Homework 2

CSE 541: Interactive Learning
Instructor: Kevin Jamieson
Due 11:59 PM on February 19, 2022

Confidence bounds

1.1 (Wald's identity) Let $X_1, X_2, \dots$ be a sequence of iid random variables. For $j \in \{0,1\}$, under $\mathbf{H}_j$ we have that $X_i \sim p_j$. Let $\mathbb{P}_j(\cdot), \mathbb{E}_j(\cdot)$ denote the probability and expectation under $\mathbf{H}_j$. Assume that the support of p_0 and p_1 are equal and furthermore, that $\sup_{x \in \text{support}(p_0)} \frac{p_1(x)}{p_0(x)} \leq \kappa$. Fix some $\delta \in (0,1)$. If $L_t = \prod_{i=1}^t \frac{p_1(X_i)}{p_0(X_i)}$ and $\tau = \min\{t : L_t > 1/\delta\}$, we showed in class that the false alarm probability $\mathbb{P}_0(L_\tau > 1/\delta) \leq \delta$. Assume that $\mathbb{E}_1[\tau] < \infty$. Show that $\frac{\log(1/\delta)}{KL(p_1||p_0)} \leq \mathbb{E}_1[\tau] \leq \frac{\log(\kappa/\delta)}{KL(p_1||p_0)}$ where $KL(p_1|p_0) = \int p_1(x) \log(\frac{p_1(x)}{p_0(x)}) dx$ is the Kullback Leibler divergence between p_1 and p_0.

1.2 (Method of Mixtures) Let $X_1, X_2, \dots$ be a sequence of iid random variables where $X_1 \sim \mathcal{N}(\mu, 1)$. Under $\mathbf{H}_0$ we have $\mu = 0$ and under $\mathbf{H}_1$ assume $\mu = \theta$.

- Define $L_t(\theta) := \exp(S_t \theta - t\theta^2/2)$ where $S_t = \sum_{s=1}^t X_s$. Show that $\prod_{i=1}^t \frac{p_1(X_i)}{p_0(X_i)} = L_t(\theta)$.
- Now suppose the sequence $X_1, X_2, \dots$ is still iid and $\mathbb{E}[X_1] = \mu$ in the above notation, but now the distribution of X_1 is unknown other than the knowledge that $\mathbb{E}[\exp(\lambda(X_1 - \mu))] \leq e^{\lambda^2/2}$. Show that $L_t(\theta)$ defined above is a super-martingale under $\mathbf{H}_0$.
- Assume the setting of the previous step. Define $\bar{L_t} = \int_\theta L_t(\theta) h(\theta) d\theta$ where $h(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\theta^2}{2\sigma^2}}$ and $\sigma > 0$. Show that $\bar{L_t}$ is a super-martingale under $\mathbf{H}_0$.
- Fix $\delta \in (0,1)$. Show that $\mathbb{P}_0(\exists t \in \mathbb{N} : \bar{L_t} > 1/\delta) \leq \delta$.
- Conclude that if $Z_1, Z_2, \dots$ are iid random variables with $\mathbb{E}[\exp(\lambda(Z_1 - \mathbb{E}[Z_1])] \leq \exp(\lambda^2/2)$, then for any $\sigma > 0$

$$\mathbb{P}\left(\exists t \in \mathbb{N} : |\frac{1}{t}\sum_{s=1}^t (Z_s - \mathbb{E}[Z_s])| > \sqrt{1 + \frac{1}{t\sigma^2}} \sqrt{\frac{2\log(1/\delta) + \log(t\sigma^2 + 1)}{t}} \right) \leq \delta. \tag{1}$$

1.3 (Best arm identification) Consider an n-armed multi-armed bandit problem where the jth pull of the ith arm yields a random variable $X_{i,j} \sim \mathcal{N}(\theta_i, 1)$. The objective of the player is to strategically pull arms until getting to a point that they can predict the index of the arm with the highest mean, at which time they stop and output this estimated arm. Suppose an oracle told the player that θ, the true means they are playing against, is equal to $\Delta \mathbf{e}_j$ for some $j = 1, \dots, n$ where $\mathbf{e}_j$ is a vector of all zeros except a 1 in the jth location. Note that while $\Delta > 0$ is known to the player, which is the true j is unknown.

