A Supplementary material: Proofs

Before proving Theorems 1 and 2, we provide some preliminary results presented sections A.1 and A.2

A.1 Tail inequalities for vector-valued martingales

We need the following result about vector-valued martingales, extracted from [12].

Lemma 1. *Let $(\mathcal{F}_k; k \geq 0)$ be a filtration, $(m_k; k \geq 0)$ be an $\mathbb{R}^d$-valued stochastic process adapted to $(\mathcal{F}_k)$, $(\eta_k; k \geq 1)$ be a real-valued martingale difference process adapted to $(\mathcal{F}_k)$. Assume that η_k is conditionally sub-Gaussian in the sense that there exists some $R > 0$ such that for any $\gamma \geq 0$, $k \geq 1$,*

$$\mathbb{E}[\exp(\gamma \eta_k) \,|\, \mathcal{F}_{k-1}] \leq \exp\left(\frac{\gamma^2 R^2}{2}\right) \quad \text{a.s.} \tag{9}$$

Consider the martingale $\xi_t = \sum_{k=1}^t m_{k-1}\eta_k$ and the process $M_t = \sum_{k=1}^t m_{k-1}m_{k-1}'$. Assume that with probability one the smallest eigenvalue of M_d is lower bounded by some positive constant λ_0 and that $\|m_k\|_2 \leq c_m$ holds a.s. for any $k \geq 0$.

The following hold true: Let

$$\kappa = \sqrt{3 + 2\log(1 + 2c_m^2/\lambda_0)}. \tag{10}$$

For any $x \in \mathbb{R}^d$, $0 < \delta \leq 1/e$, $t \geq \max(d, 2)$, with probability at least $1 - \delta$,

$$|x'\xi_t| \leq \kappa R \sqrt{2\log t}\,\sqrt{\log(1/\delta)}\ \|x\|_{M_t}. \tag{11}$$

Further, for any $0 < \delta < \min(1, d/e)$, $t \geq \max(d, 2)$, with probability at least $1 - \delta$,

$$\|\xi_t\|_{M_t^{-1}} \leq \kappa R \sqrt{2\,d\log t}\,\sqrt{\log(d/\delta)}. \tag{12}$$

The proof of (11) is based on an exponential inequality of [16] and is adopted from that of Lemma B.4 of [17]. Given (11), inequality (12) follows by some algebra from (11).

Proof. In order to prove (11), we shall use Corollary 2.2 of [16] which states the following: Pick some random variables A and $B \geq 0$ such that

$$\mathbb{E}\left[\exp\left\{\gamma A - \frac{\gamma^2}{2}B^2\right\}\right] \leq 1 \quad \text{for all } \gamma \in \mathbb{R}\,. \tag{13}$$

Then, for all $c \geq \sqrt{2}$, and all $y > 0$,

$$\mathbb{P}\left(|A| \geq c\sqrt{(B^2 + y)\left(1 + \frac{1}{2}\log\left(\frac{B^2}{y} + 1\right)\right)}\right) \leq \exp\left\{-\frac{c^2}{2}\right\}\,. \tag{14}$$

We apply this inequality to the random variables $A = x'\xi_t/R$ and $B = \|x\|_{M_t}$, where $x \in \mathbb{R}^d$ is some fixed vector. We first check if the so-defined A, B satisfy (13). Pick any $\gamma \in \mathbb{R}$. We first study $\gamma A - (\gamma B)^2/2$. We have

$$\gamma A - (\gamma B)^2/2 = \frac{\gamma\, x'\xi_t}{R} - \frac{\gamma^2\, x' M_t x}{2} = \sum_{k=1}^t D_k\,,$$

where

$$D_k = \tfrac{\gamma}{R}\, x' m_{k-1}\eta_k - \tfrac{\gamma^2}{2} x'\, m_{k-1}m_{k-1}'\, x = \tfrac{\gamma}{R}\, x' m_{k-1}\eta_k - \tfrac{\gamma^2}{2}(x' m_{k-1})^2\,.$$

