A LEAST SQUARES MONTE CARLO METHOD FOR PROBLEMS OF OPTIMAL STOCHASTIC CONTROL WITH CONVEX VALUE FUNCTIONS

JURI HINZ*

We present a method for optimal policy calculation of stochastic control problems whose value functions are convex. Problems of this type appear in many applications and encompass important examples arising in the area of optimal stopping and in the framework of control, based on partial observations. Given convexity of value functions, we suggest a basis-free modification of the classical least-squares approach.

1. Introduction. Complex industrial problems are frequently formulated as sequential decisions under uncertainty. Im mathematical terms, these problems can be addressed within the so-called *Markov decision theory*, or more generally, as *discrete-time stochastic control*. Although these fields provide theoretical foundations to relevant decision problems, the complexity of typical real-world questions usually goes beyond what is computationally feasible. From practical view, any good approximation is of great interest, if exact solution is not reachable. With this perspective, the theory of *approximate dynamic programming* aims to provide a generalized view on theoretical insights, working solutions, and well-performing heuristics in the area of stochastic control. The interested reader will find in the book [10] an in-depth review of current challenges from industrial practice and on the state-of-art in the theory of approximate stochastic programming. This comprehensive source also contains detailed bibliographical references to the most important work in this field.

Crucial challenges in stochastic control are originated from high-dimensionality of the state space, of the observation space, and of the space of available actions, as pointed out in [10]. In the present work, we focus on the *first* dimensionality course. Assuming a Markovian state evolution whose dynamics can be controlled by a finite number of actions, we restrict ourselves to problems where only the state space is possible high-dimensional. Given such complex state space, any practical solution of a given control problem requires either a finite discretization of the state space or an efficient approximation of value functions. In the latter case, diverse approximation methods of continuous-value functions are linked to diverse statistical methodologies, including the so-called least squares Monte Carlo method.

Motivated by applications in financial mathematics, the least squares Monte Carlo method has attracted increasing attention. Utilizing [2] and [12], the work [6] has popularized this approach. Subsequent research focused on its theoretical justification. For instance, in [3], convergence issues of the least squares method were addressed and later generalized in [11]. Trying to capture the local behavior of value function, advanced statistical techniques like kernel-based regression methods [7], [8], local polynomial regression [4], and neural networks [1] have been investigated. For the particular case of partially observable Markov decision processes [9], diverse specific approaches been suggested.

*National University of Singapore, Department of Mathematics, 10 Lower Kent Ridge Road, Singapore 119076

A survey [5] gives an overview on these methods, with applications to autonomous robot navigation. In the case of partially observable Markov decision processes the value functions are convex, which is related to our approach.

In this work, we suggest an *approximate* calculation of the optimal control policy based on a modification of the classical least squares Monte Carlo method. Under the assumption that the value function of the original problem is convex, we develop a *basis-free* regression method which is, at least theoretically, not restricted to the course of dimensionality. Note that the applicability of Monte Carlo techniques is independent on state space dimension. Thus, we hope that our approach could help treating large-scale optimal control problems, whose value functions are convex.

This paper is organized as follows. In Section 2, we introduce a controlled system, consisting of two parts. The first part (system position) is on the total deterministic control of the agent and takes a finite number values, whereas the other part (system state) follows a free uncontrolled Markovian evolution, realized on a possible high-dimensional state space. In Section 3, an adaptation of the least squares method under convexity assumption is discussed, whose theoretical justification is given in Section 6.

2. Optimal control of a Markov system. Consider an agent who is confronted with a sequential decision under uncertainty, where a certain facility interacts with a random environment and needs to be re-adjusted at any time in order to achieve a position which is optimal with respect to the present cost caused by the interaction, considering revenue/costs from the potential re-positioning, which may be required in the future. As an illustration, the reader could imagine a typical example of storage facility management, where a commodity stock needs to be controlled at any time. Given storage costs and random price fluctuation, the agent has to decide when to withdraw the commodity from the storage in order to sell it at the market price and when to buy in order to top up the storage level.

Let us introduce the ingredients of our model. Consider a finite set $\mathcal{P}$ of *positions* representing all possible states of the system, which are under agent's control. For instance, in the above case of storage, this set naturally corresponds to all possible storage levels. Further, introduce a finite set $\mathcal{A}$ of possible actions which can be taken to change position and consider the function

$$\alpha : \mathcal{P} \times \mathcal{A} \to \mathcal{P}, \qquad (p, a) \mapsto \alpha(p, a) \tag{2.1}$$

which represents position change, with the interpretation that $\alpha(p, a)$ describes a new position which is reached from the previous position p by taking the action a. In our inventory example, the action a may describe one of the decisions to withdraw or to accommodate a unit of stock, which yields a transition from the previous storage level p to the new level $\alpha(p, a)$.

Further, let us consider the cost of position change. Here, we suppose that due to a random interaction with the environment, the reward from position control is uncertain, driven by a time homogenous Markov process $(Z_t)_{t \in \mathbb{N}}$ on a state space $\mathcal{Z}$. At any time $t \in \mathbb{N}$, the instantaneous reward depends on the position p, on the action a, and on the current realization $z \in \mathcal{Z}$ of the state variable Z_t and is described by a pre-specified function

$$R_t : \mathcal{P} \times \mathcal{Z} \times \mathcal{A} \to \mathbb{R}, \qquad (p, z, a) \to R_t(p, z, a).$$

In this context, $R_t(p, z, a)$ stands for the revenue/costs caused by the action a at time $t \in \mathbb{N}$ depending on the current position p and on the system state z.

To continue the illustration by inventory, the reader may imagine $(Z_t)_{t \in \mathbb{N}}$ as a process, which describes the Markovian evolution of the market price of the underlying good. Realistically, only one of the components of Z_t stores the market price listed at time t, whereas other components may contain latent variables (representing market situation) which are needed to ensure the Markovian property of the dynamics. In the inventory example, the value $R_t(p, z, a)$ describes the cash flow associated with the decision a to buy or to sell the good at time t. Note that the value $R_t(p, z, a)$ depends not only on the action a and on the market price through the corresponding price-component of z, but may also depend on the recent inventory level p. For instance, in the case of gas storage, the injection/withdawal costs depend on the storage level through the gas pressure.

Let us introduce the control at time $t = 0, 1, 2, \ldots,$ by

$$\pi_t : \mathcal{P} \times \mathcal{Z} \to \mathcal{A}, \qquad (p, z) \to \pi_t(p, z)$$

which stands for a rule to chose at time t in the position $p \in \mathcal{P}$ and in the system state $z \in \mathcal{Z}$ the action $a := \pi_t(p, z)$. A control policy $\pi = (\pi_t)_{t=0}^{N}$ is given by a sequence of decision rules, where each control π_t is applied at corresponding time t. Following π, the system positions $(p_t^\pi)_{t=0}^{N}$ and actions $(a_t^\pi)_{t=0}^{N}$ evolve as stochastic processes, recursively given by

$$a_t^\pi := \pi_t(p_t^\pi, Z_t), \qquad p_{t+1}^\pi := \alpha(p_t^\pi, a_t^\pi), \qquad t = 0, \ldots, N,$$

started at time $t = 0$ in the state Z_0 with the initial position $p_0 = p_0^\pi$. In the storage example, $a = \pi_t(p, z)$ could denote the decision a to buy or to sell a unit of stock at time t depending on the present inventory level p and on the current market situation z.