- Due to the symmetry of the known problem setup, it is conceivable that the following algorithm is optimal: play every arm the same number of times (say, τ times), and then declare that the arm with the highest empirical mean is best. Provide a sufficient condition on τ such that such a procedure correctly identifies the location of the true j index with probability at least $1 - \delta$ (don't forget the union bound!).
- Argue that if each arm is pulled the same number of times this value of τ up to a constant factor (ignoring dependence on δ) is necessary. Hint[1]
- The previous two parts suggest that any algorithm that pulls every arm the same number of times and identifies the best arm requires essentially $n\Delta^{-2}\log(n/\delta)$ total pulls. We will beat this with an adaptive procedure that requires just $O(n\Delta^{-2}\log(1/\delta))$ pulls.

  Algorithm: Initialize $S_1 = [n]$. At each around $t \geq 1$, while $|S_t| > 1$, pull every arm in S_t and then set $S_{t+1} = \{i \in S_t : \sum_{s=1}^t (X_{i,s} - \Delta) \geq -\Delta t/2 - \frac{\log(1/\delta)}{\Delta}\}$.

[1] If $Z_i \sim \mathcal{N}(0, \sigma^2)$ for $i = 1, \dots, n$ show that $\mathbb{P}(\max_{i=1,\dots,n} Z_i \geq \sqrt{\sigma^2 \log(n)}) > c$ for some absolute constant c.

We showed in class that if $Z_s \sim \mathcal{N}(0,1)$ then for any $\alpha > 0$ and $\rho \in (0,1)$ we have

$$\max\left\{\mathbb{P}\left(\bigcup_{t=1}^{\infty}\{\sum_{s=1}^{t} Z_s < -\frac{\alpha t}{2} - \frac{\log(1/\rho)}{\alpha}\}\right), \mathbb{P}\left(\bigcup_{t=1}^{\infty}\{\sum_{s=1}^{t} Z_s > \frac{\alpha t}{2} + \frac{\log(1/\rho)}{\alpha}\}\right)\right\} \leq \rho. \tag{2}$$

Conclude that if j is the index of the best-arm (so that $X_{j,s} \sim \mathcal{N}(\Delta, 1)$) then with probability at least $1-\delta$ arm j remains in S_t for all $t \geq 1$.

- For $i \neq j$ (so that $X_{i,s} \sim \mathcal{N}(0,1)$) define the random variables

  $$\rho_i := \sup\left\{\rho \in (0,1) : \bigcap_{t=1}^{\infty}\{\sum_{s=1}^{t} X_{i,s} \leq \frac{\Delta t}{4} + \frac{\log(1/\rho)}{\Delta/2}\}\right\}.$$

  If $T_i = \max\{t : i \in S_t\}$ show that $T_i \leq 4\Delta^{-2}\log(1/\delta) + 8\Delta^{-2}\log(1/\rho_i)$.

- Note that by (2) we have $\mathbb{P}(\rho_i \leq \rho) \leq \rho$ for any $\rho \in (0,1)$. Use this fact to show that $\mathbb{E}\left[\sum_{i=1}^{n} T_i\right] \leq cn\Delta^{-2}\log(1/\delta)$ for some absolute constant $c > 0$. While not necessary for this problem, it is also possible to show that the right hand side holds with probability at least $1-\delta$ (see sub-Gamma random variables).

In general, when the means are unknown, curved boundaries with a UCB-like algorithm [1] can be used to identify the best arm using just $O(\log(1/\delta)\sum_{i=2}^{n}\Delta^{-2}\log(\log(\Delta^{-2})))$ total pulls, where $\Delta_i = \max_j \theta_j - \theta_i$ using a similar analysis.

