Optimal Stopping of Partially Observable Markov Processes: A Filtering-Based Duality Approach

Fan Ye, and Enlu Zhou, *Member, IEEE*

***Abstract*—In this note we develop a numerical approach to the problem of optimal stopping of discrete-time continuous-state partially observable Markov processes (POMPs). Our motivation is to find approximate solutions that provide lower and upper bounds on the value function such that the gap between the bounds can provide a practical measure of the quality of the solutions. To this end, we develop a filtering-based duality approach, which relies on the martingale duality formulation of the optimal stopping problem and the particle filtering technique. We show that this approach complements an asymptotic lower bound derived from a suboptimal stopping time with an asymptotic upper bound on the value function. We carry out error analysis and illustrate the effectiveness of our method on an example of pricing American options under partial observation of stochastic volatility.**

***Index Terms*—Partially observable, optimal stopping, particle filtering, martingale duality, American option pricing, stochastic volatility.**

I. Introduction

Optimal stopping of a partially observable Markov process (POMP) is a sequential decision making problem under partial observation of the underlying state. This type of problems arise in a number of applications, including change point detection in a production line, launching of a new technology under incomplete information of the market, and selling of an asset or a financial derivative. Optimal stopping of a POMP is more challenging than its counterpart of a fully observable process, since the inference of the hidden state and the choice of an optimal action should be accomplished at the same time. As a special class of the partially observable Markov decision processes (POMDPs), optimal stopping of a POMP can be transformed to a fully observable optimal stopping problem by introducing a new state variable, often referred to as the filtering distribution. However, this concise representation does not reduce the complexity of the problem, because the filtering distribution is usually infinite dimensional when the unobserved state takes values in a continuous space. In addition, the problem also suffers from the so-called "curse of dimensionality" of dynamic programming that is common in solving continuous-state Markov decision processes. Numerical solutions to optimal stopping of POMPs have been studied by [4], [8], [10], [9], mostly in the setting of pricing American options under partial observation of stochastic volatility. These methods can be viewed as a combination of dimension reduction on the filtering distribution and approximate dynamic programming, whereas [14] avoids the filtering step to approximate the value function. Some of the aforementioned approaches are proven to converge asymptotically to the true value function. However, in practice with a finite amount of computation resource, the difference between their approximate solutions and the true value function is usually unknown and hard to quantify.

In view of the lack of performance guarantee and computational complexity of the aforementioned methods, in this note we focus on developing a lower-and-upper-bound approach with moderate computational cost. The motivation is that the gap between the lower and upper bounds gives an indication of the quality of the approximate solutions. To guarantee a high-quality approximate solution, we can increase the computation effort until the gap between the two bounds decreases to a desirable tolerance level. To this end, we propose a filtering-based duality approach that complements a suboptimal stopping time (hence an asymptotic lower bound) with an asymptotic upper bound on the value function. Since our approach does not tie to a particular model and only involves Monte Carlo simulation, it can be generalized to any POMP as long as the particle filtering technique can be applied. Our method relies on the martingale duality formulation of the fully observable optimal stopping problem, which is proposed by [11] and [5] in the setting of pricing American options under constant volatility.

From the perspective of modeling fidelity versus computational complexity, it is not trivial to compare optimal stopping of POMPs with its counterpart in fully observable Markov processes. In particular, the difference of their value functions cannot be quantified in general and is problem dependent, so we are also interested in learning the features that influence this difference in the underlying probabilistic model. Indeed, as an example, our numerical experiments on pricing American options under partially observable stochastic volatility show that our asymptotic upper bound is strictly less than the option price of the model where the volatility is treated directly observable, and the difference is especially obvious when the effect of the volatility is dominant. This in turn shows that our method provides a better criterion to evaluate the performance of a suboptimal policy in the partially observable model.

The rest of the note is organized as follows. In Section II, we describe the general problem formulation of optimal stopping of POMPs and the transformation to an equivalent fully observable optimal stopping problem. In Section III, we develop the filtering-based duality approach, and its error analysis and convergence result are presented in Section IV. We present some numerical examples in Section V, and finally conclude in Section VI. All the proofs are contained in the Appendix.

II. Problem Formulation

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Consider a hidden Markov model $\{(X_t, Y_t), t = 0, 1, \cdots, T\}$ satisfying the following equations

$$X_{t+1} = f(X_t, Z^1_{t+1}), \quad t = 0, 1, \cdots, T-1; \tag{1a}$$

$$Y_0 = h_0(X_0, Z^2_0); \tag{1b}$$

$$Y_{t+1} = h(X_{t+1}, Y_t, Z^2_{t+1}), \quad t = 0, 1, \cdots, T-1; \tag{1c}$$

where the unobserved state X_t is in a continuous state space $\mathscr{X} \subseteq \mathbb{R}^{n_x}$, the observation Y_t is in a continuous observation space $\mathscr{Y} \subseteq \mathbb{R}^{n_y}$. The noises $\{(Z^1_t, Z^2_t), t = 1, \cdots, T\}$, which are independent of the initial state X_0 and the initial observation Y_0, are independent random vectors with known distributions, but the components of each vector can be correlated. Equations (1a) and (1b)-(1c) are often referred to as the state equation and the observation equation respectively. Note that $\{(X_t, Y_t)\}$ is a bivariate Markov process adapted to the filtration $\left\{\mathscr{F}_t \triangleq \sigma\{(X_i, Y_i); i = 0, \ldots, t\}\right\}$.

Let $\mathscr{J} \triangleq \{1, \cdots, T\}$. Denote by $\left\{\mathscr{F}^Y_t \triangleq \sigma\{Y_0, \ldots, Y_t\}\right\}$ the filtration generated by the processes (1b)-(1c). A random variable $\tau : \Omega \to \mathscr{J}$ is an $\mathscr{F}^Y_t$-stopping time if $\{\tau \leq t\} \in \mathscr{F}^Y_t$ for every $t \in \mathscr{J}$. We define $\mathscr{T}^Y$ as the set of $\mathscr{F}^Y_t$-stopping times that take values in $\mathscr{J}$. Assume that the initial Y_0 is a known constant, and the initial X_0 follows a known distribution π_0, which is derived from the historical data (including Y_0). We consider the finite-horizon partially observable optimal stopping problem

$$V_0(\pi_0, y_0) = \sup_{\tau \in \mathscr{T}^Y} \mathbb{E}[g(\tau, X_\tau, Y_\tau) | X_0 \sim \pi_0, Y_0 = y_0], \tag{2}$$

where $g : \mathscr{J} \times \mathscr{X} \times \mathscr{Y} \to \mathbb{R}$ is the reward function. In this setting the decision maker has access to only state Y_t so that her decision at time t is made purely depending on the observation history up to time t,

F. Ye and E. Zhou are with the Department of Industrial & Enterprise Systems Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, 61801 USA e-mail:{fanye2, enluzhou}@illinois.edu.

This work was supported by the National Science Foundation under Grants ECCS-0901543 and CMMI-1130273, and by the Air Force Office of Scientific Research under YIP Grant FA-9550-12-1-0250.

i.e.,$\{Y_0,\cdots,Y_t\}$. For convenience, in the following we use $g(X_t,Y_t)$ and $g(X_\tau,Y_\tau)$ in short for $g(t,X_t,Y_t)$ and $g(\tau,X_\tau,Y_\tau)$ respectively.

