A Examples of TS distributions

Example 1: Uniform distribution $\eta \sim \mathcal{U}_{B_d(0,\sqrt{d})}$. The uniform distribution satisfies the concentration property with constants $c = 1$ and $c' = \frac{e}{d}$ by definition. Since the set $\{\eta | u^\mathsf{T}\eta \geq 1\} \cap B_d(0, \sqrt{d})$ is an hyper-spherical cap for any direction u of $\mathbb{R}^d$, the the anti-concentration property is satisfied provided that the ratio between the volume of an hyper-spherical cap of height $\sqrt{d} - 1$ and the volume of the ball of radius $\sqrt{d}$ is constant (i.e., independent from d). Using standard geometric results (see Prop. 9), one has that for any vector $\|u\| = 1$

$$\mathbb{P}(u^\mathsf{T}\eta \geq 1) = \frac{1}{2} I_{1-\frac{1}{d}}\Big(\frac{d+1}{2}, \frac{1}{2}\Big), \tag{9}$$

where $I_x(a, b)$ is the incomplete regularized beta function. In Prop. 10 we prove that

$$I_{1-\frac{1}{d}}\Big(\frac{d+1}{2}, \frac{1}{2}\Big) \geq \frac{1}{8\sqrt{6\pi}},$$

and hence we obtain $p = \frac{1}{16\sqrt{6\pi}}$. □.

Example 2: Gaussian case $\eta \sim \mathcal{N}(0, I_d)$. The concentration property comes directly from the Chernoff bound for standard Gaussian random variable together with union bound argument. For any $\alpha > 0$, we have

$$\mathbb{P}(\|\eta\| \leq \alpha\sqrt{d}) \geq \mathbb{P}(\forall 1 \leq i \leq d,\ |\eta_i| \leq \alpha) \geq 1 - d\mathbb{P}(|\eta_i| \geq \alpha).$$

Standard concentration inequality for Gaussian random variable gives, $\forall \alpha > 0$,

$$\mathbb{P}(|\eta_i| \geq \alpha) \leq 2e^{-\alpha^2/2}.$$

Plugging everything together with $\alpha = \sqrt{2\log\frac{2d}{\delta}}$ gives the desired result with $c = c' = 2$. Let η_i be the i-th component of η for any $1 \leq i \leq d$. Then $\eta_i \sim \mathcal{N}(0, 1)$. Since η is rotationally invariant, for any direction u of $\mathbb{R}^d$ and an appropriate choice of basis, we have $\mathbb{P}(u^\mathsf{T}\eta \geq 1) \geq \mathbb{P}(\eta_1 \geq 1)$. From standard Gaussian properties (see Thm 2 of Chang et al. [2011]) we have

$$\mathbb{P}(\eta_1 \geq 1) = \frac{1}{2}\text{erfc}\left(\frac{1}{\sqrt{2}}\right) \geq \frac{1}{4\sqrt{e\pi}}$$

which ensures the anti-concentration property with $p = \frac{1}{4\sqrt{e\pi}}$. □

B Properties of convex function

Proposition 4. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function and C be a closed convex set of $\mathbb{R}^d$. Then, on C, f reaches its maximum on the boundary of C.*

Proof. Let's denote as $int(C)$ and $bound(C)$ the interior and the boundary of the closed convex set C respectively. Assume that $\exists x^\star \in int(C)$ such that $f(x^\star) > f(x)$ for any $x \in bound(C)$ and $f(x^\star) \geq f(y)$ for any $y \in int(C)$.

Then define $y = x^\star + \epsilon(x^\star - x)$ for some $x \in bound(C)$. By definition of the open set $int(C)$, $\exists \epsilon > 0$ such that $y \in int(C)$. Moreover, $x^\star \in [y, x]$ e.g.

$$x^\star = (1-t)x + ty, \qquad t = \frac{1}{1+\epsilon} \in]0, 1[$$

Using the convexity of f on has

$$\begin{aligned} f(x^\star) &\leq (1-t)f(x) + tf(y) < (1-t)f(x^\star) + tf(y) \\ f(x^\star) &< f(y) \end{aligned}$$

which is impossible by assumption. □

Proposition 5. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function. Let $B_d(0,1)$ be the unit $d-$dimensional ball and $S_d(0,1)$ the associated unit sphere.*
Given a point $x \in S_d(0,1)$, define as $\mathcal{H}(x)$ the hyperplan tangent to $B_d(0,1)$ at the point x. $\mathcal{H}(x)$ split $\mathbb{R}^d$ into two complementary subspace $\mathcal{G}(x)$ and $\mathcal{G}^{\perp}(x)$ where $\mathcal{G}(x)$ does not contain the unit ball by convention.
Then for any $x^\star \in S_d(0,1)$ such that $f(x^\star) \geq f(x)$ for all $x \in B_d(0,1)$, one has

$$\forall y \in \mathcal{G}(x^\star), \quad f(y) \geq f(x^\star)$$

Proof. We first notice that from Proposition 4 $x^\star$ is well defined since the maximum is reached on the boundary. The associated subspace $\mathcal{G}(x^\star)$ is then

$$\mathcal{G}(x^\star) := \{y = x^\star + u, u \in \mathbb{R}^d \ \mid \ u^\mathsf{T} x^\star \geq 0\}.$$

