A Control theory

A.1 Proof of Prop. 2

1. When $\theta^\top = (A, B)$ is not stabilizable, there exists no linear control K such that the controlled process $x_{t+1} = Ax_t + BKx_t + \epsilon_{t+1}$ is stationary. Thus, the positiveness of Q and R implies $J(\theta) = \mathrm{Tr}(P(\theta)) = +\infty$. As a consequence, $\theta^\top \notin \mathcal{S}$.
2. The mapping $\theta \to \mathrm{Tr}(P(\theta))$ is continuous (see Lem. 1). Thus, $\mathcal{S}$ is compact as the intersection between a closed and a compact set.
3. The continuity of the mapping $\theta \to K(\theta)$ together with the compactness of $\mathcal{S}$ justifies the finite positive constants ρ and C. Moreover, since every $\theta \in \mathcal{S}$ are stabilizable pairs, $\rho < 1$.

A.2 Proof of Lem. 1

Let $\theta^\mathsf{T} = (A, B)$ where A and B are matrices of size $n \times n$ and $n \times d$ respectively. Let $\mathcal{R} : \mathbb{R}^{n+d,n} \times \mathbb{R}^{n,n} \to \mathbb{R}^{n,n}$ be the Riccati operator defined by:

$$\mathcal{R}(\theta, P) := Q - P + A^\mathsf{T} P A - A^\mathsf{T} P B (R + B^\mathsf{T} P B)^{-1} B^\mathsf{T} P A, \tag{12}$$

where Q, R are positive definite matrices. Then, the solution $P(\theta)$ of the Riccati equation of Thm. 1 is the solution of $\mathcal{R}(\theta, P) = 0$. While Prop. 2 guarantees that there exists a unique admissible solution as soon as $\theta \in \mathcal{S}$, addressing the regularity of the function $\theta \to P(\theta)$ requires the use of the implicit function theorem.

Theorem 2 (Implicit function theorem). *Let E and F be two banach spaces, let $\Omega \subset E \times F$ be an open subset. Let $f : \Omega \to F$ be a C^1-map and let (x_0, y_0) be a point of Ω such that $f(x_0, y_0) = 0$. We denote as $d_y f(x_0, y_0) : F \to F$ the differential of the function f with respect to the second argument at point (x_0, y_0). Assume that this linear transformation is bounded and invertible. Then, there exists*

1. *two open subsets U and V such that $(x_0, y_0) \in U \times V \subset \Omega$,*
2. *a function $g : U \to V$ such that $g(x) = y$ for all $(x, y) \in U \times V$.*

Moreover, g is C^1 and $dg(x) = -d_y f(x, g(x))^{-1} d_x f(x, g(x))$ for all $(x, y) \in U \times V$.

Since R is positive definite, the Riccati operator is clearly a C^1-map. Moreover, thanks to Thm. 1, to any $\theta \in \mathcal{S}$, there exists an admissible P such that $\mathcal{R}(\theta, P) = 0$. Thanks to Thm. 2, a sufficient condition for $\theta \to P(\theta)$ to be C^1 on $\mathcal{S}$ is that the linear map $d_P\mathcal{R}(\theta, P(\theta)) : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ is a bounded invertible transformation i.e.

- **Bounded.** There exists M such that, for any $P \in \mathbb{R}^{n\times n}$, $\|d_P\mathcal{R}(\theta, P(\theta))(P)\| \le M\|P\|$.
- **Invertible.** There exists a bounded linear operator $S : \mathbb{R}^{n\times n} \to R^{n\times n}$ such that $SP = I_{n,n}$ and $PS = I_{n,n}$.

Lemma 5. *Let $\theta^\mathsf{T} = (A, B)$ and $\mathcal{R}$ be the Riccati operator defined in equation* (12)*. Then, the differential of $\mathcal{R}$ w.r.t P taken in $(\theta, P(\theta))$ denoted as $d_P\mathcal{R}(\theta, P(\theta))$ is defined by:*

$$d_P\mathcal{R}(\theta, P(\theta))(\delta P) := A_c^T \delta P A_c - \delta P, \quad \textit{for any } \delta P \in \mathbb{R}^{n\times n},$$

where $A_c = A - B(R + B^\mathsf{T} P B)^{-1} B^\mathsf{T} P(\theta) A$.

Proof. The proof is straightforward using the standard composition/multiplication/inverse operations for the differential operator together with an appropriate rearranging. □

Clearly, $d_P\mathcal{R}(\theta, P(\theta))$ is a bounded linear map. Moreover, thanks to the Lyapunov theory, for any stable matrix $\|A_c\|_2 < 1$ and for any matrix Q, the Lyapunov equation $A_c^T X A_c - X = Q$ admits a unique solution. From Thm. 1, the optimal matrix $P(\theta)$ is such that the corresponding A_c is stable. This implies that $d_P\mathcal{R}(\theta, P(\theta))$ is an invertible operator, and $\theta \to P(\theta)$ is C^1 on $\mathcal{S}$.

Therefore, the differential of $\theta \to P(\theta)$ can be deduced from the implicit function theorem. After tedious yet standard operations, one gets that for any $\theta \in \mathcal{S}$ and direction $\delta\theta \in \mathbb{R}^{(n+d)\times n}$:

$$dJ(\theta)(\delta\theta) = \mathrm{Tr}(dP(\theta)(\delta\theta)) = \mathrm{Tr}(\nabla J(\theta)^\mathsf{T}\delta\theta),$$

where $\nabla J(\theta) \in \mathbb{R}^{(n+d)\times n}$ is the jacobian matrix of J in θ. For any $\delta\theta \in \mathbb{R}^{(n+d)\times n}$, one has:

$$\nabla J(\theta)^\mathsf{T}\delta\theta = A_c(\theta)^\mathsf{T}\nabla J(\theta)^\mathsf{T}\delta\theta A_c(\theta) + C(\theta,\delta\theta) + C(\theta,\delta\theta)^\mathsf{T}, \quad \text{where} \quad C(\theta,\delta\theta) = A_c(\theta)^\mathsf{T}P(\theta)\delta\theta^\mathsf{T}H(\theta). \tag{13}$$

Proposition 5. *For any $\theta \in \mathcal{S}$ and any positive definite matrix V, one has the following inequality for the weighted norm of the gradient of J:*

$$\|\nabla J(\theta)\|_V \leq \|A_c(\theta)\|_2^2\|\nabla J(\theta)\|_V + 2\|P(\theta)\|\|A_c(\theta)\|_2\|H(\theta)\|_V.$$

