On Regression-Based Stopping Times

Benjamin Van Roy
Stanford University

October 14, 2009

Abstract

We study approaches that fit a linear combination of basis functions to the continuation value function of an optimal stopping problem and then employ a greedy policy based on the resulting approximation. We argue that computing weights to maximize expected payoff of the greedy policy or to minimize expected squared-error with respect to an invariant measure is intractable. On the other hand, certain versions of approximate value iteration lead to policies competitive with those that would result from optimizing the latter objective.

1 Introduction

There is a growing literature on computational methods for approximating solutions to high-dimensional optimal stopping problems, motivated primarily by their use in the analysis of financial derivatives (see [4] for a review of methods and this application area). A sizeable segment of this literature focusses on approximate dynamic programming approaches and, in particular, methods that fit a linear combination of basis functions to a continuation value function and then employ a greedy policy based on the resulting approximation. Such methods have proved to be useful in addressing optimal stopping problems when exact solution methods become computationally unmanageable. The case study presented in [7], for example, demonstrates substantial financial value in their application to the exercising of American-style swaptions with a multi-factor term structure model. When applying such a method, one selects a set of basis functions and

then executes an algorithm that computes basis function weights. In this paper, we study issues involving the computation of these weights.

Possibly the first issue that comes to mind when considering how to compute weights is what objective to pursue. One possibility is to aim at computing weights that maximize expected payoff from the resulting greedy policy. However, in most cases of practical interest, it is not clear how to do this efficiently. Further, as we will establish, even when the underlying optimal stopping problem is tractable, computing weights to optimize this objective, or even to approximate within a logarithmic factor, is NP-hard in the strong sense. The complexity of maximizing expected payoff encourages consideration of alternative objectives.

Another possibility is to minimize approximation error relative to the continuation value function. This approach and a choice of error metric can be motivated by a performance bound from [9], as we now explain. Consider a discrete-time infinite-horizon optimal stopping problem where the underlying state follows a time-homogeneous Markov process. The value function J^* provides the optimal expected payoff as a function of current state, while the continuation value function Q^* provides the optimal expected payoff contingent on continuing for at least one time period. For any stopping policy τ, let J^τ denote expected payoff as a function of state. Let π be an invariant measure over states of the Markov process, and define a norm $\|J\|_\pi^2 = \int J^2(x)\pi(dx)$. The following bound, established in [9], relates approximation error to performance loss:

$$\|J^* - J^{\tilde{\tau}}\|_\pi \leq \frac{1}{1-\alpha}\|Q^* - \tilde{Q}\|_\pi,$$

for all $\tilde{Q}$, where $\tilde{\tau}$ denotes the greedy policy with respect to $\tilde{Q}$. The left-hand-side represents a measure of performance loss from using $\tilde{\tau}$ instead of an optimal stopping policy. It is natural to aim at computing weights that minimize the right-hand-side.

Computing basis function weights to minimize the approximation error is equivalent to projecting Q^* onto the span of basis functions, where projection is defined with respect to $\|\cdot\|_\pi$. The performance loss of the resulting greedy policy $\hat{\tau}$ is bounded according to

$$\|J^* - J^{\hat{\tau}}\|_\pi \leq \frac{1}{1-\alpha}\|Q^* - \Pi Q^*\|_\pi,$$

where Π is the projection operator. In this paper, we establish that this bound is sharp. In particular, for any α, $1/(1-\alpha)$ is the smallest

coefficient for which the bound holds uniformly across problem instances; a problem instance here is characterized by an optimal stopping and a set of basis functions.

As we will discuss, computing weights to minimize approximation error is likely to be intractable. As such, we consider as an alternative objective *competitiveness* in a certain sense with what minimizing approximation error would accomplish. In particular, given an algorithm that generates an approximation with greedy policy $\tilde{\tau}$, we assess the algorithm based on the smallest value L^* such that

$$\|J^* - J^{\tilde{\tau}}\|_\pi \leq \frac{L^*}{1-\alpha}\|Q^* - \Pi Q^*\|_\pi,$$

for all problem instances. For an algorithm that computes the projection ΠQ^*, L^* is 1. For other algorithms that compute different basis fuction weights, L^* may be larger, possibly even infinite.

The main contribution of this paper is to show that for the approximate value iteration algorithm proposed in [9], $L^* < 2.17$. This improves on the previous performance loss bound from [9], which does not imply finiteness of L^*. It suggests that carrying out approximate value iteration is almost as effective as minimizing approximation error. The result applies as well to variations of approximate value iteration that more efficiently compute the same weights [2, 10, 12].

Other kinds of algorithms have also been proposed for computing basis function weights to approximate solutions to optimal stopping problems. In particular, there are variations of approximate policy iteration [6, 3] and approximate linear programming [1]. The reason for our focus on approximate value iteration is the theoretical bound we are able to establish. Whether similar bounds hold for other approaches remains an open issue.

2 Preliminaries

We begin by introducing the problem formulation and notation that we will be working with.

2.1 Problem Formulation

We consider a stationary Markov process $\{x_t | t = 0, 1, 2, \ldots\}$ evolving in a measurable state space $(\mathcal{S}, \mathbb{F})$ according to a transition probability

function $P : \mathcal{S} \times \mathbb{F} \mapsto [0,1]$. With some abuse of notation, we also use P to denote the single-step expectation operator:

$$(PJ)(x) = E[J(x_{t+1}) \mid x_t = x] = \int_{y \in \mathcal{S}} J(y) P(x, dy),$$

for all $J : \mathcal{S} \mapsto \Re$ for which the expectation is well-defined. All functions on $\mathcal{S}$ that we introduce are assumed to be $\mathbb{F}$-measurable.

We assume that there is an invariant measure π on $(\mathcal{S}, \mathbb{F})$, so that

$$\pi(A) = \int_{x \in \mathcal{S}} P(x, A) \pi(dx).$$

Let $L_2(\pi)$ be the Hilbert space of real-valued functions on $\mathcal{S}$ with inner product $\langle J, \overline{J} \rangle_\pi = \int J(x) \overline{J}(x) \pi(dx)$ and norm $\|J\|_\pi = \langle J, J \rangle_\pi^{1/2}$.

