A. Using the OLS Estimator

Here we construct an example problem to demonstrate how using the standard OLS estimator can fail in the semiparametric setting. While not a comprehensive proof against all asymptotically biased approaches, similar examples can be constructed for related estimators.

Consider a two-dimensional problem with two actions and no stochastic noise, where $\theta = e_2$, the second standard basis vector. On the even rounds, the actions are $z_1 = (1,1), z_2 = (1, 1/3)$ and the confounding term is $f = -1$. On the odd rounds, the actions are $z_1 = z_2 = (1,0)$ and the confounding term is $f = 1$. For any policy for selecting actions, the OLS estimator before round t (for even t) is the solution to the following optimization problem:

$$\text{minimize}_{w \in \mathbb{R}^2}\ \alpha(w_1 + w_2)^2 + (1-\alpha)(w_1 + w_2/3 + 2/3)^2 + (w_1 - 1)^2 = L(w)$$

where $\alpha \in [0,1]$ corresponds to the fraction of the even rounds (up to round t) where the policy chose z_1. We will argue that, for any α, the solution to this problem $\hat{w}$ has $\hat{w}_2 < 0$. Since there is no stochastic noise, there is no need for confidence bounds once the covariance is full rank, which happens after the second round. Together, this implies that any sensible policy based on $\hat{w}$ will prefer z_2 to z_1 on the even rounds, but z_1 yields higher reward by a fixed constant. Thus using OLS in a confidence-based approach leads to linear regret.

We now show that $\hat{w}_2$ is strictly negative. We have

$$\frac{\partial L(w)}{\partial w_1} = 2\alpha(w_1 + w_2) + 2(1-\alpha)(w_1 + w_2/3 + 2/3) + 2(w_1 - 1),$$
$$\frac{\partial L(w)}{\partial w_2} = 2\alpha(w_1 + w_2) + \frac{2}{3}(1-\alpha)(w_1 + w_2/3 + 2/3).$$

Setting both equations equal to zero yields the following system:

$$4w_1 + (2/3 + 4\alpha/3)w_2 = 2/3 + 4\alpha/3, \qquad (2/3 + 4\alpha/3)w_1 + (2/9 + 16\alpha/9)w_2 = 4\alpha/9 - 4/9.$$

The solution to this system is

$$w_1 = \frac{(2\alpha+1)^2}{-4\alpha^2 + 12\alpha + 1}, \qquad w_2 = \frac{4\alpha^2 + 5}{4\alpha^2 - 12\alpha - 1},$$

provided that $4\alpha^2 \neq 12\alpha + 1$, which is not possible with $\alpha \in [0,1]$. In the interval $[0,1]$ we have that $4\alpha^2 - 12\alpha - 1 < 0$, and hence $w_2 < 0$. Thus, the OLS estimator incorrectly predicts that z_2 receives higher reward than z_1 on the even rounds. Since confidence intervals are not needed, the algorithm suffers linear reget.

B. Proof of Proposition 3

We consider two possible values for the true parameter: $\theta_1 = e_1 \in \mathbb{R}^2, \theta_2 = e_2 \in \mathbb{R}^2$. At all rounds, the context $x_t = \{e_1, e_2\}$ contains just two actions, and we further assume that the noise term $\xi_t = 0$ almost surely. Since the action a_t is a deterministic function of the history, it can also be computed by the adaptive adversary at the beginning of the round, and the adversary chooses

$$f_t(x_t) = -\mathbf{1}\{a_t = \underset{a}{\text{argmax}} \langle \theta, z_{t,a}\rangle\}.$$

We show that $r_t(a_t) = 0$ for all rounds t. Assume the parameter is θ_1 so the optimal action is $a_t^\star = e_1$ and the suboptimal action e_2 has $\langle \theta, e_2 \rangle = 0$. If the learner chooses action e_2, then the adversary sets $f_t(x_t) = 0$, so $r_t(a_t) = 0$. On the other hand, if the learner chooses action e_1, then the adversary sets $f_t(x_t) = -1$ so the reward is also zero. Similarly, if $\theta = \theta_2$, the observed reward is always zero. Since the algorithm is deterministic, it behaves identically regardless of whether the parameter is θ_1 or θ_2. In one of these instances the algorithm must choose the suboptimal action at least $T/2$ times, leading to the lower bound.

C. Proof for the Two-Action Case

We first focus on the simpler two action case. Before turning to the main analysis, we prove two supporting lemmas. The first is an algebraic inequality relating matrix determinants to traces. This inequality also appears in Abbasi-Yadkori et al. (2011).

Lemma 9. *Let $X_1, \ldots, X_n$ denote vectors in $\mathbb{R}^d$ with $\|X_i\|_2 \le L$ for all $i \in [n]$. Define $\Gamma \triangleq \lambda I + \sum_{i=1}^n X_i X_i^\top$. Then*

$$\det(\Gamma) \le (\lambda + nL^2/d)^d.$$

Proof. We will apply the following standard argument:

$$\det(\Gamma)^{1/d} \le \frac{1}{d}\operatorname{tr}(\Gamma) = \frac{1}{d}\operatorname{tr}(\lambda I) + \frac{1}{d}\sum_{i=1}^n \operatorname{tr}(X_i X_i^\top) = \lambda + \frac{1}{d}\sum_{i=1}^n \|X_i\|_2^2 \le \lambda + nL^2/d.$$

The first step is a spectral version of the AM-GM inequality and the remaining steps use linearity of the trace operator and the boundedness conditions. □

The second lemma is a new self-normalized concentration inequality for vector valued martingales.

