Proof Note for R&Z 2023.

Theorem 1. *Fix any s_0, d and T. Let $K = \log(T/s_0)$ and consider the problem $x_{t,a} \sim \mathcal{N}(0, I_d)$, $\forall a \in [K]$, $\forall t \in [T]$, where the contexts are independence across t. Then for any $M \le T$ and any dynamic batch learning algorithm* **Alg***, we have*

$$\sup_{\theta^*: \|\theta^*\|_2 \le 1, \|\theta^*\|_0 \le s_0} \mathbb{E}_{\theta^*}[R_T(\mathbf{Alg})] \ge c \cdot \max\left(M^{-4} 2^{-7M/2} \cdot \sqrt{T s_0} \cdot \left(\frac{T}{s_0}\right)^{\frac{1}{2(2^M - 1)}}, \sqrt{T s_0} \right), \tag{3}$$

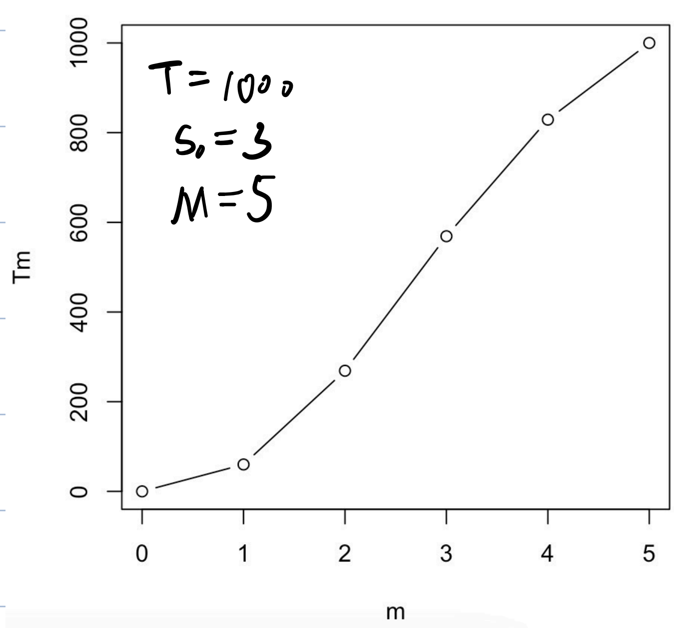
where $\mathbb{E}_{\theta^*}$ denotes taking expectation w.r.t. the distribution based on the parameter θ^*, and $c > 0$ is a numerical constant independent of (T, M, d, s_0).

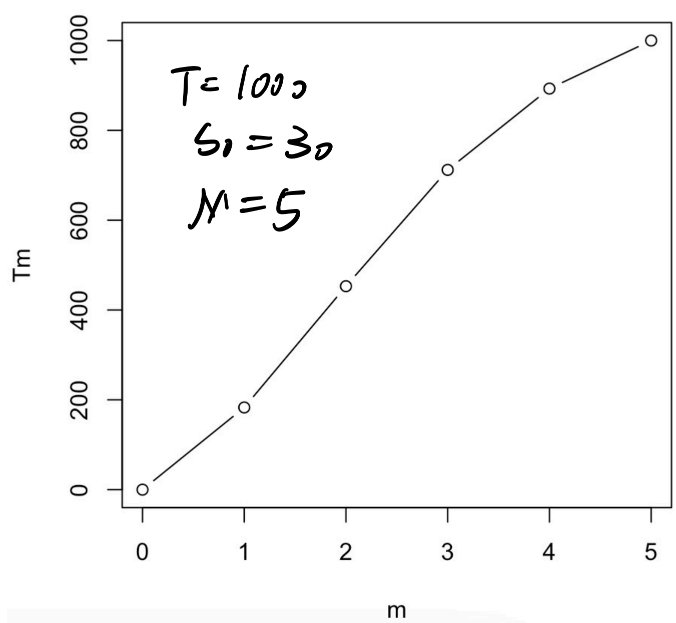
① Proof Outline

Consider $K = 2^M$ arms, define a "bad" grid $\{T_1, T_2 \cdots T_M\}$, where

I $\quad T_m = \left\lfloor s_0 \cdot \left(\frac{T}{s_0}\right)^{\frac{1-2^{-m}}{1-2^{-M}}} \right\rfloor$

Claim: # observations before T_{m-1} is too few to learn policy effectively ?





Potential Motivation:

Bypassing the Monster: A Faster and Simpler Optimal Algorithm for Contextual Bandits under Realizability

David Simchi-Levi Yunzong Xu

Institute for Data, Systems, and Society, Massachusetts Institute of Technology

dslevi@mit.edu, yxu@mit.edu

Algorithm 1 FAst Least-squares-regression-oracle CONtextual bandits (`FALCON`)

input epoch schedule $0 = \tau_0 < \tau_1 < \tau_2 < \cdots$, confidence parameter δ, tuning parameter c

for epoch $m = 1, 2, \ldots$ **do**

  Let $\gamma_m = c\sqrt{K\tau_{m-1}/\log(|\mathcal{F}|\log(\tau_{m-1})m/\delta)}$ (for epoch 1, $\gamma_1 = 1$).

  Compute $\widehat{f}_m = \arg\min_{f\in\mathcal{F}} \sum_{t=1}^{\tau_{m-1}} (f(x_t, a_t) - r_t(a_t))^2$ via the **offline least squares oracle**.

  for round $t = \tau_{m-1} + 1, \cdots, \tau_m$ **do**

    Observe context $x_t \in \mathcal{X}$.