We want to show that $f(y) \geq f(x^\star)$ for any $y \in \mathcal{G}(x^\star)$. We introduce the increasing sequence of subspace

$$\mathcal{G}_n = \left\{ y = x^\star + u, u \in \mathbb{R}^d \ \mid \ u^\mathsf{T} x^\star \geq \frac{\|u\|}{2(n-1)} \right\}, \quad n \geq 2.$$

For any $y = x^\star + u$ in $\mathcal{G}_n$, we associate

$$x = x^\star - \frac{1}{2(n-1)} \frac{u}{\|u\|}.$$

By definition of y (and hence u), we have

$$\begin{aligned}
\|x\|^2 &= 1 + \frac{1}{2(n-1)}^2 - \frac{1}{2(n-1)\|u\|} u^\mathsf{T} x^\star \\
&= 1 + \frac{1}{2(n-1)} \left[\frac{1}{2(n-1)} - \frac{u^\mathsf{T}}{\|u\|} x^\star \right] \\
&\leq 1,
\end{aligned}$$

which means that $x \in \mathcal{B}_d(0,1)$. Moreover let $t = [2(n-1)\|u\| + 1]^{-1}$, $t \in]0,1[$ one has $x^\star = (1-t)x + ty$. Since $x \in \mathcal{B}_d(0,1)$ then

$$\begin{aligned}
f(x^\star) &\leq (1-t)f(x) + tf(y) \\
&\leq (1-t)f(x^\star) + tf(y) \\
\Rightarrow & f(x^\star) \leq f(y).
\end{aligned}$$

Since the statement of the proposition holds for any $\mathcal{G}_n$, then we obtain the desired result for $\mathcal{G}$ by continuity of f. Let $y \in \mathcal{G}(x^\star)$, $y = x^\star + u$. If $u^\mathsf{T} x^\star > 0$, then $\exists n \geq 2$ such that $y \in \mathcal{G}_n$ and the proposition is satisfied. Otherwise, if $u^\mathsf{T} x^\star = 0$, we introduce the sequences $\{u_n\}$ and $\{y_n\}$ defined as:

$$\begin{aligned}
u_n &= u + \frac{\|u\|}{\sqrt{1 - \frac{1}{2(n-1)}^2}} \frac{x^\star}{2(n-1)} \\
&= u + \frac{\|u_n\|}{2(n-1)} x^\star, \\
y_n &= x^\star + u_n.
\end{aligned}$$

By construction, $y_n \in \mathcal{G}_n$ and $y_n \to y$ as $n \to \infty$. Since the $f(y_n) \geq f(x^\star)$ for any $n \geq 2$ we obtain the desired result taking the limit since f is continuous as a convex function on $\mathbb{R}^d$. □

Theorem 2 (A.D. Alexandrov)**.** *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function, then it is twice differentiable almost everywhere with respect to the Lebesgue's measure.*

Proof. This result is an extension of the Rademacher's theorem for convex functions. A proof can be found in Niculescu and Persson [2006], theorem 3.11.2. □

C Properties of support function (proof of Proposition 3 and Lemma 2)

We study the *support function* of a set C, which is a function $f_C : \mathbb{R}^d \to \mathbb{R}$ such that

$$f_C(\theta) = \sup_{x \in C} x^\mathsf{T} \theta \tag{10}$$

Those functions are at the core of convex geometry analysis.

Proposition 6. *Let $C \subset \mathbb{R}^d$ be a non-empty compact set and f_C the associated support function. Then,*

1. *f_C is real-valued and $\sup_{x\in C} x^\mathsf{T}\theta$ is attained in C,*
2. *f_C is convex,*
3. *f_C is continuous on $\mathbb{R}^d$ and twice differentiable almost everywhere with respect to the Lebesgue's measure.*

Proof. 1. This comes directly from the compactness of C: since C is bounded, the support function is real-valued and since C is closed, the supremum is attained in C,

2. Let θ_1, θ_2 two vectors of $\mathbb{R}^d$, and $t \in (0, 1)$. By definition of the supremum, since f_C is real-valued:

$$f_C(t\theta_1 + (1-t)\theta_2) = \sup_{x\in C} \left(tx^\mathsf{T}\theta_1 + (1-t)x^\mathsf{T}\theta_2 \right) \le t \sup_{x\in C} x^\mathsf{T}\theta_1 + (1-t) \sup_{x\in C} x^\mathsf{T}\theta_2$$

3. The continuity is consequence of the convexity of f_C on the open convex set $\mathbb{R}^d$ and the second order differentiability comes from Alexandrov's theorem 2.

□

Proposition 7. *Let $x(\theta) \in \arg\sup_{x\in C} x^\mathsf{T}\theta$, denote as $\nabla f_C(\theta)$ and $\partial f_C(\theta)$ the gradient (when it is uniquely defined) and the sub-gradient of f_C in $\theta \in \mathbb{R}^d$. Then,*

1. *for all $\theta \in \mathbb{R}^d$, $x(\theta) \in \partial f_C(\theta)$,*
2. *their exists a null set $\mathcal{N}$ with respect to the Lebesgue's measure such that $x(\theta) = \nabla f_C(\theta)$ for all $\theta \in \mathbb{R}^d \setminus \mathcal{N}$,*
3. *equivalentely, $x(\theta) = \nabla f_C(\theta)$ where the equality holds in the sense of the distribution.*

