The Annals of Applied Probability
2005, Vol. 15, No. 2, 1396–1432
DOI 10.1214/105051605000000043

MONTE CARLO ALGORITHMS FOR OPTIMAL STOPPING AND STATISTICAL LEARNING

BY DANIEL EGLOFF

Zurich Cantonal Bank

We extend the Longstaff–Schwartz algorithm for approximately solving optimal stopping problems on high-dimensional state spaces. We reformulate the optimal stopping problem for Markov processes in discrete time as a generalized statistical learning problem. Within this setup we apply deviation inequalities for suprema of empirical processes to derive consistency criteria, and to estimate the convergence rate and sample complexity. Our results strengthen and extend earlier results obtained by Clément, Lamberton and Protter [*Finance and Stochastics* **6** (2002) 449–471].

1. Introduction. The problem of arbitrage-free pricing American options has renewed the interest in efficient methods for numerically solving high-dimensional optimal stopping problems. In this paper we explain how to solve a discrete-time, finite-horizon optimal stopping problem by restating it as a generalized statistical learning problem. We give a unified treatment of the Longstaff–Schwartz and the Tsitsiklis–Van Roy algorithm. They use both Monte Carlo simulation and linearly parameterized approximation spaces. We introduce a new class of algorithms which interpolate between the Longstaff–Schwartz and Tsitsiklis–Van Roy algorithm and relax the linearity assumption of the approximation spaces.

Learning an optimal stopping rule differs from the standard setup in statistical and machine learning in the sense that it requires a series of learning tasks, one for every time step, starting at the terminal horizon and proceeding backward. The individual learning tasks are connected by the dynamic programming principle. At each time step, the result depends on the outcome of the previous learning tasks. Connecting the subsequent learning tasks to a recursive sequence of learning problems leads to an error propagation. We control the error propagation by using a Lipschitz property and a suitable error decomposition which relies on the convexity of the approximation spaces. Finally, we estimate the sample error with exponential tail bounds for the supremum of empirical processes. To apply these techniques, we need to calculate the covering numbers of certain function classes. An important type of function class for which good estimates on the

Received October 2003; revised June 2004.

AMS 2000 subject classifications. Primary 91B28, 60G40, 93E20; secondary 65C05, 93E24, 62G05.

Key words and phrases. Optimal stopping, American options, statistical learning, empirical processes, uniform law of large numbers, concentration inequalities, Vapnik–Chervonenkis classes, Monte Carlo methods.

covering numbers exist are the so called Vapnik–Chervonenkis (VC) classes, see [1] or [51]. We prove that payoff functions evaluated at Markov stopping times parameterized by a VC-class of functions is again a VC-class. The covering number estimate of Haussler [23] then gives the required bounds. Our approach is conceptually different from Clément, Lamberton and Protter [15], which is purely tailored to the classical Longstaff–Schwartz algorithm with linear approximation. By exploiting convexity and fundamental properties of VC-classes, we can prove convergence and derive error estimates under less restrictive conditions, also if both the dimension of the approximating spaces and the number of samples tends to infinity.

This paper is structured as follows. The next background section discusses recent developments in numerical techniques for optimal stopping problems and summarizes the probabilistic tools which we use in this work. Section 3 reviews discrete-time optimal stopping problems. Section 4 shows how to restate optimal stopping as a statistical learning problem and introduces the dynamic look-ahead algorithm. In Section 5 we state and comment on our main results: a general consistency result for convergence, estimates of the overall error, the convergence rate and the sample complexity. The focus of the work lies in estimating the sample error. The proofs are deferred to Section 6 where we also introduce the necessary tools of the Vapnik–Chervonenkis theory.

2. Background.

Optimal stopping problems naturally arise in the context of games where a player wants to determine when to stop playing a sequence of games to maximize his expected fortune. The first systematic theory of optimal stopping emerged with Wald and Wolfowitz [57] on the sequential probability ratio test. The monographs by Chow, Robbins and Siegmund [14] and Shiryayev [46] provide an extensive treatment of optimal stopping theory.

The general no-arbitrage valuation of American options in terms of an optimal stopping problem begins with Bensoussan [5] and Karatzas [26]. Nowadays, American option valuation is an important application of optimal stopping theory. For more background on American options and financial aspects of the related optimal stopping problem, we refer to [27].

2.1. *Algorithms for solving optimal stopping problems.*

Optimal stopping problems generally cannot be solved in closed form. Therefore, several numerical techniques have been developed. Barone–Adesi and Whaley [2] propose a semi-analytical approximation. The binomial tree algorithm of Cox, Ross and Rubinstein [16] directly implements the dynamic programming principle. Other approaches comprise Markov chain approximations (see [30]) direct integral equation and PDE methods. The PDE methods are based on variational inequalities, developed in [6] or [25], the linear complementary problem (see [24]) or the free boundary value problem (see [52]). However, the viability of any of these methods is prohibited by the curse of dimensionality. For these algorithms the computing cost and storage needs grow exponentially with the dimension of the underlying state space.

To address this limitation, new Monte Carlo algorithms have been proposed. The first landmark papers in this direction are [9, 49] and [11]. Longstaff and Schwartz [36] introduce a new algorithm for Bermudan options in discrete time. It combines Monte Carlo simulation with multivariate function approximation. They show how to solve the optimal stopping problem algorithmically by a nested sequence of least-square regression problems and briefly outline a convergence proof. Tsitsiklis and Van Roy [50] independently propose an alternative parametric approximation algorithm on the basis of temporal-difference learning. Their approach relies on stochastic approximation of fixed points of contraction maps. They prove almost sure convergence by using stochastic approximation techniques as developed in [7, 31] or [32]. The Longstaff–Schwartz, as well as the Tsitsiklis–Van Roy algorithm, approximate the value function or the early exercise rule and, therefore, provide a lower bound for the true optimal stopping value. Rogers [43] proposes a method based on the dual problem which results in upper bounds. The overview paper [12] describes the state of development of Monte Carlo algorithms for optimal stopping as of 1998. A more recent reference is the book of Glasserman [20]. A comparative study of various Monte Carlo algorithms for optimal stopping can be found in [33].

Despite the contributions of Tsitsiklis and Roy [50], Longstaff and Schwartz [36] and Rogers [43], many aspects of Monte Carlo algorithms for optimal stopping, such as convergence and error estimates, remain unanswered. Clément, Lamberton and Protter [15] provide a complete convergence proof and a central limit theorem for the Longstaff–Schwartz algorithm. But there are so far no results on more general possibly nonlinear approximation schemes, the rate of convergence or error estimates. These problems are the main topics addressed in this paper.

2.2. *Probabilistic tools.* The main probabilistic tools which we apply in this paper are exponential deviation inequalities for suprema of empirical processes. These tail bounds have been developed by Vapnik and Chervonenkis [55], Pollard [40], Talagrand [48], Ledoux [34], Massart [37], Rio [42] and many others. Compared to central limit theorems, they are nonasymptotic and provide meaningful results already for a finite sample size. Deviation inequalities, together with combinatorial estimates of covering numbers in terms of the Vapnik–Chervonenkis dimension, are the cornerstones of statistical learning by empirical risk minimization. For additional details on statistical learning theory, we refer to [1, 17, 22, 38, 39, 53, 54, 56].

2.3. *Basic notation.* The following terminology and notation will be used throughout this paper. If μ is a measure on a measurable space $(M, \mathcal{A})$, we denote by $L_p(M, \mu)$ the usual L_p-spaces endowed with the norm $\|\cdot\|_{p,\mu}$. If we need to indicate the measure space, we write $\|\cdot\|_{p,M,\mu}$. Let $d_{p,\mu}$ be the induced metric $d_{p,\mu}(f,g)=\|f-g\|_{p,\mu}$.

Let (M,d) be a metric space. If $U\subset M$ is an arbitrary subset, we define the covering number

$$N(\varepsilon,U,d)=\inf\Big\{n\in\mathbb{N}\Big|\exists\{x_1,\dots,x_n\}\subset M \text{ such that } \forall x\in U \min_{i=1,\dots,n} d(x,x_i)\le\varepsilon\Big\}, \tag{2.1}$$

which is the minimum number of closed balls of radius ε required to cover U. The logarithm of the covering number is called the entropy. The growth rate of the entropy for $\varepsilon\to 0$ is a measure for the compactness of the metric space U.

Let $X,X_1,X_2,\dots$ be i.i.d. random elements on a measurable space $(M,\mathcal{A})$ with distribution P. The empirical measure of a random sample $X_1,\dots,X_n$ is the discrete random measure given by

$$P_n(A)=\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{X_i\in A\}},\qquad A\in\mathcal{A}, \tag{2.2}$$

or, if g is a function on M,

$$P_n g=\frac{1}{n}\sum_{i=1}^{n} g(X_i). \tag{2.3}$$

The empirical measure is a random measure supported on $(M^\infty, P^\infty, \mathcal{A}^\infty)$, where $M^\infty=\prod_{\mathbb{N}} M$ is the product space of countably many copies of M, P^∞ the product measure and $\mathcal{A}^\infty$ the product σ-algebra. The random variables X_i can now be identified with the ith coordinate projections.

3. Review of discrete time optimal stopping. Let $\mathbf{X}=(X_t)_{t=0,\dots,T}$ be a discrete time $\mathbb{R}^m$-valued Markov process. We assume $\mathbf{X}$ is canonically defined on the path space $\mathcal{X}=\mathbb{R}^m\times\cdots\times\mathbb{R}^m$ of $T+1$ factors and identify X_t with the projection onto the factor t. We endow $\mathcal{X}$ with the Borel σ-algebra $\mathcal{B}$. Let $\mathcal{F}_t$ be the smallest σ-algebra generated by $\{X_s|s\le t\}$ and $\mathbb{F}=(\mathcal{F}_t)_{t=0,\dots,T}$ the corresponding filtration.

Let P be the law of $\mathbf{X}$ on $\mathcal{X}$ and $\mu_t=P_{X_t}$ the law of X_t on $\mathbb{R}^m$. We introduce the spaces of Markov L_p-functions

$$L_p(\mathbf{X})=\{h=(h_0,\dots,h_T)|h_t\in L_p(\mathbb{R}^m,\mu_t),\ \forall t=0,\dots,T\}, \tag{3.1}$$

with norm

$$\|h\|_p=\sum_{t=0}^{T}\|h_t\|_{p,\mu_t}=\sum_{t=0}^{T}E[|h_t(X_t)|^p]^{1/p}. \tag{3.2}$$

For brevity, we drop the measures P, μ_t and the coordinate projections X_t in our notation whenever no confusion is possible. Also, if $h \in L_p(\mathbf{X})$ and $\mathbf{x} = (x_0, \dots, x_T) \in \boldsymbol{\mathcal{X}}$ is a point of the path space, we introduce the shorthand notation

$$h(\mathbf{x})_t \equiv h_t(x_t). \tag{3.3}$$

3.1. *Discrete time optimal stopping.* In the following $f \in L_1(\mathbf{X})$ is a nonnegative reward or payoff function. The optimal stopping problem consists of finding the value process

$$V_t = \operatorname*{ess\,sup}_{\tau \in \mathcal{T}(t,\dots,T)} E[f_\tau(X_\tau)|\mathcal{F}_t], \tag{3.4}$$

where the supremum is taken over the family $\mathcal{T}(t, \dots, T)$ of all $\mathbb{F}$-stopping times with values in $t, \dots, T$. Adding a positive constant ε to the payoff f just increases V_t by ε. We therefore can assume without loss of generality that $f \in L_1(\mathbf{X})$ is a positive payoff function. A stopping rule $\tau_t^* \in \mathcal{T}(t, \dots, T)$ is optimal for time t if it attains the optimal value

$$V_t = E\big[f_{\tau_t^*}\big(X_{\tau_t^*}\big)|\mathcal{F}_t\big]. \tag{3.5}$$

Once the value process is known, an optimal stopping rule at time t is given by

$$\tau_t^* = \inf\{s \geq t | V_s \leq f_s(X_s)\}. \tag{3.6}$$

To exploit the Markov property of the underlying process X_t, we introduce the value function

$$v_t(x) = \sup_{\tau \in \mathcal{T}(t,\dots,T)} E[f_\tau(X_\tau)|X_t = x]. \tag{3.7}$$

The Markov property implies $V_t = v_t(X_t)$. Closely related to the value process V_t is the process

$$Q_t = \operatorname*{ess\,sup}_{\tau \in \mathcal{T}(t+1,\dots,T)} E[f_\tau(X_\tau)|\mathcal{F}_t] = E\big[f_{\tau_{t+1}^*}\big(X_{\tau_{t+1}^*}\big)|\mathcal{F}_t\big], \tag{3.8}$$

which is defined for all $t = 0, \dots, T-1$. Again, by the Markov property, we get the representation $Q_t = q_t(X_t)$, where

$$q_t(x) = \sup_{\tau \in \mathcal{T}(t+1,\dots,T)} E[f_\tau(X_\tau)|X_t = x] = E\big[f_{\tau_{t+1}^*}\big(X_{\tau_{t+1}^*}\big)|X_t = x\big]. \tag{3.9}$$

We extend the definition of q_t up to the horizon T and set $q_T = f_T$. The function q_t is referred to as the continuation value. It represents the optimal value at time t, subject to the constraint of not stopping at t. The value function and the continuation value are related by

$$v_t(X_t) = \max\big(f_t(X_t), q_t(X_t)\big), \qquad q_t(X_t) = E[v_{t+1}(X_{t+1})|X_t]. \tag{3.10}$$

The dynamic programming principle implies a recursive expression for the value, the continuation value and the optimal stopping times. The recursion starts at the horizon T with $v_T(X_T) = q_T(X_T) = f_T(X_T)$ and proceeds backward for $t = T-1, \dots, 0$ according to

$$v_t(X_t) = \max\bigl(f_t(X_t), E[v_{t+1}(X_{t+1})|X_t]\bigr), \tag{3.11}$$

respectively,

$$q_t(X_t) = E\bigl[\max\bigl(f_{t+1}(X_{t+1}), q_{t+1}(X_{t+1})\bigr)|X_t\bigr]. \tag{3.12}$$

Similarly, the recursion for the optimal stopping rules τ_t^* starts at the horizon T with $\tau_T^* = T$. Given v_t, respectively, q_t and the optimal stopping rule τ_{t+1}^* at time $t+1$, the optimal stopping rule τ_t^* is determined by

$$\begin{aligned}\tau_t^* &= t\mathbb{1}_{\{v_t(X_t)=f_t(X_t)\}} + \tau_{t+1}^*\mathbb{1}_{\{v_t(X_t)>f_t(X_t)\}}\\ &= t\mathbb{1}_{\{q_t(X_t)\le f_t(X_t)\}} + \tau_{t+1}^*\mathbb{1}_{\{q_t(X_t)>f_t(X_t)\}}.\end{aligned} \tag{3.13}$$

From a theoretical point of view, the value function v_t and the continuation value q_t are equivalent since they both provide a solution to the optimal stopping problem. However, from an algorithmic point of view, the continuation value is preferred. Indeed, q_t tends to be smoother than v_t because the max operation introduces a kink in the value function. We note that in continuous time this kink disappears, since by the smooth fit principle, the value function connects C^1-smoothly to the payoff function along the optimal stopping boundary.