    Compute $\widehat{f}_m(x_t, a)$ for each action $a \in \mathcal{A}$. Let $\widehat{a}_t = \max_{a\in\mathcal{A}} \widehat{f}_m(x_t, a)$. Define

$$p_t(a) = \begin{cases} \frac{1}{K+\gamma_m\left(\widehat{f}_m(x_t,\widehat{a}_t)-\widehat{f}_m(x_t,a)\right)}, & \text{for all } a \neq \widehat{a}_t, \\ 1 - \sum_{a\neq\widehat{a}_t} p_t(a), & \text{for } a = \widehat{a}_t. \end{cases}$$

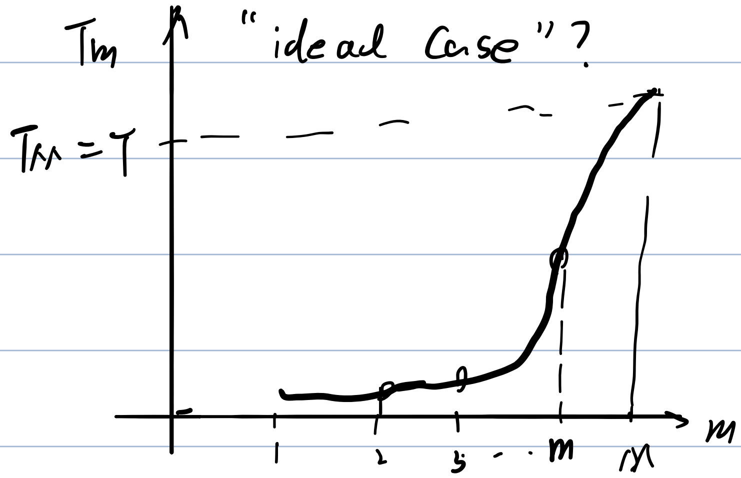
    Sample $a_t \sim p_t(\cdot)$ and observe reward $r_t(a_t)$.

Our algorithm runs in an epoch schedule to reduce oracle calls, i.e., it only calls the oracle at certain pre-specified rounds $\tau_1, \tau_2, \tau_3, \ldots$. For $m \in \mathbb{N}$, we refer to the rounds from $\tau_{m-1} + 1$ to τ_m as epoch m. As a concrete example, consider $\tau_m = 2^m$, then for any (possibly unknown) T, our algorithm runs in $O(\log T)$ epochs. As another example, when T is known, consider $\tau_m = \left\lfloor 2T^{1-2^{-m}} \right\rfloor$, then our algorithm runs in $O(\log\log T)$ epochs. We allow very general epoch schedules; in particular, calling the oracle more frequently does not affect the regret analysis.

At the start of each epoch m, our algorithm makes two updates. First, it updates a (epoch-

epoch = batch.

"ideal case"?



II Construct a prior Q for θ^*

(to be specified later)

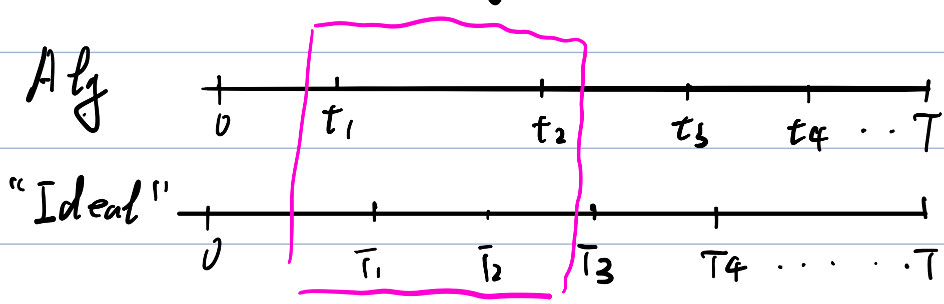
III Under Q, the regret at m-th batch is $(T_m - T_{m-1})\Delta_m$

where $\Delta_m = \frac{1}{24 M^2 2^{3M}} \left(\frac{T}{S_0}\right)^{-\frac{1-2^{1-m}}{2(1-2^{-M})}}$.

st $(T_m - T_{m-1})\Delta_m$ is large.

Interpretation: Divide 2^M arms into 2^{M-1} pairs, the difference between each pair is at the scale of Δ_m. (why?)

IV Consider $\forall$ Alg with grid design: (and "ideal" T_m).



(why?)

"A Bad Event" B_m: $\{t_{m-1} \le T_{m-1} < T_m < t_m\}$

(·X·) Claim: If at least 1 bad event B_m happens with large Prob. will get the desired bound.

(V) Summary: $\begin{cases} \text{Construct prior } Q. \\ \text{L.B when bad event happens with large P.} \\ \text{Bad event happens with large P.} \end{cases}$

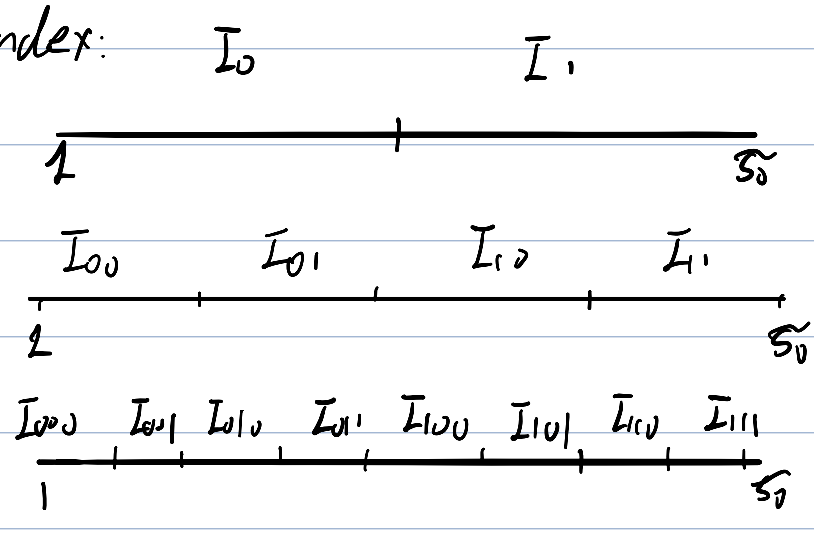
② Prior Construction.

$$\theta = (\underbrace{\theta_1, \theta_2 \cdots \theta_{s_0}}_{s_0 \text{ elements}}, 0 \cdots \quad 0).$$

Approximately divide s_0 into 2^M subgroups:

$\tilde{s}_0 := \lfloor s_0 2^{-M} \rfloor \cdot 2^M$. divide $\tilde{s}_0$ instead.

