Target Tracking for Contextual Bandits: Application to Demand Side Management

Supplementary material

Margaux Brégère **Pierre Gaillard** **Yannig Goude** **Gilles Stoltz**

We provide the proofs in order of appearance of the corresponding result:

– The proof of Lemma 1 in Appendix A
– The proof of Proposition 1 in Appendix B
– The proof of Lemma 2 in Appendix C
– The proof of Lemma 3 in Appendix D
– The proof of Theorem 2 in Appendix E

We also give more details on the numerical expression of the covariance matrix Γ built in the experiments (see Section 5.1) based on real data:

– Details on the covariance matrix Γ in Appendix F.

A. Proof of Lemma 1

The proof below relies on Laplace's method on supermartingales, which is a standard argument to provide confidence bounds on a self-normalized sum of conditionally centered random vectors. See Theorem 2 of Abbasi-Yadkori et al. (2011) or Theorem 20.2 in the monograph by Lattimore & Szepesvári (2018). Under Model 1 and given the definition of V_t, we have the rewriting

$$\begin{aligned}
\widehat{\theta}_t &= V_t^{-1} \sum_{s=1}^{t} \phi(x_s, p_s) Y_{s,p_s} \\
&= V_t^{-1} \sum_{s=1}^{t} \phi(x_s, p_s) \big(\phi(x_s, p_s)^{\mathrm{T}} \theta + p_s^{\mathrm{T}} \varepsilon_s \big) \\
&= V_t^{-1} \big((V_t - \lambda \mathrm{I}_d)\theta + M_t \big) = \theta - \lambda V_t^{-1} \theta + V_t^{-1} M_t \,,
\end{aligned}$$

where we introduced

$$M_t = \sum_{s=1}^{t} \phi(x_s, p_s) p_s^{\mathrm{T}} \varepsilon_s \,,$$

which is a martingale with respect to $\mathcal{F}_t = \sigma(\varepsilon_1, \ldots, \varepsilon_t)$. Therefore, by a triangle inequality,

$$\begin{aligned}
\big\| V_t^{1/2} \big(\widehat{\theta}_t - \theta \big) \big\| &= \big\| -\lambda V_t^{-1/2} \theta + V_t^{-1/2} M_t \big\| \\
&\leqslant \lambda \big\| V_t^{-1/2} \theta \big\| + \big\| V_t^{-1/2} M_t \big\| \,.
\end{aligned}$$

On the one hand, given that all eigenvalues of the symmetric matrix V_t are larger than λ (given the $\lambda \mathrm{I}_d$ term in its definition), all eigenvalues of $V_t^{-1/2}$ are smaller than $1/\sqrt{\lambda}$ and thus,

$$\lambda \big\| V_t^{-1/2} \theta \big\| \leqslant \lambda \frac{1}{\sqrt{\lambda}} \|\theta\| = \sqrt{\lambda} \|\theta\| \,.$$

We now prove, on the other hand, that with probability at least $1 - \delta$,

$$\big\| V_t^{-1/2} M_t \big\| \leqslant \rho \sqrt{2 \ln \frac{1}{\delta} + d \ln \frac{1}{\lambda} + \ln \det(V_t)} \,,$$

which will conclude the proof of the lemma.

Step 1: Introducing super-martingales. For all $\nu \in \mathbb{R}^d$, we consider

$$S_{t,\nu} = \exp\left(\nu^{\mathrm{T}} M_t - \frac{\rho^2}{2}\nu^{\mathrm{T}} V_t \nu\right)$$

and now show that it is an $\mathcal{F}_t$–super-martingale. First, note that since the common distribution of the $\varepsilon_1, \varepsilon_2, \ldots$ is ρ–sub-Gaussian, then for all $\mathcal{F}_{t-1}$–measurable random vectors ν_{t-1},

$$\mathbb{E}\left[\mathrm{e}^{\nu_{t-1}^{\mathrm{T}}\varepsilon_t} \,\middle|\, \mathcal{F}_{t-1}\right] \leqslant \mathrm{e}^{\rho^2\|\nu_{t-1}\|^2/2}\,. \tag{14}$$

Now,

$$S_{t,\nu} = S_{t-1,\nu}\, \exp\bigg(\nu^{\mathrm{T}}\phi(x_t,p_t)p_t^{\mathrm{T}}\varepsilon_t - \frac{\rho^2}{2}\nu^{\mathrm{T}}\phi(x_t,p_t)\phi(x_t,p_t)^{\mathrm{T}}\nu\bigg)$$

where, by using the sub-Gaussian assumption (14) and the fact that $\sum_j p_{j,t}^2 \leqslant 1$ for all convex weight vectors p_t,

$$\mathbb{E}\left[\exp\bigl(\nu^{\mathrm{T}}\phi(x_t,p_t)p_t^{\mathrm{T}}\varepsilon_t\bigr) \,\middle|\, \mathcal{F}_{t-1}\right] \leqslant \exp\left(\frac{\rho^2}{2}\nu^{\mathrm{T}}\phi(x_t,p_t)\underbrace{p_t^{\mathrm{T}}p_t}_{\leqslant 1}\phi(x_t,p_t)^{\mathrm{T}}\nu\right).$$

This implies $\mathbb{E}\bigl[S_{t,\nu}\big|\mathcal{F}_{t-1}\bigr] \leqslant S_{t-1,\nu}$.

Note that the rewriting of $S_{t,\nu}$ in its vertex form is, with $m = V_t^{-1}M_t/\rho^2$:

$$\begin{aligned} S_{t,\nu} &= \exp\left(\frac{1}{2}(\nu-m)^{\mathrm{T}}\rho^2 V_t\,(\nu-m) + \frac{1}{2}m^{\mathrm{T}}\rho^2 V_t\, m\right) \\ &= \exp\left(\frac{1}{2}(\nu-m)^{\mathrm{T}}\rho^2 V_t\,(\nu-m)\right) \\ &\qquad \times \exp\left(\frac{1}{2\rho^2}\bigl\|V_t^{-1/2}M_t\bigr\|^2\right). \end{aligned}$$

Step 2: Laplace's method—integrating $S_{t,\nu}$ over $\nu \in \mathbb{R}^d$. The basic observation behind this method is that (given the vertex form) $S_{t,\nu}$ is maximal at $\nu = m = V_t^{-1}M_t/\rho^2$ and then equals $\exp\bigl(\bigl\|V_t^{-1/2}M_t\bigr\|^2/(2\rho^2)\bigr)$, which is (a transformation of) the quantity to control. Now, because the exp function quickly vanishes, the integral over $\nu \in \mathbb{R}^d$ is close to this maximum. We therefore consider

$$\overline{S}_t = \int_{\mathbb{R}^d} S_{t,\nu}\,\mathrm{d}\nu\,.$$

We will make repeated uses of the fact that the Gaussian density functions,

$$\nu \longmapsto \frac{1}{\sqrt{\det(2\pi C)}}\,\exp\Bigl((\nu-m)^{\mathrm{T}}C^{-1}(\nu-m)\Bigr),$$

where $m \in \mathbb{R}^d$ and C is a (symmetric) positive-definite matrix, integrate to 1 over $\mathbb{R}^d$. This gives us first the rewriting

$$\overline{S}_t = \sqrt{\det\bigl(2\pi\rho^{-2}V_t^{-1}\bigr)}\,\exp\left(\frac{1}{2\rho^2}\bigl\|V_t^{-1/2}M_t\bigr\|^2\right).$$

