A Deterministic regret upper bound

In this section, we prove Corollary 2.3.1 in which we provide the explicit regret bounds for online ridge regression and forward regression in the adversarial case. First, recall the following result.

Theorem. *(Theorem 11.8 of Cesa-Bianchi & Lugosi [6]) For all $T \geq 1, (x_t)_{1\leq t\leq T} \in \mathbb{R}^d$, $(y_t)_{1\leq t\leq T} \in [-Y, Y]$ such that $\|x_t\|_2 \leq X$,*

$$\text{for } \mathcal{A} \in \{\boldsymbol{r}, \boldsymbol{f}\} \qquad \bar{R}_T^{\mathcal{A}} \leq c^{\mathcal{A}} (Y^{\mathcal{A}})^2 d \ln\left(1 + \frac{TX^2}{\lambda d}\right) + \lambda \|\theta_T\|_2^2,$$

where $c^{\boldsymbol{r}} = 4, c^{\boldsymbol{f}} = 1, Y^{\boldsymbol{r}} = \max\{Y, \max_{1\leq t\leq T} |x_t^\top \theta_{t-1}^{\boldsymbol{r}}|\}, Y^{\boldsymbol{f}} = Y$ and $\theta_T = \arg\min_\theta L_T(\theta)$.

We can now derive the explicit regret bound we seek by bounding the norm of the parameter θ_T.

Corollary. *(Corollary 2.3.1) For all $T \geq 1, (x_t)_{1\leq t\leq T} \in \mathbb{R}^d, (y_t)_{1\leq t\leq T} \in [-Y, Y]$ such that $\|x_t\|_2 \leq X$,*

$$\text{for } \mathcal{A} \in \{\boldsymbol{r}, \boldsymbol{f}\} \qquad \bar{R}_T^{\mathcal{A}} \leq c^{\mathcal{A}} (Y^{\mathcal{A}})^2 d \ln\left(1 + \frac{TX^2}{\lambda d}\right) + \frac{\lambda\, (Y^{\mathcal{A}})^2 T}{\lambda_{r_T}(G_T(0))}$$

where $r_T = \operatorname{rank}(G_t(0))$ and λ_{r_T} is its smallest positive eigenvalue, $c^{\boldsymbol{r}} = 4, c^{\boldsymbol{f}} = 1, Y^{\boldsymbol{r}} = \max\{Y, \max_{1\leq t\leq T} |x_t^\top \theta_{t-1}^{\boldsymbol{r}}|\}, Y^{\boldsymbol{f}} = Y$.

Proof. Consider (w.l.o.g) ridge regression, denote X_T the design matrix and y_T the labels, then:

$$\|\theta_T\|_2 = \left\|G_T(0)^\dagger \mathbf{b}_T\right\|_2 = \sqrt{y_T^\top X_T^\top G_T(0)^\dagger G_T(0)^\dagger X_T y_T} \leq \sqrt{\frac{y_T^\top X_T^\top G_T(0)^\dagger X_T y_T}{\lambda_{r_T}(G_T)}}$$
$$\leq Y^{\mathbf{r}} \sqrt{\frac{T}{\lambda_{r_T}(G_T)}},$$

where $G_T(0)^\dagger$ is the pseudo-inverse of $G_T(0)$, the last inequality is because $X_T^\top G_T(0)^\dagger X_T$ is an orthogonal projection on $\operatorname{Im}(X^\top)$. Injecting in the previous theorem finishes the proof, these bounds hold for arbitrary bounded sequences. The proof for the forward algorithm proceeds in the same way by replacing G_T by G_{T+1} and $Y^{\mathbf{r}}$ by $Y^{\mathbf{f}}$. ☐

B Regret definition

In this section, we prove that with high probability, $\bar{R}_T$ and R_T yield the same first order high probability bounds for online regression algorithms.

Theorem. *(Regret equivalence) For all $\delta > 0$, with probability at least $1 - \delta$, for $T > 0$ such that $\sum_{s=1}^t x_s x_s^\top$ is non-singular:*

$$R_T = \bar{R}_T + o(\log(T)^2)$$

Note that this is enough to prove that R_T and $\bar{R}_T$ are equal in first order because the upper bound on $\bar{R}_T$ is of order $\log(T)^2$.

Denote $\forall T \geq 1 : \theta_T = \arg\min_{\theta\in\mathbb{R}^d} L_T(\theta)$, then:

$$R_T - \bar{R}_T = L_T(\theta_*) - L_T(\theta_T) = 2\sum_{t=1}^T \epsilon_t (\theta_T - \theta_*)^\top x_t - \sum_{t=1}^T \left((\theta_T - \theta_*)^\top x_t\right)^2. \tag{8}$$

Denote $S_T = \sum_{t=1}^T \epsilon_t (\theta_T - \theta_*)^\top x_t, A_T = \sum_{t=1}^T \left((\theta_T - \theta_*)^\top x_t\right)^2$, we prove that $S_T = o(A_T)$.

Lemma B.1. *(Tail inequality) For all $\delta > 0, \sigma' > 0$, with probability at least $1 - \delta$, for all $T > 0$:*

$$|S_T| \leq \sqrt{2(A_T + 1/\sigma'^2) \log\left(\frac{\sqrt{\sigma'^2 A_T + 1}}{\delta}\right)}$$

Proof. We use the method of mixtures, denote

$$M_t^\lambda = \exp\left(\lambda\epsilon_t(\theta_T - \theta_*)^\top x_t - \frac{\lambda^2}{2}\left((\theta_T - \theta_*)^\top x_t\right)^2\right).$$

Without loss of generality, we can assume that $(\epsilon_s)_{s\geq 1}$ is 1-sub-Gaussian (this can be achieved by scaling features appropriately), then $\mathbb{E}[M_t^\lambda] \leq 1$.
Let $\Lambda \sim \mathbb{N}(0, \sigma'^2)$ be a Gaussian random variable and define $M_t = \mathbb{E}[M_t^\Lambda | F^\infty]$. We have $\mathbb{E}[M_t] = \mathbb{E}[\mathbb{E}[M_t^\Lambda|\Lambda]] \leq 1$. By making explicit M_t and using Markov's inequality we get that for any stopping time τ, for all $\delta > 0$, with probability at least $1 - \delta$:

$$\frac{|S_\tau|^2}{1/\sigma'^2 + A_\tau} \leq 2\sigma^2 \log\left(\frac{\sqrt{1 + \sigma'^2 A_\tau}}{\delta}\right).$$

We conclude using the same stopping time construction in Proof. C. □

From Lemma. B.1 and equation. (8) we get that for all $\sigma', \delta > 0$ with probability at least $1 - \delta$:

$$\begin{aligned} R_T - \bar{R}_T &\leq \sqrt{2(A_T + 1/\sigma'^2)\log\left(\frac{\sqrt{\sigma'^2 A_T + 1}}{\delta}\right)} - A_T \\ &\leq \sqrt{(A_T + 1/\sigma'^2)\left(\log(\sigma'^2 A_T + 1) + 2\log(1/\delta)\right)} - A_T \\ &\leq \sqrt{A_T + 1/\sigma'^2}\left(\sqrt{\log(\sigma'^2 A_T + 1)} + \sqrt{2\log(1/\delta)}\right) - A_T \\ &\leq \frac{1}{\sigma'^2} + \sqrt{2(A_T + 1/\sigma'^2)\log(1/\delta)} \end{aligned} \tag{9}$$

The next step is to the use confidence intervals of Maillard [15] which hold once the design matrix is singular.

