## A Projection Lemma

**Proposition A.1.** *For any PSD matrix $A$ with dimension $d$, any closed convex set $\mathcal{B}$ in the Euclidian space $\mathbb{R}^d$, and $\widehat{\boldsymbol{x}} \in \mathbb{R}^d$, let*

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}\in\mathcal{B}}{\operatorname{argmin}}\, g(\widehat{\boldsymbol{x}}, \boldsymbol{x})$$

*where*

$$g(\boldsymbol{u}, \boldsymbol{v}) \triangleq (\boldsymbol{u}-\boldsymbol{v})^\intercal A(\boldsymbol{u}-\boldsymbol{v}),$$

*then*

$$g(\boldsymbol{x}^*, \boldsymbol{x}_0) \leq g(\widehat{\boldsymbol{x}}, \boldsymbol{x}_0) \qquad \forall \boldsymbol{x}_0 \in \mathcal{B}.$$

*More generally,*

$$g(\boldsymbol{x}^*, \boldsymbol{z}_0) \leq g(\widehat{\boldsymbol{x}}, \boldsymbol{z}_0) + \min_{\boldsymbol{x}\in\mathcal{B}} g(\boldsymbol{z}_0, \boldsymbol{x}) \qquad \forall \boldsymbol{z}_0 \in \mathbb{R}^d.$$

*Proof.* We first provide a proof for $\boldsymbol{x}_0 \in \mathcal{B}$. For any $\alpha \in [0,1]$, let

$$\boldsymbol{x}_\alpha \triangleq \alpha \boldsymbol{x}^* + (1-\alpha)\boldsymbol{x}_0.$$

By convexity, we have $\boldsymbol{x}_\alpha \in \mathcal{B}$ for any $\alpha$. Note that $g(\widehat{\boldsymbol{x}}, \boldsymbol{x}_\alpha)$ is differentiable. By the definition of $x^*$, we have

$$(\boldsymbol{x}^* - \widehat{\boldsymbol{x}})^\intercal A(\boldsymbol{x}^* - \boldsymbol{x}_0) = \frac{1}{2}\frac{\partial}{\partial\alpha} g(\widehat{\boldsymbol{x}}, \boldsymbol{x}_\alpha)\Big|_{\alpha=1} \leq 0.$$

Therefore,

$$g(\boldsymbol{x}^*, \boldsymbol{x}_0) = g(\widehat{\boldsymbol{x}}, \boldsymbol{x}_0) + 2(\boldsymbol{x}^* - \widehat{\boldsymbol{x}})^\intercal A(\boldsymbol{x}^* - \boldsymbol{x}_0) - g(\boldsymbol{x}^*, \widehat{\boldsymbol{x}}) \leq g(\widehat{\boldsymbol{x}}, \boldsymbol{x}_0),$$

where the last inequality uses the PSD property of $A$.

Now we consider the more general case and let $\boldsymbol{x}$ be any vector in $\mathcal{B}$. Following the same steps in the earlier case, we have

$$(\boldsymbol{x}^* - \widehat{\boldsymbol{x}})^\intercal A(\boldsymbol{x}^* - \boldsymbol{x}) \leq 0.$$

Hence,

$$\begin{aligned} g(\boldsymbol{x}^*, \boldsymbol{z}_0) - g(\widehat{\boldsymbol{x}}, \boldsymbol{z}_0) &= 2(\boldsymbol{x}^* - \widehat{\boldsymbol{x}})^\intercal A(\boldsymbol{x}^* - \boldsymbol{z}_0) - g(\boldsymbol{x}^*, \widehat{\boldsymbol{x}}) \\ &\leq 2(\boldsymbol{x}^* - \widehat{\boldsymbol{x}})^\intercal A(\boldsymbol{x} - \boldsymbol{z}_0) - g(\boldsymbol{x}^*, \widehat{\boldsymbol{x}}) \\ &= g(\boldsymbol{z}_0, \boldsymbol{x}) - (\boldsymbol{x} - \boldsymbol{z}_0 - \boldsymbol{x}^* + \widehat{\boldsymbol{x}})^\intercal A\,(\boldsymbol{x} - \boldsymbol{z}_0 - \boldsymbol{x}^* + \widehat{\boldsymbol{x}}) \\ &\leq g(\boldsymbol{z}_0, \boldsymbol{x}). \end{aligned}$$

Note that the above inequality holds for any $\boldsymbol{x} \in \mathcal{B}$. The proposition is proved by taking the minimum over $\boldsymbol{x}$.