Finally, let us turn to the definition of the control target. We suppose that the reward from following a policy π is accumulated within the entire time horizon $0, 1, 2, \ldots, N$ whose expected value is given by

$$E(\sum_{t=0}^{N} R_t(p_t^\pi, Z_t, a_t^\pi)). \tag{2.2}$$

The problem of the optimal stochastic control deals with the the calculation of the policy π^* which maximizes the expected total reward (2.2).

According to maximum principle of the stochastic control, the calculation of the optimal policy π^* is addressed in terms of the so-called value functions. The value function $V_t^\pi(p, z)$ stands for the remaining revenue, expected at time $t = 0, \ldots, N$, if the if the policy $\pi = (\pi_s)_{s=0}^{N}$ is supposed to be followed until the end of the time horizon, conditioned on the position $p \in \mathcal{P}$ and on the state $z \in \mathcal{Z}$, in other words

$$V_t^\pi(p, z) = E(\sum_{s=t}^{N} R_s(p_s^\pi, Z_s, a_s^\pi) \mid Z_t = z, p_t^\pi = p)$$

for $t = 0, \ldots, N$, $p \in \mathcal{P}$, and $z \in \mathcal{Z}$. The value functions $(V_t^\pi(p, z))_{t=0}^{N}$ of the policy $\pi = (\pi_s)_{s=0}^{N}$ obey a recursive relation

$$V_t^\pi(p, z) = R_t(p, z, \pi_t(p, z)) + \int_{\mathcal{Z}} V_{t+1}^\pi(\alpha(p, \pi_t(p, z)), z') \mathcal{K}(z, dz'), \tag{2.3}$$

for all $(p, z) \in \mathcal{P} \times \mathcal{Z}$. Here $\mathcal{K}(z, dz')$ denotes the transition kernel of the Markovian time-homogeneous process $(Z_t)_{t \in \mathbb{N}}$. Note that if the recursion (2.3) can be followed backward through time, starting from $t = N$ with convention $V_{N+1}^\pi = 0$. This backward calculation method yields all value functions $(V_t^\pi(p, z))_{t=0}^N$. The policy optimization is addressed in terms of value functions. Namely, the policy π^* is called optimal among policies from appropriately specified class, if π^* maximizes the total reward (2.2), in other words, if

$$V_0^{\pi^*}(p, z) \geq V_0^\pi(p, z) \qquad \text{for all } (p, z) \in \mathcal{P} \times \mathcal{Z}$$

holds for each policy π chosen from this class. Under specific assumptions on the state space $\mathcal{Z}$, an optimal policy π^* is characterized by the following optimality equations

$$V_t^*(p, z) = \max_{a \in \mathcal{A}} \left(R_t(p, z, a) + \int_{\mathcal{Z}} V_{t+1}^*(\alpha(p, a), z') \mathcal{K}(z, dz') \right), \tag{2.4}$$

$$\pi_t^*(p, z) = \text{argmax}_{a \in \mathcal{A}} \left(R_t(p, z, a) + \int_{\mathcal{Z}} V_{t+1}^*(\alpha(p, a), z') \mathcal{K}(z, dz') \right) \tag{2.5}$$

for all $p \in \mathcal{P}$, $z \in \mathcal{Z}$ and $t = 0, \dots, N$. Theoretically, these equations provide a recursive scheme (called backward induction) to determine an optimal policy π^*, starting at $t = N$ with $V_{N+1}^* \equiv 0$.

3. Least-squares method and convexity. To determine an optimal policy π^* by backward induction, one needs to calculate in (2.4) and (2.5) the integrals of the type

$$\int_{\mathcal{Z}} f(z) \mathcal{K}(z, dz') \qquad z \in \mathcal{Z}.$$

Such integration can be very involving, particularly if the state space $\mathcal{Z}$ is high-dimensional. Note also that this calculation must be performed for any $z \in \mathcal{Z}$. This fact makes the numerics of high-dimensional optimal control problems persistently challenging. However, although there is no sound solution to all type of situations, the theory of *approximate dynamic programming* provides diverse techniques to tackle several important situations.

For the reminder of this work, suppose that the state space $\mathcal{Z}$ is not countable, in which case the calculation of the optimal policy becomes difficile. Here, further results can be achieved only under additional assumptions.

In what follows, we focus on one of the promising directions in the area of approximate stochastic control, on the so-called *least squares* approach and suggest an improvement to this technique, which works under specific assumption that

$$\text{the value functions are convex.} \tag{3.1}$$

Under this restriction, we consider the transition operator associated with the Markov process $(Z_t)_{t \in \mathbb{N}}$

$$(Tf)(z) = \int_{\mathcal{Z}} f(z') \mathcal{K}(z, dz') = E(f(Z_{t+1}) \,|\, Z_t = z), \qquad z \in \mathcal{Z} \tag{3.2}$$

which acts on appropriate functions $f : \mathcal{Z} \to \mathbb{R}$. In terms of the operator T, the characterization of the optimal policy by (2.4) and (2.5) is given by

$$V_t^*(p, z) = \max_{a \in \mathcal{A}} \left(R_t(p, z, a) + TV_{t+1}^*(\alpha(p, a), \cdot)(z) \right) \tag{3.3}$$

$$\pi_t^*(p, z) = \text{argmax}_{a \in \mathcal{A}} \left(R_t(p, z, a) + TV_{t+1}^*(\alpha(p, a), \cdot)(z) \right) \tag{3.4}$$

for all $p \in \mathcal{P}$, $z \in \mathcal{Z}$, and $t = 0, \ldots, N$. Clearly, an appropriate approximation of T is the key to any numerical implementation.

Let us demonstrate how the convexity assumption (3.1) can be utilized. For this, we return to the probabilistic interpretation of the Markov transition operator T and highlight the philosophy of the standard least squares Monte Carlo projection. Thereby, we explain problems which occur if T is approximated naively. Consider the $\mathcal{Z}$-valued random variables Z, Z', realized on a probability space $(\Omega, \mathcal{F}, P)$, whose regular conditional distribution satisfies

$$P(Z' \in dz' \mid Z = z) = \mathcal{K}(z, dz')$$

where $\mathcal{K}$ is the transition kernel defining the time homogenous Markov process $(Z_t)_{t \in \mathbb{N}}$. Consider a finite sample $\mathcal{S} \subset \mathcal{Z} \times \mathcal{Z}$ which consists of a finite number of point pairs $(z, z') \in \mathcal{Z} \times \mathcal{Z}$ obtained as a realizations of independent copies of (Z, Z'). By strong law of large numbers, the combination of the Dirac measures

$$\frac{1}{|\mathcal{S}|} \sum_{(z,z') \in \mathcal{S}} \delta_{(z,z')} \text{ approximates of the distribution of } (Z, Z'). \tag{3.5}$$