Theorem. *(Theorem 3.3 of [15]) (Ordinary Least-squares) Assume that N is a stopping time adapted to the filtration of the past. Then in the sub-Gaussian streaming regression model, for any $\delta > 0$, with probability at least $1 - \delta, \forall T \geq 1$ if $|G_T(0)| > 0$:*

$$\|\theta_* - \theta_T\|^2_{G_T(0)} \leq 2(1+\kappa)(1+\alpha)\sigma^2 \log\frac{\kappa_d(e^2\lambda_{max}(G_T))}{\delta}$$

where $\kappa_d(x)$ is function of κ and α, $\kappa_d(x) = \frac{2}{3}\pi^2\log(x/e)^2\left\lceil\frac{\log(x)}{2}\right\rceil\left[(12(d+1)\sqrt{d})^d x^d + d\right]$ for $\kappa = \alpha = 1$.

For bounded features $\|x\| \leq X$, we bound $\lambda_{max}(G_T(0)) \leq TX^2$. Denote $T_0 = \inf_{t\geq 1}\{|G_t| > 0\}$, and for $t \geq T_0 : \beta_t = 2(1+\kappa)(1+\alpha)\sigma^2 \log\frac{\kappa_d(e^2\lambda_{max}(G_T))}{\delta}$, then for all $\delta > 0$ with probability at least $1 - \delta$:

$$\begin{aligned} A_T &= \sum_{t=1}^T \left((\theta_T - \theta_*)^\top x_t\right)^2 \leq A_{T_0} + \sum_{t=T_0}^T \left((\theta_T - \theta_*)^\top x_t\right)^2 \\ &\leq A_{T_0} + \sum_{t=T_0}^T \beta_t \|x_t\|^2_{G_t(0)^{-1}} \leq A_{T_0} + \beta_T \sum_{t=T_0}^T \|x_t\|^2_{G_t(0)^{-1}} \end{aligned} \tag{10}$$

Then we bound the sum of features.

Lemma B.2. *(Technical inequality) For all sequences $\{x_t\}_t \in \mathbb{R}^d$ such that $\forall t, \|x_t\|_2 \leq X$, for all $\lambda \in \mathbb{R}_+, T_0, T \in \mathbb{N}$*

$$\sum_{t=T_0}^T \|x_t\|^2_{G_t^{-1}} \leq d\log\left(1 + TX^2/\lambda_{min}(G_{T_0})d\right)$$

where $G_t = G_t(\lambda)$.

Proof. Using the Weinstein–Aronszajn identity: $\|x_t\|^2_{G_t^{-1}} = 1 - \frac{|G_{t-1}|}{|G_t|}$, and that $z - 1 \geq \log(z)$ leads to:

$$\sum_{t=T_0}^{T} \|x_t\|^2_{G_t^{-1}} \leq \sum_{t=1}^{T} - \log \frac{|G_{t-1}|}{|G_t|} = \log\left(\frac{|G_T|}{|G_{T_0}|}\right).$$

Since $\|x_t\|_2 \leq X$, using the AM-GM inequality:

$$\sum_{t=T_0}^{T} \log\left(1 + \|x_t\|^2_{G_{t-1}^{-1}}\right) \leq d \log\left(1 + TX^2/\lambda_{min}(G_{T_0})d\right).$$

□

From equation (9), using $A_T \leq \sum_{t=1}^{T} \|\theta_T - \theta_*\|^2_{G_t} \|x_t\|^2_{G_t^{-1}} \leq \sum_{t=1}^{T} \|\theta_T - \theta_*\|^2_{G_T} \|x_t\|^2_{G_t^{-1}}$, then injecting Lemma B.2 with $\lambda = 0$, we find that for all $\delta > 0$, with probability at least $1 - \delta$:

$$A_T \leq A_{T_0} + \beta_T d \log\left(1 + TX^2/\lambda_{min}(G_{T_0}(0))d\right)$$

Then injecting this last inequality in equation (8) gives, for all $\delta, \sigma' > 0$, with probability at least $1 - \delta$:

$$R_T - \bar{R}_T \leq \frac{1}{\sigma'^2} + \sigma' \sqrt{2 \log(1/\delta) \left(\beta_T d \log\left(1 + TX^2/\lambda_{min}(G_{T_0})d\right) + 1\right)}.$$

We also know -by definition- that $R_T \geq \bar{R}_T$. This concludes the proof for the equivalence of the two regret definitions.

C Ridge regression analysis

Here we prove a high probability time-uniform upper bound for online ridge regression. Let's recall the statement of the theorem that we prove.

Theorem. *(Theorem 3.2) For any $\delta > 0$, with probability at least $1 - \delta$, for all $T > 0$:*

$$\bar{R}_T^{\mathbf{r}} \leq (d\sigma)^2 \frac{X^2/\lambda}{\log(1 + X^2/\lambda)} \log\left(\frac{1 + TX^2/\lambda d}{\delta/2}\right) \log\left(1 + TX^2/\lambda d\right) + o(\log(T)^2)$$

See Eq. 14 for an explicit bound. In particular, the $o(\log(T)^2)$ term is $O(\log(T)^{3/2})$.

Let's write the instantaneous regret:

$$\bar{r}_t = \ell_t(\theta_{t-1}) - \ell_t(\theta_*) = (\theta_{t-1}^\top x_t - \theta_*^\top x_t)^2 + 2\epsilon_t(\theta_{t-1}^\top x_t - \theta_*^\top x_t) \tag{11}$$

The proof proceeds in three steps, that we detail hereafter and then we explain how to combine them for the final result.

First step: Confidence bound to control the concentration of θ_{t-1} around θ_*. For this we use the confidence ellipsoid from Abbasi-Yadkori et al. [1].

