

Supplementary for the paper “A Convergent Off-Policy Temporal Difference Algorithm”

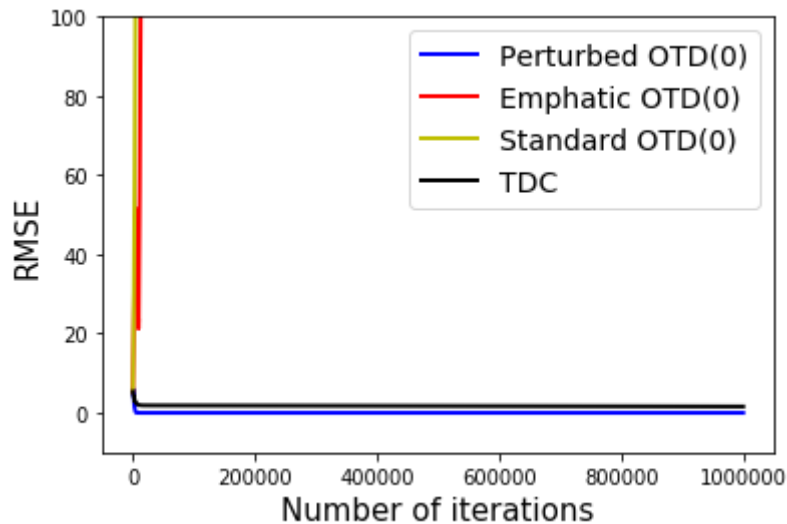
In this document, we present the results of our experiments ran with various step-sizes. In the paper, for the Experiment -2 and 3, step-sizes are fixed to 0.0001. The following are the results obtained with other step-sizes.

1. Experiment 2: Baird's Counter Example -

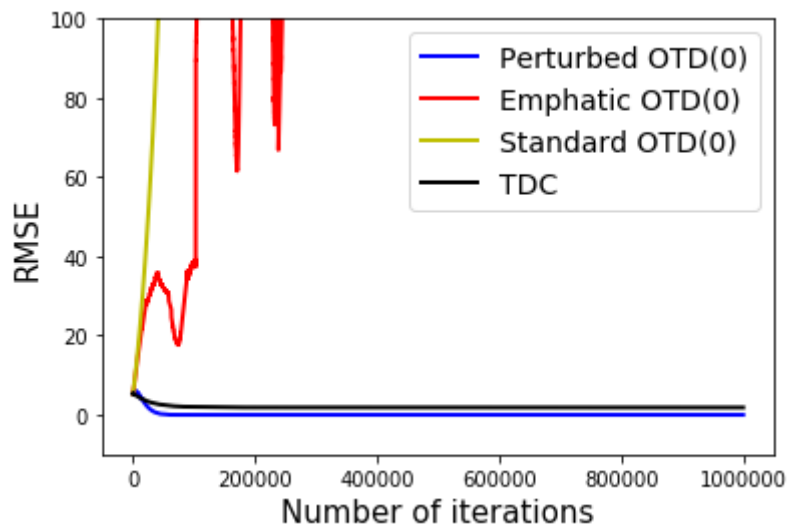
Number of Independent Runs	10
η parameter	1
Number of iterations	1000000
β (step-size for TDC update)	10 * step_size.

Table 1 : Parameters of the Experiment-2.

