Mathematical Programming Series A manuscript No.
(will be inserted by the editor)

A note on the convergence of the SDDP algorithm

Kengy Barty

the date of receipt and acceptance should be inserted later

Abstract In this paper we are interested in the convergence analysis of the Stochastic Dual Dynamic Algorithm (SDDP) algorithm in a general framework, and regardless of whether the underlying probability space is discrete or not. We consider a convex stochastic control program not necessarily linear and the resulting dynamic programming equation. We prove under mild assumptions that the approximations of the Bellman functions using SDDP converge uniformly to the solution of the dynamic programming equation. We prove also that the distribution of the state variable along iterations converges in distribution to a steady state distribution, which turn out to be the optimal distribution of the state variable. We make use of epi-convergence to assert that the sequence of policies along iterations can be sought among continuous policies and converges pointwise to the optimal policy of the problem. All the mentioned results are provided almost surely with respect to a distribution derived from the optimal policy. Furthermore we propose original proofs of these results which naturally are not based on the finite representation of randomness. It seems that these latter results are new so far.

Keywords Dynamic programming, epi-convergence, locally simplicial, convergence of distributions

Mathematics Subject Classification (2000) 90C15 · 90C39

1 Introduction

In the last three decades, substantial progress have been done concerning the possibility to overcome the well known curse of dimensionality in dynamic programming, specially for convex stochastic control optimization problems. The Stochastic Dual Dynamic Programming algorithm (SDDP) [7,8] of Pereira and Pinto is designed to approximate the Bellman functions of a dynamic programming equation. This algorithm consists in iterating two passes, the first pass selects a trajectory in the space of feasible states of the control problem, and the second pass improves the representation of the Bellman functions. The supposed advantage of this method is to learn along iterations, the optimal distribution of the state variable. By the way, some implementations consider many trajectories for the simulation pass [10], the aim being to learn as quickly as possible the optimal distribution of the state variable. We consider a stochastic dynamic programming equation, the random variables invoked in the formulation are not assumed to be discrete, and their discretization is beyond the scope of this paper. Nevertheless in case of continuous distribution a practical approach to tackle this problem could be to formulate a deterministic equivalent problem using the Stochastic Average Approximation (SAA) and then to apply the SDDP algorithm [15]. Given the stochastic dynamic programming equation, one can consider to formulate a deterministic equivalent of the Bellman operators [14] and then apply the SDDP algorithm to solve the resulting discrete dynamic programming equation. If the underlying probability distribution turns out to be a scenario tree, the SDDP algorithm has been studied for piecewise linear problem in [9] and for

K. Barty
EDF R&D, 1 avenue du Général de Gaulle, F-92141 Clamart Cedex, France E-mail: kengy.barty@edf.fr

a general class of convex problems by [6]. We organized the paper in five sections, the first one being this section. In the second section we introduce the general setup and we formulate the stochastic convex program that we shall study. We recall also the dynamic programming equation and some properties of the Bellman operator. The third section is devoted to the SDDP algorithm formulation, and we introduce the norm which will be employed to analyze the convergence. The fourth section gathers the mains results of the paper, and their proofs, the last section discusses these results.

2 General framework

An extended real-valued function $V : \mathbb{R}^d \to \overline{\mathbb{R}}$ takes values which may be real numbers or may be $+\backslash-\infty$. We denote $\mathrm{dom}(V)$ the effective domain of V which reads $\mathrm{dom}(V) := \{x \in \mathbb{R}^d \mid V(x) < +\infty\}$. We say that V is proper if it nowhere takes the value $-\infty$ and is not identically equal to $+\infty$. Let K be a compact subset of $\mathbb{R}^d$, the space of all continuous real-valued functions of K is denoted $C(K,\mathbb{R})$ and for $V \in C(K,\mathbb{R})$, the supremum norm of V denoted $\|V\|_K$ is defined by $\|V\|_K := \sup_{x\in K} |V(x)|$. When $C(K,\mathbb{R})$ is equipped with the supremum norm, it is a Banach space. Let $V : \mathbb{R}^d \to \overline{\mathbb{R}}$ be an extended real-valued function, if the restriction function $V_{|K} : x \in K \mapsto V(x)$ is such that $V_{|K} \in C(K,\mathbb{R})$, we shall denote in order to avoid cumbersome notations, $\|V\|_K$ instead of $\left\|V_{|K}\right\|_K$. The uniform convergence is the convergence with respect of the supremum norm, it is generally a stronger property than the pointwise convergence, but for a sequence of monotone functions, the uniform convergence can be derived from the pointwise convergence under mild assumptions.

Proposition 1 (Dini) *Let K be a compact subset of $\mathbb{R}^d$, $h^k \in C(K,\mathbb{R})$ and $h \in C(K,\mathbb{R})$, if $h^k \leq h^{k+1}$ and $\forall x \in K$, $\sup_{k\in\mathbb{N}} h^k(x) = h(x)$, then h^k converges uniformly to h on K, that is to say $\lim_{k\to+\infty} \left\|h^k - h\right\|_K = 0$.*

Definition 1 *A function V of $\mathbb{R}^d$ into $\mathbb{R} \cup \{+\infty\}$ is said to be subdifferentiable at a point $x \in \mathbb{R}^d$ if it has a continuous affine minorant which is exact at x. The slope $p \in \mathbb{R}^d$ of such a minorant is called subgradient of V at x, and the set of subgradient of V at x is called subdifferential at x and is denoted $\partial V(x)$. If V is not subdifferentiable at x, we have $\partial V(x) = \emptyset$.*

Under some mild assumptions the subdifferential is not empty.

Proposition 2 *Let V be a convex function of $\mathbb{R}^d$ into $\mathbb{R} \cup \{+\infty\}$, finite and continuous at the point $x \in \mathbb{R}^d$. Then $\partial V(x) \neq \emptyset$ for all $x \in \mathrm{int}(\mathrm{dom}(V))$, and in particular $\partial V(x) \neq \emptyset$.*

Assuming continuity of the objective function in a convex setup is not a restrictive condition because convex functions are very smooth in general, this is what states the next Proposition.

Proposition 3 ([5]) *Every proper convex function on a space of finite dimension is continuous on the interior of its effective domain.*

Let $(\Omega, \mathcal{F}, \mu)$ be a probability space where Ω is an abstract topological space equipped with a σ-field $\mathcal{F}$ of subsets of Ω and $\mu : \mathcal{F} \to [0,1]$ a probability measure. Let $\xi = (\xi_t)_{t=0}^{T-1}$ be a random variable such that $\xi_t = (\xi_{t,i})_{i=1,\dots,q} \in L^\infty(\Omega, \mathcal{F}, \mu; \mathbb{R}^q)$, the latter space being equipped with the infinite norm $\|\xi_t\|_\infty := \min\{\alpha \geq 0 \mid \sum_{i=1}^q |\xi_{t,i}| \leq \alpha \ \mu \text{ a.s.}\}$ and $(\xi_t)_{t=0}^{T-1}$ are stagewise independents. The complete σ-field generated by $(\xi_0, \dots, \xi_t)$ is denoted $\mathcal{F}_t$. For any subset $F \subset \mathbb{R}^d$ the indicator function δ_F maps points of F into $\{0\}$ and points of $\mathbb{R}^d \backslash F$ (set difference) into $\{+\infty\}$. The support of a probability distribution is the smallest measurable set of mass one for this probability distribution. In the sequel we shall consider functions $F : \mathbb{R}^d \times \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ (finite values) such that $F(\cdot,\cdot,\xi)$ is continuous for each ξ and $\xi \mapsto F(x,u,\xi)$ is measurable for each (x,u), such functions are called Carathéodory. For Carathéodory functions $\xi \mapsto F(x(\xi), u(\xi), \xi)$ is measurable when $(x(\xi), u(\xi))$ depends measurably of ξ. A set valued mapping $S : \mathbb{R}^d \rightrightarrows \mathbb{R}^p$ maps $\mathbb{R}^d$ into subsets of $\mathbb{R}^p$. Such mappings have an inverse $S^{-1} : \mathbb{R}^p \rightrightarrows \mathbb{R}^d$ and $S^{-1}(u) := \{x \mid u \in S(x)\}$, the domain of S is defined by $\mathrm{dom}(S) := \{x \mid S(x) \neq \emptyset\}$. We say that S is measurable if for every open sets $\mathcal{O} \subset \mathbb{R}^p$ the set $S^{-1}(\mathcal{O}) \subset \mathbb{R}^d$ is Borel measurable. The inner semicontinuity of S means that $S(\overline{x}) \subset \liminf_{x\to\overline{x}} S(x)$ with:

$$\liminf_{x\to\overline{x}} S(x) := \left\{ u \mid \forall x^k \to \overline{x},\ \exists N \in \mathcal{N}_\infty,\ \lim_{k\to+\infty, k\in N} u^k = u \text{ with } u^k \in S(x^k) \right\},$$

where $\mathcal{N}_\infty := \{N \subset \mathbb{N} \mid \mathbb{N}\backslash N \text{ finite }\}$, and outer semicontinuous if $\limsup_{x\to\overline{x}} S(x) \subset S(x)$ where:

$$\limsup_{x\to\overline{x}} S(x) = \left\{u \mid \exists x^k \to x, \quad \exists u^k \to u \text{ with } u^k \in S(x^k)\right\}.$$

The mapping S is called continuous if both inner and outer semincontinuity condition hold.

2.1 Optimization model

We consider now $f_t : (x, u, \xi) \in \mathbb{R}^d \times \mathbb{R}^p \times \mathbb{R}^q \to f_t(x, u, \xi) \in \mathbb{R}^d$ a linear mapping with respect to (x, u), $c_t : (x, u, \xi) \in \mathbb{R}^d \times \mathbb{R}^p \times \mathbb{R}^q \to c_t(x, u, \xi) \in \mathbb{R}$ jointly convex function relative to (x, u) and measurable with respect to ξ, and $\Phi_T : \mathbb{R}^d \to \mathbb{R}$ a convex function. It would not bring any additional difficulty to connect the dimensions d, p and q with the time step t, we merely have avoided this possibility in order to alleviate the notations. Let U_t^f be a measurable correspondence $\xi_t \rightrightarrows U_t^f(\xi_t) \subset \mathbb{R}^d \times \mathbb{R}^p$, supposed to be multi-compact-convex-valued, from which we define the correspondence $M_t^f(x, \xi_t) := \left\{u \in \mathbb{R}^p \mid (x, u) \in U_t^f(\xi_t)\right\}$ assumed to be multi-compact-valued, and the correspondence $U_t^1(\xi_t) := \left\{x \in \mathbb{R}^d \mid M_t^f(x, \xi_t) \neq \emptyset\right\}$. At a given x of $U_t^1(\xi_t)$ and for $t = 0, \ldots, T-1$ the optimal value of the following problem:

$$\begin{cases} \min \mathbb{E}\left[\sum_{s=t}^{T-1} c_t(X_s, U_s, \xi_s) + \Phi_T(X_T)\right], \\ \quad X_{s+1} = f_s(X_s, U_s, \xi_s), \quad \forall s = t, \ldots, T-1, \\ \quad X_t = x, \\ \quad U_s \in M_s^f(X_s, \xi_s), \quad \forall s = t, \ldots T-1, \\ \quad U_s \in L^\infty((\mathbb{R}^q)^{s+1}, \mathcal{F}_s; \mathbb{R}^p) \quad \forall s = t, \ldots, T-1, \end{cases} \tag{1}$$

is denoted $V_t^*(x)$, and this value is set to $+\infty$ when the problem is not feasible. A policy is a rule describing which decision we are willing to take in any situation, such rules are also called in the literature *decision rules, feedback control or recourse decision*. By feasible policy is meant a policy π_t which complies with the constraints of the problem. The Markov Decision theory asserts that it is enough to consider only Markov Decision Policies, it means in our framework, policies $\pi_t(X_t, \xi_t)$ depending measurably of the couple (X_t, ξ_t), and we shall prove that the latter policy is in fact continuous. Given $x_0 \in U_0^1(\xi_0)$ the goal is to compute both $V_0^*(x_0)$ and to come up with a feasible policy in order to attain the optimal cost by simulation. Let's us consider the following technical assumptions:

Assumption 1 $\forall t = 0, \ldots, T-1$*:*

1. *The functions $c_t(x, u, \xi_t)$ are strictly convex with respect to u and continuous in ξ_t relatively to the support of ξ_t;*
2. *There exists an integrable function denoted m of ξ_t such that $c_t(x, u, \xi_t) \geq m(\xi_t)$ a.s. and $\phi_T(x) \geq m_T$;*
3. *$u \in M_t^f(x, \xi_t)$ a.s. $\Rightarrow M_{t+1}^f(f_t(x, u, \xi_t), \xi_{t+1}) \neq \emptyset$ a.s.;*
4. *$x \in U_t^1(\xi_t)$ a.s. $\Leftrightarrow x \in \mathrm{dom}(V_t^*)$;*
5. *The correspondence M_t^f is continuous relatively to the support of ξ_t;*
6. *The function $f(x, u, \xi_t)$ is continuous in ξ_t relatively to the support of ξ_t.*

Henceforth we shall assume that all the assumptions above hold true.

Remark 1 *One can make the following observations:*

1. *The statements (1-3-4) of the foregoing assumption are independents of the probability distribution of ξ_t;*
2. *V_t^* is a proper function, for $x \in U_t^1(\xi_t)$ we have:*

$$\mathbb{E}\left[\sum_{s=t}^{T-1} m(\xi_s)\right] + m_T \leq V_t^*(x) < +\infty.$$

3. *Let* $\text{supp}(\xi_t)$ *be the support of the random variable* ξ_t*, then:*

$$\text{dom}(M_t^f) = \text{dom}(V_t^*) \times \text{supp}(\xi_t).$$

Let V be a Borel measurable real extended valued function of $\mathbb{R}^d$, we define the Bellman operator Γ_t by:

$$V \mapsto \Gamma_t(V) \text{ where } x \in \mathbb{R}^d \mapsto \Gamma_t(V)(x) := \mathbb{E}\left[\min_{u \in M_t^f(x,\xi_t)} c_t(x,u,\xi_t) + V(f_t(x,u,\xi_t))\right]. \quad (2)$$

It is well known [2] that functions $(V_t^*)_{t=0}^{T-1}$ satisfy the Bellman equation that reads:

$$V_t^* = \Gamma_t(V_{t+1}^*), \quad V_T^* := \Phi_T. \quad (3)$$

Let's drop the optimization problem (1) and we let's focus on the numerical resolution of the dynamic equation hereafter:

$$\text{find} \quad (V_t)_{t=0}^T \text{ such that } \Gamma_t(V_{t+1}) = V_t \text{ and } V_T = \Phi_T. \quad (4)$$

We have gathered some useful properties of the Bellman operators Γ_t in the following Proposition.

Proposition 4 *Let* V, V' *be continuous real-valued functions from* $\mathbb{R}^d$ *such that* $V' \leq V$ *then* $\forall t = 0, \dots, T-1$ *we have:*

1. *If* V *is a convex function then so is* $\Gamma_t(V)$*;*
2. *Monotonicity:* $\Gamma_t(V) \geq \Gamma_t(V')$*;*
3. *If* $\text{dom}(V_t^*)$ *is compact and* $V' \leq V \leq V_{t+1}^*$ *then:*

$$\|\Gamma_t(V) - \Gamma_t(V')\|_{\text{dom}(V_t^*)} \leq \|V - V'\|_{\text{dom}(V_{t+1}^*)}$$

.

