Technical Report of Trading Vector Data via Hierarchical Bandits

APPENDIX A

Proof of Theorem 1. We prove the regret bound by establishing a concentration inequality on the estimated reward for any given configuration. The main difficulty arises from the fact that rewards generated by a configuration are not independent across time and may be highly correlated due to the shared pricing mechanism. Classical Chernoff bounds are thus inapplicable in this setting. To overcome this, we leverage martingale-based concentration techniques.

Specifically, we demonstrate that the deviation between cumulative observed rewards and their expected values forms a martingale sequence. This allows us to apply Azuma-Hoeffding-type inequalities.

Since each cluster maintains an independent learning process, we omit explicit indexing by ς_t for brevity. Consider a fixed configuration e, and let t_ℓ denote the time step when e is selected for the ℓ^{th} time. Suppose $t_1, t_2, \dots$ are fixed. Define $z_e(0) = 0$ and for $\ell > 0$,

$$z_e(\ell) = z_e(\ell-1) + r_e(\ell) - \mu\left(e, p^\ell\right) \tag{1}$$

where p^ℓ is the pricing arm that is played during the ℓ^{th} instance of configuration e.

We claim that the random sequence $\{z_e(\ell) | \ell = 1, 2, \dots\}$ is a martingale. Indeed, for $\ell \geq 1$, we have:

$$\begin{aligned}
&\mathbb{E}[z_e(\ell) \mid z_e(\ell-1) = z] \\
&= z + \mathbb{E}[r_e(\ell) - \mu(e, p^\ell) \mid z_e(\ell-1) = z] \\
&= z + \int_{p \in \mathcal{P}} f^\ell(p \mid z_e(\ell-1) = z) \cdot \\
&\qquad \mathbb{E}[r_e(\ell) - \mu(e,p) \mid z_e(\ell-1) = z,\, p^\ell = p]\, dp \\
&= z
\end{aligned} \tag{2}$$

where $f^\ell(p \mid z_e(\ell-1) = z)$ is the conditional probability density function over prices $p \in \mathcal{P}$, given the historical state $z = z_e(\ell-1)$ at the ℓ-th play. It is easy to verify that $\forall \ell$, $|z_e(\ell) - z_e(\ell-1)| < 1$, hence by Azuma-Hoeffding inequality:

$$\Pr[|z_e(\ell)| \geq v] = \Pr[|z_e(\ell) - z_e(0)| \geq v] \leq 2\exp\left(\frac{-v^2}{2\ell}\right) \tag{3}$$

Consider $e = e^*$.We note that since $t_1, t_2, \dots$ are fixed, the number of times e^* is selected, denoted by $n_{e^*}(t)$, is also fixed. From (3), we have $\forall t$:

$$\begin{aligned}
&\Pr[U_{e^*}(t) \leq l_{e^*}] \\
&= \Pr[(n_{e^*}(t)) U_{e^*}(t) \leq (n_{e^*}(t)) l_{e^*}] \\
&\leq \Pr\left[\sum_{\ell=1}^{n_{e^*}(t)} r_{e^*}(\ell) + \sqrt{2(n_{e^*}(t))\log(t)} \leq \sum_{p | p \in \mathcal{P}} n^p_{e^*}(t) \mu(e^*, p)\right] \\
&= \Pr\left[z_{e^*}(n_{e^*}(t)) \leq -\sqrt{2(n_{e^*}(t))\log(t)}\right] \\
&\leq \frac{1}{t}
\end{aligned} \tag{4}$$

where $\mathcal{P}_{e^*}(t)$ is the set of prices observed for the configuration e^* up to time t and $n^p_{e^*}(t)$ is the number of time p appears. Now consider $e \neq e^*$, and any fixed t such that $n_e(t) = \eta \geq \gamma_e(t) = \frac{8\log(t)}{\delta_e^2}$. From (3), we have

