Unsupervised Sequential Sensor Acquisition: With Contextual Information

1 Unsupervised Sensor Selection (USS)

Preliminaries and Notation: The set of real numbers is denoted by $\mathbb{R}$. For positive integer n, we let $[n] = \{1, \dots, n\}$. We let $M_1(\mathcal{X})$ to denote the set of probability distributions over some set $\mathcal{X}$. When $\mathcal{X}$ is finite with a cardinality of $d \doteq |\mathcal{X}|$, $M_1(\mathcal{X})$ denotes the d-dimensional probability simplex.

We first consider the *unsupervised, stochastic, cascaded sensor selection* problem ignoring any side information. We cast it as a special case of stochastic partial monitoring problem (SPM). We will then study the case when side information is available through context. Formally, a problem instance is specified by a pair $\theta = (P, c)$, where P is a distribution over the $K+1$ dimensional hypercube, and c is a K-dimensional, nonnegative valued vector of costs. While c is known to the learner from the start, P is initially unknown. Henceforth we identify problem instance θ by P. The instance parameters specify the learner-environment interaction as follows: In each round for $t = 1, 2, \dots$, the environment generates a $K+1$-dimensional binary vector $Y = (Y_t, Y_t^1, \dots, Y_t^K)$ chosen at random from P. Here, Y_t^i is the output of sensor i, while Y_t is a (hidden) label to be guessed by the learner. Simultaneously, the learner chooses an index $I_t \in [K]$ and observes the sensor outputs $Y_t^1, \dots, Y_t^{I_t}$, i.e., the learner goes through the first I_t sensors and observes their output. The sensors are known to be ordered from least accurate to most accurate, i.e., $\gamma_k \doteq \gamma_k(\theta) \doteq \mathbb{P}\left(Y_t \neq Y_t^k\right)$ is decreasing with k increasing. Knowing this, the learner's choice of I_t

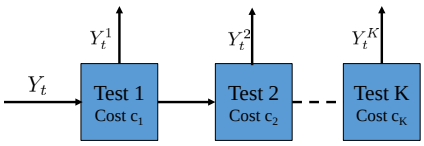

Figure 1: Cascaded Unsupervised Sequential Sensor Selection. Y_t is the hidden state of the instance and $Y_t^1, Y_t^2 \dots$ are test outputs. Not shown are features that a sensor could process to produce the output.

also indicates that he/she chooses I_t to predict the unknown label Y_t. Observing sensors is costly: The cost of choosing I_t is $C_{I_t} \doteq c_1 + \dots + c_{I_t}$. The total cost suffered by the learner in round t is thus $C_{I_t} + \mathbb{I}\{Y_t \neq Y_t^{I_t}\}$. The goal of the learner is to compete with the best choice given the hindsight of the values $(\gamma_k)_k$. Let $c(k, \theta) = \mathbb{E}\left[C_k + \mathbb{I}\{Y_t \neq Y_t^k\}\right]$ $(= C_k + \gamma_k)$ and $c^*(\theta) = \min_k c(k, \theta)$. The expected regret of learner up to the end of round T is $\mathfrak{R}_T(\theta) = \left(\sum_{t=1}^T \mathbb{E}\left[c(I_t, \theta)\right]\right) - T c^*(\theta)$. The parameter θ determines γ_k for all $k \in [K]$. Henceforth we use θ and vector $\gamma \doteq (\gamma_k)_k$ interchangeably.

Sublinear Regret: The quantification of the learning speed is given by the expected regret $\mathfrak{R}_T$, which, for brevity and when it does not cause confusion, we will just call regret. A sublinear expected regret, i.e., $\mathfrak{R}_T / T \to 0$ as $T \to \infty$ means that the learner in the long run collects almost as much reward on expectation as if the optimal action was known to it.

In the following we denote the optimal action as $a^*(\theta)$.

