \bar{f}^-)}^{\text{sf}}} \int_{\mathbb{W}_{(s^+, \bar{f})}^{\text{sf}}} \rho_{(s^+, \bar{f}^-):(s^+, \bar{f})}^{\text{sf}}(dw_{(s^+, \bar{f})}^{\text{sf}} | w_{s^+}^s, w_{(s^+, \underline{f}^+)}^{\text{sf}}, \cdots, w_{(s^+, \bar{f}^-)}^{\text{sf}}) \\
&\quad \left(\Lambda_s(x_s^s, u_s^s, w_{s^+}^s, \cdots, u_{(s^+, \bar{f}^-)}^{\text{sf}}, w_{(s^+, \bar{f})}^{\text{sf}}) \right. \\
&\quad \left. + V_{s^+}(\mathcal{F}_{s:s^+}(x_s^s, u_s^s, w_{s^+}^s, \cdots, u_{(s^+, \bar{f}^-)}^{\text{sf}}, w_{(s^+, \bar{f})}^{\text{sf}})) \right), \tag{30}
\end{aligned}$$

where $\mathcal{F}_{s:s^+}$ is the composition $\mathcal{F}_{s:s^+} = \mathcal{F}_{(s^+, \bar{f}^-)}^{\text{sf}} \circ \cdots \circ \mathcal{F}_{(s^+, \underline{f})}^{\text{sf}} \circ \mathcal{F}_s^s$ associated with the state dynamics defined in (22). Then, the value of the optimization problem (4) is given by $V_{\underline{s}^-}(x_{\underline{s}^-}^s)$.

Proof. The proof is an application of Theorem 7 with the help of Remarks 5 and 6. First, we have re-framed in §4.2 the two time scale optimization problems described in §4.1 in the formalism of §2.2 with the help of the extended timeline (21). Second, as we are given state dynamics (27) on the extended timeline and thanks to Remark 5, we obtain a state reduction at times $\{(s, \bar{f})\}_{s \in \mathcal{S}}$ by composition of the state dynamics. Moreover, as the slow time scale kernels given by Equation (25a) are constant, the state reduction across the slow time scale is compatible with the stochastic kernels (see Remark 6). Third, the case of a time additive criterion has been considered in Corollary 8. We are thus able to apply Theorem 7 and obtain the slow time scale Bellman recursion (30) as a special case of Equation (17). \square

The slow time scale Bellman equation (30) is as difficult to solve as the Bellman equation on the extended timeline. However, the interest of (30) lies elsewhere. Imagine that one is able to obtain, in a relatively easy way, lower \underline{V}_s and upper \overline{V}_s approximations of V_s in (30). Then, by replacing the last term V_{s^+} of (30) by either \underline{V}_{s^+} or \overline{V}_{s^+} , one can now solve a (lower or upper) surrogate of Equation (30) by any suitable method. For instance, one could use scenario decomposition methods, like progressive hedging [11], that do not require statistical independence of noises within the slow time interval $[s, s^+]$. Thus, the two time scale optimization problem as formulated in §4.1 and §4.2 can be approximatively solved, from below and from above, by a mix of slow time scale dynamic programming and of (for example) progressive hedging (or any other method, including dynamic programming). This is the approach we followed to numerically tackle a battery management problem over 20

years involving both the battery operating (with a fast time step of 1 minutes) and the battery replacement (with a slow time step of one day). Despite involving several millions time steps, we were able to solve the problem using a reasonable CPU time [10].

4.4 Link with the classical framework of stochastic optimal control

The property that the stochastic kernels (25) do not depend on any decision variable makes it possible to build a probability $\rho_{(\underline{s}, \underline{f}):(\bar{s}, \bar{f})}$ on the product space $\mathbb{W}_{(\underline{s}, \underline{f}):(\bar{s}, \bar{f})}$ by

$$\begin{aligned} & \rho_{(\underline{s}, \underline{f}):(\bar{s}, \bar{f})} \\ &= \bigotimes_{s \in \bar{S}} \left(\rho_{s:s^+}^s(dw_{s^+}^s) \right. \\ & \quad \otimes \rho_{(s^+, \underline{f}): (s^+, \underline{f}^+)}^{\text{sf}}(dw_{(s^+, \underline{f}^+)}^{\text{sf}} | w_{s^+}^s) \otimes \cdots \\ & \quad \left. \otimes \rho_{(s^+, \bar{f}^-): (s^+, \bar{f})}^{\text{sf}}(dw_{(s^+, \bar{f})}^{\text{sf}} | w_{s^+}^s, w_{(s^+, \underline{f}^+)}^{\text{sf}}, \cdots, w_{(s^+, \bar{f}^-)}^{\text{sf}}) \right). \end{aligned} \quad (31)$$