Proof. For any $\theta \in \mathcal{S}$ and any positive definite matrix $V \in \mathbb{R}^{(n+d)\times(n+d)}$. Applying (13) to $\delta\theta = V\nabla J(\theta)$ leads to:

$$\nabla J(\theta)^\mathsf{T}V\nabla J(\theta) = A_c(\theta)^\mathsf{T}\nabla J(\theta)^\mathsf{T}V\nabla J(\theta)A_c(\theta) + C(\theta, V\nabla J(\theta)) + C(\theta, V\nabla J(\theta))^\mathsf{T},$$

where $C(\theta, V\nabla J(\theta))^\mathsf{T} = \big(V^{1/2}H(\theta)\big)^\mathsf{T}V^{1/2}\nabla J(\theta)P(\theta)A_c(\theta)$. Let $\langle A, B\rangle = \mathrm{Tr}A^\mathsf{T}B$ be the Frobenius inner product, then taking the trace of the above equality, one gets:

$$\|\nabla J(\theta)\|_V^2 = \|\nabla J(\theta)A_c(\theta)\|_V^2 + 2\big\langle V^{1/2}H(\theta), V^{1/2}\nabla J(\theta)P(\theta)A_c(\theta)\big\rangle.$$

Using the Cauchy-Schwarz inequality and that the Frobenius norm is sub-multiplicative together with $\mathrm{Tr}(M_1M_2) \leq \|M_1\|_2\mathrm{Tr}(M_2)$ for any M_1, M_2 symmetric positive definite matrices, one obtains:

$$\|\nabla J(\theta)\|_V^2 \leq \|A_c(\theta)\|_2^2\|\nabla J(\theta)\|_V^2 + 2\|H(\theta)\|_V\|P(\theta)\|\|A_c(\theta)\|_2\|\nabla J(\theta)\|_V.$$

Finally, dividing by $\|\nabla J(\theta)\|_V$ provides the desired result. □

B Material

Theorem 3 (Azuma's inequality)**.** *Let $\{M_s\}_{s\geq 0}$ be a super-martingale such that $|M_s - M_{s-1}| \leq c_s$ almost surely. Then, for all $t > 0$ and all $\epsilon > 0$,*

$$\mathbb{P}\big(|M_t - M_0| \geq \epsilon\big) \leq 2\exp\Big(\frac{-\epsilon^2}{2\sum_{s=1}^t c_s^2}\Big).$$

Lemma 6 (Lemma. 8 from Abbasi-Yadkori and Szepesvári [1])**.** *Let K^{det} be the number of changes in the policy of Algorithm 1 due to the determinant trigger $\det(V_t) \geq 2\det(V_0)$. Then, on E, K^{det} is at most*

$$K^{den} \leq (n+d)\log_2(1 + TX^2(1+C^2)/\lambda).$$

Corollary 2. *Let K be the number of policy changes of Algorithm 1, K^{det} be defined as in Lem. 6 and $K^{len} = K - K^{det}$ be the number of policy changes due to the length trigger $t \geq t_0 + \tau$. Then, on E, K is at most*

$$K \leq K^{det} + K^{len} \leq (n+d)\log_2(1 + TX^2(1+C^2)/\lambda) + T/\tau.$$

Moreover, assuming that $T \geq \frac{\lambda}{X^2(1+C^2)}$, one gets $K \leq (n+d)\log_2(1 + TX^2(1+C^2)/\lambda)T/\tau$.

Lemma 7 (Chernoff bound for Gaussian r.v.)**.** *Let $X \sim \mathcal{N}(0,1)$. For any $0 < \delta < 1$, for any $t \geq 0$, then,*

$$\mathbb{P}(|X| \geq t) \leq 2\exp\big(-\frac{t^2}{2}\big).$$

Proof of Lem.2. Let $\delta' = \delta/8T$.

1. From Prop. 3, $\mathbb{P}\big(\|\widehat{\theta}_t - \theta_*\|_{V_t} \leq \beta_t(\delta')\big) \geq 1 - \delta'$. Hence,

$$\begin{aligned}
\mathbb{P}(\widehat{E}) &= \mathbb{P}\Big(\bigcap_{t=0}^{T}\big(\|\widehat{\theta}_t - \theta_*\|_{V_t} \leq \beta_t(\delta')\big)\Big) \\
&= 1 - \mathbb{P}\Big(\bigcup_{t=0}^{T}\big(\|\widehat{\theta}_t - \theta_*\|_{V_t} \geq \beta_t(\delta')\big)\Big) \\
&\geq 1 - \sum_{t=0}^{T}\mathbb{P}\big(\|\widehat{\theta}_t - \theta_*\|_{V_t} \geq \beta_t(\delta')\big) \\
&\geq 1 - T\delta' \geq 1 - \delta/8
\end{aligned}$$

2. From Lem. 7, let $\eta \sim \mathcal{D}^{\mathrm{TS}}$ then, for any $\epsilon > 0$, making use of the fact that $\|\eta\| \leq n\sqrt{n+d}\max_{i\leq n+d, j\leq n}|\eta_{i,j}|$,

$$\mathbb{P}\big(\|\eta\| \leq \epsilon\big) \geq \mathbb{P}\big(n\sqrt{n+d}\max_{i,j}|\eta_{i,j}| \leq \epsilon\big) \geq 1 - \prod_{i,j}\mathbb{P}\big(|\eta_{i,j}| \geq \frac{\epsilon}{n\sqrt{n+d}}\big) \geq 1 - n(n+d)\mathbb{P}_{X\sim\mathcal{N}(0,1)}\big(|X| \geq \frac{\epsilon}{n\sqrt{n+d}}\big).$$

   Hence,

$$\begin{aligned}
\mathbb{P}(\widetilde{E}) &= \mathbb{P}\Big(\bigcap_{t=0}^{T}\big(\|\widetilde{\theta}_t - \widehat{\theta}_t\|_{V_t} \leq \gamma_t(\delta')\big)\Big) = 1 - \mathbb{P}\Big(\bigcup_{t=0}^{T}\big(\|\widetilde{\theta}_t - \widehat{\theta}_t\|_{V_t} \geq \gamma_t(\delta')\big)\Big) \\
&\geq 1 - \sum_{t=0}^{T}\mathbb{P}\big(\|\widetilde{\theta}_t - \widehat{\theta}_t\|_{V_t} \geq \gamma_t(\delta')\big) \geq 1 - \sum_{t=0}^{T}\mathbb{P}\big(\|\eta\| \geq \gamma_t(\delta')/\beta_t(\delta')\big) \\
&\geq 1 - \sum_{t=0}^{T}\mathbb{P}\Big(\|\eta\| \geq n\sqrt{2(n+d)\log\big(2n(n+d)/\delta'\big)}\Big) \\
&\geq 1 - T\delta' \geq 1 - \delta/8.
\end{aligned}$$