Let us denote the first component of this sample by Ξ

$$\Xi = \{z \in \mathcal{Z} : \text{there exists } z' \in \mathcal{Z} \text{ such that } (z, z') \in \mathcal{S}\}.$$

Given real-valued basis functions $(\psi_j)_{j=1}^m$ on $\mathcal{Z}$, the least-squares Monte Carlo transition operator $\tilde{T}$ acts on any function $f : \mathcal{Z} \to \mathbb{R}$ as

$$\tilde{T}f = \sum_{j=1}^m \tilde{\lambda}_j \psi_j, \tag{3.6}$$

where the coefficients $(\tilde{\lambda}_j)_{j=1}^m \in \mathbb{R}^m$ are obtained from the solution of the following problem:

$$\begin{array}{c} \text{determine a minimizer } (\tilde{\lambda}_j)_{j=1}^m \text{ of the sum of squared} \\ \text{errors } \sum_{(z,z') \in \mathcal{S}} |f(z') - \sum_{j=1}^m \lambda_j \psi_j(z)|^2 \text{ over } (\lambda_j)_{j=1}^m \in \mathbb{R}^m. \end{array} \tag{3.7}$$

Theoretically, the Monte-Carlo transition $\tilde{T}$ approximates the tue Markov transition operator T if the basis dimension m and the sample size $|\mathcal{S}|$ are chosen sufficiently large (see [11]). Here, we come to the main problems:

$$\begin{array}{c} \text{the increase of basis space may cause oscillation} \\ \text{of } \tilde{T}f, \text{ if the sample size is too low,} \end{array} \tag{3.8}$$

furthermore,

$$\begin{array}{c} \text{an appropriate choice of basis functions turns out to be} \\ \text{difficult, particularly for high-dimensional state space.} \end{array} \tag{3.9}$$

In what follows, we suggest a solution to both problems under the standing assumption (3.1).

3.1. Adaptive choice of convex basis functions. Having in mind the assumption (3.1), suppose that we have ensured that Tf is convex. In this case, the construction of the least-squares projection should be adapted accordingly.

We suggest to address(3.8) by searching in (3.7) for a minimizer of squared errors within a pre-specified cone of convex functions rather than within a given linear space. Based on this idea, we consider a modification of the Monte-Carlo transition operator $\tilde{T}$, introducing the *convex transition* $\breve{T}$.

Assume that

$$\text{the basis functions } (\psi_j)_{j=1}^m \text{ are convex,} \tag{3.10}$$

then $\breve{T}$ acts similarly to $\tilde{T}$, by

$$\breve{T}f = \sum_{j=1}^m \breve{\lambda}_j \psi_j, \tag{3.11}$$

but the coefficients $(\breve{\lambda}_j)_{j=1}^m$ are obtained by solution of a *constrained* quadratic optimization problem

$$\begin{aligned} &\text{minimize } \sum_{(z,z') \in \mathcal{S}} |f(z') - \sum_{j=1}^m \lambda_j \psi_j(z)|^2 \\ &\text{subject to } \lambda_j \in [0, \infty[\text{ for } j = 1, \ldots, m. \end{aligned} \tag{3.12}$$

Note that the non-negativity restriction ensures convexity of the resulting projection. From numerical viewpoint, the problem (3.12) is as simple as the classic problem (3.7), since efficient algorithms for quadratic minimization under linear constraints are available. The restriction to a cone adds a notable stability to the problem in the following sense: Too many basis functions can not give oscillations, due to the convexity constrained.

Next, let us turn to the second problem (3.9). In the applications of Monte Carlo methodology, the choice of a basis could be one of the most crucial problems. Thereby, this problem addresses the choice of shape of the basis functions ψ_j, $j = 1, \ldots, m$ and their number m. At first glance, the intuition may suggest that the dimension of the basis space should be sufficiently high. Still, at least theoretically, acceptable results can be obtained also with low-dimensional basis spaces, if the basis is chosen properly. Ideally, the shape of the basis functions should be similar to the targeted projection Tf. As a limiting case, the reader may imagine that an excellent result can be obtained with one dimensional basis space, provided it is spanned by the basis function $\psi_1 = Tf$. Of course, such a basis choice not feasible in practice, since Tf is not known in advance. However, after calculations of projection with preliminary basis, the new basis can be chosen such that it contains elements whose shape is similar to what is obtained from the preliminary calculation. Such procedure may be applied repeatedly, improving the projection step by step.

In what follows, we present a methodology which captures this idea of the *basis improvement operator* on a formal level. Given a finite sample $\mathcal{S} \subset \mathcal{Z} \times \mathcal{Z}$, consider the Hilbert space $\mathcal{H}$ of real-valued functions on $\mathcal{Z} \times \mathcal{Z}$ equipped with the scalar product

$$\langle u, v \rangle = \sum_{(z,z') \in \mathcal{S}} u(z, z') v(z, z'),$$

and denote by $\| \cdot \|$ the corresponding norm. For each $h : \mathcal{Z} \to \mathbb{R}$, we write $h \otimes \mathbb{I}$ and $\mathbb{I} \otimes h$ to denote functions on $\mathcal{Z} \times Z$:

$$(h \otimes \mathbb{I})(z, z') = h(z), \qquad (\mathbb{I} \otimes h)(z, z') = h(z') \qquad (z, z') \in \mathcal{Z} \times \mathcal{Z}.$$

Similarly, for each convex cone Ψ of functions on $\mathcal{Z}$, we assume that

$$\Psi \otimes \mathbb{I} = \{h \otimes \mathbb{I} : h \in \Psi\},$$

is the corresponding convex cone of functions on $\mathcal{Z} \times \mathcal{Z}$. Finally, we agree that $\Pi_{\Psi \otimes \mathbb{I}}$ stands for the projector onto the cone $\Psi \otimes \mathbb{I}$, acting on the Hilbert space $\mathcal{H}$. With these notations, the approximative transition $\check{T}_\Psi$ can be re-defined as

$$\check{T}_\Psi f \otimes \mathbb{I} = \Pi_{\Psi \otimes \mathbb{I}}(\mathbb{I} \otimes f) \qquad \text{for each } f : \mathcal{Z} \to \mathbb{R}. \tag{3.13}$$

and is characterized by the minimum distance property

$$\check{T}_\Psi f = \operatorname{argmin}_{\psi \in \Psi} \|\psi \otimes \mathbb{I} - \mathbb{I} \otimes f\|.$$

Suppose that for each convex function $\phi : \mathcal{Z} \to \mathbb{R}$ a specific procedure $\Psi(\cdot)$ determines $\Psi(\phi)$ where

$$\begin{aligned} &\Psi(\phi) \text{ is a convex cone which consist of convex functions} \\ &\text{on } \mathcal{Z}, \text{ contains } \phi \text{ and all constant functions.} \end{aligned} \tag{3.14}$$