Second, by the Fubini-Tonelli theorem and the super-martingale property

$$\mathbb{E}\bigl[S_{t,\nu}\bigr] \leqslant \mathbb{E}\bigl[S_{0,\nu}\bigr] = \exp\bigl(-\lambda\rho^2\|\nu\|^2/2\bigr)\,,$$

we also have

$$\begin{aligned} \mathbb{E}\bigl[\overline{S}_t\bigr] &\leqslant \int_{\mathbb{R}^d} \exp\bigl(-\lambda\rho^2\|\nu\|^2/2\bigr)\,\mathrm{d}\nu \\ &= \sqrt{\det\bigl(2\pi\rho^{-2}\lambda^{-1}\mathrm{I}_d\bigr)}\,. \end{aligned}$$

Combining the two statements, we proved

$$\mathbb{E}\left[\exp\left(\frac{1}{2\rho^2}\bigl\|V_t^{-1/2}M_t\bigr\|^2\right)\right] \leqslant \sqrt{\frac{\det(V_t)}{\lambda^d}}\,.$$

Step 3: Markov-Chernov bound. For $u > 0$,

$$\begin{aligned} &\mathbb{P}\Bigl[\bigl\|V_t^{-1/2}M_t\bigr\| > u\Bigr] \\ &= \mathbb{P}\left[\frac{1}{2\rho^2}\bigl\|V_t^{-1/2}M_t\bigr\|^2 > \frac{u^2}{2\rho^2}\right] \\ &\leqslant \exp\left(-\frac{u^2}{2\rho^2}\right)\mathbb{E}\left[\exp\left(\frac{1}{2\rho^2}\bigl\|V_t^{-1/2}M_t\bigr\|^2\right)\right] \\ &\leqslant \exp\left(-\frac{u^2}{2\rho^2} + \frac{1}{2}\ln\frac{\det(V_t)}{\lambda^d}\right) = \delta \end{aligned}$$

for the claimed choice

$$u = \rho\sqrt{2\ln\frac{1}{\delta} + d\ln\frac{1}{\lambda} + \ln\det(V_t)}\,.$$

B. Proof of Proposition 1

Comment: The main difference with the regret analysis of LinUCB provided by Chu et al. (2011) or Lattimore & Szepesvári (2018) is in the first part of *Step 1*, as we need to deal with slightly more complicated quantities: not just with linear quantities of the form $\phi(x_t, p)^{\mathrm{T}}\theta$. Steps 2 and 3 are easy consequences of Step 1.

We show below (*Step 1*) that for all $t \geqslant 2$, if

$$\big\| V_{t-1}^{1/2}\big(\widehat{\theta}_{t-1} - \theta\big) \big\| \leqslant B_{t-1}(\delta t^{-2}) \quad \text{and} \quad \big\| \Gamma - \widehat{\Gamma}_t \big\|_\infty \leqslant \gamma\,, \tag{15}$$

then

$$\forall p \in \mathcal{P}, \qquad \big|\ell_{t,p} - \widehat{\ell}_{t,p}\big| \leqslant \alpha_{t,p}\,. \tag{16}$$

Property (16), for those t for which it is satisfied, entails (*Step 2*) that the corresponding instantaneous regrets are bounded by

$$r_t \stackrel{\text{def}}{=} \ell_{t,p_t} - \min_{p \in \mathcal{P}} \ell_{t,p} \leqslant 2\alpha_{t,p_t}\,.$$

It only remains to deal (*Step 3*) with the rounds t when (16) does not hold; they account for the $1-\delta$ confidence level.

Step 1: Good estimation of the losses. When the two events (15) hold, we have

$$\begin{aligned}
&\big|\ell_{t,p} - \widehat{\ell}_{t,p}\big| \\
&= \bigg| \Big(\phi(x_t, p)^{\mathrm{T}}\theta - c_t\Big)^2 + p^{\mathrm{T}}\Gamma p \\
&\qquad - \Big(\big[\phi(x_t, p)^{\mathrm{T}}\widehat{\theta}_{t-1}\big]_C - c_t \Big)^2 + p^{\mathrm{T}}\widehat{\Gamma}_t p \bigg| \\
&\leqslant \big|p^{\mathrm{T}}\Gamma p - p^{\mathrm{T}}\widehat{\Gamma}_t p\big| \\
&\qquad + \bigg| \Big(\phi(x_t, p)^{\mathrm{T}}\theta - c_t\Big)^2 - \Big(\big[\phi(x_t, p)^{\mathrm{T}}\widehat{\theta}_{t-1}\big]_C - c_t \Big)^2 \bigg|.
\end{aligned}$$

On the one hand, $\big|p^{\mathrm{T}}\Gamma p - p^{\mathrm{T}}\widehat{\Gamma}_t p\big| \leqslant \gamma$ while on the other hand,

$$\begin{aligned}
&\bigg| \Big(\phi(x_t, p)^{\mathrm{T}}\theta - c_t\Big)^2 - \Big(\big[\phi(x_t, p)^{\mathrm{T}}\widehat{\theta}_{t-1}\big]_C - c_t \Big)^2 \bigg| \\
&= \Big| \phi(x_t, p)^{\mathrm{T}}\theta - \big[\phi(x_t, p)^{\mathrm{T}}\widehat{\theta}_{t-1}\big]_C \Big| \\
&\qquad \times \Big| \phi(x_t, p)^{\mathrm{T}}\theta + \big[\phi(x_t, p)^{\mathrm{T}}\widehat{\theta}_{t-1}\big]_C - 2c_t \Big|,
\end{aligned}$$

where by the boundedness assumptions (5), all quantities in the final inequality lie in $[0, C]$, thus

$$\Big| \phi(x_t, p)^{\mathrm{T}}\theta + \big[\phi(x_t, p)^{\mathrm{T}}\widehat{\theta}_{t-1}\big]_C - 2c_t \Big| \leqslant 2C\,.$$

Finally,

$$\begin{aligned}
&\Big| \phi(x_t, p)^{\mathrm{T}}\theta - \big[\phi(x_t, p)^{\mathrm{T}}\widehat{\theta}_{t-1}\big]_C \Big| \\
&\leqslant \big| \phi(x_t, p)^{\mathrm{T}}\theta - \phi(x_t, p)^{\mathrm{T}}\widehat{\theta}_{t-1} \big| \\
&\leqslant \Big\| V_{t-1}^{1/2}\big(\theta - \widehat{\theta}_{t-1}\big) \Big\| \; \big\| V_{t-1}^{-1/2}\phi(x_t, p) \big\|\,,
\end{aligned} \tag{17}$$

where we used the Cauchy-Schwarz inequality for the second inequality, and the fact that $\big|y - [x]_C\big| \leqslant |y - x|$ when $y \in [0, C]$ and $x \in \mathbb{R}$ for the first inequality. Collecting all bounds together, we proved

$$\begin{aligned}
&\bigg| \Big(\phi(x_t, p)^{\mathrm{T}}\theta - c_t\Big)^2 - \Big(\big[\phi(x_t, p)^{\mathrm{T}}\widehat{\theta}_{t-1}\big]_C - c_t \Big)^2 \bigg| \\
&\qquad \leqslant 2C \underbrace{\Big\| V_{t-1}^{1/2}\big(\theta - \widehat{\theta}_{t-1}\big) \Big\|}_{\leqslant B_{t-1}(\delta t^{-2})} \big\| V_{t-1}^{-1/2}\phi(x_t, p) \big\|\,,
\end{aligned}$$

but of course, this term is also bounded by the quantity L introduced in Section 3.5. This concludes the proof of the claimed inequality (16).