Then Problem (4) may be rewritten using this probability as

$$\begin{aligned} V_{\underline{s}^-}(x_{\underline{s}^-}^s) &= \inf_{\gamma} \int_{\mathbb{W}_{(\underline{s}, \underline{f}):(\bar{s}, \bar{f})}} \left(\Lambda_{\bar{s}}(x_{\bar{s}}^s) + \right. \\ & \quad \left. \sum_{s \in \bar{S}} \Lambda_s(x_{s^-}^s, u_{s^-}^s, w_s^s, \{x_{(s,f)}^{\text{sf}}, u_{(s,f)}^{\text{sf}}, w_{(s,f)^+}^{\text{sf}}\}_{f \in \mathbb{F} \setminus \{\bar{f}\}}) \right) \\ & \quad \rho_{(\underline{s}, \underline{f}):(\bar{s}, \bar{f})}(dw_{\underline{s}}^s, dw_{(\underline{s}, \underline{f}^+)}^{\text{sf}} \cdots dw_{(\bar{s}, \bar{f}^-)}^{\text{sf}}, dw_{(\bar{s}, \bar{f})}^{\text{sf}}) \end{aligned} \quad (32a)$$

$$\begin{aligned} s.t. \quad x_{(s,f)^+}^{\text{sf}} &= \mathcal{F}_{(s,f)}^{\text{sf}}(x_{(s,f)}^{\text{sf}}, u_{(s,f)}^{\text{sf}}, w_{(s,f)^+}^{\text{sf}}), \\ \forall s \in \bar{S}, \quad \forall f \in \mathbb{F} \setminus \{\bar{f}\}, \end{aligned} \quad (32b)$$

$$x_{(s^+, \underline{f})}^s = \mathcal{F}_s^s(x_s^s, u_s^s, w_{s^+}^s), \quad \forall s \in \bar{S} \setminus \{\bar{s}\}, \quad (32c)$$

$$u_s^s = \gamma_s(\{u_{(s', f')}, w_{(s', f')^+}\}_{(s', f') \prec (s, \bar{f})}), \quad \forall s \in \bar{S} \setminus \{\bar{s}\}, \quad (32d)$$

$$\begin{aligned} u_{(s,f)}^{\text{sf}} &= \gamma_{(s,f)}(\{u_{(s', f')}, w_{(s', f')^+}\}_{(s', f') \prec (s, f)}), \\ \forall s \in \bar{S}, \quad \forall f \in \mathbb{F} \setminus \{\bar{f}\}. \end{aligned} \quad (32e)$$

The integral cost given in the right hand side of Equation (32a) can be reformulated as an expectation, denoted by \mathbb{E} , with respect to the probability $\rho_{(\underline{s}, \underline{f}):(\bar{s}, \bar{f})}$ by introducing random variables for the exogeneous noises as projection mappings from $\mathbb{W}_{(\underline{s}, \underline{f}):(\bar{s}, \bar{f})}$ to $\mathbb{W}_{(s, f)}$ for all $(s, f) \in \bar{S} \times \mathbb{F}$

$$\mathbf{W}_{(s, f)} : \mathbb{W}_{(\underline{s}, \underline{f}):(\bar{s}, \bar{f})} \rightarrow \mathbb{W}_{(s, f)}, \quad \forall (s, f) \in \bar{S} \times \mathbb{F}, \quad (33)$$

and obtaining random variables for the states and the control through the dynamics equations (32b)–(32c) and the feedback equations (32d)–(32e).

This leads to a reformulation of Problem 32 as a classical stochastic optimal control problem

$$\inf \mathbb{E} \left[\sum_{s \in \mathcal{S}} \Lambda_s(\mathbf{X}_{s^-}^s, \mathbf{U}_{s^-}^s, \mathbf{W}_s, \{\mathbf{X}_{(s,f)}^{\text{sf}}, \mathbf{U}_{(s,f)}^{\text{sf}}, \mathbf{W}_{(s,f)^+}^f\}_{f \in \mathcal{F} \setminus \{\bar{f}\}}) + \Lambda_{\bar{s}}(\mathbf{X}_{\bar{s}}^s) \right] \quad (34a)$$

$$s.t. \mathbf{X}_{(s,f)^+}^{\text{sf}} = \mathcal{F}_{(s,f)}^{\text{sf}}(\mathbf{X}_{(s,f)}^{\text{sf}}, \mathbf{U}_{(s,f)}^{\text{sf}}, \mathbf{W}_{(s,f)^+}^{\text{sf}}), \quad \forall s \in \mathcal{S}, \quad \forall f \in \mathcal{F} \setminus \{\bar{f}\}, \quad (34b)$$

$$\mathbf{X}_{(s^+,f)}^{\text{sf}} = \mathcal{F}_s^s(\mathbf{X}_s^s, \mathbf{U}_s^s, \mathbf{W}_{s^+}^s), \quad \forall s \in \mathcal{S} \setminus \{\bar{s}\}, \quad (34c)$$