3. Finally, a union bound argument ensures that $\mathbb{P}(\widehat{E} \cap \widetilde{E}) \geq 1 - \delta/4$.

Proof of Cor. 1. This result comes directly from Sec. 4.1. and App. D of Abbasi-Yadkori and Szepesvári [1]. The proof relies on the fact that, on $\widehat{E}$, because $\widetilde{\theta}_t$ is chosen within the confidence ellipsoid $\mathcal{E}_t^{\mathrm{RLS}}$, the number of time steps the true closed loop matrix $A_* + B_*K(\widetilde{\theta}_t)$ is unstable is small. Intuitively, the reason is that as soon as the true closed loop matrix is unstable, the state process explodes and the confidence ellipsoid is drastically changed. As the ellipsoid can only shrink over time, the state is well controlled expect for a small number of time steps.

Since the only difference is that, on $\widehat{E}\cap\widetilde{E}$, $\widetilde{\theta}_t \in \mathcal{E}_t^{\mathrm{TS}}$, the same argument applies and the same bound holds replacing β_t with γ_t. Therefore, there exists appropriate problem dependent constants X, X' such that $\mathbb{P}(\bar{E}|\widehat{E}\cap\widetilde{E}) \geq 1 - \delta/4$. Finally, a union bound argument ensures that $\mathbb{P}(\widehat{E} \cap \widetilde{E} \cap \bar{E}) \geq 1 - \delta/2$.

C Proof of Lem. 3

We prove here that, on E, the sampling $\widetilde{\theta} \sim \mathcal{R}_{\mathcal{S}}(\widehat{\theta}_t + \beta_t(\delta')V_t^{1/2})$ guarantees a fixed probability of sampling an optimistic parameter, i.e. which belongs to $\Theta_t^{\mathrm{opt}} := \{\theta \in \mathbb{R}^d \mid J(\theta) \leq J(\theta^\star)\}$. However, our result only holds for the 1−dimensional case as we deeply leverage on the geometry of the problem. Figure 2 synthesizes the properties of the optimal value function and the geometry of the problem w.r.t the probability of being optimistic.

1. First, we introduce a simpler subset of optimistic parameters which involves hyperplanes rather than complicated J level sets. Without loss of generality we assume that $A_* + B_*K_* = \rho_* \geq 0$ and introduce $H_* = \begin{pmatrix} 1 \\ K_* \end{pmatrix} \in \mathbb{R}^2$ so that $A_* + B_*K_* = \theta^\mathsf{T} H_*$. Let $\Theta^{lin,\mathrm{opt}} = \{\theta \in \mathbb{R}^d \mid |\theta^\mathsf{T} H_*| \leq \rho_*\}$. Intuitively, $\Theta^{lin,\mathrm{opt}}$ consists in the set of systems θ which are more stable under control K_*. The following proposition ensures those systems to be optimistic.

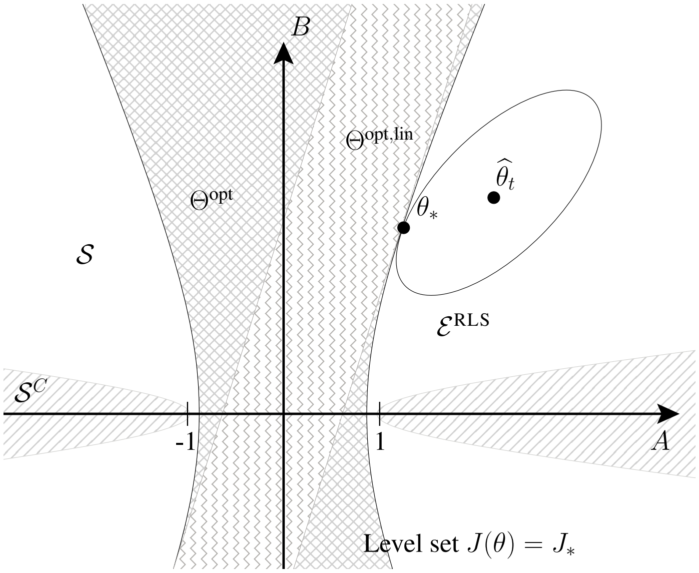

Figure 2: **Optimism and worst case configuration. 1)** In 1-D, the Riccati solution is well-defined expect for $\{(A, B) \in]-\infty, -1] \cup [1, \infty[\times \{0\}\}$. The rejection sampling procedure into $\mathcal{S}$ ensures $P(\widetilde{\theta}_t)$ to be well-defined. Moreover, $\mathcal{S}^c$ does not overlap with Θ^{opt}. **2)** The introduction of the subset $\Theta^{lin,\text{opt}}$ prevents using the actual - yet complicated - optimistic set Θ^{opt} to lower bound the probability of being optimistic. **3)** Even if the event $\mathcal{E}^{\text{RLS}}$ holds, there exists an ellipsoid configuration which does not contain any optimistic point. This justifies the over-sampling to guarantee a fixed probability of being optimistic.

Proposition 6. $\Theta^{lin,opt} \subset \Theta_t^{opt}$.

Proof. Leveraging on the expression of J, one has when $n = d = 1$,

$$J(\theta) = \text{Tr}(P(\theta)) = P(\theta) = \lim_{T \to \infty} \sum_{t=0}^{T} x_t^2 (Q + K(\theta)^2 R) = (Q + K(\theta)^2 R)\mathbb{V}(x_t),$$

where $\mathbb{V}(x_t) = (1 - |\theta^\mathsf{T} H(\theta)|^2)^{-1}$ is the steady-state variance of the stationary first order autoregressive process $x_{t+1} = \theta^\mathsf{T} H(\theta) x_t + \epsilon_{t+1}$ where ϵ_t is zero mean noise of variance 1 and $H(\theta) = \begin{pmatrix} 1 \\ K(\theta) \end{pmatrix}$. Thus,

$$J(\theta) = \big(Q + K(\theta)^2 R\big)\big(1 - |\theta^\mathsf{T} H(\theta)|^2\big)^{-1}.$$