$$\begin{aligned}
&\Pr\left[U_e(t) \geq u_e + \delta_e \mid n_e(t) = \eta\right] \\
&= \Pr\left[\eta U_e(t) \geq \eta(u_e + \delta_e) \mid n_e(t) = \eta\right] \\
&\leq \Pr\left[\sum_{\ell=1}^{\eta} r_e(\ell) + \sqrt{2\eta\log(t)} \geq \right. \\
&\qquad \left. \sum_{p | p \in \mathcal{P}} n^p_e(t)\mu(e,p) + \eta\delta_e \mid n_e(t) = \eta\right] \\
&= \Pr\left[z_e(\eta) \geq \eta\delta_e - \sqrt{2\eta\log(t)} \mid n_e(t) = \eta\right] \\
&= \Pr\left[z_e(\eta) \geq \sqrt{2\eta\log(t)} \mid n_e(t) = \eta\right] \\
&\leq \frac{1}{t}
\end{aligned} \tag{5}$$

Suppose we have an event $\mathcal{A}$ such that $\mathcal{A} \Rightarrow \mathcal{B} \vee \mathcal{C} \vee \mathcal{D}$. Then for any event $\mathcal{E}$:

$$\mathcal{A} \Rightarrow (\mathcal{A} \wedge \mathcal{E}) \vee (\mathcal{A} \wedge \neg\mathcal{E}) \Rightarrow (\mathcal{A} \wedge \mathcal{E}) \vee ((\mathcal{B} \vee \mathcal{C} \vee \mathcal{D}) \wedge \neg\mathcal{E})$$

Hence,

$$\begin{aligned}
\Pr[\mathcal{A}] &\leq \Pr[\mathcal{B} \wedge \neg\mathcal{E}] + \Pr[\mathcal{C} \wedge \neg\mathcal{E}] + \Pr[\mathcal{A} \wedge \mathcal{E}] + \Pr[\mathcal{D} \wedge \neg\mathcal{E}] \\
&\leq \Pr[\mathcal{B}] + \Pr[\mathcal{C}] + \Pr[\mathcal{A} \wedge \mathcal{E}] + \Pr[\mathcal{D} \wedge \neg\mathcal{E}]
\end{aligned} \tag{6}$$

Let e be a suboptimal configuration. The process will initialize by playing configuration e at the beginning of the online phase (define this event as $\mathcal{B}$).

- $\mathcal{A} = \mathbb{1}(e_t = e)$
- $\mathcal{C} = \mathbb{1}(U_{e^*}(t) \leq l_{e^*})$

- $\mathcal{D} = \mathbb{1}(U_e(t) \geq u_e + \delta_e)$
- $\mathcal{E} = \mathbb{1}(n_e(t) < \gamma_e(t))$

where $\mathbb{1}(\cdot)$ is the indicator function.

We now have:

$$\begin{aligned}
&\mathbb{E}[n_e(T_\varsigma)] \\
&= \sum_{t=1}^{T_\varsigma} \Pr[e_t = e] \\
&\leq 1 + \sum_{t=1}^{T_\varsigma} \Pr[U_{e^*}(t) \leq l_{e^*}] + \sum_{t=1}^{T_\varsigma} \Pr[e_t = e \mid n_e(t) < \gamma_e(t)] \\
&+ \sum_{t=1}^{T_\varsigma} \Pr[U_e(t) \geq u_e + \delta_e \mid n_e(t) \geq \gamma_e(t)] \\
&\mathbb{E}[n_e(T_\varsigma)] \leq 1 + \gamma_e(t) + \sum_{t=1}^{T_\varsigma} \Pr[U_{e^*}(t) \leq l_{i^*}] \\
&+ \sum_{t=1}^{T_\varsigma} \Pr[U_e(t) \geq u_e + \delta_e \mid n_e(t) \geq \gamma_e(t)]
\end{aligned} \tag{7}$$

We now have:

$$\begin{aligned}
\mathbb{E}[n_e(T_\varsigma)] &\leq 1 + \gamma_e(T_\varsigma) + \sum_{t=1}^{T_\varsigma} \frac{1}{t} + \sum_{t=1}^{T_\varsigma} \frac{1}{t} \\
&\leq 3 + \gamma_e(T_\varsigma) + 2\ln(T_\varsigma)
\end{aligned} \tag{8}$$

Since $r_e(t) \in [0, \bar{r}]$, summing over inequality (8) completes the proof.