□

## B Proof of Proposition 3.3

*Proof.* We first prove for the case where $\boldsymbol{Z}$ is deterministic. Let $\mu_{\boldsymbol{Z}}$ denote the conditional expectation of $\theta$. By Cauchy's inequality,

$$\mathbb{E}[(\theta-\mu_{\boldsymbol{Z}})^2|\boldsymbol{Z}] \cdot \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln f_{\boldsymbol{Z}}(\theta)\right)^2 \middle| \boldsymbol{Z}\right] \geq \mathbb{E}\left[\left|(\theta-\mu_{\boldsymbol{Z}}) \cdot \frac{\partial}{\partial\theta}\ln f_{\boldsymbol{Z}}(\theta)\right| \middle| \boldsymbol{Z}\right]^2. \qquad (22)$$

The quantity on the RHS above can be bounded as follows.

$$\begin{aligned}
\mathbb{E}\left[\left|(\theta-\mu_{\boldsymbol{Z}})\cdot\frac{\partial}{\partial\theta}\ln f_{\boldsymbol{Z}}(\theta)\right|\middle|\boldsymbol{Z}\right] &= \int\left|(\theta-\mu_{\boldsymbol{Z}})\cdot\frac{\partial}{\partial\theta}\ln f_{\boldsymbol{Z}}(\theta)\right| f_{\boldsymbol{Z}}(\theta)d\theta \\
&= \int\left|(\theta-\mu_{\boldsymbol{Z}})\cdot\frac{\partial}{\partial\theta} f_{\boldsymbol{Z}}(\theta)\right| d\theta \\
&\geq \limsup_{T\to+\infty}\left|\int_{-T}^{T}(\theta-\mu_{\boldsymbol{Z}})\cdot\frac{\partial}{\partial\theta} f_{\boldsymbol{Z}}(\theta)d\theta\right| \\
&= \limsup_{T\to+\infty}\left|\left((\theta-\mu_{\boldsymbol{Z}})f_{\boldsymbol{Z}}(\theta)\Big|_{\theta=-T}^{\theta=T}\right)-\mathbb{P}[\theta\in[-T,T]|\boldsymbol{Z}]\right| \\
&\geq 1,
\end{aligned}$$

where the last inequality uses the integrability of $f_{\boldsymbol{Z}}$, which implies

$$\liminf_{T\to+\infty}(\theta-\mu_{\boldsymbol{Z}})f_{\boldsymbol{Z}}(\theta)\Big|_{\theta=-T}^{\theta=T}\leq 0.$$

Then we evaluate the second factor on the LHS of inequality (22). Recall that $\frac{\partial^2}{\partial\theta^2}\ln f_{\boldsymbol{Z}}(\theta)$ is integrable, the following limit exists.

$$\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\ln f_{\boldsymbol{Z}}(\theta)\middle|\boldsymbol{Z}\right]=\lim_{T\to+\infty}\int_{-T}^{T} f_{\boldsymbol{Z}}(\theta)\frac{\partial^2}{\partial\theta^2}\ln f_{\boldsymbol{Z}}(\theta)d\theta.$$

Then by positivity, we also have

$$\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln f_{\boldsymbol{Z}}(\theta)\right)^2\middle|\boldsymbol{Z}\right]=\lim_{T\to+\infty}\int_{-T}^{T} f_{\boldsymbol{Z}}(\theta)\left(\frac{\partial}{\partial\theta}\ln f_{\boldsymbol{Z}}(\theta)\right)^2 d\theta.$$

If we focus the non-trivial case where the first limit is not $-\infty$, the above two equation implies the existence of the following limit.

$$\begin{aligned}
&\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\ln f_{\boldsymbol{Z}}(\theta)\middle|\boldsymbol{Z}\right]+\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln f_{\boldsymbol{Z}}(\theta)\right)^2\middle|\boldsymbol{Z}\right] \\
&=\lim_{T\to+\infty}\int_{-T}^{T} f_{\boldsymbol{Z}}(\theta)\left(\frac{\partial^2}{\partial\theta^2}\ln f_{\boldsymbol{Z}}(\theta)+\left(\frac{\partial}{\partial\theta}\ln f_{\boldsymbol{Z}}(\theta)\right)^2\right)d\theta \\
&=\lim_{T\to+\infty} f_{\boldsymbol{Z}}(\theta)\frac{\partial}{\partial\theta}\ln f_{\boldsymbol{Z}}(\theta)\Big|_{\theta=-T}^{\theta=T} \\
&=\lim_{T\to+\infty}\frac{\partial}{\partial\theta} f_{\boldsymbol{Z}}(\theta)\Big|_{\theta=-T}^{\theta=T}.
\end{aligned}$$