2 Weak Dominance

Let Θ_{SA} be the set of all stochastic, cascaded sensor acquisition problems. Thus, $\theta \in \Theta_{\mathrm{SA}}$ such that if $Y \sim \theta$ then $\gamma_k(\theta) := \mathbb{P}\left(Y \neq Y^k\right)$ is a decreasing sequence. Given a subset $\Theta \subset \Theta_{\mathrm{SA}}$, we say that Θ is *learnable* if there exists a learning algorithm $\mathfrak{A}$ such that for any $\theta \in \Theta$, the expected regret $\mathbb{E}\left[\mathfrak{R}_n(\mathfrak{A}, \theta)\right]$ of algorithm $\mathfrak{A}$ on instance θ is sublinear. A subset Θ is said to be a maximal learnable problem class if it is learnable and for any $\Theta' \subset \Theta_{\mathrm{SA}}$ superset of Θ, Θ' is not learnable.

Definition 1 (Weak Dominance (WD))**.** *An instance $\theta \in \Theta_{\mathrm{SA}}$ is said to satisfy the* weak dominance, property *if for $i = a^*(\theta)$,*

$$\rho = \min_{j > i} \frac{C_j - C_i}{\mathbb{P}\left(Y^i \neq Y^j\right)} \geq 1 \tag{1}$$

We denote the set of all instances in Θ_{SA} that satisfies this condition by Θ_{WD}.

Theorem 1. *The set Θ_{WD} is essentially a maximal learnable set.*

Define a set $\mathcal{A}$ as follows:

$$\mathcal{A} = \Big\{ i \in [K] : \forall j < i \ : \ C_i - C_j < \mathbb{P}\left(Y^i \neq Y^j\right) \text{ and } \forall j > i \ : \ C_j - C_i \geq \mathbb{P}\left(Y^i \neq Y^j\right) \Big\} .$$

Lemma 1. *Let the WD conditions holds. Then the set $\mathcal{A}$ contains an optimal action, and it is a singleton set.*

Proof. It is clear that under WD property $\mathcal{A}$ contains an optimal action. We prove the second part by contradiction. Assume that $i_1^\star, i_2^\star \in \mathcal{A}$ and $i_1^\star \neq i_2^\star$. WLOG, assume that $i_1^\star < i_2^*$. Since $i_1^\star \in \mathcal{A}$, we have $C_{j_1} - C_{i_1^\star} \geq \mathbb{P}\left(Y^{i_1^\star} \neq Y^{j_1}\right)$ for all $i_1^\star < j_1$. Also, since $i_2^\star \in \mathcal{A}$, we have $C_{i_2^\star} - C_{j_2} < \mathbb{P}\left(Y^{i_2^\star} \neq Y^{j_2}\right)$ for all $j_2 < i_2^\star$.

Now, setting $j_1 = i_2^\star$ and $j_2 = i_1^\star$ above, we get $C_{i_2^\star} - C_{i_1^\star} \geq \mathbb{P}\left(Y^{i_1^\star} \neq Y^{i_2^\star}\right)$ and $C_{i_2^\star} - C_{i_1^\star} < \mathbb{P}\left(Y^{i_2^\star} \neq Y^{i_1^\star}\right)$. Hence a a contradiction. □

3 Algorithms

Define for any $i \neq j, \gamma_{ij} \doteq \Pr\{Y^i \neq Y^j\}$ and $\Delta_{ij} = C_j - C_i - \gamma_{ij}$. Our key insight for developing the algorithm is based on the fact that (see Def. 3), under WD property, the set $\{i \in [K] : \min_{j>i} \Delta_{ij} \geq 0\}$ includes the optimal arm. We use this idea in Algorithm 1 to identify the optimal arm when an instance of USS satisfies WD.

