Appendix A. Additional Notation

For a random variable X, we denote by $\|X\|_{\psi_2}$ the subgaussian norm of X, i.e.,

$$\|X\|_{\psi_2} = \inf\{\sigma > 0 : \ \mathbb{E}[\exp(X^2/\sigma^2)] \leq 2\}.$$

For a random vector X, the sugaussian norm is $\|X\|_{\psi_2} = \sup_{v \in \mathcal{S}^{d-1}} \|v^\top X\|_{\psi_2}$. For two reals a, b we denote $a \vee b$ and $a \wedge b$ the maximum and minimum respectively.

Appendix B. One-dimensional Example

Note that $x_{t+1} = \sum_{s=0}^{t-1} a_*^s w_{t-s}$, so $x_{t+1} \sim \mathcal{N}(0, \sum_{s=0}^{t-1} a_*^{2s})$. We will need the following claim.

Claim 4 *Under Assumption 2 with $\beta = \Omega(1)$, we must have $\lambda \leq O\left(\frac{1}{\sqrt{1-a_*}}\right)$.*

Proof Fix a time t. Since $x_{t+1} \sim \mathcal{N}(0, \sum_{s=0}^{t-1} a_*^{2s})$, we have $\mathrm{Var}[x_{t+1}] = O\left(\frac{1}{1-a_*}\right)$, and so from Gaussian concentration: $\mathbb{P}\left[x_{t+1} \geq \kappa/\sqrt{1-a_*}\,\right] \leq \exp(-\Omega(\kappa^2))$. By choosing large enough $\kappa = \Theta(\sqrt{\log(1/\beta)})$, we get that $\mathbb{P}\left[x_{t+1} \geq \kappa/\sqrt{1-a_*}\,\right] \leq \beta/4$. Thus,

$$\mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}\{x_{t+1} \geq \kappa/(1-a_*)\}\right] \leq \beta T/4. \tag{11}$$

Now, suppose that $\lambda < \kappa/\sqrt{1-a_*}$. Then, $\mathbb{E}[|\mathcal{O}|] = \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}\{x_{t+1} \geq \lambda\}\right] \leq \beta T/4$, and so from Markov's inequality, $\mathbb{P}\left[|\mathcal{O}| \geq \beta T/2\right] \leq 1/2$, which contradicts Assumption 2. Thus, $\lambda < \kappa/\sqrt{1-a_*}$, and since $\kappa = \Theta(\sqrt{\log(1/\beta)})$, and $\beta = \Omega(1)$, we are done. ■

The claim implies that $\lambda \leq O\left(\frac{1}{1-a_*}\right)$. Let C denote the constant on this bound, i.e., $\lambda \leq \frac{C}{1-a_*}$. We consider two cases. First, if $x_t \leq \frac{C}{1-a_*}$, then $a_* x_t + C \geq x_t$, i.e, if $x_t \in \mathcal{S}$, then $x_{t+1} \in \mathcal{S}$ with probability $\exp(-O(C^2)) = \Omega(1)$. If $x_t > \frac{C}{1-a_*}$, then $a_* x_t + C \geq \frac{C}{1-a_*} \geq \lambda$. Again, since $C = O(1)$, we are done.

Appendix C. Proof of Proposition 4

We will need some definitions. We define the probabilities $\alpha_t^* := \mathcal{N}(A_* x_t, I; \mathcal{S}_{t+1})$, and the events $\mathcal{E}_t^\alpha := \{\alpha_t^* \geq \alpha\}$, and $\mathcal{O}_t := \{x_t \in \mathcal{S}_t\}$. Also, consider the process

$$B_t := \Big(\mathbb{1}\{\mathcal{O}_{t+1}\} - \alpha\Big) \cdot \mathbb{1}\{\mathcal{O}_t \wedge \mathcal{E}_t^\alpha\},$$

and the filtration $G_t := \sigma(w_0, w_1, \ldots, w_t, \mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_{t+2})$. Then,

$$\begin{aligned}\mathbb{E}[B_t \mid G_{t-1}] &= \mathbb{E}\left[\Big(\mathbb{1}\{\mathcal{O}_{t+1}\} - \alpha\Big) \cdot \mathbb{1}\{\mathcal{O}_t \wedge \mathcal{E}_t^\alpha\} \,\middle|\, G_{t-1}\right] \\ &= \mathbb{P}\left[\mathcal{O}_{t+1} \,\middle|\, G_{t-1}, \mathcal{O}_t \wedge \mathcal{E}_t^\alpha\right] - \alpha \geq 0,\end{aligned} \tag{12}$$

where in the last step we use that given G_{t-1} and $\mathcal{O}_t \wedge \mathcal{E}_t^\alpha$, we have $x_{t+1} \sim \mathcal{N}(A_* x_t, I)$, and $\alpha_t^* \geq \alpha$. Now, let $M_t := B_t - \mathbb{E}[B_t|G_{t-1}]$, and observe that M_t is G_t-measurable, and that $\mathbb{E}[M_t \mid G_{t-1}] = 0$. Since $-1 \leq B_t \leq 1$, we have $-2 \leq M_t \leq 2$. Thus, from Azuma-Hoeffding inequality,

$$\mathbb{P}\left[\sum_{t=1}^{T} M_t \leq -\frac{\alpha\beta T}{4}\right] \leq \exp\left(-\frac{\alpha^2\beta^2}{128}T\right) = o(1). \tag{13}$$

Moreover,

$$\begin{aligned}\sum_{t=1}^{T} M_t &= \sum_{t=1}^{T}\Big(B_t - \mathbb{E}[B_t|G_{t-1}]\Big) \leq^{(12)} \sum_{t=1}^{T} B_t \\ &= \sum_{t=1}^{T} \mathbb{1}\{\mathcal{O}_{t+1} \wedge \mathcal{O}_t \wedge \mathcal{E}_t^\alpha\} - \alpha \sum_{t=1}^{T} \mathbb{1}\{\mathcal{O}_t \wedge \mathcal{E}_t^\alpha\} \\ &\leq |\mathcal{P}| - \alpha \sum_{t=1}^{T} \mathbb{1}\{\mathcal{O}_t \wedge \mathcal{E}_t^\alpha\}. \end{aligned} \tag{14}$$

From Assumptions 2 and 3, we have

$$\mathbb{P}\left[\sum_{t=1}^{T} \mathbb{1}\{\mathcal{O}_t \wedge \mathcal{E}_t^\alpha\} < \beta T - L\right] \;\leq\; o(1). \tag{15}$$

Finally, using the assumed lower bound on T, we get $\beta T - L \geq \frac{3\beta}{4}T$. The proposition follows after combining this with 13,14 and 15. ■

Appendix D. Detailed Proof of Theorem 6

D.1. The Generic bound

We will show the following lemma

Lemma 12 *Independently of how g_i's are chosen,*

$$\left\|\widehat{A} - A_*\right\|_\Sigma^2 \;\leq\; 1 - \underbrace{\sum_{i=1}^{N}\left(2\eta\langle g_i, A_i\widetilde{A}_{i+1} - A_*\rangle - \|(A_i - A_*)x_{t_i}\|^2\right)}_{E_1} + \eta^2 \underbrace{\sum_{i=1}^{N} \operatorname{tr}\left(g_i \Sigma_i^{-1} g_i^\top\right)}_{E_2}. \tag{16}$$

Proof For all i,

$$\widetilde{A}_{i+1} - A_* = A_i - A_* - \eta g_i \Sigma_i^{-1}. \tag{17}$$

Multiplying with Σ_i,

$$(\widetilde{A}_{i+1} - A_*)\Sigma_i = (A_i - A_*)\Sigma_i - \eta g_i. \tag{18}$$

Multiplying 18 with the transpose of 17,

$$(\widetilde{A}_{i+1} - A_*)\Sigma_i(\widetilde{A}_{i+1} - A_*)^\top = ((A_i - A_*)\Sigma_i - \eta g_i)\left(A_i - A_* - \eta g_i \Sigma_i^{-1}\right)^\top$$

Expanding and taking trace,

$$\|\widetilde{A}_{i+1} - A_*\|_{\Sigma_i}^2 = \|A_i - A_*\|_{\Sigma_i}^2 - 2\eta\langle g_i, A_i - A_*\rangle + \eta^2 \operatorname{tr}\left(g_i \Sigma_i^{-1} g_i^\top\right)$$

Since $\mathcal{K}$ is (R, ω)-accurate, $A_* \in \mathcal{K}$, and by convexity of $\mathcal{K}$, $\|A_{i+1} - A_*\|_{\Sigma_i} \leq \|\widetilde{A}_{i+1} - A_*\|_{\Sigma_i}$ (see Hazan (2019) Theorem 2.1). Thus,

$$\|A_{i+1} - A_*\|_{\Sigma_i}^2 \leq \|A_i - A_*\|_{\Sigma_i}^2 - 2\eta\langle g_i, A_i - A_*\rangle + \eta^2 \operatorname{tr}\left(g_i \Sigma_i^{-1} g_i^T\right)$$

Summing over all i,

$$0 \leq \sum_{i=1}^N \left(\|A_i - A_*\|_{\Sigma_i}^2 - \|A_{i+1} - A_*\|_{\Sigma_i}^2\right) - 2\eta \sum_{i=1}^N \langle g_i, A_i - A_*\rangle + \eta^2 \sum_{i=1}^N \operatorname{tr}\left(g_i \Sigma_i^{-1} g_i^\top\right) \quad (19)$$

We now focus on the first sum.

$$\sum_{i=1}^N \left(\|A_i - A_*\|_{\Sigma_i}^2 - \|A_{i+1} - A_*\|_{\Sigma_i}^2\right) = \|A_1 - A_*\|_{\Sigma_1}^2 + \sum_{i=2}^N \left(\|A_i - A_*\|_{\Sigma_i}^2 - \|A_i - A_*\|_{\Sigma_{i-1}}^2\right) - \|A_{N+1} - A_*\|_{\Sigma_N}^2. \quad (20)$$

Also, since $\Sigma_i = \Sigma_{i-1} + x_{t_i} x_{t_i}^T$,

$$\begin{aligned}\|A_i - A_*\|_{\Sigma_i}^2 - \|A_i - A_*\|_{\Sigma_{i-1}}^2 &= \operatorname{tr}\left((A_i - A_*)^\top (A_i - A_*)\Sigma_i\right) - \operatorname{tr}\left((A_i - A_*)^\top (A_i - A_*)\Sigma_{i-1}\right) \\ &= \operatorname{tr}\left((A_i - A_*)^\top (A_i - A_*) x_{t_i} x_{t_i}^\top)\right) = \|(A_i - A_*)x_{t_i}\|^2. \quad (21)\end{aligned}$$

Thus,

$$\begin{aligned}\sum_{i=1}^N \left(\|A_i - A_*\|_{\Sigma_i}^2 - \|A_{i+1} - A_*\|_{\Sigma_i}^2\right) &= \|A_1 - A_*\|_{\Sigma_1}^2 - \|A_{N+1} - A_*\|_{\Sigma_N}^2 + \sum_{i=2}^N \|(A_i - A_*)x_{t_i}\|^2 \\ &=^{(21)} \|A_1 - A_*\|_{\Sigma_0}^2 - \|A_{N+1} - A_*\|_{\Sigma_N}^2 + \sum_{i=1}^N \|(A_i - A_*)x_{t_i}\|^2. \quad (22)\end{aligned}$$

Combining 19 with 22,

$$\|A_{N+1} - A_*\|_{\Sigma_N}^2 \leq \|A_1 - A_*\|_{\Sigma_0}^2 - \sum_{i=1}^N \left(2\eta\langle g_i, A_i - A_*\rangle - \|(A_i - A_*)x_{t_i}\|^2\right) + \eta^2 \sum_{i=1}^N \operatorname{tr}\left(g_i \Sigma_i^{-1} g_i^\top\right).$$

Observe that $\|A_1 - A_*\|_{\Sigma_0}^2 = \|A_0 - A_*\|_{\Sigma_0}^2 \leq 1$, since $A_1 = A_0$ and $A_* \in \mathcal{K}$. The facts that $\widehat{A} = A_{N+1}$ and $\Sigma = \Sigma_N$ finish the proof. ■

D.2. Decomposition of E_1

We start with some definitions. Fix a t, and let

$$\mathcal{T}_t := \Big\{ o_t = 1 \text{ and at step } t\text{, Test returns True} \Big\}.$$

Also, let $O_t = \{o_t = 1\}$. We now define some "good" events. Let $G_t := \{t \notin B\}$,

$$C_t := \{O_t \to \{ \text{ at iteration } i(t)\text{, either Test returns True and } a_t \geq \gamma\text{, or Test returns False and } a_t \leq 4\gamma\}\},$$

$$K_t := \Big\{ \|\Sigma_0^{-1/2} x_t\| \leq R\sqrt{\log T} \Big\}$$

$G_t := \big\{\mathcal{N}(A_* x_t, I; S_t) \geq \alpha\big\}$, and observe that $|B(\alpha)| = \sum_{t=1}^T \mathbb{1}\{\neg G_t \wedge o_t = 1\}$. The next event is about the "correctness" of the Test. Let

$$C_t := \Big\{ o_t = 1 \text{ and at that step, either Test returns True and } \gamma_t \geq \gamma\text{, or Test returns False and } \gamma_t \leq 4\gamma \Big\},$$

where a_t is defined in Section 4.3. We aggregate via $\mathcal{E}_t := G_t \wedge C_t \wedge K_t \wedge O_t$. Now, given $o_t = 1$, let $V_t := 2\eta\langle g(t), A(t) - A_*\rangle - \|(A(t) - A_*)x_t\|^2$, where $A(t)$ and $g(t)$ are defined in the main text. We decompose E_1 as

$$E_1 = \sum_{t=1}^T V_t \mathbb{1}\{\mathcal{T}_t \wedge \mathcal{E}_t\} + \sum_{t=1}^T V_t \mathbb{1}\{\neg\mathcal{T}_t \wedge \mathcal{E}_t\} + \sum_{t=1}^T V_t \mathbb{1}\{\neg\mathcal{E}_t \wedge O_t\}.$$

Let Q_1, Q_2, Q_3 be the above three sums, respectively.