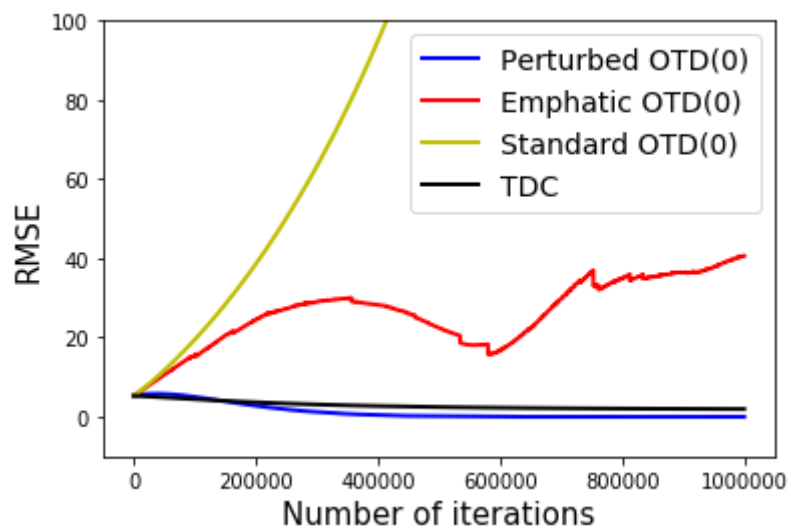
a. Step-size: 0.001



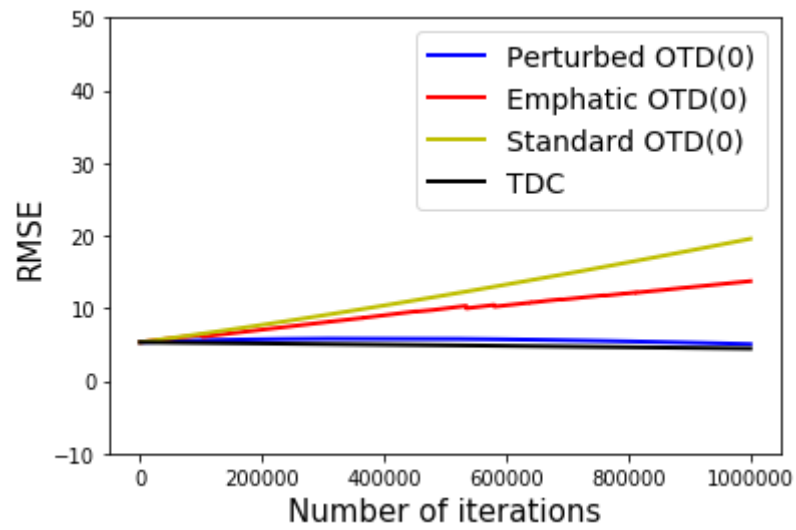
b. Step-size: 0.0001



c. Step-size: 0.00001



d. Step-size: 0.000001

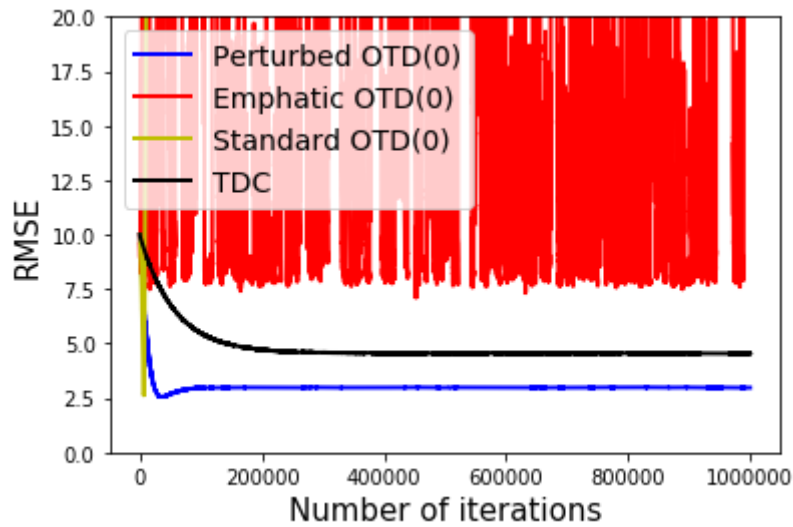


2. Experiment 3: A 3-state MDP problem

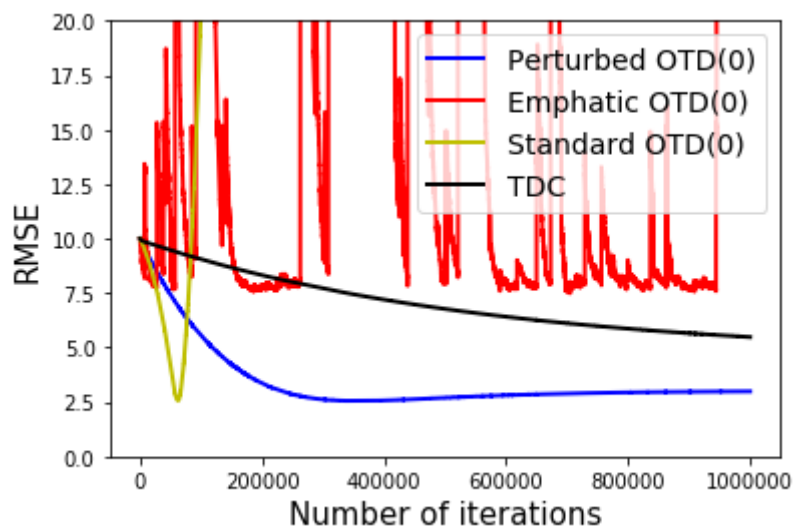
Number of Independent Runs	10
η parameter	0.5
Number of iterations	1000000
β (step-size for TDC update)	$10 * \text{step_size}$

Table 2: Parameters of the Experiment-3.

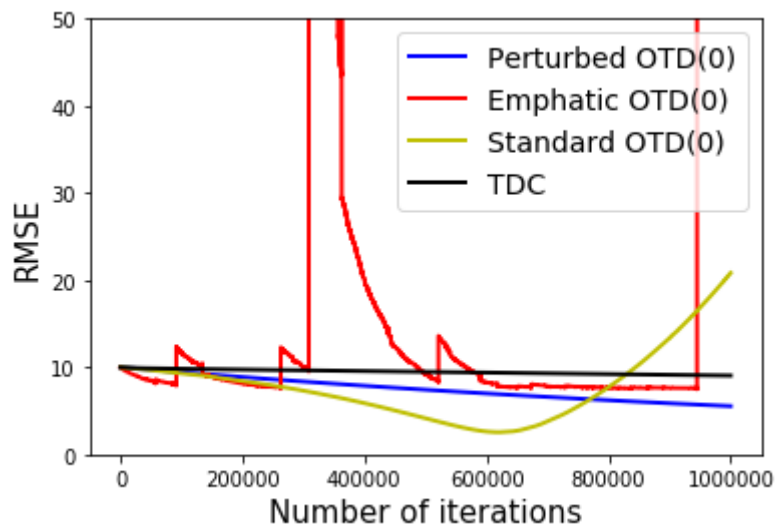
a. Step-size: 0.001



b. Step-size: 0.0001



c. Step-size: 0.00001



d. Step-size: 0.000001

