Online Supplement to "Pathwise Optimization for Optimal Stopping Problems"

Vijay V. Desai
Industrial Engineering and Operations Research
Columbia University
email: vvd2101@columbia.edu

Vivek F. Farias
Sloan School of Management
Massachusetts Institute of Technology
email: vivekf@mit.edu

Ciamac C. Moallemi
Graduate School of Business
Columbia University
email: ciamac@gsb.columbia.edu

Initial Version: August 20, 2010
Current Revision: November 22, 2011

A. Proofs

The following elementary fact will be helpful in the proofs that follow:

Fact 1. *If $y, y' \in \mathbb{R}^N$ are two sequences of real numbers, then*

$$\max_s y_s - \max_s y'_s \leq \max_s \; |y_s - y'_s|.$$

By symmetry, it holds that

$$|\max_s y_s - \max_s y'_s| \leq \max_s \; |y_s - y'_s|.$$

A.1. Proof of Lemma 1

Lemma 1 (Martingale Duality).

(i) (Weak Duality) For any $J \in \mathcal{P}$ and all $x \in \mathcal{X}$ and $t \in \mathcal{T}$, $J^*_t(x) \leq F_t J(x)$.

(ii) (Strong Duality) For all $x \in \mathcal{X}$ and $t \in \mathcal{T}$, $J^*(x)_t = F_t J^*(x)$.

Proof. (i) Note that

$$J^*_t(x_t) = \sup_{\tau_t} \; \mathsf{E}\left[\alpha^{\tau_t - t} g(x_{\tau_t}) \;\middle|\; x_t \right] \tag{A.1}$$

$$= \sup_{\tau_t} \; \mathsf{E}\left[\alpha^{\tau_t - t} g(x_{\tau_t}) - \sum_{p=t+1}^{\tau_t} \alpha^{p-t} (\Delta J)(x_p, x_{p-1}) \;\middle|\; x_t \right] \tag{A.2}$$

$$\leq \mathsf{E}\left[\max_{t \leq s \leq d} \; \alpha^{s-t} g(x_s) - \sum_{p=t+1}^{s} \alpha^{p-t} (\Delta J)(x_p, x_{p-1}) \;\middle|\; x_t \right]. \tag{A.3}$$

Here, in (A.1), τ_t is a stopping time that takes values in the set $\{t, t+1, \ldots, d\}$. (A.2) follows from the optimal sampling theorem for martingales. (A.3) follows from the fact that stopping times are non-anticipatory, and hence the objective value can only be increased by allowing policies with access to the entire sample path.

(ii) From (i) we know that $F_t J^*(x_t) \geq J_t^*(x_t)$. To see the opposite inequality,

$$\begin{aligned}
F_t J^*(x_t) &= \mathsf{E}\left[\max_{t \leq s \leq d} \ \alpha^{s-t} g(x_s) - \sum_{p=t+1}^{s} \alpha^{p-t} \, (\Delta J^*)(x_p, x_{p-1}) \;\middle|\; x_t \right] \\
&= \mathsf{E}\left[\max_{t \leq s \leq d} \ \alpha^{s-t} g(x_s) - \sum_{p=t+1}^{s} \alpha^{p-t} \left(J_p^*(x_p) - \mathsf{E}[J_p^*(x_p) | x_{p-1}] \right) \;\middle|\; x_t \right] \\
&= \mathsf{E}\left[\max_{t \leq s \leq d} \ \alpha^{s-t} g(x_s) - \alpha^{s-t} J_s^*(x_s) + J_t^*(x_t) \right. \\
&\qquad \left. + \sum_{p=t+1}^{s} \alpha^{p-t-1} \left(\alpha \mathsf{E}[J_p^*(x_p) | x_{p-1}] - J_{p-1}^*(x_{p-1}) \right) \;\middle|\; x_t \right] \\
&\leq J_t^*(x_t)
\end{aligned}$$

The last inequality follows from the Bellman equation (1). ■

A.2. Proof of Theorem 1

Theorem 1. *Let $\mathcal{N} \subset \mathbb{R}^K$ be a compact set. Fix an initial state x_0 and $\epsilon > 0$. Then, almost surely, if S is sufficiently large, for all I sufficiently large,*

$$\left| \min_{r \in \mathcal{N}} \ F_0 \Phi r(x_0) - \min_{r \in \mathcal{N}} \ \hat{F}_0^{S,I} \Phi r(x_0) \right| \leq \epsilon.$$

Proof. Given $\epsilon', \delta' > 0$, define a *finite* set $\mathcal{R} \subset \mathcal{N}$ such that for all $r \in \mathcal{N}$, there exists $r' \in \mathcal{R}$ with $\|r - r'\|_\infty < \epsilon'$. The existence of $\mathcal{R}$ is guaranteed by the compactness of $\mathcal{N}$.