Step 2: Resulting bound on the instantaneous regrets. We denote by

$$p_t^\star \in \operatorname*{arg\,min}_{p \in \mathcal{P}} \big\{ \ell_{t,p} + p^{\mathrm{T}}\Gamma p \big\} \tag{18}$$

an optimal convex vector to be used at round t. By definition (3) of the optimistic algorithm, we have that the played p_t satisfies

$$\begin{aligned}
&\widehat{\ell}_{t,p_t} - \alpha_{t,p_t} \leqslant \widehat{\ell}_{t,p_t^\star} - \alpha_{t,p_t^\star}\,, \\
\text{that is,} \qquad &\widehat{\ell}_{t,p_t} - \widehat{\ell}_{t,p_t^\star} \leqslant \alpha_{t,p_t} - \alpha_{t,p_t^\star}\,.
\end{aligned}$$

Now, for those t for which both events (15) hold, the property (16) also holds and yields, respectively for $p = p_t$ and $p = p_t^\star$:

$$\ell_{t,p_t} - \widehat{\ell}_{t,p_t} \leqslant \alpha_{t,p_t} \qquad \text{and} \qquad \widehat{\ell}_{t,p_t^\star} - \ell_{t,p_t^\star} \leqslant \alpha_{t,p_t^\star}\,.$$

Combining all these three inequalities together, we proved

$$\begin{aligned}
r_t &= \ell_{t,p_t} - \ell_{t,p_t^\star} \\
&= \big(\ell_{t,p_t} - \widehat{\ell}_{t,p_t}\big) + \big(\widehat{\ell}_{t,p_t} - \widehat{\ell}_{t,p_t^\star}\big) + \big(\widehat{\ell}_{t,p_t^\star} - \ell_{t,p_t^\star}\big) \\
&\leqslant \alpha_{t,p_t} + \big(\alpha_{t,p_t} - \alpha_{t,p_t^\star}\big) + \alpha_{t,p_t^\star} = 2\alpha_{t,p_t}\,,
\end{aligned}$$

as claimed. This yields the $2\sum \alpha_{t,p_t}$ in the regret bound, where the sum is for $t \geqslant n+1$.

Step 3: Special cases. We conclude the proof by dealing with the time steps $t \geqslant n+1$ when at least one of the events (15) does not hold. By a union bound, this happens for some $t \geqslant n+1$ with probability at most

$$\frac{\delta}{2} + \delta \sum_{t \geqslant n+1} t^{-2} \leqslant \frac{\delta}{2} + \delta \int_2^\infty \frac{1}{t^2}\,\mathrm{d}t = \delta\,,$$

where we used $n \geqslant 2$. These special cases thus account for the claimed $1-\delta$ confidence level.

C. Proof of Lemma 2

We derived the proof scheme below from scratch as we could find no suitable result in the literature for estimating Γ in our context.

We first consider the following auxiliary result.

Lemma 4. *Let $n \geqslant 1$. Assume that the common distribution of the $\varepsilon_1, \varepsilon_2, \dots$ is ρ–sub-Gaussian. Then, no matter how the provider picks the p_t, we have, for all $\delta \in (0,1)$, with probability at least $1-\delta$,*

$$\left\| \sum_{t=1}^{n} p_t p_t^{\mathrm{T}} \big(\widehat{\Gamma}_n - \Gamma\big) p_t p_t^{\mathrm{T}} \right\|_\infty \leqslant \kappa_n \sqrt{n}\,,$$

where the quantities κ_n, M_n and M'_n are defined as in Lemma 2:

$$\begin{aligned} M_n &\stackrel{\text{def}}{=} \rho/2 + \ln(6n/\delta) \\ M'_n &\stackrel{\text{def}}{=} M_n^2 \sqrt{2\ln(3K^2/\delta)} + 2\sqrt{\exp(2\rho)\delta/6} \\ \kappa_n &\stackrel{\text{def}}{=} \big(C + 2M_n\big) B_n(\delta/3) + M'_n \end{aligned}$$

Proof of Lemma 4. We can show that $\widehat{\Gamma}_n$ defined in (4) satisfies

$$\sum_{t=1}^{n} p_t p_t^{\mathrm{T}} \widehat{\Gamma}_n p_t p_t^{\mathrm{T}} = \sum_{t=1}^{n} \widehat{Z}_t^2 p_t p_t^{\mathrm{T}}\,, \tag{19}$$

where we recall that $\widehat{Z}_t \stackrel{\text{def}}{=} Y_{t,p_t} - \big[\phi(x_t,p_t)^{\mathrm{T}}\widehat{\theta}_n\big]_C$. Indeed, with,

$$\Phi\big(\widehat{\Gamma}\big) \stackrel{\text{def}}{=} \sum_{t=1}^{n} \Big(\widehat{Z}_t^2 - p_t^{\mathrm{T}}\widehat{\Gamma}p_t\Big)^2 = \sum_{t=1}^{n} \Big(\widehat{Z}_t^2 - \mathrm{Tr}\big(\widehat{\Gamma}p_t p_t^{\mathrm{T}}\big)\Big)^2,$$

using $\nabla_A \mathrm{Tr}(AB) = B$, we get

$$\nabla_{\widehat{\Gamma}} \Phi\big(\widehat{\Gamma}\big) = \sum_{t=1}^{n} 2 p_t p_t^{\mathrm{T}} \Big(\widehat{Z}_t^2 - p_t^{\mathrm{T}}\widehat{\Gamma}p_t\Big),$$

which leads to (19) by canceling the gradient and keeping in mind that $p_t^{\mathrm{T}}\widehat{\Gamma}p_t$ is a scalar value.

Let us denote

$$Z_t \stackrel{\text{def}}{=} Y_{t,p_t} - \phi(x_t,p_t)^{\mathrm{T}}\theta = p_t^{\mathrm{T}}\varepsilon_t$$

for all $t \geqslant 1$. To prove the lemma, we replace $\widehat{\Gamma}_n$ by using (19) and apply a triangular inequality:

$$\begin{aligned} &\left\| \sum_{t=1}^{n} p_t p_t^{\mathrm{T}} \big(\widehat{\Gamma}_n - \Gamma\big) p_t p_t^{\mathrm{T}} \right\|_\infty \qquad (20) \\ &\leqslant \left\| \sum_{t=1}^{n} (\widehat{Z}_t^2 - Z_t^2) p_t p_t^{\mathrm{T}} \right\|_\infty + \left\| \sum_{t=1}^{n} Z_t^2 p_t p_t^{\mathrm{T}} - p_t p_t^{\mathrm{T}} \Gamma p_t p_t^{\mathrm{T}} \right\|_\infty \end{aligned}$$

We will consecutively provide bounds for each of the two terms in the right-hand side of the above inequality, each holding with probability at least $1-\delta/3$. To do so, we focus on the event defined below where all Z_t are bounded:

$$\mathcal{E}_n(\delta) \stackrel{\text{def}}{=} \big\{\forall t = 1, \dots n, \quad |Z_t| \leqslant M_n\big\}, \tag{21}$$

with M_n defined in the statement of the lemma. We will show below that $\mathcal{E}_n(\delta)$ takes place with probability at least $1-\delta/3$. All in all, our obtained global bound will hold with probability at least $1-\delta$, as stated in the lemma.