Linear regression and experimental design
2.1 Exercise 20.2 of [SzepesvariLattimore]

Non-parametric bandits
4.1 Let $\mathcal{F}_{Lip}$ be a set of functions defined over $[0,1]$ such that for each $f \in \mathcal{F}_{Lip}$ we have $f : [0,1] \to [0,1]$ and for every $x, y \in [0,1]$ we have $|f(y) - f(x)| \leq L|y-x|$ for some known $L > 0$. At each round t the player chooses an $x_t \in [0,1]$ and observes a random variable $y_t \in [0,1]$ such that $\mathbb{E}[y_t] = f_*(x_t)$ where $f_* \in \mathcal{F}_{Lip}$. Define the regret of an algorithm after T steps as $R_T = \mathbb{E}\left[\sum_{t=1}^{T} f_*(x_\star) - f_*(x_t)\right]$ where $x_\star = \arg\max_{x\in[0,1]} f_*(x)$.

- Propose an algorithm, that perhaps uses knowledge of the time horizon T, that achieves $R_T \leq O(T^{2/3})$ regret (Okay to ignore constant, log factors).
- Argue that this is minimax optimal (i.e., improvable in general through the use of an explicit example, with math, but no formal proof necessary).

Experiments
5.1 Suppose we have random variables $Z_1, Z_2, \ldots$ that are iid with $\mathbb{E}[\exp(\lambda(Z_1 - \mathbb{E}[Z_1])] \leq \exp(\lambda^2/2)$. Then for any *fixed* $t \in \mathbb{N}$ we have the standard tail bound

$$\mathbb{P}\left(|\frac{1}{t}\sum_{s=1}^{t}(Z_s - \mathbb{E}[Z_s])| > \sqrt{\frac{2\log(2/\delta)}{t}}\right) \leq \delta. \tag{3}$$

The bound in (1) holds for all $t \in \mathbb{N}$ simultaneously (i.e., not for just a fixed t) at the cost of a slightly inflated bound. Let $\delta = 0.05$. Plot the *ratio* of the confidence bound of (1) to (3) as a function of t for values of $\sigma^2 \in \{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}\}$. What do you notice about where the ratio is smallest with respect to σ^2 (if the smallest value is at the edge, your t is not big enough–it should be in the many millions)? Suppose you are observing a stream of iid Gaussian random variables $Z_1, Z_2, \ldots$ and you are trying to determine whether their mean is positive or negative. If someone tells you they think the absolute value of the mean is about .01, how would you choose σ^2 and use the above confidence bound to perform this test using as few total observations as possible?

5.2 **G-optimal design** This problem addresses finding a G-optimal design. Fix $(x_1, \dots, x_n) \subset \mathbb{R}^d$. Let $\triangle_n = \{p \in \mathbb{R}^n : \sum_{i=1}^n p_i = 1, p_i \geq 0 \, \forall i \in [n]\}$ and $A(\lambda) = \sum_{j=1}^n \lambda_j x_j x_j^\top$. For some $\lambda \in \triangle_n$ let $f(\lambda) = \max_{i=1,\dots,n} \|x_i\|^2_{A(\lambda)^{-1}}$ and $g(\lambda) = -\log(|A(\lambda)|)$ (these are the G and D optimal designs, respectively). We wish to find a discrete allocation of size N for G-optimal design. Below are strategies to output an allocation $(I_1, \dots, I_N) \in [n]$. Choose **one** strategy and **one** $h \in \{f, g\}$ to obtain samples from a G-optimal design

1. **Greedy** For $i = 1, \dots, 2d$ select I_i uniformly at random with replacement from $[n]$. For $t = 2d + 1, \dots, N$ select $I_t = \arg\min_{k \in [n]} h(x_k x_k^\top + \sum_{j=1}^{t-1} x_{I_j} x_{I_j}^\top)$. Return $(I_1, \dots, I_N)$.