Now, observe that thanks to (9), $\mathbb{E}\left[\exp(D_k) \,|\, \mathcal{F}_{k-1}\right] \leq 1$. Let $P_k = \exp(D_k)$. Noting that P_k is $\mathcal{F}_k$-adapted,

$$\begin{aligned}
\mathbb{E}\left[\exp(\gamma A - \gamma B^2/2)\right] &= \mathbb{E}\left[P_1 \cdots P_{t-1}P_t\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[P_1 \cdots P_{t-1}P_t \,|\, \mathcal{F}_{t-1}\right]\right] = \mathbb{E}\left[P_1 \cdots P_{t-1}\, \mathbb{E}\left[P_t \,|\, \mathcal{F}_{t-1}\right]\right] \\
&\leq \mathbb{E}\left[\mathbb{E}\left[P_1 \cdots P_{t-1} \,|\, \mathcal{F}_{t-2}\right]\right] = \mathbb{E}\left[P_1 \cdots P_{t-2}\mathbb{E}\left[P_{t-1} \,|\, \mathcal{F}_{t-2}\right]\right] \\
&\ \ \vdots \\
&\leq \mathbb{E}\left[\mathbb{E}\left[P_1 \,|\, \mathcal{F}_0\right]\right] \leq 1
\end{aligned}$$

which finishes the verification of (13). Now, choose $y = \lambda_0 \|x\|_2^2$ to get from (14) that for all $0 < \delta \le 1/e$, $t \ge 1$, with probability $1-\delta$,

$$|x'\xi_t| \le R\sqrt{\left(\|x\|_{M_t}^2 + \lambda_0\|x\|_2^2\right)\left(1+\frac{1}{2}\log\left(1+\frac{\|x\|_{M_t}^2}{\lambda_0\|x\|_2^2}\right)\right)}\sqrt{2\log\left(\frac{1}{\delta}\right)}\,. \tag{15}$$

Noting that for $t \ge \max(d,2)$, $\lambda_0\|x\|_2^2 \le \|x\|_{M_t}^2 \le t\|x\|_2^2 c_m^2$, we have $\|x\|_{M_t}^2 + \lambda_0\|x\|_2^2 \le 2\|x\|_{M_t}^2$ and $1+\frac{1}{2}\log\left(1+\frac{\|x\|_{M_t}^2}{\lambda_0\|x\|_2^2}\right) \le 1+\frac{1}{2}\log\left(1+\frac{tc_m^2}{\lambda_0}\right) \le \kappa^2\log(t)/2$, thanks to the definition of κ. Indeed, it is easy to verify that the slope of function $1+\frac{1}{2}\log(1+c_m^2 t/\lambda_0)$ is below that of $\kappa^2\ \log(t)/2$ for any $t \ge 1$ provided that $\kappa \ge 1$. Hence, the last inequality holds if it holds true for $t=2$, which, after reordering the terms gives the constraint

$$\kappa \ge \sqrt{\frac{2+\log(1+2c_m^2/\lambda_0)}{\log 2}}.$$

Upper bounding $2/\log 2$ by 3 and $1/\log 2$ by 2, we get the definition of κ, which indeed satisfies $\kappa \ge 1$.

Hence, when (15) holds, it also holds that

$$|x'\xi_t| \le \kappa R\|x\|_{M_t}\sqrt{\log(t)}\sqrt{2\log\left(\frac{1}{\delta}\right)}\,. \tag{16}$$

which is exactly (11).

Now, let us turn to proving (12). Denote by S_t the symmetric, positive definite matrix such that $S_t^2 = M_t$ and, for all $1 \le i \le d$, let $\mathbf{e}_i$ be the i^{th} unit vector (i.e., for all $j \ne i$, $\mathbf{e}_{ij} = 0$ and $\mathbf{e}_{ii} = 1$). Noting that the identity matrix can be written as $I = \sum_{i=1}^d \mathbf{e}_i\mathbf{e}_i'$, we have $\|\xi_t\|_{M_t^{-1}}^2 = \xi_t' M_t^{-1}\xi_t = \xi_t' S_t^{-1} I S_t^{-1}\xi_t = \sum_{i=1}^d \xi_t' S_t^{-1}\mathbf{e}_i\mathbf{e}_i' S_t^{-1}\xi_t$. Therefore, for any constant $\tau > 0$,

$$\begin{aligned}\mathbb{P}\left[\|\xi_t\|_{M_t^{-1}}^2 \ge d\tau^2\right] &= \mathbb{P}\left[\sum_{i=1}^d \xi_t' S_t^{-1}\mathbf{e}_i\mathbf{e}_i' S_t^{-1}\xi_t \ge d\tau^2\right] \le \sum_{i=1}^d \mathbb{P}\left[\xi_t' S_t^{-1}\mathbf{e}_i\mathbf{e}_i' S_t^{-1}\xi_t \ge \tau^2\right] \\ &\le \sum_{i=1}^d \mathbb{P}\left[|\xi_t' S_t^{-1}\mathbf{e}_i| \ge \tau\right]\ .\end{aligned}$$