□

APPENDIX B

Proof of Lemma 1. We aim to show that for any small neighborhood $B = \Xi \times I$ centered at (e_t, a_{j_t}), the function $\chi^\varsigma(e,p)$ can be uniformly approximated by a polynomial of degree $n-1$ with an explicit bound on the approximation error. The proof proceeds via the multivariate Taylor expansion with the Lagrange remainder and uses the Hölder smoothness condition from Assumption 1 to bound the remainder term.

By the multivariate Taylor series with Lagrange remainder, for any $(e,p) \in B$, there exists a point $(\tilde{e}, \tilde{p}) \in B$ such that

$$\begin{aligned}
\chi^\varsigma(e,p) = &\sum_{|\boldsymbol{\alpha}| \leq n-2} \frac{\partial^{\boldsymbol{\alpha}} \chi^\varsigma(e_t, a_{j_t})}{|\boldsymbol{\alpha}|!} (e - e_t,\, p - a_{j_t})^{\boldsymbol{\alpha}} \\
&+ \sum_{|\boldsymbol{\alpha}| = n-1} \frac{\partial^{\boldsymbol{\alpha}} \chi^\varsigma(\tilde{e}, \tilde{p})}{|\boldsymbol{\alpha}|!} (e - e_t,\, p - a_{j_t})^{\boldsymbol{\alpha}}.
\end{aligned} \tag{9}$$

Define the polynomial approximation

$$P^B(e,p) = \sum_{|\boldsymbol{\alpha}| \leq n-1} \lambda_{\boldsymbol{\alpha}} (e - e_t, p - a_{j_t})^{\boldsymbol{\alpha}}, \tag{10}$$

where the coefficients are given by $\lambda_{\boldsymbol{\alpha}} = \partial^{\boldsymbol{\alpha}} f(e_t, a)/|\boldsymbol{\alpha}|!$. Under the Hölder smoothness condition in Assumption 1, all mixed partial derivatives of order $k-1$ are Lipschitz continuous with constant β. Therefore, the remainder term of order k can be bounded as follows:

$$\begin{aligned}
&\left|\chi^\varsigma(e,p)) - P^B(e,p)\right| \\
&= \left|\tfrac{1}{(n-1)!} \sum_{|\boldsymbol{\alpha}|=n-1} \left(\partial^\alpha \chi^\varsigma(\tilde{e}, \tilde{p}) - \partial^\alpha \chi^\varsigma(e_t, p)\right) (e - e_t, p - a_{j_t})^{\boldsymbol{\alpha}}\right| \\
&\leq \left|\tfrac{1}{(n-1)!} \sum_{|\boldsymbol{\alpha}|=n-1} \left(\partial^\alpha \chi^\varsigma(\tilde{e}, \tilde{p}) - \partial^\alpha \chi^\varsigma(e_t, p)\right) |e - e_t, p - a_{j_t}|^{|\boldsymbol{\alpha}|}\right| \\
&\leq \tfrac{1}{(n-1)!} \sum_{|\boldsymbol{\alpha}|=n-1} \beta |(e - e_t, p - a_{j_t})|^n \\
&\leq \frac{\beta n}{(n-1)!} \left(\eta + (\underline{p} + \bar{p})/N\right)^n,
\end{aligned} \tag{11}$$

which completes the proof.

□

APPENDIX C

Proof of Lemma 2. We aim to upper bound the cumulative regret contributed by pricing decisions within a fixed configuration e. The key idea is to apply a uniform confidence bound (via Lemma 1 in Appendix F) over all time steps and invoke the elliptical potential lemma to control the exploration terms. The bound is derived by splitting the error into a confidence bonus term and a smoothness residual term, followed by applying norm-based inequalities and bounding determinants.

Fix configuration e. Invoke Lemma 1 in Appendix F for $\tau = 1, 2, \ldots, t$ with confidence level $\delta = 1/T_\varsigma^3$.