The result of the above equation has to be zero, because the limit points of $\frac{\partial}{\partial\theta}f_{\boldsymbol{Z}}(\theta)$ must contain zero on both ends of the real line, which is implied by the integrability of $f_{\boldsymbol{Z}}$. Consequently, we have

$$\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln f_{\boldsymbol{Z}}(\theta)\right)^2\middle|\boldsymbol{Z}\right]=\mathbb{E}\left[-\frac{\partial^2}{\partial\theta^2}\ln f_{\boldsymbol{Z}}(\theta)\middle|\boldsymbol{Z}\right]. \tag{23}$$

Then, the special case of Proposition 3.3 with fixed $\boldsymbol{Z}$ is implied by inequality (22).

When $\boldsymbol{Z}$ is variable, we simply have

$$\begin{aligned}
\mathbb{E}\left[\mathrm{Var}[\theta|\boldsymbol{Z}]\right] &\geq \mathbb{E}\left[1/\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln f_{\boldsymbol{Z}}(\theta)\right)^2\middle|\boldsymbol{Z}\right]\right] \\
&\geq \frac{1}{\mathbb{E}\left[\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln f_{\boldsymbol{Z}}(\theta)\right)^2|\boldsymbol{Z}\right]\right]}.
\end{aligned}$$

Then the proposition is implied by equation (23). □

## C Proof of Theorem 2.2

We first investigate the lower bounds. Observe that the proof provided in Section 3.1 only fails when the constructed hard instances have $||\boldsymbol{x}_0||_2 > 1$. Hence, we have already covered the $T \geq \left(\sum_{k=1}^{d} \lambda_k^{-\frac{1}{2}}\right)\left(\sum_{k=1}^{d} \lambda_k^{-\frac{3}{2}}\right)$ case, i.e., when $k^* = \text{dim}A = d$. It remains to consider the other scenarios, where $k^* < d$ is satisfied.

By the assumption that $T \geq \left(\sum_{k=1}^{k^*} \lambda_k^{-\frac{1}{2}}\right)\left(\sum_{k=1}^{k^*} \lambda_k^{-\frac{3}{2}}\right)$, one can instead set the entries of $\boldsymbol{x}_0$ in the earlier proof with indices greater than $k^*$ to be zero, so that $||\boldsymbol{x}_0||_2 \leq 1$ is satisfied. Formally, let the hard-instance functions be constructed by the following set.

$$\boldsymbol{x}_0 \in \mathcal{X}_{\text{H}} \triangleq \left\{ (x_1, x_2, ..., x_{k^*}, 0, ..., 0) \,\middle|\, x_k = \pm\sqrt{\frac{\lambda_k^{-\frac{3}{2}}\left(\sum_j \lambda_j^{-\frac{1}{2}}\right)}{2T}}, \forall k \in [k^*] \right\}.$$

Then by the identical proof steps, we have $\mathfrak{R}(T; A) = \Omega\left(\left(\sum_{k=1}^{k^*} \lambda_k^{-\frac{1}{2}}\right)^2 \Big/ T\right)$.

Next, we show that $\mathfrak{R}(T; A) = \Omega\left(\lambda_{k^*+1}\right)$. We assume the non-trivial case where $\lambda_{k^*+1} \neq 0$. Note that $\mathfrak{R}(T; A)$ is non-increasing w.r.t. $T$. We can lower bound $\mathfrak{R}(T; A)$ through the above steps but by replacing $T$ with any larger quantity. Specifically, recall that $k^*$ is largest integer satisfying $T \geq \left(\sum_{k=1}^{k^*} \lambda_k^{-\frac{1}{2}}\right)\left(\sum_{k=1}^{k^*} \lambda_k^{-\frac{3}{2}}\right)$, which implies $T \leq \left(\sum_{k=1}^{k^*+1} \lambda_k^{-\frac{1}{2}}\right)\left(\sum_{k=1}^{k^*+1} \lambda_k^{-\frac{3}{2}}\right)$. We have,

$$\mathfrak{R}(T; A) \geq \mathfrak{R}\left(\left(\sum_{k=1}^{k^*+1} \lambda_k^{-\frac{1}{2}}\right)\left(\sum_{k=1}^{k^*+1} \lambda_k^{-\frac{3}{2}}\right); A\right).$$