Applying (11) with $x = S_t^{-1}\mathbf{e}_i$, and $\tau = \kappa R\left\|S_t^{-1}\mathbf{e}_i\right\|_{M_t}\sqrt{\log(t)}\sqrt{2\log\left(\frac{d}{\delta}\right)}$, $0 < \delta < \min(1, d/e)$, $t \ge \max(d,2)$, and using the fact that $\left\|S_t^{-1}\mathbf{e}_i\right\|_{M_t} = 1$, we have

$$\mathbb{P}\left[\|\xi_t\|_{M_t^{-1}}^2 \ge 2d\kappa^2 R^2\log(t)\ \log\left(\frac{d}{\delta}\right)\right] \le \delta\,,$$

thus, finishing the proof. □

Remark 1. *Note that if $\eta_k \in [\alpha_k - R, \alpha_k + R]$ holds almost surely for some $\mathcal{F}_{k-1}$-measurable random variable α_k then, using Hoeffding's lemma (see, e.g., Lemma A.1 of [3]), we get that for all $\gamma \in \mathbb{R}$,*

$$\mathbb{E}\left[\exp\{\gamma\eta_k\}\,|\,\mathcal{F}_{k-1}\right] \le \exp\{\gamma\mathbb{E}\left[\eta_k\,|\,\mathcal{F}_{k-1}\right]\}\exp\left\{\frac{4R^2\gamma^2}{8}\right\} = \exp\left\{\frac{\gamma^2 R^2}{2}\right\}\,,$$

showing that (η_k) satisfies the sub-Gaussian conditions (9). *In particular, this holds if $|\eta_k| \le R$ holds almost surely.*

A.2 A bound on the prediction error

In this section we prove some bounds on the error of predicting the mean-rewards.

We start with the following result:

Proposition 1. *Take any δ, t such that $0 < \delta < \min(1, d/e)$, $1 + \max(d, 2) \le t \le T$. Let $\tilde{A}_t$ be any* A*-valued random variable. Let*

$$\beta_t^a(\delta) = \frac{2\, k_\mu \kappa R_{\max}}{c_\mu} \|m_a\|_{M_t^{-1}} \sqrt{2\, d \log t}\, \sqrt{\log(d/\delta)}\,, \tag{17}$$

where κ is defined by (10)*. Then, with probability at least $1 - \delta$, it holds that*

$$\left|\mu(m'_{\tilde{A}_t}\theta_*) - \mu(m'_{\tilde{A}_t}\tilde{\theta}_t)\right| \le \beta_t^{\tilde{A}_t}(\delta)\,.$$

Proof. Pick a time t such that $d + 1 \le t \le T$ and an action $a \in \mathsf{A}$. We start with bounding $\left|\mu(m'_a\theta_*) - \mu(m'_a\tilde{\theta}_t)\right|$. Since μ is Lipschitz, we have $|\mu(m'_a\theta_*) - \mu(m'_a\tilde{\theta}_t)| \le k_\mu |m'_a(\theta_* - \tilde{\theta}_t)|$. By Assumption 1, ∇g_t is continuous,[5] hence, by the Fundamental Theorem of Calculus,

$$g_t(\theta_*) - g_t(\tilde{\theta}_t) = G_t(\theta_* - \tilde{\theta}_t)\,,$$

where

$$G_t = \int_0^1 \nabla g_t\left(s\theta_* + (1-s)\tilde{\theta}_t\right) ds\,.$$

Now, for any $\theta \in \Theta$, $\nabla g_t(\theta) = \sum_{k=1}^{t-1} m_{A_k} m'_{A_k} \dot{\mu}(m'_{A_k}\theta)$. Therefore, thanks to Assumption 1, we have $G_t \succeq c_\mu M_t \succeq c_\mu M_d \succ 0$, where in the last step we used that the first d actions are such that $M_d \succeq \lambda_0 I \succ 0$. Thus, G_t is positive definite and, hence, it is also non-singular. Therefore,

$$\left|\mu(m'_a\theta_*) - \mu(m'_a\tilde{\theta}_t)\right| \le k_\mu \left|m'_a G_t^{-1}(g_t(\theta_*) - g_t(\tilde{\theta}_t))\right|\,.$$