D.3. Large Survival Probability

In this step, we prove that $\mathbb{E}[V_t \mid \mathcal{T}_t \wedge \mathcal{E}_t] \geq 0$, which implies $\mathbb{E}[Q_1] \geq 0$. First of all, we condition on $\mathcal{F}_t$ and $\mathcal{T}_t \wedge \mathcal{E}_t$. Then, letting $\mu_t := A(t)x_t$ and $\mu_t^* := A_* x_t$, we get $g(t) = (z_t - y_t)x_t^\top$, where $y_t \sim \mathcal{N}(\mu_t^*, I, S_t)$ and $z_t \sim \mathcal{N}(\mu_t, I, S_t)$. So,

$$\mathbb{E}\Big[V_t \;\Big|\; \mathcal{F}_t,\ \mathcal{T}_t \wedge \mathcal{E}_t\Big] = 2\eta\langle \nu_t - \nu_t^*, \mu_t - \mu_t^*\rangle - \|\mu_t - \mu_t^*\|^2,$$

where $\nu_t = \mathbb{E}_{z\sim\mathcal{N}(\mu_t, I, S_t)}[z]$ and $\nu_t^* = \mathbb{E}_{y\sim\mathcal{N}(\mu_t^*, I, S_t)}[y]$. Also, note that on $\mathcal{T}_t \wedge \mathcal{E}_t$,

$$\mathcal{N}(\mu_t, I; S_t),\ \mathcal{N}(\mu_t^*, I; S_t)\ \geq\ \gamma$$

Now, using Lemma 10 of the main text (proven in detail there), we see that for large enough constant c_η[20], we get $\mathbb{E}[V_t \mid \mathcal{F}_t,\ \mathcal{T}_t \wedge \mathcal{E}_t] \geq 0$. Indeed, note that $\nabla_\mu L_S(\mu; \mu^*) = \nu - \nu^*$, where $\nu = \mathbb{E}_{z\sim\mathcal{N}(\mu, I, S)}[z]$ and $\nu^* = \mathbb{E}_{y\sim\mathcal{N}(\mu^*, I, S)}[z]$. Thus,

$$\begin{aligned}\mathbb{E}\Big[V_t \;\Big|\; \mathcal{F}_t,\ \mathcal{T}_t \wedge \mathcal{E}_t\Big] \geq (2\eta(\gamma/2)^c - 1)\cdot\|\mu_t - \mu_t^*\|^2 &= \left(2\,(2/\alpha)^{c_\eta}\,(\alpha/4)^{c\cdot c_\gamma} - 1\right)\cdot\|\mu_t - \mu_t^*\|^2 \\ &= \left(\frac{2^{c_\eta + 1 - 2c\cdot c_\gamma}}{a^{c_\eta - c\cdot c_\gamma}} - 1\right)\cdot\|\mu_t - \mu_t^*\|^2 \geq 0,\end{aligned}$$

for any constant $c_\eta \geq 2c \cdot c_\gamma$.

20. Remember that $\eta = (2/\alpha)^{c_\eta}$.

D.4. Small Survival Probability

Here, we prove that $\mathbb{E}[V_t \mid \neg\mathcal{T}_t \wedge \mathcal{E}_t] \geq 0$, which implies $\mathbb{E}[Q_2] \geq 0$. Given $\mathcal{F}_t$ and $\neg\mathcal{T}_t \wedge \mathcal{E}_t$, we have $g(t) = (\mu_t - y_t)x_t^T$, where $y_t \sim \mathcal{N}(\mu_t^*, I, S_t)$. Thus,

$$\mathbb{E}\Big[V_t \;\Big|\; \mathcal{F}_t,\ \neg\mathcal{T}_t \wedge \mathcal{E}_t\Big] = 2\eta\langle \mu_t - \nu_t^*, \mu_t - \mu_t^*\rangle - \|\mu_t - \mu_t^*\|^2. \tag{23}$$

Now, since $\mathcal{E}_t$ implies C_t, we have that given $\neg\mathcal{T}_t \wedge \mathcal{E}_t$, the probability $\mathcal{N}(\mu_t, I; S_t) \leq 4\gamma$. Moreover, $\mathcal{E}_t$ implies G_t, i.e., $\mathcal{N}(\mu_t^*, I; S_t) \geq \alpha$, which from Claim 1 implies $\|\nu_t^* - \mu_t^*\| \leq s(\alpha)$. We claim that $\|\mu_t - \mu_t^*\| > 2s(\alpha)$, provided that $c_\gamma = O(1)$ is sufficiently large. Indeed, suppose $\|\mu_t - \mu_t^*\| \leq 2s(\alpha)$. Combining with $\mathcal{N}(\mu_t^*, I; S_t) \geq \alpha$ and Claim 3, we get

$$4(\alpha/2)^{c_\gamma} = 4\gamma \geq \mathcal{N}(\mu_t, I, S_t) \;\geq\; (\alpha/2)^{O(1)},$$

which is a contradiction for large enough $c_\gamma = O(1)$. Using that $\|\nu_t^* - \mu_t^*\| \leq s(\alpha)$,

$$\begin{aligned}\langle \mu_t - \nu_t^*, \mu_t - \mu_t^*\rangle &= \|\mu_t - \mu_t^*\|^2 + \langle \mu_t^* - \nu_t^*, \mu_t - \mu_t^*\rangle \geq \|\mu_t - \mu_t^*\|^2 - \|\mu_t^* - \nu_t^*\| \cdot \|\mu_t - \mu_t^*\| \\ &\geq \|\mu_t - \mu_t^*\| \cdot \big(\|\mu_t - \mu_t^*\| - s(\alpha)\big) \geq \frac{\|\mu_t - \mu_t^*\|^2}{2}.\end{aligned}$$

Since $\eta \geq 1$, equation 23 implies $\mathbb{E}[V_t \mid \mathcal{F}_t, \neg\mathcal{T} \wedge \mathcal{E}_t] \geq 0$.

Handling Bad Events and Gradient Variance

For Q_3, we show that

$$\mathbb{E}\big[|Q_3|\big] \;\leq\; \widetilde{O}(1) \cdot \mathrm{poly}(1/\alpha) \cdot \big(d + R^2 + R_w^2\big) \cdot (L+1).$$

This is proven using that bad events have small probabilities, relative to V_t's. This is a purely technical proof, and we provide it in Appendix E.

The last step is to bound $E_2 = \sum_{i=1}^N \mathrm{tr}\left(g_i \Sigma_i^{-1} g_i^T\right)$. This is often called "the gradient variance" in optimization. We show that

$$\mathbb{E}\big[E_2\big] \;\leq\; \widetilde{O}(1) \cdot d(d + R^2 + R_w^2).$$

The proof uses the volume-based potential-function argument used in Hazan et al. (2007), which also appears in linear bandit theory (Lattimore and Szepesvári (2020)). We provide the proof in Appendix F. Overall,

$$\begin{aligned}\mathbb{E}\left[\left\|\widehat{A} - A_*\right\|_\Sigma^2\right] &\leq 1 - \mathbb{E}[Q_1] - \mathbb{E}[Q_2] - \mathbb{E}[Q_3] + \eta^2\mathbb{E}[E_2] \\ &\leq \widetilde{O}(1) \cdot \mathrm{poly}(1/\alpha) \cdot \big(d + R^2 + R_w^2\big) \cdot (L+1) + \widetilde{O}(1) \cdot \mathrm{poly}(1/\alpha) \cdot d(d + R^2 + R_w^2) \\ &\leq D + LD',\end{aligned}$$

where $D = dD'$, and $D' = \widetilde{O}(1) \cdot \mathrm{poly}(1/\alpha) \cdot (d + R^2 + R_w^2)$.

Appendix E. Bound on $\mathbb{E}[|Q_3|]$

Lemma 13 $\mathbb{E}[|Q_3|] \leq \widetilde{O}(1) \cdot \text{poly}(1/\alpha) \cdot (d + R^2 + R_w^2) \cdot (L+1)$.

We now prove the above lemma. We start with some definitions (and reminders). If at time t we have $o_t = 1$, then we define the following:

- Let $i(t)$ be the iteration i, which corresponds to time t.
- Let $A(t) \coloneqq A_{i(t)}$, and $\mu_t \coloneqq A(t)x_t$. Also, let $\mu_t^* \coloneqq A_* x_t$.
- Let $\gamma_t \coloneqq \mathcal{N}(\mu_t, I; S_t)$.
- Let z_t be the z at the end of SwitchGrad function, when called at iteration $i(t)$.
- If additionally, Test at iteration $i(t)$ returns True ($\mathcal{T}_t$), then let z_t' be the z' of the SwitchGrad function, at iteration $i(t)$. Observe that on $\mathcal{T}_t$, $z_t = z_t'$.

Now, observe that given $o_t = 1$,

$$V_t = 2\eta\langle z_t - y_t, \mu_t - \mu_t^*\rangle - \|\mu_t - \mu_t^*\|^2.$$

All the above quantities are defined on $O_t = \{o_t = 1\}$ (except for z_t' which is defined on $\mathcal{T}_t \subseteq O_t$). We extend all these, by defining them to be zero outside of O_t ($\mathcal{T}_t$ for z_t'). Furthermore, remember that $\mathcal{E}_t = G_t \wedge C_t \wedge K_t \wedge O_t$, where the events (along with $\mathcal{T}_t$) are defined in D.2. Moreover, in Appendix H (Claim 25), we show that $\mathbb{P}[\neg C_t] \leq 1/T^2$. About K_t, since $\mathcal{K}$ is (R, ω)-accurate, $\mathbb{P}[\neg K_t] \leq O(1/T)$. We first decompose $|Q_3|$.

Claim 5 $|Q_3| \leq \eta Q_4 + \eta Q_5$, *where*

$$Q_4 = \sum_{t=1}^{T} (8\|\mu_t - \mu_t^*\|^2 + 2\|w_t\|^2)\mathbb{1}\{\neg\mathcal{E}_t \wedge O_t\}, \quad \textit{and} \quad Q_5 = \sum_{t=1}^{T} \|z_t' - \mu_t\|^2 \mathbb{1}\{\neg\mathcal{E}_t \wedge \mathcal{T}_t\}. \tag{24}$$

Proof Let $w_t \coloneqq y_t - \mu_t^*$ (the noise). We have $|Q_3| \leq \sum_{t=1}^{T} |V_t| \mathbb{1}\{\neg\mathcal{E}_t \wedge O_t\}$, and

$$\begin{aligned}
|V_t| &\leq 2\eta \left|\langle z_t' - y_t, \mu_t - \mu_t^*\rangle\right| \cdot \mathbb{1}\{\mathcal{T}_t\} + 2\eta \left|\langle \mu_t - y_t, \mu_t - \mu_t^*\rangle\right| + \|\mu_t - \mu_t^*\|^2 \\
&\leq 2\eta\|z_t' - y_t\| \cdot \|\mu_t - \mu_t^*\| \cdot \mathbb{1}\{\mathcal{T}_t\} + 2\eta\|\mu_t - y_t\| \cdot \|\mu_t - \mu_t^*\| + \eta\|\mu_t - \mu_t^*\|^2 \\
&\leq 2\eta\|z_t' - \mu_t\| \cdot \|\mu_t - \mu_t^*\| \cdot \mathbb{1}\{\mathcal{T}_t\} + 4\eta\|\mu_t - y_t\| \cdot \|\mu_t - \mu_t^*\| + \eta\|\mu_t - \mu_t^*\|^2 \\
&\leq 2\eta\|z_t' - \mu_t\| \cdot \|\mu_t - \mu_t^*\| \cdot \mathbb{1}\{\mathcal{T}_t\} + 4\eta\|\mu_t - \mu_t^*\|^2 + 4\eta\|w_t\| \cdot \|\mu_t - \mu_t^*\| + \eta\|\mu_t - \mu_t^*\|^2 \\
&\leq \eta\|z_t' - \mu_t\|^2 \cdot \mathbb{1}\{\mathcal{T}_t\} + 6\eta\|\mu_t - \mu_t^*\|^2 + 2\eta\|\mu_t - \mu_t^*\|^2 + 2\eta\|w_t\|^2 \\
&= \eta\|z_t' - \mu_t\|^2 \cdot \mathbb{1}\{\mathcal{T}_t\} + 8\eta\|\mu_t - \mu_t^*\|^2 + 2\eta\|w_t\|^2
\end{aligned}$$

where we have used that $\eta \geq 1$, and that for all $a, b \in \mathbb{R}$, $2ab \leq a^2 + b^2$. Since $\mathcal{T}_t$ implies O_t,

$$|Q_3| \leq \eta \sum_{t=1}^{T} \|z_t' - \mu_t\|^2 \cdot \mathbb{1}\{\neg\mathcal{E}_t \wedge \mathcal{T}_t\} + \eta \sum_{t=1}^{T} (8\|\mu_t - \mu_t^*\|^2 + 2\|w_t\|^2) \cdot \mathbb{1}\{\neg\mathcal{E}_t \wedge O_t\}.$$

■

We will prove the following claims.