Expression (3.13) for the optimal stopping rule suggests that we consider stopping rules parameterized by functions $h \in L_1(\mathbf{X})$ with $h_T = f_T$. The terminal condition $h_T = f_T$ reflects the terminal boundary condition $\tau_T^* = T$. Let

$$\theta_{f,t}(h) = \theta(f_t - h_t), \qquad \theta_{f,t}^-(h) = 1 - \theta(f_t - h_t), \tag{3.14}$$

where $\theta(s) = \mathbb{1}_{\{s\ge 0\}}$ is the heaviside function. Set $\tau_T(h) = T$ and define recursively

$$\tau_t(h)(\mathbf{x}) = t\theta_{f,t}(h)(x_t) + \tau_{t+1}(h)(\mathbf{x})\theta_{f,t}^-(h)(x_t), \qquad \mathbf{x} \in \mathfrak{X}. \tag{3.15}$$

For every $h \in L_1(\mathbf{X})$, we get a valid stopping rule $\tau_t(h)$ which does not anticipate the future, because at each point in time t, the knowledge of X_t is sufficient to decide whether to stop or to continue.

DEFINITION 3.1. The family of stopping rule $\{\tau_t(h)|h \in L_1(\mathbf{X}),\ h_T = f_T\}$ is called the set of *Markov stopping rules*.

The stopping rule $\tau_t(h)$ depends only on $h_t, \dots, h_{T-1}$ and is therefore constant as a function of the arguments $x_0, \dots, x_{t-1}$. Moreover, the recursion formula (3.13) implies that the optimal stopping rule τ_t^* at time t is identical to the Markov stopping rule $\tau_t(q)$.

Applying the Markov stopping rule $\tau_t(h)$ leads to the cash flow $f_{\tau_t(h)}(X_{\tau_t(h)})$. More generally, we define, for $\mathbf{x} \in \mathfrak{X}$, any $0 \le w \le T - t$, and $h \in L_1(\mathbf{X})$ with $h_T = f_T$, the function

$$\begin{aligned}\vartheta_{t\,:\,w}(f,h)(\mathbf{x}) = {} & \sum_{s=t}^{t+w} f_s(x_s)\theta_{f,s}(h)(x_s) \prod_{r=t}^{s-1} \theta^-_{f,r}(h)(x_r) \\ & + h_{t+w}(x_{t+w}) \prod_{r=t}^{t+w} \theta^-_{f,r}(h)(x_r),\end{aligned} \tag{3.16}$$

where we follow the convention that the product over an empty index set is equal to one. The function $\vartheta_{t\,:\,w}(f,h)$ has a natural financial interpretation. It is the cash flow we would obtain by holding the American option for, at most, w periods, applying the stopping rule $\tau_t(h)$, and selling the option at time $t+w$ for the price of $h_{t+w}(X_{t+w})$, if it is not exercised before. We call $\vartheta_{t\,:\,w}(f,h)$ the cash flow function induced by h.

Equations (3.9) and (3.12) provide two different representations of q_t. In terms of $\vartheta_{t\,:\,w}(f,h)$, they can be re-expressed as follows. Because $f_{\tau^*_{t+1}}(X_{\tau^*_{t+1}}) = f_{\tau_{t+1}(q)}(X_{\tau_{t+1}(q)}) = \vartheta_{t+1\,:\,T-t-1}(f,q)$, (3.9) becomes

$$q_t(X_t) = E[\vartheta_{t+1\,:\,T-t-1}(f,q)|X_t], \tag{3.17}$$

whereas $\vartheta_{t+1\,:\,0}(f,q) = \max(f_{t+1}, q_{t+1})$ turns (3.12) into

$$q_t(X_t) = E[\vartheta_{t+1\,:\,0}(f,q)|X_t]. \tag{3.18}$$

In fact, there is a whole family of representations, parameterized by $w \in \{0, \dots, T-t-1\}$. Recursively expanding $q_{t+1}, \dots, q_{t+w}$ in (3.12) and using the Markov property, we find that

$$q_t(X_t) = E[\vartheta_{t+1\,:\,w}(f,q)|X_t], \tag{3.19}$$

for any $0 \le w \le T - t - 1$.

4. Optimal stopping as a recursive statistical learning problem. The calculation of the recursive series of nested regression problems (3.19) is becoming increasingly demanding for high-dimensional state spaces. A further complication is introduced if the transition densities of the Markov process $\mathbf{X}$ are not explicitly available. In this case, the only means to assess the distribution of the Markov process is by simulating a large number of independent sample paths $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$. These kind of problems are considered in statistical learning theory.

4.1. *Dynamic look-ahead algorithm.* Assume a payoff $f \in L_2(\mathbf{X})$. We interpret the unknown continuation value $q_t \in L_2(\mathbb{R}^m, \mu_t)$ as an approximation of the unknown optimal cash flow $\vartheta_{t+1:w}(f, q)$, in the sense that it only depends on the state of the underlying Markov process at time t. To reduce the problem further, we choose, for every $t \geq 0$, a suitable set of functions $\mathcal{H}_t$ defined on $\mathbb{R}^m$. Let

$$\mathcal{H} = \{h = (h_0, \dots, h_T) : \mathfrak{X} \to \mathbb{R}^{T+1} | h_t \in \mathcal{H}_t\}. \tag{4.1}$$

Given a finite amount of independent sample paths,

$$D_n = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}, \tag{4.2}$$

we want to find a learning rule $\hat{q}_{\mathcal{H}}$, that is, a map

$$\hat{q}_{\mathcal{H}} : D_n \mapsto \hat{q}_{\mathcal{H}}(D_n) = \big(\hat{q}_{\mathcal{H},0}(D_n), \dots, \hat{q}_{\mathcal{H},T}(D_n)\big) \in \mathcal{H}, \tag{4.3}$$

such that $\hat{q}_{\mathcal{H},t}(D_n)$ provides an accurate approximation of $\vartheta_{t+1:w}(f, q)$ in $\mathcal{H}_t$. The dynamic programming principle imposes consistency conditions on a learning rule.

DEFINITION 4.1. A learning rule $\hat{q}_{\mathcal{H}}$ is called admissible if $\hat{q}_{\mathcal{H},T}(D_n) \equiv f_T$ and $\hat{q}_{\mathcal{H},t}(D_n)$, as a function of D_n, does not depend on the sample paths up to and including time $t-1$, or, equivalently, is a function of $\{X_{i,s} | s \geq t, i = 1, \dots, n\}$ alone.

We apply empirical risk minimization to recursively define an admissible learning rule as follows. At the horizon T we set

$$\hat{q}_{\mathcal{H},T}(D_n) \equiv f_T. \tag{4.4}$$

For $t < T$, equation (3.19) suggests that we approximate the cash flow function

$$\vartheta_{t+1:w}\big(f, \hat{q}_{\mathcal{H}}(D_n)\big), \tag{4.5}$$

for some suitably selected parameter $w = w(t) \in \{0, \dots, T-t-1\}$. We choose

$$\begin{aligned} \hat{q}_{\mathcal{H},t}(D_n) &= \operatorname*{arg\,min}_{h \in \mathcal{H}_t} P_n \big| h - \vartheta_{t+1:w}\big(f, \hat{q}_{\mathcal{H}}(D_n)\big)\big|^2 \\ &= \operatorname*{arg\,min}_{h \in \mathcal{H}_t} \frac{1}{n} \sum_{i=1}^{n} \big| h(X_{i,t}) - \vartheta_{t+1:w}\big(f, \hat{q}_{\mathcal{H}}(D_n)\big)(\mathbf{X}_i)\big|^2, \end{aligned} \tag{4.6}$$

which is an element of $\mathcal{H}_t$ with minimal empirical L_2-distance from the cash flow function (4.5). Because the objective function in the optimization problem (4.6) depends solely on the functions $\hat{q}_{\mathcal{H},s}(D_n)$, $s = t+1, \dots, t+w+1$, we see by induction that the empirical risk minimization algorithm (4.6) indeed leads to an admissible learning rule.

REMARK 4.2. It is important to note that, while the function $\hat{q}_{\mathcal{H}}(D_n)$ is a function of $\mathbf{x} \in \mathcal{X}$, its choice depends on the sample D_n. Therefore, $\hat{q}_{\mathcal{H}}(D_n)$ is a random element with values in $\mathcal{H}$ which is defined on the countable product space $(\mathcal{X}^\infty, P^\infty, \mathcal{F}^\infty)$. Strictly speaking, for a sample size n, only the first n coordinates of $\mathcal{X}^\infty$ are relevant. Analogously, the expectation

$$E[\hat{q}_{\mathcal{H}}(D_n)] = \int_{\mathcal{X}} \hat{q}_{\mathcal{H}}(D_n)(\mathbf{x})\, dP(\mathbf{x}) \tag{4.7}$$

of $\hat{q}_{\mathcal{H}}(D_n)$ over the path space $\mathcal{X}$ is still a random variable on $\mathcal{X}^\infty$.

DEFINITION 4.3. The dynamic look-ahead algorithm with look-ahead parameter $w = w(t)$, $0 \le w(t) \le T - t - 1$, approximates the continuation value q_t by the empirical minimizer $\hat{q}_{\mathcal{H},t}(D_n)$ of (4.6).

The cash flow (4.5) depends on the next $w + 1$ time periods, hence, it "looks ahead" $w + 1$ periods. The algorithm is called "dynamic" because the look-ahead parameter w may be chosen time and sample dependent. We simplify our notation and drop the explicit dependency on the sample D_n, the sample size n and the look-ahead parameter w, writing $\hat{q}_{\mathcal{H},t}$ for the solution of the empirical minimization problem (4.6).

4.2. *Tsitsiklis–Van Roy and Longstaff–Schwartz algorithm.* Both the Tsitsiklis–Van Roy and the Longstaff–Schwartz algorithm are special instances of the dynamic look-ahead algorithm. The Longstaff–Schwartz algorithm is based on the cash flow function

$$\vartheta_{t+1}^{\mathrm{LS}} = f_{\tau_{t+1}(\hat{q}_{\mathcal{H}})}\big(X_{\tau_{t+1}(\hat{q}_{\mathcal{H}})}\big), \tag{4.8}$$

which corresponds to the maximal possible value $w = T - t - 1$. On the other extreme, the choice $w = 0$ in (4.5) results in the much simpler expression

$$\vartheta_{t+1}^{\mathrm{TR}} = \max(f_{t+1}, \hat{q}_{\mathcal{H},t+1}), \tag{4.9}$$

used in the Tsitsiklis–Van Roy algorithm. In its initial form, this algorithm has been developed to solve infinite horizon optimal stopping problems of ergodic Markov processes. The advantage of $\vartheta_{t+1}^{\mathrm{TS}}$ is its numerical simplicity. On the other hand, $\vartheta_{t+1}^{\mathrm{LS}}$ is better suited to approximate the optimal stopping rule because it incorporates all future time points up to the final horizon. This property is particularly important for a Markov process with slow mixing properties.

The dynamic look-ahead algorithm introduced in Definition 4.3 interpolates between the Tsitsiklis–Van Roy and the Longstaff–Schwartz algorithm. A dynamic adjustment of the look-ahead parameter $w = w(t)$ allows us to combine the algorithmic simplicity of Tsitsiklis–Van Roy and the good approximation properties of the Longstaff–Schwartz approach. For instance, we may increase $w(t)$ for the last few time steps to compensate the slow mixing of the Markov process.

5. Main results. In our definition of the dynamic look-ahead algorithm (4.6), we did not further specify the approximation scheme. The richer the set of functions $\mathcal{H}_t$, the better it can approximate the optimal cash flow. On the other hand, large sets $\mathcal{H}_t$ would require an abundance of samples to get a minimizer in (4.6) with reasonably small variance. These conflicting objectives are generally referred to as the bias-variance trade-off. To get a reasonable convergence behavior of the dynamic look-ahead algorithm, we need to impose some restrictions on the massiveness of the approximation spaces $\mathcal{H}_t$ and relate it to the number of samples which are used to calculate the minimizers in (4.6).

The massiveness of a set of functions can be measured in terms of covering and entropy numbers. The calculation of covering numbers of classes of function has a long history dating back to Kolmogorov and Tikhomirov [29] and Birman and Solomyak [8]. We refer to [13] for a modern approach and additional references. An important type of function class for which covering numbers can be estimated with combinatorial techniques are the so called Vapnik–Chervonenkis classes or VC-classes, which are, by definition, classes of functions of finite VC-dimension. Informally speaking, the VC-dimension measures the size of nonlinear sets of functions by looking at the maximum number of sign alternations of its elements. To give a precise definition, we consider a class of functions $\mathcal{G}$ defined on some set S. A set of n points $\{x_1, \dots, x_n\} \subset S$ is said to be shattered by $\mathcal{G}$ if there exists $r \in \mathbb{R}^n$ such that, for every $b \in \{0, 1\}^n$, there is a function $g \in \mathcal{G}$ such that for each i, $g(x_i) > r_i$ if $b_i = 1$, and $g(x_i) \leq r_i$ if $b_i = 0$. The VC-dimension $\mathrm{vc}(\mathcal{G})$ of $\mathcal{G}$ is defined as the cardinality of the largest set of points which can be shattered by $\mathcal{G}$. The function classes that will appear in the analysis of the fluctuations of the empirical minimizers (4.6) very well fit in the theory of Vapnik–Chervonenkis. We introduce the necessary tools of the VC-theory on the way as we prove the main results in Section 6.

Our error decomposition crucially depends on the convexity and the uniform boundedness of the class of functions $\mathcal{H}_t$. We will impose, for all $t \geq 0$, the following three conditions:

- (H_1) The class $\mathcal{H}_t$ is a closed convex subset of $L_p(\mathbb{R}^m, \mu_t)$ for some $2 \leq p \leq \infty$.
- (H_2) There exists a constant d such that the VC-dimension of $\mathcal{H}_t$ satisfies $\mathrm{vc}(\mathcal{H}_t) \leq d < \infty$.
- (H_3) The class $\mathcal{H}_t$ is uniformly bounded, that is, for some constant H, $|h_t| \leq H < \infty \ \forall h_t \in \mathcal{H}_t$.

The convexity and uniform boundedness assumptions (H_1), respectively, (H_3) are somewhat restrictive, but encompass many common approximation schemes, such as bounded convex sets in finite-dimensional linear spaces, local polynomial approximations or tensor product splines.