Since G_t^{-1} is also positive definite, we get

$$\left|\mu(m'_a\theta_*) - \mu(m'_a\tilde{\theta}_t)\right| \le k_\mu \|m_a\|_{G_t^{-1}} \left\|g_t(\theta_*) - g_t(\tilde{\theta}_t)\right\|_{G_t^{-1}}\,. \tag{18}$$

Since $G_t \succeq c_\mu M_t$ implies that $G_t^{-1} \preceq c_\mu^{-1} M_t^{-1}$, $\|x\|_{G_t^{-1}} \le \frac{1}{\sqrt{c_\mu}} \|x\|_{M_t^{-1}}$ holds for arbitrary $x \in \mathbb{R}^d$. Hence,

$$\left|\mu(m'_a\theta_*) - \mu(m'_a\tilde{\theta}_t)\right| \le \frac{k_\mu}{c_\mu} \|m_a\|_{M_t^{-1}} \left\|g_t(\theta_*) - g_t(\tilde{\theta}_t)\right\|_{M_t^{-1}}\,.$$

Now,

$$\begin{aligned}\left\|g_t(\theta_*) - g_t(\tilde{\theta}_t)\right\|_{M_t^{-1}} &\le \left\|g_t(\theta_*) - g_t(\hat{\theta}_t)\right\|_{M_t^{-1}} + \left\|g_t(\hat{\theta}_t) - g_t(\tilde{\theta}_t)\right\|_{M_t^{-1}} \\ &\le 2\left\|g_t(\theta_*) - g_t(\hat{\theta}_t)\right\|_{M_t^{-1}}\,,\end{aligned}$$

where the first inequality follows from the triangle inequality and second follows since by assumption $\theta_* \in \Theta$ and because of the optimizing property of $\tilde{\theta}_t$ within Θ.

Thanks to the definition of $\hat{\theta}_t$, and using $\epsilon_k = R_k - \mu(m'_{A_k}\theta_*)$, $\xi_t \overset{\text{def}}{=} g_t(\hat{\theta}_t) - g_t(\theta_*) = \sum_{k=1}^{t-1} m_{A_k}\epsilon_k$. Therefore,

$$\left|\mu(m'_a\theta_*) - \mu(m'_a\tilde{\theta}_t)\right| \le \frac{2\, k_\mu}{c_\mu} \|m_a\|_{M_t^{-1}} \|\xi_t\|_{M_t^{-1}}\,.$$

Since this holds simultaneously for all $a \in \mathsf{A}$, it also holds when a is replaced by *any* A-valued random variable $\tilde{A}_t$:

$$\left|\mu(m'_{\tilde{A}_t}\theta_*) - \mu(m'_{\tilde{A}_t}\tilde{\theta}_t)\right| \le \frac{2\, k_\mu}{c_\mu} \left\|m_{\tilde{A}_t}\right\|_{M_t^{-1}} \|\xi_t\|_{M_t^{-1}}\,. \tag{19}$$

[5] For all $x \in \mathbb{R}^d$, $\nabla g_t(x)$ denotes the Jacobian matrix of g_t at point x.

Now, let us use Lemma 1 to bound $\|\xi_t\|_{M_t^{-1}}$. Set $m_k = m_{A_{k+1}}$ ($k = 0, 1, \ldots$), $\eta_k = \epsilon_k$ ($k = 1, 2, \ldots$), $\mathcal{F}_k = \sigma(m_s, \eta_s; s \le k)$. Due to Assumption 3, $\mathbb{E}\left[\eta_k | \mathcal{F}_{k-1}\right] = \mathbb{E}\left[\eta_k | m_{k-1}, \eta_{k-1}, \ldots, m_1, \eta_1, m_0\right] = \mathbb{E}\left[\epsilon_k | m_{A_k}, \epsilon_{k-1}, \ldots, m_{A_2}, \epsilon_1, m_{A_1}\right] = 0$. Since by the same assumption, $|\epsilon_k| \le R_{\max}$, we may choose $R = R_{\max}$ by Remark 1. Further, by Assumption 2, $\|m_k\|_2 = \|m_{A_{k+1}}\|_2 \le \max_{a \in \mathsf{A}} \|m_a\|_2 \le c_m$, and, by the choice of the first d actions, $\sum_{k=1}^d m_{k-1} m_{k-1}' = \sum_{k=1}^d m_{A_k} m_{A_k}' \succeq \lambda_0 I$. Therefore, all the assumptions of the Lemma are met and we can conclude that for any $0 < \delta < \min(1, d/e)$, $t \ge 1 + \max(d, 2)$, with probability at least $1 - \delta$,