Hence, for any $\theta \in \Theta^{lin,\text{opt}}$, $(1 - |\theta^\mathsf{T} H_*|^2)^{-1} \leq (1 - |\theta_*^\mathsf{T} H_*|^2)^{-1}$ which implies that

$$(Q + K_*^2 R)(1 - |\theta^\mathsf{T} H_*|^2)^{-1} \leq (Q + K_*^2 R)(1 - |\theta_*^\mathsf{T} H_*|^2)^{-1} = J(\theta_*).$$

However, since $K(\theta)$ is the optimal control associated with θ,

$$\begin{aligned} J(\theta) &= (Q + K(\theta)^2 R)(1 - |\theta^\mathsf{T} H(\theta)|^2)^{-1} \\ &= \min_K (Q + K^2 R)(1 - |\begin{pmatrix} 1 & K \end{pmatrix} \theta|^2)^{-1} \\ &\leq (Q + K_*^2 R)(1 - |\theta^\mathsf{T} H_*|^2)^{-1} \\ &\leq J(\theta_*) \end{aligned}$$

□

As a result, $\mathbb{P}\big(\widetilde{\theta}_t \in \Theta^{\text{opt}} \mid \mathcal{F}_t^x, \widehat{E}_t\big) \geq \mathbb{P}\big(\widetilde{\theta}_t \in \ \Theta^{lin,\text{opt}} \mid \mathcal{F}_t^x, \widehat{E}_t\big)$ and we can focus on $\Theta^{lin,\text{opt}}$.

2. To ensure the sampling parameter to be admissible, we perform a rejection sampling until $\widetilde{\theta}_t \in \mathcal{S}$. Noticing that $\Theta^{lin,\mathrm{opt}} \subset \Theta^{\mathrm{opt}} \subset \mathcal{S}$ by construction, the rejection sampling is always favorable in terms of probability of being optimistic. Since we seek for a lower bound, we can get rid of it and consider $\widetilde{\theta}_t = \widehat{\theta}_t + \beta_t(\delta') V_t^{-1/2} \eta$ where $\eta \sim \mathcal{N}(0, I_2)$.[9]

3. On $\widehat{E}_t$, $\theta_\star \in \mathcal{E}_t^{\mathrm{RLS}}$, where $\mathcal{E}_t^{\mathrm{RLS}}$ is the confidence RLS ellipsoid centered in $\widehat{\theta}_t$. Since θ_* is fixed (by definition), we lower bound the probability by considering the worst possible $\widehat{\theta}_t$ such that $\widehat{E}_t$ holds. Intuitively, we consider the worst possible center for the RLS ellipsoid such that $\theta_\star$ still belong in $\mathcal{E}_t^{\mathrm{RLS}}$ and that the probability of being optimistic is minimal. Formally,

$$\begin{aligned}\mathbb{P}\big(\widetilde{\theta}_t \in \Theta^{lin,\mathrm{opt}} \mid \mathcal{F}_t^x, \widehat{E}_t\big) &= \mathbb{P}_{\widetilde{\theta}_t \sim \mathcal{N}(\widehat{\theta}_t, \beta_t^2(\delta') V_t^{-1})}\big(\widetilde{\theta}_t \in \Theta^{lin,\mathrm{opt}} \mid \mathcal{F}_t^x, \widehat{E}_t\big) \\ &\geq \min_{\widehat{\theta} : \|\widehat{\theta} - \theta_*\|_{V_t} \leq \beta_t(\delta')} \mathbb{P}_{\widetilde{\theta}_t \sim \mathcal{N}(\widehat{\theta}, \beta_t^2(\delta') V_t^{-1})}\big(\widetilde{\theta}_t \in \Theta^{lin,\mathrm{opt}} \mid \mathcal{F}_t^x\big)\end{aligned}$$

Moreover, by Cauchy-Schwarz inequality, for any $\widehat{\theta}$,

$$\big|(\widehat{\theta} - \theta_*)^\mathsf{T} H_*\big| \leq \|\widehat{\theta} - \theta_*\|_{V_t} \|H_*\|_{V_t^{-1}} \leq \beta_t(\delta') \|H_*\|_{V_t^{-1}},$$

thus,

$$\begin{aligned}\mathbb{P}\big(\widetilde{\theta}_t \in \Theta^{lin,\mathrm{opt}} \mid \mathcal{F}_t^x, \widehat{E}_t\big) &\geq \min_{\widehat{\theta} : \|\widehat{\theta} - \theta_*\|_{V_t} \leq \beta_t(\delta')} \mathbb{P}_{\widetilde{\theta}_t \sim \mathcal{N}(\widehat{\theta}, \beta_t^2(\delta') V_t^{-1})}\big(\widetilde{\theta}_t \in \Theta^{lin,\mathrm{opt}} \mid \mathcal{F}_t^x\big) \\ &\geq \min_{\widehat{\theta} : |(\widehat{\theta} - \theta_*)^\mathsf{T} H_*| \leq \beta_t(\delta') \|H_*\|_{V_t^{-1}}} \mathbb{P}_{\widetilde{\theta}_t \sim \mathcal{N}(\widehat{\theta}, \beta_t^2(\delta') V_t^{-1})}\big(\widetilde{\theta}_t \in \Theta^{lin,\mathrm{opt}} \mid \mathcal{F}_t^x\big) \\ &= \min_{\widehat{\theta} : |\widehat{\theta}^\mathsf{T} H_* - \rho_*| \leq \beta_t(\delta') \|H_*\|_{V_t^{-1}}} \mathbb{P}_{\widetilde{\theta}_t \sim \mathcal{N}(\widehat{\theta}, \beta_t^2(\delta') V_t^{-1})}\big(|\widetilde{\theta}_t^\mathsf{T} H_*| \leq \rho_* \mid \mathcal{F}_t^x\big)\end{aligned} \tag{14}$$

Cor. 3 provides us with an explicit expression of the worst case ellipsoid. Introducing $x = \widetilde{\theta}_t^\mathsf{T} H_*$, one has $x \sim \mathcal{N}(\bar{x}, \sigma_x^2)$ with $\bar{x} = \widehat{\theta} H_*$ and $\sigma_x = \beta_t(\delta') \|H_*\|_{V_t^{-1}}$. Applying Cor. 3 with $\alpha = \rho_*$, $\rho = \rho_*$ and $\beta = \beta_t(\delta') \|H_*\|_{V_t^{-1}}$, inequality (14) becomes

$$\begin{aligned}\mathbb{P}\big(\widetilde{\theta}_t \in \Theta^{lin,\mathrm{opt}} \mid \mathcal{F}_t^x, \widehat{E}_t\big) &\geq \min_{\widehat{\theta} : |\widehat{\theta}^\mathsf{T} H_* - \rho_*| \leq \beta_t(\delta') \|H_*\|_{V_t^{-1}}} \mathbb{P}_{\eta \sim \mathcal{N}(0, I_2)}\big(|\widehat{\theta}^\mathsf{T} H_* + \beta_t(\delta') \eta^\mathsf{T} V_t^{-1/2} H_*| \leq \rho_* \mid \mathcal{F}_t^x\big) \\ &\geq \mathbb{P}_{\eta \sim \mathcal{N}(0, I_2)}\big(|\rho_* + \beta_t(\delta') \|H_*\|_{V_t^{-1}} + \beta_t(\delta') \eta^\mathsf{T} V_t^{-1/2} H_*| \leq \rho_* \mid \mathcal{F}_t^x\big)\end{aligned}$$

