A Technical Proofs

We gather in this section the proofs omitted in the core text.

A.1 Proof of Proposition 1

Proposition 1 *The oracle greedy strategy, which plays $a_t^{\text{greedy}} = \arg\max_{a \in \mathcal{A}} \langle a, A_{t-1}\theta^* \rangle$ at time step t, can suffer linear regret, both in rotting or rising scenarios.*

Proof We build two instances of LBM, one rotting, one rising, in which the oracle greedy strategy suffers linear regret. We highlight that the other strategy exhibited, which performs better than oracle greedy, may not be optimal.

Rotting instance. Let $\mathcal{A} = \mathcal{B}_d$, $\theta^* = e_1$, $m = d - 1$, and A such that

$$A(a_1, \ldots, a_m) = \left(I_d + \sum_{s=1}^{m} a_s a_s^\top \right)^{-\gamma} ,$$

for some $\gamma > 0$ to be specified later. Oracle greedy, which plays at each time step $a_t^{\text{greedy}} = \arg\max_{a \in \mathcal{A}} \langle a, A_{t-1}\theta^* \rangle$, constantly plays e_1. After the first m pulls, it collects a reward of $1/d^\gamma$ at every time step. On the other side, the strategy that plays cyclically the block $e_1 \ldots e_d$ collects a reward of 1 every $d = m + 1$ time steps, i.e., an average reward of $1/d$ per step. Hence, up to the transitive first m puuls, the cumulative reward of oracle greedy after T rounds is T/d^γ, and that of the cyclic policy is T/d. The regret of oracle greedy is thus at least

$$T \left(\frac{1}{d} - \frac{1}{d^\gamma} \right) ,$$

which is linear for $\gamma > 1$.

Rising instance. Let $m \geq 1$, $d = 2$, $\mathcal{A} = \mathcal{B}_2$, $\theta^* = (\varepsilon, 1)$ where $\varepsilon > 0$ is to be specified later, and A such that

$$A(a_1, \ldots, a_m) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \sum_{s=1}^{m} a_s a_s^\top .$$

Oracle greedy constantly plays e_1 collecting a reward of $(m+1)\theta_1^*$ from round $m + 1$ onward. On the other side, the strategy that plays constantly e_2 collects a reward of $m\theta_2^*$ from round $m + 1$ onward. Hence, the regret of oracle greedy from round $m + 1$ onward is at least $(T - m)[m - (m + 1)\varepsilon]$, which is linear for $\varepsilon < m/(m + 1)$. $\square$

A.2 Proof of Proposition 2

Proposition 2 *For any $m, L \geq 1$, let $\widetilde{\boldsymbol{a}}$ be the block of $m + L$ actions defined in (5) and $(\widetilde{r}_t)_{t=1}^T$ be the expected rewards collected when playing cyclically $\widetilde{\boldsymbol{a}}$. We have*

$$\mathrm{OPT} - \sum_{t=1}^{T} \widetilde{r}_t \leq \frac{2mR}{m+L} T . \tag{6}$$

Proof Recall that the optimal sequence is denoted $(a_t^*)_{t=1}^T$ and collects rewards $(r_t^*)_{t=1}^T$. Let $L > 0$; by definition, there exists a block of actions of length L in $(a_t^*)_{t=1}^T$ with average expected reward higher that OPT/T. Let t^* be the first index of this block, we thus have $(1/L) \sum_{t=t^*}^{t^*+L-1} r_t^* \geq \mathrm{OPT}/T$. However, this average expected reward is realized only using the initial matrix A_{t^*-1}, generated from $a_{t^*-1}^*, \ldots, a_{t^*-m}^*$. Let $\boldsymbol{a}^* = a_{t^*-m}^*, \ldots, a_{t^*+L-1}^*$ of length $m + L$. Note that, by definition, we have that $\widetilde{r}(\widetilde{\boldsymbol{a}}) \geq \widetilde{r}(\boldsymbol{a}^*) = \sum_{t=t^*}^{t^*+L-1} r_t^* \geq L\ \mathrm{OPT}/T$. Furthermore, by (8), when playing cyclically $\widetilde{\boldsymbol{a}}$ one obtains at least a reward of $-R$ in each one of the first m pulls of the block. Collecting all the pieces, we obtain

$$\sum_{t=1}^{T} \widetilde{r}_t \geq \frac{T}{m+L} \Big(-mR + \widetilde{r}(\widetilde{\boldsymbol{a}}) \Big)$$

$$
\begin{aligned}
&\geq \frac{T}{m+L}\Big(-mR + \widetilde{r}(\boldsymbol{a}^*)\Big) \\
&\geq \frac{T}{m+L}\left(-mR + L\,\frac{\mathrm{OPT}}{T}\right) \\
&= \frac{L}{m+L}\mathrm{OPT} - \frac{mR}{m+L}T \\
&\geq \frac{L}{m+L}\mathrm{OPT} + \frac{m}{m+L}\mathrm{OPT} - \frac{mR}{m+L}T - \frac{mR}{m+L}T \qquad (14)\\
&= \mathrm{OPT} - \frac{2mR}{m+L}T\,,
\end{aligned}
$$

where (14) derives from $\mathrm{OPT} \leq RT$. □

A.3 Proof of Proposition 4

We prove the (stronger) high probability version of Proposition 4.