Claim 6 $\mathbb{E}[Q_4] \le \widetilde{O}(R^2 + R_w^2) \cdot (L+1)$.

Claim 7 $\mathbb{E}[Q_5] \le \widetilde{O}(d) \cdot (L+1)$.

Observe that these two claims complete the bound on $\mathbb{E}[|Q_3|]$. We now prove the claims.

E.1. Proof of Claim 6

We will bound $\mathbb{E}[Q_4] = \mathbb{E}\left[\sum_{t=1}^T (8\|\mu_t - \mu_t^*\|^2 + 2\|w_t\|^2)\mathbb{1}\{\neg\mathcal{E}_t \wedge O_t\}\right]$ by proving the following two claims.

Claim 8 $\sum_{t=1}^T \mathbb{E}\left[\|\mu_t - \mu_t^*\|^2 \mathbb{1}\{\neg\mathcal{E}_t \wedge O_t\}\right] \le \widetilde{O}(1) \cdot R^2(L+1)$

Claim 9 $\sum_{t=1}^T \mathbb{E}\left[\|w_t\|^2 \mathbb{1}\{\neg\mathcal{E}_t \wedge O_t\}\right] \le \widetilde{O}(1) \cdot R_w^2(L+1)$

E.1.1. Proofs of Claims 8, 9

On O_t, $A(t) = A_{i(t)}$ and so $A(t), A_* \in \mathcal{K}$, in other words $\|A_0 - A_*\|_{\Sigma_0} \le 1$ and $\|A_0 - A(t)\|_{\Sigma_0} \le 1$. From triangle inequality, $\|A(t) - A_*\|_{\Sigma_0} \le 2$, and so

$$\begin{aligned}
\|\mu_t - \mu_t^*\| &= \|(A(t) - A_*)x_t\| = \|(A(t) - A_*)\Sigma_0^{1/2}\Sigma_0^{-1/2}x_t\| \\
&\le \|(A(t) - A_*)\Sigma_0^{1/2}\|_2 \cdot \|\Sigma_0^{-1/2}x_t\| \\
&\le \|(A(t) - A_*)\Sigma_0^{1/2}\|_{\mathrm{F}} \cdot \|\Sigma_0^{-1/2}x_t\| \le 2\|x_t\|_{\Sigma^{-1}}.
\end{aligned} \tag{25}$$

Thus, since $\mathcal{K}$ is (R, ω)-accurate, $\|\mu_t - \mu_t^*\|$ is $O(R^2)$-subgaussian, which implies

$$\mathbb{P}\left[\|\mu_t - \mu_t^*\| > R\sqrt{\log(TR)}\right] \le \frac{1}{T^2R^4} \quad \text{and} \quad \mathbb{E}\left[\|\mu_t - \mu_t^*\|^4\right] \le O(R^4). \tag{26}$$

Let K_t' be the event in above probability.

$$\begin{aligned}
\mathbb{E}\left[\|\mu_t - \mu_t^*\|^2 \mathbb{1}\{\neg\mathcal{E}_t \wedge O_t\}\right] &= \mathbb{E}\left[\|\mu_t - \mu_t^*\|^2 \mathbb{1}\{\neg\mathcal{E}_t \wedge O_t \wedge (\neg K_t')\}\right] + \mathbb{E}\left[\|\mu_t - \mu_t^*\|^2 \mathbb{1}\{\neg\mathcal{E}_t \wedge O_t \wedge K_t'\}\right] \\
&\le^{(26)} \widetilde{O}(R^2) \cdot \mathbb{P}[\mathcal{E}_t \wedge O_t] + \mathbb{E}[\|\mu_t - \mu_t^*\|^2 \mathbb{1}\{K_t'\}] \\
&\le \widetilde{O}(R^2) \cdot (\mathbb{P}[\neg G_t] + \mathbb{P}[\neg K_t] + \mathbb{P}[\neg C_t]) + \sqrt{\mathbb{E}[\|\mu_t - \mu_t^*\|^4]} \cdot \sqrt{\mathbb{P}[K_t']} \\
&\le^{(26)} \widetilde{O}(R^2) \cdot (\mathbb{P}[\neg G_t] + 1/T) + O(R^2) \cdot \frac{1}{TR^2} \\
&\le \widetilde{O}(R^2) \cdot (\mathbb{P}[\neg G_t] + 1/T).
\end{aligned}$$

Summing over t and using that $\sum_{t=1}^T \mathbb{P}[\neg G_t] \le L$, we finish the proof. For Claim 9, observe that since $\|w_t\|$ is R_w^2 subgaussian, we can just use identical steps as above. ■

E.2. Proof of Claim 7

Let $\mathcal{E}_{t,5} := \neg\mathcal{E}_t \wedge \mathcal{T}_t$, and so $Q_5 = \sum_{t=1}^T \|z_t' - \mu_t\|^2 \mathbb{1}\{\mathcal{E}_{t,5}\}$

Claim 10 *Given $\mathcal{F}_t$ and $\mathcal{E}_{t,5}$, we have that $\|z_t' - \mu_t\|_2$ is $\widetilde{O}(d) \cdot (1 \vee \log(1/\gamma_t))$-subgaussian.*

Proof Given $\mathcal{F}_t$ and $\mathcal{E}_{t,5}$, we have $z'_t \sim \mathcal{N}(\mu_t, I, S_t)$. In Appendix H (Claim 22) we show that this implies $\|z'_t - \mu_t\|_{\psi_2}^2 \leq O(1 \vee \gamma_t)$. Claim 23 in Appendix H finishes the proof. ■

The claim above implies (see Vershynin (2018))

$$\mathbb{E}[\|z'_t - \mu_t\|_2^2 \mid \mathcal{F}_t, \mathcal{E}_{t,5}] \leq \widetilde{O}(d) \cdot (1 \vee \log(1/\gamma_t)). \tag{27}$$

Thus,

$$\begin{aligned}
\mathbb{E}\left[\|z'_t - \mu_t\|^2 \mathbb{1}\{\mathcal{E}_{t,5}\}\right] &= \mathbb{E}\left[\mathbb{E}\left[\|z'_t - \mu_t\|^2 \mathbb{1}\{\mathcal{E}_{t,5}\} \mid \mathcal{F}_t\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\|z'_t - \mu_t\|^2 \mathbb{1}\{\mathcal{E}_{t,5}\} \mid \mathcal{F}_t\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\|z'_t - \mu_t\|^2 \mid \mathcal{F}_t, \mathcal{E}_{t,5}\right] \cdot \mathbb{P}[\mathcal{E}_{t,5} \mid \mathcal{F}_t]\right] \\
&\leq \widetilde{O}(d) \cdot \mathbb{E}\left[(1 \vee \log(1/\gamma_t)) \cdot \mathbb{P}[\mathcal{E}_{t,5} \mid \mathcal{F}_t]\right] \\
&= \widetilde{O}(d) \cdot \mathbb{E}\left[\mathbb{E}\left[(1 \vee \log(1/\gamma_t)) \cdot \mathbb{1}\{\mathcal{E}_{t,5}\} \mid \mathcal{F}_t\right]\right] \\
&= \widetilde{O}(d) \cdot \mathbb{E}\left[(1 \vee \log(1/\gamma_t)) \cdot \mathbb{1}\{\mathcal{E}_{t,5}\}\right],
\end{aligned} \tag{28}$$

where the inequality above is justified by (27). So,

$$\mathbb{E}\left[\|z'_t - \mu_t\|^2 \mathbb{1}\{\mathcal{E}_{t,5}\}\right] \leq \widetilde{O}(d) \cdot \Big(\mathbb{P}\left[\mathcal{E}_{t,5}\right] + \mathbb{E}\left[\log(1/\gamma_t) \cdot \mathbb{1}\{\mathcal{E}_{t,5}\}\right]\Big) \tag{29}$$

We will first bound

$$\begin{aligned}
\mathbb{E}[\log(1/\gamma_t) \cdot \mathbb{1}\{\mathcal{E}_{t,5}\}] &= \mathbb{E}\left[\log(1/\gamma_t) \cdot \mathbb{1}\{C_t \wedge \mathcal{E}_{t,5}\}\right] + \mathbb{E}\left[\log(1/\gamma_t) \cdot \mathbb{1}\{\neg C_t \wedge \mathcal{E}_{t,5}\}\right] \\
&= \mathbb{E}\left[\log(1/\gamma_t) \cdot \mathbb{1}\{C_t \wedge \mathcal{T}_t\} \cdot \mathbb{1}\{\neg(G_t \wedge K_t)\}\right] + \mathbb{E}\left[\log(1/\gamma_t) \cdot \mathbb{1}\{\neg C_t \wedge \mathcal{T}_t\}\right],
\end{aligned} \tag{30}$$

where we used the definition of $\mathcal{E}_{t,5}$. Now, observe that on $C_t \wedge \mathcal{T}_t$, we have $\gamma_t \geq \gamma$. Thus,

$$\begin{aligned}
\mathbb{E}\left[\log(1/\gamma_t) \cdot \mathbb{1}\{C_t \wedge \mathcal{T}_t\} \cdot \mathbb{1}\{\neg(G_t \wedge K_t)\}\right] &\leq \log(1/\gamma) \cdot \mathbb{P}[\neg(G_t \wedge K_t)] \\
&\leq \log(1/\gamma) \cdot (\mathbb{P}[\neg G_t] + \mathbb{P}[\neg K_t]).
\end{aligned} \tag{31}$$

We now bound $\mathbb{E}\left[\log(1/\gamma_t) \cdot \mathbb{1}\{\neg C_t \wedge \mathcal{T}_t\}\right]$ via the following claim.

Claim 11 *Fix $r = O(1)$. Then,* $\mathbb{E}\left[\left(\log \frac{1}{\gamma_t}\right)^r \cdot \mathbb{1}\{\neg C_t \wedge \mathcal{T}_t\}\right] \leq \widetilde{O}(1) \cdot \frac{\gamma^{-1}}{T^2}$.

We state the claim for general r, because later we will need it for $r = 2$. Observe that for $r = 1$, we get $\mathbb{E}\left[\log(1/\gamma_t) \cdot \mathbb{1}\{\neg C_t \wedge \mathcal{T}_t\}\right] \leq \widetilde{O}(1) \cdot \frac{\gamma^{-1}}{T^2}$. We prove the claim at the end of the subsection. Combining with (30) and (31), we get

$$\mathbb{E}[\log(1/\gamma_t) \cdot \mathbb{1}\{\mathcal{E}_{t,5}\}] \leq \log(1/\gamma) \cdot (\mathbb{P}[\neg G_t] + \mathbb{P}[\neg K_t]) + \widetilde{O}(1) \cdot \frac{\gamma^{-1}}{T^2},$$

so from (29),

$$\begin{aligned}
\mathbb{E}\left[\|z'_t - \mu_t\|^2 \mathbb{1}\{\mathcal{E}_{t,5}\}\right] &\leq \widetilde{O}(d) \cdot \left(\mathbb{P}[\mathcal{E}_{t,5}] + \mathbb{P}[\neg G_t] + \mathbb{P}[\neg K_t] + \frac{\gamma^{-1}}{T^2}\right) \\
&\leq \widetilde{O}(d) \cdot \left(\mathbb{P}[\mathcal{E}_{t,5}] + \mathbb{P}[\neg G_t] + \mathbb{P}[\neg K_t] + \frac{\gamma^{-1}}{T^2}\right)
\end{aligned}$$

Observe that $\mathbb{P}[\mathcal{E}_{t,5}] \leq \mathbb{P}[\neg C_t] + \mathbb{P}[\neg K_t] + \mathbb{P}[\neg G_t]$. Also, $\sum_{t=1}^{T} Pr[\neg G_t] = \mathbb{E}[|B(\alpha)|] \leq L$, and $\mathbb{P}[\neg K_t] \leq O(1/T)$ (since $\mathcal{K}$ is (R, ω)-accurate), and $\mathbb{P}[\neg C_t] \leq 1/T^2$ (Appendix H, Claim 25). Thus, from (28),

$$\mathbb{E}[Q_5] \leq \widetilde{O}(d) \cdot \sum_{t=1}^{T} \left(\mathbb{P}[\neg G_t] + 1/T + \frac{\gamma^{-1}}{T^2} \right) \leq \widetilde{O}(d) \cdot (L+1),$$

using that $T \geq \text{poly}(1/\alpha)$ and $1/\gamma = \text{poly}(1/\alpha)$.