For any element $r \in \mathcal{N}$ and let $r' \in \mathcal{R}$ be such that $\|r - r'\|_\infty < \epsilon'$. By triangle inequality, we have

$$\begin{aligned}
\left| F_0 \Phi r(x_0) - \hat{F}_0^{S,I} \Phi r(x_0) \right| \leq & \left| F_0 \Phi r(x_0) - F_0 \Phi r'(x_0) \right| + \left| F_0 \Phi r'(x_0) - \hat{F}_0^{S,I} \Phi r'(x_0) \right| \\
& + \left| \hat{F}_0^{S,I} \Phi r'(x_0) - \hat{F}_0^{S,I} \Phi r(x_0) \right|.
\end{aligned} \tag{A.4}$$

We bound each of the quantities on the right by using Lemma 8, which is established below, to guarantee the choice of (S, I) so that

$$\left| F_0 \Phi r(x_0) - \hat{F}_0^{S,I} \Phi r(x_0) \right| \leq L\epsilon' + \epsilon' + (L + \delta')\epsilon'.$$

Since $\epsilon', \delta' > 0$ are arbitrary, the result follows. ■

Lemma 8. *Fix an initial state* $x_0 \in \mathcal{X}$.

(i) $F_0\Phi r(x_0)$ *is a Lipschitz function of* $r \in \mathbb{R}^K$, *i.e.,*

$$\left|F_0\Phi r(x_0) - F_0\Phi r'(x_0)\right| \leq L\|r - r'\|_\infty, \quad \forall\ r, r' \in \mathbb{R}^K,$$

where we denote the associated Lipschitz constant by L.

(ii) Fix $\epsilon, \delta > 0$ *and suppose that* $\mathcal{R} \subset \mathbb{R}^K$ *is a finite set. Then, almost surely, if* S *is sufficiently large, for all* I *sufficiently large, we have:*

(a) For all $r \in \mathcal{R}$, $\left|\hat{F}_0^{S,I}\Phi r(x_0) - F_0\Phi r(x_0)\right| \leq \epsilon$.

(b) $\hat{F}_0^{S,I}\Phi r(x_0)$ *is a Lipschitz function of* $r \in \mathbb{R}^K$ *with Lipschitz constant* $L + \delta$.

Proof. (i) Using Fact 1, the triangle inequality, and Jensen's inequality, we have that, for $r, r' \in \mathbb{R}^K$,

$$\begin{aligned}
&\left|F_0\Phi r(x_0) - F_0\Phi r'(x_0)\right| \\
&\leq \mathsf{E}\left[\max_{0\leq s\leq d} \left|\sum_{p=1}^{s} \alpha^p \Big((\Phi r)_p(x_p) - (\Phi r')_p(x_p) + \mathsf{E}\big[(\Phi r)_p(x_p)|x_{p-1}\big] - \mathsf{E}\big[(\Phi r')_p(x_p)|x_{p-1}\big]\Big)\right| \;\middle|\; x_0\right] \\
&\leq \mathsf{E}\left[\sum_{p=1}^{d} \alpha^p \Big(\left|(\Phi r)_p(x_p) - (\Phi r')_p(x_p)\right| + \left|\mathsf{E}\big[(\Phi r)_p(x_p)|x_{p-1}\big] - \mathsf{E}\big[(\Phi r')_p(x_p)|x_{p-1}\big]\right|\Big) \;\middle|\; x_0\right] \\
&\leq 2\mathsf{E}\left[\sum_{p=1}^{d} \left|(\Phi r)_p(x_p) - (\Phi r')_p(x_p)\right| \;\middle|\; x_0\right] \leq L\|r - r'\|_\infty,
\end{aligned}$$

where

$$L \triangleq 2\sum_{p=1}^{d}\sum_{\ell=1}^{K} \mathsf{E}\left[\left|\phi_\ell(x_p, p)\right| \;\middle|\; x_0\right] < \infty.$$

(ii-a) Fix $r \in \mathcal{R}$. By the triangle inequality,

$$\left|\hat{F}_0^{S,I}\Phi r(x_0) - F_0\Phi r(x_0)\right| \leq \left|\hat{F}_0^{S,I}\Phi r(x_0) - \hat{F}_0^{S}\Phi r(x_0)\right| + \left|\hat{F}_0^{S}\Phi r(x_0) - F_0\Phi r(x_0)\right|. \tag{A.5}$$

By the strong law of large numbers, almost surely, for all S sufficiently large, we have that

$$\left|\hat{F}_0^{S}\Phi r(x_0) - F_0\Phi r(x_0)\right| \leq \frac{\epsilon}{2}, \tag{A.6}$$

$$\left|\frac{1}{S}\sum_{i=1}^{S}\sum_{p=1}^{d}\sum_{\ell=1}^{K}\left|\phi_\ell(x_p^{(i)}, p)\right| - \frac{L}{2}\right| \leq \frac{\delta}{2}. \tag{A.7}$$