Divide & Index:



$$I_{0\cdots00} = \left\{1, \ldots, \frac{1}{2^M}\tilde{s}_0\right\}, \; I_{0\cdots01} = \left\{\frac{1}{2^M}\tilde{s}_0 + 1, \ldots, \frac{1}{2^{M-1}}\tilde{s}_0\right\}, \cdots,$$

$$I_{1\cdots11} = \left\{\tilde{s}_0 - \frac{1}{2^M}\tilde{s}_0 + 1, \ldots, \tilde{s}_0\right\}.$$

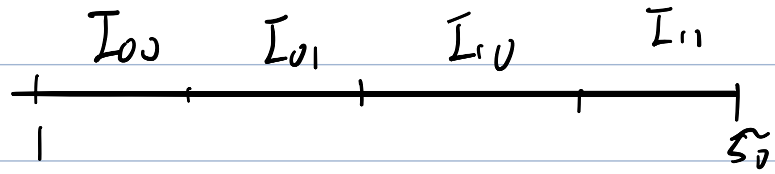
Each interval is indexed by $\delta \in \Pi(M) := \{0, 1\}^M$.

To construct Q:

- Generate $\theta_1, \dots \theta_M$ from $\mathrm{Unif}(\mathbb{S}^{\tilde{s}_0/2^M})$
- $\forall m \in [M]$, define $\tilde{\theta}_m \in \mathbb{R}^{\tilde{s}_0}$: $\tilde{\theta}_m(I_b) = \begin{cases} \theta_m & b_m = 0 \\ -\theta_m & b_m = 1. \end{cases}$

$b \in \Pi(m)$.

$M = 2$:



$b \in \{00, 01, 10, 11\}$

$m = 1$: $\tilde{\theta}_1 = (\theta_1^T, \theta_1^T, -\theta_1^T, -\theta_1^T)^T$. b_1

$m = 2$: $\tilde{\theta}_2 = (\theta_2^T, -\theta_2^T, \theta_2^T, -\theta_2^T)$ b_2.

[flip the sign of half of the elements by their position].

- Set $\tilde{\theta} = \sum_{m=1}^{M} \Delta_m \tilde{\theta}_m \in \mathbb{R}^{\tilde{s}_0}$

$\theta \in \mathbb{R}^d$: first $\tilde{s}_0 = \tilde{\theta}$, others 0.

- Specify joint distribution of $K = 2^M$ arms:

$\forall t \in [T]$, draw $X_t \sim N(0, I_d)$.

$S = \{1, 2, \dots \tilde{s}_0\}$ $S^c = \{\tilde{s}_0 + 1, \dots d\}$

$\forall a \in [K]$ $X_{t,a}(S^c) = X_t(S^c)$

Nuisance parts across arms are same.

Constructing $X_{t,a}$ for $\forall a \in M$:

$X_{t,a}(S) = \{ X_{t,a}^{(1)}, X_{t(a)}^{(2)} \cdots\cdots X_{t(a)}^{(\widetilde{S}_0)} \}$ → Permute it!

2^{M} intervals: $I_{00\cdots0}$ $\cdots\cdots$ $I_{11\cdots1}$

$a = 1 + \sum_{m=1}^{M} a_m \cdot 2^{m-1}$ $\quad a_m \in \{0,1\}$

Define a mapping $M_a: \Pi(M) \to \Pi(M)$

$M_a(b)_m = (1-a_m) \cdot b_m + a_m(1-b_m)$.

Interpretation: for the index code $b \in \mathbb{R}^{M}$, flip the element according to the position.

$a_m = 1$: flip b_m, otherwise don't.

Set $X_{t,a}(b) = X_t(M_a(b))$.

Example. $M=2$. $2^2 = 4$ Intervals.

I_{00} I_{01} I_{10} I_{11}

$b = \{00, 01, 10, 11\}$.

Consider $K=4$. $a \in \{1,2,3,4\}$

$$\begin{array}{lllll}
 & & a_1 & a_2 & \\
a=1: & 1 = 1 + & 0\cdot 1 + & 0\cdot 2 & 00 \\
a=2: & 2 = 1 + & 1\cdot 1 + & 0\cdot 2 & 10 \\
a=3: & 3 = 1 + & 0\cdot 1 + & 1\cdot 2 & 01 \\
a=4: & 4 = 1 + & 1\cdot 1 + & 1\cdot 2 & 11
\end{array}$$

$G = \{00, 01, 10, 11\}$

$M_{a=1}(G) = \{00, 01, 10, 11\} \Rightarrow X_{t,1}(\tilde{S}_0) = X_t(\tilde{S}_0)$

$M_{a=2}(G) = \{10, 11, 00, 01\}$

$\Rightarrow X_{t,2}(\tilde{S}_0) = [X_t(I_{10}), X_t(I_{11}), X_t(I_{00}), X_t(I_{01})]$

$M_{a=3}(G) = \{01, 00, 11, 10\}$

$\Rightarrow X_{t,3}(\tilde{S}_0) = \{X_t(I_{01}), X_t(I_{00}), X_t(I_{11}), X_t(I_{10})\}$

$M_{a=4}(G) = \{11, 10, 01, 00\}$

$\Rightarrow X_{t4}(\tilde{S}_0) = \{X_t(I_{11}), X_t(I_{10}), X_t(I_{01}), X_t(I_{00})\}$

Recall: the difference between each pairs is at the order of Δm:

$$\theta^\top (X_{t,1} - X_{t,2}) = \sum_{m=1}^{M} \Delta m \, \tilde{\theta}_m^\top \begin{pmatrix} X_t(I_0) - X_t(I_1) \\ X_t(I_1) - X_t(I_0) \end{pmatrix}$$

$\in \mathbb{R}^{\tilde{S}_0}$

③ Notation for Regret Decomposion.

$$\sup_{\theta^*: \|\theta^*\|_2 \le 1, \|\theta^*\|_0 \le s_0} \mathbb{E}_{\theta^*}[R_T(\mathbf{Alg})] \ge \mathbb{E}_Q \mathbb{E}_\theta [R_T(\mathbf{Alg})]$$

$$= \sum_{t=1}^{T} \mathbb{E}_Q \left(\mathbb{E}_x \mathbb{E}_{P^t_{\theta,x}} \left[\max_{a \in [K]} x_{t,a}^\top \theta - x_{t,a_t}^\top \theta \right] \right),$$

E_x: taking expectation to all contexts at all time

$E_{P^t_{\theta,x}}$: taking expectation to observed rewards before time at the current batch for given X & θ.

batch.