Let $\mathcal{F}_t$ denote the σ-algebra generated by $\{x_0, \ldots, x_t\}$. We define a stopping time to be a random variable τ that takes values in $\{0, 1, \ldots, \infty\}$ such that, for each t, the event $\{\tau \leq t\}$ is $\mathbb{F}_t$-measurable. We denote the set of stopping times by $\mathcal{T}$. Let $g \in L_2(\pi)$ represent a payoff function and $\alpha \in [0, 1)$ a discount factor. We consider an optimal stopping problem of the form

$$\sup_{\tau \in \mathcal{T}} E\left[\alpha^\tau g(x_\tau)\right].$$

Many versions of optimal stopping problems reduce to the above formulation and naturally give rise to invariant distributions. A few examples include:

1. **Ergodic Processes.** When the Markov process is ergodic, there is a unique invariant probability measure given by

   $$\pi(A) = \lim_{t \to \infty} E[\mathbf{1}_A(x_t) | x_0 = x],$$

   for all $x \in \S$, where $\mathbf{1}_A$ is the indicator of the set A. In this case, the requirement that $g \in L_2(\pi)$ ensures that the payoff has a finite second moment in steady state.

2. **Independent Increments.** Consider the case where the state space $\mathcal{S}$ is a Euclidean space $\Re^d$ and $\mathbb{F}$ consists of the Borel sets. For an independent increments process, that is, where $P(x, A)$ only depends on the difference $A - x = \{y - x | y \in A\}$, the

Lebesque measure is invariant. This is easy to show:

$$\begin{aligned}\int_{x\in\Re^d} P(x,A)dx &= \int_{x\in\Re^d} P(0,A-x)dx \\ &= \int_{y\in\Re^d} P(0,dy)\int_{x\in\Re^d} 1_{A-x}(y)dx \\ &= \int_{y\in\Re^d} P(0,dy)\int_{x\in A} dx \\ &= \int_{x\in A} dx.\end{aligned}$$

3. **Intermediate Payoffs.** Consider a situation where, in addition to the terminal payoff $g(x_\tau)$, an intermediate payoff $\hat{g}(x_t)$ is received each time a decision is made to continue while at state x_t. The objective becomes

$$\sup_{\tau\in\mathcal{T}} E\left[\sum_{t=0}^{\tau-1}\alpha^t\hat{g}(x_t)+\alpha^\tau g(x_\tau)\right].$$

This optimal stopping problem is equivalent to one with only a terminal payoff: $\sup_{\tau\in\mathcal{T}} E\left[\alpha^\tau\tilde{g}(x_\tau)\right]$, where $\tilde{g}=g-\sum_{t=0}^{\infty}\alpha^t P^t\hat{g}$. This is because

$$\begin{aligned}E\left[\alpha^\tau\tilde{g}(x_\tau)\right] &= E\left[\alpha^\tau\left(g(x_\tau)-\sum_{t=0}^{\infty}\alpha^t(P^t\hat{g})(x_\tau)\right)\right] \\ &= E\left[\alpha^\tau g(x_\tau)-\sum_{t=0}^{\infty}\alpha^{\tau+t}\hat{g}(x_{\tau+t})\right] \\ &= E\left[\sum_{t=0}^{\tau-1}\alpha^t\hat{g}(x_t)+\alpha^\tau g(x_\tau)\right]-E\left[\sum_{t=0}^{\infty}\alpha^t\hat{g}(x_t)\right],\end{aligned}$$

and the final term does not depend on τ.

4. **Finite Horizon Problems.** Suppose we start in an initial state is $x_0=\overline{x}$ and a termination decision is forced at time h. The objective becomes

$$\sup_{\tau\in\mathcal{T}} E\left[\alpha^{\tau\wedge h}g(x_{\tau\wedge h})\right].$$

We will reduce this problem to an infinite horizon one without forced termination by allowing restart at time h at a high cost.

The states of our infinite horizon problem will consist of state-time pairs $z = (x, t)$ where $x \in \mathcal{S}$ and $t \in \{0, 1, \ldots, h\}$. With some abuse of notation, let $g(z) = g(x)$ for $z = (x, t)$. We also introduce a continuation reward $\hat{g}(x, t)$ which is equal to

$$E\left[\sum_{s=0}^{h} \alpha^s |g(x_s)|\right] - g(x)$$

if $t = h$ and zero otherwise. This continuation reward makes it optimal to terminate at time h even if not forced. Hence, an equivalent objective is given by

$$\sup_{\tau \in \mathcal{T}} E\left[\sum_{t=0}^{\infty} \alpha^t \hat{g}(z_t) + \alpha^\tau g(z_\tau)\right].$$

Define a new transition operator $\tilde{P}$ on the state space $\mathbb{F} \times \{0, \ldots, h\}$ so that transitions are consistant with P when $t = 0, \ldots, h-1$, and when $t = h$, the process always transitions to $(\overline{x}, 0)$. It is easy to show that

$$\pi(A) = E\left[\sum_{t=0}^{h} \mathbf{1}_{A_t}(x_t) \Big| x_0 = \overline{x}\right],$$

is an invariant distribution, where $A_t = \{x | (x, t) \in A\}$ and $\mathbf{1}_A$ is the indicator of a set A [11].

2.2 The Value Function

Denote the expected value generated by a stopping time τ given an initial state x by

$$J^\tau(x) = E\left[\alpha^\tau g(x_\tau) | x_0 = x\right].$$

The *value function* is defined by

$$J^*(x) = \sup_{\tau \in \mathcal{T}} J^\tau(x).$$

It is easy to show [11] that the value function is in $L_2(\pi)$ and satisfies Bellman's equation:

$$J^* = \max\left(g, \alpha P J^*\right), \tag{2.1}$$

where the maximization is carried out pointwise.

When there is a tractable finite number of states, it is easy to compute an optimal stopping time. This can be done, for example, using the value iteration algorithm, which computes the value function by carrying out iterations according to (2.1). Then, an optimal stopping time is given by $\tau^* = \inf\{t \mid g(x_t) \geq J_t^*(x_t)\}$ [11]. Note that we take τ^* to be infinite when the condition $g(x_t) \geq J_t^*(x_t)$ is never met.