Now, using Fact 1 and the triangle inequality,

$$\left|\hat{F}_0^{S,I}\Phi r(x_0) - \hat{F}_0^{S}\Phi r(x_0)\right| \leq \frac{1}{S}\sum_{i=1}^{S}\sum_{p=1}^{d}\left|\mathsf{E}\left[(\Phi r)_p(x_p)\middle|\, x_{p-1}^{(i)}\right] - \frac{1}{I}\sum_{j=1}^{I}(\Phi r)_p(x_p^{(i,j)})\right|. \tag{A.8}$$

Suppose that S is sufficiently large so that (A.6)–(A.7) hold. Using the strong law of large numbers, almost surely, for all I sufficiently large, we have that

$$\left|\frac{1}{I}\sum_{j=1}^{I}(\Phi r)_p(x_p^{(i,j)}) - \mathsf{E}\left[(\Phi r)_p(x_p)|\, x_{p-1}^{(i)}\right]\right| \leq \frac{\epsilon}{2d}, \quad \forall\ 1 \leq i \leq S, \tag{A.9}$$

$$\left|\frac{1}{I}\sum_{j=1}^{I}\sum_{p=1}^{d}\sum_{\ell=1}^{K}\left|\phi_\ell(x_p^{(i,j)}, p)\right| - \frac{L}{2}\right| \leq \frac{\delta}{2}, \quad \forall\ 1 \leq i \leq S. \tag{A.10}$$

Equations (A.8) and (A.9) together imply that

$$\left|\hat{F}_0^{S,I}\Phi r(x_0) - \hat{F}_0^{S}\Phi r(x_0)\right| \leq \frac{\epsilon}{2}. \tag{A.11}$$

Using (A.5), (A.6) and (A.11), we obtain the result for a fixed r. Since $\mathcal{R}$ is a finite set, S and I can be chosen sufficiently large so that the result holds for all $r \in \mathcal{R}$.

(ii-b) The result holds using the same argument as in part (i), along with the choice of (S, I) from part (ii-a) that guarantees (A.7) and (A.10). ■

A.3. Proofs of Section 5

Lemma 5. *Suppose that the state space $\mathcal{X}$ is finite. Then,*

$$\lambda(P) = \sqrt{\rho(I - P^*P)},$$

where $\rho(\cdot)$ is the spectral radius. Further, if P is time-reversible (i.e., if $P = P^*$), then

$$\lambda(P) = \sqrt{\rho(I - P^2)} \leq \sqrt{2\rho(I - P)}.$$

Proof. Note that, from (16),

$$\begin{aligned}
\mathsf{E}_\pi\left[\mathrm{Var}\left(J(x_1) \mid x_0\right)\right] &= \mathrm{Var}_\pi(J) - \mathrm{Var}_\pi(PJ) \\
&= \langle J, J\rangle_\pi - (\mathsf{E}_\pi[J(x_0)])^2 - \langle PJ, PJ\rangle_\pi + (\mathsf{E}_\pi[PJ(x_0)])^2 \\
&= \langle J, J\rangle_\pi - \langle PJ, PJ\rangle_\pi = \langle J, J\rangle_\pi - \langle J, P^*PJ\rangle_\pi = \langle J, (I - P^*P)J\rangle_\pi.
\end{aligned}$$

Observe that $I - P^*P$ is self-adjoint, and hence must have real eigenvalues. Let $\sigma_{\min}$ and $\sigma_{\max}$ be the smallest and largest eigenvalues, respectively. By the Courant-Fischer variational characterization of eigenvalues,

$$\begin{aligned}
\sigma_{\max} = \sup_{J\in\mathcal{P},\ \|J\|_{2,\pi}=1} \langle J, (I - P^*P)J\rangle_\pi &= \sup_{J\in\mathcal{P},\ \|J\|_{2,\pi}=1} \langle J, J\rangle_\pi - \langle J, P^*PJ\rangle_\pi \\
= 1 - \inf_{J\in\mathcal{P},\ \|J\|_{2,\pi}=1} \langle PJ, PJ\rangle_\pi &= 1 - \inf_{J\in\mathcal{P},\ \|J\|_{2,\pi}=1} \|PJ\|_{2,\pi}^2 \leq 1.
\end{aligned} \tag{A.12}$$

Similarly,

$$\text{(A.13)} \qquad \sigma_{\min} = \inf_{J\in\mathcal{P},\ \|J\|_{2,\pi}=1} \langle J, (I - P^*P)J\rangle_\pi = 1 - \sup_{J\in\mathcal{P},\ \|J\|_{2,\pi}=1} \|PJ\|_{2,\pi}.$$

Now, by Jensen's inequality and the fact that π is the stationary distribution of P,

$$\|PJ\|_{2,\pi}^2 = \mathsf{E}_\pi\left[\left(\mathsf{E}\left[J(x_1)|x_0\right]\right)^2\right] \leq \mathsf{E}_\pi\left[J(x_1)^2\right] = \|J\|_{2,\pi}^2.$$

That is, P is a non-expansive under the $\|\cdot\|_{2,\pi}$ norm. Combining this fact with (A.12)–(A.13), we have that $0 \leq \sigma_{\min} \leq \sigma_{\max} \leq 1$. Then, $\rho(I - P^*P) = \max\left(|\sigma_{\min}|, |\sigma_{\max}|\right) = \sigma_{\max}$. However, observe that from (A.12), $\lambda(P)^2 = \sigma_{\max}$. The result follows.

