PRIMAL-DUAL ACCELERATED GRADIENT ALGORITHM FOR A STOCHASTIC MULTI-ARMED BANDIT GOVERNED BY A STATIONARY FINITE MARKOV CHAIN

A.V. Nazin, Trapeznikov Institute of Control Sciences, Russia, nazine@ipu.ru
B.M. Miller, Institute for Transmission Problems (Kharkevich Institute), Russia, bmiller@iitp.ru

The class of partially observable Markov decision processes with finite number of states and control values constitutes the basis of the queuing systems optimization under uncertainty. In this article, the adaptive control problem for a class of homogeneous finite Markov chains governed by a stochastic multi-armed bandit with unknown mean losses is considered. Using the problem statement as in (Nazin and Miller, AUCC 2013, Perth, Australia), one can apply a novel primal-dual algorithm which generalizes Nesterov’s accelerated gradient method for convex optimization. This algorithm may be considered as a non-trivial modification of the mirror descent randomized control algorithm which was proposed and studied in (Nazin and Miller, AUCC 2013, Perth, Australia). Here the explicit, non-asymptotic bounds for the mean losses at a given time horizon has been discussed and proved. Numerical example illustrates both theoretical and simulation results for suggested algorithm.