5.1. *Consistency and convergence.* The payoff function of an optimal stopping problem is often unbounded. For example, in option pricing even the simplest payoff functions of American put and call options increase linearly in the underlying. On the other hand, any numerical algorithm works at finite precision and tight error or convergence rate estimates rely on some sort of boundedness assumptions. We therefore introduce the truncation operator T_β, which assigns to a real valued function g the bounded function

$$T_\beta g = \begin{cases} g, & \text{if } |g| \le \beta, \\ \operatorname{sign}(g)\beta, & \text{else,} \end{cases} \tag{5.1}$$

and to $g \in L_p(\mathbf{X})$ its coordinate-wise truncation $T_\beta g = (T_\beta g_0, \dots, T_\beta g_T)$. We then replace the estimator (4.6) by

$$\hat{q}_{\mathcal{H}_n,t} = \hat{q}_{\mathcal{H}_n,t}(D_n) = \operatorname*{arg\,min}_{h \in \mathcal{H}_{n,t}} P_n \big| h - \vartheta_{t+1\,:\,w(t)}\big(T_{\beta_n} f, \hat{q}_{\mathcal{H}_n}(D_n)\big)\big|^2, \tag{5.2}$$

where $T_{\beta_n} f$ is the payoff truncated at a threshold β_n. The estimator (5.2) rests on the hypothesis that whenever $\hat{q}_{\mathcal{H}_n,s}(D_n)$ is an approximation of q_s for $s \ge t+1$, then the cash flow $\vartheta_{t+1\,:\,w}(T_{\beta_n} f, \hat{q}_{\mathcal{H}_n}(D_n))$ is a sufficiently accurate substitute for the unknown optimal cash flow $\vartheta_{t+1\,:\,w}(T_{\beta_n} f, q)$. We justify this hypothesis in Proposition 6.4 by proving a conditional Lipschitz continuity of the functional $h \mapsto \vartheta_{t+1\,:\,w}(T_{\beta_n} f, h)$ at q. The error propagation of the recursive estimation procedure is resolved in Corollary 6.2, which relies on the convexity of the approximation architecture.

The first main result provides a sufficient condition on the growth of the number of sample paths n, the VC-dimension $\operatorname{vc}(\mathcal{H}_{n,t})$ of the approximation spaces $\mathcal{H}_{n,t}$ and the truncation level β_n to ensure convergence. Let $(\mathcal{X}^\infty, P^\infty, \mathcal{F}^\infty)$ be the countable product space introduced in Remark 4.2. We use the notation $\mathbb{P} = P^\infty$ and denote by $\mathbb{E}$ the expectation with respect to $\mathbb{P}$.

THEOREM 5.1. *Assume the payoff f is in $L_2(\mathbf{X})$ and $\mathcal{H}_n$ is a sequence of approximation spaces uniformly bounded by β_n such that $\bigcup_{n=1}^\infty \mathcal{H}_n$ is dense in $L_2(\mathbf{X})$. Furthermore, assume that each $\mathcal{H}_{n,t}$ is closed, convex and $\operatorname{vc}(\mathcal{H}_{n,t}) \le d_n$. Let $\hat{q}_{\mathcal{H}_n,t}$ be the empirical L_2-minimizer from* (5.2) *for a look-ahead parameter $0 \le w(t) \le T - t - 1$. Under the assumptions*

$$\beta_n \to \infty, \qquad d_n \to \infty, \qquad \frac{d_n \beta_n^2 \log(\beta_n)}{n} \to 0 \qquad (n \to \infty), \tag{5.3}$$

it follows that

$$\|\hat{q}_{\mathcal{H}_n,t} - q_t\|_2 \to 0, \tag{5.4}$$

in probability and in $L_1(\mathbb{P})$. If, furthermore,

$$\frac{\beta_n^2 \log(n)}{n} \to 0, \tag{5.5}$$

then the convergence in (5.4) *holds almost surely.*

For the proof see Section 6.3.

Theorem 5.1 proves convergence of the truncated version (5.2) of the dynamic look-ahead algorithm. It generalizes previous results in two directions. First, the number of samples, the size of the approximation architecture (measured in terms of the VC-dimension) and the truncation threshold are increased simultaneously. Glasserman and Yu [21] address the same question for the Longstaff–Schwartz algorithm with linear finite-dimensional approximation. They avoid truncation by imposing fourth-order moment conditions and find that the number samples must grow surprisingly fast. For example, if X_t is log-normally distributed and n denotes the dimension of the linear approximation space, the number of samples must be proportional to $\exp(n^2)$. Second, Theorem 5.1 covers approximation architectures of bounded VC-dimension and does not depend on the law of the underlying Markov process. For instance, the convergence proof of Clément, Lamberton and Protter [15] relies on the additional assumption $P(q = f) = 0$.

In (5.2) we reduce unbounded to bounded payoffs by truncating at a suitable cutoff level. The next result bounds the approximation error in terms of the cutoff level.

PROPOSITION 5.2. *Let $1 \leq p < \infty$ and $f \in L_p(\mathbf{X})$ be a nonnegative payoff function. If $\bar{q}_\beta$ is the continuation value of the truncated payoff $T_\beta f$, it follows that*

$$\|q_t - \bar{q}_{\beta,t}\|_p \to 0, \tag{5.6}$$

for $\beta \to \infty$, and if $1 < r < p$, then

$$\|q_t - \bar{q}_{\beta,t}\|_r \leq \sum_{s=t+1}^{T} \left(r \int_\beta^\infty u^{r-1} P(f_{t+1} > u)\, du \right)^{1/r} \leq O\big(\beta^{(r-p)/r}\big). \tag{5.7}$$

For the proof see Section 6.5.

The bound (5.7) can be refined in terms of Orlicz norms. The Orlicz norm of a random variable Y is defined as

$$\|Y\|_\psi = \inf\{C > 0 | E[\psi(|Y|C^{-1})] \leq 1\}, \tag{5.8}$$

where ψ is a nondecreasing, convex function with $\psi(0) = 0$. Note that $\psi(y) = y^p$ reduces to the usual L_p-norms. If $\|f_{t+1}\|_\psi < \infty$, Markov's inequality implies the tail bound

$$P(f_{t+1} > u) \leq \frac{1}{\psi(u\|f_{t+1}\|_\psi^{-1})}, \tag{5.9}$$

which we then can apply to the middle term in (5.7). In particular, $\psi_p(x) = \exp(x^p) - 1$ leads to the exponential bound

$$P(f_{t+1} > u) \leq \exp\big(-u^p \|f_{t+1}\|_{\psi_p}^{-1}\big)\big(1 - \exp\big(-\beta^p \|f_{t+1}\|_{\psi_p}^{-1}\big)\big)^{-1}, \tag{5.10}$$

for all $u \geq \beta$. In financial applications a typical situation is $f_{t+1} = f(\exp(X_{t+1}))$, where X_{t+1} is normally distributed and $f(y) \leq Cy^q$ has polynomial growth. The tail estimate

$$P(f_{t+1} > u) \leq O\left(\frac{1}{\log(u)} \exp\bigl(-\log(u)^2\bigr)\right) \tag{5.11}$$

is a direct consequence of the well-known asymptotic expansion

$$1 - \Phi(u) \leq \phi(u)u^{-1}\left(1 - \frac{1}{u^2} + \frac{3}{u^4} + O(u^{-6})\right) \tag{5.12}$$

for the tail of the standard normal distribution Φ with density ϕ. (5.11) improves the rate of order $O(\beta^{1-p/r})$ in (5.7) considerably, despite the logarithmic terms in the exponent.

5.2. *Error estimate and sample complexity.* Theorem 5.1 shows that simultaneously increasing the truncation threshold, the VC-dimension of the approximation architecture and the number of samples at a proper rate, the resulting estimator (5.2) converges to the solution of the optimal stopping problem. Proposition 5.2 quantifies the error of an initial truncation at a fixed threshold. We continue the error analysis of the dynamic look-ahead algorithm by truncating unbounded payoffs at a sufficiently large threshold Θ and considering a single approximation architecture $\mathcal{H}$. The second main result bounds the overall error for bounded payoff functions in terms of the approximation error and the sample error, generalizing the familiar bias-variance trade-off in nonparametric regression and density estimation.

THEOREM 5.3. *Consider a payoff $f \in L_\infty(\mathbf{X})$ with $\|f_t\|_\infty \leq \Theta$. Assume that each $\mathcal{H}_t$ is a closed convex set of functions, uniformly bounded by H, with* $\mathrm{vc}(\mathcal{H}_t) \leq d$. *Let $\hat{q}_{\mathcal{H},t}(D_n)$ be the empirical L_2-minimizer from* (4.6) *for a look-ahead parameter* $0 \leq w(t) \leq T - t - 1$. *Set* $\beta = \max(\Theta, H)$. *Then,*

$$\begin{aligned} &\mathbb{E}[\|\hat{q}_{\mathcal{H},t}(D_n) - q_t\|_2^2] \\ &\quad \leq 2 \cdot 16^{w(t)} \max_{s=t,\dots,t+w(t)+1} \inf_{h \in \mathcal{H}_s} \|h - q_s\|_2^2 \\ &\qquad + 2 \cdot 16^{w(t)}\bigl(w(t)+2\bigr)\left(\frac{6998\beta^2 + \log(6998K\beta^2)}{n} + \frac{v\log(n)}{n}\right), \end{aligned} \tag{5.13}$$

where

$$v = 2d\bigl(c(w(t)) + 1\bigr), \qquad K = 6e^4(d+1)^2\bigl(c(w(t))d + 1\bigr)^2(1024e\beta)^v$$

and

$$c(w(t)) = 2\bigl(w(t)+2\bigr)\log_2\bigl(e(w(t)+2)\bigr).$$

For the proof see Section 6.3.

The effectiveness of a learning algorithm can be quantified by the number of samples which are required to produce with high confidence $1-\delta$ an almost minimizer

$$\|\hat{q}_{\mathcal{H},t}(D_n)-q_t\|_2^2 \le \inf_{h_t\in\mathcal{H}_t}\|h_t-q_t\|_2^2+\varepsilon \qquad \forall t=0,\dots,T-1, \tag{5.14}$$

for a certain error accuracy ε. In (5.14) the error is measured relative to the minimal approximation error at time step t. It is evident from (5.13) that an accurate estimate is only obtained if the approximation error in all previous learning tasks is small as well. To disentangle sample complexity and approximation error, we measure the performance of the learning rule relative to the overall approximation error in (5.13).

COROLLARY 5.4. *Assume $f\in L_\infty(\mathbf{X})$ with $\|f_t\|_\infty\le\Theta$ and let $\mathcal{H}$ be as in Theorem* 5.3. *The sample complexity*

$$\begin{aligned} c(\varepsilon,\delta)=\min\Big\{n_0\Big|\forall n\ge n_0,&\\ \mathbb{P}\Big(\|\hat{q}_{\hbar,t}(d_n)-q_t\|_2^2&\\ \ge 2\cdot 16^{w(t)}\max_{s=t,\dots,t+w(t)+1}\inf_{h\in\hbar_s}\|h-q_s\|_2^2+\varepsilon\Big)\le\delta\Big\}& \end{aligned} \tag{5.15}$$

of the empirical L_2-minimizer (4.6) *is bounded by*

$$c(\varepsilon,\delta)\le 2\cdot 13996(w(t)+2)16^{w(t)}\beta^2\max\left(\frac{1}{\varepsilon}\log\left(\frac{K}{\delta}\right),v\log\left(\frac{1}{\varepsilon}\right)\right), \tag{5.16}$$

where β, v and K are as in Theorem 5.3.

For the proof see Section 6.3.

Theorem 5.3 and Corollary 5.4 estimate the sample error for a fixed approximation scheme and truncation threshold. The bound (5.13) and the complexity estimate (5.16) hold uniformly for any law of $\mathbf{X}$ and payoff function f with $\|f\|_\infty\le\Theta$. Hence, the bounds are independent of the distribution of the underlying Markov process, the optimal stopping time and the smoothness of the continuation value. The asymptotic rate $O(\log(n)n^{-1})$ of the sample error [the second term on the right-hand side of (5.13)] is typical for nonparametric least square estimates with approximation schemes of finite VC-dimension, see, for example, [22], Theorem 11.5.

If we impose additional assumptions on the smoothness of the continuation value q, the approximation errors $\inf_{h\in\mathcal{H}_{n,s}}\|h-q_s\|_2^2$ in (5.13) can be estimated further by approximation theory. Smoothness assumptions are not unreasonable.

Although for many financial applications the payoff is only continuous or piecewise continuous, the continuation value is often smooth. The degree of smoothness of q is crucial for how to choose approximation spaces $\mathcal{H}_n$ to get the most favorable rate of convergence by properly balancing the approximation error and the sample error.

Smoothness is often measured in terms of Sobolev spaces $W^k(L_p(\Omega, \lambda))$, where $\Omega \subset \mathbb{R}^m$ is a domain in $\mathbb{R}^m$ and λ is the Lebesgue measure on Ω. These are functions $g \in L_p(\Omega, \lambda)$ which have all their distributional derivatives of order up to k in $L_p(\Omega, \lambda)$. The Sobolev (semi-)norm $\|g\|_{p,k,\Omega,\lambda}$ may be regarded as a measure of smoothness for a function $g \in W^k(L_p(\Omega, \lambda))$.

In practical applications of the Longstaff–Schwartz algorithm, approximation by polynomials performs rather well. Let $\mathcal{P}_r$ be the space of multivariate polynomials on $\mathbb{R}^m$ with coordinate wise degree at most $r - 1$. For simplicity, we assume X_t is localized to a sufficiently large cube $I \subset \mathbb{R}^m$. This assumption can be satisfied by applying a truncation argument similar to the one developed in Proposition 5.2.

COROLLARY 5.5. *Assume that X_t is localized to a cube $I \subset \mathbb{R}^m$, $f \in L_\infty(\mathbf{X})$, and that the continuation value q_t is in the Sobolev space $W^k(L_\infty(I, \lambda))$ for all t. Define the sequence of approximation architectures*

$$\mathcal{H}_{n,t} = \{p \in \mathcal{P}_{n^{1/(m+2k)}} \,|\, \|p\|_{\infty,I,\lambda} \leq 2\|q_t\|_{\infty,k,I,\lambda}\}. \tag{5.17}$$

Then,

$$\mathbb{E}\big[\|\hat{q}_{\mathcal{H}_{n,t}}(D_n) - q_t\|_2^2\big] \leq O\big(\log(n) n^{-2k/(2k+m)}\big). \tag{5.18}$$

If μ_t has a bounded density with respect to the Lebesgue measure and $q_t \in W^k(L_p(I, \lambda))$ for some $p \geq 2$, the same result holds if we replace $\mathcal{H}_{n,t}$ in (5.17) *by*

$$\mathcal{H}_{n,t} = \big\{p \in \mathcal{P}_{n^{1/(m+2k)}} \,|\, \|p\|_{p,I,\lambda} \leq 2\|q_t\|_{p,k,I,\lambda}\big\}. \tag{5.19}$$

PROOF. The result essentially follows from Jackson-type estimates, Theorem 6.2 in Chapter 7 of [18]. See Section 6.4. □

Corollary 5.5 is a prototypical application of Theorem 5.3 to global approximation by polynomials. Other approximation schemes can be treated similarly, as long as the conditions (H_1)–(H_3) are satisfied. To get the rate stated in Corollary 5.5, the dimension $n^{m/(m+2k)}$ of the polynomial approximation architecture (5.17) has to grow with increasing sample size, such that the approximation error and the sample error are balanced. The rate (5.18) is up to a logarithmic term the lower minimax rate of convergence for estimating regression functions; see [47].