Proposition 5 *Let $\lambda \geq 1$, $\delta \in (0,1)$, and $\boldsymbol{a}_\tau$ be the blocks of actions in $\mathbb{R}^{d(m+L)}$ associated to the $\boldsymbol{b}_\tau$ defined in (9). Then, with probability at least $1-\delta$ we have*

$$
\sum_{\tau=1}^{T/(m+L)} \widetilde{r}(\widetilde{\boldsymbol{a}}) - \widetilde{r}(\boldsymbol{a}_\tau) \leq 4L(m+1)^{\gamma^+}\sqrt{Td\ \ln\left(1+\frac{T(m+1)^{2\gamma^+}}{d(m+L)\lambda}\right)} \cdot \left(\sqrt{\lambda L} + \sqrt{\ln\left(\frac{1}{\delta}\right) + d(m+L)\ \ln\left(1+\frac{T(m+1)^{2\gamma^+}}{d(m+L)\lambda}\right)}\right).
$$

Proof The proof essentially follows that of (Abbasi-Yadkori et al., 2011, Theorem 3). The main difference is that our version of OFUL operates at the block level. This implies a smaller time horizon, but also and increased dimension and an instantaneous regret $\langle \widetilde{\boldsymbol{b}}, \boldsymbol{\theta}^*\rangle - \langle \boldsymbol{b}_\tau, \boldsymbol{\theta}^*\rangle$ upper bounded by $2L(m+1)^{\gamma^+}$ instead of 1. We detail the main steps of the proof for completeness. Recall that running OFUL in our case amounts to compute at every block time step τ

$$
\widehat{\boldsymbol{\theta}}_\tau = \boldsymbol{V}_\tau^{-1}\left(\sum_{\tau'=1}^{\tau} \boldsymbol{y}_{\tau'}\,\boldsymbol{b}_{\tau'}\right),
$$

where

$$
\boldsymbol{V}_\tau = \sum_{\tau'=1}^{\tau} \boldsymbol{b}_{\tau'}\boldsymbol{b}_{\tau'}^\top + \lambda I_{d(m+L)}\,, \qquad \text{and} \qquad \boldsymbol{y}_\tau = \sum_{i=m+1}^{m+L} y_{\tau,i}\,,
$$

since we associate with a block of actions the sum of rewards obtained after time step m. Note that by the determinant-trace inequality, see e.g., (Abbasi-Yadkori et al., 2011, Lemma 10), with actions $\boldsymbol{b}_\tau$ that satisfy $\|\boldsymbol{b}_\tau\|_2^2 \leq m + L(m+1)^{2\gamma^+}$ we have

$$
\frac{|\boldsymbol{V}_\tau|}{|\lambda I_{d(m+L)}|} \leq \left(1 + \frac{\tau(m+L(m+1)^{2\gamma^+})}{d(m+L)\lambda}\right)^{d(m+L)} \leq \left(1+\frac{\tau(m+1)^{2\gamma^+}}{d\lambda}\right)^{d(m+L)}. \qquad (15)
$$

The action played at block time step τ is the block $\boldsymbol{a}_\tau \in \mathcal{B}_d^{m+L}$ associated with

$$
\boldsymbol{b}_\tau = \operatorname*{arg\,max}_{\boldsymbol{b}\in\boldsymbol{\mathcal{B}}}\ \sup_{\boldsymbol{\theta}\in\boldsymbol{\mathcal{C}}_{\tau-1}} \langle \boldsymbol{b}, \boldsymbol{\theta}\rangle\,, \qquad (16)
$$

where

$$
\boldsymbol{\mathcal{C}}_\tau = \left\{\boldsymbol{\theta}\in\mathbb{R}^{d(m+L)} \colon \big\|\widehat{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}\big\|_{\boldsymbol{V}_\tau} \leq \beta_\tau(\delta)\right\},
$$

with

$$\boldsymbol{\beta}_\tau(\delta) = \sqrt{2\ln\left(\frac{1}{\delta}\right) + d(m+L)\,\ln\left(1 + \frac{\tau(m+1)^{2\gamma^+}}{d\lambda}\right)} + \sqrt{\lambda L}\,. \tag{17}$$

Applying (Abbasi-Yadkori et al., 2011, Theorem 2) to $\boldsymbol{\theta}^* \in \mathbb{R}^{d(m+L)}$ which satisfies $\|\boldsymbol{\theta}^*\|_2 \le \sqrt{L}$ we have that $\boldsymbol{\theta}^* \in \boldsymbol{\mathcal{C}}_\tau$ for every τ with probability at least $1-\delta$. Denoting by $\widetilde{\boldsymbol{\theta}}_\tau$ the model that maximizes (16), we thus have that with probability at least $1-\delta$, the inequality $\langle \widetilde{\boldsymbol{b}}, \boldsymbol{\theta}^*\rangle \le \langle \boldsymbol{b}_\tau, \widetilde{\boldsymbol{\theta}}_\tau\rangle$ holds for every τ, and consequently

$$\begin{aligned}
&\sum_{\tau=1}^{T/(m+L)} \langle \widetilde{\boldsymbol{b}}, \boldsymbol{\theta}^*\rangle - \langle \boldsymbol{b}_\tau, \boldsymbol{\theta}^*\rangle \\
&\qquad\le \sum_{\tau=1}^{T/(m+L)} \min\left\{2L(m+1)^{\gamma^+}, \langle \boldsymbol{b}_\tau, \widetilde{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}^*\rangle\right\} \\
&\qquad\le \sum_{\tau=1}^{T/(m+L)} \min\left\{2L(m+1)^{\gamma^+}, \|\widetilde{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}^*\|_{\boldsymbol{V}_{\tau-1}} \|\boldsymbol{b}_\tau\|_{\boldsymbol{V}_{\tau-1}^{-1}}\right\} \\
&\qquad\le \sum_{\tau=1}^{T/(m+L)} \min\left\{2L(m+1)^{\gamma^+}, 2\boldsymbol{\beta}_\tau(\delta)\,\|\boldsymbol{b}_\tau\|_{\boldsymbol{V}_{\tau-1}^{-1}}\right\} \\
&\qquad\le 2L(m+1)^{\gamma^+}\,\boldsymbol{\beta}_{T/(m+L)}(\delta) \sum_{\tau=1}^{T/(m+L)} \min\left\{1, \|\boldsymbol{b}_\tau\|_{\boldsymbol{V}_{\tau-1}^{-1}}\right\} \\
&\qquad\le 2L(m+1)^{\gamma^+}\,\boldsymbol{\beta}_{T/(m+L)}(\delta) \sqrt{\frac{T}{m+L}\sum_{\tau=1}^{T/(m+L)} \min\left\{1, \|\boldsymbol{b}_\tau\|^2_{\boldsymbol{V}_{\tau-1}^{-1}}\right\}} \\
&\qquad\le 2\sqrt{2}L(m+1)^{\gamma^+}\,\boldsymbol{\beta}_{T/(m+L)}(\delta) \sqrt{\frac{T}{m+L}\,\ln\frac{|\boldsymbol{V}_{T/(m+L)}|}{|\lambda I_{d(m+L)}|}} \\
&\qquad\le 4L(m+1)^{\gamma^+}\sqrt{Td\,\ln\left(1+\frac{T(m+1)^{2\gamma^+}}{d(m+L)\lambda}\right)} \\
&\qquad\qquad\cdot\left(\sqrt{\lambda L} + \sqrt{\ln\left(\frac{1}{\delta}\right) + d(m+L)\,\ln\left(1+\frac{T(m+1)^{2\gamma^+}}{d(m+L)\lambda}\right)}\right),
\end{aligned}$$