2.3 The Continuation Value Function

In the design of approximation algorithms, it is often more convenient to work with the *continuation value function*, defined by

$$Q^*(x) = \sup_{\tau \in \mathcal{T}} E\left[\alpha^\tau g(x_\tau) \mid x_0 = x, \tau > 0\right].$$

The continuation value function evaluates the optimal expected discounted future payoff in the event that the process continues for at least one time period. This function is also in $L_2(\pi)$ and satisfies a variation of Bellman's equation

$$Q^* = \alpha P \max\left(g, Q^*\right). \tag{2.2}$$

Further, the continuation and standard value functions relate according to

$$J^* = \max(g, Q^*) \quad \text{and} \quad Q^* = \alpha P J^*.$$

There is a variation of value iteration that computes the continuation value function by carrying out iterations according to (2.2). An optimal stopping time is given by $\tau^* = \inf\{t \mid g(x_t) \geq Q_t^*(x_t)\}$.

2.4 Regression-Based Stopping Times

Our focus is on cases where the state space $\mathcal{S}$ is intractably large, possibly infinite. A prototypical case is $\mathcal{S} = \Re^d$. When d is small, say less than 5, a close approximation to the value and/or continuation value function can often be efficiently computed over a discrete grid in $\Re^d$. The resulting approximation can then be used to generate a near-optimal stopping time. The methods we will discuss, however, are intended for cases where d is large and the grid resulting from discretization would not be manageable.

We consider use of a finite collection of basis functions $\phi_1, \ldots, \phi_K \in L_2(\pi)$ to approximate the continuation value function. For any $r \in$

$\Re^K$, let $\Phi r = \sum_{k=1}^{K} r_k \phi_k$. The selection of basis functions is an important topic in its own right but not one we will treat in this paper. Rather, we take the basis functions as given and study approaches to computing weight vectors r so that Φr approximates Q_t^*. This leads to a suboptimal stopping time

$$\tau_r = \inf\{t \mid g(x_t) \geq (\Phi r)(x_t)\}.$$

3 Objectives

In this section, we study objectives one might seek to optimize in computing weight vectors. The first involves maximizing expected payoff and the second minimizing approximation error. We argue that both optimization problems are intractable and consider as an alternative competing in a certain sense with policies that would be generated if the latter objective were optimized.

3.1 Maximizing Expected Payoff

Given an initial state x, it is natural to consider the problem of optimizing over r the expected payoff generated by stopping time τ_r:

$$\sup_{r \in \Re^K} J^{\tau_r}(x). \tag{3.1}$$

This problem is intractable in a certain sense that we will now explain.

There is no finite encoding for our class of problems. This is because there are no finite encodings that can represent elements of our problem data such as the transition operator P. The intention of our formulation was to capture many classes of problems associated with different finite encodings. However, in order to relate to the framework of computational complexity theory and prove an intractability result in that context, we must specify a narrower class of problems and a specific encoding of input data. In this spirit, we define the *finite-state regression-based stopping problem*. Each problem instance is characterized by a sextuple $(\mathcal{S}, P, g, \alpha, K, \Phi)$, where $\mathcal{S}$ is a finite set, P is a $|\mathcal{S}| \times |\mathcal{S}|$ transition matrix, $g \in \Re^{|\mathcal{S}|}$, $\alpha \in [0, 1)$, K is the number of basis elements, and $\Phi \in \Re^{|\mathcal{S}| \times K}$. The objective is as in (3.1). The following result characterizes the complexity of this problem.

Theorem 1 *The finite-state regression-based stopping problem is NP-hard in the strong sense, even to approximate within a logarithmic factor.*

Our proof of Theorem 1 relies on analysis of an auxiliary problem, which we will refer to as the *matrix thresholding problem.* The problem data consists of a matrix $A \in \Re^{M \times N}$. Each vector $\theta \in \Re^{M-1}$ constitutes a feasible solution. The objective is

$$\max_{\theta \in \Re^{M-1}} \sum_{j=1}^{N} A_{i_j(\theta),j}, \text{ where } i_j(\theta) = \min\{M, i | A_{ij} \geq \theta_i\}. \tag{3.2}$$

The following lemma characterizes the complexity of this problem.

Lemma 1 *The matrix thresholding problem is NP-hard in the strong sense, even to approximate within a logarithmic factor.*

Proof: We will transform the minimum set cover problem to the matrix thresholding problem. First let us review the minimum set cover problem. Consider a nonempty finite set S and collection C of nonempty subsets $S_1, \ldots, S_n$. A set cover is a subset $C' \subseteq C$ such that each element of S belongs to at least one element of C'. The objective of the minimal set cover problem is to find a set cover of minimal cardinality. This problem is NP-hard, even to approximate to within a logarithmic factor [8].

We will define a matrix thresholding problem that solves the minimum set cover problem. Let the number of rows M be $n+1$ and the number of columns N be $2|S|+n$. The elements of the matrix are

$$A_{ij} = \begin{cases} \mathbf{1}(j \in S_i) - \kappa & \text{if } i \leq n, j \leq |S| \\ \mathbf{1}(j - |S| \in S_i) - \kappa & \text{if } i \leq n, |S| < j \leq 2|S| \\ \mathbf{1}(j - 2|S| = i) - \kappa & \text{if } i \leq n, 2|S| < j \\ -\kappa & \text{if } i = n+1, j \leq 2|S| \\ 2 - \kappa & \text{if } i = n+1, 2|S| < j, \end{cases}$$

where $\kappa = 2(|S|+n)/(2|S|+n)$. The set of objective values that can be obtained by solutions $\theta \in \Re^{M+1}$ can be obtained by $\theta \in \{-\kappa, 1-\kappa, 2-\kappa\}^{M+1}$. Further, at any optimal solution no components of θ can be set to $-\kappa$. So, without loss of generality, we restrict θ to be in $\theta \in \{1-\kappa, 2-\kappa\}^{M+1}$.