For the second part, suppose that $\zeta_1 \leq \zeta_2 \leq \cdots \leq \zeta_{|\mathcal{X}|}$ are the eigenvalues of the self-adjoint matrix P. By the same arguments as in before, $0 \leq \zeta_i \leq 1$ for each i. Then,

$$\rho(I - P^2) = \max_i\ 1 - \zeta_i^2 = \max_i\ (1-\zeta_i)(1+\zeta_i) \leq \max_i\ 2(1-\zeta_i) = 2\rho(I - P).$$

■

Lemma 6. *For any pair of functions $J, J' \in \mathcal{P}$,*

$$\|FJ - FJ'\|_{2,\pi} \leq \frac{R(\alpha)\alpha}{\sqrt{1-\alpha}}\, \lambda(P)\, \sqrt{\mathrm{Var}_\pi(J - J')},$$

where $R\colon [0,1) \to \left[1, \sqrt{5/2}\right]$ is a bounded function given by

$$R(\alpha) \triangleq \min\left\{\frac{1}{\sqrt{1-\alpha}}, \frac{2}{\sqrt{1+\alpha}}\right\}.$$

Proof. We can apply Fact 1 and the monotone convergence theorem to the pathwise maximization in the F operator to obtain that, for all $x_0 \in \mathcal{X}$,

$$FJ(x_0) - FJ'(x_0) \leq \mathsf{E}\left[\sup_{s\geq 0}\ \left|\sum_{t=1}^{s} \alpha^t\left(\Delta J(x_t, x_{t-1}) - \Delta J'(x_t, x_{t-1})\right)\right|\ \middle|\ x_0\right].$$

By symmetry,

$$|FJ(x_0) - FJ'(x_0)| \leq \mathsf{E}\left[\sup_{s\geq 0}\ \left|\sum_{t=1}^{s} \alpha^t\left(\Delta J(x_t, x_{t-1}) - \Delta J'(x_t, x_{t-1})\right)\right|\ \middle|\ x_0\right].$$

Using Jensen's inequality,

$$
\begin{aligned}
\text{(A.14)} \qquad |FJ(x_0) - FJ'(x_0)|^2 &\leq \mathsf{E}\left[\sup_{s\geq 0} \left| \sum_{t=1}^{s} \alpha^t \left(\Delta J(x_t, x_{t-1}) - \Delta J'(x_t, x_{t-1})\right) \right|^2 \,\middle|\, x_0 \right] \\
&\leq \mathsf{E}\left[\left(\sum_{t=1}^{\infty} \alpha^t \left|\Delta J(x_t, x_{t-1}) - \Delta J'(x_t, x_{t-1})\right| \right)^2 \,\middle|\, x_0 \right].
\end{aligned}
$$

Taking an expectation over x_0 and again applying Jensen's inequality,

$$
\begin{aligned}
\text{(A.15)} \qquad \|FJ - FJ'\|_{2,\pi}^2 &\leq \left(\frac{\alpha}{1-\alpha}\right)^2 \mathsf{E}_\pi\left[\left(\frac{1-\alpha}{\alpha} \sum_{t=1}^{\infty} \alpha^t \left|\Delta J(x_t, x_{t-1}) - \Delta J'(x_t, x_{t-1})\right| \right)^2 \right] \\
&\leq \left(\frac{\alpha}{1-\alpha}\right)^2 \mathsf{E}_\pi\left[\frac{1-\alpha}{\alpha} \sum_{t=1}^{\infty} \alpha^t \left|\Delta J(x_t, x_{t-1}) - \Delta J'(x_t, x_{t-1})\right|^2 \right] \\
&= \left(\frac{\alpha}{1-\alpha}\right)^2 \|\Delta J - \Delta J'\|_{2,\pi}^2 .
\end{aligned}
$$

Here, the norm in the final equality is defined in Lemma 4, and we have used the fact that π is the stationary distribution.