Recall that for each $j \in [2^M]$, we write $j = 1 + \sum_{m=1}^{M} j_m \cdot 2^{m-1}$. Then for each $t \in [T]$ and any $m \in [M]$,

$\{a_t = j + 2^{m-1}\} \Leftrightarrow \{j_m = 1\}$

$$
\begin{aligned}
\max_{a\in[K]}(x_{t,a}^\top\theta - x_{t,a_t}^\top\theta) &= \sum_{j\in[K]} \mathbf{1}\{a_t = j\}\cdot \max_{a\in[K]}(x_{t,a}^\top\theta - x_{t,j}^\top\theta) \\
&\overset{(a)}{=} \sum_{j\in[K]:j_m=0} \mathbf{1}\{a_t = j\}\cdot \max_{a\in[K]}(x_{t,a}^\top\theta - x_{t,j}^\top\theta) + \mathbf{1}\{a_t = j+2^{m-1}\} \\
&\quad \cdot \max_{a\in[K]}(x_{t,a}^\top\theta - x_{t,j+2^{m-1}}^\top\theta) \\
&\ge \sum_{j\in[K]:j_m=0} \mathbf{1}\{a_t = j\}\cdot \underline{\max_{a\in\{j,j+2^{m-1}\}}}(x_{t,a}^\top\theta - x_{t,j}^\top\theta) \\
&\quad + \mathbf{1}\{a_t = j+2^{m-1}\}\cdot \underline{\max_{a\in\{j,j+2^{m-1}\}}}(x_{t,a}^\top\theta - x_{t,j+2^{m-1}}^\top\theta) \\
&= \sum_{j\in[K]:j_m=0} \mathbf{1}\{a_t = j\}\cdot (x_{t,j+2^{m-1}}^\top\theta - x_{t,j}^\top\theta)_+ + \mathbf{1}\{a_t = j+2^{m-1}\} \\
&\quad \cdot (x_{t,j+2^{m-1}}^\top\theta - x_{t,j}^\top\theta)_-, \qquad (4)
\end{aligned}
$$

$X_+ = \max(X, 0)$

$X_- = \max(-X, 0)$.

For the group with $j_m = 0$:

$$
\begin{aligned}
x_{t,j+2^{m-1}}^\top\theta - x_{t,j}^\top\theta &= \sum_{\sigma\in\Pi(M)} x_{t,j+2^{m-1}}(I_\sigma)^\top\theta(I_\sigma) - x_{t,j}(I_\sigma)^\top\theta(I_\sigma) \\
&= \sum_{\sigma\in\Pi(M):\sigma_m=0} x_{t,j+2^{m-1}}(I_\sigma)^\top\theta(I_\sigma) - x_{t,j}(I_\sigma)^\top\theta(I_\sigma) \\
&\quad + \sum_{\sigma\in\Pi(M):\sigma_m=1} x_{t,j+2^{m-1}}(I_\sigma)^\top\theta(I_\sigma) - x_{t,j}(I_\sigma)^\top\theta(I_\sigma) \\
&= 2\Delta_m\cdot\theta_m^\top\left(\sum_{\sigma\in\Pi(M):\sigma_m=1} x_t(I_\sigma) - \sum_{\sigma\in\Pi(M):\sigma_m=0} x_t(I_\sigma)\right).
\end{aligned}
$$

$\cdot\, \theta_{S^c} = 0$. $X_{t,a}(S^c)$

Same across arms.

$$
\begin{pmatrix} X_{t,a}(I_{00}) \\ X_{t,a}(I_{01}) \\ X_{t,a}(I_{10}) \\ X_{t,a}(I_{11}) \end{pmatrix}
\begin{pmatrix} \theta(I_{00}) \\ \theta(I_{01}) \\ \theta(I_{10}) \\ \theta(I_{11}) \end{pmatrix}
$$

To simplify the notation, we define

$$
d_{m,t} = \sum_{\sigma\in\Pi(M):\sigma_m=2} x_t(I_\sigma) - \sum_{\sigma\in\Pi(M):\sigma_m=1} x_t(I_\sigma),\quad u_{m,t} = \frac{d_{m,t}}{\|d_{m,t}\|_2},
$$

Recall: $\tilde\theta_m(I_\sigma) = \begin{cases} \theta_m & \sigma_m = 0, \\ -\theta_m & \sigma_m = 1, \end{cases}$ and $\tilde\theta = \sum \Delta_m \tilde\theta_m$

flip signs only for θ_m. $\rightarrow$ Reg lower bound!

$$(4) = 2\Delta_m \sum_{j\in\mathcal{A}_m} \mathbf{1}\{a_t = j\}\cdot(d_{m,t}^\top\theta_m)_+ + \mathbf{1}\{a_t = j + 2^{m-1}\}\cdot(d_{m,t}^\top\theta_m)_-$$

$$= 2\Delta_m\cdot\mathbf{1}\{j\in\mathcal{A}_m\}\cdot(d_{m,t}^\top\theta_m)_+ + \mathbf{1}\{j\in\mathcal{A}_m^c\}\cdot(d_{m,t}^\top\theta_m)_-.$$

$A_m : \{j \in [K] ; j_m = 0\}$

As a result, we have

$$\mathbb{E}_Q\mathbb{E}_{P_{\theta,x}^t}\left[\max_{a\in[K]}(x_{t,a}^\top\theta - x_{t,a_t}^\top\theta)\right]$$
$$\geq 2\Delta_m\cdot\mathbb{E}_Q[(d_{m,t}^\top\theta_m)_+\cdot\mathbb{E}_{P_{\theta,x}^t}[\mathbf{1}\{\underline{a_t\in\mathcal{A}_m}\}] + (d_{m,t}^\top\theta_m)_- \cdot\mathbb{E}_{P_{\theta,x}^t}[\mathbf{1}\{\underline{a_t\in\mathcal{A}_m^c}\}]], \quad (5)$$

Change Prob. measure through Radon - Nykodym:

$$\frac{dQ_{m,t}^+}{dQ}(\theta) = \frac{(d_{m,t}^\top\theta_m)_+}{Z_m(d_{m,t})}, \quad \frac{dQ_{m,t}^-(\theta)}{dQ} = \frac{(d_{m,t}^\top\theta_m)_-}{Z_m(d_{m,t})}$$

$$Z_m(d_{m,t}) = \mathbb{E}_Q(d_{m,t}^\top\theta_m)_+ = \mathbb{E}_Q(d_{m,t}^\top\theta_m)_-.$$

$$\mathbb{E}_Q\mathbb{E}_{P_{\theta,x}^t}\left[\max_{a\in[K]}(x_{t,a}^\top\theta - x_{t,a_t}^\top\theta)\right]$$
$$\geq 2\Delta_m Z_m(d_{m,t})\cdot\Big(\mathbb{E}_{P_{\theta,x}^t\circ Q_{m,t}^+}[\mathbf{1}\{a_t\in\mathcal{A}_m\}]$$
$$+\mathbb{E}_{P_{\theta,x}^t\circ Q_{m,t}^-}[\mathbf{1}\{a_t = \mathcal{A}_m^c\}]\Big),$$

Next, deal with measures $P_{\theta,x}^t \circ Q_{m,t}^-$ and $P_{\theta,x}^t\circ Q_{m,t}^+$ respectively.