Let us call such $\Psi(\cdot)$ a *cone envelope operator* in what follows. Given a convex function ϕ^0, the preliminary projection is calculated by $\phi^1 := \check{T}_{\Psi(\phi^0)} f$. Continuing, the projections can be re-calculated subsequently

$$\phi^{k+1} = \check{T}_{\Psi(\phi^k)} f, \qquad k \geq 1 \tag{3.15}$$

which gradually improves the approximation

$$\begin{aligned} \|\phi^{k+1} \otimes \mathbb{I} - \mathbb{I} \otimes f\| &= \|\check{T}_{\Psi(\phi^k)} f \otimes \mathbb{I} - \mathbb{I} \otimes f\| \\ &= \|\Pi_{\Psi(\phi^k) \otimes \mathbb{I}}(\mathbb{I} \otimes f) - \mathbb{I} \otimes f\| \\ &\leq \|\phi^k \otimes \mathbb{I} - \mathbb{I} \otimes f\| \end{aligned}$$

where the last estimate holds since $\phi^k \in \Psi(\phi^k)$ by assumption (3.14). Clearly, the success of the above projection improvement procedure relies on a lucky choice of the cone envelope operator $\Psi(\cdot)$. To clarify this aspect, we call a convex function ϕ *non-improvable projection* of f with respect to the operator $\Psi(\cdot)$ if holds

$$\phi = \check{T}_{\Psi(\phi)} f.$$

The most desirable property of $\Psi(\cdot)$ would be that the sequence of the improved projections $(\phi^k)_{k \geq 1}$ converges to a non-improvable projection ϕ and that the non-improvability means that the best approximation among the entire cone $\mathcal{C}$ of convex functions is reached. In the subsequent sections, we introduce a concrete cone envelope operator and study sufficient conditions for this to hold.

3.2. An algorithm for adaptive convex projection. Let us discuss a particular cone envelope operator. Write $\mathcal{L}$ to denote the linear space of all affine linear functions on $\mathcal{Z}$. Given a convex function ϕ on $\mathcal{Z}$, for each $l \in \mathcal{L}$ let us agree that

$$\begin{aligned} &\Psi_l(\phi) \text{ is the convex cone spanned} \\ &\text{by constant functions, } \phi \text{ and } \phi \vee l. \end{aligned}$$

The reader may imagine ϕ as a candidate for the targeted projection $\phi \sim \breve{T}_{\mathcal{C}} f$, whereas $\phi \vee l$ should be considered a possible improvement, suggesting to cut off one of the edges of ϕ. Apparently, if edge smoothing does not contribute to the projection improvement, we may expect that the best approximation among all convex functions is already reached:

$$\text{If } \phi = \breve{T}_{\Psi_l(\phi)} f \text{ holds for each} \atop \text{affine linear } l \in \mathcal{L}, \text{ then } \phi = \breve{T}_{\mathcal{C}} f. \tag{3.16}$$

Let us postpone the discussion of this issue to the next sections. A practical use would be implemented as follows: Given a modification $\phi \vee l$ of ϕ by affine linear $l \in \mathcal{L}$, calculate $\breve{T}_{\Psi_l(\phi)} f$. If the result $\breve{T}_{\Psi_l(\phi)} f$ coincides with the candidate ϕ for each $l \in \mathcal{L}$, then ϕ is the required projection.

Based on this idea, we suggest a stylized procedure to approach the desired result by a recursive improvement of two dimensional basis spaces. Such method requires the following steps:

0) Given f, chose a convex function ϕ.

1) For an affine linear $l \in \mathcal{L}$ calculate $\breve{T}_{\Psi_l(\phi)} f$.

2) If the result coincides $\breve{T}_{\Psi_l(\phi)} f = \phi$ with the original basis function ϕ, then repeat the step 1) with the same ϕ but another affine linear $l \in \mathcal{L}$.

3) In the other case $\breve{T}_{\Psi_l(\phi)} f \neq \phi$ repeat the step 1) with the new basis function $\breve{T}_{\Psi_l(\phi)} f$ and the same l.

4) The algorithm terminates if the steps 1)–2) are followed repeatedly for sufficiently many affine linear functions. Upon termination, the procedure returns the result ϕ.

In a practical implementation, it turns to be more efficient to work with more that two dimensions. To improve the algorithm performance, we suggest to replace two functions ϕ and $\phi \vee l$ by a number $m := S + D + 1$ $(S, D \in \mathbb{N})$ of convex functions

$$\varphi_1, \ldots, \varphi_S, \ \phi, \ \phi \vee l_1, \ldots, \phi \vee l_D.$$

Thereby, the basis elements $\phi_1, \ldots, \phi_S$ are static, since they do not change during the basis improvement, whereas all other elements are altered. Static elements must be chosen such that the convex cone, spanned by basis functions, includes all constant functions in order to fulfill (3.14). In each step, the preliminary projection ϕ is re-calculated and modified by affine linear functions $l_1, \ldots, l_D \in \mathcal{L}$, which also change from step to step. The algorithm consists of the following parts:

- **Step 0 (Initialization)** Given f and a finite sample $\mathcal{S}$, calculate the realizations $\beta := (f(z'))_{(z,z') \in \mathcal{S}}$ of f on all image points. Specify $S \in \mathbb{N}$ convex static basis elements $\{\varphi_1, \ldots, \varphi_S\}$. Given $D \in \mathbb{N}$ affine linear functions $\{l_1^{(0)}, \ldots, l_D^{(0)}\} \in \mathcal{L}$ and a convex $\phi^{(0)}$, define the basis as

$$\{\psi_1^{(0)}, \ldots, \psi_m^{(0)}\} = \{\varphi_1, \ldots, \varphi_S, \phi^{(0)}, \phi^{(0)} \vee l_1, \ldots, \phi^{(0)} \vee l_D\}.$$

- **Step 1 (Minimization)** For $k \geq 0$, calculate the matrix $M^{(k)}$ from the realizations of the basis elements on the sample

$$M_{z,j}^{(k)} := \psi_j^{(k)}(z), \qquad z \in \Xi, \ j = 1, \ldots, m.$$

Determine the coefficient vector $\lambda^{(k)} = (\lambda_i^{(k)})_{j=1}^m \in [0, \infty[^M$ as the minimizer to the constrained problem

$$[0, \infty[^m \to \mathbb{R}, \quad \lambda \mapsto \lambda^\top M^{(k)\top} M^{(k)} \lambda - 2\lambda^\top M^{(k)\top} \beta.$$

and calculate new projection function

$$\phi^{(k+1)} := \sum_{j=1}^{m} \lambda_j^{(k)} \psi_j^{(k)}.$$

- **Step 2 (Test for Termination)** Determine the projection error

$$E^{(k+1)} = \left(\sum_{(z,z') \in \mathcal{S}} |\phi^{(k+1)}(z) - f(z')|^2 \right)^{\frac{1}{2}}.$$

If the improvement in the projection error falls below a certain threshold $\varepsilon > 0$

$$E^{(k+1)} - E^{(k)} < \varepsilon \quad \text{for } k \geq 0$$

then
  2a) Interrupt if 2a) is entered repeatedly for a critical number of past steps, else chose other linear functionals $\{l_1^{(k+1)}, \ldots, l_D^{(k+1)}\}$.
else
  2b) proceed with the same $\{l_1^{(k+1)}, \ldots, l_D^{(k+1)}\} = \{l_1^{(k)}, \ldots, l_D^{(k)}\}$.