Let

$$\Lambda_\tau = I + \sum_{\tau' < \tau} \phi(e_{\tau'}, \hat{p}_{\tau'}) \phi(e_{\tau'}, \hat{p}_{\tau'})^\top \in \mathbb{R}^{d \times d} \tag{12}$$

be the matrix accumulated at the τ-th call of SUBROUTINE.

Let ρ_∞ be an upper bound on the exploration bonus across all t calls:

$$\rho_\infty = \beta\sqrt{\kappa} + \Upsilon\sqrt{t} + \sqrt{2\kappa \ln(4\kappa T^4)} \leq \beta\sqrt{\kappa} + \Upsilon\sqrt{t} + 2\sqrt{2\kappa \ln(\kappa T)}. \tag{13}$$

We aim to upper bound:

$$\sum_{\tau \leq t} \min\left\{1, \rho_\infty \sqrt{\phi_{e_\tau}(\hat{p}_\tau)^\top \Lambda_\tau^{-1} \phi_{e_\tau}(\hat{p}_\tau)} + \Upsilon\right\}. \tag{14}$$

By the elliptical potential lemma (see, e.g., Lemma 11 of [1]), we have:

$$\begin{aligned}
&\sum_{\tau \leq t} \min\left\{1, \phi(e_\tau, \hat{p}_\tau)^\top \Lambda_\tau^{-1} \phi(e_\tau, \hat{p}_\tau)\right\} \\
&\leq 2 \ln \det(\Lambda_{t+1}) \leq 2\kappa \ln(\kappa t + 1).
\end{aligned} \tag{15}$$

Hence,

$$\begin{aligned}
&\sum_{\tau\le t}\min\left\{1,\rho_\infty\sqrt{\phi(e_\tau,\hat{p}_\tau)^\top\Lambda_\tau^{-1}\phi(e_\tau,\hat{p}_\tau)}+\Upsilon\right\}\\
&\quad\le \Upsilon t+\rho_\infty\sum_{\tau\le t}\min\left\{1,\sqrt{\phi(e_\tau,\hat{p}_\tau)^\top\Lambda_\tau^{-1}\phi(e_\tau,\hat{p}_\tau)}\right\}\\
&\quad\le \Upsilon t+\rho_\infty\sqrt{t}\cdot\sqrt{\sum_{\tau\le t}\min\left\{1,\phi(e_\tau,\hat{p}_\tau)^\top\Lambda_\tau^{-1}\phi(e_\tau,\hat{p}_\tau)\right\}}\\
&\quad\le \Upsilon t+(\beta\sqrt{\kappa}+\Upsilon\sqrt{t}+2\sqrt{2\kappa\ln(\kappa T)})\cdot\sqrt{t}\cdot\sqrt{2\kappa\ln(\kappa t+1)}\\
&\quad\le 2\Upsilon t\sqrt{2\kappa\ln(\kappa t+1)}+\beta\kappa\sqrt{2t\ln(\kappa t+1)}\\
&\quad+4\sqrt{td\ln(\kappa T)\ln(dt+1)}\\
&\quad\le 2\sqrt{2}\kappa\ln(\kappa T_\varsigma+1)\left(\Upsilon t+\beta\sqrt{t}+\sqrt{2t}\right)
\end{aligned}\tag{16}$$

Dividing both sides by t, we complete the proof of Lemma 2. □

APPENDIX D

Proof of Theorem 3. To bound the cumulative pricing regret, we follow three steps. First, we invoke Lemma 2 along with standard concentration inequalities to obtain high-probability confidence bounds on the average reward estimates within each price interval. Second, we use a UCB-style argument to show that the per-round regret is at most twice the confidence width. Third, summing these bounds over all rounds and leveraging the structure of the confidence terms yields a sublinear bound on the total regret.