The optimal stopping problem of a POMP can be transformed to an equivalent fully observable optimal stopping problem by introducing a new state variable Π_t, often referred to as the *filtering distribution*, which is the conditional distribution of X_t given the observations $Y_{0:t} \triangleq \{Y_0,\dots,Y_t\}$. More specifically, given a set A in the Borel σ-algebra over $\mathscr{X}$, define

$$\Pi_t(A) \triangleq \text{Prob}(X_t \in A|Y_0,\dots,Y_t), \quad t=0,\dots,T.$$

Given a realization of the observations $y_{0:t} \triangleq \{y_0,\dots,y_t\}$, the probability density π_t of the filtering distribution Π_t evolves as follows:

$$\pi_t(x_t) = \frac{\int_{\mathscr{X}} p(x_t,y_t|x_{t-1},y_{t-1})\pi_{t-1}(x_{t-1})\,dx_{t-1}}{\int_{\mathscr{X}} p(y_t|x_{t-1},y_{t-1})\pi_{t-1}(x_{t-1})\,dx_{t-1}},\ t=1,\dots,T, \quad (3)$$

where the conditional probability density functions $p(x_t,y_t|x_{t-1},y_{t-1})$ and $p(y_t|x_{t-1},y_{t-1})$ are induced by (1a), (1c), and the distributions of Z_t^1 and Z_t^2. Noticing that π_t only depends on π_{t-1}, y_{t-1}, and y_t, and letting the realization $y_{0:t}$ be replaced by the random variables $Y_{0:t}$, we can abstractly rewrite the filtering recursion (3) as

$$\Pi_t = \Phi(\Pi_{t-1},Y_{t-1},Y_t), \quad t=1,2,\dots,T.$$

Then problem (2) can be transformed to an equivalent optimal stopping problem (see, e.g., Chapter 5 in [3]) with fully observable state (Π_t,Y_t):

$$V_0(\pi_0,y_0) = \sup_{\tau\in\mathscr{T}^Y} \mathbb{E}[\tilde{g}(\Pi_\tau,Y_\tau)|X_0\sim\pi_0,Y_0=y_0],$$

where

$$\tilde{g}(\Pi_t,Y_t) \triangleq \mathbb{E}[g(X_t,Y_t)|\mathscr{F}_t^Y] = \int g(x_t,Y_t)\Pi_t(x_t)\,dx_t.$$

Theoretically, we can solve (2) following the dynamic programming recursion:

$$V_t(\Pi_t,Y_t) = \max\left(\tilde{g}(\Pi_t,Y_t),C_t(\Pi_t,Y_t)\right), \quad t=T,\dots,1, \qquad (4)$$

where $C_t(\Pi_t,Y_t)$ is the *continuation value* at time t defined as

$$\begin{aligned} C_T(\Pi_T,Y_T) &\triangleq \tilde{g}(\Pi_T,Y_T); \\ C_t(\Pi_t,Y_t) &\triangleq \mathbb{E}[V_{t+1}(\Pi_{t+1},Y_{t+1})|\Pi_t,Y_t], \quad t=T-1,\dots,0. \end{aligned}$$

Here $\mathbb{E}[\cdot|\Pi_t,Y_t]$ is interpreted as $\mathbb{E}[\cdot|X_t\sim\Pi_t,Y_t]$. Then $V_0=C_0$ and the optimal stopping time is

$$\tau^* = \min\{t\in\mathscr{J} \mid \tilde{g}(\Pi_t,Y_t)\geq C_t(\Pi_t,Y_t)\}.$$

We also define its associated t-indexed stopping time τ_t^* for each $t\in\mathscr{J}$:

$$\tau_t^* \triangleq \min\{i\in\mathscr{J}_t \mid \tilde{g}(\Pi_i,Y_i)\geq C_i(\Pi_i,Y_i)\} \qquad (5)$$

with $\mathscr{J}_t \triangleq \{t,t+1,\dots,T\}$. The above recursion also shows that (Π_t,Y_t) are the sufficient statistics that determine the optimal stopping time. The process $\{V_t \triangleq V_t(\Pi_t,Y_t)\}$ defined in (4) is called the Snell envelope process (see, e.g., Chapter 2 in [6]) of the process $\{\tilde{g}(\Pi_t,Y_t)\}$, which is the smallest $\mathscr{F}_t^Y$-supermartingale that dominates $\tilde{g}$ in the sense that $V_t(\Pi_t,Y_t)\geq\tilde{g}(\Pi_t,Y_t)$. In particular, by shifting the time index in (2) we can interpret V_t as

$$\begin{aligned} V_t(\pi_t,y_t) &= \sup_{\tau\in\mathscr{T}^Y,\ t\leq\tau\leq T} \mathbb{E}[g(X_\tau,Y_\tau)|X_t\sim\pi_t,Y_t=y_t] \\ &= \mathbb{E}[g(X_{\tau_t^*},Y_{\tau_t^*})|X_t\sim\pi_t,Y_t=y_t], \quad t=1,\dots,T. \qquad (6) \end{aligned}$$

However, it is often impossible to solve the problem exactly following (4) due to two main difficulties. One is that in general the filtering distribution Π_t is infinite dimensional and the filtering recursion (3) cannot be computed exactly. The other difficulty lies in the accurate estimation of the continuation value $C_t(\Pi_t,Y_t)$ that leads to the optimal stopping time τ^*. So we develop an approximation method in the next section.

III. Filtering-Based Martingale Duality Approach

In this section, we construct a dual problem to the original optimal stopping of POMPs, and develop a numerical method that yields an asymptotic upper bound on the value function. Our dual formulation is a straightforward extension of the dual formulation for the optimal stopping problem proposed in [11], [5], and [1], by replacing the filtration with $\mathscr{F}_t^Y$.

Theorem 1 (c.f. (5) in [1])**.** *Let $\mathscr{M}$ represent the space of $\mathscr{F}_t^Y$-adapted martingales $\{M_t\}$ with $M_0=0$ and $\sup_{t\in\mathscr{J}}\mathbb{E}|M_t|<\infty$. Then*

$$V_0(\pi_0,y_0) = \min_{M\in\mathscr{M}}\left\{\mathbb{E}[\max_{t\in\mathscr{J}}\{\tilde{g}(\Pi_t,Y_t)-M_t\}|X_0\sim\pi_0,Y_0=y_0]\right\}. \qquad (7)$$

The optimal martingale $\{M_t^*\}$ that achieves the minimum on the right hand side of (7) is of the form

$$M_t^* = \sum_{i=1}^t \Delta_i^*, \qquad (8)$$

where$\{\Delta_t^*\}$ is the martingale difference sequence defined as

$$\Delta_t^* \triangleq \mathbb{E}[V_t|\mathscr{F}_t^Y] - \mathbb{E}[V_t|\mathscr{F}_{t-1}^Y],\ t\in\mathscr{J}. \qquad (9)$$

In addition, the following equality holds pathwisely in the almost sure sense, i.e.,

$$V_0(\pi_0,y_0) = \max_{t\in\mathscr{J}}(\tilde{g}(\Pi_t,Y_t)-M_t^*) \quad a.s..$$

The proof of Theorem 1 follows the same line in [1] and hence is omitted here. Theorem 1 characterizes a strong duality relation between the primal problem (2) and its dual problem on the right side of (7); this duality suggests that any $\mathscr{F}_t^Y$-adapted martingale $\{M_t\}$ can lead to an upper bound on $V_0(\pi_0,y_0)$ and that the optimal martingale (8) is derived from the Doob-Meyer decomposition of the supermartingale $\{V_t\}$. In particular, we can rewrite (9) as

$$\begin{aligned} \Delta_t^* &= \mathbb{E}[V_t|\Pi_t,Y_t] - \mathbb{E}[V_t|\Pi_{t-1},Y_{t-1}] \qquad (10a) \\ &= \mathbb{E}[g(X_{\tau_t^*},Y_{\tau_t^*})|\Pi_t,Y_t] - \mathbb{E}[g(X_{\tau_t^*},Y_{\tau_t^*})|\Pi_{t-1},Y_{t-1}]. \qquad (10b) \end{aligned}$$