5.3. *Discussion and remarks.* The Longstaff–Schwartz algorithm and its generalization, the dynamic look-ahead algorithm, perform surprisingly well for many practical applications, such as pricing American options which are not too far in or out of the money. This empirical observation can be explained as follows. It follows from (3.19) that an approximation of the optimal cash flow $\vartheta_{t+1:w}(f,q)$ can be used to estimate the continuation value at time t. A closer look at definition (3.16) shows that for the maximal possible value $w = T - t - 1$ the cash flow $\vartheta_{t+1:w}(f,h)$ is close (in the L_2-sense) to the optimal $\vartheta_{t+1:w}(f,q)$ if the signs of $f - h$ and $f - q$ disagree only on a subset of the path space with small probability, or, equivalently, if the probability of the symmetric difference,

$$P(\{f - h > 0\}\Delta\{f - q > 0\}), \tag{5.20}$$

is small. Note that a small probability (5.20) does not necessarily entail that the functions h and q are close in the L_2-sense. If the look-ahead parameter w satisfies $w < T - t - 1$, then $\vartheta_{t+1:w}(f,h)$ is a good approximation of the optimal cash flow if, in addition to a small probability (5.20), also the L_2-distance between h_{t+w+1} and the unknown continuation value q_{t+w+1} is small. Consequently, a look-ahead parameter $0 \le w < T - t - 1$ requires good approximations for $q_{w+1}, \dots, q_{T-1}$. Determining accurate and stable estimators for q_t with t close to 1 may be difficult to achieve, in particular, if the samples of the Markov process do not cover sufficiently large parts of the state space. This explains why the Tsitsiklis–Van Roy algorithm (corresponding to $w = 0$) may perform badly for finite horizon problems.

As opposed to the empirically demonstrated efficiency of the Longstaff–Schwartz algorithm, the results of Theorem 5.3 and Corollary 5.4 are somewhat pessimistic. For practical parameter values ε, δ, d, w and large enough cutoff level β, the sample complexity bound (5.16) leads to a very large sample size. The reason for the pessimistic sample size estimates is twofold. First, the estimator $\hat{q}_{\mathcal{H}}$ is sensitive to error propagation effects caused by the backward induction. This leads to error estimates such as (5.13) which depend exponentially on the number of look-ahead periods $w(t)$. The minimal choice $w = 0$ would resolve the exponential dependence but, as explained above, may have limited capabilities to approximate the optimal cash flow. Another reason is the generality of our error estimates. We already observed that $\hat{q}_{\mathcal{H}}$ leads to an accurate approximation of the optimal cash flow if the probability of the symmetric difference $P(\{f - \hat{q}_{\mathcal{H}} > 0\}\Delta\{f - q > 0\})$ is small. However, it is difficult to derive error estimates which take this effect into account without imposing additional assumptions on the smoothness of the payoff and the distribution of the stopping time in the neighborhood of $\{q = f\}$.

We considered in this work estimators based on straightforward empirical L_2-risk minimization. A deficiency of the simple estimator considered in Corollary 5.5 is that the degree of smoothness and an upper bound for $\|q_t\|_{\infty,k,I,\lambda}$ has to be known. There exists a variety of advanced nonparametric regression estimators which have been developed to cope with the shortcomings of the basic empirical risk minimization procedure. The main generalizations in this direction are sieve estimators studied, for example, by Shen and Wong [45], Shen [44] and Brige and Massart [10], adaptive methods such as complexity regularization, penalization and model selection; see [3, 22] and the references therein.

The benefit of conditions (H_1)–(H_3) is that convexity arguments and VC-techniques lead to error estimates without the necessity of imposing further assumptions on the Markov process $\mathbf{X}$. On the downside, some important commonly used approximation schemes are excluded. For instance, condition (H_2) conflicts with approximation in Sobolev or Besov balls, which have infinite VC-dimension, and the convexity condition (H_3) is incompatible with many interesting nonlinear approximation schemes, such as n-term approximation, wavelet thresholding or neural network architectures.

A promising approach to extend and refine the results of this work is to approximate the cash flow $\vartheta_{t:w}(f,h)$ by a suitably smoothed version with better Lipschitz continuity properties. We then can express the massiveness of the approximation schemes directly in terms of covering numbers and exploit the dependency of the covering numbers on the radius of the function class. The additional step of first bounding the VC-dimension becomes unnecessary. However, this approach is of less generality because it depends on the additional assumptions that the probability $P(\{|q-f|<\varepsilon\})$ decays to zero as $\varepsilon \to 0$ and the semi-group generated by the Markov process $\mathbf{X}$ has good smoothing properties.

Once we have selected a sequence of approximation architectures $\mathcal{H}_{n,t}$, the final step toward an implementation is to determine a computationally efficient algorithm that minimizes the empirical L_2-risk (5.2) over $\mathcal{H}_{n,t}$ in a polynomial number of time steps. Unfortunately, for many approximation spaces, such as certain neural network architectures, constructing a solution which nearly minimizes the empirical L_2-risk turns out to be NP-complete or even NP-hard. Thus, there might still exist serious complexity theoretic barriers to efficient numerical implementations of specific approximation schemes.

6. Proofs.

The proof of the main results, Theorems 5.1 and 5.3, is divided into tree steps. The strategy is as follows. First, we prove in Corollary 6.2 an error decomposition in terms of an approximation error and an expected centered loss (6.3). The second step is to estimate the covering numbers of the so called centered loss class (6.28), see Corollary 6.10. The last step is to apply empirical process techniques to bound the fluctuation of the expected centered loss in terms of the covering numbers.

6.1. *Error decomposition.*

We assume from now on without further mentioning that $\mathcal{H} \subset L_2(\mathbf{X})$ and that all approximation spaces $\mathcal{H}_t$ are closed and convex.

Before we can state our main error decomposition we need to introduce some more notation. Let

$$\pi_{\mathcal{H}_t} : L_2(\mathbb{R}^m, \mu_t) \to \mathcal{H}_t \tag{6.1}$$

denote the projection onto the closed convex subset $\mathcal{H}_t \subset L_2(\mathbb{R}^m, \mu_t)$ and set

$$\mathrm{pr}_{\mathcal{H}_t} = \pi_{\mathcal{H}_t} \circ E[\cdot | X_t = \cdot] : L_2(\boldsymbol{\mathcal{X}}, P) \to \mathcal{H}_t. \tag{6.2}$$

For any $h = (h_0, \dots, h_T) : \boldsymbol{\mathcal{X}} \to \mathbb{R}^{T+1}$ with $h_T = f_T$, we introduce the centered loss

$$l_t(h) = |h_t - \vartheta_{t+1:w}(f, h)|^2 - \left|\mathrm{pr}_{\mathcal{H}_t} \vartheta_{t+1:w}(f, h) - \vartheta_{t+1:w}(f, h)\right|^2. \tag{6.3}$$

In favor of a more compact notation, we have dropped the dependency of $l_t(h)$ on the look-ahead parameter w. Note that the centered loss $l_t(h)$ only depends on $h_t, \dots, h_{T-1}$ and can take on negative values. However, $E[l_t(h)] \geq 0$, as we will see in Lemma 6.3.

We decompose the overall error into an approximation error, a sample error and a third term which captures the error propagation caused by the recursive definition of the dynamic look-ahead estimator.

PROPOSITION 6.1. *Assume that $\hat{q}_{\mathcal{H}}$ is the result of an admissible learning rule. Then*

$$\|\hat{q}_{\mathcal{H},t} - q_t\|_2 \leq \inf_{h \in \mathcal{H}_t} \|h - q_t\|_2 + E[l_t(\hat{q}_{\mathcal{H}})]^{1/2} + 3 \sum_{s=t+1}^{t+w+1} \|\hat{q}_{\mathcal{H},s} - q_s\|_2. \tag{6.4}$$

In general, we cannot approximate $\vartheta_{t+1:w}(f, \hat{q}_{\mathcal{H}})$ by functions $h_t \in L_2(\mathbb{R}^d, \mu_t)$ arbitrarily well and, therefore,

$$\inf_{h_t \in \mathcal{H}_t} E[|h_t - \vartheta_{t+1:w}(f, \hat{q}_{\mathcal{H}})|^2] > 0. \tag{6.5}$$

For this reason we base our error decomposition (6.4) on the more complicated centered loss function, which expresses the sample error relative to the optimal one-step expected loss

$$E\left[\left|\mathrm{pr}_{\mathcal{H}_t} \vartheta_{t+1:w}(f, \hat{q}_{\mathcal{H}}) - \vartheta_{t+1:w}(f, \hat{q}_{\mathcal{H}})\right|^2\right]. \tag{6.6}$$

The first term on the right-hand side of (6.4) is the approximation error, a deterministic quantity, which can be analyzed by approximation theory. The second term $E[l_t(\hat{q}_{\mathcal{H}})]^{1/2}$ is usually referred to as the sample error. The last term in (6.4) collects the error propagation introduced by the previous learning tasks through the dynamic programming backward recursion.

COROLLARY 6.2. *Let*

$$e_t = \inf_{h \in \mathcal{H}_t} \|h - q_t\|_2 + E[l_t(\hat{q}_{\mathcal{H}})]^{1/2} \tag{6.7}$$

denote the one-step error. Then,

$$\|\hat{q}_{\mathcal{H},t} - q_t\|_2 \le e_t + 3 \sum_{s=t+1}^{t+w+1} 4^{s-t-1} e_s, \tag{6.8}$$

and

$$\|\hat{q}_{\mathcal{H},t} - q_t\|_2 \le 4^{w+1} \max_{s=t,\dots,t+w+1} \Big(\inf_{h \in \mathcal{H}_s} \|h - q_s\|_2 + E[l_s(\hat{q}_{\mathcal{H}})]^{1/2} \Big). \tag{6.9}$$

PROOF. This follows at once from (6.4) by recursively inserting the error estimate (6.4) for $s \ge t+1$. $\square$

The proof of the error decomposition (6.4) crucially relies on the convexity of the approximation spaces, Lemma 6.3 and a Lipschitz estimate for $\vartheta_{t+1:w}(f,h)$ as a function of h, Proposition 6.4.

LEMMA 6.3. *Denote by*

$$\rho_t(h)(x) = E[\vartheta_{t+1:w}(f,h) | X_t = x] \tag{6.10}$$

the regression function of $\vartheta_{t+1:w}(f,h)$. For any $h \in \mathcal{H}$ with $h_T = f_T$,

$$\|h_t - \pi_{\mathcal{H}_t} \rho_t(h)\|_2^2 = \|h_t - \mathrm{pr}_{\mathcal{H}_t} \vartheta_{t+1:w}(f,h)\|_2^2 \le E[l_t(h)]. \tag{6.11}$$

In particular, $E[l_t(h)] \ge 0$.

PROOF. The proof is identical to the proof of Lemma 5 in [17]. Because $\rho_t(h)$ is the regression function of $\vartheta_{t+1:w}(f,h)$, which only depends on $h_{t+1}, \dots, h_{T-1}$, we have, for all $h_t \in L_2(\mathbb{R}^d, \mu_t)$,

$$\|h_t - \rho_t(h)\|_2^2 = E[|h_t - \vartheta_{t+1:w}(f,h)|^2 - |\rho_t(h) - \vartheta_{t+1:w}(f,h)|^2]. \tag{6.12}$$

Let $h \in \mathcal{H}$ be arbitrary. Since $\mathcal{H}_t$ is convex and since $\mathrm{pr}_{\mathcal{H}_t} \vartheta_{t+1:w}(f,h) = \pi_{\mathcal{H}_t} \rho_t(h)$ minimizes the distance to $\rho_t(h)$, it follows that

$$\langle \rho_t(h) - \pi_{\mathcal{H}_t} \rho_t(h), h_t - \pi_{\mathcal{H}_t} \rho_t(h) \rangle \le 0. \tag{6.13}$$

Therefore,

$$\begin{aligned} \|\mathrm{pr}_{\mathcal{H}_t} \vartheta_{t+1:w}(f,h) - h_t\|_2^2 &= \|\pi_{\mathcal{H}_t} \rho_t(h) - h_t\|_2^2 \\ &\le \|\rho_t(h) - h_t\|_2^2 - \|\rho_t(h) - \pi_{\mathcal{H}_t} \rho_t(h)\|_2^2. \end{aligned} \tag{6.14}$$

Because both h_t and $\pi_{\mathcal{H}_t}\rho_t(h)$ are in $\mathcal{H}_t$ we can apply (6.12) twice, which shows that the right-hand side of (6.14) is equal to

$$\begin{aligned}\|\rho_t(h)-h_t\|_2^2 - \|\rho_t(h)-\pi_{\mathcal{H}_t}\rho_t(h)\|_2^2 \\ = E\big[|h_t-\vartheta_{t+1:w}(f,h)|^2 - |\pi_{\mathcal{H}_t}\rho_t(h)-\vartheta_{t+1:w}(f,h)|^2\big].\end{aligned} \tag{6.15}$$

□

For $w=0$, we immediately obtain from $|\max(a,x)-\max(a,y)|\le|x-y|$ and Jensen's inequality the uniform Lipschitz bound

$$\|E[\vartheta_{t+1:0}(f,g)-\vartheta_{t+1:0}(f,h)|X_t\,]\|_p \le \|g_{t+1}-h_{t+1}\|_p. \tag{6.16}$$

More generally, we have the following conditional Lipschitz continuity at the continuation value.

PROPOSITION 6.4. *For every $h\in L_p(\mathbf{X})$ with $h_T=f_T$ and $0\le w\le T-t$,*

$$\begin{aligned}\|E[\vartheta_{t+1:w}(f,h)|X_t]-q_t\|_p \\ = \|E[\vartheta_{t+1:w}(f,h)-\max(f_{t+1},q_{t+1})|X_t]\|_p \\ = \|E[\vartheta_{t+1:w}(f,h)-\vartheta_{t+1:w}(f,q)|X_t]\|_p.\end{aligned} \tag{6.17}$$

Furthermore,

$$\|E[\vartheta_{t+1:w}(f,h)-\vartheta_{t+1:w}(f,q)|X_t]\|_p \le \sum_{s=t+1}^{t+w+1}\|h_s-q_s\|_p. \tag{6.18}$$

A similar estimate for the special case $w=T-t-1$ can also be found in [15]. Note that the uniform Lipschitz estimate (6.16) does not extend to $w>0$. Proposition 6.4 only provides a Lipschitz estimate at the continuation value.