Lemma 10 (Symmetric self-normalized inequality)**.** *Let $\{\mathcal{F}_t\}_{t=1}^T$ be a filtration and let $\{(Z_t, \zeta_t)\}_{t=1}^T$ be a stochastic process with $Z_t \in \mathbb{R}^d$ and $\zeta_t \in \mathbb{R}$ such that (1) (Z_t, ζ_t) is $\mathcal{F}_t$ measurable, (2) $|\zeta_t| \le M$ for all $t \in [T]$, (3) $Z_t \perp\!\!\!\perp \zeta_t | \mathcal{F}_t$, (4) $\mathbb{E}[Z_t|\mathcal{F}_t] = 0$, and (5) for all $x \in \mathbb{R}^d$, $\mathcal{L}(\langle x, Z_t\rangle \mid \mathcal{F}_t) = \mathcal{L}(-\langle x, Z_t\rangle \mid \mathcal{F}_t)$ where $\mathcal{L}$ denotes the probability law, so that Z_t is conditionally symmetric. Let $\Sigma \triangleq \sum_{t=1}^T Z_t Z_t^\top$. Then for any positive definite matrix Q we have*

$$\mathbb{P}\left[\left\|\sum_{t=1}^T Z_t\zeta_t\right\|^2_{(Q+M^2\Sigma)^{-1}} \ge 2\log\left(\frac{1}{\delta}\sqrt{\frac{\det(Q+M^2\Sigma)}{\det(Q)}}\right)\right] \le \delta.$$

Proof. The proof follows the recipe in de la Peña et al. (2009) (See also de la Peña et al. (2008) for a more comprehensive treatment including the univariate case). We start by applying the Chernoff method. Let $\bar\Sigma \triangleq Q + M^2\Sigma$. We can write

$$\begin{aligned}\mathbb{P}\left[\left\|\sum_{t=1}^T Z_t\zeta_t\right\|^2_{\bar\Sigma^{-1}} \ge 2\log\left(\frac{1}{\delta}\sqrt{\frac{\det(\bar\Sigma)}{\det(Q)}}\right)\right] &= \mathbb{P}\left[\exp\left(\frac{1}{2}\left\|\sum_{t=1}^T Z_t\zeta_t\right\|^2_{\bar\Sigma^{-1}}\right) \ge \frac{1}{\delta}\sqrt{\frac{\det(\bar\Sigma)}{\det(Q)}}\right]\\ &\le \delta\mathbb{E}\left[\sqrt{\frac{\det(Q)}{\det(\bar\Sigma)}}\exp\left(\frac{1}{2}\left\|\sum_{t=1}^T Z_t\zeta_t\right\|^2_{\bar\Sigma^{-1}}\right)\right].\end{aligned}$$

Therefore, if we prove that this latter expectation is at most one, we will arrive at the conclusion. A similar statement appears in Theorem 1 of de la Peña et al. (2009), but our process is slightly different due to the presence of ζ_t. To bound this latter expectation, fix some $\lambda \in \mathbb{R}^d$ and consider an exponentiated process with the increments

$$D_t^\lambda \triangleq \exp\left(\langle\lambda, Z_t\zeta_t\rangle - \frac{M^2\langle\lambda, Z_t\rangle^2}{2}\right).$$

Observe that $\mathbb{E}[D_t^\lambda|\mathcal{F}_t] \le 1$ since by the conditional symmetry of Z_t, we have

$$\begin{aligned}\mathbb{E}[D_t^\lambda|\mathcal{F}_t] &= \mathbb{E}\left[\mathbb{E}\left[D_t^\lambda \mid \mathcal{F}_t, \zeta_t\right] \mid \mathcal{F}_t\right]\\ &= \mathbb{E}\left[\mathbb{E}\left[\exp\left(\frac{-M^2\langle\lambda, Z_t\rangle^2}{2}\right) \times \frac{1}{2}\left(\exp(\langle\lambda, Z_t\zeta_t\rangle) + \exp(-\langle\lambda, Z_t\zeta_t\rangle\right) \mid \mathcal{F}_t, \zeta_t\right] \mid \mathcal{F}_t\right]\\ &= \mathbb{E}\left[\mathbb{E}\left[\exp\left(\frac{-M^2\langle\lambda, Z_t\rangle^2}{2}\right) \times \cosh(\langle\lambda, Z_t\zeta_t) \mid \mathcal{F}_t, \zeta_t\right] \mid \mathcal{F}_t\right]\\ &\le \mathbb{E}\left[\mathbb{E}\left[\exp\left(\frac{-M^2\langle\lambda, Z_t\rangle^2}{2} + \frac{\langle\lambda, Z_t\zeta_t\rangle^2}{2}\right) \mid \mathcal{F}_t, \zeta_t\right] \mid \mathcal{F}_t\right] \le 1.\end{aligned}$$

This argument first uses the conditional symmetry of Z_t and the conditional independence of Z_t, ζ_t, then the identity $(e^x + e^{-x})/2 = \cosh(x)$ and finally the analytical inequality $\cosh(x) \le e^{x^2/2}$. Finally in the last step we use the bound $|\zeta_t| \le M$. This implies that the martingale $U_t^\lambda \triangleq \prod_{\tau=1}^t D_\tau^\lambda$ is a super-martingale with $\mathbb{E}[U_t^\lambda] \le 1$ for all t, since by induction

$$\mathbb{E}[U_t^\lambda] = \mathbb{E}[U_{t-1}^\lambda \mathbb{E}[D_t^\lambda|\mathcal{F}_t]] \le \mathbb{E}[U_{t-1}^\lambda] \le \ldots \le 1. \qquad (6)$$

Now we apply the method of mixtures. In a standard application of the Chernoff method, we would choose λ to maximize $\mathbb{E}[U_T^\lambda]$, but since we still have an expectation, we cannot swap expectation and maximum. Instead, we integrate the inequality $\mathbb{E}[U_T^\lambda] \le 1$, which holds for any λ, against λ drawn from a Gaussian distribution with covariance Q^{-1}. By Fubini's theorem, we can swap the expectations to obtain

$$
\begin{aligned}
1 \ge \mathbb{E}_{\lambda\sim\mathcal{N}(0,Q^{-1})}\mathbb{E}[U_T^\lambda] &= \mathbb{E}\int U_T^\lambda (2\pi)^{-d/2}\sqrt{\det(Q)}\exp(-\lambda^\top Q\lambda/2)d\lambda \\
&= \mathbb{E}\int (2\pi)^{-d/2}\sqrt{\det(Q)}\exp\left(\sum_{t=1}^T\langle\lambda, Z_t\zeta_t\rangle - \frac{M^2\lambda^\top(\sum_{t=1}^T Z_tZ_t^\top)\lambda + \lambda^\top Q\lambda}{2}\right)d\lambda \\
&= \mathbb{E}\int (2\pi)^{-d/2}\sqrt{\det(Q)}\exp\left(\langle\lambda, S\rangle - \frac{M^2\lambda^\top\Sigma\lambda + \lambda^\top Q\lambda}{2}\right)d\lambda,
\end{aligned}
$$

where $S \triangleq \sum_{t=1}^T Z_t\zeta_t$ and recall that $\Sigma \triangleq \sum_{t=1}^T Z_tZ_t^\top$. By completing the square, the term in the exponent can be rewritten as