2. **Frank Wolfe** For $i = 1, \dots, 2d$ select I_i uniformly at random with replacement from $[n]$ and let $\lambda^{(2d)} = \frac{1}{2d} \sum_{i=1}^{2d} \mathbf{e}_{I_i}$. For $k = 2d + 1, \dots, N$ select $I_k = \arg\min_{j \in [n]} \frac{\partial h(\lambda)}{\partial \lambda_j}|_{\lambda = \lambda^{(k-1)}}$ and set $\lambda^{(k)} = (1 - \eta_k)\lambda^{(k-1)} + \eta_k \mathbf{e}_{I_k}$ where $\eta_k = 2/(k+1)$. Return $(I_1, \dots, I_N)$. You will need to derive the partial gradients $\frac{\partial h(\lambda)}{\partial \lambda_i}$. Hint[2].

Let $\widehat{\lambda} = \frac{1}{N} \sum_{i=1}^N \mathbf{e}_{I_i}$. We are going to evaluate $f(\widehat{\lambda})$ in a variety of settings. For $a \geq 0$ and $n \in \mathbb{N}$ let $x_i \sim \mathcal{N}(0, \text{diag}(\sigma^2))$ with $\sigma_j^2 = j^{-a}$ for $j = 1, \dots, d$ with $d = 10$. On a single plot with $n \in \{10 + 2^i\}_{i=1}^{10}$ on the x-axis, plot your chosen method for $a \in \{0, 0.5, 1, 2\}$ as separate lines with $N = 1000$.

5.3 **Experimental design for function estimation**.
Let $f : [0,1]^d \to \mathbb{R}$ be some unknown function and fix some locations of interest $\mathcal{X} = \{x_i\}_{i=1}^n$. We wish to output an estimate $\widehat{f}$ of f uniformly well over $\mathcal{X}$ in the sense that $\max_{x \in \mathcal{X}} \mathbb{E}[(\widehat{f}(x) - f(x))^2]$ is small. To build such an estimate, we can make $N = 1000$ measurements. If we measure the function at location $X \in \mathcal{X}$ we assume we observe $Y = f(X) + \eta$ where $\eta \sim \mathcal{N}(0,1)$. While you can choose any N measurement locations in X you'd like (with repeats) you have to choose all locations before the observations are revealed. This problem emphasizes that even though we're studying linear functions and linear bandits in class, this can encode incredibly rich functions using non-linear transformations.

1. **Linear functions** Assume $f(x) = \langle x, \theta_* \rangle$ for some $\theta_* \in \mathbb{R}^d$. Propose a strategy to select your N measurements and fit $\widehat{f}$. What guarantees can you make about $\max_{x \in \mathcal{X}} \mathbb{E}[(\widehat{f}(x) - f(x))^2]$? What about the random quantity $\max_{x \in \mathcal{X}} (\widehat{f}(x) - f(x))^2$?

2. **Sobolev functions** For simplicity, let $d = 1$ and assume $f(x)$ is absolutely continuous and $\int_0^1 |f'(x)|^2 dx < \infty$. This class of functions is known as the First-order Sobolev space[3]. Remarkably, this set of functions is equivalent to a reproducing kernel Hilbert space (RKHS) for the kernel $k(x, x') = 1 + \min\{x, x'\}$. While a discussion of all its properties are beyond the scope of this class (see [2, Ch.12] for an excellent treatment) it follows that there exists a sequence of orthogonal functions $\phi(x) := (\phi_1(x), \phi_2(x), \phi_3(x), \dots)$ such that each $\phi_k : [0,1] \to \mathbb{R}$ and

   - for all $i, j \in \mathbb{N}$ we have $\int_0^1 \phi_i(x)\phi_j(x) dx = \beta_i \mathbf{1}\{i = j\}$ for a non-increasing sequence $\{\beta_i\}_i$
   - $k(x, x') = \langle \phi(x), \phi(x') \rangle := \sum_{k=1}^\infty \phi(x)_k \phi_k(x')$,
   - and any function f in this class can be written as $f(x) = \sum_{k=1}^\infty \alpha_k \phi_k(x)$ with $\sum_{k=1}^\infty \alpha_k^2 / \beta_k < \infty$.