Proof Convexity of $\Gamma_t(V)$ arises from the joint convexity of the application defined by:

$$(x,u) \mapsto c_t(x,u,\xi_t) + V(f_t(x,u,\xi_t)) + \delta_{U_t^f(\xi_t)}(x,u). \quad (5)$$

The monotonicity is straight. To show the last point we derive from the relation $V' \leq V \leq V_{t+1}^*$ both that $\text{dom}(V_{t+1}^*) \subset \text{dom}(V) \subset \text{dom}(V')$ and that $\Gamma_t(V') \leq \Gamma_t(V) \leq \Gamma_t(V_{t+1}^*) = V_t^*$ and therefore $\text{dom}(V_t^*) = \text{dom}(\Gamma_t(V_{t+1}^*)) \subset \text{dom}(\Gamma_t(V)) \subset \text{dom}(\Gamma_t(V'))$. We deduce that if $x \in \text{dom}(V_t^*)$ then $x \in U_t(\xi_t)$ and by item three of assumptions 1 $\forall u \in M_t^f(x,\xi_t)$, $f_t(x,u,\xi_t) \in U_{t+1}^1(\xi_{t+1})$, by fourth item of assumption 1 $f_t(x,u,\xi_t) \in \text{dom}(V_{t+1}^*)$. In other words whatever the policy π_t such that $\pi_t(x,\xi_t) \in M_t^f(x,\xi_t)$ we have $f_t(x,\pi_t(x,\xi_t),\xi_t) \in \text{dom}(V_{t+1}^*)$ and therefore $\forall x \in \text{dom}(V_t^*)$:

$$|\Gamma_t(V)(x) - \Gamma_t(V')(x)| \leq \sup_{u \in M_t^f(x,\xi_t)} |V(f_t(x,u,\xi_t)) - V'(f_t(x,u,\xi_t))| \leq \|V - V'\|_{\text{dom}(V_{t+1}^*)}. \quad (6)$$

The existence of optimal policies, is related to the notion of epi-convergence that we shall address it in section 4.

3 The SDDP algorithm

Basically the SDDP algorithm collects informations along iterations, and constructs for each t both a non-Markovian model of the Bellman function V_t^* and an approximation of the optimal policy $\pi_t^*(X_t,\xi_t)$. There are many approaches in order to implement the SDDP algorithm, our aim is not to provide a convergence proof for all of them but merely to studied in detail one variant. The variant we consider is based on three steps and proceeds as follows:

– At the first step we "Run" the current policy, this step consists in a simulation. For the initialization we have to come up with a feasible policy;

– At the second step, we "Integrate" the Bellman equation (4) only at the visited states of the first step. The integration of the dynamic programming equation gives new subgradients and hence new cuts;
– The last step is performed after the complete integration of the Bellman function of the second step, and consists in "Maximizing" for each time-step the bundle of cuts "After" the complete integration of the dynamic programming equation. This operation updates the Bellman functions and therefore the policies.

Let's agree to call this variant SDDP-RIMA, notice that the Bellman operators Γ_t are very convenient to provide a compact description of how the SDDP-RIMA proceeds and hence avoids the use of cumbersome notations.

Firstly we describe the initialization procedure, let $V_T^0 = \Phi_T$ and $\forall t = 0, \dots, T-1$:

1. Pick $x_t^0 \in \mathrm{dom}(V_t^*)$;
2. Pick $\rho_t^0 \in \Gamma_t(V_{t+1}^0)(x_t^0)$;
3. $V_t^0(\cdot) := \langle \rho_t^0, \cdot - x_t^0 \rangle + \Gamma_t(V_{t+1}^0)(x_t^0)$.

Remark that the above initialization yields necessarily functions V_t^0 such that $V_t^0 \leq V_t^*$. Let $W_t \in L^\infty(\Omega, \mathcal{F}; \mathbb{R}^q)$ be a random variable, we shall describe in the sequel how proceeds the SDDP-RIMA algorithm.

Algorithm 1 *The SDDP-RIMA loops:*

Initialization: according to items (1-2-3) above;
Begin of the loop: we have π_t^k *and* $V_t^k \forall t = 0, \dots, T-1$ *and* $V_T^k := \Phi_T$*;*
Drawing: Draw $W^{k+1} := (W_0^{k+1}, \dots, W_{T-1}^{k+1})$ *independently from the past drawings;*
Simulation pass: $X_{t+1}^{k+1} := f_t(X_t^{k+1}, \pi_t^k(X_t^{k+1}, W_t^{k+1}), W_t^{k+1}), \quad X_0^{k+1} := x_0$*;*
Optimization pass: Pick $\rho_t^{k+1} \in \partial \Gamma_t(V_{t+1}^k)(X_t^{k+1}), \quad \forall t = 0, \dots, T-1$*;*
Update Bellman functions: $V_t^{k+1}(x) := \max\left(V_t^k(x), \langle \rho_t^{k+1}, x - X_t^{k+1} \rangle + \Gamma_t(V_{t+1}^k)(X_t^{k+1})\right)$*;*
Update policies: π_t^{k+1}*;*
Loop: $k \leftarrow k+1$ *and loop.*

We stress on the simulation procedure which is performed with a different probability distribution of the one used to evaluate the Bellman operator.

Assumption 2 *For each* t *the random variables* $(W_0, \dots, W_t)$ *and* $(\xi_0, \dots, \xi_t)$ *have the same Borels sets of null mass.*

The foregoing assumption ensures that $(W_0, \dots, W_{T-1})$ and $(\xi_0, \dots, \xi_{T-1})$ have the same support denoted $\mathrm{supp}(W)$, and for any measurable function $\varphi : \mathrm{supp}(W) \to \mathbb{R}$, if $\varphi(\xi_0, \dots, \xi_{T-1}) = 0$ holds true a.s., then $\varphi(W_0, \dots, W_{T-1}) = 0$ a.s. and reversely. Similarly if $M_t^f(x, \xi_t) \neq \emptyset$ *a.s.*, we have also $M_t^f(x, W_t) \neq \emptyset$ *a.s.* We shall assume that $\Gamma_t(V_{t+1}^k)$ has bounded subgradients and hence ρ_t^{k+1} exists at any step of algorithm 1. Inasmuch we shall study the convergence of a stochastic algorithm it is convenient to introduce the samples space $(\Lambda, \mathcal{B}, \mathbb{P}) := (\Xi, \mathcal{B}_\Xi, \mu_W)^{\otimes \mathbb{N}}$, where $\Xi := (\mathbb{R}^q)^T$, $\mathcal{B}_\Xi$ is the Borel σ-field of open subsets of Ξ and $\mu_W := \mu_{W_0} \otimes \ldots \otimes \mu_{W_{T-1}}$. We denote $\mathcal{F}^0 := \sigma(\emptyset, \Xi)$ and $\forall k \in \mathbb{N}^* := \mathbb{N} \setminus \{0\}, \mathcal{F}^k$ the Borel σ-field of open subsets of Ξ^k Under these notations V_t^k is $\mathcal{F}^k$-measurable, and $\mathbb{F} := (\mathcal{F}^k)_{k \in \mathbb{N}}$ is a filtration of $\mathcal{B}$ it means that $\mathcal{F}^k \subset \mathcal{F}^{k+1} \subset \mathcal{B}$. It is important to notice that if π_t^k is $\mathcal{F}^0$-measurable, its means $\pi_t^k = \pi_t$ is a deterministic policy. The drawings $(W^k)_{k \in \mathbb{N}^*}$ in algorithm 1 are independents and identically distributed (iid). We shall prove that algorithm 1 converges with respect to the norms induced by the policy $\pi := (\pi_0, \dots, \pi_{T-1})$ and denoted $\|\cdot\|_{\pi_t}$ which are defined by $\|V\|_{\pi_t} := \mathbb{E}\left[|V(X_t^{\pi_t})|\right]$ where $X_t^{\pi_t}$ is the random variable defined dynamically with the equation $X_{t+1}^{\pi_{t+1}} = f_t(X_t^{\pi_t}, \pi_t(X_t^{\pi_t}, W_t), W_t)$ and $X_0^{\pi_0} = x_0$. The distribution of the random variable $X_t^{\pi_t}$ will be denoted $\mu_{X_t^{\pi_t}}$ and for any measurable function $V : \mathbb{R}^d \to \mathbb{R}$ we define $\|V\|_{\pi_t} := \int |V| \, \mu_{X_t^{\pi_t}}$. We can remark that if π_t^k is a $\mathcal{F}^k$-measurable policy then, $\mathbb{E}\left(\left|V_t^k(X_t^{k+1})\right| \mid \mathcal{F}^k\right) = \left\|V_t^k\right\|_{\pi_t^k}$. If we consider X_t^{k+1} as a measurable function of $\pi_0^k, \dots, \pi_{t-1}^k, W_1^{k+1}, \dots, W_{t-1}^{k+1}$ that is to say $X_t^{k+1} = F_t(\pi_0^k, \dots, \pi_{t-1}^k, W_1^{k+1}, \dots, W_{t-1}^{k+1})$ $\mathbb{P}$-almost surely, where F_t is measurable with respect to its arguments. Then for each t the succession of equalities below hold true $\mathbb{P}$-almost surely:

$$\begin{aligned}\mathbb{E}\left(\left|V_t^k(X_t^{k+1})\right| \mid \mathcal{F}^k\right) &= \mathbb{E}\left(\left|V_t^k(F_t(\pi_0^k,\dots,\pi_{t-1}^k, W_1^{k+1}\dots, W_{t-1}^{k+1}))\right| \mid \mathcal{F}^k\right),\\ &= \mathbb{E}\left[\left|V_t(F_t(\pi_0,\dots,\pi_{t-1}, W_1\dots, W_{t-1}))\right|\right]_{|V_t=V_t^k,\pi=\pi^k},\\ &= \|V_t\|_{\pi\,|V_t=V_t^k,\pi=\pi^k},\\ &= \left\|V_t^k\right\|_{\pi_t^k}.\end{aligned}$$

This equation means that we perform first the conditional expectation with respect to $\mathcal{F}^k$ without averaging V_t^k and π^k, which turn out to be an expectation. Thereafter we substitute V_t (resp. π) with V_t^k (resp. π^k) which is $\mathcal{F}^k$ measurable.

4 Convergence analysis

In this section we study the convergence of the SDDP-RIMA algorithm, we proceed in our analysis in two parts. The first part is concerned with the algebraic properties of the functions V_t^k. We prove in Proposition 5 hereafter that the latter sequence of functions is non-decreasing (straight) and converges uniformly. The Proposition 6 states a fundamental relation between different errors terms involved the algorithm, and which is still valid along iterations. The purpose is to match these two results and to prove that they are agreeing. For that purpose we shall use the concept of epi-convergence in a second part of our analysis. We shall prove in Proposition 9 the policy which governs the evolution equation of the state variable, converges pointwise to a steady policy and hence by Proposition 10 that the state variable converges along iterations to a steady state. The Corollary 1 makes the link between the mentioned results and asserts that the uniform limit of the sequence V_t^k is a solution of the dynamic programming equation(4). Before stating these results we need to recall the definition of a locally simplicial set. Indeed that latter notion plays a role in the continuity of the Belman functions on their effective domains.

Definition 2 *A subset S is said to be locally simplicial if for each $x \in S$ there exists a finite collection of simplices $T_1,\dots,T_m$ contained in S such that:*

$$U \cap (T_1 \cup \ldots \cup T_m) = U \cap S. \tag{7}$$

for some neighborhood U of x.

Assumption 3 *For each t the set $\mathrm{dom}(V_t^*)$ is:*

1. *compact;*
2. *locally simplicial.*

Remark 2 *All polytopes and polyhedral convex sets are locally simplicials, the relatively open convex sets are so too.*

Proposition 5 *The following properties holds true $\mathbb{P}$-almost surely $\forall t = 0,\dots,T-1, \quad \forall k \in \mathbb{N}$:*

- *V_t^k is a real valued proper convex function;*
- *The sequence V_t^k converges pointwise to $V_t^\sharp$ on $\mathrm{dom}(V_t^*)$ the latter function being convex lsc and such that $\mathrm{dom}(V_t^*) = \mathrm{dom}(V_t^\sharp)$;*
- *$V_t^k \le V_t^{k+1} \le V_t^\sharp$;*
- *$V_t^{k+1}(X_t^{k+1}) = \Gamma_t(V_{t+1}^k)(X_t^{k+1})$.*

If in addition Assumption 3 holds true, then sequence V_t^k converges uniformly to $V_t^\sharp$ on $\mathrm{dom}(V_t^\sharp)$.

Proof 1. The initialization provides for each t a function V_t^0 which is affine, therefore it is proper and convex. If we assume V_t^k is real valued proper and convex, then so is V_t^{k+1} as the maximum of two real valued proper and convex functions. Therefore $\forall k \in \mathbb{N}, \quad \forall t = 0,\dots,T, \quad V_t^k$ are convex continuous functions.

2. Let's consider the statement $P(k) \Leftrightarrow \forall t\ V_t^k \leq V_t^*$, we then reason by induction in k. We have clearly $P(0)$ holds true, then we assume $P(k)$ holds true so what happen with $P(k+1)$. Since V_{t+1}^k is convex then $\mathit{\Gamma}_t(V_{t+1}^k)$ is convex and $\mathit{\Gamma}_t(V_{t+1}^k)(x) \geq \langle \rho_t^{k+1}, x - X_t^{k+1}\rangle + \mathit{\Gamma}_t(V_{t+1}^k)(X_t^{k+1})$ and by the induction hypothesis we have $V_{t+1}^k \leq V_{t+1}^*$. Furthermore by virtue of monotonicity of $\mathit{\Gamma}_t$ we have $\mathit{\Gamma}_t(V_{t+1}^k) \leq \mathit{\Gamma}_t(V_{t+1}^*) = V_t^*$, therefore $V_t^{k+1} \leq V_t^*$ and $P(k+1)$ holds true.
   The sequence $(V_t^k)_{k\in\mathbb{N}}$ is nondecreasing and upper bounded on $\mathrm{dom}(V_t^*)$. Let's define $V_t^\sharp(x)$ equal to $\sup_{k\in\mathbb{N}} V_t^k(x)$ if $x \in \mathrm{dom}(V_t^*)$ and $+\infty$ otherwise. Hence $\mathrm{dom}(V_t^\sharp) = \mathrm{dom}(V_t^*)$ and $V_t^\sharp$ is a proper convex [5, Proposition 2.2] function, which is furthermore bounded above by V_t^* on its effective domain.
3. By construction V_t^k is bounded above by $V_t^\sharp$.
4. By virtue of the recursive definition of V_t^k we have:

$$V_t^{k+1}(x) = \max\left\{\max_{\ell\leq k}\left\{\langle \rho_t^{\ell+1}, x - X_t^{\ell+1}\rangle + \mathit{\Gamma}_t(V_{t+1}^\ell)(X_t^{\ell+1})\right\}, V_t^0(x)\right\},$$