Let $U = \{i|\theta_i = 1-\kappa\}$ and $\overline{U} = \{i|\theta_i = 2-\kappa\}$. Let $S_U = \{j \in S_i | i \in U\}$ and $\overline{S}_U = S \setminus S_U$. Then,

$$\begin{aligned}\sum_{j=1}^{N} A_{i_j(\theta),j} &= 2(1-\kappa)|S_U| - 2\kappa|\overline{S}_U| - (1-\kappa)|U| + (2-\kappa)|\overline{U}| \\ &= -(2|\overline{S}_U| + |U|).\end{aligned}$$

Note that solving $\max_{U\subseteq\{1,\ldots,n\}}(2|\overline{S}_U|+|U|)$ results in $\overline{S}_U$ being empty and U being the index set for a minimum set cover. It follows that the matrix thresholding problem is NP-hard.

Now suppose we approximate the solution to the matrix thresholding problem within a factor of c. This offers a factor of c approximation for $\max_{U\subseteq\{1,\ldots,n\}}(2|\overline{S}_U| + |U|)$. If the associated set $\overline{S}_U$ is nonempty, for each element $j \in \overline{S}_U$, augment U with the index of a set that contains j. This results in a set cover U that is within a factor of c from optimal. It follows that finding a logarithmic factor approximate solution to the matrix thresholding problem is NP-hard in the strong sense. ■

Proof of Theorem 1: We show how the matrix thresholding problem can be transformed to a finite-state regression-based stopping problem with forced termination at time $h = M$. This problem can in turn be transformed to the finite-state regression-based stopping problem using the method discussed at the end of Section 2.1.

Let the Markov process start in a distinguished state $x = 0$, and let each other state correspond to an element of the matrix A. If $x_t = 0$ then x_{t+1} is sampled uniformly at random from among $(1,1),(1,2),\ldots,(1,N)$. If $x_t = (i,j)$ with $i < M$ then $x_{t+1} = (i+1,j)$. If $x_t = (i,j)$ with $i = M$ then $x_{t+1} = 0$. The invariant probability measure assigns probability $1/(M+1)$ to state 0 and $1/(N(M+1))$ to each other state.

Let the discount factor be $\alpha = 1/2$ and the payoff function be given by $g(0) = -1$ and $g(i,j) = \alpha^{-i}A_{ij}$. Use a set of $M-1$ basis function, each kth basis function is an indicator that is positive if $x = (i,j)$ with $i = k$. Note that because the approximate continuation value at state 0 is 0 and the termination payoff there is -1, $\tau_r > 0$.

The objective of the optimal stopping problem we have formulated is

$$\max_r E\left[\alpha^{\tau_r\wedge h} g(x_{\tau_r\wedge h})\right] = \max_r \frac{1}{N}\sum_{j=1}^{N} A_{i_j(r)j},$$

where $i_j(r) = \min\{M, i | A_{ij} \geq r_i\}$. Hence, the finite-state regression-based stopping problem we have formulated is equivalent to the matrix thresholding problem. The result follows. ■

Given a finite-state regression-based stopping problem, if we ignore the basis functions, the remaining input data define an optimal stopping problem. An optimal stopping time for this problem can be computed in pseudo-polynomial time using, for example, linear programming. As such, Theorem 1 implies that, given a collection of basis functions, optimizing performance – or even coming within a logarithmic factor of that – is harder than solving the original optimal stopping problem.

3.2 Minimizing Approximation Error

An alternative, less direct, objective is to minimize approximation error relative to the continuation value function Q^*. This can be motivated by a performance loss bound originally established in [9]. The following lemma will be used to establish the bound. Recall that $\|J\|_\pi^2 = \int J^2(x)\pi(dx)$, and for any stopping policy τ, let $Q^\tau = \alpha P J^\tau$.

Lemma 2 *For all* $(\mathcal{S}, \mathbb{F})$*,* g*,* P*,* $\tilde{Q}$*, and* $\alpha \in [0, 1)$*,*

$$\|J^* - J^{\tilde{\tau}}\|_\pi \leq \|Q^* - \tilde{Q}\|_\pi + \|Q^* - Q^{\tilde{\tau}}\|_\pi,$$

where $\tilde{\tau} = \inf\{t | g(x_t) \geq \tilde{Q}(x_t)\}$.

Proof: Note that

$$J^*(x) - J^{\tilde{\tau}}(x) = \begin{cases} 0 & \text{if } g(x) \geq Q^*(x), g(x) \geq \tilde{Q}(x) \\ Q^*(x) - g(x) & \text{if } g(x) < Q^*(x), g(x) \geq \tilde{Q}(x) \\ Q^*(x) - Q^{\tilde{\tau}}(x) & \text{if } g(x) < Q^*(x), g(x) < \tilde{Q}(x) \\ g(x) - Q^{\tilde{\tau}}(x) & \text{if } g(x) \geq Q^*(x), g(x) < \tilde{Q}(x) \end{cases}$$

$$\leq \begin{cases} 0 \\ Q^*(x) - \tilde{Q}(x) \\ Q^*(x) - Q^{\tilde{\tau}}(x) \\ \tilde{Q}(x) - Q^*(x) + Q^*(x) - Q^{\tilde{\tau}}(x). \end{cases}$$

It follows that

$$\|J^* - J^{\tilde{\tau}}\|_\pi \leq \|Q^* - \tilde{Q}\|_\pi + \|Q^* - Q^{\tilde{\tau}}\|_\pi.$$

■

We will also make use of the following lemma which establishes that P is a nonexpansion on $L_2(\pi)$.

Lemma 3 *For any P and Q, $\|PQ\|_\pi \leq \|Q\|_\pi$.*

Proof: For any Q, we have

$$\begin{aligned}
\|PQ\|_\pi^2 &= \int \pi(dx) \left(\int P(x,dy)Q(y) \right)^2 \\
&\leq \int \pi(dx) \int P(x,dy)Q^2(y) \\
&= \int \left(\int \pi(dx)P(x,dy) \right) Q^2(y) \\
&= \int \pi(dy)Q^2(y) \\
&= \|Q\|_\pi^2
\end{aligned}$$

by Jensen's inequality and the invariance of π. ■

Using the previous lemmas, we establish the bound.

Theorem 2 *For all $(\mathcal{S}, \mathbb{F})$, g, P, $\tilde{Q}$, and $\alpha \in [0,1)$,*

$$\|J^* - J^{\tilde{\tau}}\|_\pi \leq \frac{1}{1-\alpha} \|Q^* - \tilde{Q}\|_\pi,$$

where $\tilde{\tau} = \inf\{t | g(x_t) \geq \tilde{Q}(x_t)\}$.