$$\|\xi_t\|_{M_t^{-1}} \le \kappa R_{\max} \sqrt{2\, d \log t}\, \sqrt{\log(d/\delta)}, \tag{20}$$

where κ is defined by (10).

By chaining (19) and (20), we get that on the event when (20) holds, we also have

$$\left| \mu(m_{\tilde{A}_t}' \theta_*) - \mu(m_{\tilde{A}_t}' \tilde{\theta}_t) \right| \le \frac{2\, k_\mu \kappa R_{\max}}{c_\mu} \|m_{\tilde{A}_t}\|_{M_t^{-1}} \sqrt{2\, d \log t}\, \sqrt{\log(d/\delta)}\,,$$

finishing the proof. □

Proposition 1 implies the following bound on the immediate mean regret:

Proposition 2. *For all δ such that $0 < \delta \le \min(1, 2Td/e)$, simultaneously for all $t \in \{1 + \max(d, 2), \ldots, T\}$,*

$$\mu(m_{a_*}' \theta_*) - \mu(m_{A_t}' \theta_*) \le 2\, \beta_t^{A_t} \left(\tfrac{\delta}{2T}\right).$$

holds with probability at least $1 - \delta$.

Proof. Fix $t \in \{1 + \max(d, 2), \ldots, T\}$ and let δ be as in the statement. Consider the decomposition

$$\begin{aligned} \mu(m_{a_*}' \theta_*) - \mu(m_{A_t}' \theta_*) &= \left(\mu(m_{a_*}' \theta_*) - \mu(m_{a_*} \tilde{\theta}_t) \right) \\ &\quad + \left(\mu(m_{a_*} \tilde{\theta}_t) - \mu(m_{A_t} \tilde{\theta}_t) \right) + \left(\mu(m_{A_t} \tilde{\theta}_t) - \mu(m_{A_t}' \theta_*) \right). \end{aligned}$$

Now, according to Proposition 1, outside of an event of measure bounded by $\delta/(2T)$,

$$\mu(m_{a_*}' \theta_*) - \mu(m_{a_*}' \tilde{\theta}_t) \le \beta_t^{a_*}(\delta/(2T))\,.$$

Also, outside of an event of measure bounded by $\delta/(2T)$,

$$\mu(m_{A_t}' \theta_*) - \mu(m_{A_t}' \tilde{\theta}_t) \le \beta_t^{A_t}(\delta/(2T))\,.$$

Further, by the definition of A_t,

$$\begin{aligned} \mu(m_{a_*} \tilde{\theta}_t) - \mu(m_{A_t} \tilde{\theta}_t) &= \mu(m_{a_*} \tilde{\theta}_t) + \beta_t^{a_*}(\delta/(2T)) - \mu(m_{A_t} \tilde{\theta}_t) - \beta_t^{a_*}(\delta/(2T)) \\ &\le \mu(m_{A_t} \tilde{\theta}_t) + \beta_t^{A_t}(\delta/(2T)) - \mu(m_{A_t} \tilde{\theta}_t) - \beta_t^{a_*}(\delta/(2T)) \\ &= \beta_t^{A_t}(\delta/(2T)) - \beta_t^{a_*}(\delta/(2T)). \end{aligned}$$