   therefore for $x = X_t^{k+1}$ we have $V_t^{k+1}(X_t^{k+1}) \geq \mathit{\Gamma}_t(V_{t+1}^k)(X_t^{k+1})$. Since every V_{t+1}^ℓ are convex, then so $\mathit{\Gamma}_t(V_{t+1}^\ell)$ by Proposition 4 therefore, $\mathit{\Gamma}_t(V_{t+1}^\ell)(x) \geq \langle \rho_t^{\ell+1}, x - X_t^{\ell+1}\rangle + \mathit{\Gamma}_t(V_{t+1}^\ell)(X_t^{\ell+1})$. Moreover $\forall \ell \leq k,\ V_t^\ell \leq V_t^k$ then $\mathit{\Gamma}_t(V_{t+1}^\ell) \leq \mathit{\Gamma}_t(V_{t+1}^k)$ and $\mathit{\Gamma}_t(V_{t+1}^k)(X_t^{k+1}) \geq \mathit{\Gamma}_t(V_{t+1}^\ell)(X_t^{k+1}) \geq \langle \rho_t^{\ell+1}, X_t^{k+1} - X_t^{\ell+1}\rangle + \mathit{\Gamma}_t(V_{t+1}^\ell)(X_t^{\ell+1})$. We have $V_t^0 \leq \mathit{\Gamma}_t(V_{t+1}^0)$, moreover $V_{t+1}^0 \leq V_{t+1}^k$, then by the monotonicity property of $\mathit{\Gamma}_t$, we also have $V_t^0 \leq \mathit{\Gamma}_t(V_{t+1}^k)$. Therefore $V_t^{k+1}(X_t^{k+1}) \leq \mathit{\Gamma}_t(V_{t+1}^k)(X_t^{k+1})$.

Under Assumption 3 the set $\mathrm{dom}(V_t^*)$ is locally simplicial, then so is $\mathrm{dom}(V_t^\sharp)$, but since $V_t^\sharp$ is also convex lsc and proper, then by [12, Theorem 10.2] $V_t^\sharp$ is continuous on its effective domain. By Proposition 1 the convergence of the sequence V_t^k to $V_t^\sharp$ is uniform on the effective domain of $V_t^\sharp$.

Remark 3 *In the foregoing proposition we have assumed the simplicial localness of* $\mathrm{dom}(V_t^*)$ *in order to derive the continuity of* $V_t^\sharp$ *on its effective domain. The continuity of a propositioner convex lsc function on its effective domain has been studied in [1], where authors provide characterizations.*

Proposition 6 *The equality below hold true* $\forall k$*:*

$$\left\|V_t^\sharp - \mathit{\Gamma}_t(V_{t+1}^k)\right\|_{\pi_t^k} = \left\|V_t^\sharp - V_t^k\right\|_{\pi_t^k} - \left\|V_t^k - \mathit{\Gamma}_t(V_{t+1}^k)\right\|_{\pi_t^k} \quad \mathbb{P}\text{-almost surely.} \tag{8}$$

Proof Using the relation $\min(a,b) = \frac{1}{2}(a+b) - \frac{1}{2}|a-b|$ yields for each t the equation:

$$\begin{aligned}\forall x \in \mathrm{dom}(V_t^\sharp),\quad (V_t^\sharp - V_t^{k+1})(x) =& \frac{1}{2}\Big(V_t^\sharp(x) - V_t^k(x) + V_t^\sharp(x) - \langle \rho_t^{k+1}, x - X_t^{k+1}\rangle - \mathit{\Gamma}_t(V_{t+1}^k)(X_t^{k+1})\Big)\ldots\\ &- \frac{1}{2}\left|V_t^k(x) - \langle \rho_t^{k+1}, x - X_t^{k+1}\rangle - \mathit{\Gamma}_t(V_{t+1}^k)(X_t^{k+1})\right|.\end{aligned} \tag{9}$$

For $x = X_t^{k+1}$, we have $X_t^{k+1} \in \mathrm{dom}(V_t^\sharp)$ then:

$$\begin{aligned}V_t^\sharp(X_t^{k+1}) - V_t^{k+1}(X_t^{k+1}) =& \frac{1}{2}\Big(V_t^\sharp(X_t^{k+1}) - V_t^k(X_t^{k+1}) + V_t^\sharp(X_t^{k+1}) - \mathit{\Gamma}_t(V_{t+1}^k)(X_t^{k+1})\Big)\ldots\\ &- \frac{1}{2}\left|V_t^k(X_t^{k+1}) - \mathit{\Gamma}_t(V_{t+1}^k)(X_t^{k+1})\right|.\end{aligned} \tag{10}$$

By Proposition 5 we have the relation:

$$\begin{aligned}V_t^\sharp(X_t^{k+1}) - \mathit{\Gamma}_t(V_{t+1}^k)(X_t^{k+1}) =& \frac{1}{2}\Big(V_t^\sharp(X^{k+1}) - V_t^k(X^{k+1}) + V_t^\sharp(X^{k+1}) - \mathit{\Gamma}_t(V_{t+1}^k)(X_t^{k+1})\Big)\ldots\\ &- \frac{1}{2}\left|V_t^k(X^{k+1}) - \mathit{\Gamma}_t(V_{t+1}^k)(X_t^{k+1})\right|.\end{aligned} \tag{11}$$

Which in turn implies that:

$$V_t^\sharp(X_t^{k+1}) - \mathit{\Gamma}_t(V_{t+1}^k)(X_t^{k+1}) = \Big(V_t^\sharp(X_t^{k+1}) - V_t^k(X_t^{k+1})\Big) - \left|V_t^k(X_t^{k+1}) - \mathit{\Gamma}_t(V_{t+1}^k)(X_t^{k+1})\right|. \tag{12}$$

and:

$$(V_t^\sharp(X_t^{k+1}) - \Gamma_t(V_{t+1}^k)(X_t^{k+1})) = \Big(V_t^\sharp(X_t^{k+1}) - V_t^k(X_t^{k+1})\Big) \\ - \left|V_t^k(X_t^{k+1}) - \Gamma_t(V_{t+1}^k)(X_t^{k+1})\right| . \qquad (13)$$

And we take the conditional expectation:

$$\mathbb{E}\left(V_t^\sharp(X_t^{k+1}) - \Gamma_t(V_{t+1}^k)(X_t^{k+1}) \mid \mathcal{F}^k\right) = \mathbb{E}\left(V_t^\sharp(X_t^{k+1}) - V_t^k(X_t^{k+1}) \mid \mathcal{F}^k\right) \\ - \mathbb{E}\left(\left|V_t^k(X_t^{k+1}) - \Gamma_t(V_{t+1}^k)(X_t^{k+1})\right| \mid \mathcal{F}^k\right) . \qquad (14)$$

From $V_t^\sharp - V_t^k \geq 0$ we deduce that $\Gamma_t(V_{t+1}^k)(X_t^{k+1}) \leq V_t^{k+1}(X_t^{k+1}) \leq V_t^\sharp(X_t^{k+1})$, and the observation that $\Gamma_t(V_t^k)$, V_t^k and π_t^k are $\mathcal{F}^k$ measurable functions, yields the desired result:

$$\left\|V_t^\sharp - \Gamma_t(V_{t+1}^k)\right\|_{\pi_t^k} = \left\|V_t^\sharp - V_t^k\right\|_{\pi_t^k} - \left\|V_t^k - \Gamma_t(V_{t+1}^k)\right\|_{\pi_t^k} . \qquad (15)$$

Remark 4 *We can observe that if we substitute Γ_t with an unbiased estimator Γ_t^k, the foregoing result would failed due to the non linearity of the absolute value function.*