Notice that this change of sampling time allows us to apply the earlier lower bound with $k^*$ incremented by 1.

$$\begin{aligned}
\mathfrak{R}(T; A) &\geq \Omega\left(\frac{\left(\sum_{k=1}^{k^*+1} \lambda_k^{-\frac{1}{2}}\right)^2}{\left(\sum_{k=1}^{k^*+1} \lambda_k^{-\frac{1}{2}}\right)\left(\sum_{k=1}^{k^*+1} \lambda_k^{-\frac{3}{2}}\right)}\right) \\
&= \Omega\left(\frac{\sum_{k=1}^{k^*+1} \lambda_k^{-\frac{1}{2}}}{\sum_{k=1}^{k^*+1} \lambda_k^{-\frac{3}{2}}}\right) = \Omega(\lambda_{k^*+1}).
\end{aligned}$$

To conclude,

$$\mathfrak{R}(T; A) = \Omega\left(\max\left\{\frac{\left(\sum_{k=1}^{k^*} \lambda_k^{-\frac{1}{2}}\right)^2}{T}, \lambda_{k^*+1}\right\}\right) = \Omega\left(\frac{\left(\sum_{k=1}^{k^*} \lambda^{-\frac{1}{2}}\right)^2}{T} + \lambda_{k^*+1}\right),$$

which completes the proof of the lower bounds.

The needed upper bounds can be obtained by only estimating the first $k^*$ entries of $\boldsymbol{x}_0$.

**Remark C.1.** *The requirement of $T > 3\,\text{dim}A$ in the Theorem statement is simply due to the integer constraints for the achievability bounds. Indeed, when $\lambda_{\text{dim}A}$ is large, it requires at least $\Omega(\text{dim}A)$ samples to achieve $O(1)$ expected simple regret.*

## D Proof Details for Theorem 2.4

### D.1 Truncation Method and Its Applications

The truncation method is based on the following facts.

**Proposition D.1.** *For any sequence of independent random variables $X_1, X_2, ..., X_n$ and any fixed parameter $m$ satisfying $m > \max_k |\mathbb{E}[X_k]|$. Let $Z_k = \max\{\min\{X_k, m\}, -m\}$ for any $k \in [n]$, we have*

$$|\mathbb{E}[Z_k] - \mathbb{E}[X_k]| \leq \frac{1}{4} \cdot \frac{\text{Var}[X_k]}{m - |\mathbb{E}[X_k]|}, \tag{24}$$

$$\text{Var}[Z_k] \leq \mathbb{E}\left[(Z_k - \mathbb{E}[X_k])^2\right] \leq \text{Var}[X_k]. \tag{25}$$

*Moreover, for any $z > 0$, we have*

$$\mathbb{P}\left[\left|\sum_k Z_k - \sum_k \mathbb{E}[X_k]\right| \geq z\right] \leq 2\exp\left(\sum_k \frac{\text{Var}[X_k]}{m(m - |\mathbb{E}[X_k]|)} - \frac{z}{m}\right). \tag{26}$$

*Proof.* The first inequality is proved by expressing the LHS with piecewise linear functions. Note that by the definition of $Z_k$, we have

$$\begin{aligned} |\mathbb{E}[Z_k] - \mathbb{E}[X_k]| &= |\mathbb{E}[\max\{-m - X_k, 0\}] - \mathbb{E}[\max\{X_k - m, 0\}]| \\ &\leq |\mathbb{E}[\max\{-m - X_k, 0\}]| + |\mathbb{E}[\max\{X_k - m, 0\}]| \\ &= \mathbb{E}[\max\{|X_k| - m, 0\}]. \end{aligned}$$

We apply the following inequalities, which holds for any $m \geq |\mathbb{E}[X_k]|$.

$$|X_k| - m \leq |X_k - \mathbb{E}[X_k]| - m + \mathbb{E}[X_k] \leq \frac{1}{4} \cdot \frac{|X_k - \mathbb{E}[X_k]|^2}{m - \mathbb{E}[X_k]}.$$

Therefore,

$$\begin{aligned} |\mathbb{E}[Z_k] - \mathbb{E}[X_k]| &\leq \mathbb{E}\left[\frac{1}{4} \cdot \frac{|X_k - \mathbb{E}[X_k]|^2}{m - \mathbb{E}[X_k]}\right] \\ &= \frac{1}{4} \cdot \frac{\text{Var}[X_k]}{m - |\mathbb{E}[X_k]|}. \end{aligned}$$

The second inequality is due to the following elementary facts,

$$\mathbb{E}[(Z_k - \mathbb{E}[X_k])^2] \leq \mathbb{E}[(X_k - \mathbb{E}[X_k])^2] = \text{Var}[X_k],$$

where the inequality step is implied by the definition of $Z_k$ and the condition $m > \max_k |\mathbb{E}[X_k]|$.