PROOF. First note that, from the Markov property,

$$E[\vartheta_{t+1:w}(q)-\vartheta_{t+1:w}(h)|X_t] = E[\vartheta_{t+1:w}(q)-\vartheta_{t+1:w}(h)|\mathcal{F}_t]. \tag{6.19}$$

Equation (6.17) follows directly from the recursive definition of q_t. The case $w=0$ is covered in (6.16). For $w>0$, it follows from the definition of $\vartheta_{t+1:w}$ that

$$\begin{aligned}&\|E[\vartheta_{t+1:w}(q)-\vartheta_{t+1:w}(h)|\mathcal{F}_t]\|_p \\ &\quad\le \big\|E\big[f_{t+1}\big(\theta_{f,t+1}(q)-\theta_{f,t+1}(h)\big) \\ &\qquad + \theta^-_{f,t+1}(q)\vartheta_{t+2:w-1}(q)-\theta^-_{f,t+1}(h)\vartheta_{t+2:w-1}(h)|\mathcal{F}_t\big]\big\|_p.\end{aligned}$$

Adding and subtracting the term $q_{t+1}(\theta_{f,t+1}(q)-\theta_{f,t+1}(h))$, the triangle inequality implies

$$
\begin{aligned}
&\|E[\vartheta_{t+1:w}(q)-\vartheta_{t+1:w}(h)|\mathcal{F}_t]\|_p \\
&\qquad \le \big\|E\big[(f_{t+1}-q_{t+1})\big(\theta_{f,t+1}(q)-\theta_{f,t+1}(h)\big)|\mathcal{F}_t\big]\big\|_p \\
&\qquad\quad + \big\|E\big[\theta^-_{f,t+1}(q)\vartheta_{t+2:w-1}(q) \\
&\qquad\qquad\quad - \theta^-_{f,t+1}(h)\vartheta_{t+2:w-1}(h) + q_{t+1}\big(\theta_{f,t+1}(q)-\theta_{f,t+1}(h)\big)|\mathcal{F}_t\big]\big\|_p.
\end{aligned}
$$

Now

$$
\begin{aligned}
\theta_{f,t+1}(q)-\theta_{f,t+1}(h) &= \mathbb{1}_{\{f_{t+1}\ge q_{t+1}\}} - \mathbb{1}_{\{f_{t+1}\ge h_{t+1}\}} \\
&= \mathbb{1}_{\{0\le f_{t+1}-q_{t+1}<h_{t+1}-q_{t+1}\}} - \mathbb{1}_{\{h_{t+1}-q_{t+1}\le f_{t+1}-q_{t+1}<0\}},
\end{aligned}
$$

which leads to

$$
\begin{aligned}
&(f_{t+1}-q_{t+1})\big(\mathbb{1}_{\{0\le f_{t+1}-q_{t+1}<h_{t+1}-q_{t+1}\}} - \mathbb{1}_{\{h_{t+1}-q_{t+1}\le f_{t+1}-q_{t+1}<0\}}\big) \\
&\qquad \le (h_{t+1}-q_{t+1})\mathbb{1}_{\{h_{t+1}-q_{t+1}>0\}} - (h_{t+1}-q_{t+1})\mathbb{1}_{\{h_{t+1}-q_{t+1}<0\}} \\
&\qquad \le |h_{t+1}-q_{t+1}|.
\end{aligned}
$$

By the Markov property, $q_{t+1}(X_{t+1}) = E[\vartheta_{t+2:w-1}(q)|\mathcal{F}_{t+1}]$. Because $\theta_{f,t+1}(q)$ and $\theta_{f,t+1}(h)$ are $\sigma(X_{t+1})$-measurable, it follows that

$$
\begin{aligned}
&E\big[q_{t+1}\big(\theta_{f,t+1}(q)-\theta_{f,t+1}(h)\big)|\mathcal{F}_t\big] \\
&\qquad = E\big[E[\vartheta_{t+2:w-1}(q)|\mathcal{F}_{t+1}]\big(\theta_{f,t+1}(q)-\theta_{f,t+1}(h)\big)|\mathcal{F}_t\big] \\
&\qquad = E\big[\vartheta_{t+2:w-1}(q)\big(\theta_{f,t+1}(q)-\theta_{f,t+1}(h)\big)|\mathcal{F}_t\big].
\end{aligned}
\tag{6.20}
$$

By Jensen's inequality, this leads to

$$
\begin{aligned}
&\|E[\vartheta_{t+1:w}(q)-\vartheta_{t+1:w}(h)|\mathcal{F}_t]\|_p \\
&\qquad \le \|q_{t+1}-h_{t+1}\|_p \\
&\qquad\quad + \big\|E\big[\vartheta_{t+2:w-1}(q)\big(1-\theta_{f,t+1}(h)\big) - \vartheta_{t+2:w-1}(h)\theta^-_{f,t+1}(h)|\mathcal{F}_t\big]\big\|_p \\
&\qquad = \|q_{t+1}-h_{t+1}\|_p + \big\|E\big[\big(\vartheta_{t+2:w-1}(q)-\vartheta_{t+2:w-1}(h)\big)\theta^-_{f,t+1}(h)|\mathcal{F}_t\big]\big\|_p \\
&\qquad \le \|q_{t+1}-h_{t+1}\|_p + \|E[\vartheta_{t+2:w-1}(q)-\vartheta_{t+2:w-1}(h)|\mathcal{F}_{t+1}]\|_p.
\end{aligned}
$$

The proof is completed by induction. □

PROOF OF PROPOSITION 6.1. Introduce the regression function

$$
\bar{\rho}_{\mathcal{H},t}(x) = E[\vartheta_{t+1:w}(f,\hat{q}_{\mathcal{H}})|X_t = x]
\tag{6.21}
$$

of $\vartheta_{t+1:w}(f,\hat{q}_{\mathcal{H}})$ and let

$$
\bar{q}_{\mathcal{H},t} = \pi_{\mathcal{H}_t}\bar{\rho}_{\mathcal{H},t} = \mathrm{pr}_{\mathcal{H}_t}\,\vartheta_{t+1:w}(f,\hat{q}_{\mathcal{H}})
\tag{6.22}
$$

be its projection onto $\mathcal{H}_t$. By the triangle inequality,

$$\|\hat{q}_{\mathcal{H},t} - q_t\|_2 \le \|\hat{q}_{\mathcal{H},t} - \bar{q}_{\mathcal{H},t}\|_2 + \|\bar{q}_{\mathcal{H},t} - \bar{\rho}_{\mathcal{H},t}\|_2 + \|\bar{\rho}_{\mathcal{H},t} - q_t\|_2. \tag{6.23}$$

Again, by the triangle inequality and because $\mathcal{H}_t$ is convex so that the projection $\pi_{\mathcal{H}_t}$ from $L_2(\mathbb{R}^m, \mu_t)$ onto $\mathcal{H}_t$ is distance decreasing,

$$\begin{aligned}\|\bar{q}_{\mathcal{H},t} - \bar{\rho}_{\mathcal{H},t}\|_2 &= \big\|\pi_{\mathcal{H}_t}\bar{\rho}_{\mathcal{H},t} - \bar{\rho}_{\mathcal{H},t}\big\|_2 \\ &\le \big\|\pi_{\mathcal{H}_t}\bar{\rho}_{\mathcal{H},t} - \pi_{\mathcal{H}_t}q_t\big\|_2 + \big\|\pi_{\mathcal{H}_t}q_t - q_t\big\|_2 + \|q_t - \bar{\rho}_{\mathcal{H},t}\|_2 \\ &\le \big\|\pi_{\mathcal{H}_t}q_t - q_t\big\|_2 + 2\|q_t - \bar{\rho}_{\mathcal{H},t}\|_2.\end{aligned} \tag{6.24}$$

Inserting (6.24) back into (6.23) gives

$$\|\hat{q}_{\mathcal{H},t} - q_t\|_2 \le \inf_{h \in \mathcal{H}_t} \|h - q_t\|_2 + \|\hat{q}_{\mathcal{H},t} - \bar{q}_{\mathcal{H},t}\|_2 + 3\|\bar{\rho}_{\mathcal{H},t} - q_t\|_2. \tag{6.25}$$

By Lemma 6.3,

$$\|\hat{q}_{\mathcal{H},t} - \bar{q}_{\mathcal{H},t}\|_2 = \big\|\hat{q}_{\mathcal{H},t} - \pi_{\mathcal{H}_t}\bar{\rho}_{\mathcal{H},t}\big\|_2 \le E[l_t(\hat{q}_{\mathcal{H}})]^{1/2}. \tag{6.26}$$

For the third term in (6.25), by Proposition 6.4,

$$\begin{aligned}\|\bar{\rho}_{\mathcal{H},t} - q_t\|_2 &= \|E[\vartheta_{t+1:w}(f, \hat{q}_{\mathcal{H}}) - \vartheta_{t+1:w}(f, q)|X_t]\|_2 \\ &\le \sum_{s=t+1}^{t+w+1} \|\hat{q}_{\mathcal{H},s} - q_s\|_2.\end{aligned} \tag{6.27}$$

□

6.2. *Covering number bounds.* We define the so-called centered loss class

$$\mathcal{L}_t(\mathcal{H}) = \{l_t(h) | h \in \mathcal{H}\}. \tag{6.28}$$

To bound the fluctuations of the sample error $E[l_t(\hat{q}_{\mathcal{H}})]^{1/2}$ later on in Section 6.3, we require bounds on the empirical L_1-covering numbers $N(\varepsilon, \mathcal{L}_t(\mathcal{H}), d_{1,P_n})$ of the centered loss class.

The first step is to bound the covering numbers of $\mathcal{L}_t(\mathcal{H})$ in terms of the covering numbers of $\mathcal{H}_t$ and the cash flow class which is defined as

$$\mathcal{G}_t = \{\vartheta_{t+1:w}(f, h) | h \in \mathcal{H}\}. \tag{6.29}$$

LEMMA 6.5. *Let $1 \le p \le \infty$. If $\mathcal{H}_t$ is uniformly bounded by H and the cash flow class $\mathcal{G}_t$ by Θ, then, for $w \ge 0$,*

$$N\big(8(H + \Theta)\varepsilon, \mathcal{L}_t(\mathcal{H}), d_{p,P_n}\big) \le N\big(\varepsilon, \mathcal{H}_t, d_{p,P_n}\big)^2 N\big(\varepsilon, \mathcal{G}_t, d_{p,P_n}\big)^2. \tag{6.30}$$

For $w = 0$, the estimate (6.30) *simplifies to*

$$N\big(8(H + \Theta)\varepsilon, \mathcal{L}_t(\mathcal{H}), d_{p,P_n}\big) \le N\big(\varepsilon, \mathcal{H}_t, d_{p,P_n}\big)^2 N\big(\varepsilon, \mathcal{H}_{t+1}, d_{p,P_n}\big)^2. \tag{6.31}$$

Note that if the payoff function f is in $L_\infty(\mathbf{X})$ and the approximation spaces $\mathcal{H}_t$ are uniformly bounded by H, then $\vartheta_{t+1:w}(f,h) \le \Theta \equiv \max(\|f\|_\infty, H)$ and the assumptions of Lemma 6.5 are satisfied.

PROOF OF LEMMA 6.5. We first recall some basic properties of covering numbers. If $\mathcal{F}$ and $\mathcal{G}$ are two classes of functions and $\mathcal{F} \pm \mathcal{G} = \{f \pm g | f \in \mathcal{F}, g \in \mathcal{G}\}$ is the class of formal sums or differences, then, for all $1 \le p \le \infty$,

$$N(\varepsilon, \mathcal{F} \pm \mathcal{G}, d_{p,P_n}) \le N\left(\frac{\varepsilon}{2}, \mathcal{F}, d_{p,P_n}\right) N\left(\frac{\varepsilon}{2}, \mathcal{G}, d_{p,P_n}\right). \tag{6.32}$$

Furthermore, if $\mathcal{G}$ class of functions uniformly bounded by G, it follows from $\|g_1^2 - g_2^2\|_{p,P_n}^p = P_n(g_1 - g_2)^p(g_1+g_2)^p \le (2G)^p \|g_1 - g_2\|_{p,P_n}^p$ that

$$N(\varepsilon, \mathcal{G}^2, d_{p,P_n}) \le N\left(\frac{\varepsilon}{2G}, \mathcal{G}, d_{p,P_n}\right), \tag{6.33}$$

Enlarging a class increases the covering numbers. Now

$$\mathcal{L}_t(\mathcal{H}) \subset (\mathcal{H}_t - \mathcal{G}_t)^2 - (\operatorname{pr}_{\mathcal{H}_t} \mathcal{G}_t - \mathcal{G}_t)^2. \tag{6.34}$$

Because $\operatorname{pr}_{\mathcal{H}_t} \mathcal{G}_t \subset \mathcal{H}_t$, it is sufficient to bound the covering number of the slightly larger class

$$\tilde{\mathcal{L}}_t(\mathcal{H}) = (\mathcal{H}_t - \mathcal{G}_t)^2 - (\mathcal{H}_t - \mathcal{G}_t)^2. \tag{6.35}$$

If $\mathcal{H}_t$ is uniformly bounded by $H < \infty$ and $\vartheta_{t+1,w}(f,h) \le \Theta$, we get from (6.32) and (6.33)

$$\begin{aligned} &N(\varepsilon, \tilde{\mathcal{L}}_t(\mathcal{H}), d_{p,P_n}) \\ &\quad \le N\left(\frac{\varepsilon}{8(H+\Theta)}, \mathcal{H}_t, d_{p,P_n}\right)^2 N\left(\frac{\varepsilon}{8(H+\Theta)}, \mathcal{G}_t, d_{p,P_n}\right)^2. \end{aligned} \tag{6.36}$$

For $w = 0$, the Lipschitz bound (6.16) directly leads to

$$N(\varepsilon, \mathcal{G}_t, d_{p,P_n}) \le N(\varepsilon, \mathcal{H}_{t+1}, d_{p,P_n}). \tag{6.37}$$

(6.31) follows directly from (6.36) and (6.37). □

A simple example for which tight covering number bounds exists are subsets of linear vector spaces. If $\mathcal{H}_t = \{h \in \mathcal{K} | \|h\|_\infty \le R\}$ and $\mathcal{K}$ is a linear vector space of dimension d, then

$$N(\varepsilon, \mathcal{H}_t, d_{2,P_n}) \le N(\varepsilon, \{h \in \mathcal{K} | P_n h^2 \le R^2\}, d_{2,P_n}) \le \left(\frac{4R+\varepsilon}{\varepsilon}\right)^d. \tag{6.38}$$

The first inequality in (6.38) is obvious because $\mathcal{H}_t$ is a subset of $\{h \in \mathcal{K} | P_n h^2 \le R^2\}$. The second inequality is standard and can be found, for instance, in [13] or [51].

Inequality (6.38) would provide uniform covering number estimates for (6.31) in case of linear approximation spaces and $w = 0$. We can not apply (6.38) to upper bound the right-hand side of (6.30) in the general situation $w > 0$ because the cash flow class $\mathcal{G}_t$ is not anymore a subset of a linear space, even if the underlying approximation space $\mathcal{H}_t$ is a finite-dimensional linear vector space. This is where the Vapnik–Chervonenkis theory comes into play.