Proof. Thanks to proposition 6, we know that the supremum is attained in $x(\theta) \in C$. Moreover, Alexandrov's theorem guarantee that $\mathcal{N}$ is a null-set. Since the sub-gradient is reduced to a singleton where the function is differentiable e.g. $\partial f_C(\theta) = \{\nabla f_C(\theta)\}$ for all $\theta \in \mathbb{R}^d \setminus \mathcal{N}$, one just need to show to $x(\theta) \in \partial f_C(\theta)$ for all $\theta \in \mathbb{R}^d$. Since $f_C(\theta) = \max_{x\in C} x^\mathsf{T}\theta$, their exist at least one $x(\theta) \in C$ for which the maximum is attained i.e. $x(\theta)^\mathsf{T}\theta = f_C(\theta)$. Moreover, for any $\bar\theta \in \mathbb{R}^d$, $f_C(\bar\theta) \ge x(\theta)^\mathsf{T}\bar\theta$ by definition. Therefore,

$$\begin{aligned} f_C(\bar\theta) - x(\theta)\bar\theta \ge 0 := f_C(\theta) - x(\theta)^\mathsf{T}\theta \\ f_C(\bar\theta) \ge f_C(\theta) + x(\theta)^\mathsf{T}\left(\bar\theta - \theta\right), \quad \forall \bar\theta \in \mathbb{R}^d \end{aligned}$$

which is the definition of the sub-gradient.

□

D Regret Proofs

We collect here the main tools that we need to derive the proof. We first recall the Azuma's concentration inequality for super-martingale.

Proposition 8. *If a super-martingale $(Y_t)_{t\ge0}$ corresponding to a filtration $\mathcal{F}_t$ satisfies $|Y_t - Y_{t-1}| < c_t$ for some constant c_t for all $t = 1, \dots, T$ then for any $\alpha > 0$,*

$$\mathbb{P}(Y_T - Y_0 \ge \alpha) \le 2e^{-\frac{\alpha^2}{2\sum_{t=1}^T c_t^2}}$$

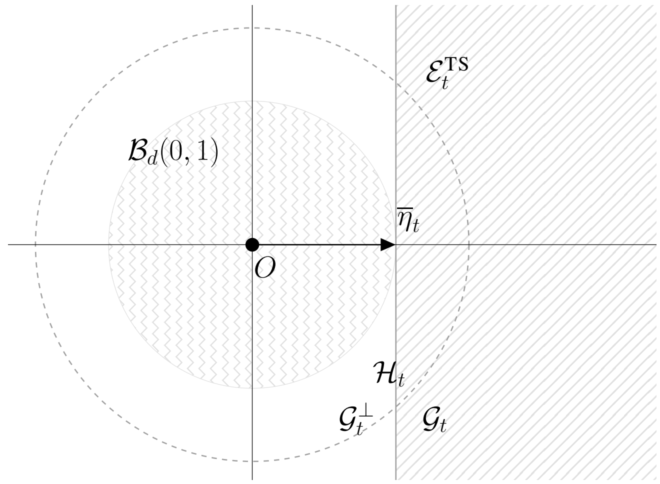

Figure 5: Illustration of the probability of selecting an optimistic $\widetilde{\theta}_t$.

Proof of Lemma 1. We first bound the two events separately.

Bounding $\widehat{E}$. This bound is a straightforward application of Proposition 1 together with a union bound argument. Let $\delta' = \delta/(4T)$, then

$$\begin{aligned}
\forall 1 \le t \le T, \quad & \mathbb{P}\left(\|\widehat{\theta}_t - \theta^\star\|_{V_t} \le \beta_t(\delta')\right) \ge 1 - \delta' \\
\text{from union bound,} \quad & \mathbb{P}\left(\bigcap_{t=1}^T \left\{\|\widehat{\theta}_t - \theta^\star\|_{V_t} \le \beta_t(\delta')\right\}\right) \ge 1 - \sum_{t=1}^T \mathbb{P}\left(\|\widehat{\theta}_t - \theta^\star\|_{V_t} \ge \beta_t(\delta')\right) \\
\Rightarrow \quad & \mathbb{P}\left(\bigcap_{t=1}^T \left\{\|\widehat{\theta}_t - \theta^\star\|_{V_t} \le \beta_t(\delta')\right\}\right) \ge 1 - \sum_{t=1}^T \delta' \\
\Rightarrow \quad & \mathbb{P}(\widehat{E}) \ge 1 - T\delta' = 1 - \frac{\delta}{4}.
\end{aligned}$$

Bounding $\widetilde{E}$. This bound comes directly from the concentration property of the TS sampling distribution. From the expression of $\widetilde{\theta}_t = \widehat{\theta}_t + \beta_t(\delta')V_t^{-1/2}\eta_t$ where η_t is drawn i.i.d. from $\mathcal{D}^{\mathrm{TS}}$, we have

$$\forall 1 \le t \le T, \quad \mathbb{P}\left(\|\widetilde{\theta}_t - \widehat{\theta}_t\|_{V_t} \le \beta_t(\delta')\sqrt{cd\log\frac{c'd}{\delta'}}\right) = \mathbb{P}\left(\|\eta_t\| \le \sqrt{cd\log\frac{c'd}{\delta'}}\right).$$

Then from Definition 1, we have

$$\mathbb{P}\left(\|\eta_t\| \le \sqrt{cd\log\frac{c'd}{\delta'}}\right) \ge 1 - \delta'.$$