Note that it is impossible to compute the optimal martingale $\{M_t^*\}$, since the martingale difference term (10a) (or (10b)) involves the intractable filtering distribution Π_t and the Snell envelop process $\{V_t\}$ (or the optimal stopping time τ_t^*). Therefore, we need to introduce approximation schemes to address both aspects. On the one hand, the intractable filtering distribution Π_t can be approximated by a discrete distribution using particle filtering, which will be stated in Section III-A. On the other hand, (10a) and (10b) suggest that we approximate Δ_t^* using either approximate value functions of V_t or suboptimal $\mathscr{F}_t^Y$-stopping times that approximate τ_t^*. In addition, some other heuristic constructions can be considered. For example, we can take $\Delta_t = \mathbb{E}[U_t(X_t,Y_t)|\mathscr{F}_t^Y] - \mathbb{E}[U_t(X_t,Y_t)|\mathscr{F}_{t-1}^Y]$, where $U_t(X_t,Y_t)$ is the value function to the corresponding optimal stopping problem with fully observable state (X_t,Y_t):

$$U_t(x_t,y_t) = \sup_{\kappa\in\mathscr{T}_t} \mathbb{E}[g(X_\kappa,Y_\kappa)|X_t=x_t,Y_t=y_t], \qquad (11)$$

where $\mathscr{T}_t$ is the set of $\mathscr{F}_t$-stopping times κ that take values in $\mathscr{J}_t$; or equivalently we can take $\Delta_t = \mathbb{E}[g(X_{\kappa_t^*},Y_{\kappa_t^*})|\Pi_t,Y_t] - \mathbb{E}[g(X_{\kappa_t^*},Y_{\kappa_t^*})|\Pi_{t-1},Y_{t-1}]$, where κ_t^* is the optimal $\mathscr{F}_t$-stopping time to problem (11). Even if the explicit forms of U_t and κ_t^* are not known, their approximations can be used in Δ_t and its martingale difference property can still be preserved. The advantage of approximating U_t or κ_t^* is their simple structure as functions of only (X_t,Y_t), whereas either V_t or τ_t^* is a function of $(Y_0,\cdots,Y_t)$. Thus, it may be easier to generate martingale difference terms based on approximate U_t or κ_t^*, even though they may yield less optimal values.

In the rest of this section we focus on approximating Δ_t^* in (10b) by the following Δ_t^m based on a fixed stopping time τ (see, e.g., (16) in Section III-B), which is either $\mathscr{F}_t^Y$ or $\mathscr{F}_t$-adapted:

$$\Delta_t^m \triangleq \mathbb{E}[g(X_{\tau_t}, Y_{\tau_t})|\Pi_t^m, Y_t] - \mathbb{E}[g(X_{\tau_t}, Y_{\tau_t})|\Pi_{t-1}^m, Y_{t-1}], \tag{12}$$

where τ_t is the t-indexed stopping time associated with τ, and Π_t^m (see details in Section III-A) is the approximate filtering distribution at time t obtained by particle filtering (the superscript m in Π_t^m denotes the number of particles), which will be elaborated in the next section. A lower-case notation π_t^m denotes the corresponding approximate filtering distribution based on a realization of the observations $y_{0:t}$. Then we define $\{M_t^m\}$ as

$$M_0^m = 0; \quad M_t^m = \Delta_1^m + \ldots + \Delta_t^m, \quad t \in \mathscr{J}. \tag{13}$$

Incorporating the above ideas, we propose the following algorithm that yields an asymptotic upper bound on V_0.

Algorithm 1. *Filtering-Based Martingale Duality Approach*
Step 1. For $k = 1, 2, \ldots, N$, **do**
- Generate a path of observations $y_{1:T}^{(k)}$ according to the processes (1a)-(1c) with initial condition $Y_0 = y_0$ and $X_0 \sim \pi_0$, and then follow Algorithm 2 (particle filtering) to generate the approximate filtering distribution $\{\pi_1^{m(k)}, \ldots, \pi_T^{m(k)}\}$.
- For $t = 1, \ldots, T$, use Algorithm 3 to compute $\tilde{\Delta}_t^{m(k)}$, which is an approximation for

$$\Delta_t^{m(k)} = \mathbb{E}[g(X_{\tau_t}, Y_{\tau_t})|\pi_t^{m(k)}, y_t^{(k)}] - \mathbb{E}[g(X_{\tau_t}, Y_{\tau_t})|\pi_{t-1}^{m(k)}, y_{t-1}^{(k)}]. \tag{14}$$

- Sum the approximate martingale differences to obtain

$$\tilde{M}_t^{m(k)} = \tilde{\Delta}_1^{m(k)} + \ldots + \tilde{\Delta}_t^{m(k)}, \; t = 1, \ldots, T.$$

- Evaluate $V^{(k)} = \max_{t \in \mathscr{J}} \left(\tilde{g}(\pi_t^{m(k)}, y_t^{(k)}) - \tilde{M}_t^{m(k)} \right)$. **end**
Step 2. Set $V_N^\tau = \frac{1}{N} \sum_{k=1}^N V^{(k)}$. V_N^τ is an asymptotic upper bound on the value function $V_0(\pi_0, y_0)$.

In the next two subsections, we will discuss how to generate approximate filtering distribution using particle filtering via Algorithm 2 and how to compute the approximate martingale difference via Algorithm 3.

A. Particle Filtering

We approximate π_t using particle filtering, which is a successful and versatile numerical method for solving nonlinear filtering problem. A good introduction on particle filtering can be found in the book [2]. The particle filtering method approximates π_t by a finite number (say m) of particles $\{x_t^{(1)}, \ldots, x_t^{(m)}\}$, i.e., a discrete distribution π_t^m written as follows

$$\pi_t^m = \frac{1}{m} \sum_{i=1}^m \delta_{x_t^{(i)}}, \tag{15}$$

where δ is the Dirac measure. As the number of particles m goes to infinity, it can be ensured that π_t^m converges to π_t in certain sense.
Algorithm 2. *Particle Filtering*
Input: $X_0 \sim \pi_0$ and a sequence of observations $y_{0:T}$.
Output: The approximate filtering distribution $\pi_0^m, \ldots, \pi_T^m$.
Step 1. Initialization: Set $t = 0$. Draw m i.i.d. samples $\{x_0^{(1)}, \ldots, x_0^{(m)}\}$ from the distribution π_0. Set $\pi_0^m = \frac{1}{m} \sum_{i=1}^m \delta_{x_0^{(i)}}$.
Step 2. For $t = 1, \ldots, T$, **do**
− Prediction: For each $i = 1, \ldots, m$, draw one sample $\bar{x}_t^{(i)}$ from $P(X_t | X_{t-1} = x_{t-1}^{(i)})$.
− Bayes' Updating: Compute $w_t^{(i)} = \frac{p(y_t | \bar{x}_t^{(i)}, y_{t-1})}{\sum_{i=1}^m p(y_t | \bar{x}_t^{(i)}, y_{t-1})}, \quad i = 1, \ldots, m.$
− Resampling: Draw i.i.d. samples $\{x_t^{(1)}, \ldots, x_t^{(m)}\}$ from the discrete distribution $\{\text{Prob}(\bar{x}_t^{(i)}) = w_t^{(i)}, \, i = 1, \ldots, m\}$. Set $\pi_t^m = \frac{1}{m} \sum_{i=1}^m \delta_{x_t^{(i)}}$. **end**

B. Approximate Martingale Difference

The remaining issue is how to compute the martingale difference (14). Throughout this subsection we assume a suboptimal stopping time τ of the form,

$$\tau = \min\{t \in \mathscr{J} \,|\, g(X_t, Y_t) \geq \tilde{C}_t(X_t, Y_t)\}, \tag{16}$$

where $\{\tilde{C}_t, t \in \mathscr{J}\}$ is a sequence of approximate continuation functions of U_t. The approximate continuation functions $\tilde{C}_t$ can be derived, for example, by regression on some basis functions as suggested by [7] and [13]. We choose an $\mathscr{F}_t$-stopping time τ of the form (16) only for ease of exposition, though Algorithm 3 can be adjusted using any other $\mathscr{F}_t$(or $\mathscr{F}_t^Y$)-stopping time with the same principle.