$$
\langle\lambda, S\rangle - \frac{M^2\lambda^\top\Sigma\lambda + \lambda^\top Q\lambda}{2} = \frac{1}{2}\left(-(\lambda - \bar{\Sigma}^{-1}S)^\top\bar{\Sigma}(\lambda - \bar{\Sigma}^{-1}S) + S^\top\bar{\Sigma}^{-1}S\right),
$$

where recall that $\bar{\Sigma} \triangleq M^2\Sigma + Q$. As such we obtain

$$
\begin{aligned}
1 &\ge \mathbb{E}\left[\exp\left(S^\top\bar{\Sigma}^{-1}S/2\right) \times \int (2\pi)^{-d/2}\sqrt{\det(Q)}\exp\left(\frac{-(\lambda - \bar{\Sigma}^{-1}S)^\top\bar{\Sigma}(\lambda - \bar{\Sigma}^{-1}S)}{2}\right)\right]d\lambda \\
&= \mathbb{E}\sqrt{\frac{\det(Q)}{\det(\bar{\Sigma})}}\exp\left(S^\top\bar{\Sigma}^{-1}S\right).
\end{aligned}
$$

This proves the lemma. □

Equipped with the two lemmas, we can now turn to the analysis of the influence-adjusted estimator.

Lemma 11 (Restatement of Lemma 5)**.** *Under Assumption 1 and Assumption 2, with probability at least $1-\delta$, the following holds simultaneously for all $t \in [T]$:*

$$
\|\hat{\theta}_t - \theta\|_{\Gamma_t} \le \sqrt{\lambda} + \sqrt{9d\log(1 + T/(d\lambda)) + 18\log(T/\delta)}.
$$

Proof. Recall that we define $\hat{\theta}_t, \Gamma_t$ to be the estimator and matrix used in round t, based on $t-1$ examples. Fixing a round t, we start by expanding the definition of $\hat{\theta}_t$. We use the shorthand $z_\tau \triangleq z_{\tau,a_\tau}$, $\mu_\tau \triangleq \mathbb{E}_{b\sim\pi_\tau}[z_{\tau,b}]$, and $r_\tau \triangleq r_\tau(a_\tau)$.

$$
\begin{aligned}
\hat{\theta}_t &= \Gamma_t^{-1}\sum_{\tau=1}^{t-1}(z_\tau - \mu_\tau)r_\tau = \Gamma_t^{-1}\sum_{\tau=1}^{t-1}(z_\tau - \mu_\tau)(\langle\theta, z_\tau\rangle + f_\tau(x_\tau) + \xi_\tau) \\
&= \Gamma_t^{-1}\sum_{\tau=1}^{t-1}(z_\tau - \mu_\tau)(\langle\theta, z_\tau - \mu_\tau\rangle + \langle\theta, \mu_\tau\rangle + f_\tau(x_\tau) + \xi_\tau) \\
&= (\Gamma_t)^{-1}(\Gamma_t - \lambda I)\theta + \Gamma_t^{-1}\sum_{\tau=1}^{t-1}(z_\tau - \mu_\tau)(\langle\theta, \mu_\tau\rangle + f_\tau(x_\tau) + \xi_\tau).
\end{aligned}
$$

Let $Z_\tau \triangleq z_\tau - \mu_\tau$ and $\zeta_\tau \triangleq \langle\theta, \mu_\tau\rangle + f_\tau(x_\tau) + \xi_\tau$. With this expansion, we can write

$$
\|\hat{\theta}_t - \theta\|_{\Gamma_t} = \| - \lambda\Gamma_t^{-1}\theta + \Gamma_t^{-1}\sum_{\tau=1}^{t-1}Z_\tau\zeta_\tau\|_{\Gamma_t} \le \|\lambda\theta\|_{\Gamma_t^{-1}} + \left\|\sum_{\tau=1}^{t-1}Z_\tau\zeta_\tau\right\|_{\Gamma_t^{-1}} \le \sqrt{\lambda} + \left\|\sum_{\tau=1}^{t-1}Z_\tau\zeta_\tau\right\|_{\Gamma_t^{-1}}.
$$

To finish the proof, we apply Lemma 10 to this last term. To verify the preconditions of the lemma, let $\mathcal{F}_\tau \triangleq \sigma(x_1, \ldots, x_\tau, a_1, \ldots, a_{\tau-1}, \xi_1, \ldots, \xi_{\tau-1})$ denote the σ-algebra corresponding to the τ^{th} round, after observing the context x_τ. Then the policy π_τ and hence the action a_τ are $\mathcal{F}_\tau$ measurable and so is the noise term ξ_τ. Therefore,

$Z_\tau = z_{\tau,a_\tau} - \mathbb{E}_{a\sim\pi_\tau}[z_{\tau,a}]$ is measurable, which verifies the first precondition. Using the boundedness properties in Assumption 2, we know that $|\zeta_\tau| \le 3 \triangleq M$, and by construction of the random variables, we have $Z_\tau \perp\!\!\!\perp \zeta_\tau | \mathcal{F}_\tau$ and $\mathbb{E}[Z_\tau|\mathcal{F}_\tau] = 0$. Finally, for the symmetry property, either $Z_\tau|\mathcal{F}_\tau \equiv 0$ if one action is eliminated, or otherwise we have $\mu_\tau = \frac{1}{2}(z_{\tau,1} + z_{\tau,2})$ since there are only two actions. In this case the random variable $Z_\tau|\mathcal{F}_\tau = \epsilon_\tau(z_{\tau,1} - z_{\tau_2})/2$ where ϵ_τ is a Rademacher random variable. By inspection this is clearly conditionally symmetric. As such, we may apply Lemma 10, which reveals that with probability at least $1-\delta$,

$$\left\|\sum_{\tau=1}^{t-1} Z_\tau\zeta_\tau\right\|^2_{\Gamma_t^{-1}} = M^2 \left\|\sum_{\tau=1}^{t-1} Z_\tau\zeta_\tau\right\|^2_{(M^2\Gamma_t)^{-1}} \le 2M^2 \log\left(\frac{1}{\delta}\sqrt{\frac{\det(M^2\Gamma_t)}{\det(M^2\lambda I)}}\right)$$
$$= 18\log\left(\sqrt{\lambda^{-d}\det(\Gamma_t)}/\delta\right).$$