where we have used (Abbasi-Yadkori et al., 2011, Lemma 11), as well as (15) and (17). Note that in the stationary case, i.e., when $m=0$ and $L=1$, we exactly recover (Abbasi-Yadkori et al., 2011, Theorem 3). Proposition 4 is obtained by setting $\lambda \in [1,d]$, $L \ge m$, and $\delta = 1/T$. □

A.4 Proof of Proposition 3

Proof Let $d = m+1$, $\mathcal{A} = \{0_d\} \cup (e_k)_{k\le d}$, $\theta^* = (1/\sqrt{d}, \dots, 1/\sqrt{d})$, and $\gamma \le 0$. For simplicity, we note the basis modulo d, i.e., $e_{k+d} = e_k$ for any $k \in \mathbb{N}$. Note that for any $a_1, \dots, a_{m+1} \in \mathcal{A}$ we have $\left|\langle a_{m+1}, A_m\theta^*\rangle\right| \le \|a_{m+1}\|_1\, \|A_m\theta^*\|_\infty \le 1/\sqrt{d}$, such that one can take $R = 1/\sqrt{d}$. Observe now that the strategy which plays cyclically $e_1, \dots, e_d$ collects a reward of $1/\sqrt{d}$ at each time step, which is optimal, such that $\mathrm{OPT} = T/\sqrt{d}$. Further, it is easy to check that block $\widetilde{\boldsymbol{a}}$, composed of m pulls of 0_d followed by $e_1, \dots, e_L$ satisfies $\widetilde{r}(\widetilde{\boldsymbol{a}}) = L/\sqrt{d}$, which is optimal for similar reasons. Playing cyclically $\widetilde{\boldsymbol{a}}$, one gets a reward of $L/\sqrt{d}$ every $m+L$ pulls. In other terms, we have

$$\mathrm{OPT} - \sum_{t=1}^{T}\widetilde{r}_t = \frac{T}{\sqrt{d}} - \frac{L}{m+L}\frac{T}{\sqrt{d}} = \frac{m}{m+L}\frac{T}{\sqrt{d}} = \frac{mR}{m+L}T\,.$$

□

A.5 Proof of Theorem 1

We prove the high probability version of Theorem 1, obtained by setting $\lambda \in [1, d]$, and $\delta = 1/T$.

Theorem 2 *Let $\lambda \geq 1$, $\delta \in (0,1)$, and $\boldsymbol{a}_\tau$ be the blocks of actions in $\mathbb{R}^{d(m+L)}$ defined in (11). Then, with probability at least $1-\delta$ we have*

$$\sum_{\tau=1}^{T/(m+L)} \widetilde{r}(\widetilde{\boldsymbol{a}}) - \widetilde{r}(\boldsymbol{a}_\tau) \leq 4L(m+1)^{\gamma^+} \sqrt{Td \, \ln\left(1 + \frac{T(m+1)^{2\gamma^+}}{d\lambda}\right)} \cdot \left(\sqrt{\lambda} + \sqrt{\ln\left(\frac{1}{\delta}\right) + d \, \ln\left(1 + \frac{T(m+1)^{2\gamma^+}}{d(m+L)\lambda}\right)}\,\right).$$

Let $m \geq 1$, $T \geq m^2 d^2 + 1$, and set $L = \lceil \sqrt{m/d}\, T^{1/4} \rceil - m$. Let r_t be the rewards collected when playing $\boldsymbol{a}_\tau$ as defined in (11). Then, with probability at least $1-\delta$ we have

$$\mathrm{OPT} - \sum_{t=1}^{T} r_t \leq 4\sqrt{d}\,(m+1)^{\frac{1}{2}+\gamma^+}\, T^{3/4}\left[1 + 2\sqrt{\ln\left(1+\frac{T(m+1)^{2\gamma^+}}{d\lambda}\right)} \cdot \left(\sqrt{\frac{\lambda}{d}} + \sqrt{\frac{\ln(1/\delta)}{d} + \ln\left(1 + \frac{T(m+1)^{2\gamma^+}}{d\lambda}\right)}\,\right)\right].$$

Proof The proof is along the lines of OFUL's analysis. The main difficulty is that we cannot use the elliptical potential lemma, see e.g., (Lattimore & Szepesvári, 2020, Lemma 19.4) due to the delay accumulated by V_τ, which is computed every $m+L$ round only. Let

$$\beta_\tau(\delta) = \sqrt{2\ln\left(\frac{1}{\delta}\right) + d\,\ln\left(1 + \frac{\tau(m+1)^{2\gamma^+}}{d\lambda}\right)} + \sqrt{\lambda}\,. \tag{18}$$