Given a realization of observations $y_{0:T}$, we employ nested simulation to obtain the estimate of Δ_t^m in (14). Note that π_t^m in Algorithm 1 is of the form (15). Therefore,

$$\begin{aligned}\Delta_t^m = &\frac{1}{m} \sum_{i=1}^m \mathbb{E}[g(X_{\tau_t}, Y_{\tau_t}) | X_t = x_t^{(i)}, Y_t = y_t] \\ &- \frac{1}{m} \sum_{i=1}^m \mathbb{E}[g(X_{\tau_t}, Y_{\tau_t}) | X_{t-1} = x_{t-1}^{(i)}, Y_{t-1} = y_{t-1}],\end{aligned}$$

where τ_t is the t-indexed stopping time associated with τ defined as

$$\tau_t = \min\{i \in \mathscr{J}_t \,|\, g(X_i, Y_i) \geq \tilde{C}_i(X_i, Y_i)\}.$$

To estimate $\mathbb{E}[g(X_{\tau_t}, Y_{\tau_t}) | x_t^{(i)}, y_t]$ (resp., $\mathbb{E}[g(X_{\tau_t}, Y_{\tau_t}) | x_{t-1}^{(i)}, y_{t-1}]$), we generate l subpaths that are stopped according to τ_t with initial condition $X_t = x_t^{(i)}, Y_t = y_t$ (resp., $X_{t-1} = x_{t-1}^{(i)}, Y_{t-1} = y_{t-1}$) for each i and t, and we average $g(X_{\tau_t}, Y_{\tau_t})$ over these subpaths. So there are a total number of $m \cdot l$ subpaths generated to estimate each expectation term in (14). The details of the nested simulation are presented below.

Algorithm 3. *Estimation of* Δ_t^m *Using Nested Simulation*
Input: y_{t-1}, y_t, $\pi_{t-1}^m = \frac{1}{m} \sum_{i=1}^m \delta_{x_{t-1}^{(i)}}$ *and* $\pi_t^m = \frac{1}{m} \sum_{i=1}^m \delta_{x_t^{(i)}}$ *from Algorithm 1 and Algorithm 2.*
(Step 1 - Step 2 are used to estimate $\mathbb{E}[g(X_{\tau_t}, Y_{\tau_t}) | \pi_{t-1}^m, y_{t-1}]$.)
Step 1. For $i = 1, \ldots, m$, **do**
- Simulate $\{(x_t^{(ij)}, y_t^{(ij)}), \ldots, (x_T^{(ij)}, y_T^{(ij)})\}_{j=1}^l$ from the processes (1a)-(1c) with the initial condition $X_{t-1} = x_{t-1}^{(i)}$ and $Y_{t-1} = y_{t-1}$.
- To apply τ_t on these sample paths, find

$$t_{ij} = \min\left\{k \in \mathscr{J}_t : g(x_k^{(ij)}, y_k^{(ij)}) \geq \tilde{C}_k(x_k^{(ij)}, y_k^{(ij)})\right\}.$$

- Set $b_i = \frac{1}{l} \sum_{j=1}^l g(x_{t_{ij}}^{(ij)}, y_{t_{ij}}^{(ij)})$. **end**
Step 2. Set $G_{t-1,t}^{m,l} \triangleq \frac{1}{m} \sum_{i=1}^m b_i$, which is an unbiased estimator of $\mathbb{E}[g(X_{\tau_t}, Y_{\tau_t}) | \pi_{t-1}^m, y_{t-1}]$.
(Step 3 - Step 4 is used to estimate $\mathbb{E}[g(X_{\tau_t}, Y_{\tau_t}) | \pi_t^m, y_t]$.)
Step 3. For $i = 1, \ldots, m$, **do**
If $g(x_t^{(i)}, y_t) \geq \tilde{C}_t(x_t^{(i)}, y_t)$, i.e., $(x_t^{(i)}, y_t)$ is in the stopping region, set $\tilde{b}_i = g(x_t^{(i)}, y_t)$. Otherwise, repeat Step 1 with the initial condition $X_t = x_t^{(i)}$ and $Y_t = y_t$ to obtain $\tilde{b}_i$. **end**
Step 4. Set $G_{t,t}^{m,l} \triangleq \frac{1}{m} \sum_{i=1}^m \tilde{b}_i$, which is an unbiased estimator of $\mathbb{E}[g(X_{\tau_t}, Y_{\tau_t}) | \pi_t^m, y_t]$.
Step 5. Set $\tilde{\Delta}_t^m = G_{t,t}^{m,l} - G_{t-1,t}^{m,l}$.

IV. Error Analysis

In this section, we analyze the error bound and asymptotic convergence of our algorithm. To lighten the notations, we use $\mathbb{E}_0[\cdot]$ to denote $\mathbb{E}[\cdot | X_0 \sim \pi_0, Y_0 = y_0]$ in the rest of note. The following assumption is used throughout our analysis.

Assumption 1.
i. $\| g \|_\infty \triangleq \max_{t \in \mathscr{J}} \| g(t, \cdot, \cdot) \|_\infty < \infty$.
ii. For any observation sequence $y_{0:T}$,

$$\sup_{x_t \in \mathscr{X}} p(y_t | x_t, y_{t-1}) < \infty, \quad \forall t \in \mathscr{J}.$$

We first introduce an $\mathscr{F}_t^Y$-adapted martingale difference sequence $\{\Delta_t^\tau\}$ and martingale $\{M_t^\tau\}$ induced by an $\mathscr{F}_t$(or $\mathscr{F}_t^Y$)-stopping time τ:

$$\Delta_t^\tau = \mathbb{E}[g(X_{\tau_t}, Y_{\tau_t})|\Pi_t, Y_t] - \mathbb{E}[g(X_{\tau_t}, Y_{\tau_t})|\Pi_{t-1}, Y_{t-1}],$$
$$M_0^\tau \triangleq 0; \quad M_t^\tau \triangleq \Delta_1^\tau + \ldots + \Delta_t^\tau, \quad t \in \mathscr{J}.$$

Since M_t^τ is an $\mathscr{F}_t^Y$-adapted martingale, then $\mathbb{E}_0[\max_{t\in\mathscr{J}}(\tilde{g}(\Pi_t, Y_t) - M_t^\tau)]$ is an upper bound on $V_0(\pi_0, y_0)$ by Theorem 1.

Recall that the approximate martingale difference Δ_t^m based on a realization of observations $y_{0:t}$ is

$$\Delta_t^m = \mathbb{E}[g(X_{\tau_t}, Y_{\tau_t})|\pi_t^m, y_t] - \mathbb{E}[g(X_{\tau_t}, Y_{\tau_t})|\pi_{t-1}^m, y_{t-1}].$$

In Algorithm 3 the empirical estimates of $\mathbb{E}[g(X_{\tau_t}, Y_{\tau_t})|\pi_t^m, y_t]$ and $\mathbb{E}[g(X_{\tau_t}, Y_{\tau_t})|\pi_{t-1}^m, y_{t-1}]$ are denoted by $G_{t,t}^{m,l}$ and $G_{t-1,t}^{m,l}$, respectively. Therefore, we use

$$\tilde{\Delta}_t^m = G_{t,t}^{m,l} - G_{t-1,t}^{m,l} \quad \text{and} \quad \tilde{M}_t^m = \sum_{i=1}^t \tilde{\Delta}_i^m$$

to approximate Δ_t^m and M_t^m. Instead of obtaining $\max_{t\in\mathscr{J}}\{\tilde{g}(\pi_t, y_t) - M_t^\tau\}$ exactly along each path of the observations $y_{0:T}$, we compute $\max_{t\in\mathscr{J}}\{\tilde{g}(\pi_t^m, y_t) - \tilde{M}_t^m\}$. Note that conditional on a fixed observation sequence, the former term is a constant, while the latter one is a random term due to sampling. The difference between these two terms is due to two sources of noise: One is from the difference of the deterministic density π_t and the random measure π_t^m, and this gap will go to zero (in expectation) by increasing the number of particles m under Assumption 1; another difference is from the variability of the nested (Monte Carlo) simulation, which can be eliminated by increasing the number of sample paths $m \cdot l$.