We now give the proof of Claim 11.

Proof Let $P_t = \{\gamma_t \geq 1/T^2\}$. Then,

$$\mathbb{E}\Big[\left(\log \frac{1}{\gamma_t}\right)^r \cdot \mathbb{1}\{\neg C_t \wedge \mathcal{T}_t\} \Big] = \mathbb{E}\Big[\left(\log \frac{1}{\gamma_t}\right)^r \cdot \mathbb{1}\{\neg C_t \wedge \mathcal{T}_t \wedge P_t\} \Big] + \mathbb{E}\left[\left(\log \frac{1}{\gamma_t}\right)^r \cdot \mathbb{1}\left\{\neg C_t \wedge \mathcal{T}_t \wedge (\neg P_t)\right\} \right]. \tag{32}$$

Since $\mathcal{T}_t \wedge (\neg P_t)$ implies $\neg C_t$, we have

$$\begin{aligned}
\mathbb{E}\left[\left(\log \frac{1}{\gamma_t}\right)^r \cdot \mathbb{1}\{\neg C_t \wedge \mathcal{T}_t \wedge (\neg P_t)\} \right] &= \mathbb{E}\left[\left(\log \frac{1}{\gamma_t}\right)^r \cdot \mathbb{1}\{\mathcal{T}_t \wedge (\neg P_t)\} \right] \\
&\leq \mathbb{E}\left[\left(\log \frac{1}{\gamma_t}\right)^r \cdot \mathbb{1}\{\mathcal{T}_t\} \,\middle|\, \neg P_t \right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\left(\log \frac{1}{\gamma_t}\right)^r \cdot \mathbb{1}\{\mathcal{T}_t\} \,\middle|\, \gamma_t, \neg P_t \right] \,\middle|\, \neg P_t \right] \\
&= \mathbb{E}\left[\left(\log \frac{1}{\gamma_t}\right)^r \cdot \mathbb{P}\left[\mathcal{T}_t \mid \gamma_t, \neg P_t\right] \,\middle|\, \neg P_t \right].
\end{aligned}$$

Combining with (32), we get

$$\begin{aligned}
\mathbb{E}\Big[\left(\log \frac{1}{\gamma_t}\right)^r \cdot \mathbb{1}\left\{\neg C_t \wedge \mathcal{T}_t\right\} \Big] &\leq \left(\log T^2\right)^r \cdot \mathbb{E}\Big[\mathbb{1}\{\neg C_t \wedge \mathcal{T}_t\}\Big] \\
&\quad + \mathbb{E}\left[\left(\log \frac{1}{\gamma_t}\right)^r \cdot \mathbb{P}\left[\mathcal{T}_t \mid \gamma_t, \neg P_t\right] \,\middle|\, \neg P_t \right].
\end{aligned}$$

From Appendix H (Claim 25), we get $\mathbb{E}\left[\mathbb{1}\{\neg C_t \wedge \mathcal{T}_t\}\right] \leq \mathbb{E}\left[\mathbb{1}\{\neg C_t \wedge O_t\}\right] \leq 1/T^2$. Also, given γ_t, the probability of $\mathcal{T}_t$ is at most $1 - (1-\gamma_t)^k$ (if no ξ_j hits S_t, then the Test definitely returns False), where k is defined in Algorithm 3. From Bernoulli's inequality, $1 - (1-\gamma_t)^k \leq \gamma_t k$, and so

$$\begin{aligned}
\mathbb{E}\left[\left(\log \frac{1}{\gamma_t}\right)^r \cdot \mathbb{P}\left[\mathcal{T}_t \mid \gamma_t, \neg P_t\right] \,\middle|\, \neg P_t \right] &\leq k \cdot \mathbb{E}\left[\left(\log \frac{1}{\gamma_t}\right)^r \cdot \gamma_t \,\middle|\, \gamma_t < 1/T^2 \right] \\
&\leq k \cdot \frac{\left(\log\left(T^2\right)\right)^r}{T^2} \leq \widetilde{O}(1) \cdot \frac{\gamma^{-1}}{T^2},
\end{aligned}$$

where we used that $r = O(1)$ and that T is lower-bounded by a large constant. ■

Appendix F. Bound on Gradient Variance

Lemma 14 $\mathbb{E}[E_2] \leq \widetilde{O}(1) \cdot d \cdot (d + R^2 + R_w^2)$.

In this section, we prove Lemma 14.

$$E_2 = \sum_{i=1}^{N} \mathrm{tr}(g_i \Sigma_i^{-1} g_i^T) \leq \sum_{i=1}^{N} \|z_{t_i} - y_{t_i}\|^2 \cdot (x_{t_i}^T \Sigma_i^{-1} x_{t_i}) \leq \max_i \left\{\|z_{t_i} - y_{t_i}\|^2\right\} \cdot \sum_{i=1}^{N} x_{t_i}^T \Sigma_i^{-1} x_{t_i}$$

Claim 12 $\sum_{i=1}^{N} x_{t_i}^T \Sigma_i^{-1} x_{t_i} \leq d\left(\log |\Sigma_N|^{1/d} + \log(1/\omega)\right)$, *where* $|\Sigma_N|$ *denotes the determinant.*

This Claim is (implicitly) proven in Hazan et al. (2007). We provide a proof for completeness.
Proof We will use the following inequality: if A, B are PSD matrices, then $\langle A^{-1}, A - B\rangle \leq \log \frac{|A|}{|B|}$. For a proof of this see Hazan (2019) Lemma 4.6.

$$\sum_{i=1}^{N} x_{t_i}^\top \Sigma_i^{-1} x_{t_i} = \sum_{i=1}^{N} \langle \Sigma_i^{-1}, x_{t_i} x_{t_i}^\top \rangle = \sum_{i=1}^{N} \langle \Sigma_i^{-1}, \Sigma_i - \Sigma_{i-1} \rangle \leq \sum_{i=1}^{N} \log \frac{|\Sigma_i|}{|\Sigma_{i-1}|} = \log \frac{|\Sigma_N|}{|\Sigma_0|}.$$

The Assumption that $\mathcal{K}$ is (R, ω)-accurate, and so $\Sigma_0 \succcurlyeq \omega \cdot I$ completes the proof. ■

By Cauchy-Schwarz,

$$\mathbb{E}[E_2] \leq O(d) \cdot \sqrt{\mathbb{E}\left[\max_i \|z_{t_i} - y_{t_i}\|^4\right]} \cdot \sqrt{\mathbb{E}\left[\log^2 |\Sigma_N|^{1/d}\right] + \widetilde{O}(1)}.$$

The following claims complete the proof.

Claim 13 $\mathbb{E}\left[\log^2 |\Sigma_N|^{1/d}\right] \leq \log^2\left(e + T \cdot R_x^2\right)$.

Claim 14 $\mathbb{E}\left[\max_i \|z_{t_i} - y_{t_i}\|^4\right] \leq \widetilde{O}(1) \cdot (d^2 + R^4 + R_w^4)$

We now prove these two claims.

F.1. Proof of Claim 14

Let $M := \max_i \|z_{t_i} - y_{t_i}\|$, and also let

1. $M_1 := \max_t \{\|z_t' - \mu_t\| \cdot \mathbb{1}\{C_t \wedge \mathcal{T}_t\}\}$
2. $M_2 := \max_t \{\|z_t' - \mu_t\| \cdot \mathbb{1}\{\neg C_t \wedge \mathcal{T}_t\}\}$
3. $M_3 := \max_t \|{\Sigma_0}^{-1/2} x_t\|$
4. $M_4 := \max_t \|w_t\|$

Proposition 15 $M \leq M_1 + M_2 + 4M_3 + 2M_4$

Proof $M_1 = \max_t \{\|z_t - y_t\| \cdot \mathbb{1}\{O_t\}\}$. Now,

$$\begin{aligned}
\|z_t - y_t\| \cdot \mathbb{1}\{O_t\} &= \|z_t - y_t\| \cdot \mathbb{1}\{\mathcal{T}_t\} + \|z_t - y_t\| \cdot \mathbb{1}\{\neg\mathcal{T}_t \wedge O_t\} \\
&= \|z'_t - y_t\| \cdot \mathbb{1}\{\mathcal{T}_t\} + \|\mu_t - y_t\| \cdot \mathbb{1}\{\neg\mathcal{T}_t \wedge O_t\} \\
&\leq \|z'_t - y_t\| \cdot \mathbb{1}\{\mathcal{T}_t\} + \|\mu_t - y_t\| \cdot \mathbb{1}\{O_t\} \\
&\leq \big(\|z'_t - \mu_t\| + \|\mu_t - y_t\|\big) \cdot \mathbb{1}\{\mathcal{T}_t\} + \|\mu_t - y_t\| \cdot \mathbb{1}\{O_t\} \\
&\leq \|z'_t - \mu_t\| \cdot \mathbb{1}\{\mathcal{T}_t\} + 2\|\mu_t - y_t\| \cdot \mathbb{1}\{O_t\} \\
&\leq \|z'_t - \mu_t\| \cdot \mathbb{1}\{\mathcal{T}_t\} + 2\|\mu_t - \mu_t^*\| \cdot \mathbb{1}\{O_t\} + 2\|w_t\| \\
&\leq \|z'_t - \mu_t\| \cdot \mathbb{1}\{\mathcal{T}_t\} + 4\|\Sigma_0^{-1/2} x_t\| + 2\|w_t\| \\
&= \|z'_t - \mu_t\| \cdot \mathbb{1}\{C_t \wedge \mathcal{T}_t\} + \|z'_t - \mu_t\| \cdot \mathbb{1}\{\neg C_t \wedge \mathcal{T}_t\} + 4\|\Sigma_0^{-1/2} x_t\| + 2\|w_t\|,
\end{aligned}$$

where in the second-to-last step we used (25). ■

We will prove the following claims, which finish the proof of Claim 14.

Claim 15 $\mathbb{E}[M_1^4] \leq \widetilde{O}(d^2)$.

Claim 16 $\mathbb{E}[M_2^4] \leq \widetilde{O}(1) \cdot \frac{d^2}{\gamma T^2}$.

Claim 17 $\mathbb{E}[M_3^4] \leq \widetilde{O}(R^4)$, *and* $\mathbb{E}[M_4^4] \leq \widetilde{O}(R_w^4)$.

First, Claim 17 follows immediately from (a) $\|w_t\|$ being R_w^2-subgaussian, (b) $\|\Sigma_0^{-1/2} x_t\|$ being R^2-subgaussian, and (c) Claim 24 (Appendix H). We now prove the other two claims.

F.1.1. Proof of Claim 15

Let $\sigma^2 := (1 \vee \log(1/\gamma)) \cdot d \cdot \log(2d)$. We will show that for all t, $\|z'_t - \mu_t\| \cdot \mathbb{1}\{C_t \wedge \mathcal{T}_t\}$ is $O(\sigma^2)$-subgaussian. Given this, Claim 24 (Appendix H) finishes the proof. Fix an $r > 0$.

$$\mathbb{P}\left[\|z'_t - \mu_t\| \cdot \mathbb{1}\{C_t \wedge \mathcal{T}_t\} \geq \sigma \cdot r\right] \leq \mathbb{P}\left[\|z'_t - \mu_t\| \geq \sigma \cdot r \;\middle|\; C_t \wedge \mathcal{T}_t\right]$$

Observe given $C_t \wedge \mathcal{T}_t$, we have $\gamma_t \geq \gamma$. Let $\sigma_t^2 := (1 \vee \log(1/\gamma_t)) \cdot d \cdot \log(2d)$.

$$\begin{aligned}
\mathbb{P}\left[\|z'_t - \mu_t\| \cdot \mathbb{1}\{C_t \wedge \mathcal{T}_t\} \geq \sigma \cdot r\right] &\leq \mathbb{P}\left[\|z'_t - \mu_t\| \geq \sigma_t \cdot r \;\middle|\; C_t \wedge \mathcal{T}_t\right] \\
&= \mathbb{E}\left[\mathbb{P}[\|z'_t - \mu_t\| \geq \sigma_t \cdot r \mid \mathcal{F}_t, \mathcal{T}_t] \mid C_t \wedge \mathcal{T}_t\right].
\end{aligned}$$

But, given $\mathcal{F}_t$ and $\mathcal{T}_t$, we have $z'_t \sim \mathcal{N}(\mu_t, I, S_t)$, and so from Claims 22 and 23 (Appendix H), $\|z'_t - \mu_t\|$ is $O(\sigma_t^2)$-subgaussian. Thus, $\mathbb{P}\left[\|z'_t - \mu_t\| \cdot \mathbb{1}\{C_t \wedge \mathcal{T}_t\} \geq \sigma \cdot r\right] \leq e^{-\Omega(r^2)}$. ■