Algorithm 1 Algorithm for USS with WD property

Play action K and observe $Y^1, \ldots, Y^K$
Set $\hat{\gamma}_{ij}(1) \leftarrow \mathbb{I}_{\{Y^i \neq Y^j\}}$ for all $i < j$
$n_i(1) \leftarrow \mathbb{I}_{\{i=K\}}\ \forall i \in [K]$
for $t = 2, 3, \ldots$ **do**
  $U_{ij}(t)) \leftarrow \hat{\gamma}_{ij}(t-1) + \sqrt{\frac{1.5\log(t)}{n_j(t-1)}}$ for all $i < j$
  $L_{ij}(t) \leftarrow \hat{\gamma}_{ij}(t-1) - \sqrt{\frac{1.5\log(t)}{n_j(t-1)}}$ for all $i < j$
  $\hat{\Delta}_{ij}(t)) \leftarrow C_j - C_i - L_{ij}(t)$ for all $i < j$
  $A_t \leftarrow \left\{ i \in [K] : \min_{j>i} \hat{\Delta}_{ij}(t) \geq 0 \right\}$
  Set $\hat{I}_t \leftarrow \arg\min A_t \cup \{K\}$
  $B_t \leftarrow \left\{ i > \hat{I}_t : C_i - C_{\hat{I}_t} - U^t_{\hat{I}_t i} \leq 0 \right\}$
  $I_t \leftarrow \arg\min B_t \cup \{\hat{I}_t\}$
  Play I_t and observe $Y^1, \ldots, Y^{I_t}$.
  for $i = 1, \ldots, I_t - 1$ **do**
    $n_i(t) \leftarrow n_i(t-1) + 1$
    for $j = i+1, \ldots, I_t$ **do**
  $\hat{\gamma}_{ij}(t) \leftarrow \left(1 - \frac{1}{n_j(t)}\right) \hat{\gamma}_{ij}(t-1) + \frac{1}{n_j(t)} \mathbb{I}_{\{Y^j \neq Y^i\}}$
    end for
  end for
end for

The algorithm works as follows. In each round, t, based on history, we keep track of estimates, $\hat{\gamma}_{ij}(t)$, of disagreements between sensor i and sensor j. In the first round, the algorithm plays arm K and initializes its values. In each subsequent round, the algorithm computes the upper confidence value of $\hat{\gamma}_{ij}(t)$ denoted as $U_{ij}(t)$ (5) for all pairs (i, j) and orders the arms: i is considered better than arm j if $C_j - C_i \geq U_{ij}(t)$. Specifically, the algorithm plays an arm i that satisfies $C_j - C_i \geq U_{ij}(t)$ for all $j > i$ (8). If no such arm is found, then it plays arm K. $n_j(t), j \in [K]$ counts the total number of observation of pairs (Y^i, Y^j), for all $i < j$, till round t and uses it to update the estimates $\hat{\gamma}_{ij}(t)$ (**??**).

4 With Contextual Information

When the contextual information is available, the problem of the unsupervised, stochastic, cascaded sensor selection can be specified by a sequence $(X_t, Y_t, (Y_t^a)_{a\in[K]})$ of iid random variables, where K is the number of sensors, $X_t \in \mathcal{X} \subset \mathcal{R}'$ is the context at time t, Y_t is the associated label and Y_t^a is label predicted by sensor $a \in [K]$. The $K(> 1)$ sensors are known to be ordered from least accurate to most accurate, i.e., expected error for a context x, $\gamma_a(x) := \Pr\{Y \neq Y^a | X = x\}$ of sensor a is a decreasing function of a. As the setting considered is unsupervised, we do not have any knowledge about expected error $\gamma_a(x)$ except for their ordering which is assumed to remain the same for all contexts. Hence, given a context x, we need to make some assumptions for optimally choosing best option.

For a given context, the total cost for selecting sensor i is $C_i + \gamma_i(x)$, where $C_i = c_1 + \cdots + c_i$. The goal of the learner is to select an action that has smallest total cost (henceforth simply referred as cost) for a given context. For a given sequence of contexts $\{x_t\}_1^n$, we evaluate the learning performance a policy/algorithm that selects action $I_t := I_t(x_t)$ in round t in terms of the regret defined as follows:

$$R_n = \sum_{t=1}^{n} \left(C_{I_t} + \gamma_{I_t}(x_t) - (C_{i_t^\star} + \gamma_{i_t^\star}(x_t)) \right) \tag{2}$$

where $i_t^\star = \text{argmin}_i (C_i + \alpha\gamma_i(x_t))$. The goal of the learner is to learn a policy that has lowest sub-linear regret, i.e., $R_n/n \to 0$ as $n \to \infty$ means that the learner in the long run collects almost as much reward on expectation as when the optimal action is known.

