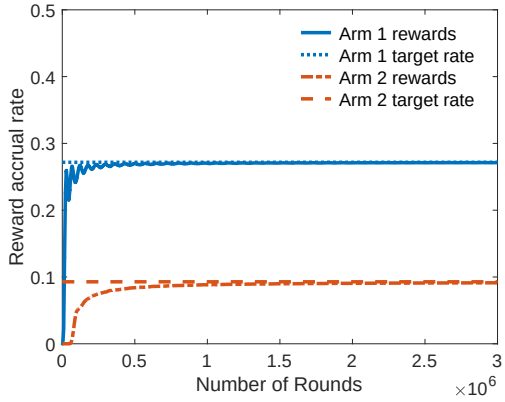

Figure 1: Reward accrual rates in the full-information setting

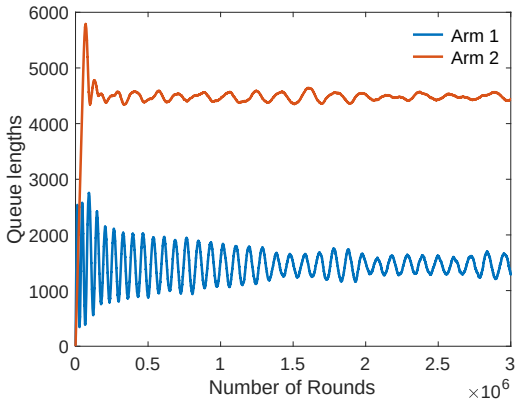

Figure 2: Queue lengths in the full-information setting

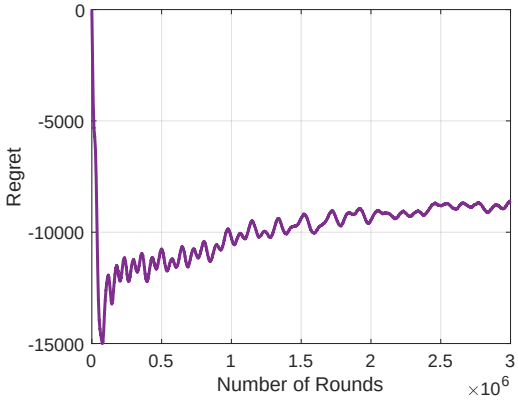

Figure 3: Regret of BANDITQ in the full-information setting

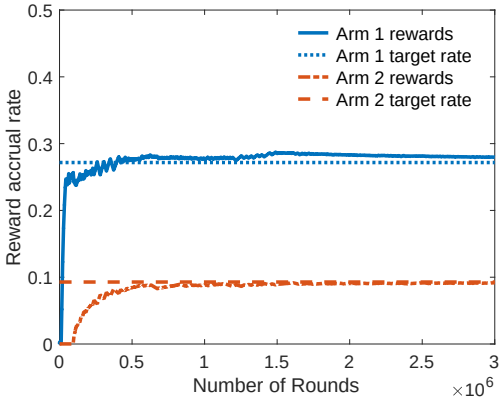

Figure 4: Reward accrual rates in the bandit feedback

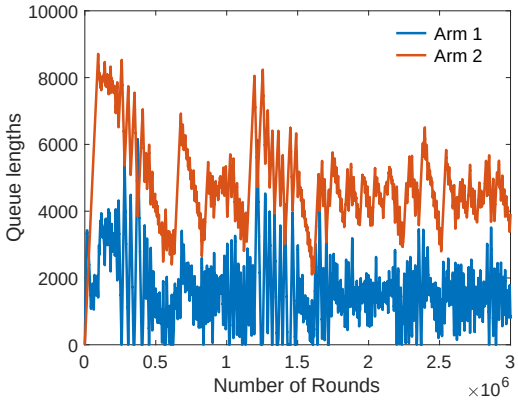

Figure 5: Queue lengths in the bandit feedback setting

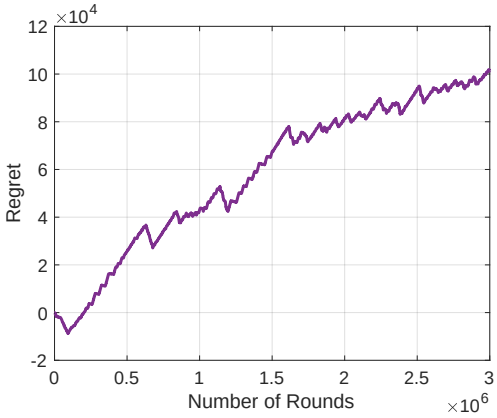

Figure 6: Regret of BANDITQ in the bandit feedback setting

Figure 7: Performance of the BanditQ policy with $N = 1000$ arms in both full information and bandit feedback setting