An important type of function class for which good uniform estimates on the covering numbers exist without assuming any linear structure are the so called Vapnik–Chervonenkis classes or VC-classes, introduced in [55] for classes of indicator functions, that is, classes of sets. Let $\mathcal{C}$ be a class of subsets of a set S. We say that the class $\mathcal{C}$ picks out a subset A of a set $\sigma_n = \{x_1, \dots, x_n\} \subset S$ of n elements if $A = C \cap \sigma_n$ for some $C \in \mathcal{C}$. The class $\mathcal{C}$ is said to shatter σ_n if each of its 2^n subset can be picked out by $\mathcal{C}$. The VC-dimension of $\mathcal{C}$ is the largest integer n such that there exists a set of n points which can be shattered by $\mathcal{C}$, that is,

$$\mathrm{vc}(\mathcal{C}) = \sup\{n \,|\, \Delta_n(\mathcal{C}) = 2^n\}, \tag{6.39}$$

where

$$\Delta_n(\mathcal{C}) = \max_{\{x_1,\dots,x_n\}} \operatorname{card}\big\{C \cap \{x_1, \dots, x_n\} | C \in \mathcal{C}\big\} \tag{6.40}$$

is the so-called growth or shattering function. A class $\mathcal{C}$ is called a Vapnik–Chervonenkis or VC-class if $\mathrm{vc}(\mathcal{C}) < \infty$. A VC-class of dimension d shatters no set of $d + 1$ points. The "richer" the class $\mathcal{C}$ is, the larger the cardinality of sets which still can be shattered. We illustrate it by a simple example. The class of left open intervals $\{(-\infty, c] | c \in \mathbb{R}\}$ cannot shatter any two-point set because it cannot pick out the largest of the two points and therefore has VC-dimension one. By similar reasoning, the class of intervals $\{(-a, b] | a, b \in \mathbb{R}\}$ shatters two-point sets, but fails to shatter three-point sets: it cannot pick out the largest and the smallest point of a three-point set. On the contrary, the collection of closed convex subsets of $\mathbb{R}^2$ has infinite VC-dimension: Consider a set σ_n of n points on the unit circle. Every subset $A \subset \sigma_n$ of the 2^n subsets can be picked out by the closed convex hull $\overline{\mathrm{co}}(A)$ of A. A peculiar property of a VC-class is that the shattering function of VC-classes grows only polynomially in n, more precisely, we have the following result which is due to Sauer, Vapnik–Chervonenkis and Shelah; see [51], Corollary 2.6.3, or [19].

LEMMA 6.6 (Sauer's lemma). *If $\mathcal{C}$ is a VC-class with VC-dimension $d = \mathrm{vc}(\mathcal{C})$, then*

$$\Delta_n(\mathcal{C}) \le \sum_{i=0}^{d} \binom{n}{i} \le 1.5 \frac{n^d}{d!} \le \left(\frac{en}{d}\right)^d. \tag{6.41}$$

VC-classes have a variety of permanence properties which allow the construction of new VC-classes from basic VC-classes by simple operations such as complements, intersections, unions or products. We again refer to [51], Section 2.6.5, or [19].

The concept of VC-classes of sets can be extended to classes of functions in several ways. A common approach is to associate to a class of functions its subgraph class. More precisely, the subgraph of a real-valued function g on an arbitrary set S is defined as

$$\mathrm{Gr}(g) = \{(x,t) \in S \times \mathbb{R} | t \leq g(x)\}. \tag{6.42}$$

A class of real-valued functions $\mathcal{G}$ on S is called a VC-subgraph class, or just VC-class, if its class of subgraphs is a VC-class and the VC-dimension of $\mathcal{G}$ is defined as

$$\mathrm{vc}(\mathcal{G}) = \mathrm{vc}\big(\{\mathrm{Gr}(g) | g \in \mathcal{G}\}\big). \tag{6.43}$$

An equivalent definition is obtained by extending the notion of shattering. A class of real-valued functions $\mathcal{G}$ is said to shatter a set $\{x_1, \dots, x_n\} \subset S$ if there is $r \in \mathbb{R}^n$ such that for every $b \in \{0,1\}^n$, there is a function $g \in \mathcal{G}$ such that for each i, $g(x_i) > r_i$ if $b_i = 1$, and $g(x_i) \leq r_i$ if $b_i = 0$. The definition

$$\mathrm{vc}(\mathcal{G}) = \sup\{n | \exists\, \{x_1, \dots, x_n\} \subset S \text{ shattered by } \mathcal{G}\} \tag{6.44}$$

agrees with (6.43). For the proof note that a set is shattered by the subgraph class $\{\mathrm{Gr}(g) | g \in \mathcal{G}\}$ if and only if it is shattered by the class of indicator functions $\{\theta(g(x) - t) | g \in \mathcal{G}\}$, where $\theta(s) = \mathbb{1}_{\{s \geq 0\}}$. The VC-dimension (6.44) for classes of functions is often called pseudo-dimension, see [23] and [41]. An alternative generalization is obtained by so called VC-major classes, originally introduced by Vapnik. For more details on the relation of the two concepts, we refer to [19].

LEMMA 6.7. *Let $\mathcal{G}$ be a finite-dimensional real vector space of measurable real-valued functions. Then, the class of sets $\mathcal{G}^+ = \{\{g \geq 0\} | g \in \mathcal{G}\}$ is a VC-class with* $\mathrm{vc}(\mathcal{G}^+) \leq \dim(\mathcal{G})$. *If g_0 is a fixed function, then* $\mathrm{vc}((g_0 + \mathcal{G})^+) = \mathrm{vc}(\mathcal{G}^+)$. *Finally, $\mathcal{G}$ is a VC-class and* $\mathrm{vc}(\mathcal{G}) = \dim(\mathcal{G})$.

PROOF. For the first two statements we refer to [19], Theorem 4.2.1, or [51], Section 2.6. The last statement follows from the first two: Let $g_0(x,t) = -t$ and consider the affine class of functions $g_0 + \mathcal{G}$ on $S \times \mathbb{R}$. Then, the subgraph class of $\mathcal{G}$ is precisely $(g_0 + \mathcal{G})^+$. $\square$

An important property of VC-classes is that their covering numbers $N(\varepsilon, \mathcal{G}, d_{p,\mu})$ are polynomial in ε^{-1} for $\varepsilon \to 0$. More precisely, we have the following estimates for the covering numbers of VC-classes due to Haussler [23]; see also [51], Theorem 2.6.7.

LEMMA 6.8. *Let $\mathcal{G} \subset L_p(\mu)$ be a class of functions with an envelope $G \in L_p(\mu)$, that is, $g \le G$ for all $g \in \mathcal{G}$. Then,*

$$N(\varepsilon \|G\|_{p,\mu}, \mathcal{G}, d_{p,\mu}) \le e\big(\mathrm{vc}(\mathcal{G}) + 1\big) 2^{\mathrm{vc}(\mathcal{G})} \left(\frac{2e}{\varepsilon}\right)^{p\,\mathrm{vc}(\mathcal{G})}. \tag{6.45}$$

After this short digression on VC-theory, we continue estimating the empirical L_1-covering numbers of the centered loss class $\mathcal{L}_t(\mathcal{H})$. The next result is fundamental to generalize the estimate (6.31) to a strictly positive look-ahead parameter $w > 0$. It bounds the VC-dimension of $\mathcal{G}_t$ in terms of the VC-dimension of the approximation spaces $\mathcal{H}_{t+1}, \ldots, \mathcal{H}_{t+w+1}$.

PROPOSITION 6.9. *Assume that, for all $s \ge t$, $\mathcal{H}_s$ are VC-classes of functions with $\mathrm{vc}(\mathcal{H}_s) \le d$. Then $\mathcal{G}_t$ is a VC-class with VC-dimension*

$$\mathrm{vc}(\mathcal{G}_t) \le c(w) d, \tag{6.46}$$

where $c(w) = 2(w+2)\log_2(e(w+2))$.

Inequalities (6.30), (6.31), (6.45) and (6.46) finally lead to explicit uniform bounds for the empirical L_1-covering numbers of the centered loss class $\mathcal{L}_t(\mathcal{H})$.

COROLLARY 6.10. *Assume that all $\mathcal{H}_s$ are classes of functions uniformly bounded by H and with bounded VC-dimension $\mathrm{vc}(\mathcal{H}_s) \le d$. If the cash flow function satisfies $\vartheta_{t+1:w}(f, h) \le H$, then*

$$\begin{aligned} &N\big(\varepsilon, \mathcal{L}_t(\mathcal{H}), d_{1,P_n}\big) \\ &\quad \le \begin{cases} e^4 (d+1)^2 \big(c(w)d + 1\big)^2 \left(\dfrac{64eH}{\varepsilon}\right)^{2d(c(w)+1)}, & \text{for } w \ge 1, \\ e^4 (d+1)^4 \left(\dfrac{64eH}{\varepsilon}\right)^{4d}, & \text{for } w = 0. \end{cases} \end{aligned} \tag{6.47}$$

Optimal stopping is a particular stochastic control problem with a simple control space. The proof of Proposition 6.9 relies on the observation that the VC-dimension of the class of indicator functions $\mathcal{C}_s = \{\theta_{f,s}(h) | h_s \in \mathcal{H}_s\}$, which appear in the definition of $\tau_t(h)$ and $\vartheta_{t+1:w}(h)$, is bounded by $\mathrm{vc}(\mathcal{H}_s)$. It is an interesting question how Proposition 6.9 can be extended to more general stochastic control problems.

Before we proceed to the proof of Proposition 6.9, we add a remark on VC-classes and their VC-dimension. Let $\mathcal{A}$ be a class of sets. The class of indicator functions $\{\mathbb{1}_A | A \in \mathcal{A}\}$ is a VC-class in the sense that its subgraph class is a VC-class if and only if $\mathcal{A}$ is a VC-class and $\mathrm{vc}(\mathcal{A}) = \mathrm{vc}(\{\mathbb{1}_A | A \in \mathcal{A}\})$. Let

$\theta(x) = \mathbb{1}_{\{x \geq 0\}}$. If $\mathcal{A}$ is a VC-class, $\mathrm{vc}(\mathcal{A}) = d$, then by Sauer's Lemma 6.6, for $x_1, \dots, x_n$ and all $t \in \mathbb{R}^n$,

$$\operatorname{card}\{(\theta(\mathbb{1}_A(x_i) - t_i))_{i=1,\dots,n} | A \in \mathcal{A}\} \leq \left(\frac{en}{d}\right)^d. \tag{6.48}$$

Conversely, if we find a polynomial bound like (6.48), $\mathcal{A}$ must be a VC-class and we can bound its VC-dimension.

To prove Proposition 6.9, we first establish the following general result on VC-classes.

LEMMA 6.11. *Let $\mathcal{X}$, $\mathcal{Y}$ be two sets and $\mathcal{A}$, $\mathcal{B}$ VC-classes of subsets of $\mathcal{X}$ (resp. $\mathcal{Y}$). Assume that* $\mathrm{vc}(\mathcal{A}) \leq d$, $\mathrm{vc}(\mathcal{B}) \leq d$. *Let $f : \mathcal{X} \to \mathbb{R}$ and $g : \mathcal{Y} \to \mathbb{R}$ be nonnegative functions. Define the class of functions*

$$\begin{aligned} &\mathcal{F}(\mathcal{A}, \mathcal{B}) \\ &\quad = \{F_{A,B}(x, y) = \mathbb{1}_A(x) f(x) + \mathbb{1}_{A^c}(x)\mathbb{1}_B(y)g(y) | A \in \mathcal{A}, B \in \mathcal{B}\}. \end{aligned} \tag{6.49}$$

Then $\mathcal{F}(\mathcal{A}, \mathcal{B})$ is a VC-subgraph class, its growth function is bounded by

$$\Delta_n(\mathcal{F}(\mathcal{A}, \mathcal{B})^+) \leq \left(\frac{en}{d}\right)^{2d} \tag{6.50}$$

and

$$\mathrm{vc}(\mathcal{F}(\mathcal{A}, \mathcal{B})) \leq 2d \log_2(e). \tag{6.51}$$

The estimates (6.50) *and* (6.51) *generalize to*

$$\mathcal{F}(\mathcal{A}, \mathcal{H}) = \{F_{A,h}(x, y) = \mathbb{1}_A(x) f(x) + \mathbb{1}_{A^c}(x)h(y) | A \in \mathcal{A}, h \in \mathcal{H}\}, \tag{6.52}$$

where $\mathcal{H}$ is a VC-class of function with $\mathrm{vc}(\mathcal{H}) = \mathrm{vc}(\mathcal{H}^+) \leq d$.

PROOF. Given points $(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}$ and $t_i \in \mathbb{R}$, $i = 1, \dots, n$, we need to bound the cardinality of

$$\{(\theta(F_{A,B}(x_i, y_i) - t_i))_{i=1,\dots,n} | A \in \mathcal{A}, B \in \mathcal{B}\}, \tag{6.53}$$

as a subset of the binary cube $\{0, 1\}^n$. Because

$$F_{A,B}(x, y) = \mathbb{1}_B(y_i)(g(y_i) - \mathbb{1}_A(x_i)g(y_i)) + \mathbb{1}_A(x_i) f(x_i),$$

and $(g(y_i) - \mathbb{1}_A(x_i)g(y_i)) \geq 0$, we find that

$$\begin{aligned} &\theta(F_{A,B}(x_i, y_i) - t_i) \\ &\quad = \begin{cases} \theta(\mathbb{1}_B(y_i) - \tau_i(A)), & \text{on } S_+(A) = \{(x_j, y_j) | \mathbb{1}_{A^c}(x_j)g(y_j) > 0\}, \\ \theta(\mathbb{1}_A(x_i) f(x_i) - t_i), & \text{on } S_0(A) = \{(x_j, y_j) | \mathbb{1}_{A^c}(x_j)g(y_j) = 0\}, \end{cases} \end{aligned} \tag{6.54}$$

where

$$\tau_i(A) = \frac{t_i - \mathbb{1}_A(x_i) f(x_i)}{g(y_i) - \mathbb{1}_A(x_i) g(y_i)}. \tag{6.55}$$

Fix A and vary B over $\mathcal{B}$. Because $\mathrm{vc}(\mathcal{B}) \le d$, we see from (6.54) and Sauer's lemma, that the binary set

$$\{(\theta(F_{A,B}(x_i, y_i) - t_i))_{i=1,\dots,n} | B \in \mathcal{B}\} \tag{6.56}$$

has cardinality K bounded above by $(end^{-1})^d$. Let

$$b_1(A), \dots, b_K(A) \tag{6.57}$$

enumerate the distinct elements of (6.56) generated by sets B_k. For $(x_i, y_i) \in S_0(A)$, we have

$$b_{k,i}(A) = \theta(\mathbb{1}_A(x_i) f(x_i) - t_i), \tag{6.58}$$

and if $(x_i, y_i) \in S_+(A)$,

$$\begin{aligned} b_{k,i}(A) &= \theta(\mathbb{1}_{B_k}(y_i) - \tau_i(A)) \\ &= \begin{cases} \theta(\mathbb{1}_A(x_i) - \tau_i(B_k)), \\ \qquad \text{on } S_+(B_k) = \{(x_j, y_j) | f(x_j) - \mathbb{1}_B(x_j) g(y_j) > 0\}, \\ 1 - \theta(\mathbb{1}_A(x_i) - \tau_i(B_k)), \\ \qquad \text{on } S_-(B_k) = \{(x_j, y_j) | f(x_j) - \mathbb{1}_B(x_j) g(y_j) < 0\}, \\ \theta(\mathbb{1}_{B_k}(y_i) g(y_i) - t_i), \\ \qquad \text{on } S_0(B_k) = \{(x_j, y_j) | f(x_j) - \mathbb{1}_B(x_j) g(y_j) = 0\}. \end{cases} \end{aligned} \tag{6.59}$$