   Because $\phi(x)$ could be infinite dimensional, its not immediately clear why this is convenient. For finite datasets though (i.e., $n < \infty$) there always exists a finite dimensional $\phi : \mathcal{X} \to \mathbb{R}^{|\mathcal{X}|}$ where $k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$. Precisely, if one computes the matrix $[\mathbf{K}]_{i,j} = k(x_i, x_j)$ and $\mathbf{K} = \sum_{k=1}^n v_k v_k^\top e_k$ is the resulting eigenvalue decomposition where $\{e_k\}_k$ is a non-increasing sequence, then for all $i \in [n]$ we have $[\phi_i]_j := [\phi(x_i)]_j := v_{i,j} \sqrt{e_j}$ and there exists a $\theta \in \mathbb{R}^n$ such that

$$f(x_i) = \langle \theta, \phi_i \rangle = \sum_{j=1}^n \theta_j v_{i,j} \sqrt{e_j}.$$

[2] $I = A(\lambda)A(\lambda)^{-1}$. Thus, $0 = \frac{\partial}{\partial \lambda_i} I = \left(\frac{\partial}{\partial \lambda_i} A(\lambda)\right) \cdot A(\lambda)^{-1} + A(\lambda) \cdot \left(\frac{\partial}{\partial \lambda_i} A(\lambda)^{-1}\right)$. $\frac{d}{dX} \log(|X|) = (A^{-1})^T$.

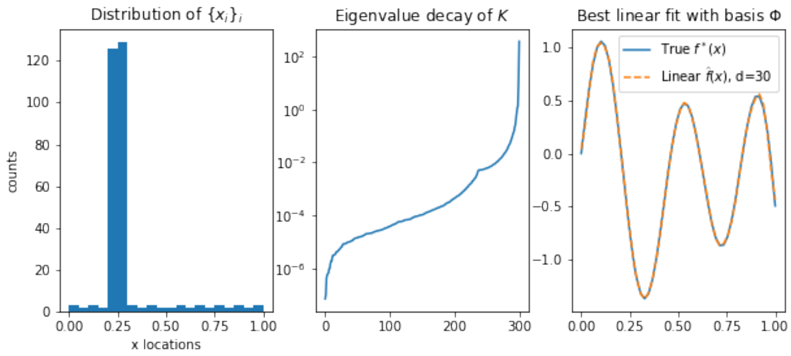
[3] This class is quite rich including functions as diverse as all polynomials $a + bx + cx^8$, trigonometric functions $\sin(ax) + \cos(bx)$, exponential functions $e^{ax} + b^x$, and even non-smooth functions like $\max\{1 - x, 3x\}$. However, it does not include functions like $1/x$ or $\sqrt{x}$ because their derivatives blow up at $x = 0$

Because the e_j are rapidly going to zero, we usually do not include all n, but cut the sum off at some $d \ll n$ so that $\phi_i \in \mathbb{R}^d$ and $f(x_i) \approx \sum_{j=1}^{d} \theta_j v_{i,j} \sqrt{e_j}$. To summarize, for each $i \in [n]$ we have defined a map $x_i \to \phi_i$ with $\phi_i \in \mathbb{R}^d$ and $f(x) \approx \langle \phi_i, \theta_* \rangle$ for some unknown θ_*. Here is a piece of code that generates our set $\mathcal{X} = \{x_i\}_{i=1}^n \subset [0,1]$ and these features $\Phi = \{\phi_i\}_{i=1}^n \subset \mathbb{R}^d$:

```python
import numpy as np
n=300
X = np.concatenate( (np.linspace(0,1,50), 0.25+ 0.01*np.random.randn(250) ), 0)
X = np.sort(X)

K = np.zeros((n,n))
for i in range(n):
    for j in range(n):
        K[i,j] = 1+min(X[i],X[j])
e, v = np.linalg.eigh(K) # eigenvalues are increasing in order
d = 30
Phi = np.real(v @ np.diag(np.sqrt(np.abs(e))) )[:,(n-d)::]
```