On the other hand, following Chen and Glasserman (2007), Doob's maximal quadratic inequality and the orthogonality of martingale differences imply that, for every time $T \geq 1$,

$$
\begin{aligned}
&\mathsf{E}_\pi\left[\sup_{0\leq s\leq T} \left| \sum_{t=1}^{s} \alpha^t \left(\Delta J(x_t, x_{t-1}) - \Delta J'(x_t, x_{t-1})\right) \right|^2 \right] \\
&\leq 4\mathsf{E}_\pi\left[\left| \sum_{t=1}^{T} \alpha^t \left(\Delta J(x_t, x_{t-1}) - \Delta J'(x_t, x_{t-1})\right) \right|^2 \right] \\
&\leq 4\mathsf{E}_\pi\left[\sum_{t=1}^{T} \alpha^{2t} \left|\Delta J(x_t, x_{t-1}) - \Delta J'(x_t, x_{t-1})\right|^2 \right] \\
&= 4\alpha^2 \frac{1-\alpha^{2T-1}}{1-\alpha^2} \|\Delta J - \Delta J'\|_{2,\pi}^2 .
\end{aligned}
$$

Using the monotone convergence theorem to take the limit as $T \to \infty$ and comparing with (A.14), we have that

$$
\text{(A.16)} \qquad \|FJ - FJ'\|_{2,\pi}^2 \leq \frac{4\alpha^2}{1-\alpha^2} \|\Delta J - \Delta J'\|_{2,\pi}^2 .
$$

Combining the upper bounds of (A.15) and (A.16), we have that

$$
\text{(A.17)} \qquad \|FJ - FJ'\|_{2,\pi} \leq \frac{R(\alpha)\alpha}{\sqrt{1-\alpha}} \|\Delta J - \Delta J'\|_{2,\pi} .
$$

Applying Lemma 4, the result follows. ∎

B. The Non-stationary Case

In this section, we will outline a finite horizon and non-stationary version of the theoretical results presented in Section 5. Our setting here follows that of Section 2: Assume a state space $\mathcal{X} \subset \mathbb{R}^n$. Consider a discrete-time Markov chain with state $x_t \in \mathcal{X}$ at each time $t \in \mathcal{T} \triangleq \{0, 1, \ldots, d\}$. Without loss of generality, assume that the transition probabilities are time-invariant, and denote by P the transition kernel of the chain. Let $\mathcal{F} \triangleq \{\mathcal{F}_t\}$ be the natural filtration generated by the process, i.e., for each time t, $\mathcal{F}_t \triangleq \sigma(x_0, x_1, \ldots, x_t)$.

For this section, we assume a *fixed* initial state $x_0 \in \mathcal{X}$, we are interested in the stopping problem over the finite time horizon $\mathcal{T}$ with payoff function $g \colon \mathcal{X} \to \mathbb{R}$. Further, without loss of generality, we will assume that $\alpha = 1$, i.e., that the problem is finite-horizon and undiscounted.

For each $t \in \mathcal{T}$, define $\mathcal{S}_t$ to be the set of measurable functions $J_t \colon \mathcal{X} \to \mathbb{R}$ with $\mathsf{E}[J_t(x_t)^2 \mid x_0] < \infty$. Assume that $g \in \mathcal{S}_t$, for all t. Define $\mathcal{P}$ to be the set of functions $J \colon \mathcal{X} \times \mathcal{T} \to \mathbb{R}$ such that, for each $t \in \mathcal{T}$, $J_t \triangleq J(\cdot, t)$ is contained in the set $\mathcal{S}_t$. In other words, $\mathcal{P}$ is the set of Markovian processes (i.e., time-dependent functionals of the state) that possess second moments.

Given $J \in \mathcal{P}$, define

$$(\Delta J)_t \triangleq \begin{cases} 0 & \text{if } t = 0, \\ J_t(x_t) - \mathsf{E}[J_t(x_t) | x_{t-1}] & \text{otherwise,} \end{cases}$$

for all $t \in \mathcal{T}$. Note that ΔJ is a martingale difference process.

Define the predictability of the Markov chain by

$$\lambda(P) \triangleq \max_{1 \leq t \leq d} \sup_{J_t \in \mathcal{S}_t, J_t \neq \mathbf{0}} \left(\frac{\mathsf{E}\left[\operatorname{Var}\left(J_t(x_t) \mid x_{t-1}\right) \mid x_0\right]}{\operatorname{Var}\left(J_t(x_t) \mid x_0\right)} \right)^{1/2}. \tag{B.1}$$

Applying the law of total variance as in (16), it is easy to see that $\lambda(P) \in [0, 1]$. Analogous to (15), $\lambda(P)$ captures the worst case uncertainty in $J(x_t)$ conditioned on the previous state x_t, relative to the prior uncertainty (i.e., the uncertainty conditioned only on the initial state x_0), over all functionals $J_t \in \mathcal{S}_t$ and all times $1 \leq t \leq d$. As before, when $\lambda(P) \approx 0$, the previous state reveals significant information on the subsequent value of any functional, hence we interpret the Markov chain as predictable.