Consequently, Sauer's lemma again implies that for each fixed k the binary set

$$\{b_k(A) | A \in \mathcal{A}\} \tag{6.60}$$

has cardinality at most $(end^{-1})^d$. This proves (6.50). Again, by Sauer's lemma, very $n_0 > 0$ such that

$$\mathrm{card}\{(\theta(F_{A,B}(x_i, y_i) - t_i))_{i=1,\dots,n} | A \in \mathcal{A}, B \in \mathcal{B}\} \le \left(\frac{en}{d}\right)^{2d} < 2^n, \tag{6.61}$$

for all $n > n_0$ is an upper bound of $\mathrm{vc}(\mathcal{F}(\mathcal{A}, \mathcal{B})^+)$. To find n_0, we look for solutions $n_0 = dj$ that are multiples of d. (6.61) leads to the condition

$$\log_2(ej) < j,$$

which is satisfied, for example, by $j = 2\log_2(e)$. The extension to $\mathcal{F}(\mathcal{A}, \mathcal{H})$ is straightforward. Replace $\theta(\mathbb{1}_B(y_i) - \tau_i(A))$ in (6.54) by $\theta(h(y_i) - \tau_i(A))$, where $\tau_i(A) = (t_i - f(x_i))/\mathbb{1}_{A^c}(x_i)$ and follow the same lines of reasoning. $\square$

PROOF OF PROPOSITION 6.9. Recall definition (3.16) of the cash flow function, according to which

$$\begin{aligned}\vartheta_{t+1:w}(f,h) &= \theta_{f,t+1}(h) f_{t+1} + \cdots + \theta_{f,t+w+1}(h) \prod_{r=t+1}^{t+w} \theta^-_{f,r}(h) f_{t+w+1} \\ &\quad + \prod_{r=t+1}^{t+w+1} \theta^-_{f,r}(h) h_{t+w+1}.\end{aligned} \tag{6.62}$$

Because the classes of indicator functions

$$\begin{aligned}\mathcal{C}_s &= \{\theta_{f,s}(h) = \mathbb{1}_{\{f_s - h_s \ge 0\}} | h_s \in \mathcal{H}_s\}, \\ \mathcal{C}_s^- &= \{\theta^-_{f,s}(h) = \mathbb{1}_{\{f_s - h_s < 0\}} | h_s \in \mathcal{H}_s\},\end{aligned} \tag{6.63}$$

are VC classes with VC-dimension

$$\operatorname{vc}(\mathcal{C}_s^-) = \operatorname{vc}(\mathcal{C}_s) = \operatorname{vc}\big((f_s - \mathcal{H}_s)^+\big) = \operatorname{vc}(\mathcal{H}_s^+) = \operatorname{vc}(\mathcal{H}_s) \le d, \tag{6.64}$$

we can recursively apply Lemma 6.11 to derive the bound

$$\operatorname{card}\{\big(\theta\big(\vartheta_{t+1:w}(f,h)(\mathbf{x}_i) - t_i\big)\big)_{i=1,\dots,n} | h \in \mathcal{H}\} \le \left(\frac{en}{d}\right)^{d(w+2)}. \tag{6.65}$$

The VC-dimension of $\mathcal{G}_t$ is then estimated as in the proof of Lemma 6.11. This completes the proof of Proposition 6.9. □

6.3. *Proofs of Theorems* 5.1 *and* 5.3. The centered loss $l_t(\hat{q}_{\mathcal{H}})$ depends on the sample D_n. To control the fluctuations of the random variable $E[l_t(\hat{q}_{\mathcal{H}})]$, we need uniform estimates over the whole centered loss class $\mathcal{L}_t(\mathcal{H})$. The usual procedure is to apply exponential deviation inequalities for the empirical process

$$\{\sqrt{n}(E[l] - P_n l) | l \in \mathcal{L}_t(\mathcal{H})\} \tag{6.66}$$

indexed by $\mathcal{L}_t(\mathcal{H})$, which are closely related to the uniform law of large numbers. For background, we refer to [22, 40, 48, 51].

The application of standard deviation inequalities to the whole centered loss class $\mathcal{L}_t(\mathcal{H})$ is not efficient since the empirical minimizer is close to the actual L_2-minimizer with high probability. Therefore, the random element $l_t(\hat{q}_{\mathcal{H}})$ is with high probability in a small subset of $\mathcal{L}_t(\mathcal{H})$. To get sharper estimates, the empirical process needs to be localized such that more weight is assigned to these loss functions. Lee, Bartlett and Williamson [35] proved the following localized deviation inequality.

THEOREM 6.12 ([35], Theorem 6). *Let $\mathcal{L}$ be a class of functions such that $|l| \le K_1$, $E[l] \ge 0$, and for some $K_2 \ge 1$,*

$$E[l^2] \le K_2 E[l] \qquad \forall l \in \mathcal{L}. \tag{6.67}$$

Let $a, b > 0$ *and* $0 < \delta < \frac{1}{2}$. *Then, for all*

$$n \geq \max\left(\frac{4(K_1 + K_2)}{\delta^2(a+b)}, \frac{K_1^2}{\delta^2(a+b)}\right), \tag{6.68}$$

$$\begin{aligned} &\mathbb{P}\left(\sup_{l \in \mathcal{L}} \frac{E[l] - P_n(l)}{E[l] + a + b} \geq \delta\right) \\ &\quad \leq 2 \sup_{x_1,\dots,x_{2n} \in \mathcal{X}^{2n}} N\left(\frac{\delta b}{4}, \mathcal{L}, d_{1,P_{2n}}\right) \exp\left(-\frac{3\delta^2 an}{4K_1 + 162K_2}\right) \\ &\qquad + 4 \sup_{x_1,\dots,x_{2n} \in \mathcal{X}^{2n}} N\left(\frac{\delta b}{4K_1}, \mathcal{L}, d_{1,P_{2n}}\right) \exp\left(-\frac{\delta^2 an}{2K_1^2}\right), \end{aligned} \tag{6.69}$$

where P_{2n} *is the empirical measure supported at* $(x_1, \dots, x_{2n})$.

A similar bound has been obtained by Cucker and Smale ([17], Proposition 7) for L_∞-covering numbers. Theorem 6.12 has been improved in [28] by applying chaining techniques, and in [4] by using concentration properties of local Rademacher averages. For additional background on related bounds, we refer to [34, 37, 42, 48]. The advantage of Theorem 6.12, as compared to the Pollard's deviation inequality, is that it improves the quadratic dependence on ε in standard deviation inequalities to a linear dependence.

The centered loss has a special structure which allows it to bound its variance in terms of its expectation.

LEMMA 6.13. *Let* $\mathcal{H}_t$ *be convex, uniformly bounded by* $H < \infty$, *and assume that* $\vartheta_{t+1:w}(f,h) \leq \Theta$ *for some constant* $\Theta < \infty$. *Then the centered loss class* $\mathcal{L}_t(\mathcal{H})$ *is uniformly bounded and, for all* $l \in \mathcal{L}_t(\mathcal{H})$,

$$\begin{aligned} |l| &\leq 4H(\Theta + H), \\ E[l^2] &\leq 4(\Theta + H)^2 E[l]. \end{aligned} \tag{6.70}$$

PROOF. We get from the definition (6.3) of $l_t(h)$ that

$$\begin{aligned} l_t(h) &= \big(h_t - \mathrm{pr}_{\mathcal{H}_t} \vartheta_{t+1:w}(f,h)\big) \\ &\quad \times \big(h_t + \mathrm{pr}_{\mathcal{H}_t} \vartheta_{t+1:w}(f,h) - 2\vartheta_{t+1:w}(f,h)\big) \\ &\leq 2(\Theta + H)\big(h_t - \mathrm{pr}_{\mathcal{H}_t} \vartheta_{t+1:w}(f,h)\big). \end{aligned} \tag{6.71}$$

Therefore,

$$E[l_t(h)^2] \leq 4(\Theta + H)^2 E\big[\big|h_t - \mathrm{pr}_{\mathcal{H}_t} \vartheta_t\big|^2\big] \leq 4(\Theta + H)^2 E[l_t(h)],$$

where the last step follows form Lemma 6.3. $\square$

Our plan is to apply Theorem 6.12 to a suitably scaled loss class

$$\lambda\mathcal{L}_t(\mathcal{H}) = \{\lambda l \mid l \in \mathcal{L}_t(\mathcal{H})\}, \tag{6.72}$$

where we choose λ such that $|\lambda l| \le 1$ [the scaling gives a term β_n^2 in the consistency condition (5.3) instead of β_n^4]. Because an empirical risk minimizer satisfies $P_n(l_t(\hat{q}_{\mathcal{H}})) \le 0$, it follows that, for any $\varepsilon > 0$ and scaling factor $\lambda > 0$,

$$\begin{aligned}\mathbb{P}\big(E[l_t(\hat{q}_{\mathcal{H}})] \ge \varepsilon\big) &\le \mathbb{P}\big(E[l_t(\hat{q}_{\mathcal{H}})] \ge 2P_n\big(l_t(\hat{q}_{\mathcal{H}})\big) + \varepsilon\big)\\ &= \mathbb{P}\left(\frac{E[l_t(\hat{q}_{\mathcal{H}})] - P_n(l_t(\hat{q}_{\mathcal{H}}))}{E[l_t(\hat{q}_{\mathcal{H}})] + \varepsilon} \ge \frac{1}{2}\right)\\ &\le \mathbb{P}\left(\sup_{l \in \lambda\mathcal{L}_t(\mathcal{H})} \frac{E[l] - P_n(l)}{E[l] + \lambda\varepsilon} \ge \frac{1}{2}\right).\end{aligned} \tag{6.73}$$

Assume that the conditions of Lemma 6.13 are satisfied and set $\beta = \max(\Theta, H)$. If we choose the scaling factor $\lambda = 1/(8\beta^2)$, the scaled class $\lambda\mathcal{L}_t(\mathcal{H})$ satisfies

$$\begin{aligned}|\lambda l| &\le 1,\\ E[(\lambda l)^2] &\le 2E[\lambda l].\end{aligned} \tag{6.74}$$

Theorem 6.12 applied with $\mathcal{L} = \lambda\mathcal{L}_t(\mathcal{H})$, $K_1 = 1$, $K_2 = 2$, $a = b = \varepsilon/(16\beta^2)$, $\delta = 1/2$ implies

$$\begin{aligned}&\mathbb{P}\big(E[l_t(\hat{q}_{\mathcal{H}})] \ge \varepsilon\big)\\ &\quad\le 6 \sup_{x_1,\dots,x_{2n}} N\left(\frac{\varepsilon}{128\beta^2}, \frac{1}{8\beta^2}\mathcal{L}_t(\mathcal{H}), d_{1,P_{2n}}\right) \exp\left(-\frac{n\varepsilon}{6998\beta^2}\right),\end{aligned} \tag{6.75}$$

for $n \ge 382\beta^2/\varepsilon$. The ε-covering number of $\lambda\mathcal{L}_t(\mathcal{H})$ is the same as the $(\lambda^{-1}\varepsilon)$-covering number of the unscaled class $\mathcal{L}_t(\mathcal{H})$. If the VC-dimension of $\mathcal{H}_s$, $s \ge t$ are bounded by d, the covering number bound (6.47) shows that

$$\mathbb{P}\big(E[l_t(\hat{q}_{\mathcal{H}})] \ge \varepsilon\big) \le K\left(\frac{1}{\varepsilon}\right)^v \exp\left(-\frac{n\varepsilon}{6998\beta^2}\right), \tag{6.76}$$

where

$$v = v(w, d) = 2d\big(c(w) + 1\big) \tag{6.77}$$

and

$$K = K(d, w, \beta) = 6e^4(d+1)^2\big(c(w)d + 1\big)^2(1024e\beta)^{v(d,w)}. \tag{6.78}$$

PROOF OF THEOREM 5.1. β_n is a sequence of truncation thresholds tending to infinity. If $\bar{q}_{\beta_n,t}$ is the continuation value for the truncated payoff $T_{\beta_n} f$, we get from (5.6) that $\|q_t - \bar{q}_{\beta_n,t}\|_2 \to 0$. The error decomposition (6.4) separates the approximation error and the sample error. The denseness assumption implies that the approximation error $\inf_{h\in\mathcal{H}_{n,t}} \|h-\bar{q}_{\beta_n,t}\|_2$ tends to zero if $n\to\infty$. It remains to analyze the sample error $E[l_t(\hat{q}_{\mathcal{H}_n})]$ for underlying payoff $T_{\beta_n} f$.

We apply (6.76) to $\mathcal{H}=\mathcal{H}_n$, for which $d=d_n$ and $\beta=\beta_n$. There exists a constant $C(\varepsilon,w)$ such that, for every fixed $\varepsilon>0$,

$$\mathbb{P}\big(E\big[l_t(\hat{q}_{\mathcal{H}_n})\big]\geq\varepsilon\big)\leq C(\varepsilon,w)\exp\biggl(d_n\log(\beta_n)-\frac{n\varepsilon}{6998\beta_n^2}\biggr). \tag{6.79}$$

The right-hand side converges to zero for every fixed $\varepsilon>0$ if n/β_n^2 diverges to infinity faster than $d_n\log(\beta_n)$ or if $d_n\beta_n^2\log(\beta_n)n^{-1}\to 0$. Convergence in probability follows from (6.4) by induction. Convergence in $L_1(\mathbb{P})$ is shown by evaluating

$$\mathbb{E}\big[E\big[l_t(\hat{q}_{\mathcal{H}_n})\big]\big]\leq\varepsilon+\int_\varepsilon^\infty \mathbb{P}\big(E\big[l_t(\hat{q}_{\mathcal{H}_n})\big]>t\big)\,dt, \tag{6.80}$$

using the estimate (6.79). Conditions (5.3) and (5.5) imply

$$\begin{aligned}
&\sum_{n=1}^\infty \mathbb{P}\big(E\big[l_t(\hat{q}_{\mathcal{H}_n})\big]\geq\varepsilon\big)\\
&\qquad\leq C(\varepsilon,w)\sum_{n=1}^\infty \exp\biggl(d_n\log(\beta_n)-\frac{n\varepsilon}{6998\beta_n^2}\biggr)\\
&\qquad= C(\varepsilon,w)\sum_{n=1}^\infty n^{-n/(\log(n)\beta_n^2)((\varepsilon/6998)-d_n\beta_n^2\log(\beta_n)/n)}<\infty.
\end{aligned} \tag{6.81}$$

Almost sure convergence follows from the Borel–Cantelli lemma. $\square$

PROOF OF THEOREM 5.3. Integrating (6.76) over ε shows that, for any $\kappa\geq\frac{382\beta^2}{n}$,

$$\begin{aligned}
\mathbb{E}\big[E[l_t(\hat{q}_{\mathcal{H}})]\big]&=\int_0^\infty \mathbb{P}\big(E[l_t(\hat{q}_{\mathcal{H}})]\geq\varepsilon\big)\,d\varepsilon\\
&\leq\kappa+Kn^v\int_\kappa^\infty \exp\biggl(-\frac{n\varepsilon}{6998\beta^2}\biggr)\,d\varepsilon\\
&\leq\kappa+Kn^{v-1}6998\beta^2\exp\biggl(-\frac{n\kappa}{6998\beta^2}\biggr).
\end{aligned} \tag{6.82}$$