For each interval I_j, define $\mu_e^*(I_j)=\max_{p\in I_j}p\chi(e,p)$. Invoking Lemma 2 and standard concentration inequalities, we have that, with probability $1-O(T^{-1})$, it holds uniformly for all j that

$$\mu_e^*(I_j)\le\tau_j/n_j+\mathrm{CI}_j\le\mu_e^*(I_j)+2\mathrm{CI}_j.\tag{17}$$

The rest of this proof will be conditioned on the success event in which Equation (17) holds. Suppose at time t the Algorithm 4 is invoked with $j_t\in[N]$ and that a total number of T_j time periods are allocated to interval I_j throughout the T time periods. Let also $j^*:=\arg\max_{j\in[N]}\mu_e^*(I_j)$ be the interval in which the optimal price resides. Using standard UCB analysis, the cumulative regret of Algorithm 1 is upper bounded by

$$\begin{aligned}
\mu_e^*(I_{j^*})-\mu_e^*(I_{j_t})&\le(\tau_{j^*}/n_{j^*}+\mathrm{CI}_{j^*})-(\tau_{j_t}/n_{j_t}+\mathrm{CI}_{j_t})\\
&\quad+(\tau_{j_t}/n_{j_t}+\mathrm{CI}_{j_t})-\mu_e^*(I_{j_t})\\
&\le(\tau_{j_t}/n_{j_t}+\mathrm{CI}_{j_t})-\mu_e^*(I_{j_t})\\
&\le 2\mathrm{CI}_{j_t},
\end{aligned}\tag{18}$$

where the second inequality holds because j_t maximizes $\tau_j/n_j+\mathrm{CI}_j$ and the third inequality comes from Equation (17).

Furthermore,

$$\begin{aligned}
&\sum_{t=1}^{T}\mathrm{CI}_{j_t}\\
&\le 2\sqrt{2}\kappa\ln(\kappa T+1)\times\left[\Upsilon T+(\beta+\sqrt{2})\sum_{j=1}^{N}\sum_{\ell=1}^{T_j}\frac{1}{\sqrt{\ell}}\right]\\
&\le 2\sqrt{2}\kappa\ln(\kappa T+1)\times\left[\Upsilon T+(\beta+\sqrt{2})\sum_{j=1}^{N}2\sqrt{T_j}\right]\\
&\le 2\sqrt{2}\kappa\ln(\kappa T+1)\times\left[\Upsilon T+2(\beta+\sqrt{2})\sqrt{N}\cdot\sqrt{\sum_{j=1}^{N}T_j}\right]\\
&2\sqrt{2}\kappa\ln(\kappa T+1)\left[\Upsilon T+2(\beta+\sqrt{2})\sqrt{TN}\right]\\
&\le 2\sqrt{2}\kappa\ln(\kappa T+1)\times\left[\beta\cdot T^{\frac{k+1}{2k+1}}+2(\hat{\Upsilon}+\sqrt{2})T^{\frac{k+1}{2k+1}}\right].
\end{aligned}\tag{19}$$

Combining Lemma 2 and Equations (18) and (19), we have that, with probability $1-O(T_\varsigma^{-1})$,

$$\begin{aligned}
&\sum_{t=1}^{T_\varsigma}\left(\max_{p\in\mathcal{P}}\ p\chi(e,p)-p_t\chi(e,p_t)\right)\\
&=\sum_{t=1}^{T_\varsigma}\left(\mu_e^*(I_{j^*})-\mu_e^*(I_{j_t})+\max_{p\in I_{j_t}}p\chi(e,p)-\hat{p}_t f(\hat{p}_t,e)\right)\\
&\le 6\sqrt{2}\kappa\ln(\kappa T_\varsigma+1)\\
&\times\left[\beta\cdot T_\varsigma^{\frac{k+1}{2k+1}}+2(\Upsilon+\sqrt{2})T_\varsigma^{\frac{k+1}{2k+1}}\right].
\end{aligned}\tag{20}$$

Finally, using the Azuma-Hoeffding inequality, we complete the proof of Theorem 3. □

APPENDIX E

Proof of Theorem 4. We analyze the per-round time complexity of VTHB (Algorithm 1), which invokes two sub-learners at each round.

CCB (Algorithm 2) selects the retrieval configuration e_t based on historical reward signals aggregated per $e\in\mathcal{E}$. With incremental maintenance of reward averages and counts, this step takes time $\mathcal{O}(|\mathcal{E}|)$ per round.