④ Regret Lower Bound When Bad Event Happens with High Probability.

Lemma 2. *If there exists $m \in [M]$, such that*

$$\sum_{t=T_{m-1}+1}^{T_m} \mathbb{E}_x\left[Z_m(d_{m,t}) \cdot \mathbb{E}_{P_{\theta,x}\circ Q^+_{m,t}}[\mathbf{1}\{B_m\}]\right] \ge \frac{T_m - T_{m-1}}{8 \cdot 2^{\frac{M}{2}} M^2}, \tag{6}$$

then there eixsts a numerical constant $c > 0$, independent of (T, M, d, s_0), such that,

$$\sup_{\theta^*:\|\theta^*\|_2\le 1, \|\theta^*\|_0\le s_0} \mathbb{E}_{\theta^*}[R_T(\mathbf{Alg})] \ge \frac{c}{M^4 2^{3M}} \sqrt{Ts_0}\left(\frac{T}{s_0}\right)^{\frac{1}{2(2^M-1)}}.$$

$$\begin{aligned}
&\sup_{\theta^*:\|\theta^*\|_2\le 1,\, \|\theta^*\|_0\le s_0} \mathbb{E}_{\theta^*}[R_T(\mathbf{Alg})] \\
&\ge 2\Delta_m \sum_{t=T_{m-1}+1}^{T_m} \mathbb{E}_x\Big[Z_m(d_{m,t}) \cdot \Big(\mathbb{E}_{P^t_{\theta,x}\circ Q^+_{m,t}}[\mathbf{1}\{a_t \in \mathcal{A}_m\}] \\
&\quad + \mathbb{E}_{P^t_{\theta,x}\circ Q^-_{m,t}}[\mathbf{1}\{a_t \in \mathcal{A}^c_m\}]\Big)\Big] \\
&\overset{(a)}{\ge} 2\Delta_m \sum_{t=T_{m-1}+1}^{T_m} \mathbb{E}_x[Z_m(d_{m,t}) \cdot (1 - \mathsf{TV}(P^t_{\theta,x}\circ Q^+_{m,t}, P^t_{\theta,x}\circ Q^-_{m,t}))] \\
&\overset{(b)}{\ge} 2\Delta_m \sum_{t=T_{m-1}+1}^{T_m} \mathbb{E}_x[Z_m(d_{m,t}) \cdot (1 - \mathsf{TV}(P^{T_m}_{\theta,x}\circ Q^+_{m,t}, P^{T_m}_{\theta,x}\circ Q^-_{m,t}))],
\end{aligned}$$

$P^t_{\theta,x}\circ Q^+_{m,t}$: P^t

$P^t_{\theta,x}\circ Q^-_{m,t}$: Q^t.

TV: Total variation

$:= \sup_{A\in\mathcal{F}} |P^t(A) - Q^t(A)|$.

for $(\Omega, \mathcal{F})$

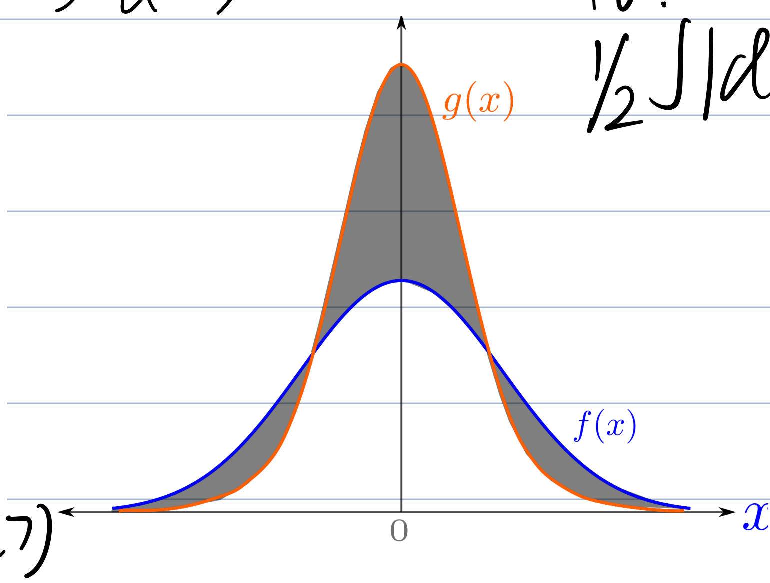
· $P^t(A) + Q^t(A^c) = 1 + P^t(A) - Q^t(A) \ge 1 - TV(P^t, Q^t)$.

· P^t, Q^t: p. m on $\mathcal{F}_t \subset \mathcal{F}_{T_m}$.