By (Abbasi-Yadkori et al., 2011, Theorem 2), we have with probability at least $1-\delta$ that $\theta^* \in \mathcal{C}_\tau$ for every τ. It follows directly that $\boldsymbol{\theta}^* \in \boldsymbol{\mathcal{D}}_\tau$ for any τ, such that $\langle \widetilde{\boldsymbol{b}}, \boldsymbol{\theta}^* \rangle \leq \langle \boldsymbol{b}_\tau, \widetilde{\boldsymbol{\theta}}_\tau \rangle$, where $\widetilde{\boldsymbol{\theta}}_\tau = (0_d, \dots, 0_d, \widetilde{\theta}_\tau, \dots, \widetilde{\theta}_\tau)$ with $\widetilde{\theta}_\tau \in \mathbb{R}^d$ that maximizes (11) over $\mathcal{C}_{\tau-1}$. It can be shown that the regret is upper bounded by $\sum_\tau \sum_{i=m+1}^{m+L} \langle b_{\tau,i}, \widetilde{\theta}_\tau - \theta^* \rangle$. Following the standard analysis, one could then use

$$\left\langle b_{\tau,i}, \widetilde{\theta}_\tau - \theta^* \right\rangle \leq \|b_{\tau,i}\|_{V_{\tau-1}^{-1}} \left\|\widetilde{\theta}_t - \theta^*\right\|_{V_{\tau-1}}.$$

While the confidence set gives $\left\|\widetilde{\theta}_t - \theta^*\right\|_{V_{\tau-1}} \leq 2\beta_{\tau-1}(\delta)$, the quantity $\sum_{i=m+1}^{m+L} \|b_{\tau,i}\|_{V_{\tau-1}^{-1}}$ is much more complex to bound. Indeed, the elliptical potential lemma allows to bound $\sum_t \|a_t\|^2_{V_{t-1}^{-1}}$ when $V_t = \sum_{s \leq t} a_s a_s^\top + \lambda I_d$. However, recall that in our case we have $V_\tau = \sum_{\tau'=1}^{\tau} \sum_{i=m+1}^{m+L} b_{\tau',i} b_{\tau',i}^\top + \lambda I_d$, which is only computed every $m+L$ rounds. As a consequence, there exists a "delay" between $V_{\tau-1}$ and the action $b_{\tau,i}$ for $i \geq m+2$, preventing from using the lemma. Therefore, we propose to use instead

$$\left\langle b_{\tau,i}, \widetilde{\theta}_\tau - \theta^* \right\rangle \leq \|b_{\tau,i}\|_{V_{\tau,i-1}^{-1}} \left\|\widetilde{\theta}_t - \theta^*\right\|_{V_{\tau,i-1}}, \quad \text{where} \quad V_{\tau,i} = V_{\tau-1} + \sum_{j=m+1}^{i} b_{\tau,j} b_{\tau,j}^\top\,. \tag{19}$$

By doing so, the elliptical potential lemma applies. On the other hand, one has to control $\left\|\widetilde{\theta}_t - \theta^*\right\|_{V_{\tau,i-1}}$, which is not anymore bounded by $2\beta_{\tau-1}(\delta)$ since the subscript matrix is $V_{\tau,i-1}$ instead of $V_{\tau-1}$. Still, one can show that for any $i \leq m+L$ we have

$$\begin{aligned}
&\left\|\widetilde{\theta}_t - \theta^*\right\|^2_{V_{\tau,i-1}} \\
&\quad = \mathrm{Tr}\left(V_{\tau,i-1}\,\left(\widetilde{\theta}_t - \theta^*\right)\left(\widetilde{\theta}_t - \theta^*\right)^\top\right)
\end{aligned}$$

$$
\begin{aligned}
&= \mathrm{Tr}\left(\left(V_{\tau-1} + \sum_{j=m+1}^{i-1} b_{\tau,j} b_{\tau,j}^\top\right) \left(\widetilde{\theta}_t - \theta^*\right)\left(\widetilde{\theta}_t - \theta^*\right)^\top\right) \\
&= \mathrm{Tr}\left(\left(I_d + \sum_{j=m+1}^{i-1} \left(V_{\tau-1}^{-1/2} b_{\tau,j}\right)\left(V_{\tau-1}^{-1/2} b_{\tau,j}\right)^\top\right) V_{\tau-1}^{1/2}\left(\widetilde{\theta}_t - \theta^*\right)\left(\widetilde{\theta}_t - \theta^*\right)^\top V_{\tau-1}^{1/2}\right) \\
&\leq \left\| I_d + \sum_{j=m+1}^{i-1} \left(V_{\tau-1}^{-1/2} b_{\tau,j}\right)\left(V_{\tau-1}^{-1/2} b_{\tau,j}\right)^\top \right\|_* \mathrm{Tr}\left(V_{\tau-1}^{1/2}\left(\widetilde{\theta}_t - \theta^*\right)\left(\widetilde{\theta}_t - \theta^*\right)^\top V_{\tau-1}^{1/2}\right) \\
&\leq \left(1 + \sum_{j=m+1}^{i-1} \left\|V_{\tau-1}^{-1/2} b_{\tau,j}\right\|_2^2\right) \left\|\widetilde{\theta}_t - \theta^*\right\|_{V_{\tau-1}}^2 \\
&\leq \left(1 + (L-1)(m+1)^{2\gamma^+}\right) \left\|\widetilde{\theta}_t - \theta^*\right\|_{V_{\tau-1}}^2 \\
&\leq L(m+1)^{2\gamma^+} \left\|\widetilde{\theta}_t - \theta^*\right\|_{V_{\tau-1}}^2 .
\end{aligned} \tag{20}
$$