CPB (Algorithm 3) scans N price intervals to identify the most promising one based on upper confidence bounds. The UCB estimation over all intervals takes $\mathcal{O}(N)$ time, assuming aggregated statistics are maintained per (ς,e,j). *CPB* then invokes *LAB* (Algorithm 4), which performs a local regression using a κ-dimensional Taylor feature. With incremental updates of the Gram matrix and its inverse—e.g., via rank-one Sherman–Morrison or Cholesky updates—the cost per call to *LAB* is reduced from $\mathcal{O}(\kappa^3)$ to $\mathcal{O}(\kappa^2)$.

Therefore, the total time complexity per round is

$$\mathcal{O}(|\mathcal{E}|+N+\kappa^2).\tag{21}$$

Over T rounds, the total time complexity of *VTHB* becomes

$$\mathcal{O}\left(T(|\mathcal{E}|+N+\kappa^2)\right),\tag{22}$$

as claimed.

□

Appendix F: Extra Lemmas

Lemma 1. *Suppose $\chi^\varsigma(e,p) \in \Sigma^n(\mathcal{E}\times\mathcal{P};\beta)$ and $|\mathcal{D}| = t$ for I_{j_t}. If the output of Algorithm 4 is $\hat{p}$, then with probability $1-\delta$, the following holds:*

$$\max_{p\in I_{j_t}} p\chi^\varsigma(e,p) - \hat{p}\chi^\varsigma(e,\hat{p}) \le 2\bar{p}\min\left\{1, \rho\sqrt{\phi(e,\hat{p})^\top\Lambda^{-1}\phi(e,\hat{p})} + \Upsilon\right\}, \tag{23}$$

where

$$\rho = \beta\sqrt{\kappa} + \Upsilon\sqrt{t} + \sqrt{2\kappa\ln(4\kappa t/\delta)} + 2. \tag{24}$$

Proof of Lemma 3. To bound the pricing regret, we model the observed reward as the true value plus approximation bias and noise. We analyze the ridge regression estimator, bound its deviation from the true coefficients using norm and concentration arguments, and derive a high-probability error bound. This deviation is then translated into a regret bound via an optimistic estimator over the target interval.

Fix a configuration e and consider the interval $I_{j_t} = [a,b]$. Label the (p,y) parameters in $\mathcal{D}$ as $\{(p_i,y_i)\}_{i=1}^t$ in chronological order. Each observation satisfies:

$$y_i = \chi^\varsigma(e,p_i) + \xi_i = P^B(p_i,e) + \xi_i + m_i, \tag{25}$$

where $\mathbb{E}[\xi_i \mid p_1, y_1, \ldots, p_{i-1}, y_{i-1}, p_i] = 0$ and $|m_i| \le \Upsilon$ due to polynomial approximation bias.

Denote $\boldsymbol{y} = (y_i)_{i\le t}$, $\boldsymbol{\xi} = (\xi_i)_{i\le t}$, and $\boldsymbol{m} = (m_i)_{i\le t}$ in $\mathbb{R}^t$. Let $X = (\phi(e_t,p_i))_{i\le t} \in \mathbb{R}^{t\times y}$. The ridge regression estimator is:

$$\hat{\theta}^\varsigma = \Lambda^{-1}X^\top\boldsymbol{y} = (X^\top X + I)^{-1}X^\top\boldsymbol{y}.$$

Also, note:

$$\boldsymbol{y} = X\theta^* + \boldsymbol{\xi} + \boldsymbol{m}, \quad \Rightarrow \quad \hat{\theta}^\varsigma - \theta^* = -\Lambda^{-1}\theta^* + \Lambda^{-1}X^\top(\boldsymbol{\xi}+\boldsymbol{m}).$$