$$\mathbf{U}_s^s \in \mathbb{U}_s^s, \quad \forall s \in \bar{\mathcal{S}}, \quad (34d)$$

$$\sigma(\mathbf{U}_s^s) \subset \sigma(\{\mathbf{W}_{s'}^s\}_{s' \preceq s}, \{\mathbf{W}_{(s',f')}^{\text{sf}}\}_{(s',f') \preceq (s,\bar{f})}), \quad \forall s \in \bar{\mathcal{S}}, \quad (34e)$$

$$\mathbf{U}_{(s,f)}^{\text{sf}} \in \mathbb{U}_{(s,f)}^{\text{sf}}, \quad \forall s \in \mathcal{S}, \quad \forall f \in \mathcal{F} \setminus \{\bar{f}\}, \quad (34f)$$

$$\sigma(\mathbf{U}_{(s,f)}^{\text{sf}}) \subset \sigma(\{\mathbf{W}_{s'}^s\}_{s' \preceq s}, \{\mathbf{W}_{(s',f')}^{\text{sf}}\}_{(s',f') \preceq (s,f)}), \quad \forall s \in \mathcal{S}, \quad \forall f \in \mathcal{F} \setminus \{\bar{f}\}, \quad (34g)$$

where the two feedback constraints in (32d) and (32e) are reformulated as measurability constraints (34e) and (34g) (of course, a formal equivalence would require to be more specific about spaces to use Doob functional Lemma).

4.5 Illustration with the crude oil procurement problem

Crude oil procurement is the part of the oil supply chain that sits between the production of crude oil and its processing in a refinery. The goal of procurement is to purchase crude oil from various suppliers around the world and having it delivered in time to the refinery to be processed. As illustrated in Figure 4, every month (on the bottom line) a refinery receives crudes that have been bought during the 8 previous weeks (on the upper line).

The problem naturally displays two time scales. On the one hand, deliveries to the refinery are made at the beginning of each month, and crude consumption is set once a month. On the other hand, crude oil shipments can be purchased at the frequency of the week; every week, a selection of shipments is presented to the decision-maker who must decide which shipments to purchase. Following the construction of the extended timeline

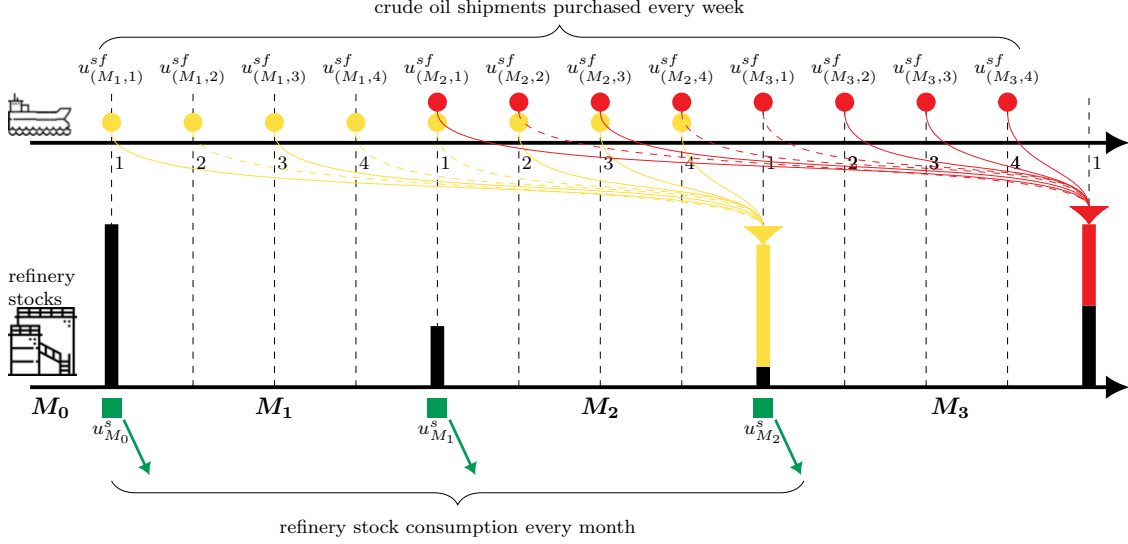


Figure 4: Procurement of crude oil over 3 months M_1 , M_2 and M_3 , where \circ denotes purchase decisions and \square denotes consumption decisions

in (21), we represent by the sequence

$$\begin{aligned}
 & (M_0, 5) \\
 & \prec (M_1, 1) \prec (M_1, 2) \prec (M_1, 3) \prec (M_1, 4) \prec (M_1, 5) \\
 & \prec (M_2, 1) \prec (M_2, 2) \prec (M_2, 3) \prec (M_2, 4) \prec (M_2, 5) \\
 & \prec (M_3, 1) \prec (M_3, 2) \prec (M_3, 3) \prec (M_3, 4) \prec (M_3, 5)
 \end{aligned} \tag{35}$$

the timeline associated with Figure 4 (notice that we consider that a month is made of 4 weeks). The initial stage $(M_0, 5)$ corresponds to the additional stage $(\underline{s}^-, \bar{f})$ in (21). The stages $(M_1, 5)$ and $(M_2, 5)$ both represent the “end of the month” when a consumption decision (slow scale decision u_s^s on the bottom line of Figure 4) is taken.