F.1.2. PROOF OF CLAIM 16

As previously, given $\mathcal{F}_t$ and $\neg C_t \wedge \mathcal{T}_t$, we have $z'_t \sim \mathcal{N}(\mu_t, I, S_t)$, and so from Claims 22 and 23 (Appendix H), $\|z'_t - \mu_t\|$ is σ_t^2-subgaussian. Thus,

$$\begin{aligned}
\mathbb{E}\Big[\|z'_t - \mu_t\|^4 \cdot \mathbb{1}\{\neg C_t \wedge \mathcal{T}_t\}\Big] &= \mathbb{P}[\neg C_t \wedge \mathcal{T}_t] \cdot \mathbb{E}\Big[\mathbb{E}\Big[\|z'_t - \mu_t\|^4 \mid \mathcal{F}_t, \neg C_t \wedge \mathcal{T}_t\Big] \;\Big|\; \neg C_t \wedge \mathcal{T}_t\Big] \\
&\le O(1) \cdot \mathbb{E}\left[\sigma_t^4 \cdot \mathbb{1}\{\neg C_t \wedge \mathcal{T}_t\}\right] \\
&\le \widetilde{O}(d^2) \cdot \mathbb{E}\left[(1 \vee \log(1/\gamma_t))^2 \cdot \mathbb{1}\{\neg C_t \wedge \mathcal{T}_t\}\right] \\
&\le \widetilde{O}(d^2) \cdot \Big(\mathbb{P}[\neg C_t] + \mathbb{E}\left[(\log(1/\gamma_t))^2 \cdot \mathbb{1}\{\neg C_t \wedge \mathcal{T}_t\}\right]\Big) \\
&\le \widetilde{O}(d^2) \cdot (1/T^2 + \gamma^{-1}/T^2),
\end{aligned}$$

where in the last step we used Claim 25 (Appendix H) and Claim 11. ■

F.2. Proof of Claim 13

The function $\log^2(x)$ is concave for $x \ge e$. Thus, from Jensen's inequality,

$$\mathbb{E}\left[\log^2 |\Sigma_N|^{1/d}\right] \le \mathbb{E}\left[\log^2\left(e + |\Sigma_N|^{1/d}\right)\right] \le \log^2\left(e + \mathbb{E}\left[|\Sigma_N|^{1/d}\right]\right)$$

Now, from AM-GM, $|\Sigma_N|^{1/d} \le \frac{1}{d} \cdot \mathrm{tr}(\Sigma_N) \le \sum_{t=1}^T \|x_t\|^2$. Thus,

$$\mathbb{E}\left[\log^2 |\Sigma_N|^{1/d}\right] \le \log^2\left(e + T \cdot R_x^2\right).$$

Appendix G. Proof of Theorem 11

For all $\mathcal{I} \subseteq [T]$, we define $X_{\mathcal{I}} \in \mathbb{R}^{d \times |\mathcal{I}|}$ be the matrix whose columns are the $\{x_t\}_{t \in \mathcal{I}}$, placed in temporal order. Also, let $X := X_{[T]}$, and $\Gamma_t := \sum_{s=0}^{t-1} (A_*^s)(A_*^s)^\top$. The following lemma (a) upper and lower bounds the "size" of the covariates, (b) certifies that our algorithms will use enough data and that $|\mathcal{B}(\alpha)|$ is small, and (c) bounds $\|w_t\|$.

Lemma 16 *Let $\mathcal{I}'_0 := \mathcal{I}_0 \setminus \mathcal{B}(\alpha)$, and $C_{\alpha,\beta} = \mathrm{poly}(1/\alpha, 1/\beta)$ (sufficiently large). Also, let*

$$\mathcal{E}_{g,1} := \{|\mathcal{B}(\alpha)| \le L\} \wedge \{|\mathcal{I}_0|, |\mathcal{I}_1| \ge \alpha\beta T/4\}, \quad \textit{and}$$

$$\mathcal{E}_{g,2} := \left\{\frac{1}{T} X X^\top \preccurlyeq \frac{d}{\delta}\Gamma_T\right\} \wedge \left\{\frac{1}{|\mathcal{I}_0|} X_{\mathcal{I}'_0} X_{\mathcal{I}'_0}^\top \succcurlyeq \frac{1}{C_{\alpha,\beta}}\Gamma_T\right\} \wedge \left\{\frac{1}{|\mathcal{I}_1|} X_{\mathcal{I}_1} X_{\mathcal{I}_1}^\top \succcurlyeq \frac{1}{C_{\alpha,\beta}}\Gamma_T\right\},$$

and $\mathcal{E}_{g,3} = \{\forall t : \|w_t\| \le \widetilde{O}(d)\}$. Let $\mathcal{E}_g := \mathcal{E}_{g,1} \wedge \mathcal{E}_{g_2} \wedge \mathcal{E}_{g,3}$. Then, $\mathbb{P}[\mathcal{E}_g] \ge 1 - \delta - o(1)$.

We will prove Lemma 16 in Appendix G.4. We will need some definitions to ease notation. Let $B := \mathcal{B}(\alpha), X_0 := X_{\mathcal{I}_0}, N_0 := |\mathcal{I}_0|, \nu_t^* := \mathbb{E}_{y \sim \mathcal{N}(A_* x_t, I, \mathcal{S}_{t+1})}[y]$, and $\zeta_t := \nu_t^* - A_* x_t$. Finally, for all $\mathcal{I} \subseteq [T]$, let $W_{\mathcal{I}}, Z_{\mathcal{I}}$ denote the $d \times |\mathcal{I}|$ matrices whose columns are $(w_t)_{t \in \mathcal{I}}$ and $(\zeta_t)_{t \in \mathcal{I}}$ respectively (in temporal order).

On $\mathcal{E}_g$, $X_0X_0^\top \succ 0$, so the least-squares solution is unique, i.e., $A_0 = \sum_{t\in\mathcal{I}_0} x_{t+1}x_t^\top \left(X_0X_0^\top\right)^{-1}$. Using $x_{t+1} = A_*x_t + w_t$, we get

$$
\begin{aligned}
(A_0 - A_*)\cdot\left(\frac{X_0X_0^\top}{N_0}\right)^{1/2} &= \frac{1}{\sqrt{N_0}}\sum_{t\in\mathcal{I}_0} w_t x_t^\top \left(X_0X_0^\top\right)^{-1/2} \\
&= \underbrace{\frac{1}{\sqrt{N_0}}\sum_{t\in\mathcal{I}_0'}(w_t-\zeta_t)x_t^\top \left(X_0X_0^\top\right)^{-1/2}}_{\Delta_1} + \underbrace{\frac{1}{\sqrt{N_0}}\sum_{t\in\mathcal{I}_0'}\zeta_t x_t^\top \left(X_0X_0^\top\right)^{-1/2}}_{\Delta_2} \\
&\quad + \underbrace{\frac{1}{\sqrt{N_0}}\sum_{t\in\mathcal{I}_0\cap B} w_t x_t^\top \left(X_0X_0^\top\right)^{-1/2}}_{\Delta_3},
\end{aligned}
$$

where remember that $\mathcal{I}_0' = \mathcal{I}_0 \setminus B$. We will control each $\|\Delta_i\|_{\mathrm{F}}$ separately.

Claim 18 *On $\mathcal{E}_g$, $\|\Delta_3\|_F^2 \le \widetilde{O}(dL/N_0)$.*

Claim 19 *On $\mathcal{E}_g$, $\|\Delta_2\|_F^2 \le O(\log(1/\alpha)+1)$.*

Claim 20 *On $\mathcal{E}_g$, with probability at least $1-\delta$, we have $\|\Delta_1\|_F^2 \le \widetilde{O}(d^2/N_0)$.*

Given these claims, we have that on $\mathcal{E}_g$, with probability at least $1-\delta$,

$$
\left\|(A_0 - A_*)\cdot\left(\frac{X_0X_0^T}{N_0}\right)^{1/2}\right\|_{\mathrm{F}}^2 \le \widetilde{O}(1)\cdot\frac{d^2+dL}{N_0} + O(\log(1/\alpha)+1)
$$

But, on $\mathcal{E}_g$, $N_0 \ge \alpha\beta T/4$ and $\frac{X_0X_0^\top}{N_0} \succcurlyeq \frac{1}{\mathrm{poly}\left(\frac{1}{\alpha\beta}\right)}\Gamma_T$. Setting $\delta = 1/T$, and using the assumed lower bound for T, we have that with probability $1-o(1)$, $\|A_0 - A_*\|_{\Gamma_T} \le O(\sqrt{\log(1/\alpha)}+1)$. Moreover, in Appendix H.4, we show that $\|\Gamma_T^{-1/2}x_t\|$ is $O(d)$-subgaussian. Combining with Lemma 16, we complete the proof of Theorem 11. We now prove the three claims.

G.1. Proof of Claim 18

Let $B' = \mathcal{I}_0 \cap B$. We will show that on $\mathcal{E}_g$, $\|\Delta_3\|_{\mathrm{F}}^2 \le \frac{1}{N_0}\sum_{t\in B'}\|w_t\|^2$. Observe that from Lemma 16, this immediately gives the claim.

$$
\begin{aligned}
\|\Delta_3\|_{\mathrm{F}}^2 &= \frac{1}{N_0}\mathrm{tr}\left(W_{B'}X_{B'}^\top(X_0X_0^\top)^{-1}X_{B'}W_{B'}^\top\right) \\
&= \frac{1}{N_0}\left\langle W_{B'}^\top W_{B'}, X_{B'}^\top(X_0X_0^\top)^{-1}X_{B'}\right\rangle \\
&= \frac{1}{N_0}\left\langle W_{B'}^\top W_{B'}, X_{B'}^\top P_{X_{B'}}(X_0X_0^\top)^{-1}P_{X_{B'}}X_{B'}\right\rangle
\end{aligned}
$$

where $P_{X_{B'}}$ is the projection matrix for the columnspace of $X_{B'}$. Now, note that $X_0X_0^T \succcurlyeq X_{B'}X_{B'}^T$. This implies that

$$
P_{X_{B'}}\left(X_0X_0^\top\right)^{-1}P_{X_{B'}} \preccurlyeq (X_{B'}X_{B'}^\top)^+,
$$

where $(X_{B'}X_{B'}^T)^+$ denotes the pseudoinverse of $X_{B'}X_{B'}^T$. For a proof of this implication, see Arora et al. (2019) Lemma E.1. Thus,

$$\|\Delta_3\|_{\mathrm{F}}^2 \leq \frac{1}{N_0}\Big\langle W_{B'}^\top W_{B'}, X_{B'}^\top (X_{B'}X_{B'}^\top)^+ X_{B'}\Big\rangle.$$

But, $X_{B'}^T(X_{B'}X_{B'}^\top)^+X_{B'}$ is just the projection matrix for the rowspace of $X_{B'}$, and so $\|\Delta_3\|_{\mathrm{F}}^2 \leq \frac{1}{N_0}\|W_{B'}\|_{\mathrm{F}}^2$, and we are done. ■

G.2. Proof of Claim 19

First we show that on $\mathcal{E}_g$, if $t \in \mathcal{I}_0'$, then $\|\zeta_t\| \leq \sqrt{2\log(1/\alpha)} + 1$. Consider the filtration $\mathcal{G}_t = \sigma(w_0, \ldots, w_{t-1}, \mathcal{S}_1, \ldots, \mathcal{S}_{t+1})$, and (similarly to the time-series case) the events $O_t = \{x_t \in \mathcal{S}_t\}$, and

$$G_t \coloneqq \{O_t \to \{\mathcal{N}(A_* x_t, I; \mathcal{S}_{t+1}) \geq \alpha\}\}$$

Notice that $t \in \mathcal{I}_0'$ implies $O_t \wedge G_t$. Now, given $\mathcal{G}_t$ and $O_t \wedge G_t$, we have $x_{t+1} \sim \mathcal{N}(A_* x_t, I, \mathcal{S}_{t+1})$, and $\mathcal{N}(A_* x_t, I; \mathcal{S}_{t+1}) \geq \alpha$. Using Claim 1, we get $\|\zeta_t\| = \|\nu_t^* - A_* x_t\| \leq \sqrt{2\log(1/\alpha)} + 1$. Now,

$$\|\Delta_2\|_{\mathrm{F}}^2 = \frac{1}{N_0}\Big\langle Z_{\mathcal{I}_0'}^\top Z_{\mathcal{I}_0'}, X_{\mathcal{I}_0'}^\top (X_0 X_0^\top)^{-1} X_{\mathcal{I}_0'}\Big\rangle,$$

and since $X_{\mathcal{I}_0'}X_{\mathcal{I}_0'}^\top \preccurlyeq (X_0X_0^\top)$, the exact same arguments as in proof of Claim 18 show that $\|\Delta_3\|_{\mathrm{F}}^2 \leq \frac{1}{N_0}\|Z_{\mathcal{I}_0'}\|_{\mathrm{F}}^2 \leq O(\log(1/\alpha) + 1)$. ■

G.3. Proof of Claim 20

First, remember that on $\mathcal{E}_g$, $X_0X_0^\top \succcurlyeq X_{\mathcal{I}_0'}X_{\mathcal{I}_0'}^\top \succ 0$. So,

$$\|\Delta_1\|_{\mathrm{F}} \leq \sqrt{d/N_0} \cdot \left\| \left(W_{\mathcal{I}_0'} - Z_{\mathcal{I}_0'}\right) X_{\mathcal{I}_0'}^\top \left(X_0 X_0^\top\right)^{-1/2} \right\|_2 \tag{33}$$

$$\leq \sqrt{d/N_0} \cdot \left\| \left(W_{\mathcal{I}_0'} - Z_{\mathcal{I}_0'}\right) X_{\mathcal{I}_0'}^\top \left(X_{\mathcal{I}_0'} X_{\mathcal{I}_0'}^\top\right)^{-1/2} \right\|_2 \tag{34}$$