$TV(P^t, Q^t) \le TV(P^{T_m}, Q^{T_m})$

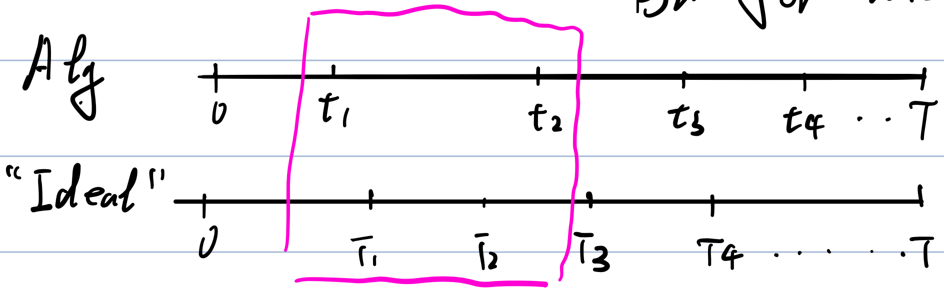
TV: $\frac{1}{2}\int |dp - dq|$

$$\begin{aligned}
&1 - \mathsf{TV}\left(P^{T_m}_{\theta,x}\circ Q^+_{m,t}, P^{T_m}_{\theta,x}\circ Q^-_{m,t}\right) \\
&= \int \min\left(dP^{T_m}_{\theta,x}\circ Q^+_{m,t}, dP^{T_m}_{\theta,x}\circ Q^-_{m,t}\right) \\
&\ge \int_{B_m} \min\left(dP^{T_m}_{\theta,x}\circ Q^+_{m,t}, dP^{T_m}_{\theta,x}\circ Q^-_{m,t}\right) \\
&= \frac{1}{2}\int_{B_m} \Big(dP^{T_m}_{\theta,x}\circ Q^+_{m,t} + dP^{T_m}_{\theta,x}\circ Q^-_{m,t} \\
&\quad - |dP^{T_m}_{\theta,x}\circ Q^+_{m,t} - dP^{T_m}_{\theta,x}\circ Q^-_{m,t}|\Big) \\
&= \frac{1}{2}\int_{B_m} \Big(dP^{T_{m-1}}_{\theta,x}\circ Q^+_{m,t} + dP^{T_{m-1}}_{\theta,x}\circ Q^-_{m,t} \\
&\quad - |dP^{T_{m-1}}_{\theta,x}\circ Q^+_{m,t} - dP^{T_{m-1}}_{\theta,x}\circ Q^-_{m,t}|\Big), \quad (7)
\end{aligned}$$



- $\min(A, B) = \frac{1}{2}(A + B - |A - B|)$.
- Recall the def of "Bad Event".

B_m for $m = 2$.



- Sample data by $P_{\theta,x}^{t_2}$ in $\mathcal{F}_{t_2} \supset \mathcal{F}_{T_2} \supset \mathcal{F}_{T_1}$.

T_2, T_1 at the same Batch $\rightarrow dP_{\theta,x}^{T_2} = dP_{\theta,x}^{T_1}$

$$(7) = \frac{1}{2}\left(\mathbb{E}_{P_{\theta,x}^{T_{m-1}} \circ Q_{m,t}^{+}}[\mathbf{1}\{B_m\}] + \mathbb{E}_{P_{\theta,x}^{T_{m-1}} \circ Q_{m,t}^{-}}[\mathbf{1}\{\cancel{A}_m\}]\right) \longrightarrow B_m$$

$$\geq ? \quad - \mathsf{TV}\left(dP_{\theta,x}^{T_{m-1}} \circ Q_{m,t}^{+}, dP_{\theta,x}^{T_{m-1}} \circ Q_{m,t}^{-}\right)$$

$$\geq \mathbb{E}_{P_{\theta,x} \circ Q_{m,t}^{+}}[\mathbf{1}\{B_m\}] - \frac{3}{2}\mathsf{TV}\left(P_{\theta,x}^{T_{m-1}} \circ Q_{m,t}^{+}, P_{\theta,x}^{T_{m-1}} \circ Q_{m,t}^{-}\right).$$

- $P^t(A) + Q^t(A) = 2P^t(A) + Q^t(A) - P^t(A)$
$\geq 2P^t(A) - 2\mathsf{TV}(P^t, Q^t)$.

Lemma A.3 (Pinsker's Inequality). *Let P and Q be any two probability measures on the same measurable space. Then* $\mathsf{TV}(P, Q) \leq \sqrt{\frac{1}{2} \cdot D_{\mathrm{KL}}(P\|Q)}$.

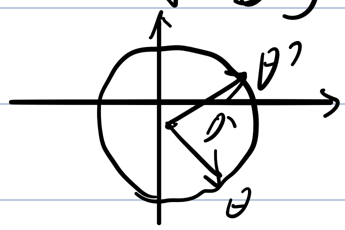
$\Rightarrow$

$$\mathsf{TV}\left(P_{\theta,x}^{T_{m-1}} \circ Q_{m,t}^{+}, P_{\theta,x}^{T_{m-1}} \circ Q_{m,t}^{-}\right) \leq \sqrt{\frac{1}{2} D_{\mathrm{KL}}\left(P_{\theta,x}^{T_{m-1}} \circ Q_{m,t}^{+} \| P_{\theta,x}^{T_{m-1}} \circ Q_{m,t}^{-}\right)}. \tag{8}$$

Goal: Simplify (8).

- Rotational invariance of uniform distribution.

$\theta \sim \mathrm{unif}(S^p)$ $\quad R^T R = I$. $\quad R\theta \overset{d}{\rightsquigarrow} \theta$.



Result:

Lemma A.4 (Joint Convexity of the KL divergence (Cover and Thomas 2006)). *The KL divergence $D_{\mathrm{KL}}(P||Q)$ is jointly convex in its argument P and Q: let P_1, P_2, Q_1, Q_2 be distributions on $\mathcal{X}$, then for any $\lambda \in [0,1]$,*

$$D_{\mathrm{KL}}(\lambda P_1 + (1-\lambda)P_2 \| \lambda Q_1 + (1-\lambda)Q_2) \le \lambda D_{\mathrm{KL}}(P_1\|Q_1) + (1-\lambda)D_{\mathrm{KL}}(P_2\|Q_2).$$

$\Rightarrow$

$$(8) = \sqrt{\frac{1}{2}D_{\mathrm{KL}}\left(P_{\theta,x}^{T_{m-1}} \circ Q_{m,t}^{+} \| P_{\theta',x}^{T_{m-1}} \circ Q_{m,t}^{+}\right)} \le \sqrt{\frac{1}{2}\mathbb{E}_{Q_{m,t}^{+}}\left[D_{\mathrm{KL}}\left(P_{\theta,x}^{T_{m-1}} \| P_{\theta',x}^{T_{m-1}}\right)\right]},$$

$E[D_{KL}(P_{\theta,x}^{T_{m-1}} \| P_{\theta',x}^{T_{m-1}})]$ can be calculated explicitly.