We will show in the next theorem (with proof in the Appendix) that $\mathbb{E}_0[\max_{t\in\mathscr{J}}\{\tilde{g}(\Pi_t^m, Y_t) - \tilde{M}_t^m\}]$ converges to $\mathbb{E}_0[\max_{t\in\mathscr{J}}\{\tilde{g}(\Pi_t, Y_t) - M_t^\tau\}]$ when the particle number m increases to infinity. Hence, $\mathbb{E}_0[\max_{t\in\mathscr{J}}\{\tilde{g}(\Pi_t^m, Y_t) - \tilde{M}_t^m\}]$ is an asymptotic (as $m \to \infty$) upper bound on $V_0(\pi_0, y_0)$. Moreover, the gap between $\mathbb{E}_0[\max_{t\in\mathscr{J}}\{\tilde{g}(\Pi_t, Y_t) - M_t^\tau\}]$ and $V_0(\pi_0, y_0)$ is purely due to the suboptimal stopping time τ.

Theorem 2. *Suppose τ is an $\mathscr{F}_t$(or $\mathscr{F}_t^Y$)-stopping time. Then*

$$\lim_{m\to\infty} \mathbb{E}_0[\max_{t\in\mathscr{J}}\{\tilde{g}(\Pi_t^m, Y_t) - \tilde{M}_t^m\}] = \mathbb{E}_0[\max_{t\in\mathscr{J}}\{\tilde{g}(\Pi_t, Y_t) - M_t^\tau\}]. \quad (17)$$

Moreover, we have the following inequalities:

$$\begin{aligned}
&\mathbb{E}_0[\max_{t\in\mathscr{J}}\{\tilde{g}(\Pi_t, Y_t) - M_t^\tau\}] - V_0(\pi_0, y_0) \\
\le& 2\sqrt{\sum_{t=1}^T \mathbb{E}_0[(\Delta_t^* - \Delta_t^\tau)^2]} \\
\le& 2\sqrt{\sum_{t=1}^T \mathbb{E}_0\left[\left(\mathbb{E}[g(X_{\tau_t^*}, Y_{\tau_t^*})|\Pi_t, Y_t,] - \mathbb{E}[g(X_{\tau_t}, Y_{\tau_t})|\Pi_t, Y_t]\right)^2\right]}. \quad (18)
\end{aligned}$$

From (17), the output V_N^τ in Algorithm 1 is an asymptotic (as the sample path number $N \to \infty$ and the particle number $m \to \infty$) upper bound on the true value function V_0. According to (18), a large m will lead to a tight upper bound provided that the martingale $\{M_t^\tau\}$ induced by the stopping time τ does not differ too much from the optimal $\{M_t^*\}$, or more intuitively, the suboptimal stopping time τ_t does not differ too much from the optimal τ_t^*.

V. Numerical Examples

We apply our method to price American put options under stochastic volatility. Following the model in [10] we considered a d_S-dimensional process of asset price $\{S_t, t = 0:T\}$:

$$S_{t+1}^i = S_t^i \exp\left\{\left(r - \frac{(\sigma_{t+1}^i)^2}{2}\right)\delta + \sigma_{t+1}^i\sqrt{\delta}Z_{t+1}^{i,1}\right\}, \ i = 1, \ldots, d_S, \quad (19)$$

where r is the constant interest rate, δ is the time period between the equally-spaced time points, $\{Z_t^{i,1}, t = 1:T\}, i = 1, \ldots, d_S$ are independent sequences of Gaussian random variables with $Z_t^{i,1} \sim \mathscr{N}(0,1)$, and the volatility $\sigma_t^i \triangleq \exp(X_t^i)$ is a deterministic function of a $d_X(= d_S)$-dimensional process $\{X_t, t = 0:T\}$ that evolves as a discretized Ornstein-Uhlenbeck process:

$$X_{t+1}^i = X_t^i e^{-\lambda_i\delta} + \theta_i(1 - e^{-\lambda_i\delta}) + \gamma_i\sqrt{\frac{1 - e^{-2\lambda_i\delta}}{2\lambda_i}} Z_{t+1}^{i,2}, \ i = 1, \ldots, d_X, \quad (20)$$

where the positive constant θ_i is the mean reversion value, the constant λ_i is the mean reversion rate, the constant γ_i is a measure of the process volatility, and $\{Z_t^{i,2}, t = 1:T\}, i = 1, \ldots, d_X$ are independent sequences of Gaussian random variables with $Z_t^{i,2} \sim \mathscr{N}(0, \mu_i^2)$, which are also independent of $\{Z_t^{i,1}\}$. Here μ_i is used to control the observation noise. For simplicity, in our numerical experiments we use $\lambda_i = \lambda$, $\theta_i = \theta$, $\gamma_i = \gamma$, $\mu_i = \mu$ for all $i = 1, \ldots, d_X$. Assume that only the asset price is observed, and exercise opportunities take place at $t = 1, \ldots, T$. We consider the put option on the minimum of d_S assets, i.e., the payoff function is of the form

$$g(t, S_t) = \max\left\{e^{-r\delta t}\left(K - \min\{S_t^1, \ldots, S_t^{d_S}\}\right), 0\right\}.$$

In the rest of this section, "exercise policy" simply means "stopping time" in the general optimal stopping problem.

Remark 1. *In this example, the conditional probability density function*

$$p(S_t|X_t, S_{t-1}) = \prod_{i=1}^{d_X} p(S_t^i|X_t^i, S_{t-1}^i)$$

where

$$p(S_t^i|X_t^i, S_{t-1}^i) = \frac{\exp\left\{-\frac{\left(\ln(S_t^i/S_{t-1}^i) - (r - \exp^2(X_t^i)/2)\delta\right)^2}{2\exp^2(X_t^i)\delta\mu^2}\right\}}{S_t^i\sqrt{2\pi\exp^2(X_t^i)\delta\mu^2}}.$$

It can be shown that $p(S_t|X_t, S_{t-1})$ satisfies Assumption 1(ii) and that Assumption 1(i) is also trivially satisfied.

Since the stochastic volatility cannot be directly observed in reality but can be "partially observable" through the inference from the observed asset price, pricing American option under the above model (19)-(20) falls into the framework of optimal stopping of POMPs. We illustrate our algorithm through a series of numerical experiments with $d_S = 1$ (one asset) and $d_S = 2$ (two assets). In particular, we are interested in how the variance of the volatility (corresponding to the parameters $(\theta, \lambda, \gamma)$) and observation noise (corresponding to the parameter μ) influence the price difference due to the difference between the fully observable and partially observable volatilities. We list the parameter sets in Table I. To compute option prices under both full and partial observations, we implement our algorithm as well as the Least-Squares Monte Carlo (LSMC) method of [7], which provides suboptimal exercise policies, and the primal-dual (PD) method of [1], which parallels our method in the fully observable models. The numerical results of the option prices under different parameter sets are listed in Table II (for one asset) and Table III (for two assets), where "*LB*" represents the lower bound obtained by the LSMC method for the fully/partially observable model with the following two sets of basis functions for the one-asset and two-asset problems respectively:

$$\begin{aligned}
H_1 =& \{L_0(S_t^1), L_0^2(S_t^1), L_1(S_t^1), L_1^2(S_t^1), L_0(S_t^1)L_1(S_t^1), 1\}, \\
H_2 =& \{L_0(S_t^1), L_0^2(S_t^1), L_0(S_t^2), L_0^2(S_t^2), L_0(S_t^1)L_0(S_t^2), L_2(S_t^1, S_t^2), L_2^2(S_t^1, S_t^2), 1\},
\end{aligned}$$