The inequality here is Lemma 10 with $Q = M^2\lambda I$, and for the last equality we use that $\det(cQ) = c^d\det(Q)$ for a $d\times d$ positive semidefinite matrix Q. As two final steps, we apply Lemma 9 and take a union bound over all rounds T. Combining these, we get that for all T,

$$\|\hat{\theta}_t - \theta\|_{\Gamma_t} \le \sqrt{\lambda} + \left\|\sum_{\tau=1}^{t-1} Z_\tau\zeta_\tau\right\|_{\Gamma_t^{-1}} \le \sqrt{\lambda} + \sqrt{18\left(\log(\sqrt{\lambda^{-d}\det(\Gamma_t)}) + \log(T/\delta)\right)}$$
$$\le \sqrt{\lambda} + \sqrt{9d\log(1+T/(d\lambda)) + 18\log(T/\delta)}. \qquad \square$$

Therefore, with $\gamma(T) \triangleq \sqrt{\lambda} + \sqrt{9d\log(1+T/(d\lambda)) + 18\log(T/\delta)}$ we can apply Lemma 6 to bound the regret by

$$\text{Reg}(T) \le \sqrt{2T\log(1/\delta)} + 2\gamma(T)\sum_{t=1}^{T}\sqrt{\text{tr}(\Gamma_t^{-1}\operatorname*{Cov}_{b\sim\pi_t}(z_{t,b}))}.$$

Via a union bound, this inequality holds with probability at least $1-2\delta$. To finish the proof we need to analyze this latter term. This is the contents of the following lemma. A related statement, with a similar proof, appears in Abbasi-Yadkori et al. (2011).

Lemma 12. *Let $X_1,\ldots,X_T$ be a sequence of vectors in $\mathbb{R}^d$ with $\|X_t\|_2 \le 1$ and define $\Gamma_1 \triangleq \lambda I$, $\Gamma_t \triangleq \Gamma_{t-1} + X_{t-1}X_{t-1}^\top$. Then*

$$\sum_{t=1}^{T}\sqrt{\text{tr}(\Gamma_t^{-1}X_tX_t^\top)} \le \sqrt{Td(1+1/\lambda)\log(1+T/(d\lambda))}.$$

Proof. First, apply the Cauchy-Schwarz inequality to the left hand side to obtain

$$\sum_{t=1}^{T}\sqrt{\text{tr}(\Gamma_t^{-1}X_tX_t^\top)} \le \sqrt{T}\sqrt{\sum_{T=1}^{T}\text{tr}(\Gamma_t^{-1}X_tX_t^\top)}.$$

For the remainder of the proof we work only with the second term. Let us start by analyzing a slightly different quantity, $\text{tr}(\Gamma_{t+1}^{-1}X_tX_t^\top)$. By concavity of $\log\det(M)$, we have

$$\log\det(\Gamma_t) \le \log\det(\Gamma_{t+1}) + \text{tr}(\Gamma_{t+1}^{-1}(\Gamma_t - \Gamma_{t+1})),$$

which implies

$$\text{tr}(\Gamma_{t+1}^{-1}X_tX_t^\top) = \text{tr}(\Gamma_{t+1}^{-1}(\Gamma_{t+1} - \Gamma_t)) \le \log\det(\Gamma_{t+1}) - \log\det(\Gamma_t)$$

As such, we obtain a telescoping sum

$$\sum_{t=1}^{T}\text{tr}(\Gamma_{t+1}^{-1}X_tX_t^\top) \le \log\det(\Gamma_{T+1}) - \log\det(\Gamma_1) \le d\log(\lambda + T/d) - d\log\lambda = d\log(1+T/(d\lambda))$$

The first inequality here uses the concavity argument and the second uses Lemma 9. To finish the proof, we must translate back to Γ_t^{-1}. For this, we use the Sherman-Morrison-Woodbury identity, which reveals that

$$X_t^\top \Gamma_{t+1}^{-1} X_t = X_t^\top (\Gamma_t + X_t X_t^\top)^{-1} X_t = X_t^\top \left(\Gamma_t^{-1} - \frac{\Gamma_t^{-1} X_t X_t^\top \Gamma_t^{-1}}{1 + \|X_t\|^2_{\Gamma_t^{-1}}} \right) X_t$$
$$= \frac{\|X_t\|^2_{\Gamma_t^{-1}}}{1 + \|X_t\|^2_{\Gamma_t^{-1}}} \geq (1 + 1/\lambda)^{-1} \|X_t\|^2_{\Gamma_t^{-1}}.$$

Here in the last step we use that $\|X_t\|^2_{\Gamma_t^{-1}} \leq \|X_t\|^2_{(\lambda I)^{-1}} \leq 1/\lambda$. Overall, we obtain

$$\sum_{t=1}^{T} \operatorname{tr}(\Gamma_t^{-1} X_t X_t^\top) \leq (1 + 1/\lambda) d \log(1 + T/(d\lambda)),$$

and combined with the first application of Cauchy-Schwarz, this proves the lemma. □

Combining the lemmas, we have that with probability at least $1 - 2\delta$, the regret is at most

$$\begin{aligned}
\operatorname{Reg}(T) &\leq \sqrt{2T \log(1/\delta)} + 2\gamma(T) \sqrt{Td(1 + 1/\lambda) \log(1 + T/(d\lambda))} \\
&= \sqrt{2T \log(1/\delta)} + 2\sqrt{Td(1 + 1/\lambda) \log(1 + T/(d\lambda))} \left(\sqrt{\lambda} + \sqrt{9d \log(1 + T/(d\lambda)) + 18 \log(T/\delta)} \right).
\end{aligned}$$

With $\lambda = 1$, this bound is $O\left(\sqrt{Td \log(T/\delta) \log(T/d)} + d\sqrt{T} \log(T/d) \right)$.

D. Proof for the General Case

We now turn to the more general case. We need several additional lemmas.