Proof: By Lemmas 2 and 3,

$$\begin{aligned}
\|J^* - J^{\tilde{\tau}}\|_\pi &\leq \|Q^* - \tilde{Q}\|_\pi + \alpha \|P(J^* - J^{\tilde{\tau}})\|_\pi \\
&\leq \|Q^* - \tilde{Q}\|_\pi + \alpha \|J^* - J^{\tilde{\tau}}\|_\pi,
\end{aligned}$$

and therefore,

$$\|J^* - J^{\tilde{\tau}}\|_\pi \leq \frac{1}{1-\alpha} \|Q^* - \tilde{Q}\|_\pi.$$

■

The left-hand-side of the above bound represents a measure of performance loss incurred by using $\tilde{\tau}$ rather than an optimal stopping policy. The right-hand-side is a multiple of approximation error, measured in terms of the norm $\|\cdot\|_\pi$. It is natural to aim at minimizing the right-hand-side:

$$\min_{r \in \Re^K} \|Q^* - \Phi r\|_\pi.$$

However, there is no efficient algorithm for this problem, which is challenging for reasons we now explain.

Suppose there were an efficient algorithm for minimizing $\|Q^* - \Phi r\|_\pi$. for any choice of basis functions. Then, this algorithm can efficiently compute optimal stopping decisions. as follows. Given a current state x, select the indicator of x as a sole basis function. Minimizing $\|Q^* - \Phi r\|_\pi$ results in $(\Phi r)(x) = Q^*(x)$. If $(\Phi r)(x) \leq g(x)$ termination is optimal, otherwise continuation is.

The above line of reasoning suggests that minimizing $\|Q^* - \Phi r\|_\pi$ is at least as hard as solving the optimal stopping problem. However, our motivation for approximation in the first place is an inability to efficiently solve the optimal stopping problem, and as such, it is unlikely that we will be able to efficiently minimize $\|Q^* - \Phi r\|_\pi$.

3.3 Competing with the Best Approximation

The linear combination of basis functions that minimizes approximation error is the projection ΠQ^* of Q^* onto the span of basis functions with respect to $\|\cdot\|_\pi$. Theorem 2 tells us that

$$\|J^* - J^{\tilde{\tau}}\|_\pi \leq \frac{1}{1-\alpha}\|Q^* - \Pi Q^*\|_\pi, \tag{3.3}$$

if $\tilde{\tau} = \inf\{t | g(x_t) \geq (\Pi Q^*)(x_t)\}$.

If the projection ΠQ^* cannot be computed efficiently, we must resort to algorithms that compute alternative basis functions weights. Consider an algorithm that takes as input the problem data $(\mathcal{S}, \mathbb{F}, P, g, \Phi)$ and generates basis function weights r that lead to a greedy policy τ_r. We propose rating such an algorithm based on the smallest value L^* such that

$$\|J^* - J^{\tau}\|_\pi \leq \frac{L^*}{1-\alpha}\|Q^* - \Pi Q^*\|_\pi,$$

for all problem instances. Based on 3.3, for an algorithm that minimizes approximation, $L^* \leq 1$. The following result establishes that $L^* = 1$ in this case.

Theorem 3 *For all $\mathcal{S}$ with $|\mathcal{S}| > 1$ and $\alpha \in [0, 1)$,*

$$\sup_{\mathbb{F},P,g,\Phi} \frac{\|J^* - J^{\tilde{\tau}}\|_\pi}{\|Q^* - \Pi Q^*\|_\pi} \geq \frac{1}{1-\alpha},$$

where $\tilde{\tau} = \inf\{t | g(x_t) \geq (\Pi Q^*)(x_t)\}$.

Proof: Without loss of generality, consider a two-state problem with

$$g = \begin{bmatrix} 1 \\ 1+\gamma \end{bmatrix}, \qquad P = \begin{bmatrix} 1-\delta & \delta \\ \epsilon\delta & 1-\epsilon\delta \end{bmatrix}, \qquad \Phi = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

for some positive scalars ϵ, δ, and γ. We will only consider cases where $\delta < (1-\alpha)/(\alpha\gamma)$. In such cases, it is always optimal to stop.

Some simple algebra gives us the continuation value function and its projection:

$$Q^* = \begin{bmatrix} \alpha(1+\delta\gamma) \\ \alpha(1+\gamma-\epsilon\delta\gamma) \end{bmatrix}, \qquad \Pi Q^* = \begin{bmatrix} \alpha + \frac{\alpha\gamma}{1+\epsilon} \\ \alpha + \frac{\alpha\gamma}{1+\epsilon} \end{bmatrix}.$$

For $\gamma > (1-\alpha)(1+\epsilon)/\alpha$, we have $(\Pi Q^*)(1) > 1$, and therefore $x_{\tilde{\tau}} = 2$ with certainty. The value function associated with $\tilde{\tau}$ is therefore

$$J^{\tilde{\tau}} = \begin{bmatrix} \frac{\alpha\delta(1+\gamma)}{1-\alpha(1-\delta)} \\ 1+\gamma \end{bmatrix}.$$

Some algebra leads to

$$\begin{aligned} \lim_{\delta \downarrow 0} \|Q^* - \Pi Q^*\|_\pi &= \alpha\gamma \left(\frac{\epsilon + \epsilon^2}{(1+\epsilon)^3} \right)^{1/2} \\ \lim_{\delta \downarrow 0} \|J^* - J^{\tilde{\tau}}\|_\pi &= \left(\frac{\epsilon}{1+\epsilon} \right)^{1/2}. \end{aligned}$$

It follows that

$$\lim_{\epsilon \downarrow 0} \lim_{\delta \downarrow 0} \frac{\|J^* - J^{\tilde{\tau}}\|_\pi}{\|Q^* - \Pi Q^*\|_\pi} = \frac{1}{\alpha\gamma}.$$

Since this works for any $\gamma > (1-\alpha)(1+\epsilon)/\alpha$, the result follows. ■

For each algorithm that computes weights, there is a value L^*. Lower values are more desirable with minimization of approximation error resulting in $L^* = 1$. Hence, L^* can be viewed as a measure of how competitive an algorithm is with one that minimizes approximation error.