where $L_0(x) = x$, $L_1(x) = \max\{K - x, 0\}$ and $L_2(x, y) = \max\{K - \min\{x, y\}, 0\}$. Please note that the basis functions only depend on the asset price S_t not the volatility $\exp(X_t)$, so the suboptimal policy is $\mathscr{F}_t^Y$-adapted and the results are guaranteed to be lower bounds for the partially observable model. In the tables, "*UB*" represents the corresponding upper bound yielded by our filtering-based duality method for the partially observable model, and "$Full.\widetilde{UB}$" represents the corresponding upper bound yielded by the PD method for the fully observable model. It is clear that we can improve the exercise policy for the fully observable model by employing more basis functions that use the information of the volatility $\exp(X_t)$: "$Full.LB$" and "$Full.UB$" are the lower bound and upper bound for the fully observable model, still obtained by the LSMC method and PD method with additional basis functions for each problem:

$$H_1^{add} = \{L_0(e^{X_t^1}), L_0(e^{X_t^1})L_1(S_t^1)\}$$
$$H_2^{add} = \{L_0(e^{X_t^1}), L_0^2(e^{X_t^1}), L_0(e^{X_t^2}), L_0^2(e^{X_t^2}), L_0(e^{X_t^1})L_2(S_t^1,S_t^2), L_0(e^{X_t^2})L_2(S_t^1,S_t^2)\}.$$

Each entry in Table II and Table III shows the sample average and the standard error (in parentheses) of the numerical results of 20 independent runs using the following procedure: we implement the LSMC method with 50000 sample paths to obtain a suboptimal policy τ, and then apply this policy on another independent set of 50000 paths to get the lower bound LB; the dual upper bound UB is obtained by implementing Algorithm 1 using the suboptimal policy τ with the number of sample paths $N = 500$, number of particles $m = 500$, and number of subpaths $l = 10$; to investigate the option prices under the fully observable stochastic volatility, we use the PD method with 500 sample paths and 5000 subpaths in nested simulation (which is equal to $m \cdot l$) to obtain an upper bound $Full.\widetilde{UB}$, since the policy τ obtained before is also a suboptimal policy for the fully observable model. Except the new sets of basis functions, the LSMC and PD methods are implemented exactly the same way as before to generate another set of lower bound $Full.LB$ and upper bound $Full.UB$ for the fully observable model. In practice we often use the average of LB and UB, and the average of $Full.LB$ and $Full.UB$ as estimates of the option prices to the partially observable and fully observable problems, respectively.

TABLE I
PARAMETER SETS

#	(θ,λ,γ)	μ
1	(log(0.1), 1.0, 1.0)	0.3
2	(log(0.1), 1.0, 1.0)	1.0
3	(log(0.2), 0.5, 1.0)	0.3
4	(log(0.2), 0.5, 1.0)	1.0
5	(log(0.2), 1.5, 1.0)	0.3
6	(log(0.2), 1.5, 1.0)	1.0
7	(log(0.2), 1.0, 0.5)	0.3
8	(log(0.2), 1.0, 0.5)	1.0
9	(log(0.3), 2.0, 0.3)	0.3
10	(log(0.3), 2.0, 0.3)	1.0

TABLE II
AMERICAN PUT OPTION PRICES ON ONE ASSET ($r = 0.05$, $K = 40$, $\delta = 0.1$, $T = 10$, $S_0 = 36$, $X_0 = \theta$)

	Volatility not observable		Volatility directly observable		
#	LB	UB	$Full.\widetilde{UB}$	$Full.LB$	$Full.UB$
1	3.820(0.000)	3.820(0.000)	3.825(0.001)	3.820(0.000)	3.821(0.000)
2	3.853(0.001)	3.887(0.001)	3.954(0.003)	3.905(0.002)	3.912(0.001)
3	3.892(0.001)	4.019(0.003)	4.321(0.005)	4.197(0.003)	4.209(0.001)
4	5.009(0.006)	5.216(0.005)	5.368(0.009)	5.297(0.005)	5.328(0.001)
5	3.881(0.001)	3.898(0.001)	3.995(0.004)	3.928(0.002)	3.938(0.001)
6	4.842(0.003)	4.935(0.002)	5.028(0.003)	4.973(0.004)	4.997(0.001)
7	3.869(0.001)	3.870(0.000)	3.876(0.001)	3.871(0.001)	3.872(0.000)
8	4.632(0.002)	4.653(0.001)	4.704(0.002)	4.679(0.003)	4.689(0.001)
9	4.010(0.001)	4.022(0.001)	4.049(0.001)	4.030(0.001)	4.044(0.001)
10	5.881(0.003)	5.902(0.001)	5.907(0.001)	5.896(0.005)	5.904(0.001)

The numerical results are divided into two categories: the first six rows report the numerical results under the dominant volatility effects, i.e., γ is comparatively large and λ is comparatively small; the last four rows report the results under moderate/weak volatility effects. It can be seen from the tables that $[Full.LB, Full.UB]$ is usually a tighter interval than $[LB, Full.\widetilde{UB}]$ for the fully observable option price, since more information is used to determine a better exercise policy. To differentiate the option prices under full and partial observations of stochastic volatility, [10] pointed out that the partial observation of stochastic volatility has an impact especially when the effect of the volatility $(i.e., \frac{\gamma^2}{2\lambda})$ is high. Our numerical results also support their viewpoints in terms of the differences between UB and $Full.\widetilde{UB}$, which demonstrate the effectiveness of introducing the filtering step. In particular, it can be observed that we can reduce relatively more overpricing for problems with dominant volatility (i.e., the first category). Considering the differences between LB and $Full.UB$, partially observable and fully observable option prices have relatively small gaps under moderate/weak volatility effects compared with the gaps in the first category. Larger observation noise μ challenges the performance of suboptimal exercise policy and also deteriorates the performance of particle filtering, so it generally increases the gap between $Full.LB$ and $Full.UB$ and the gap between LB and UB. Compared with [10] and [8], whose approaches provide asymptotic lower bounds on the option prices, our main contribution is to provide an asymptotic upper bound on the option price, which is less than or similar to the lower bound $(Full.LB)$ of the corresponding fully observable option price in the first category. Hence, our method provides a better criterion to evaluate the performance of LB: the smaller the gap between UB and LB, the better the bounds. If the gap between UB and LB is small enough, they can be both regarded as approximate option prices under partial observation. Otherwise, improvement on the exercise policy should be considered.

TABLE III
AMERICAN PUT OPTION PRICES ON THE MINIMUM OF TWO ASSETS ($r = 0.05$, $K = 40$, $\delta = 0.1$, $T = 10$, $S_0 = (36,36)^\top$, $X_0 = (\theta,\theta)^\top$)

	Volatility not observable		Volatility directly observable		
#	LB	UB	$Full.\widetilde{UB}$	$Full.LB$	$Full.UB$
1	4.027(0.002)	4.032(0.001)	4.068(0.002)	4.039(0.001)	4.043(0.001)
2	5.004(0.006)	5.147(0.004)	5.256(0.006)	5.143(0.005)	5.222(0.003)
3	5.274(0.005)	5.378(0.002)	5.565(0.004)	5.467(0.004)	5.489(0.001)
4	8.045(0.006)	8.171(0.004)	8.289(0.006)	8.188(0.010)	8.268(0.003)
5	4.641(0.002)	4.782(0.001)	4.918(0.005)	4.833(0.006)	4.870(0.001)
6	7.531(0.006)	7.638(0.002)	7.723(0.007)	7.606(0.007)	7.704(0.002)
7	4.429(0.002)	4.456(0.001)	4.514(0.001)	4.477(0.002)	4.500(0.001)
8	6.984(0.004)	7.042(0.003)	7.074(0.004)	6.997(0.007)	7.080(0.001)
9	5.417(0.002)	5.428(0.001)	5.449(0.001)	5.431(0.003)	5.447(0.001)
10	9.084(0.006)	9.130(0.002)	9.138(0.002)	9.071(0.009)	9.133(0.002)

VI. CONCLUSION

In this note we propose a numerical approach to solve for the value function of the partially observable optimal stopping problem. We represent the value function as a solution of a dual minimization problem, based on which we develop an algorithm that complements a suboptimal stopping time with an asymptotic upper bound on the value function. Our approach provides a practical way to judge whether more computational effort is needed to improve the quality of the approximate solution. We apply our approach to price American put options in stochastic volatility models, with the realistic assumption that the volatility cannot be directly observed but can be inferred from the asset prices. The numerical results confirm a higher price of the option if we alternatively assume that the volatility is directly observable. The price difference is more significant when the effect of volatility is high, indicating the importance of taking the partial observability into account.