Lemma 13 (Restatement of Lemma 8)**.** *Problem (3) is convex and always has a feasible solution. Specifically, for any vectors $z_1, \ldots, z_n \in \mathbb{R}^d$ and any positive definite matrix M, there exists a distribution $w \in \Delta([n])$ with mean $\mu_w = \mathbb{E}_{b \sim w}[z_b]$ such that*

$$\forall i \in [n], \quad \|z_i - \mu_w\|^2_M \leq \operatorname{tr}(M \operatorname*{Cov}_{b \sim w}(z_b)).$$

Proof. We analyze the minimax program

$$\min_{w \in \Delta([n])} \max_{i \in [n]} \|z_i - \mu_w\|^2_M - \operatorname{tr}(M \operatorname*{Cov}_{w}(z)).$$

The goal is to show that the value of this program is non-negative, which will prove the result. Expanding the definitions, we have

$$\begin{aligned}
&\min_{w \in \Delta([n])} \max_{i \in [n]} \|z_i - \mu_w\|^2_M - \operatorname{tr}(M \operatorname*{Cov}_{w}(z)) \\
&= \min_{w \in \Delta([n])} \max_{v \in \Delta([n])} \sum_i v_i \left(\|z_i - \mu_w\|^2_M + \mu_w^\top M \mu_w - \sum_j w_j z_j^\top M z_j \right) \\
&= \min_{v \in \Delta([n])} \max_{w \in \Delta([n])} \sum_i v_i \left(\|z_i - \mu_w\|^2_M + \mu_w^\top M \mu_w - \sum_j w_j z_j^\top M z_j \right).
\end{aligned}$$

The last equivalence here is Sion's Minimax Theorem (Sion, 1958), which is justified since both domains are compact convex subsets of $\mathbb{R}^n$ and since the objective is linear in the maximizing variable v, and convex in the minimizing variable w. This convexity is clear since μ_w is a linear in w, and hence the first two terms are convex quadratics (since M is positive definite), while the third term is linear in w. Thus Sion's theorem lets us swap the order of the minimization and maximization.

Now we upper bound the solution by setting $w = v$. This gives

$$\leq \max_{v \in \Delta([n])} \sum_i v_i \left(\|z_i - \mu_v\|_M^2 + \mu_v^\top M \mu_v - \sum_j v_j z_j^\top M z_j \right)$$
$$= \max_{v \in \Delta([n])} \sum_i v_i \left((z_i - \mu_v)^\top M (z_i - \mu_v) + \mu_v^\top M \mu_v - \sum_j v_j z_j^\top M z_j \right) = 0. \qquad \square$$

To prove the analog of Lemma 10, we need several additional tools. First, we use Freedman's inequality to derive a positive-semidefinite inequality relating the sample covariance matrix to the population matrix.

Lemma 14. *Let $X_1, \ldots, X_n$ be conditionally centered random vectors in $\mathbb{R}^d$ adapted to a filtration $\{\mathcal{F}_t\}_{t=1}^n$ with $\|X_i\|_2 \leq 1$ almost surely. Define $\hat{\Sigma} \triangleq \sum_{i=1}^n X_i X_i^\top$ and $\Sigma \triangleq \sum_{i=1}^n \mathbb{E}[X_i X_i^\top \mid \mathcal{F}_i]$. Then, with probability at least $1-\delta$, the following holds simultaneously for all unit vectors $v \in \mathbb{R}^d$:*

$$v^\top \Sigma v \leq 2 v^\top \hat{\Sigma} v + 9d \log(9n) + 8 \log(2/\delta).$$

This lemma is related to the Matrix Bernstein inequality, which can be used to control $\|\Sigma - \hat{\Sigma}\|_2$, a quantity that is quite similar to what we are controlling here. The Matrix Bernstein inequality can be used to derive a high probability bound of the form

$$\forall v \in \mathbb{R}^d, \|v\|_2 = 1, \quad v^\top (\Sigma - \hat{\Sigma}) v \leq \frac{1}{2} \|\Sigma\|_2 + c \log(dn/\delta),$$

for a constant $c > 0$. On one hand, this bound is stronger than ours since the deviation term depends only logarithmically on the dimension. However, the variance term involves the spectral norm rather than a quantity that depends on v as in our bound. Thus, Matrix Bernstein is worse when Σ is highly ill-conditioned, and since we have essentially no guarantees on the spectrum of Σ, our specialized inequality, which is more adaptive to the specific direction v, is crucial. Moreover, the worse dependence on d is inconsequential, since the error will only appear in a lower order term.

Proof. First consider a single unit vector $v \in \mathbb{R}^d$, we will apply a covering argument at the end of the proof. By assumption, the sequence of sums $\{\sum_{i=1}^\tau v^\top (X_i X_i^\top - \mathbb{E}[X_i X_i^\top \mid \mathcal{F}_i]) v\}_{\tau=1}^n$ is a martingale, so we may apply Freedman's inequality (Freedman, 1975; Beygelzimer et al., 2011), which states that with probability at least $1-\delta$

$$|v^\top (\hat{\Sigma} - \Sigma) v| \leq 2 \sqrt{\sum_{i=1}^n \mathrm{Var}(v^\top (X_i X_i^\top - \mathbb{E}[X_i X_i^\top \mid \mathcal{F}_i]) v \mid \mathcal{F}_i) \log(2/\delta)} + 2 \log(2/\delta).$$

Let us now upper bound the variance term: for each i,

$$\begin{aligned} \mathrm{Var}(v^\top (X_i X_i^\top - \mathbb{E}[X_i X_i^\top \mid \mathcal{F}_i]) v \mid \mathcal{F}_i) &\leq \mathbb{E}[(v^\top (X_i X_i^\top - \mathbb{E}[X_i X_i^\top \mid \mathcal{F}_i] \mid \mathcal{F}_i) v)^2 \mid \mathcal{F}_i] \\ &\leq \mathbb{E}[(v^\top X_i)^4 \mid \mathcal{F}_i] \leq v^\top \mathbb{E}[X_i X_i^\top \mid \mathcal{F}_i] v, \end{aligned}$$

where the last inequality follows from the fact that $\|X_i\|_2 \leq 1$ and $\|v\|_2 \leq 1$. Therefore, the cumulative conditional variance is at most $v^\top \Sigma v$. Plugging this into Freedman's inequality gives us

$$|v^\top (\hat{\Sigma} - \Sigma) v| \leq 2 \sqrt{v^\top \Sigma v \log(2/\delta)} + 2 \log(2/\delta).$$

Now, using the fact that $2\sqrt{ab} \leq \alpha a + b/\alpha$ for any $\alpha > 0$, with the choice $\alpha = 1/2$, we get

$$|v^\top (\hat{\Sigma} - \Sigma) v| \leq v^\top \Sigma v / 2 + 4 \log(2/\delta).$$

Re-arranging, this implies

$$v^\top \Sigma v \leq 2 v^\top \hat{\Sigma} v + 8 \log(2/\delta), \tag{7}$$

which is what we would like to prove, but we need it to hold simultaneously for all unit vectors v.