4.1 Some Definitions

Let Θ_{SA} be the set of all stochastic, cascaded sensor acquisition problems.

Definition 2 (Weak Dominance with Contextual Information (WDC))**.** *An instance $\theta \subset \Theta_{SA}$ is said to satisfy the weak dominance with contextual information property if for any context x for all $j \in [K]$, $j > i^\star(x)$*

$$C_j - C_{i^\star(x)} \geq \alpha \mathbb{P}(Y^{i^\star(x)} \neq Y^j | X = x) \tag{3}$$

where $i^\star(x)$ is the optimal action for context x.

Definition 3 (Sub-Gaussian Random Variables)**.** *Let (X_t) be $\{\mathcal{F}_t\}$ adapted: $\{\mathcal{F}_t\}$ is an increasing sequence of sigma fields such that X_t is $\mathcal{F}_t$-measurable with $\mathbb{E}[X_t|\mathcal{F}_{t-1}] = 0$. Then, $\{X_t\}$ is a Sub-Gaussian with parameter R if there exits $R > 0$ such that for all t and $\lambda \in \mathcal{R}$ the following holds:*

$$\mathbb{E}\left[e^{\lambda X_t}|\mathcal{F}_{t-1}\right] \leq e^{\lambda^2 R^2/2} \tag{4}$$

5 Disagreement based Model with Linear Assumption

We assume that the disagreement between observed labels of sensors i, j for a given context x satisfies

$$\mathbb{P}\left\{\widehat{Y}_i \neq \widehat{Y}_j | X = x\right\} = \phi_{ij}(x)^\top \theta^\star \tag{5}$$

where $\theta^\star \in \mathcal{R}^d$ is fixed but unknown and for all $i < j$, $\phi_{ij} : \mathcal{X} \to \mathcal{R}^d$ is a feature map. We assume that $d > d'$ and $||\theta^\star||_2 \leq S$.

5.1 Selection Criteria for Choosing Optimal sensor

The optimality of action (choosing sensor) can be captured in terms of marginal costs and marginal errors. In particular, for any context x a sensor i is optimal if for all $j > i$ the marginal increase in cost, $C_j - C_i$, is larger than the marginal decrease in error, $\gamma_i(x) - \gamma_j(x)$ i.e.,

$$\underbrace{C_j - C_i}_{\text{marginal cost}} \geq \underbrace{\gamma_i(x) - \gamma_j(x)}_{\text{marginal error}} \tag{6}$$

Since $\gamma_i(x)$ is unknown for any i and x, equation (6) does not lead to an computational algorithm for sensor selection. So, we use the relation $\gamma_i(x) - \gamma_j(x) \leq \mathbb{P}\{Y^i \neq Y^j | X = x\}$ [1][Prop. 3] where $\widehat{Y}_i$ is label given for context x by sensor i. Then using weak dominance condition of [1][Def. 3], the weaker decision rule for selecting best sensor is given by

Proposition 1. *Assume that the weak dominance property holds. Let $i^\star(x)$ be the optimal action for context x. Then,*

$$\forall j > i^\star(x), C_j - C_{i^\star(x)} \geq \mathbb{P}\{Y^{i^\star(x)} \neq Y^j | X = x\} \tag{7}$$

In the following we develop an algorithm that uses the above decision rule to select an optimal arm. The algorithm estimate the disagreement probabilities by comparing the predictions of the selected sensors. When the predictions of sensors i, j are compared, the indicator value of the disagreements can be interpreted as a noisy version of their disagreement probability as give below

$$\mathbb{1}_{\{Y^i \neq Y^j | X = x\}} = \mathbb{P}\{Y^i \neq Y^j | X = x\} + \eta_{i,j}(x)$$

where $\eta_{i,j}(x)$ is a zero-mean bounded random variable and hence sub-Gaussian with some parameter R (Need to compute value of R).