Consequently:

$$\begin{aligned}
&\sup_{\theta^*:\|\theta^*\|_2\le 1, \|\theta^*\|_0 \le s_0} \mathbb{E}_{\theta^*}[R_T(\mathbf{Alg})] \\
&\ge 2\Delta_m \sum_{t=T_{m-1}+1}^{T_m} \mathbb{E}_x\left[Z_m(d_{m,t})\cdot\left(\mathbb{E}_{P_{\theta,x}\circ Q_{m,t}^{+}}[\mathbf{1}\{B_m\}] - \frac{3}{2}\sqrt{\frac{2^{M+1}\Delta_m^2}{\tilde{s}_0}\cdot u_{m,t}^\top\left(\sum_{\tau=1}^{T_{m-1}}\sum_{j\in[K]} h_{\tau,j}h_{\tau,j}^\top\right)u_{m,t}}\right)\right] \\
&\overset{(a)}{\ge} 2\Delta_m \sum_{t=T_{m-1}+1}^{T_m} \mathbb{E}_x\left[Z_m(d_{m,t})\cdot\left(\mathbb{E}_{P_{\theta,x}\circ Q_{m,t}^{+}}[\mathbf{1}\{B_m\}] - \frac{3}{2}\sqrt{\frac{2^{3M+1}\Delta_m^2 T_{m-1}}{\tilde{s}_0}}\right)\right] \\
&\overset{(b)}{\ge} 2\Delta_m \sum_{t=T_{m-1}+1}^{T_m} \mathbb{E}_x\left[Z_m(d_{m,t})\cdot\left(\mathbb{E}_{P_{\theta,x}\circ Q_{m,t}^{+}}[\mathbf{1}\{B_m\}] - \frac{1}{2\cdot 2^{M+2}M^2}\right)\right],
\end{aligned} \tag{9}$$

Recall:

$$\sum_{t=T_{m-1}+1}^{T_m} \mathbb{E}_x\left[Z_m(d_{m,t})\cdot\mathbb{E}_{P_{\theta,x}\circ Q_{m,t}^{+}}[\mathbf{1}\{B_m\}]\right] \ge \frac{T_m - T_{m-1}}{8\cdot 2^{\frac{M}{2}}M^2}, \tag{6}$$

Finally:

$$\sup_{\theta^*:\|\theta^*\|_2\le 1, \|\theta^*\|_0 \le s_0} \mathbb{E}_{\theta^*}\cdot[R_T(\mathbf{Alg})] \ge \frac{(T_m - T_{m-1})\Delta_m}{2\cdot 2^{\frac{M}{2}+2}M^2} \ge \frac{c}{M^4 2^{7M/2}}\cdot\sqrt{s_0 T}\left(\frac{T}{s_0}\right)^{\frac{1}{2(2^M-1)}}.$$

⑤ Bad Event Happens with Large Probability.

Lemma 3. *There exists some $m \in [M]$, such that*

$$\sum_{t=T_{m-1}+1}^{T_m} \mathbb{E}_x[Z_m(d_{m,t})\cdot\mathbb{E}_{P_{\theta,x}\circ Q_{m,t}^{+}}[\mathbf{1}\{B_m\}]] \ge \frac{T_m - T_{m-1}}{2^{\frac{M}{2}+2}M^2}.$$

$P(\bigcup_{m=1}^{M} B_m) = 1$: $\Rightarrow \exists\, m$ st $P(B_m) \ge 1/M$.

- Given $\mathcal{F}_{T_{m-1}}$, $\mathbb{1}\{B_m\} \perp\!\!\!\perp \{X_{t,a}\}_{t \geq T_{m-1}}$.
- Independence between contexts
- R-N change of measure

$$\sum_{t=T_{m-1}+1}^{T_m} \mathbb{E}_x\left[Z_m(d_{m,t}) \cdot \mathbb{E}_{P_{\theta,x}\circ Q^t_{1,m}}[\mathbf{1}\{B_m\}]\right]$$
$$= \sum_{t=T_{m-1}+1}^{T_m} \mathbb{E}_x\mathbb{E}_Q\left[(d_{M,T}^\top\theta_m)_+ \cdot P_{\theta,x}(B_m)\right]$$
$$\geq \sum_{t=T_{m-1}+1}^{T_m} \mathbb{E}_x\mathbb{E}_Q\left[\min_{m'\in[M]}\{(d_{M,T}^\top\theta_{m'})_+\} \cdot P_{\theta,x}(B_m)\right]$$
$$\overset{(a)}{=} \sum_{t=T_{m-1}+1}^{T_m} \tilde{Z} \cdot \mathbb{E}_{P_{\theta,x}\circ\tilde{Q}}[\mathbf{1}\{B_m\}]$$
$$= (T_m - T_{m-1})\tilde{Z} \cdot \boxed{\mathbb{E}_{P_{\theta,x}\circ\tilde{Q}}[\mathbf{1}\{B_m\}]}, \longrightarrow \geq \frac{1}{M} \text{ for some } M$$

$\tilde{Z}$: normalizing constant.

$$\tilde{Z} \geq \frac{1}{2^{\frac{M+1}{2}} \cdot (M+1)}.$$

$$\tilde{Z} \cdot \mathbb{E}_{P_{\theta,x}\circ\tilde{Q}}[\mathbf{1}\{B_m\}] \geq \frac{\tilde{Z}}{M} \geq \frac{1}{2^{\frac{M}{2}+2}M^2}.$$

Finally for this m, $\sum_{t=T_{m-1}+1}^{T_m} \mathbb{E}_x[Z_m(d_{m,t}) \cdot \mathbb{E}_{P_{\theta,x}\circ Q^+_{m,t}}[\mathbf{1}\{B_m\}]] \geq \frac{(T_m - T_{m-1})}{2^{\frac{M}{2}+2}M^2}$.

$\boxed{\text{II}}$

Theorem 2. *Under* Model-C, *Assumptions* 1–4 *and* $M = O(\log\log(T/s_0))$, *we have*

$$\sup_{\theta^\star:\|\theta^\star\|_2\le 1,\|\theta^\star\|_0\le s_0} \mathbb{E}_{\theta^\star}[R_T(\mathbf{Alg})] \le \frac{C\cdot M^{3/2}\sqrt{\log K\log(KT)\log(dT)}}{\gamma(K)\rho(K)}\cdot\sqrt{Ts_0}\left(\frac{T}{s_0}\right)^{\frac{1}{2(2^M-1)}}, \tag{10}$$

where **Alg** *is LBGL and* $C>0$ *is a numerical constant independent of* (T,d,M,K,s_0).

① Proof Framework.