- **Step 3 (Basis change)** Define the improved basis

$$\{\psi_1^{(k+1)}, \ldots, \psi_m^{(k+1)}\} := \{\varphi_1, \ldots, \varphi_S, \phi^{(k+1)}, \phi^{(k+1)} \vee l_1^{(k+1)}, \ldots, \phi^{(k+1)} \vee l_D^{(k+1)}\}$$

and go to Step 1.

Remark Note that the errors $(E^{(k)})_{k \geq 0}$ are non-negative and non-increasing, by construction. That is, the algorithm terminates after a finite number k^* of steps. Upon algorithm termination, the projection $\phi^{(k^*)}$ is determined at all points in Ξ as

$$(\phi^{(k^*)}(z))_{z \in \Xi} = M^{k*} \lambda^{k*}. \tag{3.17}$$

However, to calculate $\phi^{(k^*)}(z)$ at another point $z \notin \Xi$ one needs to run a recursive procedure, similar to the above procedure, using all basis change coefficients and affine linear functions used by the algorithm. That is, all data $(\lambda^{(k)})_{k=0}^{k^*}$ $(l_1^{(k)}, \ldots, l_D^{(k)})_{k=0}^{k^*}$ must be stored and returned upon termination.

Example Let us consider an illustration of the above procedure. Assume that (Z, Z') satisfies $Z' = Z + X$ where Z and X are independent $N(0,1)$-distributed random variables and suppose that he function f is given by $f(z) = z^2$ for all $z \in \mathbb{R}$. Thus, the transition operator yields a convex function $z \mapsto Tf(z) = z^2 + 1$. Let us calculate an approximation to Tf using our algorithm. Introduce $\mathcal{S}$ as realizations of 200 independent copies of (Z, Z'). We decided to chose static basis elements such that each cone contains the linear space of all affine liner functions and introduced dynamic elements $l_1^{(k)}, \ldots, l_D^{(k)}$ as constants, equal to the empirical quantiles of the recent projection $(\phi^{(k)}(z))_{z \in \Xi}$ at pre-determined probability levels $0 < \alpha_1 <, \ldots, < \alpha_D < 1$. With the initial projection $\phi^{(0)} : z \mapsto |z|$, given a threshold $\varepsilon = 0.03$ and $D = 10$ quantile levels, equidistantly spaced between 0.05 and 0.95, the algorithm has terminated after $k^* = 3$ iterations and returned a result depicted in the Figure 3.1 as a thick polygon line. By construction, this function is piecewise affine linear, and as expected, approximates the quadratic polynomial $z \mapsto Tf(z) = 1 + z^2$, which is plotted as a smooth thin line.
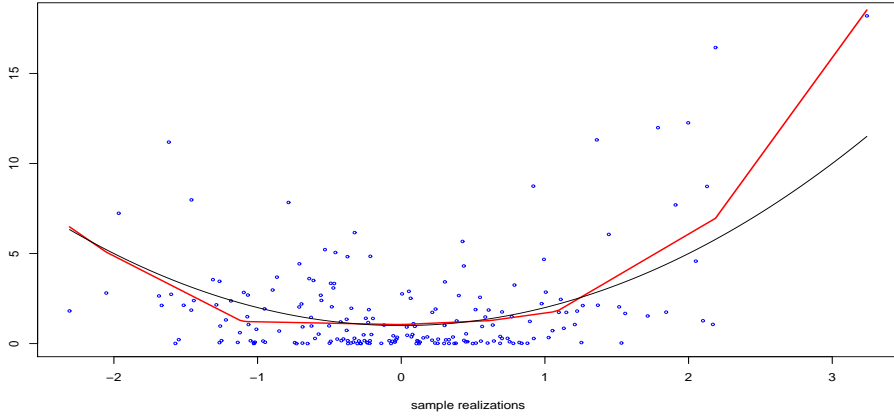
FIG. 3.1. $(z, f(z'))_{z \in \Xi}$ *(points), projection* $(z, \phi^{(k^*)}(z))_{z \in \Xi}$ *(polygone) and* $z \mapsto Tf(z)$ *(smooth line).*

The fit $\phi^{(k*)} \approx Tf$ is not perfect because the sample size $n = 200$ is relatively small. Note that our result $\phi^{(k^*)}$ gives a better fit than the exact projection Tf, which is seen from the sum of squared errors

$$\left(\sum_{(z,z') \in \mathcal{S}} |\phi^{(k^*)}(z) - f(z')|^2 \right)^{\frac{1}{2}} \approx 30.644, \qquad \left(\sum_{(z,z') \in \mathcal{S}} |Tf(z) - f(z')|^2 \right)^{\frac{1}{2}} \approx 31.913.$$

4. Approximate stochastic control. In the Section 6 we prove a sufficient condition (4.8), which ensures that the non-improvability implies that the best approximation among the entire cone $\mathcal{C}$ of all convex functions is reached:

$$\text{If (4.8) holds, then } \phi = \breve{T}_{\Psi_l(\phi)} f \text{ for each } l \in \mathcal{L} \text{ implies that } \phi = \breve{T}_{\mathcal{C}} f. \tag{4.1}$$

It turns out that (4.8) is always satisfied in one dimensional $d = 1$ situations. Although in general case $d > 1$ the validity of (4.8) is hard to control, numerical experiments suggest that our approach provides a fast an reliable scheme for approximative calculation of a convex least squares approximation, even in high dimensions. On this account, we suggest to replace the classical backward induction procedure (2.4) and (2.5) by

$$\breve{V}_t^*(p, z) = \max_{a \in \mathcal{A}} \left(R_t(p, z, a) + \phi_t(\alpha(p, a))(z) \right) \tag{4.2}$$

$$\breve{\pi}_t^*(p, z) = \text{argmax}_{a \in \mathcal{A}} \left(R_t(p, z, a) + \phi_t(\alpha(p, a))(z) \right) \tag{4.3}$$

for all $(p, z) \in \mathcal{P} \times \mathcal{Z}$, where $(\phi_t(p))_{p \in \mathcal{P}}$ are non-improvable projections of the previous value function:

$$\phi_t(p) = \breve{T}_{\Psi_l(\phi_t(p))} \breve{V}_{t+1}^*(p, \cdot) \quad \text{ for all } p \in \mathcal{P} \text{ and each } l \in \mathcal{L}. \tag{4.4}$$

Example: As an illustration, we discuss valuation of an American put option in discrete time. Introduce the discounted asset price $(Z_t)_{t=0}^N$ at time steps $0, \ldots, N$ with step size $\Delta > 0$ as a sampled geometric Brownian motion