5.2 Disagreement Model under Weak Dominance Property

Algorithm 2 Linear Disagreement Model under Weak Dominance Property

Input: $\alpha > 0, R > 0, \lambda > 0, m \in \mathbb{N}, d \in \mathbb{N}, \delta \in (0, 1)$
$\overline{V}_0 = \lambda \mathrm{I}_d$, $M_0 = \boldsymbol{e} 0_d$
for $t = 1, 2, \ldots, m$ **do**
  Observe context, $x_t \in \mathbb{R}^d$
  Choose arm $I_t = K$
  Observe labels $\{\widehat{Y}_1^t, \widehat{Y}_2^t, \ldots, \widehat{Y}_{I_t}^t\}$; $\widehat{Y}_i^t \in \{0, 1\}$
  Construct feature vector using features of input (x_t) and arm i and j; $x_{i,j}^t \in \mathbb{R}^d$
  $\overline{V}_t = \overline{V}_{t-1} + \sum_{i=1}^{I_t - 1} x_{i,I_t}^t (x_{i,I_t}^t)^\top$
  $M_t = M_{t-1} + \sum_{i=1}^{I_t - 1} \mathbb{1}_{\{\widehat{Y}_i^t \neq \widehat{Y}_{I_t}\}} x_{i,I_t}$
end for
for $t = m + 1, m + 2, \ldots$ **do**
  $\widehat{\theta}_t = \overline{V}_{t-1}^{-1} M_{t-1}$
  Observe context, x_t
  $\beta_t = R\sqrt{d \log\left(1 + \frac{tKL^2}{\lambda d}\right) + 2\log\left(\frac{1}{\delta}\right)} + \lambda^{1/2} S$

$$\mathcal{A}_t = \Big\{ i \in [K] : \forall j < i,\, C_i - C_j < \alpha \langle x_{i,j}^t, \widehat{\theta}_t \rangle - \beta_t ||x_{i,j}^t||_{\overline{V}_t^{-1}} \text{ and } \forall j > i,\, C_j - C_i \geq \alpha \langle x_{i,j}^t, \widehat{\theta}_t \rangle - \beta_t ||x_{i,j}^t||_{\overline{V}_t^{-1}} \Big\} .$$

  If $\mathcal{A}_t \neq \emptyset$, set $I_t = \arg\min \mathcal{A}_t$. Else $I_t = K$.
  Observe labels $\{\widehat{Y}_1^t, \widehat{Y}_2^t, \ldots, \widehat{Y}_{I_t}^t\}$
  $\overline{V}_t = \overline{V}_{t-1} + \sum_{i=1}^{I_t - 1} x_{i,I_t}^t (x_{i,I_t}^t)^\top$
  $M_t = M_{t-1} + \sum_{i=1}^{I_t - 1} \mathbb{1}_{\{\widehat{Y}_i^t \neq \widehat{Y}_{I_t}^t\}} x_{i,I_t}$
end for

5.3 Regret Analysis of Disagreement Model

Regret analysis is similar to that for the OFUL algorithm given in paper [2].

5.3.1 Important Lemmas and Theorems

Some of the lemma derived using results from paper [2, 1].

Lemma 2. *For any $i, j \in [K]$ and any context x the following relations hold.*

$$\gamma_i(x) - \gamma_j(x) = \Pr\{Y^i \neq Y^j | X = x\} - 2\Pr\{Y^i = Y, Y^j \neq Y | X = x\}. \quad (8)$$

$$\gamma_i(x) - \gamma_j(x) = -\Pr\{Y^i \neq Y^j | X = x\} + 2\Pr\{Y^i \neq Y, Y^j = Y | X = x\}. \quad (9)$$