We now illustrate how the crude oil procurement problem can be put in the form of a two time scale optimization problem such as presented in §4.1. For this purpose, we proceed to the identifications in Table 1.

We call s -buffer (resp. s^- -buffer), the temporary stock that is created at the beginning of the month s (resp. s^-) and that will be delivered two months after. For instance, in Figure 4, the yellow disks represent the M_1 -buffer and the red disks represent the M_2 -buffer. We introduce the state variable $x_{(s,f)}^{sf} = (s^- \text{-buffer}, s \text{-buffer}, \text{refinery stocks})$, together with the accumulation dynamics $\mathcal{F}_{(s,f)}^{sf}$ for the buffers, and the accumulation dynamics \mathcal{F}_s^s for the stocks. Supposing that the products prices are independent month by month, we represent this assumption by a family of constant kernels $\{\rho_{s:s+}^s\}_{s \in \bar{S} \setminus \{\bar{s}\}}$. By contrast, we do not assume that the crude prices are independent week by week, and the possible dependency is modeled by stochastic kernels $\rho_{(s,f):(s,f)^+}^{sf}$ for $f \in \bar{F} \setminus \{\bar{f}\}$.

Notations from §4.1	Crude oil procurement
\mathbf{S}	set of months during which we manage the refinery; in Figure 4, $\mathbf{S} = \{M_1, M_2, M_3\}$
\mathbf{F}	set of weeks in each month; in Figure 4, $\mathbf{F} = \{1, 2, 3, 4, 5\}$
\mathbb{U}_s^s	set of crude oil consumptions during the month s^+
$\mathbb{W}_{s^+}^s$	set of product prices for the month s^+
$\mathbb{U}_{(s,f)}^{sf}$	set of crude shipments purchased in week (s, f)
$\mathbb{W}_{(s,f)^+}^{sf}$	set of crude oil prices in week (s, f)
$\mathcal{F}_{(s,f)}^{sf}$	accumulation of shipments purchased in (s, f)
\mathcal{F}_s^s	delivery of orders and consumption of crude oil for the month s^+
Λ_s	operational costs during the month s (crude oil purchases during s - earnings from production)
Λ_{M_4}	end cost associated with the state $x_{M_3}^s = x_{(M_3,5)}^{sf}$ valuation of the buffers and stocks in the refinery before the beginning of the month M_4

Table 1: Identification of the elements introduced in §4.1 with elements of the crude oil procurement problem

Now that all the elements from §4.1 have been identified, Proposition 9 enables us to write a dynamic programming equation such as (30) at the scale of the month, without losing the time-dependency of crude prices inside the month. This illustration stems from a research work done in partnership with TotalEnergies, in the context of a PhD thesis [8].

5 Conclusion and perspectives

As said in the introduction, decomposition methods are appealing to tackle multistage stochastic optimization problems, as they are naturally large scale. The most common approaches are time decomposition (and state-based resolution methods, like stochastic dynamic programming, in stochastic optimal control), and scenario decomposition (like progressive hedging in stochastic programming).

This paper is part of a general research program that consists in *mixing* different decomposition bricks. Space decomposition methods have been investigated in [1] and [4]. Here, we have tackled the issue of using time blocks decomposition in such a way that stochastic dynamic programming is used at the slow time scale with an appropriate white noise assumption, whereas stochastic programming methods such as progressive hedging can be used at the fast time scale where such an independence assumption does not hold. This approach paves the way of mixing time decomposition with scenario decomposition. For this purpose, we have revisited the notion of state, and have provided a way to perform time

decomposition but only across specified time blocks.

Acknowledgements. We thank Roger Wets for fruitful discussions about the possibility of mixing stochastic dynamic programming with Progressive Hedging.

A Technical details and proofs

We introduce the notations

$$\begin{aligned}\mathbb{W}_{r:t} &= \prod_{s=r}^t \mathbb{W}_s, \quad 0 \leq r \leq t \leq T, \\ \mathbb{U}_{r:t} &= \prod_{s=r}^t \mathbb{U}_s, \quad 0 \leq r \leq t \leq T-1\end{aligned}\tag{36}$$