Let's define some arbitrary function and see how well this works!

```python
def f(x):
    return -x**2 + x*np.cos(8*x) + np.sin(15*x)

f_star = f(X)

theta = np.linalg.lstsq( Phi, f_star, rcond=None )[0]
f_hat = Phi @ theta
```



Observe that there is nearly no error in the reconstruction even though we're using a linear model in $\mathbb{R}^{30}$. This is because we defined a very good basis $\Phi \subset \mathbb{R}^d$. We could have achieved the same result by learning in kernel space and adding regularization (this is known as *Bayesian experimental design*). Instead of choosing d we would have to choose the amount of regularization, so there is still always a hyperparameter to choose.

Let's get back to estimating f from noisy samples. Now that we have made this a linear estimation problem, we are exactly in the setting of the first part of this problem. Consider the below listing:

```python
def observe(idx):
    return f(X[idx]) + np.random.randn(len(idx))

def sample_and_estimate(X, lbda, tau):
    n, d = X.shape
    reg = 1e-6 # we can add a bit of regularization to avoid divide by 0
    idx = np.random.choice(np.arange(n),size=tau,p=lbda)
    y = observe(idx)

    XtX = X[idx].T @ X[idx]
    XtY = X[idx].T @ y

    theta = np.linalg.lstsq( XtX + reg*np.eye(d), XtY, rcond=None )[0]
    return Phi @ theta, XtX

T = 1000

lbda = G_optimal(Phi)
f_G_Phi, A = sample_and_estimate(Phi, lbda, T)
conf_G = np.sqrt(np.sum(Phi @ np.linalg.inv(A) * Phi,axis=1))

lbda = np.ones(n)/n
f_unif_Phi, A = sample_and_estimate(Phi, lbda, T)
conf_unif = np.sqrt(np.sum(Phi @ np.linalg.inv(A) * Phi,axis=1))
```

Use your implementation of `G_optimal` from the previous problem and use the `sample_and_estimate` function given above. Your tasks (create a legend for all curves, label all axes, and provide a title for all plots):

(a) Plot 1: x-axis should be the x locations in $[0, 1]$. First line is the CDF of the uniform distribution over $\mathcal{X}$. The second line is the G-optimal allocation over $\mathcal{X}$. Comment on the relative shapes, and how this relates to the distribution over the x's shown in the left-most plot above.

(b) Plot 2: x-axis should be the x locations in $[0, 1]$. Plot `f_star`, `f_G_Phi`, `f_unif_Phi`.

(c) Plot 3: x-axis should be the x locations in $[0, 1]$. First line is the absolute value of `f_G_Phi` minus `f_star`, second line is `f_unif_Phi` minus `f_star`, third line is $\sqrt{d/n}$, fourth line is `conf_G`, fifth line is `conf_unif`. Comment on what these lines have to do with each other.