Chaining the inequalities and using a union bound gives the final result. □

According to the previous proposition, the behavior of the immediate regret at time step t is bounded by $2\beta_t^{A_t}(\delta/2T) = 2\rho(t)\|m_{A_t}\|_{M_t^{-1}} \le 2\rho(T)\|m_{A_t}\|_{M_t^{-1}}$. Therefore, with $t_0 = 1 + \max(d, 2)$, outside of an event of probability at most δ, we can bound the cumulated regret up to time T by

$$\text{Regret}_T \le (t_0 - 1) R_{\max} + \sum_{t=t_0}^{T} \min\left\{ \mu(m_{a_*}' \theta_*) - \mu(m_{A_t}' \theta_*), R_{\max} \right\} \tag{21}$$

$$\le (t_0 - 1) R_{\max} + 2\, \rho(T) \sum_{t=t_0}^{T} \min\left\{ \|m_{A_t}\|_{M_t^{-1}}, 1 \right\}, \tag{22}$$

where the last inequality follows from the fact that $R_{\max} \leq 2\rho(T)$ by definition of $\rho(T)$. Note that $\|m_{A_t}\|_{M_t^{-1}}$ is expected to become small as t gets large. This motivates us to bound a sum of $\|m_{A_t}\|^2_{M_t^{-1}}$. For technical reasons that will become clear later, we bound $\sum_{t=d}^T \min\left\{\|m_{A_t}\|^2_{M_t^{-1}}, 1\right\}$.

Proposition 3. *Let $t_0 \geq d+1$. Then,*

$$\sum_{t=t_0}^T \min\left\{\|m_{A_t}\|^2_{M_t^{-1}}, 1\right\} \leq 2\, d\, \log\left(\frac{c_m^2 T}{\lambda_0}\right) \quad \text{a.s.}\,.$$

Proof. This proof follows the steps of the proof of Lemma 9 of [8]. By the definition of M_{t+1}, we have

$$\begin{aligned}
\det\left(M_{t+1}\right) &= \det\left(M_t + m_{A_t} m'_{A_t}\right) = \det\left(M_t\right)\det\left(I + M_t^{-1/2} m_{A_t} (M_t^{-1/2} m_{A_t})'\right) \\
&= \det\left(M_t\right)\left(1 + \|m_{A_t}\|^2_{M_t^{-1}}\right) = \det\left(M_{t_0}\right)\prod_{k=t_0}^{t}\left(1 + \|m_{A_k}\|^2_{M_k^{-1}}\right)\,,
\end{aligned}$$

where the last line follows from the fact that $1 + \|m_{A_t}\|^2_{M_t^{-1}}$ is an eigenvalue of the matrix $I + M_t^{-1/2} m_{A_t} (M_t^{-1/2} m_{A_t})'$ and that all the other eigenvalues are equal to 1. Thus, using the fact that $x \leq 2\log(1+x)$ which holds for any $0 \leq x \leq 1$, we have

$$\begin{aligned}
\sum_{t=t_0}^T \min\left\{\|m_{A_t}\|^2_{M_t^{-1}}, 1\right\} &\leq 2\sum_{t=t_0}^T \log\left(1 + \|m_{A_t}\|^2_{M_t^{-1}}\right) \\
&= 2\log\prod_{t=t_0}^T\left(1 + \|m_{A_t}\|^2_{M_t^{-1}}\right) \\
&= 2\log\left(\frac{\det(M_{T+1})}{\det(M_{t_0})}\right)\,.
\end{aligned}$$

Note that the trace of M_{t+1} is upper-bounded by $t\, c_m^2$. Then, since the trace of the positive definite matrix M_{t+1} is equal to the sum of its eigenvalues and $\det(M_{t+1})$ is the product of its eigenvalues, we have $\det(M_{t+1}) \leq (tc_m^2)^d$. In addition, $\det(M_{t_0}) \geq \lambda_0^d$ since $t_0 \geq d+1$. Thus,

$$\sum_{t=t_0}^T \min\left\{\|m_{A_t}\|^2_{M_t^{-1}}, 1\right\} \leq 2\, d \log\left(\frac{c_m^2 T}{\lambda_0}\right)\,.$$

$\square$

A.3 Proof of the Main Theorems

A.3.1 Proof of Theorem 1

Proof. We start from (21), where $t_0 = 1 + \max(d, 2)$. According to the definition of $\Delta(\theta_*)$ whenever A_t is a suboptimal action, $\mu(m'_{a_*}\theta_*) - \mu(m'_{A_t}\theta_*) \geq \Delta(\theta_*)$, while in the other case we have $\mu(m'_{a_*}\theta_*) - \mu(m'_{A_t}\theta_*) = 0$. In both cases, we can write

$$\mu(m'_{a_*}\theta_*) - \mu(m'_{A_t}\theta_*) \leq \frac{(\mu(m'_{a_*}\theta_*) - \mu(m'_{A_t}\theta_*))^2}{\Delta(\theta_*)}\,.$$