To prove the third inequality, we first investigate the following upper bound, which is due to Markov's inequality.

$$\begin{aligned} \mathbb{P}\left[\sum_k Z_k - \sum_k \mathbb{E}[X_k] \geq z\right] &\leq \frac{\mathbb{E}[e^{\frac{1}{m}(\sum_k Z_k - \sum_k \mathbb{E}[X_k])}]}{e^{\frac{z}{m}}} \\ &= \frac{\prod_k \mathbb{E}[e^{\frac{1}{m}(Z_k - \mathbb{E}[X_k])}]}{e^{\frac{z}{m}}} \end{aligned} \tag{27}$$

The equality step above is by the fact that $Z_k$'s are jointly independent. For each $k$, using the fact that $Z_k$ is bounded, particularly, $Z_k - \mathbb{E}[X_k] \leq m + |\mathbb{E}[X_k]|$, we have the following inequality

$$e^{\frac{1}{m}(Z_k - \mathbb{E}[X_k])} - 1 - \frac{1}{m}(Z_k - \mathbb{E}[X_k]) \leq (Z_k - \mathbb{E}[X_k])^2 \cdot \frac{e^{\frac{1}{m}(m + |\mathbb{E}[X_k]|)} - 1 - \frac{1}{m}(m + |\mathbb{E}[X_k]|)}{(m + |\mathbb{E}[X_k]|)^2}.$$

For brevity, let $\theta \triangleq \frac{|\mathbb{E}[X_k]|}{m}$. We combine the above bound with inequality (24) and (25) to obtain that

$$\begin{aligned} \mathbb{E}[e^{\frac{1}{m}(Z_k - \mathbb{E}[X_k])}] &= 1 + \mathbb{E}\left[\frac{1}{m}(Z_k - \mathbb{E}[X_k])\right] + \mathbb{E}\left[e^{\frac{1}{m}(Z_k - \mathbb{E}[X_k])} - 1 - \frac{1}{m}(Z_k - \mathbb{E}[X_k])\right] \\ &\leq 1 + \frac{\text{Var}[X_k]}{m(m - |\mathbb{E}[X_k]|)} \cdot \left(\frac{1}{4} + (1 - \theta) \cdot \frac{e^{1+\theta} - 2 - \theta}{(1 + \theta)^2}\right). \end{aligned} \tag{28}$$

Recall that $\theta < 1$ as assumed in the proposion. From elementary calculus, we have

$$\begin{aligned}\mathbb{E}[e^{\frac{1}{m}(Z_k - \mathbb{E}[X_k])}] &\leq 1 + \frac{\text{Var}[X_k]}{m(m - |\mathbb{E}[X_k]|)} \\ &\leq \exp\left(\frac{\text{Var}[X_k]}{m(m - |\mathbb{E}[X_k]|)}\right).\end{aligned}$$

Therefore, recall inequality (27), we have

$$\mathbb{P}\left[\sum_k Z_k - \sum_k \mathbb{E}[X_k] \geq z\right] \leq \exp\left(\sum_k \frac{\text{Var}[X_k]}{m(m - |\mathbb{E}[X_k]|)} - \frac{z}{m}\right).$$

By symmetry, one can also prove the following bound through the same steps.

$$\mathbb{P}\left[\sum_k Z_k - \sum_k \mathbb{E}[X_k] \leq -z\right] \leq \exp\left(\sum_k \frac{\text{Var}[X_k]}{m(m - |\mathbb{E}[X_k]|)} - \frac{z}{m}\right).$$

Hence, the needed inequality is obtained by adding the two inequalities above. □

Now equation (14) for the 1D case is immediately implied by Proposition D.1. Recall the construction of $\widehat{A}$ in the proof, for any sufficiently large $T_0$, we have

$$\mathbb{P}\left[\left|\widehat{A} - A\right| \geq T_0^{-\alpha}\right] \leq 2\exp\left(3 - \frac{T_0^{0.5-\alpha}}{3}\right) = o\left(\frac{1}{T_0^{\beta}}\right).$$