Bounding the probability of the event $\mathcal{E}_n(\delta)$. Recall that p_t is $\mathcal{F}_{t-1} = \sigma(\varepsilon_1, \dots, \varepsilon_{t-1})$ measurable. For $t \in \{1, \dots, n\}$, as ε_t is a ρ–sub-Gaussian variable independent of $\mathcal{F}_{t-1}$,

$$\mathbb{E}\Big[\exp(p_t^{\mathrm{T}}\varepsilon_t) \,\Big|\, \mathcal{F}_{t-1}\Big] \leqslant \exp\!\left(\frac{\rho\|p_t\|^2}{2}\right) \leqslant \exp\!\left(\frac{\rho}{2}\right);$$

see Footnote 1 for a reminder of the definition of a ρ–sub-Gaussian variable. Using the Markov-Chernov inequality, we obtain

$$\begin{aligned} \mathbb{P}\big(Z_t \geqslant M_n \,\big|\, \mathcal{F}_{t-1}\big) &\leqslant \mathbb{E}\Big[\exp(Z_t) \,\Big|\, \mathcal{F}_{t-1}\Big] \, \exp(-M_n) \\ &\leqslant \exp\left(\frac{\rho}{2} - M_n\right) = \frac{\delta}{6n}. \end{aligned} \tag{22}$$

Symmetrically, we get that $\mathbb{P}(Z_t \leqslant -M_n) \leqslant \delta/6n$. Combining all these bounds for $t = 1, \dots, n$, the event $\mathcal{E}_n(\delta)$ happens with probability at least $1-\delta/3$.

Upper bound on the first term in (20). By Assumption (5), we have $\phi(x_t,p_t)^{\mathrm{T}}\theta \in [0,C]$, thus

$$|\widehat{Z}_t - Z_t| = \Big|\phi(x_t,p_t)^{\mathrm{T}}\theta - \big[\phi(x_t,p_t)^{\mathrm{T}}\widehat{\theta}_n\big]_C\Big| \leqslant C\,,$$

and therefore, on $\mathcal{E}_n(\delta)$,

$$\big|\widehat{Z}_t + Z_t\big| \leqslant \big|\widehat{Z}_t - Z_t\big| + \big|2Z_t\big| \leqslant C + 2M_n \stackrel{\text{def}}{=} M''_n\,.$$

Noting that all components of $p_t p_t^{\mathrm{T}}$ are upper bounded by 1,

$$\begin{aligned} \left\| \sum_{t=1}^{n} (\widehat{Z}_t^2 - Z_t^2) p_t p_t^{\mathrm{T}} \right\|_\infty &\leqslant \sum_{t=1}^{n} \big|\widehat{Z}_t^2 - Z_t^2\big| \\ &= \sum_{t=1}^{n} \big|(\widehat{Z}_t - Z_t)(\widehat{Z}_t + Z_t)\big| \\ &\leqslant M''_n \sqrt{n \sum_{t=1}^{n} (\widehat{Z}_t - Z_t)^2}\,, \end{aligned}$$

where the last inequality was obtained by $|\widehat{Z}_t + Z_t| \leqslant M''_n$ together with the Cauchy-Schwarz inequality. Using that $\big|y - [x]_C\big| \leqslant |y - x|$ when $y \in [0,C]$ and $x \in \mathbb{R}$, we note that

$$\big|\widehat{Z}_t - Z_t\big| \leqslant \Big|\phi(x_t,p_t)^{\mathrm{T}}(\widehat{\theta}_n - \theta)\Big|\,.$$

All in all, we proved so far

$$\begin{aligned}
&\Big\| \sum_{t=1}^{n} (\widehat{Z}_t^2 - Z_t^2) p_t p_t^{\mathrm{T}} \Big\|_\infty \\
&\leqslant M_n'' \sqrt{n(\widehat{\theta}_n - \theta)^{\mathrm{T}} \left(\sum_{t=1}^{n} \phi(x_t,p_t)\phi(x_t,p_t)^{\mathrm{T}} \right) (\widehat{\theta}_n - \theta)} \\
&= M_n'' \sqrt{n(\widehat{\theta}_n - \theta)^{\mathrm{T}} (V_n - \lambda I) (\widehat{\theta}_n - \theta)} \\
&\leqslant M_n'' \sqrt{n(\widehat{\theta}_n - \theta)^{\mathrm{T}} V_n (\widehat{\theta}_n - \theta)} \\
&= M_n'' \big\| V_n^{1/2} (\theta - \widehat{\theta}_n) \big\| \sqrt{n} \,,
\end{aligned}$$

where $V_n = \lambda I + \sum_{t=1}^{n} \phi(x_t,p_t)\phi(x_t,p_t)^{\mathrm{T}}$ was used for the last steps.

From Lemma 1 and the bound (6), we finally obtain that with probability at least $1-\delta/3$,

$$\Big\| \sum_{t=1}^{n} (\widehat{Z}_t^2 - Z_t^2) p_t p_t^{\mathrm{T}} \Big\|_\infty \leqslant M_n'' \, B_n(\delta/3) \sqrt{n} \tag{23}$$

$$= (C + 2M_n) \, B_n(\delta/3) \sqrt{n} \,. \tag{24}$$

Upper bound on the second term in (20). Recall that p_t is $\mathcal{F}_{t-1}$ measurable and that in Model 1, we defined $Z_t = Y_{t,p_t} - \phi(x_t,p_t)^{\mathrm{T}}\theta = p_t^{\mathrm{T}}\varepsilon_t$, which is a scalar value. These two observations yield

$$\begin{aligned}
\mathbb{E}\big[Z_t^2 p_t p_t^{\mathrm{T}} \,\big|\, \mathcal{F}_{t-1}\big] &= \mathbb{E}\big[p_t Z_t^2 p_t^{\mathrm{T}} \,\big|\, \mathcal{F}_{t-1}\big] \\
&= \mathbb{E}\big[p_t p_t^{\mathrm{T}} \varepsilon_t \varepsilon_t^{\mathrm{T}} p_t p_t^{\mathrm{T}} \,\big|\, \mathcal{F}_{t-1}\big] \\
&= p_t p_t^{\mathrm{T}} \, \mathbb{E}\big[\varepsilon_t \varepsilon_t^{\mathrm{T}} \,\big|\, \mathcal{F}_{t-1}\big] \, p_t p_t^{\mathrm{T}} = p_t p_t^{\mathrm{T}} \Gamma p_t p_t^{\mathrm{T}} \,.
\end{aligned} \tag{25}$$

We wish to apply the Hoeffding–Azuma inequality to each component of $Z_t^2 p_t p_t^{\mathrm{T}}$, however, we need some boundedness to do so. Therefore, we consider instead $Z_t^2 \mathbf{1}_{\{|Z_t| \leqslant M_n\}}$. The indicated inequality, together with a union bound, entails that with probability at least $1-\delta/3$,

$$\begin{aligned}
&\Big\| \sum_{t=1}^{n} Z_t^2 \mathbf{1}_{\{|Z_t| \leqslant M_n\}} p_t p_t^{\mathrm{T}} \\
&\qquad - \sum_{t=1}^{n} \mathbb{E}\Big[Z_t^2 \mathbf{1}_{\{|Z_t| \leqslant M_n\}} p_t p_t^{\mathrm{T}} \,\Big|\, \mathcal{F}_{t-1} \Big] \Big\|_\infty \\
&\leqslant M_n^2 \sqrt{2n \ln(3K^2/\delta)} \,.
\end{aligned} \tag{26}$$