Recalling also that $\langle \widetilde{\boldsymbol{b}}, \boldsymbol{\theta}^* \rangle - \langle \boldsymbol{b}_\tau, \boldsymbol{\theta}^* \rangle \leq 2L(m+1)^{\gamma^+}$, we have with probability at least $1-\delta$

$$
\begin{aligned}
&\sum_{\tau=1}^{T/(m+L)} \langle \widetilde{\boldsymbol{b}}, \boldsymbol{\theta}^* \rangle - \langle \boldsymbol{b}_\tau, \boldsymbol{\theta}^* \rangle \\
&\qquad \leq \sum_{\tau=1}^{T/(m+L)} \min\left\{2L(m+1)^{\gamma^+} , \langle \boldsymbol{b}_\tau, \widetilde{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}^* \rangle\right\} \\
&\qquad = \sum_{\tau=1}^{T/(m+L)} \min\left\{2L(m+1)^{\gamma^+} , \sum_{i=m+1}^{m+L} \langle b_{\tau,i}, \widetilde{\theta}_\tau - \theta^* \rangle\right\} \\
&\qquad \leq \sum_{\tau=1}^{T/(m+L)} \min\left\{2L(m+1)^{\gamma^+} , \sum_{i=m+1}^{m+L} \|b_{\tau,i}\|_{V_{\tau,i-1}^{-1}} \left\|\widetilde{\theta}_t - \theta^*\right\|_{V_{\tau,i-1}}\right\} \\
&\qquad \leq \sum_{\tau=1}^{T/(m+L)} \min\left\{2L(m+1)^{\gamma^+} , 2\sqrt{L}(m+1)^{\gamma^+} \beta_{\tau-1}(\delta) \sum_{i=m+1}^{m+L} \|b_{\tau,i}\|_{V_{\tau,i-1}^{-1}}\right\} \\
&\qquad \leq 2L(m+1)^{\gamma^+} \beta_{T/(m+L)}(\delta) \sum_{\tau=1}^{T/(m+L)} \sum_{i=m+1}^{m+L} \min\left\{1 , \|b_{\tau,i}\|_{V_{\tau,i-1}^{-1}}\right\} \\
&\qquad \leq 2L(m+1)^{\gamma^+} \beta_{T/(m+L)}(\delta) \sqrt{\frac{T\,L}{m+L} \sum_{\tau=1}^{T/(m+L)} \sum_{i=m+1}^{m+L} \min\left\{1 , \|b_{\tau,i}\|_{V_{\tau,i-1}^{-1}}^2\right\}} \\
&\qquad \leq 2\sqrt{2}L(m+1)^{\gamma^+} \beta_{T/(m+L)}(\delta) \sqrt{T \ \ln \frac{|V_{T/(m+L)}|}{|\lambda I_d|}} \\
&\qquad \leq 4L(m+1)^{\gamma^+} \sqrt{Td \ \ln\left(1 + \frac{T(m+1)^{2\gamma^+}}{d\lambda}\right)} \\
&\qquad\qquad \cdot \left(\sqrt{\lambda} + \sqrt{\ln\left(\frac{1}{\delta}\right) + d \ \ln\left(1 + \frac{T(m+1)^{2\gamma^+}}{d(m+L)\lambda}\right)}\right) ,
\end{aligned} \tag{21}
$$

where we have used (18), (19), and (20). Similarly to Proposition 5, note that in the stationary case, i.e., when $m = 0$ and $L = 1$, we exactly recover (Abbasi-Yadkori et al., 2011, Theorem 3). The first claim of Theorem 1 is obtained by setting $\lambda \in [1, d]$, and $\delta = 1/T$.

Let R_T denote the right-hand side of (21). Combining this bound with the arguments of Proposition 2, we have with probability $1-\delta$

$$
\begin{aligned}
\sum_{t=1}^{T} r_t &\geq \sum_{\tau=1}^{T/(m+L)} \widetilde{r}(\boldsymbol{a}_\tau) - \frac{m(m+1)^{\gamma^+}}{m+L}\, T && (22)\\
&= \sum_{\tau=1}^{T/(m+L)} \langle \boldsymbol{b}_\tau, \boldsymbol{\theta}^* \rangle - \frac{m(m+1)^{\gamma^+}}{m+L}\, T \\
&\geq \sum_{\tau=1}^{T/(m+L)} \langle \widetilde{\boldsymbol{b}}, \boldsymbol{\theta}^* \rangle - R_T - \frac{m(m+1)^{\gamma^+}}{m+L}\, T && (23)\\
&= \sum_{\tau=1}^{T/(m+L)} \widetilde{r}(\widetilde{\boldsymbol{a}}) - R_T - \frac{m(m+1)^{\gamma^+}}{m+L}\, T \\
&\geq \sum_{t=1}^{T} \widetilde{r}_t - R_T - \frac{2m(m+1)^{\gamma^+}}{m+L}\, T && (24)\\
&\geq \mathrm{OPT} - R_T - \frac{4m(m+1)^{\gamma^+}}{m+L}\, T && (25)\\
&\geq \mathrm{OPT} - 4(m+1)^{\gamma^+}\Bigg[\frac{mT}{m+L} + (m+L)\sqrt{Td\;\ln\left(1+\frac{T(m+1)^{2\gamma^+}}{d\lambda}\right)} \\
&\qquad\qquad \cdot \left(\sqrt{\lambda} + \sqrt{\ln\left(\frac{1}{\delta}\right) + d\ln\left(1+\frac{T(m+1)^{2\gamma^+}}{d(m+L)\lambda}\right)}\,\right)\Bigg],
\end{aligned}
$$

where (22) and (24) come from the fact that any instantaneous reward is bounded by $(m+1)^{\gamma^+}$, see (8), (23) from (21), and (25) from Proposition 2.