Let $0 \leq r \leq s \leq t \leq T$. From a history $h_t \in \mathbb{H}_t$, we can extract the $(r:s)$ -*history uncertainty part*

$$[h_t]_{r:s}^{\mathbb{W}} = (w_r, \dots, w_s) = w_{r:s} \in \mathbb{W}_{r:s}, \quad 0 \leq r \leq s \leq t, \tag{37a}$$

the $(r:s)$ -*history control part* (notice that the indices are special)

$$\begin{aligned}[h_t]_{r:s}^{\mathbb{U}} &= (u_{r-1}, \dots, u_{s-1}) \\ &= u_{r-1:s-1} \in \mathbb{U}_{r-1:s-1}, \quad 1 \leq r \leq s \leq t.\end{aligned}\tag{37b}$$

Flows Let r and t be given such that $0 \leq r < t \leq T$. For a $(r:t-1)$ -history feedback $\gamma = \{\gamma_s\}_{s=r, \dots, t-1} \in \Gamma_{r:t-1}$, we define the *flow* $\Phi_{r:t}^\gamma$ by

$$\begin{aligned}\Phi_{r:t}^\gamma : \mathbb{H}_r \times \mathbb{W}_{r+1:t} &\rightarrow \mathbb{H}_t \\ (h_r, w_{r+1:t}) &\mapsto \left(h_r, \gamma_r(h_r), w_{r+1}, \right. \\ &\quad \left. \gamma_{r+1}(h_r, \gamma_r(h_r), w_{r+1}), w_{r+2}, \right. \\ &\quad \left. \dots, \gamma_{t-1}(h_{t-1}), w_t \right).\end{aligned}\tag{38a}$$

Otherwise stated, the flow is given by

$$\Phi_{r:t}^\gamma(h_r, w_{r+1:t}) = (h_r, u_r, w_{r+1}, u_{r+1}, w_{r+2}, \dots, u_{t-1}, w_t), \tag{38b}$$

$$\text{with } h_s = (h_r, u_r, w_{r+1}, \dots, u_{s-1}, w_s), \quad r < s \leq t, \tag{38c}$$

$$\text{and } u_s = \gamma_s(h_s), \quad r \leq s \leq t-1. \tag{38d}$$

When $0 \leq r = t \leq T$, we put $\Phi_{r:r}^\gamma : \mathbb{H}_r \rightarrow \mathbb{H}_r$, $h_r \mapsto h_r$. With this convention, the expression $\Phi_{r:t}^\gamma$ makes sense when $0 \leq r \leq t \leq T$. The mapping $\Phi_{r:t}^\gamma$ gives the history at time t as a

function of the initial history h_r at time r and of the history feedbacks $\{\gamma_s\}_{s=r,\dots,t-1} \in \Gamma_{r:t-1}$. An immediate consequence of this definition are the *flow properties*:

$$\begin{aligned} \Phi_{r:t+1}^\gamma(h_r, w_{r+1:t+1}) &= \left(\Phi_{r:t}^\gamma(h_r, w_{r+1:t}), \gamma_t(\Phi_{r:t}^\gamma(h_r, w_{r+1:t})), w_{t+1} \right), \\ &0 \leq r \leq t \leq T-1, \end{aligned} \quad (39a)$$

$$\begin{aligned} \Phi_{r:t}^\gamma(h_r, w_{r+1:t}) &= \Phi_{r+1:t}^\gamma((h_r, \gamma_r(h_r), w_{r+1}), w_{r+2:t}), \quad 0 \leq r < t \leq T. \end{aligned} \quad (39b)$$

Definition 10 *Let r and t be given such that $0 \leq r \leq t \leq T$.*

- *When $0 \leq r < t \leq T$, for a $(r:t-1)$ -history feedback $\gamma = \{\gamma_s\}_{s \in \llbracket r, t-1 \rrbracket} \in \Gamma_{r:t-1}$, and for a family $\{\rho_{s-1:s}\}_{r+1 \leq s \leq t}$ of stochastic kernels $\rho_{s-1:s} : \mathbb{H}_{s-1} \rightarrow \Delta(\mathbb{W}_s)$, $s \in \llbracket r+1, t \rrbracket$, we define a stochastic kernel $\rho_{r:t}^\gamma : \mathbb{H}_r \rightarrow \Delta(\mathbb{H}_t)$ such that, for any numerical function $\varphi \in \mathbb{L}_+^0(\mathbb{H}_t, \mathcal{H}_t)$ ⁶, we have that*

$$\begin{aligned} &\int_{\mathbb{H}_t} \varphi(h'_r, h'_{r+1:t}) \rho_{r:t}^\gamma(dh'_t | h_r) \\ &= \int_{\mathbb{W}_{r+1:t}} \varphi(\Phi_{r:t}^\gamma(h_r, w_{r+1:t})) \\ &\quad \prod_{s=r+1}^t \rho_{s-1:s}(dw_s | \Phi_{r:s-1}^\gamma(h_r, w_{r+1:s-1})). \end{aligned} \quad (40)$$

- *When $0 \leq r = t \leq T$, we define $\rho_{r:r}^\gamma : \mathbb{H}_r \rightarrow \Delta(\mathbb{H}_r)$ by $\rho_{r:r}^\gamma(dh'_r | h_r) = \delta_{h_r}(dh'_r)$.*