The following lemma is analogous to Lemma 4, and provides a bound on the operator norm of the martingale difference operator Δ:

Lemma 9. *Given functions $J, J' \in \mathcal{P}$, define a distance between the martingale differences ΔJ, $\Delta J'$ by*

$$\|\Delta J - \Delta J'\|_{2,x_0} \triangleq \sqrt{\frac{1}{d} \mathsf{E}\left[\sum_{t=1}^{d} |\Delta J_t - \Delta J_t'|^2 \;\middle|\; x_0 \right]}.$$

Then,

$$\|\Delta J - \Delta J'\|_{2,x_0} \leq \lambda(P) \sqrt{\operatorname{Var}_{x_0}(J - J')},$$

where

$$\overline{\mathrm{Var}}_{x_0}(J-J') \triangleq \frac{1}{d}\sum_{t=1}^{d} \mathrm{Var}\left(J_t(x_t) - J'_t(x_t) \mid x_0\right)$$

is the average variance between the processes J and J' over the time horizon d.

Proof. Set $W \triangleq J - J'$, and observe that

$$\|\Delta W\|_{2,x_0}^2 = \frac{1}{d}\mathsf{E}\left[\sum_{t=1}^{d} \left|W_t(x_t) - \mathsf{E}[W_t(x_t) \mid x_{t-1}]\right|^2 \;\middle|\; x_0\right] = \frac{1}{d}\mathsf{E}\left[\sum_{t=1}^{d} \mathrm{Var}\left(W_t(x_t) \mid x_{t-1}\right) \;\middle|\; x_0\right]$$
$$\leq \frac{\lambda(P)^2}{d}\sum_{t=1}^{d} \mathrm{Var}\left(W_t(x_t) \mid x_0\right) = \lambda(P)^2 \overline{\mathrm{Var}}_{x_0}(W).$$

The result follows. ■

Now, given a function $J \in \mathcal{P}$, define the martingale upper bound $F_0 J(x_0)$ by

$$(F_0 J)(x_0) \triangleq \mathsf{E}\left[\max_{0\leq s\leq d}\; g(x_s) - \sum_{t=1}^{s} \Delta J_t \;\middle|\; x_0\right].$$

Consider the following analog of Lemma 6:

Lemma 10. *For any pair of functions $J, J' \in \mathcal{P}$,*

$$\left|F_0 J(x_0) - F_0 J'(x_0)\right| \leq 2\sqrt{d}\lambda(P)\sqrt{\overline{\mathrm{Var}}_{x_0}(J-J')}.$$

Proof. Following (A.14) in the proof of Lemma 6, observe that, using Fact 1 and Jensen's inequality,

$$\left|F_0 J(x_0) - F_0 J'(x_0)\right|^2 \leq \mathsf{E}\left[\max_{0\leq s\leq d}\left|\sum_{t=1}^{s}\left(\Delta J_t - \Delta J'_t\right)\right|^2 \;\middle|\; x_0\right].$$

Using Doob's maximal quadratic inequality and the orthogonality of martingale differences,

$$\left|F_0 J(x_0) - F_0 J'(x_0)\right|^2 \leq 4\mathsf{E}\left[\left|\sum_{t=1}^{d}\left(\Delta J_t - \Delta J'_t\right)\right|^2 \;\middle|\; x_0\right] = 4\mathsf{E}\left[\sum_{t=1}^{d}\left|\Delta J_t - \Delta J'_t\right|^2 \;\middle|\; x_0\right]$$
$$= 4d\,\|\Delta J - \Delta J'\|_{2,x_0}^2.$$

The result follows by applying Lemma 9. ■

Taking $J' = J^*$ to be the optimal value function in Lemma 10, we immediately obtain the following analog of Theorem 2:

Theorem 6. *For any function $J \in \mathcal{P}$,*

$$\left|F_0 J(x_0) - J^*(x_0)\right| \leq 2\sqrt{d}\lambda(P)\sqrt{\overline{\mathrm{Var}}_{x_0}(J-J^*)}. \tag{B.2}$$

Theorem 6 provides approximation guarantee for martingale duality upper bounds in the finite horizon, non-stationary case. Comparing with the bound of Theorem 2 in the infinite horizon, stationary case, we see that the bounds have qualitatively similar dependence on the structural features of the optimal stopping problem:

- **Value Function Approximation Quality.** The bounds in both (17) and (B.2) depend on the quality of the function J as an approximation to J^*, measured in a root mean squared sense.
- **Time Horizon.** The bounds in both (17) and (B.2) have a square root dependence on the time horizon. In the case of (B.2) this is explicit, in the case of (17) the dependence is on the square root of the effective time horizon (19).
- **Predictability.** The bounds in both (17) and (B.2) depend linearly on the predictability of the underlying Markov chain.