To do so, we now apply a covering argument. Let N be an ϵ-covering of the unit sphere in the projection pseudo-metric $d(u,v) = \|uu^\top - vv^\top\|_2$, with covering number $\mathcal{N}(\epsilon)$. Then via a union bound, a version of (7) holds simultaneously for all $v \in N$, where we rescale $\delta \to \delta/\mathcal{N}(\epsilon)$.

Consider another unit vector u and let v be the covering element. We have

$$\begin{aligned} u^\top \Sigma u &= \mathrm{tr}(\Sigma(uu^\top - vv^\top)) + v^\top \Sigma v \le \mathrm{tr}(\Sigma(uu^\top - vv^\top)) + 2v^\top \hat{\Sigma} v + 8\log(2\mathcal{N}(\epsilon)/\delta) \\ &= \mathrm{tr}((\Sigma - 2\hat{\Sigma})(uu^\top - vv^\top)) + 2u^\top \hat{\Sigma} u + 8\log(2\mathcal{N}(\epsilon)/\delta) \\ &\le \|\Sigma - 2\hat{\Sigma}\|_\star \epsilon + 2u^\top \hat{\Sigma} u + 8\log(2\mathcal{N}(\epsilon)/\delta). \end{aligned}$$

Here $\|\cdot\|_\star$ denotes the nuclear norm, which is dual to the spectral norm $\|\cdot\|_2$. Since all vectors are bounded by 1, we obtain

$$\|\Sigma - 2\hat{\Sigma}\|_\star \le d\lambda_{\max}(\Sigma - 2\hat{\Sigma}) \le 3dn.$$

Overall, the following bound holds simultaneously for all unit vectors $v \in \mathbb{R}^d$, except with probability at most δ:

$$v^\top \Sigma v \le 3dn\epsilon + 2v^\top \hat{\Sigma} v + 8\log(2\mathcal{N}(\epsilon)/\delta).$$

The last step of the proof is to bound the covering number $\mathcal{N}(\epsilon)$. For this, we argue that a covering of the unit sphere in the Euclidean norm suffices, and by standard volumetric arguments, this set has covering number at most $(3/\epsilon)^d$. To see why this suffices, let u be a unit vector and let v be the covering element in the Euclidean norm, which implies that $\|u - v\|_2 \le \epsilon$. Further assume that $\langle u, v\rangle > 0$, which imposes no restriction since the projection pseudo-metric is invariant to multiplying by -1. By definition we also have $\langle u, v\rangle \le 1$. Note that the projection norm is equivalent to the sine of the principal angle between the two subspaces, which once we restrict to vectors with non-negative inner product means that $\|uu^\top - vv^\top\|_2 = \sin\angle(u,v)$. Now

$$\begin{aligned} \sin\angle(u,v) = \sqrt{1 - \langle u,v\rangle^2} &= \sqrt{(1 + \langle u,v\rangle)(1 - \langle u,v\rangle)} \\ &\le \sqrt{2(1 - \langle u,v\rangle)} = \sqrt{\|u\|_2^2 + \|v\|_2^2 - 2\langle u,v\rangle} = \|u - v\|_2 \le \epsilon. \end{aligned}$$

Using the standard covering number bound, we now have

$$v^\top \Sigma v \le 3dn\epsilon + 2v^\top \hat{\Sigma} v + 8d\log(3/\epsilon) + 8\log(2/\delta).$$

Setting $\epsilon = 1/(3n)$ gives

$$v^\top \Sigma v \le d + 2v^\top \hat{\Sigma} v + 8d\log(9n) + 8\log(2/\delta) \le 2v^\top \hat{\Sigma} v + 9d\log(9n) + 8\log(2/\delta). \qquad \square$$

With the positive semidefinite inequality, we can work towards a self-normalized martingale concentration bound. The following is a restatement of Lemma 7 from de la Peña et al. (2009).

Lemma 15 (Lemma 7 of de la Peña et al. (2009))**.** *Let $\{X_i\}_{i=1}^n$ be a sequence of conditionally centered vector-valued random variables adapted to the filtration $\{\mathcal{F}_i\}_{i=1}^n$ and such that $\|X_i\|_2 \le B$ for some constant B. Then*

$$U_n(\lambda) = \exp\left(\lambda^\top \sum_{i=1}^n X_i - \lambda^\top \left(\sum_{i=1}^n X_i X_i^\top + \mathbb{E}[X_i X_i^\top | \mathcal{F}_i]\right)\lambda/2\right)$$

is a supermartingale with $\mathbb{E}[U_n(\lambda)] \le 1$ for all $\lambda \in \mathbb{R}^d$.

The lemma is related to (6), but does not require that conditional probability law for X_i is symmetric, which we used previously. To remove the symmetry requirement, it is crucial that the quadratic self-normalization has both empirical and population terms. With this lemma, the same argument as in the proof of Lemma 10, yields a self-normalized tail bound.

Lemma 16. *Let $\{\mathcal{F}_t\}_{t=1}^T$ be a filtration and let $\{(Z_t, \zeta_t)\}_{t=1}^T$ be a stochastic process with $Z_t \in \mathbb{R}^d$ and $\zeta_t \in \mathbb{R}$ such that (1) (Z_t, ζ_t) is $\mathcal{F}_t$ measurable, (2) $|\zeta_t| \le M$ for all $t \in [T]$, (3) $Z_t \perp\!\!\!\perp \zeta_t | \mathcal{F}_t$, and (4) $\mathbb{E}[Z_t | \mathcal{F}_t] = 0$. Let $\hat{\Sigma} \triangleq \sum_{t=1}^T Z_t Z_t^\top$ and $\Sigma \triangleq \sum_{t=1}^T \mathbb{E}[Z_t Z_T^\top | \mathcal{F}_t]$. Then for any positive definite matrix Q we have*

$$\mathbb{P}\left[\left\|\sum_{t=1}^T Z_t \zeta_t\right\|^2_{(Q + M^2(\hat{\Sigma}+\Sigma))^{-1}} \ge 2\log\left(\frac{1}{\delta}\sqrt{\frac{\det(Q + M^2(\hat{\Sigma}+\Sigma))}{\det(Q)}}\right)\right] \le \delta.$$

Proof. The proof is identical to Lemma 10, but uses Lemma 15 in lieu of (6). □

We can now analyze the influence-adjusted estimator.