Restricted Eigenvalues:

$$\phi_{\min}(s,A) \triangleq \min_{v\in\mathbb{R}^d:\|v\|_0\le s}\left\{\frac{v^\top Av}{\|v\|_2^2}\right\},$$

$$\phi_{\max}(s,A) \triangleq \max_{v\in\mathbb{R}^d:\|v\|_0\le s}\left\{\frac{v^\top Av}{\|v\|_2^2}\right\}.$$

Two supporting Lemmas for Lasso Estimation:

Lemma 5. *Suppose Assumptions* 1–4 *hold. Given a sparsity parameter* s*, with probability at least* $1-2M^2\exp(O(s\log d)-\Omega(\rho^2(K)\cdot\sqrt{Ts_0}/M))$*, for any* $j,m\in[M]$*,*

$$\phi_{\max}\left(s,\frac{D_{m,j}}{|T_m^{(j)}|}\right)\le 16\log K,$$

$$\phi_{\min}\left(s,\frac{D_{m,j}}{|T_m^{(j)}|}\right)\ge\frac{\gamma(K)\rho(K)}{4}.$$

$\frac{D_{m,j}}{|T_m^{(j)}|}$: (Part of) Hessian of Lasso Regression at m-th Batch.

Lemma 6. *Under Assumptions 1–4, with probability at least* $1 - M\exp(\log d - \log K \cdot \Omega(\sqrt{Ts_0}/M)) - 2M^2 \cdot \exp\Big(O\Big(s_0 \cdot \frac{\log K \log d}{\gamma(K)\rho(K)}\Big) - \Omega(\rho^2(K)\sqrt{Ts_0})\Big) - M \cdot T^{-2}$, *for any* $m \in [M]$,

$$\|\hat{\theta}_m - \theta^\star\|_2 \leq \frac{800\sqrt{2}}{\gamma(K)\rho(K)} \cdot \boxed{\sqrt{s_0}} M \cdot \sqrt{\frac{\log K \cdot (2\log T + \log d)}{t_m}}.$$

Standard Lasso error bound: $\|\hat{\theta}-\theta^*\|_2 = O\left(\sqrt{\frac{s\log p}{n}}\right)$ w.p.a.1.

Corollary 9.20 *Suppose that, in addition to the conditions of Theorem 9.19, the optimal parameter* θ^* *belongs to* $\mathbb{M}$. *Then any optimal solution* $\widehat{\theta}$ *to the optimization problem* (9.3) *satisfies the bounds*

$$\Phi(\widehat{\theta} - \theta^*) \leq 6\frac{\lambda_n}{\kappa}\Psi^2(\bar{\mathbb{M}}), \tag{9.49a}$$

$$\|\widehat{\theta} - \theta^*\|^2 \leq 9\frac{\lambda_n^2}{\kappa^2}\Psi^2(\bar{\mathbb{M}}). \tag{9.49b}$$

(9.3): $\hat{\theta} \in \arg\min_{\theta} \mathcal{L}(\theta) + \mathcal{R}(\theta)$

Lasso: $\frac{1}{2n}\sum_{i=1}^{n}(y_i - x_i^T\theta)^2 + a|\theta|_1$.

a_n: $C\cdot\sqrt{\log p/n}$. Ψ: $\sqrt{s}$.

Key idea: $\hat{\theta} \approx \theta^*$:

upper bound of regret $\Rightarrow$ upper bound of est. error.

② Analyzing Regret Upper Bound.

Greedy choice of a_t: $\max_{a\in[K]} X_{a,t}^\top \hat{\theta}_m$

$\Rightarrow$ at m-th batch:

$$\max_{a\in[K]} (x_{t,a}-x_{t,a_t})^\top \theta^* \le \max_{a\in[K]} (x_{t,a}-x_{t,a_t})^\top(\theta^*-\hat{\theta}_{m-1})$$

$$\le 2\max_{a\in[K]} |x_{t,a}^\top(\theta^*-\hat{\theta}_{m-1})|.$$

(Sub-Gaussian maximal inequality) $\le 6\sqrt{\log(TK)}\,\|\theta^*-\hat{\theta}_{m-1}\|_2$ w.p.a 1.

(Take union bound over batch m:)

$$\max_{a\in[K]} (x_{t,a}-x_{t,a_t})^\top\theta^\star \le \frac{C}{\gamma(K)\rho(K)}\cdot\sqrt{s_0 M\log(TK)}\cdot\sqrt{\frac{\log K(2\log T+\log d)}{t_{m-1}}},$$

$$\forall t\in[t_{m-1}+1, t_m],$$

Summation over $m\ge 2$:

$$\sum_{m=2}^{M}\sum_{t=t_{m-1}+1}^{t_m} \max_{a\in[K]} (x_{t,a}-x_{t,a_t})^\top\theta^\star$$

$$\le \frac{C}{\gamma(K)\rho(K)}\cdot bM^{3/2}\cdot\sqrt{s_0\log K\log(TK)(\log d+2\log T)}$$

$$\le \frac{C'}{\gamma(K)\rho(K)}\cdot M^{3/2}$$

$$\cdot\sqrt{\log K\log(TK)(\log d+2\log T)}\sqrt{Ts_0}\left(\frac{T}{s_0}\right)^{\frac{1}{2(2^M-1)}},$$

For first batch when $\hat{\theta} = 0$: use a crude bound:

$$\sum_{t=1}^{t_1} \max_{a \in [K]} (x_{t,a} - x_{t,a_t})^\top \theta^\star \leq 2 \sum_{t=1}^{t_1} \left(\max_{a \in [K]} x_{t,a}^\top \theta^\star \right).$$

Applying a sub-Gaussian maximal inequality and a union bound over all $t \in [t_1]$, we have with probability at least $1 - T^{-2}$,

$$\sum_{t=1}^{t_1} \max_{a \in [K]} (x_{t,a} - x_{t,a_t})^\top \theta^\star \leq 6\sqrt{\log(KT)} \cdot t_1$$

$$= \Theta\left(\sqrt{\log(KT)} \sqrt{Ts_0} \cdot \left(\frac{T}{s_0} \right)^{\frac{1}{2(2^M - 1)}} \right).$$

Putting things together:

$$\mathbb{E}_{\theta^\star}[R_T(\mathbf{Alg})] \leq \frac{C'''}{\gamma(K)\rho(K)} \cdot M^{3/2} \cdot \sqrt{\log K \log(KT) \log(dT)} \cdot \sqrt{Ts_0} \cdot \left(\frac{T}{s_0} \right)^{\frac{1}{2(2^M - 1)}},$$

where $C''' > 0$ is a numerical constant.