4.1 Epi-convergence

We start this section recalling the epi-convergence concept and some of its important properties. For a comprehensive treatment of epi-convergence and related issues together with the proofs of the following Proposition, see [13]. Let $(h^k)_{k\in\mathbb{N}}$ be a sequence of real valued functions, the epigraph of h^k is the set denoted $\text{epi}(h^k)$ which reads:

$$\text{epi}(h^k) := \left\{(\alpha, u) \mid \alpha \geq h^k(u)\right\} . \qquad (16)$$

We say that (h^k) epi converges to h and we denote $h^k \xrightarrow{e} h$ if the lower epi-limit $(e - \liminf h^k)$ is equal to the upper epi-limit $(e - \limsup h^k)$, where lower epi-limit is defined by:

$$\text{epi}(e - \liminf_k h^k) := \limsup_k(\text{epi}(h^k))$$

and upper epi-limit is defined by:

$$\text{epi}(e - \limsup_k h^k) := \liminf_k(\text{epi}(h^k)).$$

This convergence refers to the convergence of epigraphs (16) for the so called integrated set distance [13, Theorem 7.58]. The sequence of functions $(h^k)_{k\in\mathbb{N}}$ is said to be eventually level-bounded if if for each $\alpha \in \mathbb{R}$ the sequence of sets $\text{lev}_\alpha h^k$ is eventually bounded. It is well known that the sequence $(h^k)_{k\in\mathbb{N}}$ is eventually level-bounded if there is a level-bounded function ℓ such that eventually $h^k \geq \ell$ or if the sequence of set $\text{dom}(h^k)$ is eventually bounded. Under eventually level boundedness concept the epi-convergence entails convergence of minimizers.

Proposition 7 *[13, Theorem. 7.33] Suppose that h^k is eventually level-bounded, and $h^k \xrightarrow{e} h$ with h^k and h lsc and proper. Then:*

$$\inf h^k \to \inf h \textit{(finite)} \qquad (17)$$

while for k in some index set $N \in \mathcal{N}_\infty$ the sets $\arg\min h^k$ are nonempty and form a bounded sequence with

$$\limsup_k(\arg\min h^k) \subset \arg\min h.$$

Indeed, for any choice of $\varepsilon^k \searrow 0$ and $x^k \in \arg\min(h^k + \varepsilon^k)$, the sequence $(x^k)_{k\in\mathbb{N}}$ is bounded and such that all its cluster points belong to $\arg\min h$. If $\arg\min h$ consists of a unique point $\overline{x}$, then for any choice of $\varepsilon^k \to 0$ and $x^k \in \arg\min\left\{h^k + \varepsilon^k\right\}$ one must actually have $x^k \to x$.

Proposition 8 *Let V be a lsc-convex-real-valued function such that $V \leq V^*_{t+1}$, then there exists a continuous function π_t of (x, ξ_t) such that:*

$$\forall x \in \mathrm{dom}(\Gamma_t(V)), \quad \Gamma_t(V)(x) = \mathbb{E}\left[c_t(x, \pi_t(x, \xi_t), \xi_t) + V(f_t(x, \pi_t(x, \xi_t), \xi_t))\right], \tag{18}$$

$$\textit{and} \quad \pi_t(x, \xi_t) \in M^f_t(x, \xi_t). \tag{19}$$

Proof Let $S_t : \mathrm{dom}(M^f_t) \rightrightarrows \mathbb{R}^p$ be the correspondence defined by:

$$S_t(x, w) := \arg\min_u g_t(x, u, w). \tag{20}$$

where $g_t(x, u, w_t) := c_t(x, u, w_t) + V(f_t(x, u, w_t)) + \delta_{M^f_t(x, w_t)}(u)$. Consider $(x^n, w^n)_n \subset \mathrm{dom}(S_t)$ a convergent sequence such that $\lim x^n = x$ and $\lim w^n = w$, firstly we want to prove that $\lim S_t(x^n, w^n) = S_t(x, w)$. Let G^n_t be the function defined by $u \in \mathbb{R}^p \mapsto c_t(x^n, u, w^n) + V(f_t(x^n, u, w^n))$, then $(G^n_t)_n$ is a sequence of finite convex functions which converges pointwise to $G_t(u) := c_t(x, u, w) + V(f_t(x, u, w))$. By virtue of [13, Theorem 7.17] we have both that $G^n_t \xrightarrow{e} G_t$ and converges uniformly on every compact subsets of $\mathbb{R}^p$. Since G_t is continuous then [13, Theorem 7.14] yields that G^n_t converges continuously to G_t. Since M^f_t is a continuous correspondence then by [13, Proposition 7.4 (f)] we have $\delta_{M^f_t(x^n, w^n)} \xrightarrow{e} \delta_{M^f_t(x, w_t)}$, and by [13, Theorem 7.46] we deduce that $g^n_t + \delta_{M^f_t(x^n, w^n)} \xrightarrow{e} G_t + \delta_{M^f_t(x, w)}$. Then epi-convergence entails the convergence of the associated sequence of minimizers, we deduce using the uniqueness of these minimizers that $u^n \in S_t(x^n, w^n)$ converges to $u \in S_t(x, w)$. The latter result means that the correspondence S_t is inner semicontinuous. The Michael Theorem (see [13, Theorem 5.58]), asserts that whenever S_t is inner semicontinuous on $\mathrm{dom}(S_t)$ as well as closed-convex-valued and that $\mathrm{dom}(S_t)$ is closed, which turns out to be the case, then S_t possesses a continuous selection. This implies that there exists a continuous function π_t such that $\pi_t(x, w_t)$ minimizes $g_t(x, \cdot, w_t)$ whenever $\arg\min g_t(x, \cdot, w_t) \neq \emptyset$.

The Proposition 5 asserts that the sequence V^k_t converges pointwise to $V^\sharp_t$ on $\mathrm{dom}(V^*_t)$. Nonetheless, despite this result, we know little about the policies behavior so far. Let g^k_t be defined by:

$$g^k_t(x, u, w) := c_t(x, u, w) + V^k_{t+1}(f_t(x, u, w)) + \delta_{M^f_t(x, w)}(u). \tag{21}$$

We have clearly the following relation:

$$\Gamma_t(V^k_{t+1})(x) = \mathbb{E}\left[\min_u g^k_t(x, u, \xi_t)\right]. \tag{22}$$

By virtue of Proposition 8 we know that there exists a policy π^k_t which depends continuously of (x, w) and such that $\min_u g^k_t(x, u, w_t) = g^k_t(x, \pi^k_t(x, w_t), w_t)$. We shall show the epi-convergence of the sequence $g^k_t(x, \cdot, w)$ and the pointwise convergence of the associated sequence of minimizers π^k_t.