Theorem. *(Confidence ellipsoid for ridge regression) For any $\delta > 0$, with probability at least $1 - \delta$, for all $t > 0$:*

$$\|\theta_t^{\mathbf{r}} - \theta_*\|_{G_t} \leq \sqrt{\beta_t(\delta)} = \sigma \sqrt{d \log\left(\frac{1 + tX^2/\lambda d}{\delta}\right)} + \lambda^{1/2} S.$$

It comes, with probability at least $1 - \delta$:

$$(\theta_{t-1} - \theta_*)^\top x_t \leq ||x_t||_{G_{t-1}^{-1}} ||\theta_{t-1} - \theta_*||_{G_{t-1}} \leq \sqrt{\beta_{t-1}(\delta)} ||x_t||_{G_{t-1}^{-1}}.$$

Then, since β_t is non-decreasing:

$$L_t - L_t^* \leq \beta_{T-1} \sum_{t=1}^{T} \|x_t\|_{\eta_{t-1}}^2 + 2 \sum_{t=1}^{T} \epsilon_t (\theta_{t-1} - \theta_*)^\top x_t. \tag{12}$$

Second step: Next we bound the sum of feature norms. The main idea here is to use linear algebra techniques to obtain a telescopic sum.
Lemma B.2 doesn't apply here because we have $||x_t||_{G_{t-1}^{-1}}$ instead of $||x_t||_{G_t^{-1}}$. We derive a similar lemma for this sum of feature norms.

Lemma C.1. *(Technical inequality) For all sequences $\{x_t\}_t \in \mathbb{R}^d$ such that $\forall t, \|x_t\|_2 \leq X$, for all $\lambda \in \mathbb{R}_+, T \in \mathbb{N}$*

$$\sum_{t=1}^{T} ||x_t||_{G_{t-1}^{-1}}^2 \leq \frac{X^2/\lambda}{\log(1 + X^2/\lambda)} d \log\left(1 + TX^2/\lambda d\right)$$

Proof. We use the Weinstein–Aronszajn identity: $\|x_t\|_{G_{t-1}^{-1}}^2 = \frac{|G_t|}{|G_{t-1}|} - 1$, which leads to:

$$\sum_{t=1}^{T} \log\left(1 + \|x_t\|_{G_{t-1}^{-1}}^2\right) = \log\left(\frac{G_T}{G_0}\right).$$

Then since $\|x_t\|_2 \leq X$ and using the AM-GM inequality:

$$\sum_{t=1}^{T} \log\left(1 + \|x_t\|_{G_{t-1}^{-1}}^2\right) \leq d \log\left(1 + TX^2/\lambda d\right).$$

This next part is what differs from Lemma B.2, using $||x_t||_{G_{t-1}^{-1}}^2 \leq \lambda_{\max}(G_{t-1}^{-1})||x_t||_2^2 \leq X^2/\lambda$ and the concavity of the function $\log$ we find:

$$\sum_{t=1}^{T} \|x_t\|_{G_{t-1}^{-1}}^2 \leq \sum_{t=1}^{T} \frac{X^2/\lambda}{\log(1 + X^2/\lambda)} \log\left(1 + \|x_t\|_{G_{t-1}^{-1}}^2\right).$$

The last inequality can also be proved by noting that $x \to x/\log(1+x)$ is non-decreasing which can be used to bound every feature norm. $\square$

Third step: To control the second term in the r.h.s of Eq. 11, we use Martingale inequalities similar to the ones used for the confidence intervals to derive a uniform high probability bound.

Lemma C.2. *(Tail inequality, see Corollary 8 of [2]) Define $S_t = \sum_{s=1}^t \epsilon_s(\theta_{s-1} - \theta_*)^\top x_s$ and let $(F_t)_{t\geq 0}$ be a filtration such that x_t is F_{t-1} measurable and ϵ_t is F_t measurable. Then S_t is a martingale with respect to F_t and for any $\delta > 0, \sigma' > 0$, with probability at least $1 - \delta$, for all $t \geq 0$:*

$$|S_t| \leq \sigma \sqrt{2\left(1/\sigma'^2 + \sum_{s=1}^{t} \left((\theta_{t-1} - \theta_*)^\top x_t\right)^2\right) \log\left(\frac{\sqrt{1 + \sigma'^2 \sum_{s=1}^t \left((\theta_{t-1} - \theta_*)^\top x_t\right)^2}}{\delta}\right)}$$

Proof. The proof of this result follows the same line in the proof of Theorem 1 of Abbasi-Yadkori et al. [1], first we define for $\lambda \in \mathbb{R}^d, t > 0$: $M_t^\lambda = \exp\Bigg(\sum_{s=1}^t \Big[\epsilon_s \lambda(\theta_{t-1} - \theta_*)^\top x_t - \lambda^2 \left((\theta_{t-1} - \theta_*)^\top x_t\right)^2/2\Big]\Bigg).$

Without loss of generality, we can assume that $(\epsilon_s)_{s\geq 1}$ is 1-sub-Gaussian (this can be achieved by scaling features). Let τ be a stopping time with respect to the filtration $\{F_t\}_{t=0}^\infty$. Then M_τ^λ is well-defined almost surely and

$$\mathbb{E}[M_\tau^\lambda] \leq 1.$$

Let $\Lambda \sim \mathbb{N}(0, \sigma'^2)$ be a Gaussian random variable and define $M_t = \mathbb{E}[M_t^\Lambda | F^\infty]$. We have $\mathbb{E}[M_t] = \mathbb{E}[\mathbb{E}[M_t^\Lambda|\Lambda]] \leq 1$. By expliciting M_t and using Markov's inequality we get that for $\delta > 0$, with probability $1-\delta$:

$$|S_\tau|^2 \leq \left(1/\sigma'^2 + \sum_{t=1}^{\tau} \left((\theta_{t-1}-\theta_*)^\top x_t\right)^2\right) 2\sigma^2 \log\left(\frac{\sqrt{1+\sigma'^2 \sum_{t=1}^{\tau}\left((\theta_{t-1}-\theta_*)^\top x_t\right)^2}}{\delta}\right). \tag{13}$$

Next we use a stopping time construction from Freedman [10]: Define the bad event:

$$B_t(\delta) = \left\{\omega \in \Omega : \frac{|S_t|^2}{1/\sigma'^2 + \sum_{s=1}^{t}\left((\theta_{s-1}-\theta_*)^\top x_s\right)^2} > 2\sigma^2 \log\left(\frac{\sqrt{1+\sigma'^2\sum_{s=1}^{t}\left((\theta_{s-1}-\theta_*)^\top x_s\right)^2}}{\delta}\right)\right\}$$

We are interested in bounding the probability that $\bigcup_{t>0} B_t(\delta)$ happens. Define $\tau(\omega) = \min\{t \geq 0 : \omega \in B_t(\delta)\}$, with the convention that $\min \emptyset = \infty$. Then, τ is a stopping time. Further,

$$\bigcup_{t\geq 0} B_t(\delta) = \{\omega : \tau(\omega) < \infty\}$$

Thus, by Eq. 13:

$$\Pr\left[\bigcup_{t\geq 0} B_t(\delta)\right] = \Pr[\tau < \infty] = \Pr\left[B_\tau(\delta), \tau < \infty\right] \leq \Pr\left[B_\tau(\delta)\right] \leq \delta$$

□

This proves that the second term in Eq. 11 is a of order $\sim O(\log(T)\log\log T)$. In fact, with high probability $\sum_{s=1}^{t}\left((\theta_{t-1}-\theta_*)^\top x_t\right)^2 = O(\log(T)^2)$ therefore, with high probability S_T is of order $\sim O(\log(T)\log(\log T)/\delta)$. Consequently, with high probability, S_T is second order.