According to Proposition 2, with probability $1-\delta$, simultaneously for all $t \in \{t_0, \ldots, T\}$,

$$\mu(m'_{a_*}\theta_*) - \mu(m'_{A_t}\theta_*) \leq 2\beta_t^{A_t}(\delta/(2T)) = 2\rho(t)\,\|m_{A_t}\|_{M_t^{-1}}\,.$$

Therefore, on the event when these inequalities holds, we have

$$\begin{aligned}
\sum_{t=t_0}^T \min\left\{\mu(m'_{a_*}\theta_*) - \mu(m'_{A_t}\theta_*), R_{\max}\right\} &\leq \sum_{t=t_0}^T \min\left\{4\frac{\rho(t)^2}{\Delta(\theta_*)}\|m_{A_t}\|^2_{M_t^{-1}}, R_{\max}\right\} \\
&\leq 4\frac{\rho(T)^2}{\Delta(\theta_*)}\sum_{t=t_0}^T \min\left\{\|m_{A_t}\|^2_{M_t^{-1}}, 1\right\}\,.
\end{aligned}$$

where the last inequality follows from the fact that $\Delta(\theta_*) \leq R_{\max} \leq 4\rho(T)^2/R_{\max}$ and that $\rho(.)$ is an increasing function. Combining this with the bound of Proposition 3, we get

$$\sum_{t=t_0}^{T} \min\left\{\mu(m'_{a_*}\theta_*) - \mu(m'_{A_t}\theta_*), R_{\max}\right\} \leq 8\, d\, \frac{\rho(T)^2}{\Delta(\theta_*)} \log\left(\frac{c_m^2 T}{\lambda_0}\right) .$$

Plugging in the definition of $\rho(T)$, we get that it holds with probability $1-\delta$ that

$$\begin{aligned}
\text{Regret}_T &\leq (t_0 - 1)R_{\max} + \sum_{t=t_0}^{T} \min\left\{\mu(m'_{a_*}\theta_*) - \mu(m'_{A_t}\theta_*), R_{\max}\right\} \\
&\leq (t_0 - 1)R_{\max} + \frac{32\, d^2\, \kappa^2 R_{\max}^2 k_\mu^2}{c_\mu^2 \Delta(\theta_*)} \log(T) \log(2dT/\delta) \log\left(\frac{c_m^2 T}{\lambda_0}\right) .
\end{aligned}$$

$\square$

A.3.2 Proof of Theorem 2

Proof. Let $t_0 = 1 + \max(d, 2)$. According to Proposition 2, (22) holds with probability $1-\delta$, so it remains to bound

$$\sum_{t=t_0}^{T} \min\left\{\|m_{A_t}\|_{M_t^{-1}}, 1\right\} .$$

Using the Cauchy-Schwarz inequality and Proposition 3, we have

$$\begin{aligned}
\sum_{t=t_0}^{T} \min\left\{\|m_{A_t}\|_{M_t^{-1}}, 1\right\} &\leq \sqrt{T}\sqrt{\sum_{t=t_0}^{T} \min\left\{\|m_{A_t}\|^2_{M_t^{-1}}, 1\right\}} \\
&\leq \sqrt{T}\,\sqrt{2d\log(c_m^2 T/\lambda_0)} .
\end{aligned}$$

Combining with (22) and using the definition of $\rho(\cdot)$ gives

$$\begin{aligned}
\text{Regret}_T &\leq (t_0 - 1)R_{\max} + 2\,\rho(T)\,\sqrt{2\,d\,T\log(c_m^2 T/\lambda_0)} \\
&= (t_0 - 1)R_{\max} + 8\,d\,\frac{k_\mu \kappa R_{\max}}{c_\mu}\,\sqrt{T\,\log(T)\,\log(c_m^2 T/\lambda_0)\log(2Td/\delta)} \\
&\leq (d+1)R_{\max} + 8\,d\,\frac{k_\mu \kappa R_{\max}}{c_\mu}\,\log(s\,T)\,\sqrt{T\,\log(2Td/\delta)},
\end{aligned}$$

where $s = \max\left(\frac{c_m^2}{\lambda_0}, 1\right)$, thus, finishing the proof. $\square$