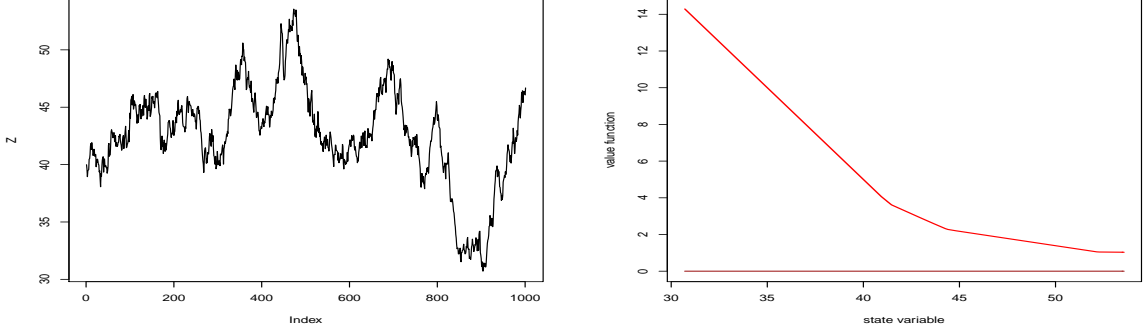$$Z_{t+1} = Z_t e^{X_{t+1}}, \quad t = 0, \ldots, N - 1, \tag{4.5}$$

10

FIG. 4.1. *Valuation of American Put Option.*

where $(X_t)_{t=1}^N$ are independent identically distributed random variables following normal distribution $N(-\sigma^2\Delta/2, \sigma^2\Delta)$ with volatility parameter $\sigma > 0$. Given short rate $r > 0$, the fair price of an American put option with strike price K and maturity date ΔN is given by

$$\sup\{E((e^{-r\Delta\tau}K - Z_\tau)^+), \qquad \tau \text{ is } \{0, 1, \ldots, N\}\text{-valued stopping time}\}.$$

This optimal stipping problem is defined by two positions and two actions $\mathcal{P} = \{1, 2\}$, $\mathcal{A} = \{1, 2\}$. The positions 'stopped' and 'goes' are represented by $p = 1$, $p = 2$ respectively and the actions 'stop' and 'go' denoted by $a = 1$ and $a = 2$. With this interpretation, the position change is given by

$$(\alpha(p, a))_{p,a} \sim \left[\begin{array}{cc} \alpha(1,1) & \alpha(2,1) \\ \alpha(1,2) & \alpha(2,2) \end{array} \right] = \left[\begin{array}{cc} 1 & 1 \\ 1 & 2 \end{array} \right].$$

The reward is paid only once, when the system transforms from the position 'goes' to 'stopped'

$$R_t(p, z, a) = (e^{-r\Delta t}K - z)^+ (p - \alpha(p, a)), \quad \text{for all } p \in \mathcal{P}, \ z \in \mathcal{Z}, \ a \in \mathcal{A}.$$

Given a path realization $(z_k)_{k=1}^{1000}$ with $z_0 = 40$, $\Delta = 1/220$ and $\sigma = 0.2$ depicted on the left hand side of the Figure 4.1, we define a sample as $\mathcal{S} = \{(z_k, z_{k+1}) : k = 1, \ldots, 999\}$. With this sample, having supposed that $r = 0.1$ and $K = 45$, an algorithm as in $(4.3) - (4.4)$ for $N = 80$ returned the value functions $\breve{V}_0^*$ illustrated on the left in the the Figure 4.1, where the upper line corresponds to $z \mapsto \breve{V}_0^*(2, z)$ whereas the lower line depicts $z \mapsto \breve{V}_0^*(1, z) = 0$.

Finally, to s explain sufficient condition (4.8) in (4.1), let us introduce some additional notions. Remember that our functions are given by their values on the sample Ξ, thus all relevant properties are inferred from them. On this account, it is convenient to treat a function $\phi : \mathcal{Z} \to \mathbb{R}$ in terms of its value vector $\phi = (\phi(z))_{z \in \Xi}$. In this setting, we make use of the characterization of convexity in terms of sub-gradients. That is, $\phi = (\phi(z))_{z \in \Xi}$ is convex if and only if there exists a finite family $(l_j)_{j \in J} \subset \mathcal{L}$ of affine linear functions such that

$$\phi(z) = \bigvee_{j \in J} l_j(z) \qquad \text{for all } z \in \Xi. \tag{4.6}$$

Let us fix a convex $\phi = (\phi(z))_{z \in \Xi}$, given in the representation (4.6) to define the sets

$$\Xi_j = \{z \in \Xi : l_j(z) = \bigvee_{i \in J} l_i(z)\}, \qquad j \in J$$

Obviously, $\cup_{j \in J} \Xi_j = \Xi$. Further, we assume that

$$\Xi_i \cap \Xi_j = \emptyset \quad \text{is satisfied for all } i, j \in J \text{ with } i \neq j. \tag{4.7}$$

In turns out that the following property of the sets $(\Xi_j)_{j \in J}$ ensures that a non-improvable projection is indeed the best approximation among all convex functions, in the sense of (4.1):

$$\left. \begin{array}{l} \text{for each } j \in J, \text{ an arbitrary convex function } (h(z))_{z \in \Xi_j} \text{ can be} \\ \text{expressed as a linear combination with non-negative coefficients of} \\ \text{affine linear functions and positive parts of affine linear functions.} \end{array} \right\} \tag{4.8}$$

In one dimensional case, $d = 1$ the assertion (4.8) holds, since affine linear functions and their positive parts span a cone, which contains all convex functions. In general situation $d \geq 1$, the validity of (4.8) depends on the dimension, on the geometry and on the cardinality of the sets Ξ_j, $j \in J$. For instance, if Ξ_j contains few elements, then the cone of affine linear functions and their positive parts is likely to span all convex functions.

5. Conclusion. We suggest a methodology for approximate solution of optimal control problems whose value functions are convex. Although convexity assumption appears restrictive, it is met by a broad class of problems, originated from many important practical applications. For instance, it is well-known [9] that diverse control problems based on partial observation yield high-dimensional stochastic control problems with convex value functions. In particular, autonomous robot navigation is connected to *such control problems*, which are also inherently high-dimensional. Namely, realistic applications require a state space, represented by a simplex of several hundert dimensions. Undoubted, treating such questions is difficult. However, since the applicability of Monte-Carlo methods is not restricted by space dimensionality, this contribution may be helpful in the practice, where where current methods of sequential decision making under uncertainty meet their computational limits.

6. Appendix: Non-improvable projections. Assume definitions and notations from the previous sections. Given a function $f : \mathcal{Z} \to \mathbb{R}$, we consider the functional F defined by

$$F(u) = \sum_{(z,z') \in \mathcal{S}} |f(z') - u(z)|^2, \qquad \text{for each } u = (u(z))_{z \in \Xi}.$$

The functional F is convex and attains its unique minimum ϕ on the cone

$$\mathcal{C} = \{ h \; : \; h = (h(z))_{z \in \Xi} \text{ is convex } \}$$

of all convex functions. In other words,

$$\begin{array}{l} \phi \in \mathcal{C} \text{ is the best convex approximation of } f \\ \text{in the sense that } F(\phi) \leq F(u) \text{ for all } u \in \mathcal{C}. \end{array} \tag{6.1}$$

The best convex approximation can be equivalently characterized in terms of the Frechet derivative $\partial F(\phi)$ of F at point ϕ, evaluated on the cone of all admissible directions at point ϕ. Namely, ϕ is the best convex approximation, if and only if holds:

$$\begin{array}{l} \partial F(\phi) \circ h \geq 0 \text{ for all } h \in \mathcal{A}(\phi) \text{ with the cone} \\ \mathcal{A}(\phi) = \text{Cone}\{ h = (h(z))_{z \in \Xi} \; : \; h = \psi - \phi \text{ where } \psi \in \mathcal{C} \}. \end{array} \tag{6.2}$$