Remark 5 *The sequence $(g^k_t(x, \cdot, w))_{k \in \mathbb{N}}$ is a sequence of random functionals and we want to studied its convergence. This issue has been studied by Robinson (see [11, Proposition 2.4]) which stresses the necessity for the effective domain of $g^k_t(x, \cdot, w)$ denoted $\mathrm{dom}(g^k_t)(x, w)$ to be almost surely constant. Indeed the effective domain is both a function the sample of $W = (W^0, \ldots, W^k, \ldots) \in \Lambda$ and of the iteration index k. This assumption clearly holds true in our problem formulation since $\mathrm{dom}(g^k_t)(x, w)$ is merely $M^f_t(x, w)$. This requirement cannot be dropped whether we want to ensure the almost sure convergence of the sequence of minimizers.*

Proposition 9 *The sequence $g^k_t(x, \cdot, w)$ epi-converges to $g^\sharp_t(x, \cdot, w)$ the latter function being lsc and proper. The function $g^k_t(x, \cdot, w)$ (resp. $g^\sharp_t(x, \cdot, w)$) has an unique minimizer, let's denote it $\pi^k_t(x, w)$ (resp. $\pi^\sharp_t(x, \cdot, w)$), then $\pi^k_t(x, w)$ converges pointwise to $\pi^\sharp_t(x, w)$.*

Proof The sequence $g^k_t(x, \cdot, w)$ is nondecreasing then [13, Proposition 7.4, (d)] asserts that $g^k_t(x, \cdot, w)$ epi-converges to $\sup_{k \in \mathbb{N}} g^k_t(x, \cdot, w)$. By Proposition 5 we have $V^\sharp_t(x) = \sup_{k \in \mathbb{N}} V^k_t(x)$ if $x \in \mathrm{dom}(V^*_t)$, then:

$$\forall u \in M^f_t(x, w), \quad \sup_{k \in \mathbb{N}} g^k_t(x, u, w) = c_t(x, u, w) + \sup_{k \in \mathbb{N}} V^k_t(f_t(x, u, w)) = c_t(x, u, w) + V^\sharp_t(f_t(x, u, w)) \tag{23}$$

Call $g_t^\sharp(x,u,w) := c_t(x,u,w) + V_t^\sharp(f_t(x,u,w)) + \delta_{M_t^f(x,w)}(u)$ then $g_t^k(x,\cdot,w) \xrightarrow{e} g_t^\sharp(x,\cdot,w)$. Due to the strict convexity of $c_t(x,\cdot,w)$ we have clearly that $g_t^k(x,\cdot,w)$ (resp. $g_t^\sharp(x,\cdot,w)$) possesses a unique minimizer $\pi_t^k(x,w)$ (resp. $\pi_t^\sharp(x,w)$) for $(x,w) \in \mathrm{dom}(M_t^f)$. Since Epi-convergence entails also the convergence of minimizers then $\pi_t^k(x,w)$ converges to $\pi_t^\sharp(x,w)$ the unique minimizer of $g_t^\sharp(x,\cdot,w)$ for $(x,w) \in \mathrm{dom}(M_t^f)$.

The main drawback with Proposition 6 arises from the fact the error term $\left\|V_t^k - V_t^\sharp\right\|_{\pi_t^k}$ is evaluated with respect to a norm depending of the current policy π_t^k, whereas we want to come up with a common norm reference along iterations in order to evaluate the convergence of the algorithm. Let $\mu_t^k := \mu_{X_t^{\pi_t^k}}$ be the probability distribution of $X_t^{\pi_t^k}$, the state variable governed by the policy π_t^k, we shall show it converges in distribution to the probability distribution $\mu_t^\sharp := \mu_{X_t^{\pi_t^\sharp}}$ the latter distribution being the optimal distribution of the state variable. We have already proved the pointwise convergence of the sequence π_t^k, we are ready to state the convergence of the probability distributions of the random variables $X_t^{\pi_t^k}$.

Proposition 10 *The sequence $X_t^{\pi_t^k}(X_0^{\pi_0^k} := x_0)$ converges in distribution to $X_t^{\pi_t^\sharp}(X_0^{\pi_0^\sharp} = 0)$.*

Proof We reason by induction, let " $\mathrm{H(t)} \Leftrightarrow \left(\mu_{X_t^{\pi_t^k}} \Rightarrow \mu_{X_t^{\pi_t^\sharp}}\right)$" be the induction hypothesis. Clearly the statement holds true for $t = 0$ since $\mu_{X_0^{\pi_0^k}}$ is the Dirac measure in x_0 for each k. We assume now H(t) hods true, let φ be a bounded continuous function (see [4, Theorem 25.8]) then, the continuity of π_t^k given by Proposition 8 yields:

$$(x,w) \in \mathrm{dom}(M_t^f) \mapsto \varphi(f_t(x,\pi_t^k(x,w),w)) \text{ and } (x,w) \in \mathrm{dom}(M_t^f) \mapsto \varphi(f_t(x,\pi_t^\sharp(x,w),w)) \tag{24}$$

are continuous and converge pointwise:

$$\lim_{k\to\infty} \varphi(f_t(x,\pi_t^k(x,w),w)) = \varphi(f_t(x,\pi_t^\sharp(x,w),w)) \tag{25}$$

We denote $\mathrm{supp}(W_t)$ the support of W_t which turns out to be bounded since $W_t \in L^\infty(\Omega,\mathcal{F};\mathbb{R}^q)$ then, $K := \overline{\mathrm{dom}(M_t^f)}$ is a compact set and:

$$\left\|\varphi(f_t(x,\pi_t^k(x,w),w)) - \varphi(f_t(x,\pi_t^\sharp(x,w),w))\right\|_K \to 0 \tag{26}$$

Since $\mu_t^k \Rightarrow \mu_t^\sharp$ then:

$$\begin{aligned}
\lim_{k\to\infty} \int \varphi(x)\mu_{X_{t+1}^{\pi_{t+1}^k}}(dx) &= \lim_{k\to\infty} \int \varphi(f(x,\pi_t^k(x,w),w))\mu_{X_t^{\pi_t^k}} \otimes \mu_{W_t}(dx,dw),\\
&= \lim_{k\to\infty} \int \Big(\varphi(f(x,\pi_t^k(x,w),w)) - \varphi(f_t(x,\pi_t^\sharp(x,w),w))\Big)\,\mu_{X_t^{\pi_t^k}} \otimes \mu_{W_t}(dx,dw)+\\
&\quad \lim_{k\to\infty} \int \varphi(f_t(x,\pi_t^\sharp(x,w),w))\mu_{X_t^{\pi_t^k}} \otimes \mu_{W_t}(dx,dw),\\
&= \int \varphi(x)\mu_{X_{t+1}^{\pi_{t+1}^\sharp}}(dx).
\end{aligned}$$

Remark 6 *The result of the latter proposition can be read also:*

$$\lim \mathbb{E}\left(\varphi(X_{t+1}^{\pi_{t+1}^k}) \mid \mathcal{F}^k\right) = \mathbb{E}\left[\varphi(X_{t+1}^{\pi_{t+1}^\sharp})\right]. \tag{27}$$

because of the following equality holds:

$$\begin{aligned}
\mathbb{E}\left(\varphi(X_{t+1}^{\pi_{t+1}^k}) \mid \mathcal{F}^k\right) =& \mathbb{E}\left(\varphi(f_t(X_t^{\pi_t^k},\pi_t^k(X_t^{\pi_t^k},W_t^{k+1}),W_t^{k+1})) \mid \mathcal{F}^k\right)\\
&= \int \varphi(f_t(x_t,\pi_t^k(x_t,w_t),w_t))\mu_{X_t^{\pi_t^k}} \otimes \mu_{W_t}(dx_t,dw_t)
\end{aligned}$$

Corollary 1 *Assume Assumption 3 holds true, then for each t we have $V_t^\sharp = \Gamma_t(V_{t+1}^\sharp)$ $\mu_t^\sharp$ a.s. on* $\mathrm{dom}(V_t^\sharp)$.