Lemma 17. *Under Assumption 1 and Assumption 2 and assuming that $\lambda \geq 4d\log(9T) + 8\log(4T/\delta)$, with probability at least $1-\delta$, the following holds simultaneously for all $t \in [T]$:*

$$\|\hat{\theta}_t - \theta\|_{\Gamma_t} \leq \sqrt{\lambda} + \sqrt{27d\log(1+2T/d) + 54\log(4T/\delta)}.$$

Proof. Using the same argument as in the proof of Lemma 5, we get

$$\|\hat{\theta}_t - \theta\|_{\Gamma_t} \leq \sqrt{\lambda} + \left\| \sum_{\tau=1}^{t-1} Z_\tau \zeta_\tau \right\|_{\Gamma_t^{-1}},$$

where $Z_\tau \triangleq z_\tau - \mu_\tau$ and $\zeta_\tau \triangleq \langle \theta, \mu_\tau \rangle + f_\tau(x_\tau) + \xi_\tau$, just as before. Now we must control this error term, for which we need both Lemma 14 and Lemma 16. Apply Lemma 14 to the vectors Z_τ, setting $\hat{\Sigma}_t \triangleq \sum_{\tau=1}^{t-1} Z_\tau Z_\tau^\top$ and $\Sigma_t \triangleq \sum_{\tau=1}^{t-1} \mathbb{E}[Z_\tau Z_\tau \mid \mathcal{F}_\tau]$. With probability at least $1-\delta/(2T)$, we have that for all unit vectors $v \in \mathbb{R}^d$

$$v^\top \Sigma_t v \leq 2v^\top \hat{\Sigma}_t v + 9d\log(9t) + 8\log(4T/\delta) \leq 2v^\top \hat{\Sigma}_t v + 9d\log(9T) + 8\log(4T/\delta).$$

This implies a lower bound on all quadratic forms involving $\hat{\Sigma}_t$, which leads to positive semidefinite inequality

$$\lambda I + \hat{\Sigma}_t \succeq (\lambda - 3d\log(9T) - 8/3\log(4T/\delta))I + (\hat{\Sigma}_t + \Sigma_t)/3.$$

This means that for any vector v, we have

$$\begin{aligned}\|v\|^2_{(\lambda I + \hat{\Sigma}_t)^{-1}} &\leq \|v\|^2_{((\lambda - 3d\log(9T) - 8/3\log(4T/\delta))I + (\hat{\Sigma}_t + \Sigma_t)/3)^{-1}} \\ &\leq 3\|v\|^2_{((3\lambda - 9d\log(9T) - 8\log(4T/\delta))I + \hat{\Sigma}_t + \Sigma_t)^{-1}}.\end{aligned}$$

Before we apply Lemma 16, we must introduce the range parameter M. Fix a round t and let $A \triangleq ((3\lambda - 9d\log(9T) - 8\log(4T/\delta))I + \hat{\Sigma}_t + \Sigma_t)$ denote the matrix in the Mahalanobis norm. Then,

$$\left\| \sum_{\tau=1}^{t-1} Z_\tau \zeta_\tau \right\|^2_{A^{-1}} = M^2 \left\| \sum_{\tau=1}^{t-1} Z_\tau \zeta_\tau \right\|^2_{(M^2 A)^{-1}}.$$

Now apply Lemma 16 with $Q \triangleq M^2(3\lambda - 9d\log(9T) - 8\log(4T/\delta))I$. Since we require $Q \succ 0$, this requires $\lambda > 3d\log(9T) - 8/3\log(4T/\delta)$, which is satisfied under the preconditions for the lemma. Under this assumption, we get

$$\begin{aligned}\| \sum_{\tau=1}^{t-1} Z_\tau \zeta_\tau \|^2_{(\lambda I + \hat{\Sigma}_t)^{-1}} &\leq 3M^2 \| \sum_{\tau=1}^{t-1} Z_\tau \zeta_\tau \|^2_{(Q + M^2(\hat{\Sigma}_t + \Sigma_t))^{-1}} \\ &\leq 6M^2 \log\left(\frac{4T}{\delta} \sqrt{\frac{\det(Q + M^2(\hat{\Sigma}_t + \Sigma_t))}{\det(Q)}} \right),\end{aligned}$$

with probability at least $1-\delta/(2T)$. With a union bound, the inequality holds simultaneously for all T, with probability at least $1-\delta$.

The last step is to analyze the determinant. Using the same argument as in the proof of Lemma 9, it is not hard to show that

$$\left(\frac{\det(Q + M^2(\hat{\Sigma}_t + \Sigma_t))}{\det(Q)} \right)^{1/d} \leq 1 + \frac{2(t-1)}{d(3\lambda - 9d\log(9T) - 8\log(4T/\delta))}.$$

If we impose the slightly stronger condition that $\lambda \geq 4d\log(9T) + 8\log(4T/\delta)$, then the term in the denominator is at least 1, and then we have that

$$\|\hat{\theta}_t - \theta\|_{\Gamma_t} \leq \sqrt{\lambda} + \sqrt{6M^2\log(4T/\delta) + 3dM^2\log(1+2T/d)}.$$

Finally, as in the two-action case, we use the fact that $|\zeta_t| \leq 3 \triangleq M$. □

Recall the setting of $\gamma(T) \triangleq \sqrt{\lambda} + \sqrt{27d\log(1+2T/d) + 54\log(4T/\delta)}$ and the definition of $\lambda \triangleq 4d\log(9T) + 8\log(4T/\delta)$. For the remainder of the proof, condition on the probability $1-\delta$ event that Lemma 17 holds. We now turn to analyzing the regret.

Lemma 18. *Let $\mu_t \triangleq \mathbb{E}_{a\sim\pi_t} z_{t,a}$ where π_t is the solution to (3) and assume the conclusion of Lemma 17 holds. Then with probability at least $1-\delta$*

$$\textit{Reg}(T) \le (1+6\gamma(T))\sqrt{2T\log(2/\delta)} + 3\gamma(T)\sqrt{T\sum_{t=1}^{T}\mathrm{tr}(\Gamma_t^{-1}(z_{t,a_t}-\mu_t)(z_{t,a_t}-\mu_t)^\top)}.$$

This lemma is slightly more complicated than Lemma 6.