The stochastic kernels $\rho_{r:t}^\gamma$ on \mathbb{H}_t , given by (40), are of the form $\rho_{r:t}^\gamma(dh'_t | h_r) = \rho_{r:t}^\gamma(dh'_r dh'_{r+1:t} | h_r) = \delta_{h_r}(dh'_r) \otimes \varrho_{r:t}^\gamma(dh'_{r+1:t} | h_r)$, where, for each $h_r \in \mathbb{H}_r$, the probability distribution $\varrho_{r:t}^\gamma(dh'_{r+1:t} | h_r)$ only charges the histories visited by the flow from $r+1$ to t . The construction of the stochastic kernels $\rho_{r:t}^\gamma$ is developed in [3, p. 190] for relaxed history feedbacks and obtained by using [3, Proposition 7.45].

Proposition 11 *The family $\{\rho_{s:t}^\gamma\}_{s=r,\dots,t}$ of stochastic kernels of Definition 10 has the flow property:*

$$\begin{aligned} \rho_{s:t}^\gamma(dh'_t | h_s) &= \int_{\mathbb{W}_{s+1}} \rho_{s:s+1}(dw_{s+1} | h_s) \\ &\quad \rho_{s+1:t}^\gamma(dh'_t | (h_s, \gamma_s(h_s), w_{s+1})), \quad \forall s < t. \end{aligned} \quad (41)$$

⁶space of universally measurable nonnegative numerical functions over \mathbb{H}_t : see Footnote 2

Proof. Let $s < t$. For any $\varphi \in \mathbb{L}_+^0(\mathbb{H}_t, \mathcal{H}_t)$, we have that

$$\int_{\mathbb{H}_t} \varphi(h'_s, h'_{s+1:t}) \rho_{s:t}^\gamma(dh'_t | h_s) \quad (42a)$$

$$= \int_{\mathbb{W}_{s+1:t}} \varphi(\Phi_{s:t}^\gamma(h_s, w_{s+1:t}))$$

$$\prod_{s'=s+1}^t \rho_{s'-1:s'}(dw_{s'} | \Phi_{s:s'-1}^\gamma(h_s, w_{s+1:s'-1})) \quad (\text{by Definition (40)})$$

$$= \int_{\mathbb{W}_{s+1:t}} \varphi(\Phi_{s:t}^\gamma(h_s, w_{s+1:t})) \rho_{s:s+1}(dw_{s+1} | h_s)$$

$$\prod_{s'=s+2}^t \rho_{s'-1:s'}(dw_{s'} | \Phi_{s:s'-1}^\gamma(h_s, w_{s+1:s'-1}))$$

$$= \int_{\mathbb{W}_{s+1:t}} \varphi(\Phi_{s+1:t}^\gamma((h_s, \gamma_s(h_s), w_{s+1}), w_{s+2:t}))$$

$$\rho_{s:s+1}(dw_{s+1} | h_s)$$

$$\prod_{s'=s+2}^t \rho_{s'-1:s'}(dw_{s'} | \Phi_{s+1:s'-1}^\gamma((h_s, \gamma_s(h_s), w_{s+1}), w_{s+2:s'-1})) \quad (\text{by the flow property (39b)})$$

$$= \int_{\mathbb{W}_{s+1}} \rho_{s:s+1}(dw_{s+1} | h_s)$$

$$\int_{\mathbb{W}_{s+2:t}} \varphi(\Phi_{s+1:t}^\gamma((h_s, \gamma_s(h_s), w_{s+1}), w_{s+2:t}))$$

$$\prod_{s'=s+2}^t \rho_{s'-1:s'}(dw_{s'} | \Phi_{s+1:s'-1}^\gamma((h_s, \gamma_s(h_s), w_{s+1}), w_{s+2:s'-1}))$$

$$\quad (\text{by Fubini Theorem [7, p.137]})$$

$$= \int_{\mathbb{W}_{s+1}} \rho_{s:s+1}(dw_{s+1} | h_s) \int_{\mathbb{H}_t} \varphi((h'_s, \gamma_s(h'_s), w'_{s+1}), h'_{s+2:t})$$

$$\rho_{s+1:t}^\gamma(dh'_t | (h_s, \gamma_s(h_s), w_{s+1})) \quad (\text{by Definition (40)})$$

$$= \int_{\mathbb{H}_t} \varphi((h'_s, \gamma_s(h'_s), w'_{s+1}), h'_{s+2:t})$$

$$\int_{\mathbb{W}_{s+1}} \rho_{s:s+1}(dw_{s+1} | h_s) \rho_{s+1:t}^\gamma(dh'_t | (h_s, \gamma_s(h_s), w_{s+1})) \quad (42b)$$

by Fubini Theorem. As the two expressions (42a) and (42b) are equal for any $\varphi \in \mathbb{L}_+^0(\mathbb{H}_t, \mathcal{H}_t)$, we deduce the flow property (41). \square

Proof of Theorem 1 We only give a sketch of the proof, as it is a variation on different results of [3], the framework of which we follow.