**Remark D.2.** *Instead of projecting to a bounded interval, the same achievability result can be obtained if the we average over any functions that map the samples to $[-T_0^{0.5}, T_0^{0.5}]$ while imposing an additional error of $o(T^{-\alpha})$ everywhere. This includes $\Theta(\ln T)$-bit uniform quantizers, which naturally appear in digital systems, over which exact computation can be performed to eliminate numerical errors. We present this simple generalization in the following corollary.*

**Corollary D.3.** *Consider the setting in Proposition D.1. Let $Y_1, ..., Y_n$ be variables that satisfy $|Y_k - Z_k| \leq b$ for all $k$ with probability 1. We have*

$$\mathbb{P}\left[\left|\sum_k Y_k - \sum_k \mathbb{E}[X_k]\right| \geq z\right] \leq 2\exp\left(\sum_k \frac{\text{Var}[X_k]}{m(m - |\mathbb{E}[X_k]|)} - \frac{z - bn}{m}\right).$$

### D.2 Proof of Proposition 4.2

*Proof.*

$$\boldsymbol{y}^{\intercal} Z \boldsymbol{y} - \boldsymbol{y}^{\intercal}\widehat{A}_0\boldsymbol{y} = \boldsymbol{y}^{\intercal}(Z - \widehat{A}_0)\boldsymbol{y} \leq ||Z - \widehat{A}_0||_{\text{F}}||\boldsymbol{y}||_2^2 \leq ||Z - \widehat{A}_0||_{\text{F}}||\widehat{A}_0^{-1}||_{\text{F}}(\boldsymbol{y}^{\intercal}\widehat{A}_0\boldsymbol{y}).$$

□

### D.3 Proof of inequality (19)

We apply Proposition 4.2 to inequality (17) and let $Z = A\widehat{A}_0^{-1}A$. Note that

$$||Z - \widehat{A}_0||_{\text{F}} \leq 2||A - \widehat{A}_0||_{\text{F}} + ||(A - \widehat{A}_0)\widehat{A}_0^{-1}(A - \widehat{A}_0)||_{\text{F}} = o(1),$$

which satisfies the condition of Proposition 4.2. Using the fact that $\widehat{A}_0^{-1}\widehat{A}_0\widehat{A}_0^{-1} = \widehat{A}_0^{-1}$, we have

$$\begin{aligned}&\mathbb{E}\left[\left(\widehat{A}_0^{-1}A\left(\widehat{\boldsymbol{x}} - \widehat{A}_0^{-1}A\boldsymbol{x}_0\right)\right)^{\intercal}\widehat{A}_0\left(\widehat{A}_0^{-1}A\left(\widehat{\boldsymbol{x}} - \widehat{A}_0^{-1}A\boldsymbol{x}_0\right)\right)\right] \\ &= \mathbb{E}\left[\left(\widehat{\boldsymbol{x}} - \widehat{A}_0^{-1}A\boldsymbol{x}_0\right)^{\intercal} Z\left(\widehat{\boldsymbol{x}} - \widehat{A}_0^{-1}A\boldsymbol{x}_0\right)\right] = O\left((\text{Tr}(A^{-\frac{1}{2}}))^2\right)/T. \qquad (29)\end{aligned}$$

Then by the triangle inequality for the PSD matrix $\widehat{A}_0$, the combination of the above inequality and inequality (18) gives

$$\mathbb{E}\left[(\widetilde{\boldsymbol{x}}_T - \boldsymbol{z})^{\intercal}\widehat{A}_0(\widetilde{\boldsymbol{x}}_T - \boldsymbol{z})\right] = O\left((\text{Tr}(A^{-\frac{1}{2}}))^2\right)/T.$$

### D.4 Proof of Proposition 4.3

*Proof.* When $A_1$ and $A$ has the same rank, the map $P_1$ is invertible over the column space of $A$. Under such condition, there exists a matrix $X$ such that $A = XP_1A$. Note that $A_1A_1^{-1} = P_1$. We have $XP_1 = XP_1A_1A_1^{-1} = AA_1^{-1}$. Therefore, the needed $A = AA_1^{-1}A$ is obtained by multiplying $A$ on the right-hand sides in the above identity. □