Proof aggregation: By combining earlier results we find for any $\delta, \sigma' > 0$, with probability at least $1-\delta$, for all $T \geq 0$:

$$\begin{aligned}\bar{R}_T^{\mathfrak{r}} \leq &\left(\sigma\sqrt{d\log\left(\frac{1+TX^2/\lambda d}{\delta/2}\right)} + \lambda^{1/2}S\right)^2 \frac{X^2/\lambda}{\log(1+X^2/\lambda)} d\log\left(1+TX^2/\lambda d\right) \\ &+ \sigma\sqrt{2\left(1/\sigma'^2 + \sum_{s=1}^{t}\left((\theta_{t-1}-\theta_*)^\top x_t\right)^2\right)\log\left(\frac{\sqrt{1+\sigma'^2\sum_{s=1}^{t}\left((\theta_{t-1}-\theta_*)^\top x_t\right)^2}}{\delta/2}\right)}.\end{aligned} \tag{14}$$

D Analysis of the forward algorithm

In this section we derive the high probability time-uniform regret bound for the forward algorithm. Let's recall the theorem.

Theorem. *(Theorem 3.3)For any $\delta > 0$, with probability at least $1-\delta$, for all $T > 0$:*

$$\bar{R}_T^{\mathfrak{f}} \leq (d\sigma)^2 \log\left(\frac{1+TX^2/\lambda d}{\delta/2}\right)\log\left(1+TX^2/\lambda d\right) + o(\log(T)^2)$$

See Eq. 15 for the explicit expression of this bound. The proof proceeds similarly to Appendix C: we need to bound the instantaneous regret.

$$\bar{r}_t = \ell_t(\theta_{t-1}) - \ell_t(\theta_*) = \left(\theta_{t-1}^\top x_t - \theta_*^\top x_t\right)^2 + 2\epsilon_t(\theta_{t-1}^\top x_t - \theta_*^\top x_t)$$

We proceed in three steps like before.

First step: We start by deriving a confidence ellipsoid for this new parameter estimate. This is a novel result.

Theorem. *(Confidence ellipsoid for the Forward algorithm) For any $\delta > 0$, with probability at least $1-\delta$, for allt > 0:*

$$\|\theta_t - \theta_*\|_{G_t} \leq \sqrt{\beta_t(\delta)} = \sigma\sqrt{d\log\left(\frac{1+tX^2/\lambda d}{\delta}\right)} + (\lambda^{1/2}+X)S.$$

Proof. Denote $X_t = (x_1^\top, \dots, x_t^\top), \varepsilon_t = (\epsilon_1, \dots, \epsilon_t)^\top$. Using

$$\begin{aligned}\theta_t &= G_{t+1}^{-1}X_t^\top\left(X\theta_* + \varepsilon_t\right) = G_{t+1}^{-1}X_t^\top\varepsilon_t + G_{t+1}^{-1}(X_t^\top X_t + \lambda I + x_{t+1}^\top x_{t+1})\theta_* - G_{t+1}^{-1}(\lambda I + x_{t+1}^\top x_{t+1})\theta_* \\ &= G_{t+1}^{-1}X_t^\top\varepsilon_t + \theta_* - G_{t+1}^{-1}(\lambda I + x_{t+1}^\top x_{t+1})\theta_*,\end{aligned}$$

we get

$$\begin{aligned}|x^\top\theta_t - x^\top\theta_*| &= |x^\top G_{t+1}^{-1}X_t\varepsilon_t - x^\top G_{t+1}^{-1}(\lambda\theta_* + x_{t+1}x_{t+1}^\top\theta_*)| \\ &\leq \|x\|_{G_{t+1}^{-1}}\Big(\|X_t^\top\varepsilon_t\|_{G_{t+1}^{-1}} + (\sqrt{\lambda}+X)\|\theta_*\|_2\Big),\end{aligned}$$

where in the last inequality we used Cauchy-Schwartz inequality and that by the Sherman-Morrison formula $x_{t+1}^\top G_{t+1}^{-1}x_{t+1} = \frac{x_{t+1}^\top G_t^{-1}x_{t+1}}{1+x_{t+1}^\top G_t^{-1}x_{t+1}} \leq 1$. We know that: $||X_t^\top\varepsilon_t||_{G_{t+1}^{-1}} \leq ||X_t^\top\varepsilon_t||_{G_t^{-1}}$ which allows us to use Theorem 1 from Abbasi-Yadkori et al. [1] that we recall just after this proof. We conclude by plugging $x = G_{t+1}(\theta_t - \theta_*)$. □

Theorem. *(Self-Normalized Bound for Vector-Valued Martingales). Let $\{F_t\}_{t=0}^\infty$ be a filtration. Let $\{\eta_t\}_{t=1}^\infty$ be a real-valued stochastic process such that η_t is F_t -measurable and η_t is conditionally R -sub-Gaussian for some $R \geq 0$ i.e.*

$$\forall\lambda\in\mathbb{R}\quad \mathbf{E}\left[e^{\lambda\eta_t}\mid F_{t-1}\right] \leq \exp\left(\frac{\lambda^2R^2}{2}\right)$$

Let $\{X_t\}_{t=1}^\infty$ be an $\mathbb{R}^d$ -valued stochastic process such that X_t is F_{t-1} -measurable. Assume that V is a $d\times d$ positive definite matrix. For any $t \geq 0$, define

$$\bar{V}_t = V + \sum_{s=1}^t X_sX_s^\top \quad S_t = \sum_{s=1}^t \eta_sX_s.$$

Then, for any $\delta > 0$, with probability at least $1-\delta$, for all $t \geq 0$,

$$\|S_t\|_{\bar{V}_t^{-1}}^2 \leq 2R^2\log\left(\frac{\det\left(\bar{V}_t\right)^{1/2}\det(V)^{-1/2}}{\delta}\right).$$

Note that the deviation of the martingale $\|S_t\|_{\bar{V}_t^{-1}}^2$ is measured by the norm weighted by the matrix $\bar{V}_t^{-1}$ which is itself derived from the martingale, hence the name "self-normalized bound".