Finally, the following theorem, an analog of Theorem 3, provides an approximation guarantee for the upper bound produced by the pathwise method in the finite horizon, non-stationary case:

Theorem 7. *Suppose that r_{PO} is an optimal solution to pathwise optimization problem*

$$\inf_{r} \; F_0 \Phi r(x_0).$$

Then,

$$\left| F_0 \Phi r_{\mathsf{PO}}(x_0) - J^*(x_0) \right| \leq 2\sqrt{d}\lambda(P) \min_{r} \; \sqrt{\overline{\mathrm{Var}}_{x_0}(\Phi r - J^*)}.$$

Proof. Observe that, for any $r \in \mathbb{R}^K$, by the optimality of r_{PO} and Lemma 1,

$$\left| F_0 \Phi r_{\mathsf{PO}}(x_0) - J^*(x_0) \right| = F_0 \Phi r_{\mathsf{PO}}(x_0) - J^*(x_0) \leq F_0 \Phi r(x_0) - J^*(x_0) = \left| F_0 \Phi r(x_0) - J^*(x_0) \right|.$$

The result follows by applying Theorem 6, and minimizing over r. ■

C. Additional Computational Results

In this section, we provide additional computational results for the optimal stopping problem of Section 4. Tables 3 and 4 show the upper and lower bounds computed as, respectively, the number of exercise opportunities d and the common asset price correlation $\rho_{jj'} = \bar{\rho}$ is varied. We also experiment with random correlation matrices. In Table 5, we report results of experiments where the correlation matrix was chosen randomly. Our setup used a random correlation matrix obtained by sampling a positive semidefinite matrix from the Wishart distribution and rescaling it so that the diagonal is identity.

(a) Upper and lower bounds, with standard errors.

n	LS-LB	S.E.	PO-LB	S.E.	DP-UB	S.E.	PO-UB	S.E.	DVF-UB	S.E.
					$d = 36$ exercise opportunities					
4	40.315	(0.004)	41.073	(0.008)	42.723	(0.016)	43.006	(0.021)	43.199	(0.009)
8	48.283	(0.004)	49.114	(0.006)	50.425	(0.019)	50.721	(0.027)	51.011	(0.008)
16	51.835	(0.003)	52.289	(0.004)	53.231	(0.009)	53.517	(0.020)	53.741	(0.006)
					$d = 54$ exercise opportunities					
4	40.797	(0.003)	41.541	(0.009)	43.587	(0.016)	43.853	(0.027)	44.017	(0.011)
8	49.090	(0.004)	50.252	(0.006)	51.814	(0.023)	52.053	(0.027)	52.406	(0.014)
16	52.879	(0.001)	53.638	(0.004)	54.883	(0.020)	55.094	(0.016)	55.450	(0.013)
					$d = 81$ exercise opportunities					
4	41.229	(0.004)	41.644	(0.017)	44.264	(0.023)	44.511	(0.030)	44.662	(0.006)
8	49.788	(0.003)	51.249	(0.004)	52.978	(0.018)	53.178	(0.027)	53.523	(0.013)
16	53.699	(0.003)	54.825	(0.005)	56.398	(0.024)	56.464	(0.007)	56.948	(0.008)

(b) Relative values of bounds.

n	(PO-LB) − (LS-LB)	(%)	(PO-UB) − (DP-UB)	(%)	(DVF-UB) − (PO-UB)	(%)
			$d = 36$ exercise opportunities			
4	0.759	1.88%	0.284	0.70%	0.192	0.48%
8	0.831	1.72%	0.297	0.61%	0.289	0.60%
16	0.454	0.88%	0.286	0.55%	0.224	0.43%
			$d = 54$ exercise opportunities			
4	0.744	1.82%	0.266	0.65%	0.164	0.40%
8	1.162	2.37%	0.239	0.49%	0.353	0.72%
16	0.759	1.43%	0.210	0.40%	0.356	0.67%
			$d = 81$ exercise opportunities			
4	0.415	1.01%	0.247	0.60%	0.151	0.37%
8	1.460	2.93%	0.201	0.40%	0.345	0.69%
16	1.126	2.10%	0.066	0.12%	0.484	0.90%

Table 3: A comparison of the lower and upper bound estimates of the PO and benchmarking methods, as a function of the number of exercise opportunities d and the number of assets n. For each algorithm, the mean and standard error (over 10 independent trials) is reported. The common initial asset price was $p_0^j = \bar{p}_0 = 100$ and the common correlation was $\rho_{jj'} = \bar{\rho} = 0$. Percentage relative values are expressed relative to the LS-LB lower bound.

(a) Upper and lower bounds, with standard errors.

n	LS-LB	S.E.	PO-LB	S.E.	DP-UB	S.E.	PO-UB	S.E.	DVF-UB	S.E.
					$\bar{\rho} = -0.05$ correlation					
4	41.649	(0.004)	42.443	(0.009)	44.402	(0.023)	44.644	(0.019)	44.846	(0.013)
8	50.077	(0.005)	51.136	(0.005)	52.581	(0.031)	52.799	(0.018)	53.163	(0.011)
16	53.478	(0.004)	54.076	(0.004)	55.146	(0.013)	55.360	(0.010)	55.708	(0.010)
					$\bar{\rho} = 0$ correlation					
4	40.797	(0.003)	41.541	(0.009)	43.587	(0.016)	43.853	(0.027)	44.017	(0.011)
8	49.090	(0.004)	50.252	(0.006)	51.814	(0.023)	52.053	(0.027)	52.406	(0.014)
16	52.879	(0.001)	53.638	(0.004)	54.883	(0.020)	55.094	(0.016)	55.450	(0.013)
					$\bar{\rho} = 0.1$ correlation					
4	39.180	(0.006)	39.859	(0.011)	42.001	(0.037)	42.187	(0.029)	42.425	(0.010)
8	47.117	(0.005)	48.371	(0.005)	50.139	(0.029)	50.362	(0.035)	50.700	(0.014)
16	51.414	(0.005)	52.498	(0.008)	54.141	(0.032)	54.217	(0.018)	54.654	(0.010)