Let us put in perspective the three objectives we have introduced and their relative merits. The first objective is to compute weights that optimize performance of the resulting greedy policy. We have shown that this is intractable, even when the underlying optimal stopping problem is tractable. The situation should only get worse in practical contexts where approximation methods are deployed, in which the

number of states is unmanageable and exact solution is intractable. The second objective is to minimize expected squared error with respect to an invariant measure. This is motivated by the fact that small squared error leads to good performance. Minimizing squared error would be tractable if the exact solution of the underlying optimal stopping problem is. However, for problems with unmanageably large state spaces, minimizing squared error appears to be intractable. The third objective is to minimize L^*, a measure of competitiveness relative to performance that would result from minimizing squared error. It is not clear whether minimizing L^* is tractable, but as we will see in the next section, approximate value iteration is guaranteed to yield $L^* \leq 2.17$. This means that performance delivered by this algorithm will be no more than a factor of 2.17 worse than what minimizing squared error would offer.

4 Approximate Value Iteration

Let us define a dynamic programming operator F for our problem by $FQ = \alpha P \max(g, Q)$ so that Bellman's Equation takes the form $Q^* = FQ^*$. The following lemma, adapted from [9], states that F is a contraction mapping with respect to $\|\cdot\|_\pi$. Among other things, this implies that F has a unique fixed point.

Lemma 4 *For any Q and $\overline{Q}$,*

$$\|FQ - F\overline{Q}\|_\pi \leq \alpha \|Q - \overline{Q}\|_\pi.$$

Proof: Note that for any scalars a, b, and c, $|\max(a, b) - \max(a, c)| \leq |b - c|$. It follows that

$$\begin{aligned} \|FQ - F\overline{Q}\|_\pi &= \|\alpha P \max(g, Q) - \alpha P \max(g, \overline{Q})\|_\pi \\ &\leq \alpha \| \max(g, Q) - \max(g, \overline{Q})\|_\pi \\ &\leq \alpha \|Q - \overline{Q}\|_\pi, \end{aligned}$$

where the first inequality follows from Lemma 3. ■

We consider computing a solution to an approximate version of Bellman's Equation

$$Q = \Pi F Q.$$

Note that the range of Π is the span of the basis functions, so any solution can be written as Φr for some r. Further, since Π is a projection with respect to $\|\cdot\|_\pi$, it is nonexpansive with respect to this norm

and therefore ΠF is a contraction. Hence, there is a unique solution to this approximate Bellman's Equation.

A variety of algorithms have been proposed for solving the equation $Q = \Pi FQ$. We refer to such algorithms as approximate value iteration algorithms. An initial analysis of this equation together with a solution algorithm were provided by [9]. Subsequent work designed new, more efficient, algorithms for solving the equation [2, 10, 12].

We now turn our attention to analysis of performance loss, and in particular, to bounding L^* for an algorithm that solves $Q = \Pi FQ$. We begin with three lemmas. The first bounds the difference that an application of F can make on the fixed point.

Lemma 5 *If $\tilde{Q} = \Pi F\tilde{Q}$ then*

$$\|F\tilde{Q} - \tilde{Q}\|_\pi \leq (1+\alpha)\|Q^* - \Pi Q^*\|_\pi.$$

Proof: Let $Q = \tilde{Q} + Q^* - \Pi Q^*$. Note that

$$\|\tilde{Q} - Q\|_\pi = \|Q^* - \Pi Q^*\|_\pi.$$

Note that $\Pi(F\tilde{Q} - Q) = 0$ whereas $\Pi(Q^* - Q) = Q^* - Q$. As such $F\tilde{Q} - Q$ and $Q^* - Q$ are orthogonal. We therefore have

$$\begin{aligned}
\|F\tilde{Q} - Q\|_\pi^2 &= \|F\tilde{Q} - Q^*\|_\pi^2 - \|Q^* - Q\|_\pi^2 \\
&\leq \alpha^2\|\tilde{Q} - Q^*\|_\pi^2 - \alpha^2\|Q^* - Q\|_\pi^2 \\
&= \alpha^2\|\tilde{Q} - Q\|_\pi^2 \\
&= \alpha^2\|Q^* - \Pi Q^*\|_\pi^2,
\end{aligned}$$

where the first and second equations make use of the Pythagorean theorem and the inequality follows from the fact that F is a contraction mapping with fixed point Q^*. It follows from the triangle inequality that

$$\begin{aligned}
\|F\tilde{Q} - \tilde{Q}\|_\pi &\leq \|F\tilde{Q} - Q\|_\pi + \|Q - \tilde{Q}\|_\pi \\
&\leq \alpha\|Q^* - \Pi Q^*\|_\pi + \|Q^* - \Pi Q^*\|_\pi \\
&= (1+\alpha)\|Q^* - \Pi Q^*\|_\pi.
\end{aligned}$$

■

The next lemma provides an error bound on the solution of $Q = \Pi FQ$.

Lemma 6 *If* $\tilde{Q} = \Pi F \tilde{Q}$ *then*

$$\|Q^* - \tilde{Q}\|_\pi \leq \frac{1}{\sqrt{1-\alpha^2}} \|Q^* - \Pi Q^*\|_\pi.$$

Proof: By the Pythagorean Theorem,

$$\begin{aligned} \|Q^* - \tilde{Q}\|_\pi^2 &= \|Q^* - \Pi Q^*\|_\pi^2 + \|\Pi Q^* - \tilde{Q}\|_\pi^2 \\ &= \|Q^* - \Pi Q^*\|_\pi^2 + \|\Pi F Q^* - \Pi F \tilde{Q}\|_\pi^2 \\ &\leq \|Q^* - \Pi Q^*\|_\pi^2 + \alpha^2 \|Q^* - \tilde{Q}\|_\pi^2. \end{aligned}$$

The result follows. ■

By Lemma 4, if $\tilde{Q} = \Pi F \tilde{Q}$, applying F to $\tilde{Q}$ brings it a factor of α closer to Q^*. The bound from our last lemma establishes that F simultaneously brings $\tilde{Q}$ a factor of α closer to $Q^{\tilde{\tau}}$, where $\tilde{\tau}$ is greedy with respect to $\tilde{Q}$.