Proof Since $X_t^{\pi_t^k} \in \mathrm{dom}(V_t^\sharp)$ *a.s.* we have:

$$\left\|(V_t^\sharp - V_t^k)\right\|_{\pi_t^k} = \int_{\mathrm{dom}(V_t^\sharp)} (V_t^\sharp - V_t^k)\mu_t^k \tag{28}$$

$$\leq \left\|V_t^\sharp - V_t^k\right\|_{\mathrm{dom}(V_t^\sharp)} \tag{29}$$

By the Proposition 5 we have $\lim_{k\to\infty} \left\|V_t^\sharp - V_t^k\right\|_{\mathrm{dom}(V_t^\sharp)} = 0$ hence $\lim_{k\to\infty} \left\|(V_t^\sharp - V_t^k)\right\|_{\pi_t^k}$=0 and by Proposition 6:

$$\lim_{k\to\infty} \left\|(V_t^\sharp - \Gamma_t(V_{t+1}^k))\right\|_{\pi_t^k} = \lim_{k\to\infty} \left\|(V_t^k - \Gamma_t(V_{t+1}^k))\right\|_{\pi_t^k} = 0. \tag{30}$$

Let $h_t^k(x) := \left|V_t^\sharp(x) - \Gamma_t(V_{t+1}^k)(x)\right|$ and $h_t(x) := \left|V_t^\sharp(x) - \Gamma_T(V_{t+1}^\sharp)(x)\right|$ then:

$$\forall x \in \mathrm{dom}(V_t^\sharp), \quad \left|(h_t^k - h_t)(x)\right| = \left|\left|V_t^\sharp(x) - \Gamma_t(V_{t+1}^k)(x)\right| - \left|V_t^\sharp(x) - \Gamma_t(V_{t+1}^\sharp)(x)\right|\right|, \tag{31}$$

$$\leq \left|\Gamma_t(V_{t+1}^k)(x) - \Gamma_t(V_{t+1}^\sharp)(x)\right|. \tag{32}$$

Applying Proposition 4 with $V' = V_{t+1}^k, V = V_{t+1}^\sharp$ then:

$$\forall x \in \mathrm{dom}(V_t^\sharp), \quad \left|\Gamma_t(V_{t+1}^k)(x) - \Gamma_t(V_{t+1}^\sharp)(x)\right| \leq \left\|\Gamma_t(V_{t+1}^k) - \Gamma_t(V_{t+1}^\sharp)\right\|_{\mathrm{dom}(V_t^\sharp)} \tag{33}$$

$$\leq \left\|V_{t+1}^k - V_{t+1}^\sharp\right\|_{\mathrm{dom}(V_{t+1}^\sharp)}. \tag{34}$$

Therefore:

$$\left\|h_t^k - h_t\right\|_{\mathrm{dom}(V_t^\sharp)} \leq \left\|V_{t+1}^k - V_{t+1}^\sharp\right\|_{\mathrm{dom}(V_{t+1}^\sharp)}. \tag{35}$$

Moreover:

$$\int_{\mathrm{dom}(V_t^\sharp)} h_t^k\mu_t^k - \int_{\mathrm{dom}(V_t^\sharp)} h_t\mu_t^\sharp = \int_{\mathrm{dom}(V_t^\sharp)} (h_t^k - h_t)\mu_t^k + \int_{\mathrm{dom}(V_t^\sharp)} h_t(\mu_t^k - \mu_t^\sharp). \tag{36}$$

By Jensen inequality:

$$\left|\int_{\mathrm{dom}(V_t^\sharp)} h_t^k\mu_t^k - \int_{\mathrm{dom}(V_t^\sharp)} h_t\mu_t\right| \leq \left\|h_t^k - h_t\right\|_{\mathrm{dom}(V_t^\sharp)} + \int_{\mathrm{dom}(V_t^\sharp)} h_t(\mu_t^k - \mu_t^\sharp). \tag{37}$$

Since $\mu_t^k \Rightarrow \mu_t^\sharp$ and V_t^k converges uniformly to $V_t^\sharp$ on $\mathrm{dom}(V_t^\sharp)$ we deduce that:

$$\lim_{k\to\infty} \left\|V_t^\sharp - \Gamma_t(V_{t+1}^k)\right\|_{\pi_t^k} = \left\|V_t^\sharp - \Gamma_t(V_{t+1}^\sharp)\right\|_{\pi_t^\sharp} = 0. \tag{38}$$

Then $V_t^\sharp \mathbb{1}_{\mathrm{dom}(V_t^\sharp)} = \Gamma_t(V_{t+1}^\sharp)\mathbb{1}_{\mathrm{dom}(V_t^\sharp)} - \mu_t^\sharp$ *a.s.*.

We are ready to state our main result.

Theorem 1 *Assume Assumption 1 holds true, then the sequence $(V_t^k)_{k\in\mathbb{N}}$ converges pointwise to $V_t^\sharp$ on $\mathrm{dom}(V_t^*)$. Furthermore for each t the sequence $g_t^k(x,\cdot,w)$ epi-converges to $g_t^\sharp(x,\cdot,w)$ and the associated sequence of minimizers π_t^k converges pointwise to the minimizer of $g_t^\sharp(x,\cdot,w)$. The corresponding distribution $\mu_{X_t^{\pi_t^k}}$ converges in distribution to the distribution $\mu_{X^{\pi_t^\sharp}}$ and $V_t^\sharp = \Gamma_t(V_{t+1}^\sharp)$ on $\mathrm{dom}(V_t^*)$.*

Proof The convergence to $V_t^\sharp$ is given by Proposition 5. By Proposition 9 we obtain the pointwise convergence of $(\pi_t^k(x,w))_{k\in\mathbb{N}}$ to $\pi_t^\sharp(x,w)$ which in turn gives the convergence of $(\mu_t^k)_{k\in\mathbb{N}}$ to $\mu_t^\sharp$ in distribution. The Corollary 1 gives the desired statement.

5 Conclusion

The results of convergence obtained are not based on any finite representation of the random variables invoked in the problem formulation. The Proposition 10 states the convergence to a steady state distribution of the state variable along iterations of the algorithm. The steady state distribution turns out to be the optimal distribution of the state variable. This result highlights the ability of the SDDP-RIMA algorithm to visit the trajectories of the state space taking into account the optimal probability distribution of the state variable. Throughout the paper we have not assumed any linearity assumption, nevertheless the second requirement of Assumption 3 can be dropped if the problem is linear and the random variables are discretes. We have assumed that all function invoked in our formulation are continuous with respect to the random variables, this requirement can be dropped to if the underlying probability space is discrete.

References

1. J. Benoist and A. Danillidis. Dual characterizations of relative continuity of convex functions. *A. Australia. Math. Soc*, pages 211–223, 2001.
2. D.P. Bertsekas and S.E. Shreve. *Stochastic Optimal Control: the discrete-time case*. Athena Scientific, Belmont, 1996.
3. D.P. Bertsekas and J.N. Tsitsiklis. Gradient convergence in gradient methods. *SIAM J. Optim.*, 10(3):627–642, 2000.
4. P. Billingsley. *Probability and Measure*. Wiley interscience, 1995.
5. I. Ekeland and R. Témam. *Convex Analysis and Variational Problems*. SIAM, 1999.
6. P. Girardeau and A. Philpott. On the convergence of decomposition methods for multistage stochastic convex program. Optimization online, April 2012.
7. M.V.F. Peirera and L.M.V.G. Pinto. Stochastic optimization of a multireservoir hydroelectric system : A decomposition approach. *Water Ressource Research*, 21(6):779–792, June 1985.
8. M.V.F. Pereira and L.M.V.G Pinto. Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming*, (52):359–375, 1991.
9. A.B. Philpott and Z. Guan. On the convergence of stochastic dual dynamic programming and related methods. (36):450–455, 2008.
10. E.G. Read and J.A. George. Dual dynamic programming for linear production/inventory systems. *Computers Math, Applic*, 19(11):29–42, 1990.
11. S.M. Robinson. Analysis of sample-path optimization. *Math. Oper. Res.*, 21(3):119–147, 1996.
12. R.T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970.
13. R.T. Rockafellar and R.J.-B. Wets. *Variational Analysis*. Springer Verlag, Berlin Heidelberg, 1998.
14. J. Rust. Using randomization to break the curse of dimensionality. *Econometrica Society*, 65(3):487–516, May 1997.
15. Alexander Shapiro. Analysis of stochastic dual dynamic programming method. *European Journal of Operational Research*, 209(1):63–72, February 2011.