Introducing the vector $u_t = \beta_t(\delta') V_t^{-1/2} H_*$, one can simplify

$$\begin{aligned}&|\rho_* + \beta_t(\delta') \|H_*\|_{V_t^{-1}} + \beta_t(\delta') \eta^\mathsf{T} V_t^{-1/2} H_*| \leq \rho_*, \\ \Leftrightarrow &- \rho_* \leq \rho_* + \|u_t\| + \eta^\mathsf{T} u_t \leq \rho_*, \\ \Leftrightarrow &- \frac{\rho_*}{\|u_t\|} - 1 \leq \eta^\mathsf{T} \frac{u_t}{\|u_t\|} \leq -1.\end{aligned}$$

Since $\eta \sim \mathcal{N}(0, I_2)$ is rotationally invariant , $\mathbb{P}\big(\widetilde{\theta}_t \in \Theta^{lin,\mathrm{opt}} \mid \mathcal{F}_t^x, \widehat{E}_t\big) \geq \mathbb{P}_{\epsilon \sim \mathcal{N}(0,1)}\big(\epsilon \in \big[1, 1 + \frac{2\rho_*}{\|u_t\|}\big] \mid \mathcal{F}_t^x, \widehat{E}_t\big)$. Finally, for all $t \leq T$, u_t is almost surely bounded: $\|u_t\| \leq \beta_T(\delta') \sqrt{(1 + C^2)/\lambda}$. Therefore,

$$\mathbb{P}\big(\widetilde{\theta}_t \in \Theta^{lin,\mathrm{opt}} \mid \mathcal{F}_t^x, \widehat{E}_t\big) \geq \mathbb{P}_{\epsilon \sim \mathcal{N}(0,1)}\big(\epsilon \in \big[1, 1 + 2\rho_* / \beta_T(\delta') \sqrt{(1 + C^2)/\lambda}\big]\big) := p$$

Corollary 3. *For any $\rho, \sigma_x > 0$, for any $\alpha, \beta \geq 0$, $\arg\min_{\bar{x} : |\bar{x} - \alpha| \leq \beta} \mathbb{P}_{x \sim \mathcal{N}(\bar{x}, \sigma_x^2)}\big(|x| \leq \rho\big) = \alpha + \beta$.*

This corollary is a direct consequence of the properties of standard gaussian r.v.

Lemma 8. *Let x be a real random variable. For any $\rho, \sigma_x > 0$ Let $f : \mathbb{R} \to [0, 1]$ be the continuous mapping defined by $f(\bar{x}) = \mathbb{P}_{x \sim \mathcal{N}(\bar{x}, \sigma_x^2)}\big(|x| \leq \rho\big)$. Then, f is increasing on $\mathbb{R}_-$ and decreasing on $\mathbb{R}_+$.*

[9] In the 1-dimensional case, η is just a 2d standard gaussian r.v.

Proof. Without loss of generality, one can assume that $\sigma_x = 1/\sqrt{2}$ (otherwise, modify ρ), and that $\bar{x} \geq 0$ (by symmetry). Denoting as Φ and erf the standard gaussian cdf and the error function, one has:

$$\begin{aligned} f(\bar{x}) &= \mathbb{P}_{x\sim\mathcal{N}(\bar{x},\sigma_x^2)}\big(-\rho \leq x \leq \rho\big), = \mathbb{P}_{x\sim\mathcal{N}(\bar{x},\sigma_x^2)}\big(x \leq \rho\big) - \mathbb{P}_{x\sim\mathcal{N}(\bar{x},\sigma_x^2)}\big(x \leq -\rho\big), \\ &= \mathbb{P}_{x\sim\mathcal{N}(\bar{x},\sigma_x^2)}\big((x-\bar{x})/\sigma_x \leq (\rho-\bar{x})/\sigma_x\big) - \mathbb{P}_{x\sim\mathcal{N}(\bar{x},\sigma_x^2)}\big((x-\bar{x})/\sigma_x \leq (-\rho-\bar{x})/\sigma_x\big), \\ &= \Phi((\rho-\bar{x})/\sigma_x) - \Phi(-(\rho+\bar{x})/\sigma_x), \\ &= \frac{1}{2} + \frac{1}{2}\mathrm{erf}((\rho-\bar{x})/\sqrt{2}\sigma_x) - \frac{1}{2} - \frac{1}{2}\mathrm{erf}(-(\rho+\bar{x})/\sqrt{2}\sigma_x), \\ &= \frac{1}{2}\big(\mathrm{erf}(\rho-\bar{x}) - \mathrm{erf}(-(\rho+\bar{x}))\big). \end{aligned}$$

Since erf is odd, one obtains $f(\bar{x}) = \frac{1}{2}\big(\mathrm{erf}(\rho-\bar{x}) + \mathrm{erf}(\rho+\bar{x})\big)$. The error function is differentiable with $\mathrm{erf}'(z) = \frac{2}{\pi}e^{-z^2}$, thus

$$\begin{aligned} f'(\bar{x}) &= \frac{1}{\pi}\Big(\exp\big(-(\rho+\bar{x})^2\big) - \exp\big(-(\rho-\bar{x})^2\big)\Big) \\ &= -\frac{2}{\pi}\sinh\big((\rho-\bar{x})^2\big) \leq 0 \end{aligned}$$

Hence, f is decreasing on $\mathbb{R}_+$ and by symmetry, is increasing on $\mathbb{R}_-$. □

D Weighted L1 Poincaré inequality (proof of Lem. 4)

This result is build upon the following theorem which links the function to its gradient in L^1 norm:

Theorem 4 (see Acosta and Durán [4]). *Let $W^{1,1}(\Omega)$ be the Sobolev space on $\Omega \subset \mathbb{R}^d$. Let Ω be a convex domain bounded with diameter D and $f \in W^{1,1}(\Omega)$ of zero average on Ω then*

$$\int_\Omega |f(x)|dx \leq \frac{D}{2}\int_\Omega ||\nabla f(x)||dx \tag{15}$$

Lem. 4 is an extension of Thm. 4. In pratice, we show that their proof still holds for log-concave weight.