Lemma 3. *Let $\overline{V}_t = \sum_{s=1}^{t}\sum_{j=1}^{I_t-1} x^s_{j,I_t}(x^s_{j,I_t})^\top + \overline{V}_0$ and $S_t = \sum_{s=1}^{t}\sum_{j=1}^{I_t-1} \eta^s_{j,I_t} x^s_{j,I_t}$. Then, for any $\delta > 0$ with probability at least $1-\delta$, for all $t \geq 0$,*

$$||S_t||^2_{\overline{V}_t^-} \leq 2R^2 \log\left(\frac{det(\overline{V}_t)^{1/2} det(\overline{V}_0)^{-1/2}}{\delta}\right)$$

Lemma 4. *Let $\overline{V}_t$ be as defined in Lemma 3 and let $\overline{V}_0 = \lambda I_d$. Assume that $||\theta^\star||_2 \leq S$. Then, for any $\delta > 0$ with probability at least $1-\delta$, for all $t \geq 0$ and x^t,*

$$|\langle x^t_{i,j}, \theta^\star\rangle - \langle x^t_{i,j}, \widehat{\theta}_t\rangle| \leq \beta_t ||x^t_{i,j}||_{\overline{V}_t^-}$$

where $\beta_t = R\sqrt{d\log\left(1 + \frac{tKL^2}{\lambda d}\right) + 2\log\left(\frac{1}{\delta}\right)} + \lambda^{1/2}S$

Lemma 5. *For any $1 \leq i < j \leq K$, $\{x^s_{i,j}\}_{s=1}^{\infty}$ be a sequence in $\mathbb{R}^d$, V is a $d \times d$ positive definite matrix and define $\overline{V}_t = V + \sum_{s=1}^{t}\sum_{j=1}^{I_t-1} x^s_{j,I_t}(x^s_{j,I_t})^\top$. Then, we have that*

$$\sum_{t=1}^{n} \min\{1, ||x^s_{i,j}||^2_{\overline{V}_{t-1}^-}\} \leq 2\log\frac{det(\overline{V}_n)}{det(V)}$$

Further if $||x^s_{i,j}||_2 \leq L$ for all t and $\lambda_{min}(V) \geq \max\{1, L^2\}$, then

$$\log\frac{det(\overline{V}_n)}{det(V)} \leq \sum_{t=1}^{n} ||x^s_{i,j}||^2_{\overline{V}_{t-1}^-} \leq 2\log\frac{det(\overline{V}_n)}{det(V)}$$

Lemma 6. *(Determinant-Trace Inequality) Suppose K is total number of sensors/tests, for any $1 \leq s \leq t$ and $1 \leq j < I_t \leq K$, $||x^s_{j,I_t}||_2 \leq L$ where $x^s_{j,I_t} \in \mathbb{R}^d$. Let $\overline{V}_t = \lambda I_d + \sum_{s=1}^{t}\sum_{j=1}^{I_t-1} x^s_{j,I_t}(x^s_{j,I_t})^\top$, for some $\lambda > 0$. Then*

$$det(\overline{V}_t) \leq (\lambda + tKL^2/d)^2$$

Proof. By inequality of arithmetic and geometric means,

$$det(\overline{V}_t) \leq (trace(\overline{V}_t)/d)^d$$

As the trace of matrix is a linear mapping i.e. $trace(A+B) = trace(A) + trace(B)$, hence

$$\begin{aligned} trace(\overline{V}_t) &= trace(\lambda I_d) + \sum_{s=1}^{t}\sum_{j=1}^{I_t-1} trace(x^s_{j,I_t}(x^s_{j,I_t})^\top) \\ &= d\lambda + \sum_{s=1}^{t}\sum_{j=1}^{I_t-1} ||x^s_{j,I_t}||_2^2 \\ &= d\lambda + \sum_{j=1}^{I_t-1}\sum_{s=1}^{t} ||x^s_{j,I_t}||_2^2 \\ \Rightarrow trace(\overline{V}_t) &\leq d\lambda + tKL^2 \text{ as } I_t \leq K \end{aligned}$$