For the first term, with probability at least $1-\delta$ for all $t \geq 0$:

$$\begin{aligned}(\theta_{t-1}-\theta_*)^\top x_t &\leq ||x_t||_{G_t^{-1}}||\theta_{t-1}-\theta_*||_{G_t} \\ &\leq \sqrt{\beta_{t-1}(\delta)}||x_t||_{G_t^{-1}} \leq \sqrt{\beta_{T-1}(\delta)}||x_t||_{G_t^{-1}}.\end{aligned}$$

Second step: We can use Lemma B.2 to bound the sum of feature norms. It comes

$$\sum_{t=1}^T(\theta_{t-1}^\top x_t - \theta_*^\top x_t)^2 \leq \beta_T(\delta)d\log\Big(1+TX^2/\lambda d\Big)$$

Third step: Again, we derive a high probability bound To control the second term in the r.h.s of (11).

Lemma D.1. *(Tail inequality) Define $S_t = \sum_{s=1}^t \epsilon_s(\theta_{s-1}-\theta_*)^\top x_s$ and let $(F_t)_{t\geq 0}$ be a filtration such that x_t is F_{t-1} measurable and ϵ_t is F_t measurable. Then S_t is a martingale with respect to F_t and for any $\delta > 0, \sigma' > 0$, with probability at least $1-\delta$, for all $t \geq 0$:*

$$|S_t| \leq \sigma\sqrt{2\left(1/\sigma'^2 + \sum_{s=1}^t\left((\theta_{t-1}-\theta_*)^\top x_t\right)^2\right)\log\left(\frac{\sqrt{1+\sigma'^2\sum_{s=1}^t\left((\theta_{t-1}-\theta_*)^\top x_t\right)^2}}{\delta}\right)}$$

Proof. The proof of this result proceeds in the exact same way as for Lemma C.2. □

Proof aggregation: We combine previous results to finish the proof of the forward algorithm regret bound. For any $\delta, \sigma' > 0$, with probability at least $1-\delta$, for all $T \geq 0$:

$$\bar{R}_T^{\mathbf{f}} \leq \left(\sigma\sqrt{d\log\left(\frac{1+TX^2/\lambda d}{\delta/2}\right)} + (\sqrt{\lambda}+X)S\right)^2 \frac{X^2/\lambda}{\log(1+X^2/\lambda)} d\log\left(1+TX^2/\lambda d\right) + \sigma\sqrt{2\left(1/\sigma'^2 + \sum_{s=1}^{t}\left((\theta_{t-1}-\theta_*)^\top x_t\right)^2\right)\log\left(\frac{\sqrt{1+\sigma'^2\sum_{s=1}^{t}\left((\theta_{t-1}-\theta_*)^\top x_t\right)^2}}{\delta}\right)}. \quad (15)$$

E The unregularized-forward algorithm

For the sake of completeness, we propose a high probability bound on the regret of a non-regularized forward algorithm -studied in the adversarial bounded case in Gaillard et al. [11]- which achieves the optimal asymptotic first order deterministic minimax bound of $dY^2\log(T)$. This algorithm is a simple yet elegant modification of forward regression, it avoids the exploding $\lambda\|\theta_T\|_2^2$ term by setting $\lambda = 0$. Consequently $\theta_t = G_{t+1}^\dagger b_t$, where $G_t^\dagger$ is the pseudo-inverse of G_t.

Theorem E.1. *(Regret of the unregularized forward) The unregularized forward regression achieves, for any $\delta > 0$, with probability at least $1-\delta$ for all $T > 0$:*

$$\bar{R}_T^{u-f} \leq 2(1+\kappa)(1+\alpha)\sigma^2\log\left(\frac{\kappa_d(1+TX^2/\gamma d)}{\delta/4}\right)\log\left(\frac{|G_T^\dagger|}{|G_{T_1}^\dagger|}\right) + 2\sigma^2\log\left(\frac{4T_1}{\delta}\right)\left(d + \sum_{1\leq t\leq T_1, t\in\mathcal{T}}\log\left(\frac{X^2}{\lambda_{r_t}(\sum_{s=1}^{t}x_t x_t^\top)}\right)\right),$$

where $\kappa, \alpha \in \mathbb{R}_+^*$ are peeling parameters (can be chosen), $\gamma = \min_{1\leq t\leq T}\|x_t\|_2$, and $\kappa_d(x) \propto x^d$ up to logarithmic factors and depends on κ and α (cf. Theorem 5.4 in Maillard [15]). $T_1 = \min\{t \geq 1, |G_t| > 0\}$ is, if it exists, the first time the design matrix is non-singular, otherwise $T_1 = T$, and $\mathcal{T}$ is the set of indices t such that rank(G_t) >rank(G_{t-1}). The last term accounts for when the design matrix is singular, and is naturally unbounded (this was also the case in the adversarial case).

Asymptotically, with probability at least $1-\delta$ the first regret term is bounded as:

$$\bar{R}_T^{u-f} \leq 2(1+\kappa)(1+\alpha)\log\left(\frac{C(\kappa,\alpha)(TX^2/\lambda d)^d}{\delta}\right)\log\left((T-T_1)X^2/\lambda d\right),$$

where $C(\kappa, \alpha)$ is a function of the peeling parameters.
We don't seek a more involved analysis to explicit this bound or improve on it, but we see that vaguely it leads to a bound similar to Theorems 3.2 and 3.3 provided that the term accounting for the singularity of the design matrix is controlled. The latter empowers the intuition that in the high probability analysis, the forward algorithm is *first order minimax optimal* even though concretely we cant be sure because we don't have access to uniform lower bounds.

Proof. The proof consists of two mains steps: the first is to use the following bound while the design matrix is singular:

Theorem E.2. *(Theorem 11 Gaillard et al. [11]) For all $T \geqslant 1$, for all sequences $x_1, \ldots, x_T \in \mathbb{R}^d$ and all $y_1, \ldots, y_T \in [-Y, Y]$, the unregularized forward algorithm achieves the regret bound*

$$R_T(\mathbf{u}) \leq Y^2\sum_{t=1}^{T}\mathbf{x}_t^{\mathrm{T}}\eta_t^\dagger\mathbf{x}_t \leqslant dY^2\log T + dY^2 + Y^2\sum_{t\in[1,T]\cap\mathcal{T}}\log\left(\frac{X^2}{\lambda_{r_t}(\sum_{s=1}^{t}x_s x_s^\top)}\right)$$

where $\forall M \in \mathcal{M}_d(\mathbb{R}), \lambda_1(M) \geq \ldots \geq \lambda_d$ *are* M*'s eigenvalues and* $r_t = \mathrm{rank}(\sum_{s=1}^t x_s x_s^\top)$ *and where the set* $\mathcal{T}$ *contains* r_T *rounds, given by the smallest* $s \geqslant 1$ *such that* x_s *is not* mull, *and all the* $s \geqslant 2$ *for which rank* $(\mathrm{G}_{s-1}) \neq \mathrm{rank}\,(\mathrm{G}_s)$.