Lemma 7 *If* $\tilde{Q} = \Pi F \tilde{Q}$ *and* $\tilde{\tau} = \inf\{t | g(x_t) \geq \tilde{Q}(x_t)\}$ *then*

$$\|F\tilde{Q} - Q^{\tilde{\tau}}\|_\pi \leq \alpha \|\tilde{Q} - Q^{\tilde{\tau}}\|_\pi.$$

Proof: Let $\tilde{V} = \max(g, \tilde{Q})$ and

$$V^{\tilde{\tau}}(x) = \begin{cases} g(x) & \text{if } g(x) \geq \tilde{Q}(x) \\ Q^{\tilde{\tau}}(x) & \text{otherwise.} \end{cases}$$

Note that $F\tilde{Q} = \alpha P \tilde{V}$ and $Q^{\tilde{\tau}} = \alpha P V^{\tilde{\tau}}$. We then have

$$\begin{aligned} \|F\tilde{Q} - Q^{\tilde{\tau}}\|_\pi &= \|\alpha P \tilde{V} - \alpha P V^{\tilde{\tau}}\|_\pi \\ &\leq \alpha \|\tilde{V} - V^{\tilde{\tau}}\|_\pi \\ &\leq \alpha \|\tilde{Q} - Q^{\tilde{\tau}}\|_\pi. \end{aligned}$$

■

Finally, our main result establishes that L^* for approximate value iteration is no greater than 2.17.

Theorem 4 *If* $\tilde{Q} = \Pi F \tilde{Q}$ *and* $\tilde{\tau} = \inf\{t | g(x_t) \geq \tilde{Q}(x_t)\}$ *then*

$$\|J^* - J^{\tilde{\tau}}\|_\pi \leq \frac{2.17}{1-\alpha} \|Q^* - \Pi Q^*\|_\pi.$$

Proof: By the triangle inequality and Lemma 7,

$$\begin{aligned}\|\tilde{Q} - Q^{\tilde{\tau}}\|_\pi &\leq \|\tilde{Q} - F\tilde{Q}\|_\pi + \|F\tilde{Q} - Q^{\tilde{\tau}}\|_\pi \\ &\leq \|\tilde{Q} - F\tilde{Q}\|_\pi + \alpha\|\tilde{Q} - Q^{\tilde{\tau}}\|_\pi,\end{aligned}$$

and it follows that

$$\|\tilde{Q} - Q^{\tilde{\tau}}\|_\pi \leq \frac{1}{1-\alpha}\|F\tilde{Q} - \tilde{Q}\|_\pi.$$

By the triangle inequality,

$$\begin{aligned}\|Q^* - Q^{\tilde{\tau}}\|_\pi &\leq \|Q^* - F\tilde{Q}\|_\pi + \|F\tilde{Q} - Q^{\tilde{\tau}}\|_\pi \\ &\leq \alpha\|Q^* - \tilde{Q}\|_\pi + \alpha\|\tilde{Q} - Q^{\tilde{\tau}}\|_\pi \\ &\leq \frac{\alpha}{\sqrt{1-\alpha^2}}\|Q^* - \Pi Q^*\|_\pi + \frac{\alpha}{1-\alpha}\|F\tilde{Q} - \tilde{Q}\|_\pi \\ &\leq \alpha\left(\frac{1+\alpha}{1-\alpha} + \frac{1}{\sqrt{1-\alpha^2}}\right)\|Q^* - \Pi Q^*\|_\pi,\end{aligned}$$

where the third inequality follows from Lemmas 6 and the last inequality follows from Lemma 5. By Lemmas 2 and 6,

$$\begin{aligned}\|J^* - J^{\tilde{\tau}}\|_\pi &\leq \|Q^* - \tilde{Q}\|_\pi + \|Q^* - Q^{\tilde{\tau}}\|_\pi \\ &\leq \frac{1}{\sqrt{1-\alpha^2}}\|Q^* - \Pi Q^*\|_\pi + \alpha\left(\frac{1+\alpha}{1-\alpha} + \frac{1}{\sqrt{1-\alpha^2}}\right)\|Q^* - \Pi Q^*\|_\pi \\ &= \frac{\sqrt{1-\alpha^2} + \alpha + \alpha^2}{1-\alpha}\|Q^* - \Pi Q^*\|_\pi.\end{aligned}$$

To complete the proof, we will show that $\sqrt{1-\alpha^2} + \alpha + \alpha^2 < 2.17$ for all $\alpha \in [0,1]$. Let $\beta = \sqrt{1-\alpha^2}$, so $\alpha = \sqrt{1-\beta^2}$. Then, $\sqrt{1-\alpha^2}+\alpha+\alpha^2 = \beta-\beta^2+1+\sqrt{1-\beta^2}$. Each term in this expression is concave over $\beta \in [0,1]$. The derivative of $\sqrt{1-\beta^2}$ is $-\beta/\sqrt{1-\beta^2}$. Hence, for any $\beta_0 \in [0,1]$,

$$\beta - \beta^2 + 1 + \sqrt{1-\beta^2} \leq \beta - \beta^2 + 1 + \sqrt{1-\beta_0^2} - \frac{\beta_0}{\sqrt{1-\beta_0^2}}(\beta - \beta_0).$$

We will maximize the upper bound:

$$\max_{\beta\in[0,1]}\left(\beta - \beta^2 + 1 + \sqrt{1-\beta_0^2} - \frac{\beta_0}{\sqrt{1-\beta_0^2}}(\beta - \beta_0)\right).$$

The optimal solution is

$$\beta^* = \frac{1}{2} - \frac{\beta_0}{2\sqrt{1-\beta_0^2}}.$$

With $\beta_0 = 1/3$, we obtain $\beta^* = (2\sqrt{2}-1)/(4\sqrt{2})$ and

$$\beta^* - (\beta^*)^2 + 1 + \sqrt{1-\beta_0^2} - \frac{\beta_0}{\sqrt{1-\beta_0^2}}(\beta^* - \beta_0) \in (2.165, 2.166).$$

The result follows. ■

To put this result in context, let us compare it to the earlier performance loss bound from [9], which assumes that π is a probability measure and takes the form

$$\int \pi(dx)(J^*(x) - J^{\tilde{\tau}}(x)) \leq \frac{2}{(1-\alpha)\sqrt{1-\alpha^2}} \|Q^* - \Pi Q^*\|_\pi.$$

The left hand side is different from the bound of Theorem 4. However, if π is a probability measure, $\int \pi(dx)(J^*(x) - J^{\tilde{\tau}}(x)) \leq \|J^* - J^{\tilde{\tau}}\|_\pi$, so the theorem implies

$$\int \pi(dx)(J^*(x) - J^{\tilde{\tau}}(x)) \leq \frac{2.17}{1-\alpha} \|Q^* - \Pi Q^*\|_\pi.$$

The left hand side given by our new theorem exhibits a more graceful dependence on α.