Over $\mathcal{E}_n(\delta)$, using (25) and applying a triangular inequality, we obtain

$$\begin{aligned}
&\Big\| \sum_{t=1}^{n} Z_t^2 p_t p_t^{\mathrm{T}} - p_t p_t^{\mathrm{T}} \Gamma p_t p_t^{\mathrm{T}} \Big\|_\infty \\
&= \Big\| \sum_{t=1}^{n} Z_t^2 \mathbf{1}_{\{|Z_t| \leqslant M_n\}} p_t p_t^{\mathrm{T}} - \sum_{t=1}^{n} \mathbb{E}\big[Z_t^2 p_t p_t^{\mathrm{T}} \,\big|\, \mathcal{F}_{t-1} \big] \Big\|_\infty \\
&\leqslant \Big\| \sum_{t=1}^{n} Z_t^2 \mathbf{1}_{\{|Z_t| \leqslant M_n\}} p_t p_t^{\mathrm{T}} \\
&\qquad - \sum_{t=1}^{n} \mathbb{E}\big[Z_t^2 p_t p_t^{\mathrm{T}} \mathbf{1}_{\{|Z_t| \leqslant M_n\}} \,\big|\, \mathcal{F}_{t-1} \big] \Big\|_\infty \\
&\qquad + \sum_{t=1}^{n} \Big\| \mathbb{E}\big[Z_t^2 p_t p_t^{\mathrm{T}} \mathbf{1}_{\{|Z_t| > M_n\}} \,\big|\, \mathcal{F}_{t-1} \big] \Big\|_\infty \,.
\end{aligned} \tag{27}$$

We just need to bound the last term of the inequality above to conclude this part. Using that $x^2 \leqslant \exp(x)$ for $x \geqslant 0$, we get

$$\begin{aligned}
&\mathbb{E}\Big[Z_t^2 \mathbf{1}_{\{|Z_t| > M_n\}} \,\Big|\, \mathcal{F}_{t-1} \Big] \\
&\qquad \leqslant \mathbb{E}\Big[\exp\big(|Z_t|\big) \mathbf{1}_{\{|Z_t| > M_n\}} \,\Big|\, \mathcal{F}_{t-1} \Big] \,.
\end{aligned}$$

Applying a conditional Cauchy-Schwarz inequality yields

$$\begin{aligned}
&\mathbb{E}\Big[\exp\big(|Z_t|\big) \mathbf{1}_{\{|Z_t| > M_n\}} \,\Big|\, \mathcal{F}_{t-1} \Big] \\
&\leqslant \sqrt{\mathbb{E}\big[\exp\big(2|Z_t|\big) \,\big|\, \mathcal{F}_{t-1} \big] \; \mathbb{E}\big[\mathbf{1}_{\{|Z_t| > M_n\}} \,\big|\, \mathcal{F}_{t-1} \big]} \,.
\end{aligned}$$

Now, thanks to the sub-Gaussian property of ε_t used with $\nu = 2p_t$ and $\nu = -2p_t$, we have

$$\begin{aligned}
&\mathbb{E}\big[\exp\big(2|Z_t|\big) \\
&\leqslant \mathbb{E}\big[\exp(2Z_t) \,\big|\, \mathcal{F}_{t-1} \big] + \mathbb{E}\big[\exp(-2Z_t) \,\big|\, \mathcal{F}_{t-1} \big] \\
&\leqslant 2\exp(2\rho) \,.
\end{aligned}$$

The bound (22) and its symmetric version indicate that

$$\mathbb{P}\big(|Z_t| \geqslant M_n \,\big|\, \mathcal{F}_{t-1} \big) \leqslant \frac{\delta}{3n} \,.$$

We therefore proved

$$\mathbb{E}\Big[\exp\big(|Z_t|\big) \mathbf{1}_{\{|Z_t| > M_n\}} \,\Big|\, \mathcal{F}_{t-1} \Big] \leqslant \sqrt{2\exp(2\rho)\,\frac{\delta}{3n}} \,.$$

Thus, we have $\mathbb{E}\big[Z_t^2 \mathbf{1}_{\{|Z_t| > M_n\}} \,\big|\, \mathcal{F}_{t-1} \big] \leqslant 2\sqrt{\exp(2\rho)\delta/(6n)}$ and as all components of the $p_t p_t^{\mathrm{T}}$ are in $[0,1]$,

$$\Big\| \mathbb{E}\big[Z_t^2 \mathbf{1}_{\{|Z_t| > M_n\}} p_t p_t^{\mathrm{T}} \,\big|\, \mathcal{F}_{t-1} \big] \Big\|_\infty \leqslant 2\sqrt{\exp(2\rho)\frac{\delta}{6n}} . \tag{28}$$

Finally , combining (27) with (26) and (28), we get with probability $1-\delta/3$

$$\begin{aligned}
&\Big\| \sum_{t=1}^{n} Z_t^2 p_t p_t^{\mathrm{T}} - p_t p_t^{\mathrm{T}} \Gamma p_t p_t^{\mathrm{T}} \Big\|_\infty \\
&\leqslant M_n^2 \sqrt{2n \ln(3K^2/\delta)} + 2n\sqrt{\exp(2\rho)\delta/(6n)} = M_n' \sqrt{n} \,,
\end{aligned}$$

where M'_n is defined in the statement of the lemma.

Combining the two upper bounds into (20). Combining the above upper bound with (20) and (24), we proved that with probability $1-\delta$,

$$\left\| \sum_{t=1}^{n} p_t p_t^{\mathrm{T}} \Big(\widehat{\Gamma}_n - \Gamma \Big) p_t p_t^{\mathrm{T}} \right\|_\infty \leqslant M'_n \sqrt{n} + M''_n B_n(\delta/3) \sqrt{n} \,,$$

which concludes the proof. □

Conclusion of the proof of Lemma 2

Remember from Section 3.3 that all vectors $p^{(i,j)}$ are played at least n_0 times in the n exploration rounds.

Proof of Lemma 2. Applying Lemma 4 together with

$$\begin{aligned} p_t p_t^{\mathrm{T}} \big(\widehat{\Gamma}_n - \Gamma\big) p_t p_t^{\mathrm{T}} &= p_t \mathrm{Tr}\Big(p_t^{\mathrm{T}} \big(\widehat{\Gamma}_n - \Gamma\big) p_t \Big) p_t^{\mathrm{T}} \\ &= \mathrm{Tr}\Big(\big(\widehat{\Gamma}_n - \Gamma\big) p_t p_t^{\mathrm{T}} \Big) p_t p_t^{\mathrm{T}} \end{aligned} \tag{29}$$

we have, with probability at least $1-\delta$, that for all pairs of coordinates $(i,j) \in E$,

$$\left| \sum_{t=1}^{n} \mathrm{Tr}\Big(\big(\widehat{\Gamma}_n - \Gamma\big) p_t p_t^{\mathrm{T}} \Big) \big[p_t p_t^{\mathrm{T}} \big]_{i,j} \right| \leqslant \kappa_n \sqrt{n} \,. \tag{30}$$

Remember that in the set E considered in Section 3.3, we only have pairs (i,j) with $i \leqslant j$. However, for symmetry reasons, it will be convenient to also consider the vectors $p^{(i,j)}$ with $i > j$, where the latter vectors are defined in an obvious way. We note that for all $1 \leqslant i,j \leqslant K$,

$$p^{(i,j)} {p^{(i,j)}}^{\mathrm{T}} = p^{(j,i)} {p^{(j,i)}}^{\mathrm{T}} \,. \tag{31}$$

Now, our aim is to control

$$\left| q^{\mathrm{T}} \big(\widehat{\Gamma}_n - \Gamma\big) q \right| = \left| \mathrm{Tr}\Big(\big(\widehat{\Gamma}_n - \Gamma\big) q q^{\mathrm{T}} \Big) \right| \tag{32}$$

uniformly over $q \in \mathcal{P}$. The proof consists of two steps: establishing such a control for the special cases where q is one of the $p^{(i,j)}$ and then, extending the control to arbitrary vectors $q \in \mathcal{P}$, based on a decomposition of qq^{T} as a weighted sum of $p^{(i,j)} {p^{(i,j)}}^{\mathrm{T}}$ vectors.