APPENDIX
PROOF OF THEOREM 2

We need the following proposition for the proof of the theorem.

Proposition 1 (Corollary 10.28, [2]). *Let $\{\pi_0^m, \ldots, \pi_T^m\}$ be the random measure generated by Algorithm 2 for the observation sequence $y_{0:T}$. Suppose that the following assumption holds:*

$$\| f \|_\infty < \infty \quad and \quad \sup_{x_t} p(y_t|x_t, y_{t-1}) < \infty, \quad t = 1, \ldots, T.$$

Then

$$\mathbb{E}\left[\left(\int_{\mathscr{X}} f(x_t)\pi_t(x_t)dx_t - \int_{\mathscr{X}} f(x_t)\pi_t^m(x_t)dx_t\right)^2\right] \leq \frac{k_t^2 \parallel f \parallel_\infty^2}{m}, \quad t = 0,\ldots,T,$$

where the constant k_t does not depend on m (but it dose depend on t and $y_{0:t}$). In particular, $k_0 = 1$.

Proof of Theorem 2: We first prove (17). Given a sample path of the observations $\{y_0,\ldots,y_T\}$, the difference of $\tilde{g}(\pi_t,y_t)$ and $\tilde{g}(\pi_t^m,y_t)$ is

$$\vartheta_t^m \triangleq \int_{\mathscr{X}} g(x_t,y_t)\pi_t(x_t)dx_t - \int_{\mathscr{X}} g(x_t,y_t)\pi_t^m(x_t)dx_t.$$

Guaranteed by Proposition 1, $\mathbb{E}[|\vartheta_t^m|] \leq \sqrt{\mathbb{E}[(\vartheta_t^m)^2]} \leq \frac{k_t\|g\|_\infty}{\sqrt{m}}$ for some constant k_t. The difference between M_t^τ and $\tilde{M}_t^m$ is the sum of the differences between Δ_t^τ and $\tilde{\Delta}_t^m$:

$$\Delta_t^\tau - \tilde{\Delta}_t^m = \chi_{t,t}^m - \chi_{t-1,t}^m + \varepsilon_{t,t}^{m,l} - \varepsilon_{t-1,t}^{m,l},$$

where

$$\begin{aligned}
\chi_{t,t}^m &\triangleq \mathbb{E}[g(X_{\tau_t},Y_{\tau_t})|\pi_t,y_t] - \mathbb{E}[g(X_{\tau_t},Y_{\tau_t})|\pi_t^m,y_t],\\
\chi_{t-1,t}^m &\triangleq \mathbb{E}[g(X_{\tau_t},Y_{\tau_t})|\pi_{t-1},y_{t-1}] - \mathbb{E}[g(X_{\tau_t},Y_{\tau_t})|\pi_{t-1}^m,y_{t-1}],\\
\varepsilon_{t,t}^{m,l} &\triangleq \mathbb{E}[g(X_{\tau_t},Y_{\tau_t})|\pi_t^m,y_t] - G_{t,t}^{m,l},\\
\varepsilon_{t-1,t}^{m,l} &\triangleq \mathbb{E}[g(X_{\tau_t},Y_{\tau_t})|\pi_{t-1}^m,y_{t-1}] - G_{t-1,t}^{m,l}.
\end{aligned}$$

The first two errors are filtering errors, since we can rewrite $\chi_{t,t}^m$ as

$$\begin{aligned}
\chi_{t,t}^m &= \mathbb{E}\left[\sum_{j=t}^T g(X_j,Y_j)1_{\{\tau_t=j\}}|\pi_t,y_t\right] - \mathbb{E}\left[\sum_{j=t}^T g(X_j,Y_j)1_{\{\tau_t=j\}}|\pi_t^m,y_t\right]\\
&= \int_{\mathscr{X}} I_t(x_t,y_t)\pi_t(x_t)dx_t - \int_{\mathscr{X}} I_t(x_t,y_t)\pi_t^m(x_t)dx_t. \qquad (21)
\end{aligned}$$

$I_t(x_t,y_t)$ is defined as the integrand of $\mathbb{E}[\sum_{j=t}^T g(X_j,Y_j)1_{\{\tau_t=j\}}|\pi_t,y_t]$, i.e.,

$$I_t(x_t,y_t) \triangleq g(x_t,y_t)1_{\{\tau_t=t\}} + \sum_{j=t+1}^T \int g(x_j,y_j)1_{\{\tau_t=j\}}p(dx_{t+1}dy_{t+1}\ldots dx_jdy_j|x_t,y_t),$$

where $p(dx_{t+1}dy_{t+1}\ldots dx_jdy_j|x_t,y_t)$ denotes the joint probability distribution of $(x_{t+1},y_{t+1},\ldots,x_j,y_j)$ conditional on (x_t,y_t). As $\{\tau_t = j\}$ are disjoint sets for each $t \leq j \leq T$, it implies $\parallel I_t \parallel_\infty \leq \parallel g \parallel_\infty$. Based on (21) and using Proposition 1 with $f = I_t$, it is ensured that $\mathbb{E}[|\chi_{t,t}^m|] \leq \frac{k_t^{'}\|g\|_\infty}{\sqrt{m}}$ for some constant $k_t^{'}$. Similarly, $\mathbb{E}[|\chi_{t-1,t}^m|] \leq \frac{b_{t-1}^{'}\|g\|_\infty}{\sqrt{m}}$ for some constant $b_{t-1}^{'}$. The latter two errors are from the sampling variability of Monte Carlo simulation (as step 1 in Algorithm 2); the error bounds are guaranteed by Proposition 1 with $t = 0$, i.e., $\mathbb{E}[|\varepsilon_{t,t}^{m,l}|] \leq \frac{\|g\|_\infty}{\sqrt{ml}}$ and $\mathbb{E}[|\varepsilon_{t-1,t}^{m,l}|] \leq \frac{\|g\|_\infty}{\sqrt{ml}}$.