Theorem 5. *Let $L > 0$ and ρ any non negative and log-concave function on $[0, L]$. Then for any $f \in W^{1,1}(0, L)$ such that*

$$\int_0^L f(x)\rho(x)dx = 0$$

one has:

$$\int_0^L |f(x)|\rho(x)dx \leq 2L\int_0^L |f'(x)|\rho(x)dx \tag{16}$$

The proof is based on the following inequality for log-concave function.

Lemma 9. *Let ρ be any non negative log-concave function on $[0, 1]$ such that $\int_0^1 \rho(x) = 1$ then*

$$\forall x \in (0,1), \quad H(\rho, x) := \frac{1}{\rho(x)}\int_0^x \rho(t)dt\int_x^1 \rho(t)dt \leq 1 \tag{17}$$

Proof. Since any non-negative log-concave function on $[0, 1]$ can be rewritten as $\rho(x) = e^{\nu(x)}$ where ν is a concave function on $[0, 1]$ and since $x \to e^x$ is increasing, the monotonicity of ν is preserved and as for concave function, ρ can be either increasing, decreasing or increasing then decreasing on $[0, 1]$.
Hence, $\forall x \in (0, 1)$, either

1. $\rho(t) \leq \rho(x)$ for all $t \in [0, x]$,

2. $\rho(t) \leq \rho(x)$ for all $t \in [x, 1]$.

Assume that $\rho(t) \leq \rho(x)$ for all $t \in [0,x]$ without loss of generality. Then,

$$\begin{aligned}
\forall x \in (0,1), \quad H(\rho,x) :&= \frac{1}{\rho(x)} \int_0^x \rho(t)dt \int_x^1 \rho(t)dt \\
&= \int_0^x \frac{\rho(t)}{\rho(x)} \int_x^1 \rho(t)dt \\
&\leq \int_0^x dt \int_x^1 \rho(t)dt \\
&\leq x \int_0^1 \rho(t)dt \leq x \leq 1
\end{aligned}$$

□

Proof of theorem 5. This proof is exactly the same as [4] where we use lemma 9 instead of a concave inequality. We provide it for sake of completeness.

A scaling argument ensures that it is enough to prove it for $L = 1$. Moreover, dividing both side of (16) by $\int_0^1 \rho(x)dx$, we can assume without loss of generality that $\int_0^1 \rho(x)dx = 1$.
Since $\int_0^1 f(x)\rho(x)dx = 0$ by integration part by part one has:

$$\begin{aligned}
f(y) &= \int_0^y f'(x) \int_0^x \rho(t)dt - \int_y^1 f'(x) \int_x^1 \rho(t)dt \\
|f(y)| &\leq \int_0^y |f'(x)| \int_0^x \rho(t)dt + \int_y^1 |f'(x)| \int_x^1 \rho(t)dt
\end{aligned}$$

Multiplying by $\rho(y)$, integrating on y and applying Fubini's theorem leads to

$$\int_0^1 |f(y)|\rho(y)dy \leq 2 \int_0^1 |f'(x)| \int_0^x \rho(t)dt \int_x^1 \rho(t)dt$$

and applying (17) of lemma 9 ends the proof. □

While theorem 5 provides a 1 dimensional weigthed Poincaré inequality, we actually seek for one in $\mathbb{R}^d$. The idea of [4] is to use arguments of [12] to reduce the $d-$dimensional problem to a $1-d$ problem by splitting any convex set Ω into subspaces Ω_i thin in all but one direction and such that an average property is preserved. We just provide their result.

Lemma 10. *Let $\Omega \subset \mathbb{R}^d$ be a convex domain with finite diameter D and $u \in L^1(\Omega)$ such that $\int_\Omega u = 0$. Then, for any $\delta > 0$, there exists a decomposition of Ω into a finite number of convex domains Ω_i satisfying*

$$\Omega_i \cap \Omega_j = \emptyset \ \text{ for } \ i \neq j, \quad \bar{\Omega} = \bigcup \bar{\Omega}_i, \quad \int_{\Omega_i} u = 0$$

and each Ω_i is thin in all but one direction i.e. in an appropriate rectangular coordinate system $(x,y) = (x, y_1, \ldots, y_{d-1})$ the set Ω_i is contained in

$$\{(x,y): \ \ 0 \leq x \leq D, \quad 0 \leq y_i \leq \delta \ \text{ for } \ i = 1, \ldots, d-1\}$$

This decomposition together with theorem 5 allow us to prove the $d-$dimensional weighted Poincaré inequality.

Proof of Lem. 4. By density, we can assume that $u \in C^\infty(\bar{\Omega})$. Hence, $up \in C^2(\bar{\Omega})$. Let M be a bound for up and all its derivative up to the second order.
Given $\delta > 0$ decompose the set Ω into Ω_i as in lemma 10 and express $z \in \Omega_i$ into the appropriate rectangular basis $z = (x,y)$, where $x \in [0, d_i]$, $y \in [0, \delta]$. Define as $\rho(x_0)$ the $d-1$ volume of the intersection between Ω_i and the hyperplan $\{x = x_0\}$. Since Ω_i is convex, ρ is concave and from the smoothness of up one has:

$$\left| \int_{\Omega_i} |u(x,y)|p(x,y)dxdy - \int_0^{d_i} |u(x,0)|p(x,0)\rho(x)dx \right| \leq (d-1)M|\Omega_i|\delta \tag{18}$$

$$\left| \int_{\Omega_i} |\frac{\partial u}{\partial x}(x,y)|p(x,y)dxdy - \int_0^{d_i} |\frac{\partial u}{\partial x}(x,0)|p(x,0)\rho(x)dx \right| \leq (d-1)M|\Omega_i|\delta \tag{19}$$

$$\left| \int_{\Omega_i} u(x,y)p(x,y)dxdy - \int_0^{d_i} u(x,0)p(x,0)\rho(x)dx \right| \leq (d-1)M|\Omega_i|\delta \tag{20}$$