As before, a union bound over the two bounds ensures that

$$\mathbb{P}(\widetilde{E}) \ge 1 - T\delta' = 1 - \frac{\delta}{4}.$$

Finally, a union bound argument between the two terms leads to

$$\mathbb{P}(\widehat{E} \cap \widetilde{E}) \ge 1 - \frac{\delta}{2}.$$

□

Proof of Lemma 3. We need to study the probability that a $\widetilde{\theta}$ drawn at time t from the TS sampling distribution is optimistic, i.e., $J(\widetilde{\theta}) \geq J(\theta^\star)$, under event $\widehat{E}_t$. More formally let

$$p_t = \mathbb{P}\big(J(\widetilde{\theta}) \geq J(\theta^\star) | \mathcal{F}_t, \widehat{E}_t\big).$$

Using the definition of $\widehat{E}_t$ we have that $\theta^\star \in \mathcal{E}_t^{\mathrm{RLS}}$ (i.e., the true parameter vector belongs to the RLS ellipsoid) and then we can replace $J(\theta^\star)$ by the supremum over the ellipsoid as

$$p_t \geq \mathbb{P}\Big(J(\widetilde{\theta}) \geq \sup_{\theta \in \mathcal{E}_t^{\mathrm{RLS}}} J(\theta) \Big| \mathcal{F}_t, \widehat{E}_t\Big).$$

By recalling the definition of the TS sampling process, we can write $\widetilde{\theta} = \widehat{\theta}_t + \beta_t(\delta') V_t^{-1/2} \eta$, where $\eta \sim \mathcal{D}^{\mathrm{TS}}$ and for notational convenience, we define the function $f_t(\eta) = J(\widehat{\theta}_t + \beta_t(\delta') V_t^{-1/2}\eta)$. Let $\overline{\theta}_t = \arg\sup_{\theta \in \mathcal{E}_t^{\mathrm{RLS}}} J(\theta)$ and $\overline{\eta}_t$ be the corresponding η (i.e., $\overline{\eta}_t$ is such that $\overline{\theta}_t = \widehat{\theta}_t + \beta_t(\delta') V_t^{-1/2} \overline{\eta}_t$). Since the supremum is taken within $\mathcal{E}_t^{\mathrm{RLS}}$, $\overline{\eta}_t$ belongs to the unit ball (i.e., $\overline{\eta}_t \in \mathcal{B}_d(0,1)$). As a result, we can rewrite the previous expression as

$$p_t \geq \mathbb{P}\Big(f_t(\eta) \geq f_t(\overline{\eta}_t) \Big| \mathcal{F}_t, \widehat{E}_t\Big).$$

Since the function f_t inherits all the properties of J, notably its convexity in η, we know that the supremum on a convex closed set is reached at least at one point $\overline{\eta}_t$ and that it belongs to the boundary (see Prop. 4), which in our case corresponds to $\|\overline{\eta}_t\| = 1$. Moreover, let $\mathcal{H}_t(\overline{\eta}_t)$ be the hyperplane tangent to $\overline{\eta}_t$. $\mathcal{H}_t(\overline{\eta}_t)$ splits $\mathbb{R}^d$ in two complementary subspaces $\mathcal{G}_t$ and $\mathcal{G}_t^\perp$ where $\mathcal{G}_t$ does not contain the unit ball by convention. Again, the convexity of f_t ensures that $f_t(\eta) \geq f_t(\overline{\eta}_t)$ for all $\eta \in \mathcal{G}_t$ as proved in Prop. 5. As illustrated in Fig. 5 the probability of being optimistic is now reduced to the probability that η drawn from $\mathcal{D}^{\mathrm{TS}}$ falls into $\mathcal{G}_t$, which corresponds to

$$p_t \geq \mathbb{P}\Big(\eta \in \mathcal{G}_t \Big| \mathcal{F}_t, \widehat{E}_t\Big).$$

Let u_t be the vector defining the hyperspace $\mathcal{H}_t(\overline{\eta}_t)$, notice that the subspace u_t is entirely defined by the filtration $\mathcal{F}_t$ and the event $\widehat{E}_t$ and it is thus independent from $\overline{\eta}_t$. As a result, we obtain

$$p_t \geq \mathbb{P}\Big(u_t^\mathsf{T} \eta \geq 1 \Big| \mathcal{F}_t, \widehat{E}_t\Big) \geq p,$$

where the last step immediately follows from property 1 of Def. 1 of the TS sampling distribution.

Finally, we show that this property is not affected, up to a second order term, by the high-probability concentration event. It relies on the fact that the chosen confidence level $\delta' = \delta/4T$ is small compared to the anti-concentration probability p of Def. 1. For sake of simplicity, we assume that $T \geq 1/2p$ which implies that $\delta' \leq p/2$.
For any events A and B, one has

$$\mathbb{P}(A \cap B) = 1 - \mathbb{P}(A^c \cup B^c) \geq \mathbb{P}(A) - \mathbb{P}(B^c)$$

Applying the previous inequality to $A := \{J(\widetilde{\theta}) \geq J(\theta^\star)\}$ and $B := \{\widetilde{\theta} \in \mathcal{E}_t^{\mathrm{TS}}\}$ where $\mathcal{E}_t^{\mathrm{TS}} = \{\theta \in \mathbb{R}^d \mid \|\theta - \widehat{\theta}_t\|_{V_t} \leq \gamma_t(\delta')\}$ leads to

$$\mathbb{P}(\widetilde{\theta}_t \in \Theta^{\mathrm{opt}} \cap \mathcal{E}_t^{\mathrm{TS}} | \mathcal{F}_t, \hat{E}_t) \geq p - \delta' \geq p/2$$

□

Proof of Theorem 1. We first bound the two regret terms $R^{\mathrm{TS}}(T)$ and $R^{\mathrm{RLS}}(T)$.