Proof. First, using the same application of Azuma's inequality as in the proof of Lemma 6, with probability $1-\delta/2$, we have

$$\mathrm{Reg}(T) \le \sqrt{2T\log(2/\delta)} + \sum_{t=1}^{T}\mathbb{E}_{a\sim\pi_t}[\langle\theta, z_{t,a_t^\star} - z_{t,a}\rangle \mid \mathcal{F}_t].$$

Now we work with this latter expected regret

$$\sum_{t=1}^{T}\mathbb{E}_{a\sim\pi_t}[\langle\theta, z_{t,a_t^\star} - z_{t,a}\rangle \mid \mathcal{F}_t] = \sum_{t=1}^{T}\langle\theta, z_{t,a_t^\star} - \mu_t\rangle \le \sum_{t=1}^{T}\langle\hat{\theta}, z_{t,a_t^\star} - \mu_t\rangle + \gamma(T)\|z_{t,a_t^\star} - \mu_t\|_{\Gamma_t^{-1}}.$$

For the first term, we use the filtration condition (2)

$$\begin{aligned}\langle\hat{\theta}, z_{t,a_t^\star} - \mu_t\rangle &= \sum_{a\in\mathcal{A}_t}\pi_t(a)\langle\hat{\theta}, z_{t,a_t^\star} - z_{t,a}\rangle \le \gamma(T)\sum_{a\in\mathcal{A}_t}\pi_t(a)\|z_{t,a_t^\star} - z_{t,a}\|_{\Gamma_t^{-1}}\\ &\le \gamma(T)\|z_{t,a_t^\star} - \mu_t\|_{\Gamma_t^{-1}} + \gamma(T)\sum_{a\in\mathcal{A}_t}\pi_t(a)\|z_{t,a} - \mu_t\|_{\Gamma_t^{-1}}.\end{aligned}$$

Applying the feasibility condition in (3), we can bound the expected regret by

$$\sum_{t=1}^{T}\mathbb{E}_{a\sim\pi_t}[\langle\theta, z_{t,a_t^\star} - z_{t,a}\rangle \mid \mathcal{F}_t] \le 3\gamma(T)\sum_{t=1}^{T}\sqrt{\mathrm{tr}(\Gamma_t^{-1}\operatorname*{Cov}_{a\sim\pi_t}(z_{t,a}))} \le 3\gamma(T)\sqrt{T\sum_{t=1}^{T}\mathrm{tr}(\Gamma_t^{-1}\operatorname*{Cov}_{a\sim\pi_t}(z_{t,a}))}.$$

To complete the proof, we need to relate the covariance, which takes expectation over the random action, with the particular realization in the algorithm, since this realization affects the term Γ_{t+1}. Let $Z_t \triangleq z_{t,a_t} - \mu_t$ denote the centered realization, then the covariance term is

$$\operatorname*{Cov}_{a\sim\pi_t}(z_{t,a}) = \mathbb{E}[Z_tZ_t^\top \mid \mathcal{F}_t]$$

In order to derive a bound on $\sum_{t=1}^{T}\mathrm{tr}(\Gamma_t^{-1}\operatorname{Cov}_{a\sim\pi_t}(z_{t,a}))$, we first consider the following

$$\sum_{t=1}^{T}\mathrm{tr}(\Gamma_t^{-1}\mathbb{E}[Z_tZ_t^\top \mid \mathcal{F}_t]) - \mathrm{tr}(\Gamma_t^{-1}Z_tZ_t^\top).$$

Observe that sequence of sums $\{\sum_{t=1}^{\tau}\mathrm{tr}(\Gamma_t^{-1}\mathbb{E}[Z_tZ_t^\top \mid \mathcal{F}_t]) - \mathrm{tr}(\Gamma_t^{-1}Z_tZ_t^\top)\}_{\tau=1}^{T}$ is a martingale. Also, each term $\mathrm{tr}(\Gamma_t^{-1}\mathbb{E}[Z_tZ_t^\top \mid \mathcal{F}_t]) - \mathrm{tr}(\Gamma_t^{-1}Z_tZ_t^\top)$ is bounded by 1 because $\Gamma_1 = \lambda I$ and $\lambda > 1$. Applying the Freedman's inequality reveals that with probability at least $1-\delta/2$

$$\begin{aligned}\sum_{t=1}^{T}\mathrm{tr}(\Gamma_t^{-1}\mathbb{E}[Z_tZ_t^\top \mid \mathcal{F}_t]) - \mathrm{tr}(\Gamma_t^{-1}Z_tZ_t^\top) &\le 2\sqrt{\sum_{t=1}^{T}\mathbb{E}[(Z_t^\top\Gamma_t^{-1}Z_t)^2 \mid \mathcal{F}_t]\log(2/\delta)} + 2\log(2/\delta)\\ &\le 2\sqrt{\sum_{t=1}^{T}\mathrm{tr}(\Gamma_t^{-1}\mathbb{E}[Z_tZ_t^\top \mid \mathcal{F}_t])\log(2/\delta)} + 2\log(2/\delta)\\ &\le \frac{1}{2}\sum_{t=1}^{T}\mathrm{tr}(\Gamma_t^{-1}\mathbb{E}[Z_tZ_t^\top \mid \mathcal{F}_t]) + 4\log(2/\delta).\end{aligned}$$

Then rearranging and plugging back into our regret bound, we have

$$\begin{aligned}\mathrm{Reg}(T) &\leq \sqrt{2T\log(2/\delta)} + 3\gamma(T)\sqrt{2T\left(\sum_{t=1}^{T}\mathrm{tr}(\Gamma_t^{-1}Z_tZ_t^\top) + 4\log(2/\delta)\right)} \\ &\leq (1+6\gamma(T))\sqrt{2T\log(2/\delta)} + 3\gamma(T)\sqrt{2T\sum_{t=1}^{T}\mathrm{tr}(\Gamma_t^{-1}Z_tZ_t^\top)}. \qquad \square\end{aligned}$$

To conclude the proof of the theorem, apply Lemma 7, which applies on the last term on the RHS of Lemma 18. Overall, with probability at least $1-2\delta$, we get

$$\mathrm{Reg}(T) \leq (1+6\gamma(T))\sqrt{2T\log(2/\delta)} + 3\gamma(T)\sqrt{2Td(1+1/\lambda)\log(1+T/(d\lambda))}.$$

Since $\lambda = \Theta(d\log(T/\delta))$ and $\gamma(T) = O(\sqrt{d\log(T)} + \sqrt{\log(T/\delta)})$, we get with probability $1-\delta$,

$$\mathrm{Reg}(T) \leq O\left(d\sqrt{T}\log(T) + \sqrt{dT\log(T)\log(T/\delta)} + \sqrt{T\log(T/\delta)\log(1/\delta)}\right).$$