Those equation allows us to switch from $d-$dimensional integral to $1-$dimensional integral for which we can apply theorem 5 at the condition that $\int_0^{d_i} u(x,0)p(x,0)\rho(x)dx = 0$ (which is not satisfied here). On the other hand, we can apply theorem 5 to

$$g(x) = u(x,0) - \int_0^{d_i} u(x,0)p(x,0)\rho(x)dx / \int_0^{d_i} p(x,0)\rho(x)dx$$

with weigthed function $x \to p(x,0)\rho(x)$. Indeed, $x \to p(x,0)$ is log-concave - as restriction along one direction of log-concave function, $x \to \rho(x)$ is log-concave - as a concave function, and so is $x \to p(x,0)\rho(x)$ - as product of log-concave function. Moreover, $g \in W^{1,1}(0,d_i)$ and $\int_0^{d_i} g(x)p(x,0)\rho(x)dx = 0$ by construction. Therefore, applying theorem 5 one gets:

$$\begin{aligned}
&\int_0^{d_i} |g(x)|p(x,0)\rho(x)dx \leq 2d_i \int_0^{d_i} |g'(x)|p(x,0)\rho(x)dx \\
&\int_0^{d_i} |u(x,0)|p(x,0)\rho(x)dx \leq 2d_i \int_0^{d_i} |\frac{\partial u}{\partial x}(x,0)|p(x,0)\rho(x)dx - \left| \int_0^{d_i} u(x,0)p(x,0)\rho(x)dx \right| \\
&\int_0^{d_i} |u(x,0)|p(x,0)\rho(x)dx \leq 2d_i \int_0^{d_i} |\frac{\partial u}{\partial x}(x,0)|p(x,0)\rho(x)dx + (d-1)M|\Omega_i|\delta
\end{aligned} \tag{21}$$

where we use equation (20) together with $\int_{\Omega_i} u(z)p(z)dz = 0$ to obtain the last inequality.

Finally, from (18)

$$\int_{\Omega_i} |u(x,y)|p(x,y)dxdy \leq \int_0^{d_i} |u(x,0)|p(x,0)\rho(x)dx + (d-1)M|\Omega_i|\delta$$

from (21)

$$\int_{\Omega_i} |u(x,y)|p(x,y)dxdy \leq 2d_i \int_0^{d_i} |\frac{\partial u}{\partial x}(x,0)|p(x,0)\rho(x)dx + (d-1)M|\Omega_i|\delta(1+2d_i)$$

from (19)

$$\begin{aligned}
&\int_{\Omega_i} |u(x,y)|p(x,y)dxdy \leq 2d_i \int_{\Omega_i} |\frac{\partial u}{\partial x}(x,y)|p(x,y)dxdy + (d-1)M|\Omega_i|\delta(1+4d_i) \\
&\int_{\Omega_i} |u(x,y)|p(x,y)dxdy \leq 2d_i \int_{\Omega_i} ||\nabla u(x,y)||p(x,y)dxdy + (d-1)M|\Omega_i|\delta(1+4d_i)
\end{aligned}$$

Summing up on Ω_i leads to

$$\int_{\Omega} |u(z)|p(z)dz \leq 2D \int_{\Omega} ||\nabla u(z)||p(z)dz + (d-1)M|\Omega|\delta(1+4D)$$

and since δ is arbitrary one gets the desired result. □

E Regret proofs

Bounding $R_1^{\mathbf{RLS}}$. On E, $\|x_t\| \leq X$ for all $t \in [0,T]$. Moreover, since $\widetilde{\theta}_t \in \mathcal{S}$ for all $t \in [0,T]$ due to the rejection sampling, $\mathrm{Tr}(P(\widetilde{\theta}_t)) \leq D$. From the definition of the matrix 2-norm, $\sup_{\|x\|\leq X} x^\mathsf{T} P(\widetilde{\theta}_t)x \leq X^2 \|P(\widetilde{\theta}_t)^{1/2}\|_2^2$. Since for any $A \in \mathbb{R}^{m,n}$, $\|A\|_2 \leq \|A\|$, one has $\|P(\widetilde{\theta}_t)^{1/2}\|_2^2 \leq \|P(\widetilde{\theta}_t)^{1/2}\|^2 = \mathrm{Tr}P(\widetilde{\theta}_t)$. As a consequence, for any $t \in [0,T]$, $\sup_{\|x\|\leq X} x^\mathsf{T} P(\widetilde{\theta}_t)x \leq X^2 D$ and the martingale increments are bounded almost surely on E by $2DX^2$. Applying Thm. 3 to R_1^{RLS} with $\epsilon = 2DX^2\sqrt{2T\log(4/\delta)}$ one obtains that

$$R_1^{\mathrm{RLS}} = \sum_{t=0}^{T} \big\{\mathbb{E}(x_{t+1}^\mathsf{T} P(\widetilde{\theta}_{t+1})x_{t+1}|\mathcal{F}_t) - x_t^\mathsf{T} P(\widetilde{\theta}_t)x_t\big\}\mathbb{1}\{E_t\} \leq 2DX^2\sqrt{2T\log(4/\delta)}$$

with probability at least $1-\delta/2$.

Bounding $R_3^{\mathbf{RLS}}$. The derivation of this bound is directly collected from Abbasi-Yadkori and Szepesvári [1]. Since our framework slightly differs, we provide it for the sake of completeness. The whole derivation is performed conditioned on the event E.

$$\begin{aligned} R_3^{\mathrm{RLS}} &= \sum_{t=0}^{T} \big\{z_t^\mathsf{T} \widetilde{\theta}_t P(\widetilde{\theta}_t)\widetilde{\theta}_t^\mathsf{T} z_t - z_t^\top \theta_* P(\widetilde{\theta}_t)\theta_*^\mathsf{T} z_t\big\} = \sum_{t=0}^{T} \big\{\|\widetilde{\theta}_t^\mathsf{T} z_t\|^2_{P(\widetilde{\theta}_t)} - \|\theta_*^\mathsf{T} z_t\|^2_{P(\widetilde{\theta}_t)}\big\}, \\ &= \sum_{t=0}^{T} \big(\|\widetilde{\theta}_t^\mathsf{T} z_t\|_{P(\widetilde{\theta}_t)} - \|\theta_*^\mathsf{T} z_t\|_{P(\widetilde{\theta}_t)}\big)\big(\|\widetilde{\theta}_t^\mathsf{T} z_t\|_{P(\widetilde{\theta}_t)} + \|\theta_*^\mathsf{T} z_t\|_{P(\widetilde{\theta}_t)}\big) \end{aligned}$$