Using this result, we get

$$\begin{aligned} det(\overline{V}_t) &\leq (trace(\overline{V}_t)/d)^d \\ &\leq \left((\lambda d + tKL^2)/d\right)^d \\ \Rightarrow det(\overline{V}_t) &\leq (\lambda + tKL^2/d)^d \end{aligned}$$

□

Lemma 7. *Assume same as in lem 6, let $A_{I_t} = \frac{1}{t}\sum_{s=1}^{t} I_s$ where I_s is the optimal sensor/test chosen by algorithm for data instance x_s.*

$$det(\overline{V}_t) \leq (\lambda + tA_{I_t}L^2/d)^2$$

Proof. As we know,

$$\begin{aligned} trace(\overline{V}_t) &= d\lambda + \sum_{j=1}^{I_t-1}\sum_{s=1}^{t} ||x^s_{j,I_t}||_2^2 \\ &\leq d\lambda + tA_{I_t}L^2 \text{ where } A_{I_t} \leq K \end{aligned}$$

By inequality of arithmetic and geometric means,

$$\begin{aligned} det(\overline{V}_t) &\leq (trace(\overline{V}_t)/d)^d \\ \Rightarrow det(\overline{V}_t) &\leq (\lambda + tA_{I_t}L^2/d)^2 \end{aligned}$$

□

5.3.2 Instantaneous Regret

The instantaneous regret of algorithm is the regret incurred for any time t.

Theorem 2. */ This proof is not complete / The instantaneous regret r_t of algorithm with probability at least $1-\delta$ is:*

$$r_t \leq 2\alpha\beta_t ||x^t_{I_t,i^\star}||_{\overline{V}^-}$$

Proof.

$$\begin{aligned} r_t &= C_{I_t} + \alpha\gamma_{I_t}(x_t) - \text{argmin}_i(C_i + \alpha\gamma_i(x_t)) \\ &= C_{I_t} + \alpha\gamma_{I_t}(x_t) - (C_{i^\star_t} + \alpha\gamma_{i^\star_t}(x_t)) \\ &= C_{I_t} - C_{i^\star_t} + \alpha(\gamma_{I_t}(x_t) - \gamma_{i^\star_t}(x_t)) \end{aligned}$$

where $i_t^\star$ is the best arm in round t. Two possibilities arises: $I_t > i_t^\star$ or $I_t \leq i_t^\star$. First consider the case $i_t^\star > I_t$. From the previous lemma we have

$$\begin{aligned}
r_t &= C_{I_t} - C_{i_t^\star} + \alpha(\gamma_{I_t}(x_t) - \gamma_{i_t^\star}(x_t)) \\
&\leq -\alpha\langle x^t_{I_i,i^\star}, \widehat{\theta}_t\rangle + \alpha\beta_t||x^t_{I_t,i^\star}||_{\overline{V}_t^-} + \\
&\quad \alpha \Pr\{Y^{I_t} \neq Y^{i_t^\star} | X = x_t\} \text{ from (??) and (8)} \\
&= \alpha(\langle x^t_{I_i,i^\star}, \theta^\star - \widehat{\theta}_t\rangle + \alpha\beta_t||x^t_{I_t,i^\star}||_{\overline{V}_t^-} \\
&\leq \alpha\beta_t||x^t_{I_t,i^\star}||_{\overline{V}_t^-} + \alpha\beta_t||x^t_{I_t,i^\star}||_{\overline{V}_t^-} \text{ from lemma 4} \\
&= 2\alpha\beta_t||x^t_{I_t,i^\star}||_{\overline{V}_t^-}
\end{aligned}$$