Now,

$$\begin{aligned}
\left\| \left(W_{\mathcal{I}_0'} - Z_{\mathcal{I}_0'}\right) X_{\mathcal{I}_0'}^\top \left(X_{\mathcal{I}_0'} X_{\mathcal{I}_0'}^\top\right)^{-1/2} \right\|_2 &= \sup_{v_1, v_2 \in \mathcal{S}^{d-1}} v_1^T \left(W_{\mathcal{I}_0'} - Z_{\mathcal{I}_0'}\right) X_{\mathcal{I}_0'}^\top \left(X_{\mathcal{I}_0'} X_{\mathcal{I}_0'}^\top\right)^{-1/2} v_2 \\
&= \sup_{v_1, v_2 \in \mathcal{S}^{d-1}} \frac{v_1^T \left(W_{\mathcal{I}_0'} - Z_{\mathcal{I}_0'}\right) X_{\mathcal{I}_0'}^\top v_2}{\left\| \left(X_{\mathcal{I}_0'} X_{\mathcal{I}_0'}^\top\right)^{1/2} v_2 \right\|_2} \\
&= \sup_{v_1, v_2 \in \mathcal{S}^{d-1}} \frac{v_1^T \left(W_{\mathcal{I}_0'} - Z_{\mathcal{I}_0'}\right) X_{\mathcal{I}_0'}^\top v_2}{\left\| X_{\mathcal{I}_0'}^\top v_2 \right\|_2} \\
&= \sup_{v_1 \in \mathcal{S}^{d-1}, v_2 \in \mathbb{R}_*^d} \frac{v_1^T \left(W_{\mathcal{I}_0'} - Z_{\mathcal{I}_0'}\right) X_{\mathcal{I}_0'}^\top v_2}{\left\| X_{\mathcal{I}_0'}^\top v_2 \right\|_2}.
\end{aligned}$$

Fix some $v_1 \in \mathcal{S}^{d-1}, v_2 \in \mathbb{R}^d_*$. Let $H_t := v_1^T(w_t - \zeta_t) \cdot \mathbb{1}\{t \in I_0'\}$, and $U_t := v_2^T x_t \cdot \mathbb{1}\{t \in I_0'\}$, and observe that on $\mathcal{E}_g$,

$$\frac{v_1^T \left(W_{\mathcal{I}_0'} - Z_{\mathcal{I}_0'}\right) X_{\mathcal{I}_0'}^\top v_2}{\left\|X_{\mathcal{I}_0'}^\top v_2\right\|_2} = \frac{\sum_{t=1}^T H_t \cdot U_t}{\sqrt{\sum_{t=1}^T U_t^2}}$$

Let $\Gamma_{\min} := \frac{1}{C_{\alpha,\beta}}\Gamma_T$, $\Gamma_{\max} := C_{\alpha,\beta}\frac{d}{\delta}\Gamma_T$, and $C_{\alpha,\beta} = \mathrm{poly}(1/\alpha, 1/\beta)$ that is large enough, so that $\Gamma_{\min} \preccurlyeq X_{\mathcal{I}_0'} X_{\mathcal{I}_0'}^T \preccurlyeq \Gamma_{\max}$ on $\mathcal{E}_g$. Fix a $\delta \in (0,1)$, and consider a large enough $K = \widetilde{\Theta}(d)$. We will bound the probability

$$p_1 := \mathbb{P}\left[\left\{\frac{\sum_{t=1}^T H_t \cdot U_t}{\sqrt{\sum_{t=1}^T U_t^2}} \geq K\right\} \wedge \left\{v_2^\top \Gamma_{\min} v_2 \leq \sum_{t=1}^T U_t^2 \leq v_2^\top \Gamma_{\max} v_2\right\}\right].$$

Consider the filtration $\mathcal{F}_t := \sigma(w_0, w_1, \ldots, w_{t-1}, \mathcal{S}_1, \ldots, \mathcal{S}_{t+1}, \mathbb{1}\{x_{t+1} \in \mathcal{S}_{t+1}\})$ (similarly to the time-series case). Given $\mathcal{F}_t$ and that $t \in \mathcal{I}_0'$, we have $A_* x_t + w_t = x_{t+1} \sim \mathcal{N}(A_* x_t, I, \mathcal{S}_{t+1})$, and $\mathcal{N}(A_* x_t, I; \mathcal{S}_{t+1}) \geq \alpha$. Combining with Claim 22 (Appendix H), we get $\|v_1^T w_t\|_{\psi_2}^2 \leq O(1 \vee \log(1/\alpha))$. Also, from Claim 1, $\|v_1^T \zeta_t\|_{\psi_2}^2 \leq \|\zeta_t\|_2^2 \leq O(1 \vee \log(1/\alpha))$. Thus, given $\mathcal{F}_t$, H_t is mean zero, and $O(1 \vee \log(1/\alpha))$-subgaussian. Furthermore, observe that H_t is $\mathcal{F}_{t+1}$-measurable, while U_t is $\mathcal{F}_t$-measurable. Using Lemma 4.2 from Simchowitz et al. (2018), we get

$$p_1 \leq \log\left(\frac{v_2^\top \Gamma_{\min} v_2}{v_2^\top \Gamma_{\max} v_2}\right) \cdot \exp\left(-\frac{K^2}{O(1 \vee \log(1/\alpha))}\right).$$

Also, observe that

$$\mathbb{P}\left[\frac{v_1^\top \left(W_{\mathcal{I}_0'} - Z_{\mathcal{I}_0'}\right) X_{\mathcal{I}_0'}^\top v_2}{\left\|X_{\mathcal{I}_0'}^\top v_2\right\|_2} \cdot \mathbb{1}\{\mathcal{E}_g\} > K\right] \leq p_1$$

Now, let $\mathcal{N}_1$ be a $1/2$-net of $\mathcal{S}^{d-1}$, and let $\mathcal{N}_2$ be a $1/4$-net of $\mathcal{S}_{\Gamma_{\min}}$ in the norm $\|\Gamma_{\max}^{1/2}(\cdot)\|_2$. Taking a union bound,

$$\mathbb{P}\left[\exists\, v_1 \in \mathcal{N}_1, v_2 \in \mathcal{N}_2 : \frac{v_1^\top \left(W_{\mathcal{I}_0'} - Z_{\mathcal{I}_0'}\right) X_{\mathcal{I}_0'}^\top v_2}{\left\|X_{\mathcal{I}_0'}^\top v_2\right\|_2} \cdot \mathbb{1}\{\mathcal{E}_g\} > K\right] \leq p_1 \cdot |\mathcal{N}_1| \cdot |\mathcal{N}_2|$$

Using the bounds on $|\mathcal{N}_1|, |\mathcal{N}_2|$ from Simchowitz et al. (2018) (Lemma D.1), we get

$$p_1 \cdot |\mathcal{N}_1| \cdot |\mathcal{N}_2| \leq \log\left(\frac{v_2^\top \Gamma_{\max} v_2}{v_2^\top \Gamma_{\min} v_2}\right) \cdot \exp\left(O(d) + 2\log\det(\Gamma_{\max}\Gamma_{\min}^{-1}) - \frac{K^2}{O(1 \vee \log(1/\alpha))}\right) \leq \delta,$$

where we used the definitions of $\Gamma_{\max}, \Gamma_{\min}$, and K. So now we have

$$\mathbb{P}\left[\left\{\sup_{v_1 \in \mathcal{N}_1, v_2 \in \mathcal{N}_2} \frac{v_1^\top \left(W_{\mathcal{I}_0'} - Z_{\mathcal{I}_0'}\right) X_{\mathcal{I}_0'}^\top v_2}{\left\|X_{\mathcal{I}_0'}^\top v_2\right\|_2} > K\right\} \wedge \mathcal{E}_g\right] \leq \delta.$$

With an argument identical to the one of Simchowitz et al. (2018) (Appendix D.2), we get that

$$\mathbb{P}\left[\left\{\sup_{v_1\in\mathcal{S}^{d-1},v_2\in\mathbb{R}^d_*}\frac{v_1^\top\left(W_{\mathcal{I}_0'}-Z_{\mathcal{I}_0'}\right)X_{\mathcal{I}_0'}^\top v_2}{\left\|X_{\mathcal{I}_0'}^\top v_2\right\|_2}>4K\right\}\wedge\mathcal{E}_g\right]\leq\delta,$$

and we are done. ■

G.4. Proof of Lemma 16

We will deduce Lemma 16 as a corollary of the following lemma.

Lemma 17 *Fix $\delta,\epsilon\in(0,1)$, and an integer $k\geq 1$. Then, there exist absolute constants $C,C'>0$, such that if*

$$\frac{T}{k}\geq C\left(\frac{d}{\epsilon^2}\log\left(\frac{d}{\delta\epsilon}\right)+\frac{1}{\epsilon^2}\log\det(\Gamma_T\Gamma_k^{-1})\right),\quad \textit{then} \tag{35}$$

$$\mathbb{P}\left[\left\{\frac{1}{T}XX^\top\preccurlyeq\frac{d}{\delta}\Gamma_T\right\}\wedge\left\{\forall\,\mathcal{I}\subseteq[T]\ \textit{s.t.}\ |\mathcal{I}|\geq\epsilon T,\ \textit{we have}\ \frac{1}{|\mathcal{I}|}X_{\mathcal{I}}X_{\mathcal{I}}^\top\succcurlyeq\frac{\epsilon^4}{C'}\Gamma_k\right\}\right]\geq 1-\delta \tag{36}$$

Given the above, we prove Lemma 16.

Proof We will use Lemma 17, with $\epsilon=\alpha\beta/8$, and $k=\widetilde{\Theta}\left(\frac{1}{1-\rho(A_*)}\right)$. We need the following proposition.

Proposition 18 *For large enough $k=\widetilde{\Theta}\left(\frac{1}{1-\rho(A_*)}\right)$, we have $2\Gamma_k\succcurlyeq\Gamma_T$.*

Proof $\|\Gamma_T-\Gamma_k\|_2\leq\sum_{s=k}^{T-1}\left\|(A_*^s)(A_*^s)^\top\right\|_2$. Also, observe that $\left\|(A_*^s)(A_*^s)^\top\right\|_2\leq\mathrm{cond}(S)^2\rho^{2k}(A_*)$. Summing up and using the scale of k, we get $\|\Gamma_T-\Gamma_k\|_2<1$. The fact that $\Gamma_k\succcurlyeq I$ completes the proof. ■

Observe that the above proposition, and the assumed lower bound on T imply that 35 is satisfied. Now, from Proposition (4), $\mathbb{P}[|\mathcal{I}_0|,|\mathcal{I}_1|\geq\alpha\beta T/4]\geq 1-o(1)$. Also, from Assumption 3, $\mathbb{P}[|\mathcal{B}(\alpha)|\leq L]\geq 1-o(1)$. Thus, $\mathbb{P}[|\mathcal{I}_0|,|\mathcal{I}_1|,|\mathcal{I}_0'|\geq\epsilon T]\geq 1-o(1)$ (remember that $L<<T$). Lemma 17 and Gaussian-norm concentration finish the proof. ■

The main part of this section is devoted to proving the following lemma, from which we can easily deduce Lemma 17.

Lemma 19 *Fix $\epsilon\in(0,1)$, $1\leq k\leq\sqrt{\epsilon}T$, and $v\in\mathcal{S}^{d-1}$. Let $c(\epsilon)=\frac{\sqrt{8/\pi}}{\epsilon}$, and $k(\epsilon)=\lfloor\frac{\epsilon k}{2}\rfloor$. Then,*

$$\mathbb{P}\left[\exists\,\mathcal{I}\subseteq[T]:\ |\mathcal{I}|\geq 5\sqrt{\epsilon}T\ \textit{and}\ \frac{1}{|\mathcal{I}|}\sum_{t\in\mathcal{I}}(v^\top x_t)^2<\frac{v^\top\Gamma_{k(\epsilon)}v}{2c^2(\epsilon)}\right]\leq\exp\left(-\frac{\lfloor T/k\rfloor}{8}\right).$$

Given Lemma 19, we can show the following corollary, from which Lemma 17 immediately follows.

Corollary 20 *Fix $\epsilon \in (0,1)$, $1 \le k \le \sqrt{\epsilon}T$, and let $c(\epsilon), k(\epsilon)$ as in Lemma 19. Then,*

$$\mathbb{P}\left[\left\{\frac{1}{T}XX^\top \preccurlyeq \frac{d}{\delta}\Gamma_T\right\} \wedge \left\{\forall\, \mathcal{I} \subseteq [T] \text{ s.t. } |\mathcal{I}| \ge 5\sqrt{\epsilon}T \text{ we have } \frac{1}{|\mathcal{I}|}X_{\mathcal{I}}X_{\mathcal{I}}^\top \succcurlyeq \Gamma_{min}\right\}\right]$$
$$\ge 1 - \delta - \exp\left(d\log 9 + \log\det(\Gamma_{max}\Gamma_{min}^{-1}) - \frac{\lfloor T/k \rfloor}{8}\right),$$

where $\Gamma_{min} = \frac{1}{4c^2(\epsilon)}\Gamma_{k(\epsilon)}$, $\Gamma_{max} = \frac{d}{5\delta\sqrt{\epsilon}}\Gamma_T$.