Similarly, the function $\phi \in \mathcal{C}$ is a projection, non-improvable by affine linear function $l \in \mathcal{L}$ if holds $F(\phi) \leq F(u)$ for all $u \in \Psi_l(u)$ which implies that $\partial F(\phi) \circ h \geq 0$ holds for all $h \in \tilde{\mathcal{A}}_l(\phi)$, where the cone $\tilde{\mathcal{A}}_l(\phi)$ of admissible directions can be written in terms of generating vectors as

$$\tilde{\mathcal{A}}_l(\phi) = \text{Cone}\{h = (h(z))_{z \in \Xi} \, : \, h = \psi - \phi \text{ with } \psi = \phi \vee l \text{ or } \psi = \phi + \lambda 1, \lambda \in \mathbb{R}\}.$$

On this account, if

$$
\begin{aligned}
&\phi \in \mathcal{C} \text{ is non-improvable projection in the sense that} \\
&F(\phi) \leq F(u) \quad \text{for all } u \in \Psi_l(\phi) \text{ and each } l \in \mathcal{L},
\end{aligned}
\tag{6.3}
$$

then

$$
\begin{aligned}
&\partial F(\phi) \circ h \geq 0 \text{ for all } h \in \tilde{\mathcal{A}}(\phi), \text{ where the directions cone } \tilde{\mathcal{A}}(\phi) \text{ is generated as} \\
&\tilde{\mathcal{A}}(\phi) = \text{Cone}\{h \, : \, h = \psi - \phi \text{ with } \psi = \phi \vee l \text{ where } l \in \mathcal{L} \text{ or } \psi = \phi + \lambda 1, \lambda \in \mathbb{R}\}.
\end{aligned}
\tag{6.4}
$$

From this, we deduce that the question, whether a non-improvable projection coincides with the best convex approximation, is now reduced to the comparison of two cones:

If $\tilde{\mathcal{A}}(\phi) \supseteq \mathcal{A}(\phi)$, then (6.3) implies that (6.1) holds.

Namely, (6.3) implies (6.4), and with $\tilde{\mathcal{A}}(\phi) \supseteq \mathcal{A}(\phi)$ also (6.2), which gives (6.1).

It the reminder of this section, we examine conditions, ensuring the inclusion $\tilde{\mathcal{A}}(\phi) \supseteq \mathcal{A}(\phi)$.

We now determine the function family, which spans the cone $\mathcal{A}(\phi)$.

LEMMA 6.1. *Given (4.7), it holds*

$$\mathcal{A}(\phi) = \text{Cone}\{h = (h(z))_{z \in \Xi} \, : \, (h(z))_{z \in \Xi_j} \text{ is convex for each } j \in J\}. \tag{6.5}$$

Proof. First, we show that $\mathcal{A}(\phi)$ is a subset of the right-hand side of (6.5). Given $h \in \mathcal{A}(\phi)$, for each $j \in J$ it holds

$$h(z) = \psi(z) - \phi(z) = \psi(z) - l_j(z) \quad \text{for all } z \in \Xi_j$$

which shows that for each $j \in J$, the restriction $(h(z))_{z \in \Xi_j}$ of h to Ξ_j yields a convex function. Now, we show the opposite inclusion. Therefore, it suffices to show for each $j \in J$ that for an arbitrary convex $h_j = (h_j(z))_{z \in \Xi_j}$ there exists an $\varepsilon_j \in]0, \infty[$ such that the function $h^j = (h^j(z))_{z \in \Xi}$ obtained by the extension of $\varepsilon_j h_j$ to the entire domain Ξ as

$$
h^j(z) = \begin{cases} \varepsilon_j h_j(z) & \text{if } z \in \Xi_j \\ 0 & \text{if } z \in \Xi \setminus \Xi_j \end{cases},
\tag{6.6}
$$

satisfies $h^j \in \mathcal{A}(\phi)$. Indeed, with this property, an arbitrary $h = (h(z))_{z \in \Xi}$ from the right hand side of (6.5) can be decomposed into convex components $h_j = (h(z))_{z \in \Xi_j}$ for all $j \in J$ giving elements $h^j \in \mathcal{A}(\phi)$ along with $\varepsilon_j \in]0, \infty[$ for all $j \in J$ which yield h as convex linear combination

$$h = \sum_{j \in J} \frac{1}{\varepsilon_j} h^j.$$

Since $\frac{1}{\varepsilon_j} h^j \in \mathcal{A}(\phi)$, so we obtain $h \in \mathcal{A}(\phi)$. That is, it remains to prove that for an arbitrary convex $h_j = (h_j(z))_{z \in \Xi_j}$, there exists $\varepsilon_j \in]0, \infty[$ such that h^j from (6.6) satisfies $h^j \in \mathcal{A}(\phi)$.

Given a convex $h_j = (h_j(z))_{z \in \Xi_j}$, determine its representation through a maximum over a family $(l_k^j)_{k \in K} \subset \mathcal{L}$ of affine linear functions

$$h_j(z) = \bigvee_{k \in K} l_k^j(z) \qquad \text{for all } z \in \Xi_j$$

and let us denote this natural extension $\bigvee_{k \in K} l_k^j$ of h_j on the entire set Ξ by the same symbol $h_j = (h_j(z))_{z \in \Xi}$. Due to (4.7), it is possible to find a sufficiently small $\varepsilon_j > 0$ such that

$$l_j(z) + \varepsilon_j h_j(z) < \bigvee_{i \in J \setminus \{j\}} l_i(z) = \phi(z) \quad z \in \Xi \setminus \Xi_j, \tag{6.7}$$

$$l_j(z) + \varepsilon_j h_j(z) > \bigvee_{i \in J \setminus \{j\}} l_i(z) \quad z \in \Xi_j. \tag{6.8}$$

Define now a convex function $\psi = (\psi(z))_{z \in \Xi}$ by

$$\psi = \left(\bigvee_{j \in J \setminus \{j\}} l_j \right) \vee (l_j + \varepsilon_j h_j)$$

which coincides with ϕ on $\Xi \setminus \Xi_j$

$$\psi(z) = \phi(z) \qquad \text{for } z \in \Xi \setminus \Xi_j, \text{ due to (6.7)}$$

and agrees with $\phi + \varepsilon_j h_j$ on Ξ_j

$$\psi(z) = l_j(z) + \varepsilon_j h_j(z) = \phi(z) + \varepsilon_j h_j(z) \qquad \text{for } z \in \Xi_j, \text{ due to (6.8)}.$$

In other words, the function $h^j := \psi - \phi \in \mathcal{A}(\phi)$ fulfills (6.6) as desired. □

For an affine linear function $l \in \mathcal{L}$, let us agree to denote its positive part by $l^+ : z \mapsto \max(l(z), 0)$. Then we characterize the cone spanned by all directions from ϕ to specific convex functions $\psi = \phi \vee l$ where l is an arbitrary affine linear function, similarly to the previous lemma:

LEMMA 6.2. *Given (4.7), the following inclusion holds:*

$$\text{Cone}\{h = (h(z))_{z \in \Xi} : h = \psi - \phi : \text{ with } \psi = \phi \vee l \text{ where } l \in \mathcal{L}\} \tag{6.9}$$

$$\supseteq \text{Cone}\{h = (h(z))_{z \in \Xi} : \forall j \in J \exists l \in \mathcal{L} \text{ such that } (h(z) = l^+(z))_{z \in \Xi_j}\}. \tag{6.10}$$

Proof.

