Constant Space Algorithm for the Stochastic Multi-armed Bandit Problem : Supplementary File

Appendix

A Proof of Lemma 5.3

Proof. Assume the contrary, i.e. at the end of round $r = \lceil (\frac{2}{\epsilon}+1)\frac{\log 2/\Delta}{\log\log 2/\Delta} + 2\rceil$, the best arm and the second best arm are still not differentiated, meaning for some $i \neq *$, we still have

$$\overline{\mu}_*^{(r)} - g_r/2 < \overline{\mu}_i^{(r)} + g_r/2$$

By Lemma 5.2, we have $g_r \leq \Delta/2$. Thus, we have

$$\mu_* \leq \overline{\mu}_*^{(r)} + g_r/2 \leq \overline{\mu}_i^{(r)} + 3g_r/2 \leq \overline{\mu}_i^{(r)} + 3\Delta/4$$

where for the second step we use Lemma 4.3. Similarly, we have$\mu_i \geq \overline{\mu}_i^{(r)} - \Delta/4$. Then, we have

$$\Delta \leq \mu_* - \mu_i \leq (\overline{\mu}_i^{(r)} + 3\Delta/4) - (\overline{\mu}_{a^{(r)}}^{(r)} - \Delta/4) < \Delta$$

which results in a contradiction. □

B Proof of Theorem 6.1

Proof. We present the algorithm in Algorithm 2. The algorithm repeatedly calls the procedure in Algorithm 1 with increasing time horizons $T_0, T_1, \ldots, T_L$, where $L \leq \log\log T$. By setting $T_l = T_{l-1}^2$, we have $T_l = T_0^{2^l}$. Then, by Theorem 4.1, we can upper bound the regret as

$$\begin{aligned}\overline{\Psi}_T \lesssim\ & \sum_{l=0}^{L}\sum_{i=1}^{K}\frac{\log(\Delta_i/\Delta)\log T_l}{\Delta_i}\\ =\ & \sum_{l=0}^{L}\sum_{i=1}^{K}\frac{2^l\log(\Delta_i/\Delta)\log T_0}{\Delta_i}\\ \lesssim\ & \sum_{i=1}^{K}\frac{2^L\log(\Delta_i/\Delta)\log T_0}{\Delta_i}\\ \lesssim\ & \sum_{i=1}^{K}\frac{\log(\Delta_i/\Delta)\log T}{\Delta_i}\end{aligned}$$

which proves the theorem. □

C Discussion of Conjecture on the Lower Bound for Stochastic Bandits

For any given round r, for some $\alpha > 0$, define:

$$R_{\text{in}}^{(r)} = \sum_{i:\Delta_i<\alpha g_r}\frac{1}{g_r}, \qquad R_{\text{out}}^{(r)} = \sum_{i:\Delta_i>\alpha g_r}\frac{1}{\Delta_i}$$

That correspond to two terms in

$$\sum_{r=1}^{r_i}\Delta_i\cdot\frac{2\log(1/\delta)}{g_r^2} + \sum_{r=r_i+1}^{r_{\max}}\Delta_i\cdot\frac{2\log(1/\delta)}{(\Delta_i - g_{r-1})^2} + \Delta_i\cdot r_{\max} \tag{13}$$

which is the total regret provided within Section 5. Consider the following example where there is a group of high-value arms and a group of low-value arms, and the size of the low-value arms is larger than the high-value arms.

Example C.1. *Assume $1 > E \gg \epsilon$, and $s > 1/2$. Let $\Delta_i = \epsilon$ for $i = 1\ldots sK$, and $\Delta_i = E$ for $i = sK+1\ldots K$.*

In this example, we can find that as $g_r < E/2$, $R_{\text{in}}^{(r)} = sK/g_r$, and $R_{\text{out}}^{(r)} = (1-s)K/E$. Since $g_r \lesssim E$ and $s > 1/2$, we can find that $R_{\text{out}}^{(r)} \lesssim R_{\text{in}}^{(r)}$. This means that Example C.1 will not harm us if we use Algorithm 1 because we know that $\sum_r R_{\text{in}}^{(r)} \lesssim \sum_i 1/\Delta_i$.

Then, we consider another example where the size of the group of the high-value arms is larger than low-value arms. Particularly, we consider

Example C.2. *Assume $1 > E \gg \epsilon$, and $s < 1/2$, where $s/(1-s) < \epsilon/E$. Let $\Delta_i = \epsilon$ for $i = 1\ldots sK$, and $\Delta_i = E$ for $i = sK+1\ldots K$.*

We can find that in this example, as long as $\epsilon \lesssim g_r \lesssim E$, $R_{\text{in}}^{(r)} \lesssim sK/\epsilon \lesssim (1-s)/E = R_{\text{out}}^{(r)}$. This means that this is the hard case for Algorithm 1 because $R_{\text{out}}^{(r)}$ is dominating. However, we can deal with this example with the following update rule

$$g_{r+1} = \frac{g_r}{2\max\{1, (1-s)/s\}}$$

which is roughly $g_{r+1} = \frac{g_r}{2\max\{1, R_{\text{out}}^{(r)}/R_{\text{in}}^{(r)}\}}$. Note that if s is unknown, we can estimate it by simply counting the number of arms not ruled out. With the new update rule, we can find that as long as $g_r \lesssim E$, we have $g_{r+1} \lesssim Es/(1-s) \lesssim \epsilon$. This means that in the next round, we are able to identify the high-value arms. Therefore, the number of rounds is a constant.

Finally, we consider the following case where we conjectured to be the hard case:

Example C.3. *Let $\Delta_i = i/K$ for $i = 1, 2, \ldots, K$.*

First note that in this example, $R_{\text{in}}^{(r)} \approx n$ and $R_{\text{out}}^{(r)} \approx n \log 1/g_r$, where we can find that $R_{\text{in}}^{(r)} \lesssim R_{\text{out}}^{(r)}$ for any r. If we use the trick we are dealing with Example C.2, we can find that the corresponding update rule becomes $g_{r+1} = \frac{g_r}{2\log 1/g_r}$. Such rule is exactly (9). Therefore, we conjecture that the additional $\frac{\log(\Delta_i/\Delta)}{\log\log(\Delta_i/\Delta)}$ factor is not improvable.