Now consider the case $i_t^\star \leq I_t$. Using Equation 9 we get

$$\begin{aligned}
r_t &= C_{I_t} - C_{i_t^\star} + \alpha(\gamma_{I_t}(x_t) - \gamma_{i_t^\star}(x_t)) \\
&= C_{I_t} - C_{i_t^\star} - \alpha \Pr\{Y_t^{I_t} \neq Y_t^{i_t^\star} | X = x_t\} + \\
&\quad 2\alpha \Pr\{Y_t^{I_t} \neq Y_t, Y_t^{i_t^\star} = Y_t | X = x_t\} \\
&\leq \alpha\langle x_{I_t,i_t^\star}, \widehat{\theta}_t\rangle + \alpha\beta_t||x^t_{I_t,i^\star}||_{\overline{V}_t^-} - \alpha\langle x^t_{I_i,i^\star}, \theta^\star\rangle + \\
&\quad 2\alpha \Pr\{Y_t^{I_t} \neq Y_t, Y_t^{i_t^\star} = Y_t | X = x_t\} \\
&\leq \alpha\langle x_{I_t,i_t^\star}, \widehat{\theta}_t - \theta^\star\rangle + \alpha\beta_t||x^t_{I_t,i^\star}||_{\overline{V}_t^-} + \\
&\quad 2\alpha \Pr\{Y_t^{I_t} \neq Y_t, Y_t^{i_t^\star} = Y_t | X = x_t\} \\
&\leq 2\alpha\beta_t||x^t_{I_t,i^\star}||_{\overline{V}_t^-} + 2\alpha \Pr\{Y_t^{I_t} \neq Y_t, Y_t^{i_t^\star} = Y_t | X = x_t\} \\
&= (??)2\alpha\beta_t||x^t_{I_t,i^\star}||_{\overline{V}_t^-} + 2\alpha \Pr\{I_t \neq i_t^\star | X = x_t\}
\end{aligned}$$

How to get rid of the second term here? □

5.3.3 Cumulative Regret

Theorem 3. *The cumulative regret R_T of disagreement model for time horizon T with probability at least $1-\delta$ is:*

$$R_T \leq 2\alpha\left(R\sqrt{d\log\left(1+\frac{KTL^2}{\lambda d}\right)+2\log\left(\frac{1}{\delta}\right)}+\lambda^{1/2}S\right)\sqrt{2Td\log\left(1+\frac{KTL^2}{\lambda d}\right)}$$

Proof.

$$\begin{aligned}
R_T &= \sum_{t=1}^{T} r_t \\
&\leq \sqrt{T\sum_{t=1}^{T} r_t^2} \\
&\leq \sqrt{T\sum_{t=1}^{T}\left[2\alpha\beta_t||x^t_{i,j}||_{\overline{V}^-}\right]^2} \text{ from theorem 2} \\
&\text{As } \beta_t \leq \beta_T \text{ because } \beta_t \text{ is not decreasing with } t \text{ (by definition} \\
&\leq 2\alpha\beta_T\sqrt{T\sum_{t=1}^{T}\left[||x^t_{i,j}||_{\overline{V}^-}\right]^2} \\
&\leq 2\alpha\beta_T\sqrt{T * 2\log\left(\frac{det(\overline{V})}{det(\lambda I_d)}\right)} \text{ from lemma 5} \\
&\leq 2\alpha\beta_T\sqrt{2Td\log\left(1+\frac{tKL^2}{\lambda d}\right)} \text{ from lemma 6} \\
&\leq 2\alpha\beta_T\sqrt{2Td\log\left(1+\frac{KTL^2}{\lambda d}\right)} \text{ as } t \leq T \\
&\text{Now using value of } \beta_T \text{ and lemma 6} \\
&\leq 2\alpha\left(R\sqrt{d\log\left(1+\frac{KTL^2}{\lambda d}\right)+2\log\left(\frac{1}{\delta}\right)}+\lambda^{1/2}S\right) \\
&\sqrt{2Td\log\left(1+\frac{KTL^2}{\lambda d}\right)}
\end{aligned}$$

□

References

[1] M. Hanawal, C. Szepesvari, and V. Saligrama, "Unsupervised sequential sensor acquisition," in *Artificial Intelligence and Statistics*, 2017, pp. 803–811.

[2] Y. Abbasi-Yadkori, D. Pál, and C. Szepesvári, "Improved algorithms for linear stochastic bandits," in *Advances in Neural Information Processing Systems*, 2011, pp. 2312–2320.