Now, assume that $m \geq 1$, $T \geq d^2m^2+1$, and let $L = \lceil \sqrt{m/d}\; T^{1/4} \rceil - m$. By the condition on T, we have $\sqrt{m/d}\; T^{1/4} > m \geq 1$, such that $L \geq 1$ and

$$
\sqrt{\frac{m}{d}}\, T^{1/4} \leq \left\lceil \sqrt{\frac{m}{d}}\, T^{1/4} \right\rceil = L+m \leq \sqrt{\frac{m}{d}}\, T^{1/4} + 1 \leq 2\sqrt{\frac{m}{d}}\, T^{1/4}\,.
$$

Substituting in the above bound, we have with probability $1-\delta$

$$
\begin{aligned}
\mathrm{OPT} - \sum_{t=1}^{T} r_t \leq 4\sqrt{d}\,(m+1)^{\frac{1}{2}+\gamma^+}\; T^{3/4}\Bigg[1 + 2\sqrt{\ln\left(1+\frac{T(m+1)^{2\gamma^+}}{d\lambda}\right)} \\
\cdot \left(\sqrt{\frac{\lambda}{d}} + \sqrt{\frac{\ln(1/\delta)}{d} + \ln\left(1+\frac{T(m+1)^{2\gamma^+}}{d\lambda}\right)}\,\right)\Bigg].
\end{aligned}
$$

The second claim of Theorem 1 is obtained by setting $\lambda \in [1,d]$, and $\delta = 1/T$. □

A.6 Proof of Corollary 1

Lemma 1 *Suppose that a block-based bandit algorithm (in our case the bandit combiner) produces a sequence of T_{bc} blocks $\boldsymbol{a}_\tau$, with possibly different cardinalities $|\boldsymbol{a}_\tau|$, such that*

$$
\sum_{\tau=1}^{T_{\mathrm{bc}}} \frac{\widetilde{r}(\widetilde{\boldsymbol{a}})}{|\widetilde{\boldsymbol{a}}|} - \sum_{\tau=1}^{T_{\mathrm{bc}}} \frac{\widetilde{r}(\boldsymbol{a}_\tau)}{|\boldsymbol{a}_\tau|} \leq F(T_{\mathrm{bc}})\,,
$$

for some sublinear function F. Then, we have

$$
\frac{\min_\tau |\boldsymbol{a}_\tau|}{\max_\tau |\boldsymbol{a}_\tau|}\left(\widetilde{r}(\widetilde{\boldsymbol{a}})\;\frac{\sum_\tau |\boldsymbol{a}_\tau|}{|\widetilde{\boldsymbol{a}}|}\right) - \sum_{\tau=1}^{T_{\mathrm{bc}}} \widetilde{r}(\boldsymbol{a}_\tau) \;\;\leq\;\; \min_\tau |\boldsymbol{a}_\tau|\, F(T_{\mathrm{bc}})\;.
$$

In particular, if all blocks have the same cardinality the last bound is just the block regret bound scaled by $|\boldsymbol{a}_\tau|$.

Proof We have

$$\begin{aligned}
\sum_{\tau=1}^{T_{\rm bc}} \widetilde{r}(\boldsymbol{a}_\tau) &\geq \min_\tau |\boldsymbol{a}_\tau| \sum_{\tau=1}^{T_{\rm bc}} \frac{\widetilde{r}(\boldsymbol{a}_\tau)}{|\boldsymbol{a}_\tau|} \\
&\geq \min_\tau |\boldsymbol{a}_\tau| \left(\sum_{\tau=1}^{T_{\rm bc}} \frac{\widetilde{r}(\widetilde{\boldsymbol{a}})}{|\widetilde{\boldsymbol{a}}|} - F(T_{\rm bc}) \right) \\
&= \frac{\min_\tau |\boldsymbol{a}_\tau|}{\max_\tau |\boldsymbol{a}_\tau|} \frac{\widetilde{r}(\widetilde{\boldsymbol{a}})}{|\widetilde{\boldsymbol{a}}|} \max_\tau |\boldsymbol{a}_\tau| \, T_{\rm bc} - \min_\tau |\boldsymbol{a}_\tau| \, F(T_{\rm bc}) \\
&\geq \frac{\min_\tau |\boldsymbol{a}_\tau|}{\max_\tau |\boldsymbol{a}_\tau|} \left(\widetilde{r}(\widetilde{\boldsymbol{a}}) \frac{\sum_\tau |\boldsymbol{a}_\tau|}{|\widetilde{\boldsymbol{a}}|} \right) - \min_\tau |\boldsymbol{a}_\tau| \, F(T_{\rm bc}) \, .
\end{aligned}$$

□

Corollary 1 *Consider an instance of LBM with unknown parameters* $(m_\star, \gamma_\star)$. *Assume a bandit combiner is run on* $N \leq d\sqrt{m_\star}$ *instances of OFUL-memory (Algorithm 2), each using a different pair of parameters* (m_i, γ_i) *from a set* $\mathcal{S} = \big\{(m_1, \gamma_1), \ldots, (m_N, \gamma_N)\big\}$ *such that* $(m_\star, \gamma_\star) \in \mathcal{S}$. *Let* $M = (\max_j m_j)/(\min_j m_j)$. *Then, for all* $T \geq (m_\star + 1)^{2\gamma_\star^+}/m_\star d^4$, *the expected rewards* $\big(r_t^{\rm bc}\big)_{t=1}^T$ *of the bandit combiner satisfy*

$$\frac{\mathrm{OPT}}{\sqrt{M}} - \mathbb{E}\left[\sum_{t=1}^T r_t^{\rm bc}\right] \;=\; \widetilde{\mathcal{O}}\Big(M\, d\, (m_\star+1)^{1+\frac{3}{2}\gamma_\star^+} \, T^{3/4} \Big) \, .$$