Part 1: The case of the $p^{(i,j)}$ vectors. Consider first the off-diagonal elements $1 \leqslant i < j \leqslant K$. Note that since p_t is of the form $p^{(i',j')}$ for all $1 \leqslant t \leqslant n$, we have

$$\big[p_t p_t^{\mathrm{T}} \big]_{i,j} = \begin{cases} 1/4 & \text{if } p_t = p^{(i,j)}, \\ 0 & \text{otherwise.} \end{cases} \tag{33}$$

Using that $p_t = p^{(i,j)}$ at least for n_0 rounds, Inequality (30) entails

$$\frac{n_0}{4} \left| \mathrm{Tr}\Big(\big(\widehat{\Gamma}_n - \Gamma\big) p^{(i,j)} {p^{(i,j)}}^{\mathrm{T}} \Big) \right| \leqslant \kappa_n \sqrt{n} \,,$$

or put differently,

$$\left| \mathrm{Tr}\Big(\big(\widehat{\Gamma}_n - \Gamma\big) p^{(i,j)} {p^{(i,j)}}^{\mathrm{T}} \Big) \right| \leqslant \frac{4 \kappa_n \sqrt{n}}{n_0} \,. \tag{34}$$

Now, let us consider the diagonal elements. Let $1 \leqslant i \leqslant K$. We have

$$\big[p_t p_t^{\mathrm{T}} \big]_{i,i} = \begin{cases} 1 & \text{if } p_t = p^{(i,i)}, \\ 1/4 & \text{if } p_t = p^{(i,j)} \text{ for some } j > i, \\ 1/4 & \text{if } p_t = p^{(k,i)} \text{ for some } k < i, \\ 0 & \text{otherwise,} \end{cases} \tag{35}$$

where we recall that the p_t are necessarily of the form $p^{(k,\ell)}$ with $k \leqslant \ell$. Therefore, Inequality (30) yields

$$\begin{aligned} n_0 \Bigg| \mathrm{Tr}\Bigg(\big(\widehat{\Gamma}_n - \Gamma\big) \bigg(& p^{(i,i)} {p^{(i,i)}}^{\mathrm{T}} + \frac{1}{4} \sum_{j>i} p^{(i,j)} {p^{(i,j)}}^{\mathrm{T}} \\ & + \frac{1}{4} \sum_{k<i} p^{(k,i)} {p^{(k,i)}}^{\mathrm{T}} \bigg) \Bigg) \Bigg| \leqslant \kappa_n \sqrt{n} \,, \end{aligned}$$

which we rewrite by symmetry—see (31)—as

$$\left| \mathrm{Tr}\bigg(\big(\widehat{\Gamma}_n - \Gamma\big) \Big(p^{(i,i)} {p^{(i,i)}}^{\mathrm{T}} + \frac{1}{4} \sum_{j \neq i} p^{(i,j)} {p^{(i,j)}}^{\mathrm{T}} \Big) \bigg) \right| \leqslant \frac{\kappa_n \sqrt{n}}{n_0} \,. \tag{36}$$

Part 2-1: Decomposing arbitrary vectors $q \in \mathcal{P}$. Now, let $q \in \mathcal{P}$. We show below by means of elementary calculations that

$$q q^{\mathrm{T}} = \sum_{i=1}^{K} \sum_{j=1}^{K} u(i,j) \; p^{(i,j)} {p^{(i,j)}}^{\mathrm{T}} \tag{37}$$

with $u(i,j) = 2 q_i q_j$ if $i \neq j$ and $u(i,i) = 2q_i^2 - q_i$.

Indeed, by identification and by imposing $u(i,j) = u(j,i)$ for all pairs i,j, the equalities (33) and the symmetry property (31) entail, for $k \neq k'$:

$$\begin{aligned} q_k q_{k'} = \big[q q^{\mathrm{T}} \big]_{k,k'} &= \sum_{i=1}^{K} \sum_{j=1}^{K} u(i,j) \; \big[p^{(i,j)} {p^{(i,j)}}^{\mathrm{T}} \big]_{k,k'} \\ &= \frac{u(k,k')}{4} + \frac{u(k',k)}{4} = \frac{u(k,k')}{2} \,, \end{aligned}$$

which can be rephrased as $u(k,k') = u(k',k) = 2 q_k q_{k'}$. Now, let us calculate the diagonal elements, by identification and by the equalities (35) as well as by the symmetry property (31):

$$
\begin{aligned}
q_k^2 = \left[qq^{\mathrm{T}}\right]_{k,k} &= \sum_{i=1}^{K}\sum_{j=1}^{K} u(i,j)\,\left[p^{(i,j)}p^{(i,j)\mathrm{T}}\right]_{k,k} \\
&= u(k,k) + \sum_{i\neq k}\frac{u(i,k)}{4} + \sum_{j\neq k}\frac{u(k,j)}{4} \\
&= u(k,k) + \frac{1}{2}\sum_{i\neq k} u(i,k) = u(k,k) + \sum_{i\neq k} q_k q_i \\
&= u(k,k) + \sum_{i=1}^{K} q_k q_i - q_k^2 = u(k,k) + q_k - q_k^2\,,
\end{aligned}
$$

which leads to $u(k,k) = 2q_k^2 - q_k$.

We introduce the notation

$$
P^{(i,j)} = p^{(i,j)}p^{(i,j)\mathrm{T}}
$$

and in light of (34) and (36), we rewrite (37) as

$$
\begin{aligned}
qq^{\mathrm{T}} = \sum_{i=1}^{K} u(i,i)&\left(P^{(i,i)} + \frac{1}{4}\sum_{j\neq i} P^{(i,j)}\right) \\
&+ \sum_{i=1}^{K}\sum_{j\neq i}\left(u(i,j) - \frac{u(i,i)}{4}\right)P^{(i,j)}\,.
\end{aligned}
$$

Part 2-2: Controlling arbitrary vectors $q \in \mathcal{P}$. Therefore, substituting this decomposition of qq^{T} into the aim (32), and using the linearity of the trace as well as the triangle inequality for absolute values, we obtain

$$
\begin{aligned}
&\left|q^{\mathrm{T}}\big(\widehat{\Gamma}_n - \Gamma\big)q\right| = \left|\mathrm{Tr}\Big(\big(\widehat{\Gamma}_n - \Gamma\big)qq^{\mathrm{T}}\Big)\right| \\
&\leqslant \sum_{i=1}^{K}\big|u(i,i)\big|\left|\mathrm{Tr}\left(\big(\widehat{\Gamma}_n - \Gamma\big)\Big(P^{(i,i)} + \frac{1}{4}\sum_{j\neq i}P^{(i,j)}\Big)\right)\right| \\
&\quad + \sum_{i=1}^{K}\sum_{j\neq i}\left|u(i,j) - \frac{u(i,i)}{4}\right|\left|\mathrm{Tr}\Big(\big(\widehat{\Gamma}_n - \Gamma\big)P^{(i,j)}\Big)\right|
\end{aligned}
$$

We then substitute the upper bounds (34) and (36) and get

$$
\begin{aligned}
&\left|q^{\mathrm{T}}\big(\widehat{\Gamma}_n - \Gamma\big)q\right| \\
&\leqslant \frac{\kappa_n\sqrt{n}}{n_0}\left(\sum_{i=1}^{K}\big|u(i,i)\big| + 4\sum_{i=1}^{K}\sum_{j\neq i}\left|u(i,j) - \frac{u(i,i)}{4}\right|\right).
\end{aligned}
$$