Setting

$$\kappa=\frac{6998\beta^2}{n}\log(6998K\beta^2n^v)\geq\frac{382\beta^2}{n}, \tag{6.83}$$

leads to the upper bound

$$\mathbb{E}\big[E[l_t(\hat{q}_{\mathcal{H}})]\big]\leq\frac{6998\beta^2+\log(6998K\beta^2)}{n}+\frac{v\log(n)}{n}. \tag{6.84}$$

Corollary 6.2 implies that

$$\mathbb{E}[\|\hat{q}_{\mathcal{H},t} - q_t\|_2^2] \\ \le 2 \cdot 16^{w+1}\Bigl(\max_{s=t,\dots,t+w+1} \inf_{h\in\mathcal{H}_s} \|h - q_s\|_2^2 + \mathbb{E}\Bigl[\max_{s=t,\dots,t+w+1} E[l_s(\hat{q}_{\mathcal{H}})]\Bigr]\Bigr).$$

But

$$\mathbb{E}\Bigl[\max_{s=t,\dots,t+w+1} E[l_s(\hat{q}_{\mathcal{H}})]\Bigr] \le (w+2) \max_{s=t,\dots,t+w+1} \mathbb{E}\bigl[E[l_s(\hat{q}_{\mathcal{H}})]\bigr].$$

Apply (6.84) to complete the proof. □

PROOF OF COROLLARY 5.4. Estimate (6.76) implies

$$\mathbb{P}\bigl(E[l_t(\hat{q}_{\mathcal{H}})] \ge \epsilon\bigr) \le K \exp\Bigl(-\frac{n\epsilon}{13996\beta^2}\Bigr)\exp\Bigl(-\frac{n\epsilon}{13996\beta^2} - \log(\epsilon)v\Bigr). \tag{6.85}$$

By straightforward calculations, the right-hand side is smaller than δ for all n satisfying

$$n \ge 13996\beta^2 \max\Bigl(\frac{1}{\epsilon}\log\Bigl(\frac{K}{\delta}\Bigr), v\log\Bigl(\frac{1}{\epsilon}\Bigr)\Bigr). \tag{6.86}$$

The sample complexity bound (5.16) follows from Corollary 6.2 and (6.85), (6.86) with $\epsilon = \varepsilon/(32(w+2)16^w)$. □

6.4. *Proof of Corollary* 5.5. Because $q_t \in W^k(L_\infty(I,\lambda))$, Jackson-type estimates imply that, for every $r > k$, there exists a polynomial $p_r \in \mathcal{P}_r$,

$$\|p_r - q_t\|_{\infty,I,\lambda} \le C_I\, r^{-k}\, \|q_t\|_{\infty,k,I,\lambda}. \tag{6.87}$$

The constant C_I only depends on I, but not on r or q_t. See, for instance, [18], Theorem 6.2, Chapter 7. Consequently,

$$\|p_r\|_{\infty,I,\lambda} \le \|p_r - q_t\|_{\infty,I,\lambda} + \|q_t\|_{\infty,I,\lambda} \le 2\|q_t\|_{\infty,k,I,\lambda} \tag{6.88}$$

for r sufficiently large. We therefore may restrict the minimization to the convex, uniformly bounded set of functions $\mathcal{H}_{n,t}$ as defined in (5.17). The VC-dimension of $\mathcal{H}_{n,t}$ is bounded by $n^{m/(m+2k)}$. Theorem 5.3 applies. Because X_t is localized to I, the approximation error in (5.13) is bounded by

$$\inf_{p\in\mathcal{H}_{n,t}} \|p - q_t\|_2^2 \le \inf_{p\in\mathcal{H}_{n,t}} \|p - q_t\|_{\infty,I,\lambda}^2 \le C_I n^{-2k/(2k+m)}\|q_t\|_{\infty,k,I,\lambda}. \tag{6.89}$$

Inserting $\mathrm{vc}(\mathcal{H}_{n,t}) \le n^{m/(m+2k)}$ into (5.13) shows that the sample error is of the order

$$O\bigl(\log(n) n^{-2k/(2k+m)}\bigr).$$

The extension to μ_t with bounded density with respect to Lebesgue measure is proved identically.

6.5. *Proof of Proposition* 5.2. Note that

$$|\max(a,x)-\max(a,y)|\le|x-y|. \tag{6.90}$$

The representation of the continuation value in terms of the transition functions gives

$$\begin{aligned}\|q_t-\bar q_{\beta,t}\|_p &= \|E[\max(f_{t+1},q_{t+1})-\max(T_\beta f_{t+1},\bar q_{\beta,t+1})|X_t]\|_p\\ &\le \|E[f_{t+1}-T_\beta f_{t+1}|X_t]\|_p+\|E[q_{t+1}-\bar q_{\beta,t+1}|X_t]\|_p\\ &\le \big\|(f_{t+1}-\beta)\mathbb{1}_{\{f_{t+1}>\beta\}}\big\|_p+\|q_{t+1}-\bar q_{\beta,t+1}\|_p.\end{aligned}$$

If $f\in L_p(\mathbf{X})$, then $\|(f_{t+1}-\beta)\mathbb{1}_{\{f_{t+1}>\beta\}}\|_p\to 0$ for $\beta\to\infty$. We first recall that for a nonnegative random variable Y and $r>1$,

$$E[Y^r]=r\int_0^\infty y^{r-1}P(Y>y)\,dy. \tag{6.91}$$

Then (5.7) follows from

$$\begin{aligned}\big\|(f_{t+1}-\beta)\mathbb{1}_{\{f_{t+1}>\beta\}}\big\|_r^r &= r\int_0^\infty u^{r-1}P\big((f_{t+1}-\beta)\mathbb{1}_{\{f_{t+1}>\beta\}}>u\big)\,du\\ &= r\int_\beta^\infty (u-\beta)^{r-1}P(f_{t+1}>u)\,du\\ &\le r\int_\beta^\infty u^{r-1}P(f_{t+1}^p>u^p)\,du\\ &\le \frac{r}{p-r}E[f_{t+1}^p]\beta^{r-p}\le O(\beta^{r-p}),\end{aligned}$$

where we have used Markov's inequality to get to the last line.

Acknowledgments. The author would like to thank Paul Glasserman, Tom Hurd, Markus Leippold, Maung Min-Oo and Paolo Vanini for helpful discussions. The detailed comments and suggestions of a referee greatly helped to improve a first version of this paper.

REFERENCES

[1] ANTHONY, M. and BARTLETT, P. L. (1999). *Neural Network Learning*. Cambridge Univ. Press.

[2] BARONE-ADESI, G. and WHALEY, R. (1987). Efficient analytic approximation of American option values. *J. Finance* **42** 301–320.

[3] BARRON, A., BRIGÉ, L. and MASSART, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113** 301–413.

[4] BARTLETT, P., BOUSQUET, O. and MENDELSON, S. (2005). Local Rademacher complexities. *Ann. Statist.* To appear.

[5] BENSOUSSAN, A. (1984). On the theory of option pricing. *Acta Appl. Math.* **2** 139–158.

[6] BENSOUSSAN, A. and LION, J. (1982). *Application of Variational Inequalities in Stochastic Control*. North-Holland, Amsterdam.

[7] BENVENISTE, A., METIVER, M. and PRIOURET, P. (1990). *Adaptive Algorithms and Stochastic Approximations*. Springer, Berlin.

[8] BIRMAN, M. S. and SOLOMYAK, M. Z. (1967). Piecewise polynomial approximation of functions of the classes w_p^α. *Math. USSR Sb.* **73** 331–355.

[9] BOESSARTS, P. (1989). Simulation estimators of optimal early exercise. Working paper, Carnegie-Mellon Univ.

[10] BRIGÉ, L. and MASSART, P. (1998). Minimum constrast estimators on sieves: Exponential bounds and rates of convergence. *Bernoulli* **4** 329–375.

[11] BROADIE, M. and GLASSERMAN, P. (1997). Pricing American-style securities using simulation. *J. Econom. Dynam. Control* **21** 1323–1352.

[12] BROADIE, M. and GLASSERMAN, P. (1998). Monte Carlo methods for pricing high-dimensional American options: An overview. In *Monte Carlo*: *Methodologies and Applications for Pricing and Risk Management* (B. Dupire, ed.) 149–161. Risk Books, London.

[13] CARL, B. and STEPHANI, I. (1990). *Entropy*, *Compactness and the Approximation of Operators*. Cambridge Univ. Press.

[14] CHOW, Y. S., ROBBINS, H. and SIEGMUND, D. (1971). *Great Expectations*: *The Theory of Optimal Stopping*. Houghton Mifflin, Boston.

[15] CLÉMENT, E., LAMBERTON, D. and PROTTER, P. (2002). An analysis of the Longstaff–Schwartz algorithm for American option pricing. *Finance and Stochastics* **6** 449–471.

[16] COX, J., ROSS, S. and RUBINSTEIN, M. (1979). Option pricing: A simplified approach. *J. Financial Economics* **7** 229–263.

[17] CUCKER, F. and SMALE, S. (2001). On the mathematical foundations of learning. *Bull. Amer. Math. Soc.* **39** 1–49.

[18] DEVORE, R. A. and LORENTZ, G. G. (1993). *Constructive Approximation*. Springer, Berlin.

[19] DUDLEY, R. M. (1999). *Uniform Central Limit Theorems*. Cambridge Univ. Press.

[20] GLASSERMAN, P. (2004). *Monte Carlo Methods in Financial Engineering*. Springer, Berlin.

[21] GLASSERMAN, P. and YU, B. (2003). Number of paths versus number of basis functions in American option pricing. Preprint.

[22] GYÖRFI, L., KOHLER, M., KRZYŻAK, A. and WALK, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. Springer, New York.

[23] HAUSSLER, D. (1995). Sphere packing numbers for subsets of the Boolean n-cube with bounded Vapnik–Chervonenkis dimension. *J. Combin. Theory Ser. A* **69** 217–232.

[24] HUANG, J. and PANG, J. (1998). Option pricing and linear complementarity. *J. Comput. Finance* **2** 31–60.

[25] JAILLET, P., LAMBERTON, D. and LAPEYRE, B. (1990). Variational inequalities and the pricing of American options. *Acta Appl. Math.* **21** 263–289.

[26] KARATZAS, I. (1988). On the pricing of American options. *Appl. Math. Optim.* **17** 37–60.

[27] KARATZAS, I. and SHREVE, S. E. (1998). *Methods of Mathematical Finance*. Springer, New York.

[28] KOHLER, M. (2000). Inequalities for uniform deviations of averages from expectations with applications to nonparametric regression. *J. Statist. Plann. Inference* **89** 1–23.

[29] KOLMOGOROV, A. N. and TIKHOMIROV, V. M. (1959). ϵ-entropy and ϵ-capacity of function spaces. *Uspekhi Mat. Nauk.* **14** 3–86.

[30] KUSHNER, H. J. (1997). Numerical methods for stochastic control in finance. In *Mathematics of Derivative Securities* (M. A. H. Dempster and S. R. Pliska, eds.) 504–527. Cambridge Univ. Press.

[31] KUSHNER, H. J. and CLARK, D. S. (1978). *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Springer, New York.

[32] KUSHNER, H. J. and YIN, G. G. (1997). *Stochastic Approximation Algorithms and Applications*. Springer, Berlin.

[33] LAPRISE, S. B., SU, Y., WU, R., FU, M. C. and MADAN, D. B. (2001). Pricing American options: A comparision of Monte Carlo simulation approaches. *J. Comput. Finance* **4** 39–88.

[34] LEDOUX, M. (1996). On Talagrand's deviation inequality for product measures. *ESIAM Probab. Statist.* **1** 63–87.

[35] LEE, W., BARTLETT, P. and WILLIAMSON, R. C. (1996). Efficient agnonstic learning of neural networks with bounded fan-in. *IEEE Trans. Inform. Theory* **42** 2118–2132.

[36] LONGSTAFF, F. A. and SCHWARTZ, E. S. (2001). Valuing American options by simulation: A simple least-square approach. *Review of Financial Studies* **14** 113–147.

[37] MASSART, P. (2000). On the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.* **28** 863–885.

[38] MENDELSON, S. (2003). A few notes on statistical learning theory. *Advanced Lectures in Machine Learning. Lecture Notes in Comput. Sci.* **2600** 1–40. Springer, Berlin.

[39] MENDELSON, S. (2003). Geometric parameters in learning theory. GAFA lecture notes.

[40] POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

[41] POLLARD, D. (1990). *Empirical Processes*: *Theory and Applications*. IMS, Hayward, CA.

[42] RIO, E. (2001). Une ineqalité de Bennet pour les maxima de processus empirques. In *Colloque en l'Honneur de J. Bretagnolle* (D. D. Castelle and I. Ibragimov, eds.).

[43] ROGERS, L. C. G. (2002). Monte Carlo valuing of American options. *Math. Finance* **12** 271–286.

[44] SHEN, X. (1997). On methods of sieves and penalization. *Ann. Statist.* **25** 2555–2591.

[45] SHEN, X. and WONG, W. H. (1994). On methods of sieves and penalization. *Ann. Statist.* **22** 580–615.

[46] SHIRYAYEV, A. N. (1978). *Optimal Stopping Rules*. Springer, New York.

[47] STONE, C. J. (1982). Optimal rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.

[48] TALAGRAND, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Probab.* **22** 28–76.

[49] TILLEY, J. A. (1993). Valuing American options in a path simulation model. *Transactions of the Society of Actuaries* **45** 83–104.

[50] TSITSIKLIS, J. N. and VAN ROY, B. (1999). Optimal stopping of Markov processes: Hilbert space theory, approximation algorithms, and an application to pricing high-dimensional financial derivatives. *IEEE Trans. Automat. Control* **44** 1840–1851.

[51] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer, New York.

[52] VAN MOERBEKE, P. L. J. (1976). On optimal stopping and free boundary problems. *Arch. Ration. Mech. Anal.* **60** 101–148.

[53] VAPNIK, V. N. (1982). *Estimation of Dependences Based on Empirical Data*. Springer, New York.

[54] VAPNIK, V. N. (2000). *The Nature of Statistical Learning Theory*, 2nd ed. Springer, New York.

[55] VAPNIK, V. N. and CHERVONENKIS, A. Y. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.* **16** 264–280.

[56] VIDYASAGAR, A. (2003). *The Theory of Learning and Generalization with Applications to Neural Networks*, 2nd ed. Springer, New York.

[57] WALD, A. and WOLFOWITZ, J. (1948). Optimum character of the sequential probability ratio tests. *Ann. Math. Statist.* **19** 326–339.

ZURICH CANTONAL BANK
P.O. BOX CH-8010
ZÜRICH
SWITZERLAND
E-MAIL: daniel.egloff@zkb.ch