The second step is a bound when the design matrix is invertible, using Theorem B. Denote $T_1 = \inf_{t \geq 1}\{|G_t| > 0\}$, using Theorem E.2:

$$\bar{R}_{T_1} \leq Y^2 \left(d\log(T_1) + d + \sum_{1 \leq t \leq T_1, t \in \mathcal{T}} \log\left(\frac{X^2}{\lambda_{r_t}(G_t)}\right)\right)$$

From standard results on sub-Gaussian noise, we also know that $\mathbb{E}[\max_{1 \leq t \leq T} \epsilon_t] \leq \sigma\sqrt{2\log(T)}$ (see *e.g.* Kamath [12]), then using the transformation of Laplace along with Markov's inequality, $\forall \delta > 0\ \mathrm{P}\left(\forall T \geq 1, Y^2 \leq 2\sigma^2 \log(T/\delta)\right) \geq 1 - \delta$, hence with probability at least $1-\delta$:

$$\bar{R}_{T_1} \leq 2d\sigma^2 \log\frac{T_1}{\delta}\log(T_1) + 2d\sigma^2\log\frac{T_1}{\delta} + 2\sigma^2\frac{\log T_1}{\delta}\sum_{1 \leq t \leq T_1, t \in \mathcal{T}} \log\left(\frac{X^2}{\lambda_{r_t}(\sum_{s=1}^t x_s x_s^\top)}\right). \tag{16}$$

And for $T > T_1$, we bound $R_T - R_{T_1}$ using the same methodology in Appendix C and Appendix D and using the confidence bounds above (*cf.* Theorem B). $\forall \delta > 0$, with probability at least $1-\delta$:

$$\forall t > T_1 : \left(\theta_{t-1}^\top x_t - \theta_*^\top x_t\right)^2 \leq \sqrt{\beta_{t-1}(\delta)}||x_t||_{G_t^\dagger}$$

We use the tail inequality. (C.2) to get, $\forall \delta > 0$, with probability at least $1-\delta, \forall T > 0$:

$$\bar{R}_T - \bar{R}_{T_1} \leq 2(1+\kappa)(1+\alpha)\sigma^2 \log\left(\frac{\kappa_d(1 + TX^2/\lambda d)}{\delta/2}\right)\log\left(\frac{|G_T^\dagger|}{|G_{T_1}^\dagger|}\right) \tag{17}$$

From (16) and (17) we obtain for all $\delta > 0$, with probability at least $1-\delta$:

$$\begin{aligned}\bar{R}_T \lesssim{}& 2(1+\kappa)(1+\alpha)\sigma^2 \log\left(\frac{\kappa_d(1 + TX^2/\lambda d)}{\delta/4}\right)\log\left(\frac{|G_T^\dagger|}{|G_{T_1}^\dagger|}\right)\\ &+ 2\sigma^2\frac{\log(T_1)}{\delta/4}\left(d + \sum_{1 \leq t \leq T_1, t \in \mathcal{T}} \log\left(\frac{X^2}{\lambda_{r_t}(\sum_{s=1}^t x_t x_t^\top)}\right)\right).\end{aligned}$$

□

F Applications

In this section, we provide technical details regarding the settings of stationary and non-stationary linear bandits.

F.1 Linear bandits (Proof of Theorem 4.1)

We start by analyzing linear bandits in the stationary setting. Let us first see how OFUL$^{\mathtt{f}}$ behaves in the "unbounded rewards" scenario.

F.1.1 OFUL with forward regression

Consider the same setting as that of Abbasi-Yadkori et al. [1], that we detailed in Section 4, we write the confidence interval $C_t(x)$ for the forward algorithm at the action x as:

$$\left\{\theta \in \mathbb{R}^d : ||\theta_t^{\mathtt{f}} - \theta||_{G_t + xx^\top} \leq \sqrt{\beta_t(x,\delta)} = (\sqrt{\lambda} + ||x||_2)S + \sigma\sqrt{2\log\left(\frac{(1+tX^2/\lambda d)^{d/2}}{\delta}\right)}\right\}$$

which gives, for all $T \geq 0$ the regret (*cf.* Theorem 4.1):

$$R_T \leq 4\sqrt{Td\log(\lambda + TX^2/d)}\Big(\lambda^{1/2}(S+X) + \sigma\sqrt{2\log(1/\delta) + d\log(1+TX^2/(\lambda d))}\Big)$$

this is equivalent to ridge in its first order, with better scaling and dependence on λ.

Proof. Lets decompose the instantaneous regret as follows:

$$\begin{aligned}
r_t &= \langle \theta_*, x_* \rangle - \langle \theta_*, x_t \rangle \leq \left\langle \tilde{\theta}_t, x_t \right\rangle - \langle \theta_*, x_t \rangle = \left\langle \tilde{\theta}_t - \theta_*, x_t \right\rangle \\
&= \left\langle \widehat{\theta}_{t-1} - \theta_*, x_t \right\rangle + \left\langle \tilde{\theta}_t - \widehat{\theta}_{t-1}, x_t \right\rangle \\
&= \left\| \widehat{\theta}_{t-1} - \theta_* \right\|_{(G_{t-1}+x_t x_t^\top)} \|X_t\|_{(G_{t-1}+x_t x_t^\top)^{-1}} + \left\| \tilde{\theta}_t - \widehat{\theta}_{t-1} \right\|_{(G_{t-1}+x_t x_t^\top)} \|x_t\|_{(G_{t-1}+x_t x_t^\top)^{-1}} \\
&\leq 2\sqrt{\beta_{t-1}(x_t,\delta)}\, \|x_t\|_{(G_{t-1}+x_t x_t^\top)^{-1}}\,, \qquad (18)
\end{aligned}$$

where $\widetilde{\theta}_t$ is the optimistic parameter estimate, *i.e.* the $\theta \in C_t(x_t)$ that maximizes the upper confidence bound on the reward of action x_t. The first inequality is since $\left(X_t, \widetilde{\theta}_t\right)$ is optimistic, and the last step holds by Cauchy-Schwarz. Using inequality (18) and the fact that $\sqrt{\beta_t(x,\delta)} \leq \sqrt{\beta_t(\delta)} = (\sqrt{\lambda}+X)S + \sigma\sqrt{2\log\left(\frac{(1+tX^2/\lambda d)^{d/2}}{\delta}\right)}$ we get that, with probability at least $1-\delta$, for all $n \geq 0$

$$\begin{aligned}
R_n &\leq \sqrt{n\sum_{t=1}^n r_t^2} \leq \sqrt{8\beta_n(\delta) n \sum_{t=1}^n \|x_t\|_{(G_{t-1}+x_t x_t^\top)^{-1}}} \\
&\leq 4\sqrt{nd\log(\lambda + nL/d)}\left((\lambda^{1/2}+X)S + \sigma\sqrt{2\log(1/\delta) + d\log(1+nL/(\lambda d))}\right)
\end{aligned}$$

where the last step follow from Lemma B.2. □

F.1.2 OFUL with ridge regression

In this section, we derive a novel regret bound for online ridge regression, one that doesn't require the bounded rewards assumption (*cf.* Assumption 1).