So given a sample path of the observations $y_{0:t}$ we have for each $t \in \mathscr{J}$,

$$\begin{aligned}
&\lim_{m\to\infty}\mathbb{E}[|(\tilde{g}(\pi_t,y_t) - M_t^\tau) - (\tilde{g}(\pi_t^m,y_t) - \tilde{M}_t^m)|]\\
&= \lim_{m\to\infty}\mathbb{E}[|\vartheta_t^m + (\sum_{i=1}^t(\tilde{\Delta}_i^m - \Delta_i^\tau))|] = 0. \qquad (22)
\end{aligned}$$

Since

$$\begin{aligned}
&|\max_{t\in\mathscr{J}}\{\tilde{g}(\pi_t,y_t) - M_t^\tau\} - \max_{t\in\mathscr{J}}\{\tilde{g}(\pi_t^m,y_t) - \tilde{M}_t^m\}|\\
\leq&\max_{t\in\mathscr{J}}\{|(\tilde{g}(\pi_t,y_t) - M_t^\tau) - (\tilde{g}(\pi_t^m,y_t) - \tilde{M}_t^m)|\}\\
\leq&\sum_{t=1}^T|(\tilde{g}(\pi_t,y_t) - M_t^\tau) - (\tilde{g}(\pi_t^m,y_t) - \tilde{M}_t^m)|,
\end{aligned}$$

by taking expectation and letting m go to infinity we have

$$\lim_{m\to\infty}\mathbb{E}[|\max_{t\in\mathscr{J}}\{\tilde{g}(\pi_t^m,y_t) - \tilde{M}_t^m\} - \max_{t\in\mathscr{J}}\{\tilde{g}(\pi_t,y_t) - M_t^\tau\}|] = 0.$$

Note that $\tilde{\Delta}_t^m$ is bounded by $2 \parallel g \parallel_\infty$ for each $t \in \mathscr{J}$, and therefore, $\tilde{g}(\Pi_t^m,Y_t) - \tilde{M}_t^m$ is bounded by $(2t+1)\cdot \parallel g \parallel_\infty$ and $\max_{t\in\mathscr{J}}\{\tilde{g}(\Pi_t^m,Y_t) - \tilde{M}_t^m\}$ is bounded by $(2T+1)\cdot \parallel g \parallel_\infty$. The same conclusions are also valid for Δ_t^τ, $\tilde{g}(\Pi_t,Y_t) - M_t^\tau$ and $\max_{t\in\mathscr{J}}\{\tilde{g}(\Pi_t,Y_t) - M_t^\tau\}$. Then

$$\begin{aligned}
&\lim_{m\to\infty}\mathbb{E}_0\big[|\max_{t\in\mathscr{J}}\{\tilde{g}(\Pi_t^m,Y_t) - \tilde{M}_t^m\} - \max_{t\in\mathscr{J}}\{\tilde{g}(\Pi_t,Y_t) - M_t^\tau\}|\big]\\
&= \lim_{m\to\infty}\mathbb{E}_0\big[\mathbb{E}\big[|\max_{t\in\mathscr{J}}\{\tilde{g}(\Pi_t^m,Y_t) - \tilde{M}_t^m\} - \max_{t\in\mathscr{J}}\{\tilde{g}(\Pi_t,Y_t) - M_t^\tau\}|\big|\mathscr{F}_T^Y\big]\big]\\
&=\mathbb{E}_0\big[\lim_{m\to\infty}\mathbb{E}\big[|\max_{t\in\mathscr{J}}\{\tilde{g}(\Pi_t^m,Y_t) - \tilde{M}_t^m\} - \max_{t\in\mathscr{J}}\{\tilde{g}(\Pi_t,Y_t) - M_t^\tau\}|\big|\mathscr{F}_T^Y\big]\big]\\
&=0,
\end{aligned}$$

where the second equality follows from the boundedness of the integrand and the dominated convergence theorem. Hence,

$$\lim_{m\to\infty}\mathbb{E}_0[\max_{t\in\mathscr{J}}\{\tilde{g}(\Pi_t^m,Y_t) - \tilde{M}_t^m\}] = \mathbb{E}_0[\max_{t\in\mathscr{J}}\{\tilde{g}(\Pi_t,Y_t) - M_t^\tau\}].$$

Now we prove (18). First we have

$$\begin{aligned}
&\mathbb{E}_0[\max_{t\in\mathscr{J}}\{\tilde{g}(\Pi_t,Y_t) - M_t^\tau\}] - V_0\\
=&\mathbb{E}_0[\max_{t\in\mathscr{J}}\{\tilde{g}(\Pi_t,Y_t) - M_t^\tau\}] - \mathbb{E}_0[\max_{t\in\mathscr{J}}\{\tilde{g}(\Pi_t,Y_t) - M_t^*\}]\\
\leq&\mathbb{E}_0[\max_{t\in\mathscr{J}}\{M_t^* - M_t^\tau\}],
\end{aligned}$$

following the fact that

$$\max_{t\in\mathscr{J}}\{\tilde{g}(\Pi_t,Y_t) - M_t^\tau\} - \max_{t\in\mathscr{J}}\{\tilde{g}(\Pi_t,Y_t) - M_t^*\} \leq \max_{t\in\mathscr{J}}\{M_t^* - M_t^\tau\}.$$

Then (18) follows from

$$\begin{aligned}
&\mathbb{E}_0[\max_{t\in\mathscr{J}}\{M_t^* - M_t^\tau\}]\\
\leq&2\sqrt{\mathbb{E}_0[(M_T^* - M_T^\tau)^2]}\\
=&2\sqrt{\sum_{t=1}^T\mathbb{E}_0\left[\left((M_t^* - M_t^\tau) - (M_{t-1}^* - M_{t-1}^\tau)\right)^2\right]}\\
=&2\sqrt{\sum_{t=1}^T\mathbb{E}_0[(\Delta_t^* - \Delta_t^\tau)^2]}\\
\leq&2\sqrt{\sum_{t=1}^T\mathbb{E}_0\left[\left(\mathbb{E}[g(X_{\tau_t^*},Y_{\tau_t^*})|\Pi_t,Y_t] - \mathbb{E}[g(X_{\tau_t},Y_{\tau_t})|\Pi_t,Y_t]\right)^2\right]},
\end{aligned}$$

where the first inequality follows from the fact that $M_t^* - M_t^\tau$ is a martingale and applying Doob's martingale inequality, and the first equality uses the orthogonality property of martingale difference (see p.331 in [12]). To show the last inequality, recall that

$$\begin{aligned}
\Delta_t^* - \Delta_t^\tau =&(\mathbb{E}[g(X_{\tau_t^*},Y_{\tau_t^*})|\mathscr{F}_t^Y] - \mathbb{E}[g(X_{\tau_t},Y_{\tau_t})|\mathscr{F}_t^Y])\\
&- (\mathbb{E}[g(X_{\tau_t^*},Y_{\tau_t^*})|\mathscr{F}_{t-1}^Y] - \mathbb{E}[g(X_{\tau_t},Y_{\tau_t})|\mathscr{F}_{t-1}^Y]);
\end{aligned}$$

then the last inequality can be shown by simple algebra and iterated expectation on $\mathscr{F}_{t-1}^Y$.

■

References

[1] L. Andersen and M. Broadie. Primal-dual simulation algorithm for pricing multidimensional American options. *Management Science*, 50(9):1222 – 1234, 2004.

[2] A. Bain and D.Crisan. *Fundamentals of Stochastic Filtering*. Springer, 2008.

[3] D.P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, 3rd edition, 2007.

[4] I. Florescu and F. Viens. Stochastic volatility: Option pricing using a multinomial recombining tree. *Applied Mathematical Finance*, 15(2):151 – 181, 2008.

[5] M. B. Haugh and L. Kogan. Pricing American options: A duality approach. *Operations Research*, 52(2):258 – 270, 2004.

[6] D. Lamberton and B. Lapeyre. *Introduction to stochastic calculus applied to finance*. Chapman & Hall/CRC, 2007.

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

[8] M. Ludkovski. A simulation approach to optimal stopping under partial information. *Stochastic Processes and Applications*, 119(12):2071 – 2087, 2009.

[9] H. Pham, W. Runggaldier, and A. Sellami. Approximation by quantization of the filter process and applications to optimal stopping problems under partial observation. *Monte Carlo Methods and Applicaitons*, 11(1):57 – 81, 2005.

[10] B. R. Rambharat and A. E. Brockwell. Sequential Monte Carlo pricing of American-style options under stochastic volatility models. *The Annals of Applied Statistics*, 4, No. 1, 222-265(1):222 – 265, 2010.

[11] L. C. G. Rogers. Monte Carlo valuation of American options. *Mathematical Finance*, 12(3):271 – 286, 2002.

[12] S.Karlin and H. Taylor. *A First Course in Stochastic Process,2nd edn*. Academic Press, San Diego, 1975.

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

[14] E. Zhou. Optimal stopping under partial observation: Near-value iteration. 2011. Forthcoming in *IEEE Transactions on Automatic Control*.