It is also worth mentioning that the line of analysis used in [9] includes a proof that

$$\|Q^* - Q^{\tilde{\tau}}\|_\pi \leq \frac{2\alpha}{(1-\alpha)\sqrt{1-\alpha^2}} \|Q^* - \Pi Q^*\|_\pi.$$

Combining this with Lemmas 2 and 6 yields

$$\begin{aligned}
\|J^* - J^{\tilde{\tau}}\|_\pi &\leq \|Q^* - \tilde{Q}\|_\pi + \|Q^* - Q^{\tilde{\tau}}\|_\pi \\
&\leq \frac{1}{\sqrt{1-\alpha^2}} \|Q^* - \Pi Q^*\|_\pi + \frac{2\alpha}{(1-\alpha)\sqrt{1-\alpha^2}} \|Q^* - \Pi Q^*\|_\pi \\
&= \frac{(1-\alpha+2\alpha)/\sqrt{1-\alpha^2}}{1-\alpha} \|Q^* - \Pi Q^*\|_\pi.
\end{aligned}$$

Unlike Theorem 4, this result does not imply finiteness of L^*, since $(1-\alpha+2\alpha)/\sqrt{1-\alpha^2}$ is unbounded.

5 Conclusion

What we have presented offers a framework for comparing algorithms that compute basis function weights. It entails bounding the value L^* associated with each particular algorithm, which is the minimal value of L such that

$$\|J^* - J^{\tilde{\tau}}\|_\pi \leq \frac{L}{1-\alpha}\|Q^* - \Pi Q^*\|_\pi,$$

for all problem instances (optimal stopping problems and sets of basis functions), where $\tilde{\tau}$ is the resulting greedy policy. An algorithm that computes weights associated with the projection of Q^* onto the span of basis functions has an L^* of 1. However, there is no known efficient algorithm that does this.

We have established that approximate value iteration leads to an L^* no greater than 2.17. This result applies to variations of the algorithm developed in [9, 2, 10, 12]. It would be interesting to characterize L^* values for alternative algorithms that have been proposed in the literature, such as those based on policy iteration [6, 3] or linear programming [1]. Like the aforementioned variations of approximate value iteration, these algorithms offer efficient methods for computing weights, and it would be interesting to better understand how the algorithms fare relative to one another. Analysis of how L^* depends on algorithm features can potentially guide algorithm design.

Though L^* provides an interesting metric by which we can assess algorithms, it differs from our end objective, which is to maximize expected payoff in the optimal stopping problem. As we have shown, computing weights that optimize expected payoff within a logarithmic factor is NP-hard. We consider L^* because of the analysis we are able to carry out. There may be alternative analytically tractable metrics that assess weight computing algorithms as effectively or more so, though that remains a topic for future investigation.

Our framework focusses on the performance resulting from weights computed for a pre-selected set of basis functions. It does not factor in compute time beyond requiring tractability. One might consider a framework where the allocation of compute cycles is more carefully accounted for, and possibly, the number of basis functions is variable, as in [5]. Here the goal could be to design algorithms that strike a desirable trade-off between compute time, which is influenced by algorithm design choices and the number of basis functions, and policy performance.

Acknowledgments

This paper was written while the author was visiting the Faculty of Commerce and Accountancy at Chulalongkorn University and supported by the Chin Sophonpanich Foundation Fund. The author thanks Amin Saberi for helpful discussions on complexity theory and the key idea for proving Lemma 1.

References

[1] V. S. Borkar, J. Pinto, and T. Prabhu. A new learning algorithm for optimal stopping. preprint, 2001.

[2] D. S. Choi and B. Van Roy. A generalized Kalman filter for fixed point approximation and efficient temporal-difference learning. *Discrete Event Dynamic Systems*, 16(2), 2006.

[3] E. Clément, D. Lamberton, and P. Protter. An analysis of a least squares regression algorithm for American option pricing. *Finance and Stochastics*, 6:449–471, 2002.

[4] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, Berlin, 2003.

[5] P. Glasserman and B. Yu. Number of paths versus number of basis functions in American option pricing. *Annals of Applied Probability*, 14(4):2090–2119, 2004.

[6] F. Longstaff and E. S. Schwartz. Valuing American options by simulation: A simple least-square approach. *Review of Financial Studies*, 14(1):113–147, 2001.

[7] F. A. Longstaff, P. Santa-Clara, and E. S. Schwartz. Throwing away a billion dollars: the cost of suboptimal exercise strategies in the swaptions market. *Journal of Financial Economics*, 62(1):39–66, 2001.

[8] R. Raz and S. Safra. A sub-constant error-probability low-degree test, and sub-constant error-probability PCP characterization of NP. In *Proceedings of the 29th Annual ACM Symposium on the Theory of Computatins*, pages 475–484. ACM, 1997.

[9] J. N. Tsitsiklis and B. Van Roy. Optimal stopping of Markov processes: Hilbert space theory, approximation algorithms, and an application to pricing high-dimensional financial derivatives.

IEEE Transactions on Automatic Control, 44(10):1840–1851, 1999.

[10] J. N. Tsitsiklis and B. Van Roy. Regression-based methods for pricing complex American-style options. *IEEE Transactions on Neural Networks*, 12(4):694–703, 2001.

[11] B. Van Roy. *Learning and Value Function Approximation in Complex Decision Processes*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1998.

[12] H. Yu and D. P. Bertsekas. A least squares Q-learning algorithm for optimal stopping problems. Technical Report LIDS-P-2731, MIT Laboratory for Information and Decision Systems, 2006.