Proof. We take the history space \mathbb{H}_t for state space, and the state dynamics

$$f(h_t, u_t, w_{t+1}) = (h_t, u_t, w_{t+1}) = h_{t+1} \in \mathbb{H}_{t+1} = \mathbb{H}_t \times \mathbb{U}_t \times \mathbb{W}_{t+1}. \quad (43)$$

Then, the family $\{\rho_{s-1:s}\}_{s \in \llbracket 1, T \rrbracket}$ of stochastic kernels (1) gives a family of disturbance kernels that do not depend on the current control. The criterion to be minimized (2) is a function of the history at time T , thus of the state at time T . The optimization problem defined by the associated value function (3) is thus a finite horizon model with a final cost and we are minimizing over the so-called state-feedbacks. Then, the proof of Theorem 1 follows from the results developed in Chap. 7, 8 and 10 of [3] in a Borel setting. Since we are considering a finite horizon model with a final cost, we detail the steps needed to use the results of [3, Chap. 8].

The final cost at time T can be turned into an instantaneous cost at time $T - 1$ by inserting the state dynamics (43) in the final cost. Getting rid of the disturbance in the expected cost by using the disturbance kernel is standard practice. Then, we can turn this non-homogeneous finite horizon model into a finite horizon model with homogeneous dynamics and costs by following the steps of [3, Chap. 10]. Using [3, Proposition 8.2], we obtain that the family of optimization problems defined by the associated value functions (3), when minimizing over the relaxed state feedbacks, satisfies the Bellman equation (6); we conclude with [3, Proposition 8.4] which covers the minimization over state feedbacks. \square

To summarize, Theorem 1 is valid under the general Borel assumptions of [3, Chap. 8] and with the specific (F^-) assumption needed for [3, Proposition 8.4]; this last assumption is fulfilled here since we have assumed that the criterion (2) is nonnegative.

Proof of Proposition 3

Proof. Let $\tilde{\varphi}_t : \mathbb{X}_t \rightarrow [0, +\infty]$ be a given measurable nonnegative numerical function, and let $\varphi_t : \mathbb{H}_t \rightarrow [0, +\infty]$ be

$$\varphi_t = \tilde{\varphi}_t \circ \theta_t . \quad (44)$$

Let $\varphi_r : \mathbb{H}_r \rightarrow [0, +\infty]$ be the measurable nonnegative numerical function obtained by applying the Bellman operator $\mathcal{B}_{t:r}$ across $(t:r)$ (see (9)) to the measurable nonnegative numerical function φ_t :

$$\varphi_r = \mathcal{B}_{t:r}\varphi_t = \mathcal{B}_{r+1:r} \circ \dots \circ \mathcal{B}_{t:t-1}\varphi_t . \quad (45)$$

We show that there exists a measurable nonnegative numerical function $\tilde{\varphi}_r : \mathbb{X}_r \rightarrow [0, +\infty]$ such that

$$\varphi_r = \tilde{\varphi}_r \circ \theta_r . \quad (46)$$

First, we show by backward induction that, for all $s \in \{r, \dots, t\}$, there exists a measurable nonnegative numerical function $\bar{\varphi}_s$ such that $\varphi_s(h_s) = \bar{\varphi}_s(\theta_r(h_r), h_{r+1:s})$. Second, we prove that the function $\tilde{\varphi}_r = \bar{\varphi}_r$ satisfies (46).

- For $s = t$, we have, by (44) and by (7c), that $\varphi_t(h_t) = \tilde{\varphi}_t(\theta_t(h_t)) = \tilde{\varphi}_t(f_{r:t}(\theta_r(h_r), h_{r+1:t}))$, so that the measurable nonnegative numerical function $\bar{\varphi}_t$ is given by $\tilde{\varphi}_t \circ f_{r:t}$.
- Assume that, at $s + 1$, the result holds true, that is, $\varphi_{s+1}(h_{s+1}) = \bar{\varphi}_{s+1}(\theta_r(h_r), h_{r+1:s+1})$.