By the triangle inequality, by the values $2q_iq_j$ of the coefficients $u(i,j)$ when $i \neq j$ and by using $|u(i,i)| \leqslant q_i$,

$$
\begin{aligned}
&\sum_{i=1}^{K}\big|u(i,i)\big| + 4\sum_{i=1}^{K}\sum_{j\neq i}\left|u(i,j) - \frac{u(i,i)}{4}\right| \\
&\leqslant K\sum_{i=1}^{K}\big|u(i,i)\big| + 4\sum_{i=1}^{K}\sum_{j\neq i}\big|u(i,j)\big| \\
&\leqslant K\sum_{i=1}^{K} q_i + 8\sum_{i=1}^{K}\sum_{j\neq i} q_i q_j \\
&= K + 8\sum_{i=1}^{K} q_i(1-q_i) \leqslant K + 8\,.
\end{aligned}
$$

Putting all elements together, we proved

$$
\sup_{q\in\mathcal{P}}\left|q^{\mathrm{T}}\big(\widehat{\Gamma}_n - \Gamma\big)q\right| \leqslant \frac{\kappa_n\sqrt{n}}{n_0}(K+8)\,,
$$

which concludes the proof of Lemma 2. □

D. Proof of Lemma 3

We recall that this lemma is a straightforward adaptation/generalization of Lemma 19.1 of the monograph by Lattimore & Szepesvári (2018); see also a similar result in Lemma 3 by Chu et al. (2011).

We consider the worst case when all summations would start at $n+1=2$.

By definition, the quantity $\overline{B}$ upper bounds all the $B_{t-1}(\delta t^{-2})$. It therefore suffices to upper bound

$$\begin{aligned}
&\sum_{t=2}^{T} \min\Big\{L,\ 2C\overline{B}\,\big\|\,V_{t-1}^{-1/2}\phi(x_t,p_t)\,\big\|\Big\} \\
&\leqslant \sqrt{T}\sqrt{\sum_{t=2}^{T}\min\Big\{L^2,\ \big(2C\overline{B}\big)^2\,\big\|\,V_{t-1}^{-1/2}\phi(x_t,p_t)\,\big\|^2\Big\}} \\
&= \sqrt{T}\sqrt{\sum_{t=2}^{T}\min\bigg\{L^2,\ \big(2C\overline{B}\big)^2\bigg(\frac{\det(V_t)}{\det(V_{t-1})}-1\bigg)\bigg\}}
\end{aligned}$$

where we applied first the Cauchy-Schwarz inequality and used second the equality

$$\begin{aligned}
&1+\big\|\,V_{t-1}^{-1/2}\phi(x_t,p_t)\,\big\|^2 \\
&\qquad = 1+\phi(x_t,p_t)^{\mathrm{T}}V_{t-1}^{-1}\phi(x_t,p_t) = \frac{\det(V_t)}{\det(V_{t-1})}\,,
\end{aligned}$$

that follows from a standard result in online matrix theory, namely, Lemma 5 below.

Now, we get a telescoping sum with the logarithm function by using the inequality

$$\forall b>0, \quad \forall u>0, \qquad \min\{b,u\} \leqslant b\,\frac{\ln(1+u)}{\ln(1+b)}\,, \tag{38}$$

which is proved below. Namely, we further bound the sum above by

$$\begin{aligned}
&\sum_{t=2}^{T}\min\bigg\{L^2,\ \big(2C\overline{B}\big)^2\bigg(\frac{\det(V_t)}{\det(V_{t-1})}-1\bigg)\bigg\} \\
&\leqslant \big(2C\overline{B}\big)^2\sum_{t=2}^{T}\min\bigg\{\frac{L^2}{\big(2C\overline{B}\big)^2},\ \frac{\det(V_t)}{\det(V_{t-1})}-1\bigg\} \\
&\leqslant \big(2C\overline{B}\big)^2\sum_{t=2}^{T}\frac{L^2/\big(2C\overline{B}\big)^2}{\ln\Big(1+L^2/\big(2C\overline{B}\big)^2\Big)}\ln\bigg(\frac{\det(V_t)}{\det(V_{t-1})}\bigg) \\
&= \frac{L^2}{\ln\Big(1+L^2/\big(2C\overline{B}\big)^2\Big)}\ln\bigg(\frac{\det(V_T)}{\det(V_2)}\bigg) \\
&\leqslant \frac{L^2}{\ln\Big(1+L^2/\big(2C\overline{B}\big)^2\Big)}\,d\ln\frac{\lambda+T}{\lambda}
\end{aligned}$$

where we used (5) and one of its consequences to get the last inequality.

Finally, we use $1/\ln(1+u) \leqslant 1/u+1/2$ for all $u \geqslant 0$ to get a more readable constant:

$$\frac{L^2}{\ln\Big(1+L^2/\big(2C\overline{B}\big)^2\Big)} \leqslant \big(2C\overline{B}\big)^2+\frac{L^2}{2}\,.$$

The proof is concluded by collecting all pieces. □

Finally, we now provide the proofs of two either straightforward or standard results used above.

D.1. A Standard Result in Online Matrix Theory

The following result is extremely standard in online matrix theory (see, among many others, Lemma 11.11 in Cesa-Bianchi & Lugosi, 2006 or the proof of Lemma 19.1 in the monograph by Lattimore & Szepesvári, 2018).

Lemma 5. *Let M a $d \times d$ full-rank matrix, let $u,\ v \in \mathbb{R}^d$ be two arbitrary vectors. Then*

$$1+v^{\mathrm{T}}M^{-1}u = \frac{\det\big(M+uv^{\mathrm{T}}\big)}{\det(M)}\,.$$

The proof first considers the case $M=\mathrm{I}_d$. We are then left with showing that $\det\big(\mathrm{I}_d+uv^{\mathrm{T}}\big)=1+v^{\mathrm{T}}u$, which follows from taking the determinant of every term of the equality

$$\begin{aligned}
&\begin{bmatrix} \mathrm{I}_d & 0 \\ v^{\mathrm{T}} & 1 \end{bmatrix}\begin{bmatrix} \mathrm{I}_d+uv^{\mathrm{T}} & u \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \mathrm{I}_d & 0 \\ -v^{\mathrm{T}} & 1 \end{bmatrix} \\
&= \begin{bmatrix} \mathrm{I}_d & u \\ 0 & 1+v^{\mathrm{T}}u \end{bmatrix}.
\end{aligned}$$

Now, we can reduce the case of a general M to this simpler case by noting that

$$\begin{aligned}
\det\big(M+uv^{\mathrm{T}}\big) &= \det(M)\ \det\Big(\mathrm{I}_d+\big(M^{-1}u\big)v^{\mathrm{T}}\Big) \\
&= \det(M)\ \big(1+v^{\mathrm{T}}M^{-1}u\big)\,.
\end{aligned}$$

D.2. Proof of Inequality (38)

This inequality is used in Lemma 19.1 of the monograph by Lattimore & Szepesvári (2018), in the special case $b=1$. The extension to $b>0$ is straightforward.

We fix $b>0$. We want to prove that

$$\forall u>0, \qquad \min\{b,u\} \leqslant b\,\frac{\ln(1+u)}{\ln(1+b)}\,. \tag{39}$$

We first note that

$$\min\{b,u\} = b\,\frac{\ln(1+u)}{\ln(1+b)} \quad \text{for } u=b$$

and that $\min\{b,u\} = b$ for $u \geqslant b$, with the right-hand side of (39) being an increasing function of u. Therefore, it suffices to prove (39) for $u \in [0,b]$, where $\min\{b,u\} = u$. Now,

$$u \longmapsto b\,\frac{\ln(1+u)}{\ln(1+b)} - u$$

is a concave and (twice) differentiable function, vanishing at $u=0$ and $u=b$, and is therefore non-negative on $[0,b]$. This concludes the proof.