Proof Using that for any $\mathcal{I} \subseteq [T]$ such that $|\mathcal{I}| \ge 5\sqrt{\epsilon}T$, we have $\frac{1}{|\mathcal{I}|}X_{\mathcal{I}}X_{\mathcal{I}}^\top \preccurlyeq \frac{1}{5\sqrt{\epsilon}}\frac{1}{T}XX^\top$, the corollary follows from Lemma 19 by using exactly the same (net-based) arguments as Simchowitz et al. (2018) (Lemma D.1., and beginning of Section 4). ■

For the rest of the section, we work on proving Lemma 19. We use the martingale small-ball technique of Simchowitz et al. (2018).

Definition 21 (Martingale Small-Ball; Simchowitz et al. (2018)) *Let $\mathcal{G}_t$ be a filtration, and let $(z_t)_{t\ge1}$ be a $\{\mathcal{G}_t\}_{t\ge1}$-adapted random process taking values in $\mathbb{R}$. We say $(z_t)_{t\ge1}$ satisfies the (k,ν,p)-block martingale small-ball (BMSB) condition if, for any $j \ge 0$, one has $\frac{1}{k}\sum_{i=1}^k \mathbb{P}(|z_{j+i}| \ge \nu|\mathcal{G}_j) \ge p$ almost surely. Given a process $(x_t)_{t\ge1}$ taking values in $\mathbb{R}^d$, we say that it satisfies the (k,Γ_{sb},p)-BMSB condition for $\Gamma_{sb} \succ 0$ if, for any fixed $v \in \mathcal{S}^{d-1}$, the process $z_t := \langle v, x_t\rangle$ satisfies $(k, \sqrt{v^\top\Gamma_{sb}v}, p)$-BMSB.*

Now, suppose that a (real) process $(z_t)_t$ satisfies the $(k,\nu,1-\epsilon)$-BMSM condition. We partition the sequence $z_1,\dots,z_T$ into $\tau := \lfloor T/k \rfloor$ blocks of size k (we discard the remaining terms). For $0 \le \ell \le \tau - 1$, we consider the events

$$\mathcal{E}_\ell := \left\{\frac{1}{k}\sum_{j=1}^k \mathbb{1}\left\{|Z_{\ell k+j}| \ge \nu\right\} \ge 1 - \sqrt{\epsilon}\right\}$$

Claim 21 $\mathbb{P}\left[\sum_{\ell=0}^{\tau-1}\mathbb{1}\{\mathcal{E}_\ell\} \le (1-\sqrt{\epsilon}/2)\tau\right] \le \exp(-\tau/8)$.

We provide a proof of Claim 21 in Appendix H (Claim 27).

APPLICATION TO LDSS

Proposition 22 *Consider the linear dynamical system $x_{t+1} = A_*x_t + w_t$, where $x_0 = 0$, $w_t \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,I)$, and let $\Gamma_t := \sum_{s=0}^{t-1}(A_*^s)(A_*^s)^\top$. Fix an $\epsilon \in (0,1)$. Then, for $1 \le k \le T$, the process $(x_t)_{t\ge1}$ satisfies the*

$$\left(k, \frac{1}{c^2(\epsilon)}\Gamma_{k(\epsilon)}, 1-\epsilon\right) \text{-block martingale small-ball condition},$$

where $c(\epsilon) = \frac{\sqrt{8/\pi}}{\epsilon}$ and $k(\epsilon) = \lfloor \frac{\epsilon k}{2} \rfloor$.

Proof Consider the filtration $\mathcal{G}_t := \sigma(w_0, w_1, \ldots, w_{t-1})$. Fix a $v \in \mathcal{S}^{d-1}$. We have

$$v^\top x_{s+j} = v^\top A_*^j x_s + v^\top \sum_{i=0}^{j-1} A_*^i w_{s+j-i-1}.$$

Given $\mathcal{G}_s$, $v^\top \sum_{i=0}^{j-1} A_*^i w_{s+j-i-1} \sim \mathcal{N}(0, v^\top \Gamma_j v)$. From Claim 26 (Appendix H), for any $c > 0$,

$$\mathbb{P}\left[|v^\top x_{s+j}| < \frac{v^\top \Gamma_j v}{c} \,\middle|\, \mathcal{G}_s\right] \leq \frac{\sqrt{2/\pi}}{c}.$$

We choose $c = \frac{2\sqrt{2/\pi}}{\epsilon}$. Using that for all $j \geq k(\epsilon)$, $\Gamma_j \succcurlyeq \Gamma_{k(\epsilon)}$, we get

$$\begin{aligned}
\frac{1}{k}\sum_{j=0}^{k} \mathbb{P}\left[|v^\top x_{s+j}| \geq \frac{\sqrt{v^\top \Gamma_{k(\epsilon)} v}}{c} \,\middle|\, \mathcal{G}_s\right] &\geq \frac{1}{k}\sum_{j=k(\epsilon)}^{k} \mathbb{P}\left[|v^\top x_{s+j}| \geq \frac{\sqrt{v^\top \Gamma_{k(\epsilon)} v}}{c} \,\middle|\, \mathcal{G}_s\right] \\
&\geq \frac{1}{k}\sum_{j=k(\epsilon)}^{k} \mathbb{P}\left[|v^\top x_{s+j}| \geq \frac{\sqrt{v^\top \Gamma_j v}}{c} \,\middle|\, \mathcal{G}_s\right] \\
&\geq \frac{k-k(\epsilon)}{k}\left(1 - \frac{\sqrt{2/\pi}}{c}\right) \geq (1-\epsilon/2)^2 \geq 1-\epsilon.
\end{aligned}$$

■

Now, fix a $v \in \mathcal{S}^{d-1}$, a $k \in [T]$, and an $\epsilon \in (0,1)$. Let $\sigma^2(v,\epsilon) := \frac{1}{c^2(\epsilon)} v^\top \Gamma_{k(\epsilon)} v$. We define the events:

$$\mathcal{E}_\ell := \left\{\frac{1}{k}\sum_{j=1}^{k} \mathbb{1}\left\{(v^\top x_{\ell k+j})^2 \geq \sigma^2(v,\epsilon)p\right\} \geq 1-\sqrt{\epsilon}\right\},$$

for all $0 \leq \ell \leq \tau - 1$. Observe that Claim 21 and Proposition 22 imply

$$\mathbb{P}\left[\sum_{\ell=0}^{\tau-1} \mathbb{1}\left\{\mathcal{E}_\ell\right\} \leq (1-\sqrt{\epsilon}/2)\tau\right] \leq \exp\left(-\frac{\tau}{8}\right),$$

which implies

$$\mathbb{P}\left[\sum_{t=1}^{T} \mathbb{1}\left\{(v^\top x_t)^2 \geq \sigma^2(v,\epsilon)\right\} \leq (1-\sqrt{\epsilon}/2)(1-\sqrt{\epsilon})k\tau\right] \leq \exp\left(-\frac{\tau}{8}\right).$$

For $k \leq \lfloor \sqrt{\epsilon} T \rfloor$, we have $T - \tau k \leq \sqrt{\epsilon} T$, and so

$$\mathbb{P}\left[\sum_{t=1}^{T} \mathbb{1}\left\{(v^\top x_t)^2 \geq \sigma^2(v,\epsilon)\right\} \leq (1-5\sqrt{\epsilon}/2)T\right] \leq \exp\left(-\frac{\tau}{8}\right).$$

Observe that this implies

$$\mathbb{P}\left[\exists\, \mathcal{I} \subseteq [T] :\ |\mathcal{I}| \geq 5\sqrt{\epsilon}T \ \text{ and } \ \frac{1}{|\mathcal{I}|}\sum_{t\in\mathcal{I}} (v^\top x_t)^2 < \frac{\sigma^2(v,\epsilon)}{2}\right] \leq \exp\left(-\frac{\tau}{8}\right). \ ■$$

Appendix H. Technical Claims

Claim 22 *Let $z \sim \mathcal{N}(\mu, I, S)$, and $\gamma = \mathcal{N}(\mu, I; S)$. Then, $\|z-\mu\|_{\psi_2}^2 \leq O(1 \vee \log(1/\gamma))$.*

Proof It suffices to show that for some absolute constant $c > 0$, and for all $v \in \mathcal{S}^{d-1}$,

$$\mathbb{P}\left[|v^\top z - v^\top \mu| \geq t\right] \leq \exp\left(-\frac{t^2}{c \cdot \left(1 \vee \log\left(\frac{1}{\gamma}\right)\right)}\right) \tag{37}$$

Observe that it suffices to show it for $v = (1, 0, \cdots, 0)$ (otherwise we can just change coordinates). So, it suffices to prove

$$\mathbb{P}\left[|z_1 - \mu_1| \geq t\right] \leq \exp\left(-\frac{t^2}{c \cdot \left(1 \vee \log\left(\frac{1}{\gamma}\right)\right)}\right)$$

Now, consider samples $\xi_1, \ldots, \xi_{\lceil \frac{1}{\gamma} \rceil} \overset{\text{i.i.d}}{\sim} \mathcal{N}(\mu, I)$. From van Handel (2014),

$$\forall x > 0,\ \mathbb{P}\left[\max_i |\xi_{i,1} - \mu_1| \geq \sqrt{2\log\frac{1}{\gamma}} + x\right] \leq 2e^{-x^2/2}.$$

Also, note that $\mathbb{P}\left[\exists i : \xi_i \in S\right] \geq 1 - (1-\gamma)^{\lceil \frac{1}{\gamma} \rceil} > 1/2$. Let

$$p \coloneqq \mathbb{P}\left[|z_1 - \mu_1| \geq t\right] = \mathbb{P}\left[|z_1 - \mu_1| \geq \sqrt{2\log\frac{1}{\gamma}} + t - \sqrt{2\log\frac{1}{\gamma}}\right].$$

For $t \geq \sqrt{2\log\frac{1}{\gamma}}$, let $x = t - \sqrt{2\log\frac{1}{\gamma}} \geq 0$, and so $p = \mathbb{P}\left[|z_1 - \mu_1| \geq \sqrt{2\log\frac{1}{\gamma}} + x\right]$, and

$$\begin{aligned}
\mathbb{P}\left[\max_i |\xi_{i,1} - \mu_1| \geq \sqrt{2\log\frac{1}{\gamma}} + x\right] &= 1 - \prod_i \mathbb{P}\left[|\xi_{i,1} - \mu_1| < \sqrt{2\log\frac{1}{\gamma}} + x\right] \\
&= 1 - \prod_i \left(1 - \mathbb{P}\left[|\xi_{i,1} - \mu_1| \geq \sqrt{2\ln\frac{1}{\gamma}} + x\right]\right) \\
&\geq 1 - \prod_i \left(1 - \mathbb{P}\left[\xi_i \in S\right] \cdot \mathbb{P}\left[|\xi_{i,1} - \mu_1| \geq \sqrt{2\ln\frac{1}{\gamma}} + x \;\middle|\; \xi_i \in S\right]\right)
\end{aligned}$$

Now, $\mathbb{P}\left[\xi_i \in S\right] = \gamma$, and given $\xi_i \in S$, we have $\xi_i \sim \mathcal{N}(\mu, I, S)$, and so observe that

$$\mathbb{P}\left[|\xi_{i,1} - \mu_1| \geq \sqrt{2\ln\frac{1}{\gamma}} + x \;\middle|\; \xi_i \in S\right] = p, \text{ Thus,}$$

Thus,

$$2e^{-x^2/2} \geq \mathbb{P}\left[\max_i |\xi_{i,1} - \mu_1| \geq \sqrt{2\ln\frac{1}{\gamma}} + x\right] \geq 1 - \prod_i (1 - \gamma p) = 1 - (1-\gamma p)^{\lceil \frac{1}{\gamma} \rceil} \geq 1 - e^{-p}$$

Now, observe that due to $e^{-\theta}$ being convex in $[0,1]$, we have $e^{-\theta} \leq 1-\theta(1-1/e)$, which allows us to conclude that $p \leq 4e^{-x^2/2}$. Let $c(\gamma) = \sqrt{2\log(1/\gamma)}$. We showed that for $t \geq c(\gamma)$,

$$\mathbb{P}\left[|z_1 - \mu_1| \geq t\right] \leq 2 \cdot \exp\left(-\frac{(t-c(\gamma))^2}{2} + \log 2\right). \tag{38}$$

From here, straightforward calculations complete the proof. ■

Claim 23 *Let X be a random vector in $\mathbb{R}^d$, such that $\|X\|_{\psi_2}^2 \leq \sigma^2$. Then, $\|X\|_2$ is $O(\sigma^2 d \log(2d))$-subgaussian.*

Proof For all i, $\mathbb{P}\left[x_i^2 \geq t^2\right] \leq 2 \cdot \exp\left(-\Omega(t^2/\sigma^2)\right)$. Letting $t = \Theta(\sigma) \cdot \sqrt{\log\left(\frac{2d}{\delta}\right)}$ and applying union-bound finish the proof. ■

Claim 24 *Let M be the maximum of T nonnegative σ^2-subgaussian random variables (not necessarily independent). Then, for all $q \geq 1$,*

$$\mathbb{E}[M^q] \leq \left(\sqrt{2\sigma^2 \log T} + C\sigma\sqrt{q}\right)^q,$$

where $C > 0$ is an absolute constant.