(b) Relative values of bounds.

n	(PO-LB) − (LS-LB)	(%)	(PO-UB) − (DP-UB)	(%)	(DVF-UB) − (PO-UB)	(%)
			$\bar{\rho} = -0.05$ correlation			
4	0.794	1.91%	0.242	0.58%	0.202	0.49%
8	1.059	2.11%	0.218	0.44%	0.364	0.73%
16	0.598	1.12%	0.214	0.40%	0.349	0.65%
			$\bar{\rho} = 0$ correlation			
4	0.744	1.82%	0.266	0.65%	0.164	0.40%
8	1.162	2.37%	0.239	0.49%	0.353	0.72%
16	0.759	1.43%	0.210	0.40%	0.356	0.67%
			$\bar{\rho} = 0.1$ correlation			
4	0.679	1.73%	0.187	0.48%	0.238	0.61%
8	1.255	2.66%	0.224	0.47%	0.338	0.72%
16	1.084	2.11%	0.076	0.15%	0.437	0.85%

Table 4: A comparison of the lower and upper bound estimates of the PO and benchmarking methods, as a function of the common correlation $\rho_{jj'} = \bar{\rho}$ and the number of assets n. For each algorithm, the mean and standard error (over 10 independent trials) is reported. The common initial price was $p_0^j = \bar{p}_0 = 100$ and the number of exercise opportunities was $d = 54$. Percentage relative values are expressed relative to the LS-LB lower bound.

(a) Upper and lower bounds, with standard errors.

$\bar{p}_0$	LS-LB	S.E.	PO-LB	S.E.	DP-UB	S.E.	PO-UB	S.E.	DVF-UB	S.E.
					$n = 4$ assets					
90	30.532	(0.554)	30.944	(0.574)	32.494	(0.570)	32.589	(0.575)	32.755	(0.570)
100	39.088	(0.624)	39.791	(0.648)	41.523	(0.618)	41.801	(0.631)	41.973	(0.613)
110	45.892	(0.575)	46.969	(0.572)	48.599	(0.519)	48.916	(0.546)	49.151	(0.534)
					$n = 8$ assets					
90	42.486	(0.287)	43.392	(0.290)	44.923	(0.277)	45.205	(0.266)	45.304	(0.275)
100	48.971	(0.212)	50.027	(0.194)	51.483	(0.178)	51.772	(0.169)	52.035	(0.178)
110	52.618	(0.124)	53.491	(0.097)	54.835	(0.081)	55.029	(0.076)	55.420	(0.072)
					$n = 16$ assets					
90	48.784	(0.146)	49.825	(0.135)	51.270	(0.130)	51.513	(0.123)	51.812	(0.123)
100	52.376	(0.085)	53.214	(0.069)	54.568	(0.058)	54.726	(0.053)	55.127	(0.054)
110	54.292	(0.048)	54.888	(0.035)	56.034	(0.021)	56.244	(0.029)	56.600	(0.017)

(b) Relative values of bounds.

$\bar{p}_0$	(PO-LB) − (LS-LB)	(%)	(PO-UB) − (DP-UB)	(%)	(DVF-UB) − (PO-UB)	(%)
			$n = 4$ assets			
90	0.412	1.35%	0.095	0.31%	0.166	0.54%
100	0.703	1.80%	0.278	0.71%	0.172	0.44%
110	1.077	2.35%	0.317	0.69%	0.235	0.51%
			$n = 8$ assets			
90	0.906	2.13%	0.282	0.66%	0.099	0.23%
100	1.056	2.16%	0.289	0.59%	0.263	0.54%
110	0.873	1.66%	0.194	0.37%	0.391	0.74%
			$n = 16$ assets			
90	1.041	2.13%	0.243	0.50%	0.299	0.61%
100	0.838	1.60%	0.158	0.30%	0.401	0.77%
110	0.596	1.10%	0.210	0.39%	0.356	0.66%

Table 5: A comparison of the lower and upper bound estimates of the PO and benchmarking methods, as a function of the common initial asset price $p_0^j = \bar{p}_0$ and the number of assets n. For each algorithm, the mean and standard error (over 10 independent trials) is reported. For each trial the correlation matrix was sampled randomly and the number of exercise opportunities was $d = 54$. Percentage relative values are expressed relative to the LS-LB lower bound.