Theorem. *(Bandits with unbounded rewards) Without Assumption 1, for all $\delta > 0$, OFUL$^{\mathfrak{r}}$ achieves with probability at least $1-\delta$, for all $T \geq 1$,*

$$R_T^{\mathfrak{r}} \leq 4\sqrt{\frac{\boldsymbol{X}^2 Td\log(1+TX^2/\lambda d)}{\boldsymbol{\lambda}\log(1+\boldsymbol{X}^2/\boldsymbol{\lambda})}}\left(\lambda^{1/2}S + \sigma\sqrt{2\log(1/\delta) + d\log(1+TX^2/(\lambda d))}\right),$$

Proof. The proof follows exactly like in Section F.1.1 except the last step (control of the norm of actions) that now proceeds using Lemma C.1. The first step is to use the confidence ellipsoid for the ridge regression parameter (see the second theorem in Section C or Theorem 2 of Abbasi-Yadkori et al. [1]). With probability at least $1-\delta$, for all $t \geq 0$, θ_* lies in the set

$$C_t = \left\{\theta \in \mathbb{R}^d : \|\theta_t^{\mathfrak{r}} - \theta\|_{G_t} \leq \sqrt{\beta_t(\delta)} = \sigma\sqrt{d\log\left(\frac{1+tX^2/\lambda d}{\delta}\right)} + \lambda^{1/2}S\right\}.$$

Then

$$\begin{aligned}
r_t &= \langle \theta_*, x_* \rangle - \langle \theta_*, x_t \rangle \leq \left\langle \tilde{\theta}_t, x_t \right\rangle - \langle \theta_*, x_t \rangle = \left\langle \tilde{\theta}_t - \theta_*, x_t \right\rangle \\
&= \left\langle \widehat{\theta}_{t-1} - \theta_*, x_t \right\rangle + \left\langle \tilde{\theta}_t - \widehat{\theta}_{t-1}, x_t \right\rangle \\
&= \left\| \widehat{\theta}_{t-1} - \theta_* \right\|_{G_{t-1}} \|X_t\|_{G_{t-1}^{-1}} + \left\| \tilde{\theta}_t - \widehat{\theta}_{t-1} \right\|_{G_{t-1}} \|x_t\|_{G_{t-1}^{-1}} \\
&\leq 2\sqrt{\beta_{t-1}(\delta)}\, \|x_t\|_{G_{t-1}^{-1}} \qquad (19)
\end{aligned}$$

where $\widetilde{\theta}_t$ is the optimistic parameter estimate, *i.e.* the $\theta \in C_t$ that maximizes the upper confidence bound on the reward of action x_t. The first inequality is since $\left(X_t, \widetilde{\theta}_t\right)$ is optimistic, and the last step holds by Cauchy-Schwarz. Using inequality (19) we get that, with probability at least $1-\delta$, for all $n \geq 0$

$$
\begin{aligned}
R_n &\leq \sqrt{n \sum_{t=1}^{n} r_t^2} \leq \sqrt{8\beta_n(\delta) n \sum_{t=1}^{n} \|x_t\|_{G_{t-1}^{-1}}} \\
&\leq 4\sqrt{\frac{ndX^2 \log(1+nX^2/\lambda d)}{\lambda \log(1+X^2/\lambda)}} \left(\lambda^{1/2} S + \sigma \sqrt{2\log(1/\delta) + d\log(1+nX^2/(\lambda d))}\right)
\end{aligned}
$$

where the last step follow from Lemma C.1. □

F.2 Non-stationary linear bandits

In this section, we study linear stochastic bandits in the non-stationary setting. We provide an experimental study of this setup in Section G. We now turn to the setting of *non-stationary stochastic linear bandits*, where the target parameter is varying with time: $\theta_* = \theta_*(t) \in \mathbb{R}^d$, assuming that $\sum_{s=1}^{T-1} \|\theta_*(s) - \theta_*(s+1)\|_2 \leq B_T$.

One of the optimal algorithms in this setting is `D-LinUCB` of [19], it defines θ_t as

$$
\theta_t = \arg\min_{\theta \in \mathbb{R}^d} \sum_{s=1}^{t} \gamma^{t-s} (y_s - \langle x_s, \theta \rangle)^2 + \lambda/2 \|\theta\|_2^2.
$$

`D-LinUCB` proceeds as follows:

Algorithm 4: `D-LinUCB`

Input: $\delta, \sigma, \lambda, X, S, \gamma > 0$, dimension $d \in \mathbb{N}^*$.
Initialization: $b = 0_{\mathbb{R}^d}$, $V = \lambda I_d$, $\widetilde{V} = \lambda I_d$, $\theta = 0_{\mathbb{R}^d}$
for $t \geq 1$ **do**
    Receive $\mathcal{X}$, compute $\beta_{t-1} = \sqrt{\lambda} S + \sigma \sqrt{2\log\left(\frac{1}{\delta}\right) + d\log\left(1 + \frac{X^2(1-\gamma^{2(t-1)})}{\lambda d(1-\gamma^2)}\right)}$
    for $a \in \mathcal{X}$ **do**
        Compute $\text{UCB}(a) = a^\top \theta + \beta_{t-1} \sqrt{a^\top V^{-1} \widetilde{V} V^{-1} a}$
    $A_t = \arg\max_a(\text{UCB}(a))$
    Play action A_t and **receive reward** X_t
    Updating phase: $V = \gamma V + x_t x_t^\top + (1-\gamma)\lambda I_d$, $\widetilde{V} = \gamma^2 \widetilde{V} + x_t x_t^\top + (1-\gamma^2)\lambda I_d$
        $b = \gamma b + Y_t X_t$, $\theta = V^{-1} b$

We recall the regret bound of standard `D-LinUCB` .

Theorem F.1. *(Theorem 3 of Russac et al. [19]) Assuming that* $\sum_{s=1}^{T-1} \|\theta_*(s) - \theta_*(s+1)\|_2 \leq B_T$ *and* $\forall x \in \mathcal{X}, t \geq 1 : \langle x, \theta_t \rangle \leq 1$*, the regret of the* `D-LinUCB` *algorithm is bounded for all* $\gamma, \delta \in (0,1)$ *and integer* $D \geq 1$*, with probability at least* $1-\delta$*, by:*

$$
R_T^r \leq 2XDB_T + \frac{4X^3 S}{\lambda} \frac{\gamma^D}{1-\gamma} T + 2\sqrt{2}\beta_T \sqrt{dT} \times \sqrt{T\log(1/\gamma) + \log\left(1 + \frac{X^2}{d\lambda(1-\gamma)}\right)},
$$

where β_T is the width of the confidence interval for $\theta_*(T)$.