As in the previous lemma, it suffices to prove that given $j \in J$ and $l \in \mathcal{L}$, there exists $\varepsilon_j \in]0, \infty[$ sucht that

$$h^j(z) = \begin{cases} \varepsilon_j l^+(z) & \text{if } z \in \Xi_j \\ 0 & \text{if } z \in \Xi \setminus \Xi_j \end{cases} \tag{6.11}$$

fulfills

$$h^j(z) = \psi(z) - \phi(z) \qquad \text{for all } z \in \Xi.$$

with

$$\psi = \phi \vee \tilde{l} \qquad \text{where } \tilde{l} \in \mathcal{L}. \tag{6.12}$$

Given l^+ and $j \in J$, by the same arguments as in the previous lemma, there exists $\varepsilon_j \in]0, \infty[$ such that

$$l_j(z) + \varepsilon_j l^+(z) < \bigvee_{i \in J \setminus \{j\}} l_i(z) \quad \text{for all } z \in \Xi \setminus \Xi_j,$$

$$l_j(z) + \varepsilon_j l^+(z) > \bigvee_{i \in J \setminus \{j\}} l_i(z) \quad \text{for all } z \in \Xi_j.$$

Define now a convex $\psi = (\psi(z))_{z \in \Xi}$ by

$$\psi(z) = \left(\bigvee_{i \in J \setminus \{j\}} l_i(z) \right) \vee (l_j(z) + \varepsilon_j l^+(z))$$

$$= \left(\bigvee_{i \in J} l_i(z) \right) \vee (l_j(z) + \varepsilon_j l(z)). \tag{6.13}$$

By construction, ψ coincides with ϕ on $\Xi \setminus \Xi_j$ and agrees with $l_j + \varepsilon_j l^+ = \phi + \varepsilon_j l^+$ on Ξ_j. That is $h^j = \psi - \phi$ is as in (6.11). Furthermore, (6.13) shows that $\psi = \phi \vee \tilde{l}$ with affine linear function $\tilde{l} = l_j + \varepsilon_j l$ as required in (6.12) $\square$

We now extend the inclusion (6.9) of (6.10) and characterize the cone of all directions from ϕ to specific convex functions ψ, where either $\psi = \phi \vee l$ with an affine linear l or $\psi = \phi + \lambda \mathbb{1}$ with $\lambda \in \mathbb{R}$.

LEMMA 6.3. *Given (4.7), the following inclusion holds:*

Cone$\{h : h = \psi - \phi : \text{with } \psi = \phi \vee l \text{ where } l \in \mathcal{L} \text{ or } \psi = \phi + \lambda \mathbb{1}, \lambda \in \mathbb{R}\}$ (6.14)

$\supseteq$ Cone$\{h : \forall j \in J \exists \lambda \in \mathbb{R} \text{ with } (h(z) = \lambda \mathbb{1})_{z \in \Xi_j} \text{ or } \exists l \in \mathcal{L} \text{ with } (h(z) = l^+(z))_{z \in \Xi_j}\}$ (6.15)

Proof. Note that due to the previous lemma, each positive constant function

$$e^j(z) = \begin{cases} 1 & \text{if } z \in \Xi_j \\ 0 & \text{if } z \in \Xi \setminus \Xi_j \end{cases} \qquad j \in J$$

is in the cone (6.9) and so in the cone (6.14), so we merely need to show only that $-e^j$ is also contained in (6.14). Setting $\psi = \phi - 1$ we conclude that $-1 = \psi - \phi = -\sum_{i \in J} e^i$ is in (6.14). With this, negative constant functions $-e^j = -1 + \sum_{i \in J \setminus \{j\}} e^i$ is in (6.14), for each $j \in J$. $\square$

Using previous lemmata, we finally address the inclusion $\tilde{\mathcal{A}}(\phi) \supseteq \mathcal{A}(\phi)$. Therefore, note that (6.14) is nothing but the cone $\tilde{\mathcal{A}}(\phi)$, which contains a cone (6.15). If in turn, the cone (6.15) includes $\mathcal{A}(\phi)$ as a sub-cone, then the targeted assertion $\tilde{\mathcal{A}}(\phi) \supseteq \mathcal{A}(\phi)$ is fulfilled. Consequently, we deduce

PROPOSITION 6.4. *If (4.7) is satisfied and the following condition holds*

Cone$\{h : \forall j \in J \exists \lambda \in \mathbb{R} \text{ with } (h(z) = \lambda \mathbb{1})_{z \in \Xi_j} \text{ or } \exists l \in \mathcal{L} \text{ with } (h(z) = l^+(z))_{z \in \Xi_j}\}$
$\supseteq$ Cone$\{h : \text{for each } j \in J, (h(z))_{z \in \Xi_j} \text{ is convex}\}$,   (6.16)

then (6.3) implies that (6.1) holds.

Although the condition (6.16) is technical, it allows a natural interpretation (4.8).

REFERENCES

[1] D. P. Bertsekas and J. N. Tsitsiklis. *Neuro-Dynamic Programming*. Athena Scientific, 1996.

[2] J. F. Carriere. Valuation of the early-exercise price for options using simulations and nonparametric regression. *Insurance: Mathematics and Economics*, 19:19–30, 1996.

[3] E. Clément, D. Lamberton, and P. Protter. An analysis of the longstaff-schwartz algorithm for american option pricing. *Finance and Stochastics*, 6(4):449–471, 2002.

[4] J. Fan and I. Gijbels. *Local Polynomial Modelling and Its Applications*. Chapman and Hall, 1996.

[5] M. Hauskrecht. Value-function approximations for partially observable markov decision processes. *Journal of Artificial Intelligence Research*, 13:33–94, 2000.

[6] F. Longstaff and E. Schwartz. Valuing american options by simulation: a simple least-squares approach. *Review of Financial Studies*, (14):113–147, 2001.

[7] D. Ormoneit and P. Glynn. Kernel-based reinforcement learning. *Machine Learning*, 49:161–178, 2002.

[8] D. Ormoneit and P. Glynn. Kernel-based reinforcement learning in average-cost problems. *IEEE Transactions in Automatic Control*, 47:1624–1636, 2002.

[9] Kaelbling L. P., Littman M. L., and A. R Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101:99–134, 1998.

[10] W. B. Powell. *Approximate dynamic programming: Solving the curses of dimensionality*. Wiley, 2007.

[11] L. Stentoft. Convergence of the least squares monte carlo approach to american option valuation. *Management Science*, 50(9):576–611, 2004.

[12] J. N. Tsitsiklis and B. Van Roy. Regression methods for pricing complex american-style options. *IEEE Transactions on Neural Networks*, 2001.