Bound on $R^{\mathbf{TS}}(T)$. We collect the bounds on each term R_t^{TS} and obtain

$$R^{\mathrm{TS}}(T) \leq \sum_{t=1}^{T} R_t^{\mathrm{TS}} \mathbb{1}\{E_t\} \leq \frac{4\gamma_T(\delta')}{p} \sum_{t=1}^{T} \mathbb{E}\big[\|x^\star(\widetilde{\theta})\|_{V_t^{-1}} | \mathcal{F}_t\big]. \tag{11}$$

Since this term contains an expectation, we cannot directly apply Proposition 2 and we first need to rewrite to the total regret $R^{\mathrm{TS}}(T)$ as

$$R^{\mathrm{TS}}(T) \leq \frac{4\gamma_T(\delta')}{p}\bigg(\sum_{t=1}^{T} \|x_t\|_{V_t^{-1}} + \underbrace{\sum_{t=1}^{T}\Big(\mathbb{E}\big[\|x^\star(\widetilde{\theta})\|_{V_t^{-1}}|\mathcal{F}_t\big] - \|x_t\|_{V_t^{-1}}\Big)}_{R_2^{\mathrm{TS}}}\bigg). \tag{12}$$

From Prop. 2, the first term is bounded as,

$$\sum_{t=1}^{T} \|x_t\|_{V_t^{-1}} \leq \sqrt{T}\bigg(\sum_{t=1}^{T} \|x_t\|^2_{V_t^{-1}}\bigg)^{1/2} \leq \sqrt{2Td\log\big(1+\frac{T}{\lambda}\big)}.$$

We now proceed applying Azuma inequality 8 to the second term which is a martingale by construction. Under assumption 1, $\|x_t\| \leq 1$ for all $t \geq 1$, so since $V_t^{-1} \leq \frac{1}{\lambda}I$ one gets,

$$\mathbb{E}\big[\|x^\star(\widetilde{\theta})\|_{V_t^{-1}}|\mathcal{F}_t\big] - \|x_t\|_{V_t^{-1}} \leq \frac{2}{\sqrt{\lambda}}, \qquad a.s.$$

This provides an upper-bound on each element of R_2^{TS} which holds with probability at least $1-\frac{\delta}{2}$ as

$$R_2^{\mathrm{TS}} \leq \sqrt{\frac{8T}{\lambda}\log\frac{4}{\delta}}.$$

Bound on $R^{\mathbf{RLS}}(T)$. The bound on R^{RLS} is derived as previous results in [Abbasi-Yadkori et al., 2011b, Agrawal and Goyal, 2012b]. We decompose the term in a *sampling prediction error* and a RLS *prediction error* as follow

$$R^{\mathrm{RLS}}(T) \leq \sum_{t=1}^{T} |x_t^{\mathsf{T}}(\widetilde{\theta}_t - \widehat{\theta}_t)|\mathbb{1}\{E_t\} + \sum_{t=1}^{T} |x_t^{\mathsf{T}}(\widehat{\theta}_t - \theta^\star)|\mathbb{1}\{E_t\}$$

By definition of the concentration event E_t,

$$|x_t^{\mathsf{T}}(\widetilde{\theta}_t - \widehat{\theta}_t)|\mathbb{1}\{E_t\} \leq \|x_t\|_{V_t^{-1}}\gamma_t(\delta'), \qquad |x_t^{\mathsf{T}}(\widehat{\theta}_t - \theta^\star)|\mathbb{1}\{E_t\} \leq \|x_t\|_{V_t^{-1}}\beta_t(\delta'),$$

so from proposition 2,

$$R^{\mathrm{RLS}}(T) \leq \big(\beta_T(\delta') + \gamma_T(\delta')\big)\sqrt{2Td\log\big(1+\frac{T}{\lambda}\big)}. \tag{13}$$

Final bound. We finally plug everything together since from lemma 1 the concentration event holds with probability at least $1-\frac{\delta}{2}$. Using the bound on $R^{\mathrm{TS}}(T)$ and a union bound argument one obtains the desired result which holds with probability at least $1-\delta$. □

E Hyperspherical cap and beta function

Proposition 9. *Let $V_d(R)$ be the volume of the $d-$dimensional ball of radius R and let $V_d^{cap}(h)$ the volume of the hyperspherical cap of heigh $h = R - r > 0$. Then,*

$$V_d^{cap}(h) = \frac{1}{2}V_d(R)I_{1-(\frac{r}{R})^2}\left(\frac{d+1}{2},\frac{1}{2}\right)$$

where $I_x(a,b)$ is the incomplete regularized beta function.

Proof. The proof can be found in Li [2011]. □

Proposition 10. *Let $I_x(a,b)$ is the incomplete regularized beta function,*

$$\forall d \geq 2, \quad I_{1-\frac{1}{d}}\left(\frac{d+1}{2},\frac{1}{2}\right) \geq \frac{1}{8\sqrt{6\pi}}$$

Proof. The incomplete regularized beta function can be expressed in terms of the beta function $B(a,b)$ and the incomplete beta function $B_x(a,b)$ where

$$\begin{aligned}B_x(a,b) &= \int_0^x t^{a-1}(1-t)^{b-1}dt \\ B(a,b) &= B_1(a,b) \\ I_x(a,b) &= \frac{B_x(a,b)}{B(a,b)}\end{aligned}$$

Hence we seek for a lower bound on $B_{1-\frac{1}{d}}\left(\frac{d+1}{2}+\frac{1}{2}\right)$ and an upper bound for $B\left(\frac{d+1}{2}+\frac{1}{2}\right)$.