5.4 **Linear bandits** Implement the elimination algorithm (use your implementation of G-optimal design), UCB, and Thompson sampling (see listings below). Use the precise setup as the previous problem and set $\mathcal{X} \leftarrow \{\phi_i\}_{i=1}^n$. For $T = 40,000$, on a single plot with $t \in [T]$ on the x-axis, plot $R_t = \max_{x\in\mathcal{X}} \sum_{s=1}^t f(x) - y_t$ with a line for each algorithm. Comment on the results–what algorithm would you recommend to minimize regret?

```
Elimination algorithm
Input: $T \in \mathbb{N}$
Initialize $\tau = 100$, $\delta = 1/T$, $\gamma = 1$, $U = 1$ (supposed to be an upper bound on $\|\theta_*\|_2$), $V_0 = \gamma I$, $S_0 = 0$, $\widehat{\mathcal{X}}_0 = \mathcal{X}$
for: $k = 1, \ldots, \lfloor T/\tau \rfloor$
    $\lambda^{(k)} = \arg\min_{\lambda \in \triangle_{\widehat{\mathcal{X}}_k}} \max_{x \in \widehat{\mathcal{X}}_k} x^\top \left( \sum_{x' \in \widehat{\mathcal{X}}_k} \lambda_{x'} x' x'^\top \right)^{-1} x$
    Draw $x_{(k-1)\tau+1}, \ldots, x_{k\tau} \sim \lambda^{(k)}$
    For $t = (k-1)\tau + 1, \ldots, k\tau$, pull arm $x_t$ and observe $y_t = f(x_t) + \eta_t$ where $\eta_t \sim \mathcal{N}(0, 1)$
    $V_k = V_{k-1} + \sum_{t=(k-1)\tau+1}^{k\tau} x_t x_t^\top$, $S_k = S_{k-1} + \sum_{t=(k-1)\tau+1}^{k\tau} x_t y_t$, $\theta_k = V_k^{-1} S_k$
    $\beta_k = \sqrt{\gamma} U + \sqrt{2\log(1/\delta) + \log(|V_k|/|V_0|)}$
    $\widehat{x}_k = \arg\max_{x' \in \widehat{\mathcal{X}}_k} \langle x', \theta_k \rangle$
    $\widehat{\mathcal{X}}_{k+1} = \widehat{\mathcal{X}}_k - \{x \in \widehat{\mathcal{X}}_k : \langle \widehat{x}_k - x, \theta_k \rangle \geq \beta_k \|\widehat{x}_k - x\|_{V_k^{-1}}\}$
```

```
UCB algorithm
Input: $T \in \mathbb{N}$
Initialize $\delta = 1/T$, $\gamma = 1$, $U = 1$ (supposed to be an upper bound on $\|\theta_*\|_2$), $V_0 = \gamma I$, $S_0 = 0$
for: $t = 0, 1, 2, \ldots, T-1$
    $\beta_t = \sqrt{\gamma} U + \sqrt{2\log(1/\delta) + \log(|V_t|/|V_0|)}$
    $\theta_t = V_t^{-1} S_t$
    $x_t = \arg\max_{x \in \mathcal{X}} \langle x, \theta_t \rangle + \|x\|_{V_t^{-1}} \beta_t$
    Pull arm $x_t$ and observe $y_t = f(x_t) + \eta_t$ where $\eta_t \sim \mathcal{N}(0, 1)$
    $V_{t+1} = V_t + x_t x_t^\top$, $S_{t+1} = S_t + x_t y_t$
```

```
Thompson sampling algorithm
Input: $T \in \mathbb{N}$
Initialize $\gamma = 1$, $V_0 = \gamma I$, $S_0 = 0$
for: $t = 0, 1, 2, \ldots, T-1$
    $\theta_t = V_t^{-1} S_t$
    $\widetilde{\theta}_t \sim \mathcal{N}(\theta_t, V_t^{-1})$
    $x_t = \arg\max_{x \in \mathcal{X}} \langle x, \widetilde{\theta}_t \rangle$
    Pull arm $x_t$ and observe $y_t = f(x_t) + \eta_t$ where $\eta_t \sim \mathcal{N}(0, 1)$
    $V_{t+1} = V_t + x_t x_t^\top$, $S_{t+1} = S_t + x_t y_t$
```

References

[1] Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lil'ucb: An optimal exploration algorithm for multi-armed bandits. In *Conference on Learning Theory*, pages 423–439, 2014.

[2] Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.