Multiplying both sides by $(\hat{\theta}^\varsigma - \theta^*)^\top\Lambda$:

$$\begin{aligned}&(\hat{\theta}^\varsigma - \theta^*)^\top\Lambda(\hat{\theta}^\varsigma) - \theta^*)\\ &= -\langle\hat{\theta}^\varsigma - \theta^*, \theta^*\rangle + \sum_{\tau\le t}(\xi_\tau + m_\tau)\langle\phi(e_\tau,p_\tau), \hat{\theta}^\varsigma - \theta^*\rangle.\end{aligned} \tag{26}$$

Using Cauchy-Schwarz on the bias term:

$$\begin{aligned}&\left|\sum_{\tau\le t} m_\tau\langle\phi(e_\tau,p_\tau), \hat{\theta}^\varsigma - \theta^*\rangle\right|\\ &\le \sqrt{\sum_{\tau\le t} m_\tau^2}\cdot\sqrt{\sum_{\tau\le t}\left|\langle\phi(e_\tau,p_\tau), \hat{\theta}^\varsigma - \theta^*\rangle\right|^2}\\ &\le \Upsilon\sqrt{t}\cdot\sqrt{(\hat{\theta}^\varsigma - \theta^*)^\top\Lambda(\hat{\theta}^\varsigma - \theta^*)}.\end{aligned} \tag{27}$$

Dividing $\sqrt{(\hat{\theta}^\varsigma - \theta^*)^\top\Lambda(\hat{\theta}^\varsigma - \theta^*)}$ from both sides of (26) and since $\Lambda \succeq I$, we have:

$$\sqrt{(\hat{\theta}^\varsigma - \theta^*)^\top\Lambda(\hat{\theta}^\varsigma - \theta^*)} \le \|\theta^*\|_2 + \Upsilon\sqrt{t} + \sup_{z\in\Phi_\Lambda}|G_t(z)|,$$

where

$$\Phi_\Lambda = \{z\in\mathbb{R}^k : z^\top\Lambda z \le 1\} \text{ and } G_t(z) = \sum_{\tau\le t}\xi_\tau\langle\phi(e_\tau,p_\tau), z\rangle. \tag{28}$$

Because each coefficient of θ^* satisfies $|\theta^*_{k'}| = |\chi^{\varsigma(k')}(a,e)|/(k')! \le C/(k')!$, we conclude:

$$\|\theta^*\|_2 \le \beta\sqrt{\kappa} \tag{29}$$

By Lemma 2 in Appendix F, with probability $1-\delta$:

$$\sqrt{(\hat{\theta}^\varsigma - \theta^*)^\top\Lambda(\hat{\theta}^\varsigma - \theta^*)} \le \beta\sqrt{\kappa} + \Upsilon\sqrt{t} + \sqrt{2\kappa\ln(4\kappa t/\delta)} + 2.$$

Now fix $p \in I$ and define the estimator $\hat{\chi}^\varsigma(e,p) = \langle\phi(e,p), \hat{\theta}^\varsigma\rangle$. Then:

$$\begin{aligned}|\hat{\chi}^\varsigma(e,p) - \chi^\varsigma(e,p)| &\le |\hat{\chi}^\varsigma(e,p) - P_{I,e}(p)| + |P^I(e,p) - \chi^\varsigma(e,p)|\\ &\le |\langle\phi(e,p), \hat{\theta}^\varsigma - \theta^*\rangle| + \Upsilon\\ &\le \sqrt{\phi(e,p)^\top\Lambda^{-1}\phi(e,p))}\cdot\rho + \Upsilon.\end{aligned}$$

Define the optimistic estimate:

$$\bar{\chi}^\varsigma(p,e) := \min\left\{1, \langle\hat{\theta}, \phi(e,p)\rangle + \rho\sqrt{\phi(e,p)^\top\Lambda^{-1}\phi(e,p)} + \Upsilon\right\}.$$

Then, with high probability $(1-\delta)$, $\bar{f}(p,e) \ge \chi(e,p)$ for all $p\in I$, and:

$$\max_{p\in I_{j_t}} p\chi^\varsigma(e,p) - \hat{p}\chi^\varsigma(e,\hat{p}) \le 2\bar{p}\min\left\{1, \rho\sqrt{\phi(e,\hat{p})^\top\Lambda^{-1}\phi(e,\hat{p})} + \Upsilon\right\}, \tag{30}$$