1. Let first find an lower bound for the incomplete beta function. Since $t \to t^{\frac{d-1}{2}}(1-t)^{-1/2}$ is positive and increasing on $[0,1]$, for any $d \geq 2$,

$$\begin{aligned}B_{1-\frac{1}{d}}\left(\frac{d+1}{2},\frac{1}{2}\right) &\geq \int_{1-\frac{3}{2d}}^{1-\frac{d}{2}} t^{\frac{d-1}{2}}(1-t)^{-1/2}dt \\ &\geq \frac{1}{2d}\left(\frac{3}{2d}\right)^{-1/2}(1-\frac{3}{2d})^{\frac{d-1}{2}} \\ &\geq \frac{1}{\sqrt{6d}}(1-\frac{3}{2d})^{\frac{d-1}{2}} \\ &\geq \frac{1}{\sqrt{6d}}(1-\frac{3}{2d})^{\frac{d}{2}}\end{aligned}$$

From the increasing property of $x \to (1-\frac{\alpha}{x})^x$ for any $\alpha < 1$ the sequence $\left\{(1-\frac{3}{2d})^{\frac{d}{2}}\right\}_{d\geq 2}$ is increasing and

$$B_{1-\frac{1}{d}}\left(\frac{d+1}{2},\frac{1}{2}\right) \geq \frac{1}{\sqrt{6d}}(1-\frac{3}{2\times 2})^{\frac{2}{2}} = \frac{1}{4\sqrt{6d}}$$

2. Now we seek for an upper bound for $B\left(\frac{d+1}{2}+\frac{1}{2}\right)$. Since $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ one has:

$$B\left(\frac{d+1}{2}+\frac{1}{2}\right) = \frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}+1\right)} = \sqrt{\pi}\frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}+1\right)}$$

From Chen and Qi [2005] we have the following inequalities for the gamma function $\forall n \geq 1$:

$$\begin{aligned}\frac{\Gamma(n+1/2)}{\Gamma(n+1)} &\leq (n+1/4)^{-1/2} \\ \frac{\Gamma(n+1/2)}{\Gamma(n+1)} &\geq (n+4/\pi-1)^{-1/2}\end{aligned}$$

Together with $\Gamma(x+1) = x\Gamma(x)$ and treating separately cases where d is even or not, one gets $\forall d \geq 2$

$$\frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}+1\right)} \leq \frac{2}{\sqrt{d}}$$

3. Using the obtained upper and lower bound we get:

$$I_{1-\frac{1}{d}}\left(\frac{d+1}{2},\frac{1}{2}\right) \geq \frac{\sqrt{d}}{2\sqrt{\pi}\times 4\sqrt{6d}} \geq \frac{1}{8\sqrt{6\pi}}$$

$\square$

F Generalized Linear Bandit

We present here how to apply our derivation to the generalized linear bandit (GLM) problem of Filippi et al. [2010]. The regret bound is obtained by basically showing that the GLM problem can be reduced to studying the linear case.

The setting. Let $\mathcal{X} \subset \mathbb{R}^d$ be an arbitrary (finite or infinite) set of arms. Every time an arm $x \in \mathcal{X}$ is pulled, a reward is generated as $r(x) = \mu(x^\mathsf{T}\theta^\star) + \xi$, where μ is the so-called *link function*, $\theta^\star \in \mathbb{R}^d$ is a fixed but unknown parameter vector and ξ is a random zero-mean noise. The value of an arm $x \in \mathcal{X}$ is evaluated according to its expected reward $\mu(x^\mathsf{T}\theta^\star)$ and for any parameter $\theta \in \mathbb{R}^d$ we denote the optimal arm and its optimal value as

$$x^\star(\theta) = \arg\max_{x\in\mathcal{X}} \mu(x^\mathsf{T}\theta), \qquad J^{\mathrm{GLM}}(\theta) = \sup_{x\in\mathcal{X}} \mu(x^\mathsf{T}\theta). \tag{14}$$

Then $x^\star = x^\star(\theta^\star)$ is the optimal arm associated with the true parameter $\theta^\star$ and $J^{GLM}(\theta^\star)$ its optimal value. At each step t, a learner chooses an arm $x_t \in \mathcal{X}$ using all the information observed so far (i.e., sequence of arms and rewards) but without knowing $\theta^\star$ and $x^\star$. At step t, the learner suffers an *instantaneous regret* corresponding to the difference between the expected rewards of the optimal arm $x^\star$ and the arm x_t played at time t. The objective of the learner is to minimize the *cumulative regret* up to a finite step T,

$$R^{\mathrm{GLM}}(T) = \sum_{t=1}^{T} \big(\mu(x^{\star,\mathsf{T}}\theta^\star) - \mu(x_t^\mathsf{T}\theta^\star)\big). \tag{15}$$

Assumptions. The assumptions associated with this more general problem are the same as in the linear bandit problem plus one regarding the link function. Formally, we require assumption 1, 2 and 3 and add:

Assumption 4 (link function)**.** *The link function $\mu : \mathbb{R} \to \mathbb{R}$ is continuously differentiable, Lipschitz with constant k_μ and such that $c_\mu = \inf_{\theta\in\mathbb{R}^d, x\in\mathcal{X}}(x^\mathsf{T}\theta) > 0$.*

Technical tools. Let $(x_1, \ldots, x_t) \in \mathcal{X}^t$ be a sequence of arms and $(r_2, \ldots, r_{t+1})$ be the corresponding observed (random) rewards, then the unknown parameter $\theta^\star$ can be estimated by GLM estimator. Following Filippi et al. [2010] one gets, for any regularization parameter $\lambda \in \mathbb{R}^+$,

$$\widehat{\theta}_t^{\mathrm{GLM}} = \arg\min_{\theta\in\mathbb{R}^d} \| \sum_{s=1}^{t-1} \big(r_{s+1} - \mu(x_s^\mathsf{T}\theta)\big)x_s\|^2_{V_t^{-1}}, \tag{16}$$

where V_t is the same design matrix as in the linear case. Similar to Prop. 1, we have a concentration inequality for the GLM estimate.

Proposition 11 (Prop. 1 in appendix.A in Filippi et al. [2010])**.** *For any $\delta \in (0,1)$, under assumptions 1, 2, 3 and 4, for any $\mathcal{F}_t^x$-adapted sequence $(x_1, \ldots, x_t, \ldots)$, the prediction returned by the GLM estimator $\widehat{\theta}_t^{GLM}$ (Eq. 16) is such that for any fixed $t \geq 1$,*

$$\|\widehat{\theta}_t^{GLM} - \theta^\star\|_{V_t} \leq \frac{\beta_t(\delta)}{c_\mu}, \tag{17}$$

and

$$\begin{aligned} \forall x \in \mathbb{R}^d, \quad \|\mu(x^\mathsf{T}\widehat{\theta}_t^{GLM}) - \mu(x^\mathsf{T}\theta^\star)\| &\leq \frac{k_\mu\beta_t(\delta)}{c_\mu}\|x\|_{V_t^{-1}}, \\ \|x^\mathsf{T}\widehat{\theta}_t^{GLM} - x^\mathsf{T}\theta^\star\| &\leq \frac{\beta_t(\delta)}{c_\mu}\|x\|_{V_t^{-1}}, \end{aligned} \tag{18}$$

with probability $1-\delta$ (w.r.t. the noise sequence $\{\xi_t\}_t$ and any other source of randomization in the definition of the sequence of arms), where $\beta_t(\delta)$ is defined as in Eq. 5.

The Asm. 4 on the link function together with the properties of the GLM estimator implies the following:

1. since the first derivative is strictly positive, μ is strictly increasing and $x^\star(\theta) = \arg\max_{x\in\mathcal{X}} x^\mathsf{T}\theta$ so we retrieve the optimal arm of the linear case (and the support function),

2. the concentration inequality of the GLM estimate involves the same ellipsoid as for the RLS (multiplied by a factor $\frac{1}{c_\mu}$).

These two facts suggest to use then exactly the same TS algorithm as for the linear case (with a β multiplied by a factor $\frac{1}{c_\mu}$).

Sketch of the proof. From the previous comments, making use of the property of μ, one just need to reduce the GLM case to the standard linear case.

$$\begin{aligned}
R^{\mathrm{GLM}}(T) &= \sum_{t=1}^{T} \big(\mu(x^\star\theta^\star) - \mu(x_t^\mathsf{T}\theta^\star)\big), \\
&= \sum_{t=1}^{T} \big(\mu(x^\star\theta^\star) - \mu(x_t^\mathsf{T}\tilde{\theta}_t)\big) + \sum_{t=1}^{T} \big(\mu(x_t^\mathsf{T}\tilde{\theta}_t) - \mu(x_t^\mathsf{T}\theta^\star)\big) \\
&\leq \sum_{t=1}^{T} \big(\mu(x^\star\theta^\star) - \mu(x_t^\mathsf{T}\tilde{\theta}_t)\big) + \sum_{t=1}^{T} k_\mu \|x\|_{V_t^{-1}} \|\tilde{\theta}_t - \theta^\star\|_{V_t}.
\end{aligned}$$

The second term is bounded exactly as $R^{\mathrm{RLS}}(T)$. To bound the first one, we make use of the fact that

$$\begin{aligned}
&\mu(x^\star\theta^\star) - \mu(x_t^\mathsf{T}\tilde{\theta}_t) \leq k_\mu\big(J(\theta^\star) - J(\tilde{\theta}_t)\big), \quad \text{if} J(\theta^\star) - J(\tilde{\theta}_t) \geq 0, \\
&\mu(x^\star\theta^\star) - \mu(x_t^\mathsf{T}\tilde{\theta}_t) \leq c_\mu\big(J(\theta^\star) - J(\tilde{\theta}_t)\big), \quad \text{otherwise.}
\end{aligned}$$

Following the proof of the linear case, with high probability, for all $t \geq 1$,

$$J(\theta^\star) - J(\tilde{\theta}_t) \leq \frac{2\gamma_t(\delta')}{c_\mu p}\mathbb{E}\big(\|x_t\|_{V_t^{-1}}|\mathcal{F}_t\big).$$

Since the r.h.s is strictly positive one can bound the first part of the regret, independently of the sign by,

$$\sum_{t=1}^{T} \big(\mu(x^\star\theta^\star) - \mu(x_t^\mathsf{T}\tilde{\theta}_t)\big) \leq \frac{2k_\mu\gamma_T(\delta')}{c_\mu p} \sum_{t=1}^{T} \mathbb{E}\big(\|x_t\|_{V_t^{-1}}|\mathcal{F}_t\big).$$

Finally, the same proof as in the linear case leads to the following bound for the Generalized Linear Bandit regret.