Proof Let $B := \sqrt{2\sigma^2 \log T}$. From van Handel (2014), $\mathbb{P}[M \geq B + x] \leq e^{-x^2/(2\sigma^2)}$. Let $Y := M \vee B - B \geq 0$. For all $x > 0$, $\mathbb{P}[Y \geq x] \leq e^{-x^2/(2\sigma^2)}$. This implies that for all $p \geq 1$, $\|Y\|_p \leq O(\sigma\sqrt{p})$ (Vershynin (2018)). Thus, $\|M\|_p \leq \|M \vee B\|_p \leq B + O(\sigma)\sqrt{p}$. ■

Claim 25 $\mathbb{P}[\neg C_t\,] \leq 1/T^2$.

Proof Given $\mathcal{F}_t$ and O_t, let $i = i(t)$, and consider p, k, and the ξ_j's of the Test at iteration i. Then $\mathbb{E}[p \mid \mathcal{F}_t, O_t] = \gamma_t$. We consider two cases, and we use standard Chernoff-bounds in both of them.

1. If $\gamma_t > 4\gamma$, then $\mathbb{P}[p < 2\gamma] \leq \exp(-k\gamma/2) \leq 1/T^2$.
2. If $\gamma_t < \gamma$, then $\mathbb{P}[p \geq 2\gamma] \leq \exp\left(-\frac{(2\gamma/\gamma_t)^2}{2+2\beta/\gamma_t} \cdot \gamma_t \cdot k\right) \leq \exp(-k\gamma) \leq 1/T^2$.

■

Claim 26 *Let $z \sim \mathcal{N}(\mu, \sigma^2)$ be a one-dimensional normal. Fix a $c > 0$. Then, $\mathbb{P}[|z| < \sigma/c] \leq \frac{\sqrt{2/\pi}}{c}$.*

Proof $\mathbb{P}[|z| \leq \sigma/c] = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\sigma/c}^{\sigma/c} e^{-(x-\mu^2)/(2\sigma^2)}dx \leq \frac{1}{\sqrt{2\pi}\sigma}\int_{-\sigma/c}^{\sigma/c} dx \leq \frac{\sqrt{2/\pi}}{c}$. ■

Claim 27 $\mathbb{P}\left[\sum_{\ell=0}^{\tau-1} \mathbb{1}\{\mathcal{E}_\ell\} \leq (1-\sqrt{\epsilon}/2)\tau\right] \leq \exp(-\tau/8)$.

Proof Let $\mathcal{H}_t = \sigma(z_1, \ldots, z_t)$. Fix a block ℓ. From $(k, \nu, 1-\epsilon)$-BMSB, $\frac{1}{k}\sum_{j=1}^{k} \mathbb{P}[|Z_{\ell k+j}| \geq \nu \mid \mathcal{H}_{\ell k}] \geq 1-\epsilon$, i.e.,

$$\frac{1}{k}\mathbb{E}\left[\sum_{j=1}^{k} \mathbb{1}\left\{|Z_{\ell k+j}| \geq \nu\right\} \;\middle|\; \mathcal{H}_{\ell k}\right] \geq 1-\epsilon \tag{39}$$

We will need the following simple proposition.

Proposition 23 *v Let Z be a random variable supported on $[0,1]$, such that $\mathbb{E}[Z] \geq 1-\epsilon$, for some $\epsilon \in (0,1)$. Then, we have $\mathbb{P}[Z \geq 1-\sqrt{\epsilon}] \geq 1-\sqrt{\epsilon}$.*

Proof Suppose it is not true, then

$$\mathbb{E}[Z] = \int_0^1 \mathbb{P}[Z \geq x]\,dx = \int_0^{1-\sqrt{\epsilon}} \mathbb{P}[Z \geq x]\,dx + \int_{1-\sqrt{\epsilon}}^{1} \mathbb{P}[Z \geq x]\,dx < 1-\sqrt{\epsilon} + \sqrt{\epsilon}\cdot(1-\sqrt{\varepsilon})$$
$$= 1-\epsilon,$$

contradiction. ■

Combining (39) with the proposition above,

$$\mathbb{P}[\mathcal{E}_\ell \mid \mathcal{F}_{\ell k}] = \mathbb{P}\left[\frac{1}{k}\sum_{j=1}^{k} \mathbb{1}\left\{|Z_{\ell k+j}| \geq \nu\right\} \geq 1-\sqrt{\epsilon} \;\middle|\; \mathcal{H}_{\ell k}\right] \geq 1-\sqrt{\epsilon}. \tag{40}$$

Now, let $\Delta_\ell = \mathbb{1}(\mathcal{E}_\ell) - \mathbb{P}[\mathcal{E}_\ell \mid \mathcal{R}_\ell]$, where $\mathcal{R}_\ell := \mathcal{H}_{\ell k}$, so that $\mathbb{E}[\Delta_\ell \mid \mathcal{R}_\ell] = 0$, and Δ_ℓ is $\mathcal{H}_\ell$-measurable. Since $-1 \leq \Delta_\ell \leq 1$, from Azuma-Hoeffding inequality,

$$\mathbb{P}\left[\sum_{\ell=0}^{\tau-1} \Delta_\ell \leq -t\right] \leq \exp\left(-\frac{t^2}{2\tau}\right)$$

Setting $t = \tau(1-\sqrt{\epsilon}/2)$, and combining with (40) completes the proof. ■

H.1. Proof of Claim 2

We will prove the following claim, which combined with Claim 29 (Appendix H.2) gives Claim 2.

Claim 28 *If $\mathcal{N}(\mu, I; S) \geq \zeta$, then $Cov_{z\sim\mathcal{N}(\mu,I,S)}[z,z] \succcurlyeq \frac{\zeta^3}{C} I$, where $C > 0$ is an absolute constant.*

Proof The proof is implicitly in Daskalakis et al. (2018). Let $R = \mathrm{Cov}_{z\sim\mathcal{N}(\mu,I,S)}[z,z]$ and

$$R' = \mathbb{E}_{z\sim\mathcal{N}(0,I)}\left[\left(z - \mathbb{E}_{z\sim\mathcal{N}(\mu,I,S)}[z]\right)\left(z - \mathbb{E}_{z\sim\mathcal{N}(\mu,I,S)}[z]\right)^T\right]$$

First, from Claim 3 in Daskalakis et al. (2018), $R' \succcurlyeq I$. Fix some $v \in S^{d-1}$. We want to relate $v^T R v$ with $v^T R' v$. Now, $v^T R v = \mathbb{E}_{z\sim\mathcal{N}(\mu,I,S)}[p_v(z)]$, and $v^T R' v = \mathbb{E}_{z\sim\mathcal{N}(\mu,I)}[p_v(z)]$, where

$p_v(z)$ is a polynomial of degree at most two, whose coefficients depend on v. Also, $p_v(z) \geq 0$, and $R' \succcurlyeq I$ implies

$$\mathbb{E}_{z\sim\mathcal{N}(\mu,I)}[p_v(z)] \geq 1$$

Let $\theta := \left(\frac{\zeta}{4C_*}\right)^2$, where C_* is the constant from Theorem 5 in Daskalakis et al. (2018). Let $\overline{S} = \{z \in \mathbb{R}^d : \ p_v(z) \leq \theta\}$. Applying the aforementioned theorem,

$$\mathcal{N}(\mu, I, \overline{S}) \leq \frac{C_* \cdot 2 \cdot \frac{\zeta}{4C_*}}{\left(\mathbb{E}_{z\sim\mathcal{N}(\mu,I)}[p_v(z)]\right)^{1/2}} \leq \zeta/2.$$

Thus, $\mathbb{E}_{z\sim\mathcal{N}(0,I,S)}[p_v(z)] \geq \zeta/2 \cdot \theta \geq \Omega(\zeta^3)$. ■

H.2. Proof of Claim 3

We prove a more general claim:

Claim 29 *If $\mathcal{N}(\mu^*, I; S) \geq \alpha$ and $\|\mu - \mu^*\| \leq r$, then $\mathcal{N}(\mu, I; S) \geq \frac{\alpha}{2} \cdot \exp\left(-\frac{r^2}{2} - r \cdot \sqrt{2\log\frac{1}{\alpha}}\right)$.*

Observe that if $r = c \cdot s(\alpha)$, we get Claim 3.

Proof Without loss of generality, suppose $\mu^* = 0$ (the general case follows from a simple affine transformation).

$$\mathcal{N}(\mu, I; S) = \mathbb{E}_{x\sim\mathcal{N}(0,I)}\left[\mathbb{1}(x \in S) \cdot \frac{\mathcal{N}(\mu, I; x)}{\mathcal{N}(0, I; x)}\right] \tag{41}$$

We will show that if $x \sim \mathcal{N}(0, I)$ then with probability at least $1 - \frac{\alpha}{2}$ the ratio $\frac{\mathcal{N}(\mu,I;x)}{\mathcal{N}(0,I;x)}$ is larger than some bound κ. Since $\mathcal{N}(0, I; S) \geq \alpha$, this implies from (41) that $\mathcal{N}(\mu, I; S) \geq \frac{\alpha}{2} \cdot \kappa$.

$$\frac{\mathcal{N}(\mu, I; x)}{\mathcal{N}(0, I; x)} = \exp\left(-\frac{\|\mu\|^2}{2} + x^\top\mu\right) \geq \exp\left(-r^2/2\right) \cdot \exp\left(x^\top\mu\right) \tag{42}$$

Observe now that since $x \sim \mathcal{N}(0, I)$, we have $x^\top\mu \sim \mathcal{N}(0, \|\mu\|^2)$, so for all $t \geq 0$,

$$\mathbb{P}\left[x^\top\mu \leq -\|\mu\|t\right] \leq \exp\left(-t^2/2\right)$$

Letting $t = \sqrt{2}\sqrt{\log\frac{2}{2\alpha}}$, we get $\mathbb{P}\left[x^\top\mu \leq -\|\mu\|t\right] \leq \frac{\alpha}{2}$, and we are done. ■

H.3. Bound on R_x

Claim 30 *For all t, $\mathbb{E}[\|x_t\|^2] \leq O\left(\frac{d}{1-\rho(A_*)} \cdot cond(U)^2\right)$*

Proof By unrolling the system, $x_t = \sum_{k=0}^{t-1} A_*^k w_{t-k-1}$, and so $x_t \sim \mathcal{N}\left(0, \sum_{k=0}^{t-1} A_*^k \left(A_*^k\right)^\top\right)$, so $\mathbb{E}[\|x_t\|^2] \leq \mathrm{tr}(\Gamma_T)$. Since $A_* = UDU^{-1}$, we have $A_*^k \left(A_*^k\right)^\top = UD^kU^{-1}U^{-*}D^kU^*$, and so

$$\begin{aligned}
\mathrm{tr}\left(A_*^k \left(A_*^k\right)^\top\right) &= \mathrm{tr}\left(UD^kU^{-1}U^{-*}D^{k*}U^*\right) = \mathrm{tr}\left(U^*UD^kU^{-1}U^{-*}D^{k*}\right) \\
&\leq \|U^*U\|_2 \cdot \mathrm{tr}\left(D^kU^{-1}U^{-*}D^{k*}\right) = \|U^*U\|_2 \cdot \mathrm{tr}\left(D^{k*}D^kU^{-1}U^{-*}\right) \\
&\leq \|U^*U\|_2 \cdot \|U^{-*}U^{-1}\|_2 \cdot \mathrm{tr}\left(D^{k*}D^k\right) = \mathrm{cond}(U)^2 \cdot \mathrm{tr}\left(D^{k*}D^k\right) \\
&\leq \mathrm{cond}(U)^2 \cdot d \cdot \rho(A_*)^{2k}.
\end{aligned}$$

Summing over t completes the proof. ■

H.4. Bound on $\left\|\Gamma_T^{-1/2} x_t\right\|$.

Claim 31 *For all t,* $\left\|\Gamma_T^{-1/2} x_t\right\|$ *is* $O(d)$*-subgaussian.*

Proof By unrolling the system, $x_t = \sum_{k=0}^{t-1} A_*^k w_{t-k-1}$, and so $x_t \sim \mathcal{N}\left(0, \sum_{k=0}^{t-1} A_*^k \left(A_*^k\right)^\top\right)$. Since $\Gamma_T \succcurlyeq \sum_{k=0}^{t-1} A_*^k \left(A_*^k\right)^T$, $\|\Gamma_t^{-1/2} x_t\|$ is stochastically dominated by a $\|z\|$, where $z \sim \mathcal{N}(0, I)$. So, from concentration of Gaussian norm (Vershynin (2018)), $\left\|\Gamma_T^{-1/2} x_t\right\|$ is $O(d)$-subgaussian. ■