Abstract

Prioritized sweeping (PS) and its variants are model-based reinforcement-learning algorithms that have demonstrated superior performance in terms of computational and experience efficiency in practice. This note establishes the first—to the best of our knowledge—formal proof of convergence to the optimal value function when they are used as *planning* algorithms. We also describe applications of this result to provably efficient model-based reinforcement learning in the PAC-MDP framework. We do not address the issue of convergence rate in the present paper.