Now we introduce `D-LinUCB` $^{\texttt{f}}$, which uses the forward algorithm and defines an action dependent θ_t as:

$$
\arg\min_{\theta \in \mathbb{R}^d} \sum_{s=1}^{t} \gamma^{t-s} (y_s - \langle x_s, \theta \rangle)^2 + \lambda/2 \|\theta\|_2^2 + \langle x, \theta \rangle^2. \tag{20}
$$

Theorem F.2. *Assuming that $\sum_{s=1}^{T-1} \|\theta_*(s) - \theta_*(s+1)\|_2 \leq B_T$, the regret of the* `D-LinUCB` ^{f} *is bounded for all $\gamma, \delta \in (0,1)$ and integer $D \geq 1$, with probability at least $1-\delta$, by*

$$R_T^f \leq 2XDB_T + \frac{4X^3S}{\lambda}\frac{\gamma^D}{1-\gamma}T + 2\beta_T\sqrt{dT}\sqrt{T\log(1/\gamma) + \log\left(1 + \frac{(2-\gamma)X^2}{d\lambda(1-\gamma)}\right)}.$$

Proof. This result is again a modification of the original proof consisting in bounding the sum of the actions' norms differently. Let us recall the notations $V_t = \sum_{s=1}^t w_s x_s x_s^\top + \lambda_t I_d + xx^\top$ and $\tilde{V}_t = \sum_{s=1}^t w_s^2 x_s x_s^\top + \mu_t I_d + xx^\top$. To summarize the difference of this analysis -that no longer requires a bounded rewards assumption- at the step where we bound the sum of actions' norms, we replace Proposition 4 of Russac et al. [19]:

$$\sum_{t=1}^T \min\left(1, \|x_t\|^2_{V_{t-1}^{-1}\tilde{V}_{t-1}V_{t-1}^{-1}}\right) \leq 2\sum_{t=1}^T \log\left(1 + \gamma^{-t}\|x_t\|^2_{V_{t-1}^{-1}}\right) \leq 2\log\left(\frac{\det(V_T)}{\lambda^d}\right),$$

that requires the predictions to lie in the same range as the rewards with this inequality for `D-LinUCB` ^{f}

$$\sum_{t=1}^T \|x_t\|^2_{V_t^{-1}\tilde{V}_t V_t^{-1}} \leq \sum_{t=1}^T \log\left(1 + \gamma^{-t}\|x_t\|^2_{V_t^{-1}}\right) \leq \log\left(\frac{\det(V_T)}{\lambda^d}\right).$$

We don't provide the full proof of this result as it is cumbersome and not of special interest for our purposes since it is similar to the analysis for `D-LinUCB` except for the inequality above. □

Remark 6. *This result is fascinating as it first allows to remove an unnecessary assumption, and further yields a better bound than* `D-LinUCB` ^{r} *which suffers the factor $\frac{\boldsymbol{X}\sqrt{2}}{\boldsymbol{\lambda}\log(1+\boldsymbol{X}/\boldsymbol{\lambda})}$ in its last regret term without assumption 1.*

G Experiments

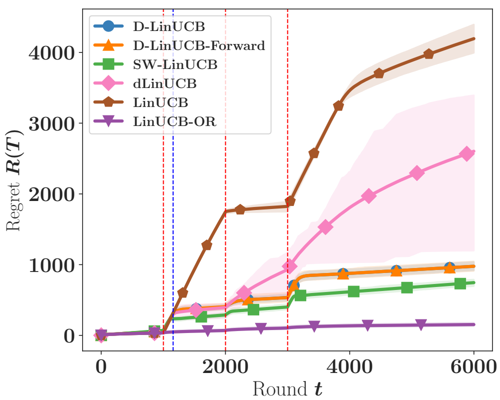
Experimental details and instructions: The experiments were run on a personal laptop with Intel Core i7-8665U, CPU 1.90GHz × 8. Code for the experiments for online regression and linear bandits is provided in the files "OnlineRegression.ipynb" and "LinearBanditsCode.ipynb". For the experiments of non-stationary linear bandits that we present next, we used an existing code from the Github page of Russac et al. [19] and we added an implementation of `D-LinUCB` ^{f} to compare with previous algorithms, this can be seen in the "WeightedLinearBandits" folder in which "`D-LinUCB` Forward_class.py" is our new algorithm; experiments for this setting can be run from the two ipynb files in the Experiments sub-folder.

Experiments for non-stationary linear bandits: We now reproduce the experiments of [19] for non-stationary linear bandits, and add `D-LinUCB` ^{f} to the pool of algorithms. We first simulate an *abruptly* changing environment of dimension 2 with 3 changes: for $t < 10^3 : \theta_* = (1,0)$; for $10^3 \leq t \leq 2.10^3 : \theta_* = (-1,0)$; for $2.10^3 < t < 3.10^3 : \theta_* = (0,1)$; for $t > 3.10^3 : \theta_* = (0,-1)$. We observe in Fig. 4a that both variants of `D-LinUCB` compare on par. Here `LinUCB-OR` denotes an oracle knowing the change points.

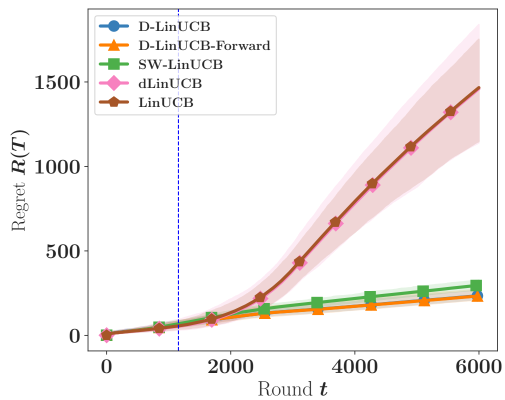
Second, we simulate a slowly changing environment where the parameter θ_* starts at $(1,0)$ and moves counter-clockwise on the unit-circle up to the position $(0,1)$ in 3.10^3 steps then remains there, $B_T = 1.57$. We see the results in Fig. 4b, where we notice that in this setting as well, `D-LinUCB` ^{f} has very similar performance to standard `D-LinUCB` .

Remark 7. *In both experiments, we also reported the performances of* `SW-LinUCB`*, that is alternative version to* `D-LinUCB`*.* `SW-LinUCB` *is better suited for abrupt changes while* `D-LinUCB` *is better suited for slow changes.*

Note that we added these final experiments to demonstrate the competitiveness of algorithms that use forward regression against their ridge counterparts in the same settings that were used by previous works. While we could have specified specific parameters to illustrate the robustness to regularization of algorithms that incorporate the forward algorithm; we estimate that the experiments presented in the main text already fulfilled this objective. Again, the purpose here is to show that using the forward algorithm improves the theoretical guarantees without deteriorating the performance.



(a) abruptly changing environment



(b) slowly varying environment

Figure 4: Performance of several algorithms in an non-stationary environments, averaged over 100 runs, shaded areas represent one standard deviation.