Lemma 4. *Under assumptions 1,2,3 and 4, the cumulative regret of* TS *over T steps is bounded as*

$$R^{GLM}(T) \leq \frac{k_\mu}{c_\mu}\big(\beta_T(\delta') + \gamma_T(\delta')(1 + 2/p)\big)\sqrt{2Td\log\left(1 + \frac{T}{\lambda}\right)} + \frac{2k_\mu\gamma_T(\delta')}{pc_\mu}\sqrt{\frac{8T}{\lambda}\log\frac{4}{\delta}} \tag{19}$$

with probability $1 - \delta$ *where* $\delta' = \frac{\delta}{4T}$.

G Regularized Linear Optimization

We consider here the Regularized Linear Optimization (RLO) problem as an extension of the Linear Bandit problem. Given a set of arms $\mathcal{X} \subset \mathbb{R}^d$ and an unknown parameter $\theta^\star \in \mathbb{R}^d$, a learner aims at each time step $t = 1, \ldots, T$ to select action $x_t \in \mathcal{X}$ which maximizes its associated reward $x_t^\mathsf{T}\theta^\star + \mu c(x_t)$ where μ is a known constant and c an arbitrary (yet known) real-valued function. Whenever arm x is pulled, the learner receives a noisy observation $y = x^\mathsf{T}\theta^\star + \xi$. As for LB, we introduce the function $f(x;\theta) = x^\mathsf{T}\theta + \mu c(x)$, and denote as $x^\star(\theta) = \arg\max_{x\in\mathcal{X}} f(x;\theta)$ and $J(\theta) = \max_{x\in\mathcal{X}} f(x;\theta)$ the optimal action and optimal reward associated with θ. The regret is therefore defined as $R^{RLO}(T) = \sum_{t=1}^{T} f(x^\star(\theta^\star);\theta^\star) - f(x_t;\theta^\star)$.
Since this problem is just the regularized extension of the Linear Bandit, the TS algorithm is similar to Alg. 1 where r_t is replaced y_t and $x_t = \arg\max_{x\in\mathcal{X}} f(x, \widetilde{\theta}_t)$. Under the same assumptions, the regret shares the same bound and our line of proof holds. First, we decompose the regret

$$R(T) = \sum_{t=1}^{T} \left[(f(x^\star(\theta^\star);\theta^\star) - f(x_t;\widetilde{\theta}_t)) + (f(x_t;\widetilde{\theta}_t) - f(x_t;\theta^\star))\right] = \underbrace{\sum_{t=1}^{T} \left[J(\theta^\star) - J(\widetilde{\theta}_t)\right]}_{=R^{\mathrm{TS}}(T)} + \underbrace{\sum_{t=1}^{T} \left[x_t^\mathsf{T}\widetilde{\theta}_t - x_t^\mathsf{T}\theta^\star\right]}_{=R^{\mathrm{RLS}}(T)}.$$

Since Prop. 1 holds thanks to the linear observations y_t, $R^{\mathrm{RLS}}(T)$ is bounded as in the LB. Finally, to bound $R^{\mathrm{TS}}(T)$, one just need to ensure that Prop. 3, Lem. 2 and Lem. 3 hold.
The convexity of the function f with respect to θ implies the convexity of J: $\forall x \in \mathcal{X}$, $\forall \theta, \theta' \in \mathbb{R}^d$, $\forall \alpha \in (0,1)$,

$$J(\alpha\theta + (1-\alpha)\theta') = \max_{x \in \mathcal{X}} f(x; \alpha\theta + (1-\alpha)\theta') \leq \max_{x \in \mathcal{X}} \big(\alpha f(x;\theta) + (1-\alpha) f(x;\theta')\big) \leq \alpha J(\theta) + (1-\alpha) J(\theta').$$

Then, J is real-valued and convex which implies its continuous differentiability thanks to Alexandrov's theorem. As a consequence, the first step of the proof holds.
The equality between the gradient $\nabla J(\theta)$ and the optimal arm $x^\star(\theta)$ can be derived as in Prop. 7: for any $\theta, \bar\theta \in \mathbb{R}^d$, by definition, $J(\theta) = f(x^\star(\theta);\theta)$ and $J(\bar\theta) \geq f(x^\star(\theta);\bar\theta)$. Then,

$$\begin{aligned}
&J(\bar\theta) - f(x^\star(\theta), \bar\theta) \geq 0 := J(\theta) - f(x^\star(\theta), \theta), \\
&J(\bar\theta) \geq J(\theta) + f(x^\star(\theta), \bar\theta) - f(x^\star(\theta), \theta) = J(\theta) + x^\star(\theta)^\mathsf{T}\left(\bar\theta - \theta\right), \quad \forall \bar\theta \in \mathbb{R}^d,
\end{aligned}$$

which is the definition of the sub-gradient. Finally, the almost everywhere differentiability of J ensures the sub-gradient to be a singleton and hence equals the gradient. Therefore, Lem. 2 holds and so is step 2.
Finally, since the optimism just relies on the convexity of J and on the over-sampling, it is satisfied in the RLO and step 3 holds. As a result, we obtain the same regret bound as in the LB.