□

Lemma 2. *Fix a configuration e, $\kappa\in\mathbb{N}$, $t\in\mathbb{N}$, and $\delta\in(0,1)$. Let Φ_Λ and G_t be defined in Equation (28) in the context of configuration e, and suppose $\|\phi(p_\tau,e)\|_2 \le \sqrt{\kappa}$ almost surely for all $\tau \le t$. Then, with probability at least $1-\delta$, it holds uniformly for all $\Lambda\in\mathbb{S}_\kappa^{++}$ that*

$$\sup_{z\in\Phi_\Lambda}|G_t(z)| \le \sqrt{2\kappa\ln(4\kappa t/\delta)} + 2. \tag{31}$$

Proof of Lemma 3. To bound the supremum of $G_t(z)$ over Φ_Λ, we construct an ϵ-cover of the unit ball, which also covers Φ_Λ due to $\Lambda\succeq I$. We apply Hoeffding's inequality on the cover and use a union bound to obtain a uniform high-probability bound, then extend it to the full set via continuity.

Let $\epsilon > 0$ be a small parameter to be specified later. Define the Λ-norm as

$$\|x\|_\Lambda := \sqrt{x^\top\Lambda x}, \tag{32}$$

and let $\mathbb{B}(1,\|\cdot\|)$ be the unit ball in $\mathbb{R}^d$ under norm $\|\cdot\|$.

Let $\mathcal{H} \subseteq \mathbb{B}(1, \|\cdot\|_2)$ be an ϵ-covering of the Euclidean unit ball such that

$$\sup_{z \in \mathbb{B}(1,\|\cdot\|_2)} \min_{z' \in \mathcal{H}} \|z - z'\|_2 \leq \epsilon. \tag{33}$$

It is known that such a covering exists with logarithmic cardinality:

$$\ln |\mathcal{H}| \leq k \ln(2/\epsilon). \tag{34}$$

Because $\Lambda \succeq I$, we have $\Phi_\Lambda \subseteq \mathbb{B}(1, \|\cdot\|_2)$, so the same $\mathcal{H}$ is also an ϵ-cover of Φ_Λ under $\|\cdot\|_2$.

Fix arbitrary $z \in \mathcal{H}$. Since $|\xi_\tau| \leq 1$ almost surely, we apply Hoeffding's inequality to get:

$$\Pr\left(|G_t(z)| \leq \sqrt{2\ln(2/\delta')} \cdot \|z\|_\Lambda\right) \geq 1 - \delta'. \tag{35}$$

Applying a union bound over all $z \in \mathcal{H}$ with total failure probability δ, we get with probability at least $1 - \delta$:

$$\sup_{z \in \mathcal{H} \cap \Phi_\Lambda} |G_t(z)| \leq \sqrt{2\ln(2|\mathcal{H}|/\delta)} \leq \sqrt{2d\ln(4/\epsilon)} + 2\ln(1/\delta). \tag{36}$$

Moreover, since $\|\phi(p_\tau, e_\tau)\|_2 \leq \sqrt{d}$ almost surely for all τ, and the regularization implies $\|\Lambda\|_{\text{op}} \leq 1 + \kappa t \leq 2\kappa t$, we get:

$$\sup_{z \in \Phi_\Lambda} |G_t(z)| \leq \sup_{z \in \mathcal{H} \cap \Phi_\Lambda} |G_t(z)| + 2dt\epsilon. \tag{37}$$

Substituting $\epsilon = 1/(\kappa t)$ into Equation (37), we obtain the desired result:

$$\sup_{z \in \Phi_\Lambda} |G_t(z)| \leq \sqrt{2d\ln(4\kappa t/\delta)} + 2. \tag{38}$$

□

REFERENCES

[1] Y. Abbasi-Yadkori, D. Pal, and C. Szepesvari, "Online-to-confidence-set conversions and application to sparse stochastic bandits," in *Artificial Intelligence and Statistics*, pp. 1–9, PMLR, 2012.