

# Homework #7

Junwu Zhang

CME 241: Reinforcement Learning for Finance

January 29, 2020

## Problem 3.

Explore/Discuss an Approximate Dynamic Programming solution as an alternative to Longstaff-Schwartz Algorithm.

*Solution.*

Longstaff-Schwartz algorithm combines 3 ideas:

- Valuation based on Monte-Carlo simulation
- Function approximation of continuation value for in-the-money states
- Backward-recursive determination of early exercise states

Another possible Approximate Dynamic Programming (ADP) method for American options is by Tsitsiklis and Van Roy on regression methods for American-style options pricing. In this method, the problem can be simplified to optimal stopping.

The Markov Decision Process (MDP) can be modeled as:

*States* are:  $x$  that represents a risk-neutral state process that is assumed to be Markov, where  $x$  is constrained to be:  $x_t \in \mathcal{X} | 0 \leq t \leq T$ .

*Actions* is:  $t$ , more specifically, the optimal time to stop (in other words, *exercise*) the option

*Rewards*: The reward equal to the intrinsic value of the option, discounted at the risk-free rate, is received at termination. The expected reward is therefore the price of the option, given by:

$$\sup_{\tau \in [0, T]} E[e^{-r\tau} g(x_\tau)] \quad (1)$$

where  $x_t$  is defined in the states section above,  $r$  is the risk-free interest rate (assumed to be a known constant),  $g(x)$  is the intrinsic value of the option when the state is  $x$  and  $T$  is the final expiration time.

In terms of *approximation*, the authors argue that the first step is to introduce a parameterized value function, so we need to choose for each  $n$ , a parameter vector  $r_n$  so that  $\tilde{J}(x, r_n) \approx J_n(x)$ . For algorithms that compute appropriate parameter values, the author discussed several possible variants of value iteration. The approximation architecture based on features are not discussed due to lack of analytical trackability. Rather, the authors discussed approximation of value function itself. Sample-based projection and Q-values are also discussed, and a version of approximate value iteration is defined. The authors looked into

the error accumulation of such approximation and how to resolve the errors, and eventually experimented the algorithms on simulated American options trading and received good results. ■