Proof Let $m_\star$ be the true memory size, and $L_\star = L(m_\star)$ the corresponding (partial) block length. Throughout the proof, $\widetilde{\boldsymbol{a}}$ denotes the block defined in (5) with length $m_\star + L_\star$. First observe that only one of the OFUL-memory instances we test is well-specified, i.e., has the true parameters $(m_\star, \gamma_\star)$. We can thus rewrite the regret bound for the Bandit Combiner (Cutkosky et al., 2020, Corollary 2), generalized to rewards bounded in $[-R, R]$ as follows

$$\mathrm{Regret}_{\rm bc} = \widetilde{\mathcal{O}} \left(C_\star T_{\rm bc}^{\alpha_\star} + C_\star^{\frac{1}{\alpha_\star}} T_{\rm bc} \eta_\star^{\frac{1-\alpha_\star}{\alpha_\star}} + R^2 T_{\rm bc} \eta_\star + \sum_{j \neq \star} \frac{1}{\eta_j} \right) \, , \tag{26}$$

where $T_{\rm bc} = T/(m_\star + L_\star)$ is the bandit combiner horizon, $C_\star$ and $\alpha_\star$ are the constants in the regret bound of the well-specified instance (see below how we determine them), and the η_j are free parameters to be tuned. We now derive $C_\star$ and $\alpha_\star$. To that end, we must establish the regret bound of the well-specified instance, and identify $C_\star$ and $\alpha_\star$ such that this bound is equal to $C_\star T_{\rm bc}^{\alpha_\star}$, where $C_\star$ may contain logarithmic factors. For the well-specified instance, the first claim of Theorem 2 gives that, with probability at least $1-\delta$, we have

$$\begin{aligned}
\sum_{\tau=1}^{T/(m_\star+L_\star)} \widetilde{r}(\widetilde{\boldsymbol{a}}) - \widetilde{r}(\boldsymbol{a}_\tau) \leq 4(m_\star + L_\star)(m_\star+1)^{\gamma_\star^+} & \sqrt{Td \, \ln\left(1 + \frac{T(m_\star+1)^{2\gamma_\star^+}}{d\lambda}\right)} \\
& \left(\sqrt{\lambda} + \sqrt{\ln\left(\frac{1}{\delta}\right) + d\, \ln\left(1 + \frac{T(m_\star+1)^{2\gamma_\star^+}}{d(m_\star+L_\star)\lambda}\right)} \right)
\end{aligned}$$

$$\begin{aligned}
\sum_{\tau=1}^{T/(m_\star+L_\star)} \frac{\widetilde{r}(\widetilde{\boldsymbol{a}})}{|\widetilde{\boldsymbol{a}}|} - \frac{\widetilde{r}(\boldsymbol{a}_\tau)}{|\boldsymbol{a}_\tau|} \leq T^{1/2}\, 4(m_\star+1)^{\gamma_\star^+} & \sqrt{d \ln\left(1 + \frac{T(m_\star+1)^{2\gamma_\star^+}}{d\lambda}\right)} \\
& \left(\sqrt{\lambda} + \sqrt{\ln\left(\frac{1}{\delta}\right) + d\, \ln\left(1 + \frac{T(m_\star+1)^{2\gamma_\star^+}}{d(m_\star+L_\star)\lambda}\right)} \right) \, ,
\end{aligned} \tag{27}$$

where we have used that $|\boldsymbol{a}_\tau| = |\widetilde{\boldsymbol{a}}| = m_\star + L_\star$ for every τ. Note that the right-hand side of (27) is expressed in terms of T, which is not the correct horizon, $T/(m_\star + L_\star)$. However, recall that we have

$$\begin{aligned}
m_\star + L_\star &\le 2\sqrt{\frac{m_\star}{d}}\, T^{1/4} \\
(m_\star + L_\star)^4 &\le \left(\frac{4m_\star}{d}\right)^2 T \\
T^3 &\le \left(\frac{4m_\star}{d}\right)^2 \left(\frac{T}{m_\star + L_\star}\right)^4 \\
T^{1/2} &\le \left(\frac{4m_\star}{d}\right)^{1/3} \left(\frac{T}{m_\star + L_\star}\right)^{2/3} ,
\end{aligned}$$

such that by substituting in (27) and identifying we have $\alpha_\star = 2/3$, and

$$\begin{aligned}
C_\star = 4\left(\frac{4m_\star}{d}\right)^{1/3} (m_\star+1)^{\gamma_\star^+} & \sqrt{d \ln\left(1 + \frac{T_{\mathrm{bc}}(m_\star + L_\star)(m_\star+1)^{2\gamma_\star^+}}{d\lambda}\right)} \\
& \left(\sqrt{\lambda} + \sqrt{\ln\left(\frac{1}{\delta}\right) + d\,\ln\left(1 + \frac{T_{\mathrm{bc}}(m_\star+1)^{2\gamma_\star^+}}{d\lambda}\right)}\,\right) .
\end{aligned}$$

Setting $\eta_j = T_{\mathrm{bc}}^{-2/3}$, and substituting in (26) with $R = (m_\star+1)^{\gamma_\star^+}$, we have that with high probability

$$\sum_{\tau=1}^{T_{\mathrm{bc}}} \frac{\widetilde{r}(\widetilde{\boldsymbol{a}})}{|\widetilde{\boldsymbol{a}}|} - \frac{\widetilde{r}(\boldsymbol{a}_\tau^{\mathrm{bc}})}{|\boldsymbol{a}_\tau^{\mathrm{bc}}|} = \widetilde{\mathcal{O}}\Big(\big(C_\star^{3/2} + N\big)\, T_{\mathrm{bc}}^{2/3} + \ (m_\star+1)^{2\gamma_\star^+}\, T_{\mathrm{bc}}^{1/3}\Big) .$$