E. Proof of Theorem 2

Comment: The key observation lies in Step 1 (and is tagged as such); the rest is standard maths.

Because of the expression for the expected losses (8) and the consequence (10) of attainability, the regret can be rewritten as

$$R_T = \sum_{t=1}^{T} \ell_{t,p_t} = \sum_{t=1}^{T} \bigl(\phi(x_t,p_t)^{\mathrm{T}}\theta - c_t\bigr)^2 .$$

We first successively prove (*Step 1*) that for $t \geqslant 2$, if the bound of Lemma 1 holds, namely,

$$\Bigl\| V_{t-1}^{1/2}\bigl(\theta - \widehat{\theta}_{t-1}\bigr)\Bigr\| \leqslant B_{t-1}(\delta t^{-2}), \tag{40}$$

then

$$\ell_{t,p_t} \leqslant 2\beta_{t,p_t} + 2\widetilde{\ell}_{t,p_t}, \tag{41}$$

$$\widetilde{\ell}_{t,p_t} \leqslant \beta_{t,p_t} + \widetilde{\ell}_{t,p_t^\star} - \beta_{t,p_t^\star}, \tag{42}$$

$$\widetilde{\ell}_{t,p_t^\star} \leqslant \beta_{t,p_t^\star}. \tag{43}$$

These inequalities collectively entail the bound $\ell_{t,p_t} \leqslant 4\beta_{t,p_t}$. Of course, because of the boundedness assumptions (5), we also have $\ell_{t,p_t} \leqslant C^2$. It then suffices to bound the sum (*Step 2*) of the ℓ_{t,p_t} by the sum of the $\min\bigl\{C^2, 4\beta_{t,p_t}\bigr\}$ and control for the probability of (40).

Step 1: Proof of (41)–(43). Inequality (42) holds by definition of the algorithm. For (43) and (41), we re-use the inequality (17) proved earlier: for all $p \in \mathcal{P}$,

$$\Bigl(\phi(x_t,p)^{\mathrm{T}}\bigl(\theta - \widehat{\theta}_{t-1}\bigr)\Bigr)^2$$

$$\leqslant \Bigl\| V_{t-1}^{1/2}\bigl(\theta - \widehat{\theta}_{t-1}\bigr)\Bigr\|^2 \bigl\| V_{t-1}^{-1/2}\phi(x_t,p)\bigr\|^2 \tag{44}$$

$$\leqslant B_{t-1}(\delta t^{-2})^2 \bigl\| V_{t-1}^{-1/2}\phi(x_t,p)\bigr\|^2 \stackrel{\text{def}}{=} \beta_{t,p}, \tag{45}$$

where we used the bound (40) for the last inequality. This inequality directly yields (43) by taking $p = p_t^\star$.

Now comes the specific improvement and our key observation: using that $(u+v)^2 \leqslant 2u^2 + 2v^2$, we have

$$\begin{aligned}
\ell_{t,p_t} &= \Bigl(\phi(x_t,p_t)^{\mathrm{T}}\theta - \phi(x_t,p_t)^{\mathrm{T}}\widehat{\theta}_{t-1} \\
&\qquad\qquad + \phi(x_t,p_t)^{\mathrm{T}}\widehat{\theta}_{t-1} - c_t\Bigr)^2 \\
&\leqslant 2\bigl(\phi(x_t,p_t)^{\mathrm{T}}\theta - \phi(x_t,p_t)^{\mathrm{T}}\widehat{\theta}_{t-1}\bigr)^2 \\
&\qquad + 2\underbrace{\bigl(\phi(x_t,p_t)^{\mathrm{T}}\widehat{\theta}_{t-1} - c_t\bigr)^2}_{=\widetilde{\ell}_{t,p_t}},
\end{aligned}$$

which yields (41) via (45) used with $p = p_t$.

Step 2: Summing the bounds. First, the bound (40) holds, by Lemma 1, with probability at least $1-\delta t^{-2}$ for a given $t \geqslant 2$.

By a union bound, it holds for all $t \geqslant 2$ with probability at least $1-\delta$. By bounding ℓ_{t,p_t} by C^2 and the $B_{t-1}(\delta t^{-2})$ by $\overline{B}$, we therefore get, from Step 1, that with probability at least $1-\delta$,

$$\overline{R}_T \leqslant C^2 + \sum_{t=2}^{T} \min\Big\{ C^2,\ 4\overline{B}^2 \left\| V_{t-1}^{-1/2} \phi(x_t, p) \right\|^2 \Big\}.$$

Now, as in the proof of Lemma 3 above (Appendix D),

$$\begin{aligned}
&\sum_{t=2}^{T} \min\Big\{ C^2,\ 4\overline{B}^2 \left\| V_{t-1}^{-1/2} \phi(x_t, p) \right\|^2 \Big\} \\
&= \sum_{t=2}^{T} \min\left\{ C^2,\ 4\overline{B}^2 \left(\frac{\det(V_T)}{\det(V_1)} - 1 \right) \right\} \\
&\leqslant 4\overline{B}^2 \sum_{t=2}^{T} \frac{C^2/\big(4\overline{B}^2\big)}{\ln\Big(1 + C^2/\big(4\overline{B}^2\big)\Big)} \ln\left(\frac{\det(V_t)}{\det(V_{t-1})} \right) \\
&= \frac{C^2}{\ln\Big(1 + C^2/\big(4\overline{B}^2\big)\Big)} \ln\left(\frac{\det(V_T)}{\det(V_1)} \right) \\
&\leqslant \left(4\overline{B}^2 + \frac{C^2}{2} \right) d \ln \frac{\lambda + T}{\lambda}.
\end{aligned}$$

This concludes the proof.

F. Numerical expression of the covariance matrix Γ built on data

The covariance matrix Γ was built based on historical data as indicated in Section 5.1. Namely, we considered the time series of residuals associated with our estimation of the consumption. The diagonal coefficients $\Gamma_{j,j}$ were given by the empirical variance of the residuals associated with tariff j, while non-diagonal coefficients $\Gamma_{j,j'}$ were given by the empirical covariance between residuals of tariffs j and j' at times t and $t \pm 48$. (A more realistic model might consider a noise which depends on the half-hour of the day).

Numerical expression obtained. More precisely, the variance terms $\Gamma_{1,1}$, $\Gamma_{2,2}$, and $\Gamma_{3,3}$ were computed with respectively 788, $15\,072$ and $1\,660$ observations, while the non-diagonal coefficients were based on fewer observations: $1\,318$ for $\Gamma_{2,3}$ and 620 for $\Gamma_{1,2}$, but only 96 for $\Gamma_{1,3}$. The resulting matrix Γ is

$$\Gamma = \sigma^2 \begin{pmatrix} 1.11 & 0.46 & 0.04 \\ 0.46 & 1.00 & 0.56 \\ 0.04 & 0.56 & 2.07 \end{pmatrix} \quad \text{with} \quad \sigma = 0.02.$$

To get an idea of the orders of magnitude at stake, we indicate that in the data set considered, the mean consumption remained between 0.08 and 0.21 kWh per half-hour and that its empirical average equals 0.46.

Off-diagonal coefficients are non-zero. We may test, for each $j \neq j'$, the null hypothesis $\Gamma_{j,j'} = 0$ using the Pearson correlation test; we obtain low p–values (smaller than something of the order of 10^{-13}), which shows that Γ is significantly different from a diagonal matrix. We may conduct a similar study to show that it is not proportional to the all-ones matrix, nor to any matrix with a special form.