Then, by (45),

$$\begin{aligned}
\varphi_s(h_s) &= (\mathcal{B}_{s+1:s}\varphi_{s+1})(h_s) \\
&= \inf_{u_s \in \mathbb{U}_s} \int_{\mathbb{W}_{s+1}} \varphi_{s+1}((h_s, u_s, w_{s+1})) \rho_{s:s+1}(dw_{s+1} | h_s) \\
&\hspace{15em} \text{(by definition (5) of the Bellman operator)} \\
&= \inf_{u_s \in \mathbb{U}_s} \int_{\mathbb{W}_{s+1}} \bar{\varphi}_{s+1}((\theta_r(h_r), (h_{r+1:s}, u_s, w_{s+1}))) \\
&\hspace{15em} \rho_{s:s+1}(dw_{s+1} | h_s) \hspace{5em} \text{(by the induction assumption)} \\
&= \inf_{u_s \in \mathbb{U}_s} \int_{\mathbb{W}_{s+1}} \bar{\varphi}_{s+1}((\theta_r(h_r), (h_{r+1:s}, u_s, w_{s+1}))) \\
&\hspace{15em} \tilde{\rho}_{s:s+1}(dw_{s+1} | (\theta_r(h_r), h_{r+1:s})) \\
&\hspace{15em} \text{(by compatibility (8) of the stochastic kernel)} \\
&= \bar{\varphi}_s(\theta_r(h_r), h_{r+1:s}),
\end{aligned}$$

where

$$\begin{aligned}
\bar{\varphi}_s(x_r, h_{r+1:s}) &= \inf_{u_s \in \mathbb{U}_s} \int_{\mathbb{W}_{s+1}} \bar{\varphi}_{s+1}((x_r, (h_{r+1:s}, u_s, w_{s+1}))) \\
&\hspace{15em} \tilde{\rho}_{s:s+1}(dw_{s+1} | (x_r, h_{r+1:s})).
\end{aligned}$$

Thus, we have shown that the result holds true at time s .

The induction implies that, at time r , the expression of $\varphi_r(h_r)$ is $\varphi_r(h_r) = \bar{\varphi}_r(\theta_r(h_r))$, since the term $h_{r+1:r}$ vanishes. Choosing $\tilde{\varphi}_r = \bar{\varphi}_r$ gives the expected result. \square

References

- [1] Kengy Barty, Pierre Carpentier, and Pierre Girardeau. Decomposition of large-scale stochastic optimal control problems. *RAIRO Operations Research*, 44(3):167–183, 2010.
- [2] R. E. Bellman. *Dynamic Programming*. Princeton University Press, Princeton, N.J., 1957.
- [3] D. P. Bertsekas and S. E. Shreve. *Stochastic Optimal Control: The Discrete-Time Case*. Athena Scientific, Belmont, Massachusetts, 1996.
- [4] Pierre Carpentier, Jean-Philippe Chancelier, Michel De Lara, and François Pacaud. Mixed spatial and temporal decompositions for large-scale multistage stochastic optimization problems. *Journal of Optimization Theory and Applications*, 186(3):985–1005, 2020.
- [5] I. V. Evstigneev. Measurable selection and dynamic programming. *Mathematics of Operations Research*, 1(3):267–272, 1976.

- [6] Onésimo Hernández-Lerma and Jean Bernard Lasserre. *Discrete-Time Markov Control Processes: Basic Optimality Criteria*. Applications of Mathematics Stochastic Modelling and Applied Probability. Springer, New York, 1996.
- [7] M. Loève. *Probability Theory I*. Springer-Verlag, New York, fourth edition, 1977.
- [8] Thomas Martin. *Stochastic Optimization for the Procurement of Crude Oil in Refineries*. PhD thesis, 2021.
- [9] M. L. Puterman. *Markov Decision Processes*. Wiley, New York, 1994.
- [10] Tristan Rigaut. *Time Decomposition Methods For Optimal Management Of Energy Storage Under Stochasticity*. PhD thesis, Université Paris-Est, 2019.
- [11] R.T. Rockafellar and R. J-B. Wets. Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of operations research*, 16(1):119–147, 1991.
- [12] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on stochastic programming: modeling and theory*. The society for industrial and applied mathematics and the mathematical programming society, Philadelphia, USA, 2009.
- [13] Jayakumar Subramanian and Aditya Mahajan. Approximate information state for partially observed systems. *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 1629–1636, 2019.
- [14] P. Whittle. *Optimization over Time: Dynamic Programming and Stochastic Control*, volume 1. John Wiley & Sons, New York, 1982.
- [15] H. S. Witsenhausen. A standard form for sequential stochastic control. *Mathematical Systems Theory*, 7(1):5–11, 1973.
- [16] H. S. Witsenhausen. The intrinsic model for discrete stochastic control: Some open problems. In A. Bensoussan and J. L. Lions, editors, *Control Theory, Numerical Methods and Computer Systems Modelling*, volume 107 of *Lecture Notes in Economics and Mathematical Systems*, pages 322–335. Springer-Verlag, 1975.
- [17] H. S. Witsenhausen. On policy independence of conditional expectations. *Information and Control*, 28(1):65–75, 1975.
- [18] H. S. Witsenhausen. Equivalent stochastic control problems. *Mathematics of Control, Signals, and Systems*, 1(1):3–7, 1988.
- [19] Serdar Yüksel. A universal dynamic program and refined existence results for decentralized stochastic control. *SIAM Journal on Control and Optimization*, 58(5):2711–2739, 2020.