By the triangular inequality, $\|\widetilde{\theta}_t^\mathsf{T} z_t\|_{P(\widetilde{\theta}_t)} - \|\theta_*^\mathsf{T} z_t\|_{P(\widetilde{\theta}_t)} \leq \|P(\widetilde{\theta}_t)^{1/2}(\widetilde{\theta}_t^\mathsf{T} z_t - \theta_*^\mathsf{T} z_t)\| \leq \|P(\widetilde{\theta}_t)\|\|(\widetilde{\theta}_t^\mathsf{T} - \theta_*^\mathsf{T})z_t\|$. Making use of the fact that $\widetilde{\theta}_t \in \mathcal{S}$ by construction of the rejection sampling, $\theta_\star \in \mathcal{S}$ by Asm. 2 and that $\sup_{t\in[0,T]} \|z_t\| \leq \sqrt{(1+C^2)X^2}$ thanks to the conditioning on E and Prop. 2, one gets:

$$R_3^{\mathrm{RLS}} \leq \sum_{t=0}^{T} \big(\sqrt{D}\|(\widetilde{\theta}_t^\mathsf{T} - \theta_*^\mathsf{T})z_t\|\big)\big(2S\sqrt{D}\sqrt{(1+C^2)X^2}\big) \leq 2SD\sqrt{(1+C^2)X^2}\sum_{t=0}^{T}\|(\widetilde{\theta}_t^\mathsf{T} - \theta_*^\mathsf{T})z_t\|$$

and one just has to bound $\sum_{t=0}^{T}\|(\widetilde{\theta}_t^\mathsf{T} - \theta_*^\mathsf{T})z_t\|$. Let $\tau(t) \leq t$ be the last time step before t when the parameter was updated. Using Cauchy-Schwarz inequality, one has:

$$\sum_{t=0}^{T}\|(\widetilde{\theta}_t^\mathsf{T} - \theta_*^\mathsf{T})z_t\| = \sum_{t=0}^{T}\|(V_\tau^{1/2}(t)(\widetilde{\theta}_{\tau(t)} - \theta_*))^\mathsf{T} V_{\tau(t)}^{-1/2} z_t\| \leq \sum_{t=0}^{T}\|\widetilde{\theta}_{\tau(t)} - \theta_*\|_{V_{\tau(t)}}\|z_t\|_{V_{\tau(t)}^{-1}}$$

However, on E, $\|\widetilde{\theta}_{\tau(t)} - \theta_*\|_{V_{\tau(t)}} \leq \|\widetilde{\theta}_{\tau(t)} - \widehat{\theta}_{\tau(t)}\|_{V_{\tau(t)}} + \|\theta_* - \widehat{\theta}_{\tau(t)}\|_{V_{\tau(t)}} \leq \beta_{\tau(t)}(\delta') + \gamma_{\tau(t)}(\delta') \leq \beta_T(\delta') + \gamma_T(\delta')$ and, thanks to the lazy update rule $\|z_t\|_{V_{\tau(t)}^{-1}} \leq \|z_t\|_{V_t^{-1}} \frac{\det(V_t)}{\det(V_{\tau(t)})} \leq 2\|z_t\|_{V_t}$. Therefore,

$$R_3^{\mathrm{RLS}} \leq 4SD\sqrt{(1+C^2)X^2}\big(\beta_T(\delta') + \gamma_T(\delta')\big)\sum_{t=0}^{T}\|z_t\|_{V_t^{-1}}.$$

Bounding $\sum_{k=1}^{K} T_k\alpha_k$. From section 4.4,

$$\sum_{k=1}^{K} T_k\alpha_k \leq 2\tau\sum_{k\in\mathcal{K}^{den}}\|\alpha_k\| + \tau\sum_{k=1}^{K}\alpha_k.$$

First, it is clear from

$$\begin{aligned} \alpha_k &= (R_{t_k}^{\mathrm{TS},1} + R_{t_k}^{\mathrm{TS},3})\{E_{t_k}\} \\ &= \big(J(\widetilde{\theta}_{t_k}) - \mathbb{E}[J(\widetilde{\theta}_{t_k})|\mathcal{F}_{t_k}^x, E_{t_k}]\big)\mathbb{1}\{E_{t_k}\}, +(\mathbb{E}\Big[\Big\|\begin{pmatrix} I \\ K(\widetilde{\theta}_{t_k})^\top \end{pmatrix}\Big\|_{V_{t_k}^{-1}}|\mathcal{F}_{t_k}^x\Big] - \Big\|\begin{pmatrix} I \\ K(\widetilde{\theta}_{t_k})^\top \end{pmatrix}\Big\|_{V_{t_k}^{-1}}), \end{aligned}$$

that the sequence $\{\alpha_k\}_{k=1}^{K}$ is a martingale difference sequence with respect to $\mathcal{F}_{t_k}^x$. Moreover, since $\widetilde{\theta}_{t_k} \in \mathcal{S}$ for all $k \in [1,K]$, $\|\alpha_k\| \leq 2D + 2\sqrt{(1+C^2)/\lambda}$. Therefore,

1. $\sum_{k\in\mathcal{K}^{den}} \|\alpha_k\| \leq \big(2D + 2\sqrt{(1+C^2)}\big)|K^{den}|$,
2. with probability at least $1-\delta/2$, Azuma's inequality ensures that $\sum_{k=1}^{K} \alpha_k \leq \big(2D + 2\sqrt{(1+C^2)}\big)\sqrt{2|K|\log(4/\delta)}$.

From Lem. 6 and Cor. 2, $|K^{det}| \leq (n+d)\log_2(1+TX^2(1+C^2)/\lambda)$ and $|K| \leq (n+d)\log_2(1+TX^2(1+C^2)/\lambda)T/\tau$. Finally, one obtains:

$$\sum_{k=1}^{K} T_k\alpha_k \leq 4\big(2D + 2\sqrt{(1+C^2)}\big)(n+d)\log_2(1+TX^2(1+C^2)/\lambda)\sqrt{\log(4/\delta)}T/\tau$$

Bounding $\sum_{t=0}^{T}\|z_t\|_{V_t^{-1}}$. On E, for all $t\in[0,T]$, $\|z_t\|^2 \leq (1+C^2)X^2$. Thus, from Cauchy-Schwarz inequality and Prop. 4,

$$\sum_{t=0}^{T}\|z_t\|_{V_t^{-1}} \leq \sqrt{T}\Big(\sum_{t=0}^{T}\|z_t\|^2_{V_t^{-1}}\Big)^{1/2} \leq \sqrt{T}\sqrt{2(n+d)(1+C^2)X^2/\lambda}\log^{1/2}\Big(1+\frac{T(1+C^2)X^2}{\lambda(n+d)}\Big).$$