Now, recall that $T_{\mathrm{bc}} = \mathcal{O}\big(\sqrt{d/m_\star}\, T^{3/4}\big)$, and that $C_\star = \widetilde{\mathcal{O}}\big((m_\star+1)^{\frac{1}{3}+\gamma_\star^+}\, d^{2/3}\big)$. Hence, $N \le d\sqrt{m_\star}$ implies $N = \mathcal{O}\big(C_j^{3/2}\big)$, and $(m_\star+1)^{\gamma_\star^+} \le d^2\sqrt{m_\star T}$ implies $(m_\star+1)^{\gamma_\star^+}\, T_{\mathrm{bc}}^{1/3} = \mathcal{O}\big(C_\star^{3/2}\, T_{\mathrm{bc}}^{2/3}\big)$. Setting $\lambda \in [1, d]$, $\delta = 1/T$, we obtain

$$\mathbb{E}\left[\sum_{\tau=1}^{T_{\mathrm{bc}}} \frac{\widetilde{r}(\widetilde{\boldsymbol{a}})}{|\widetilde{\boldsymbol{a}}|} - \frac{\widetilde{r}(\boldsymbol{a}_\tau^{\mathrm{bc}})}{|\boldsymbol{a}_\tau^{\mathrm{bc}}|}\right] = \widetilde{\mathcal{O}}\Big(d\,\sqrt{m_\star}\,(m_\star+1)^{\frac{3}{2}\gamma_\star^+}\, T_{\mathrm{bc}}^{2/3}\Big) . \tag{28}$$

Let m_τ be the memory size associated to the bandit played at block time step τ by Algorithm 2. Let $m_{\min} = \min_j m_j$ and $m_{\max} = \max_j m_j$. Finally, let $L_{\min}$ and $L_{\max}$ the (partial) block length associated with $m_{\min}$ and $m_{\max}$. We have

$$\sum_{t=1}^{T} r_t^{\mathrm{bc}} \ge \sum_{\tau=1}^{T_{\mathrm{bc}}} \Big(\widetilde{r}(\boldsymbol{a}_\tau^{\mathrm{bc}}) - m_\tau\,(m_\star+1)^{\gamma_\star^+}\Big) \ge \sum_{\tau=1}^{T_{\mathrm{bc}}} \widetilde{r}(\boldsymbol{a}_\tau^{\mathrm{bc}}) - m_{\max}\,(m_\star+1)^{\gamma_\star^+}\, T_{\mathrm{bc}} ,$$

such that by Lemma 1 and (28) we obtain

$$\begin{aligned}
\mathbb{E}&\left[\frac{\min_\tau |\boldsymbol{a}_\tau|}{\max_\tau |\boldsymbol{a}_\tau|}\left(\widetilde{r}(\widetilde{\boldsymbol{a}})\ \frac{\sum_\tau |\boldsymbol{a}_\tau|}{|\widetilde{\boldsymbol{a}}|}\right) - \sum_{t=1}^{T} r_t^{\mathrm{bc}}\right] \\
&\qquad \le m_{\max}\,(m_\star+1)^{\gamma_\star^+}\, T_{\mathrm{bc}} + \min_\tau |\boldsymbol{a}_\tau|\, \widetilde{\mathcal{O}}\Big(d\,\sqrt{m_\star}\,(m_\star+1)^{\frac{3}{2}\gamma_\star^+}\, T_{\mathrm{bc}}^{2/3}\Big) ,
\end{aligned}$$

$$\mathbb{E}\left[\frac{m_{\min} + L_{\min}}{m_{\max} + L_{\max}}\left(\frac{L_\star\,\mathrm{OPT}}{T}\ \frac{T}{m_\star + L_\star}\right) - \sum_{t=1}^{T} r_t^{\mathrm{bc}}\right]$$

$$\leq \frac{m_{\max}\,(m_\star+1)^{\gamma_\star^+}\,T}{m_{\min}+L_{\min}} + (m_{\min}+L_{\min})^{1/3}\,\widetilde{\mathcal{O}}\Big(d\,\sqrt{m_\star}\,(m_\star+1)^{\frac{3}{2}\gamma_\star^+}\,T^{2/3}\Big)\,,$$

$$\mathbb{E}\left[\sqrt{\frac{m_{\min}}{m_{\max}}}\;\mathrm{OPT}-\sum_{t=1}^{T} r_t^{\mathrm{bc}}\right] \leq \frac{m_{\max}}{m_{\min}}\sqrt{d\,m_\star}\,(m_\star+1)^{\gamma_\star^+}\,T^{3/4} + \widetilde{\mathcal{O}}\Big(d\,m_\star\,(m_\star+1)^{\frac{3}{2}\gamma_\star^+}\,T^{3/4}\Big)$$
$$= \frac{m_{\max}}{m_{\min}}\,\widetilde{\mathcal{O}}\Big(d\,m_\star\,(m_\star+1)^{\frac{3}{2}\gamma_\star^+}\,T^{3/4}\Big)\,,$$

where we have used the fact that $m_{\min}+L_{\min} = \sqrt{m_{\min}/d}\;T^{1/4}$, and $m_{\max}+L_{\max} = \sqrt{m_{\max}/d}\;T^{1/4}$. Corollary 1 is obtained by setting $M = m_{\max}/m_{\min}$. □

B Bandit Combiner

In this section we show our adaptation of the numbers C_j and target regrets R_j for the Bandit Combiner algorithm Algorithm 2 which builds on Cutkosky et al. (2020). For `O3M`(m_j, γ_j), $j = 1, \ldots, N$, the numbers C_j and target regrets R_j are defined as

$$C_j = 4\left(\frac{4m_j}{d}\right)^{1/3}(m_j+1)^{\gamma_j^+}\sqrt{d\ln\left(1+\frac{T_{\mathrm{bc}}(m_j+L_j)(m_j+1)^{2\gamma_j^+}}{d\lambda}\right)} \left(\sqrt{\lambda}+\sqrt{\ln\left(\frac{1}{\delta}\right)+d\,\ln\left(1+\frac{T_{\mathrm{bc}}(m_j+1)^{2\gamma_j^+}}{d\lambda}\right)}\,\right)\,, \tag{29}$$

$$R_j = C_j T_{\mathrm{bc}}^{\alpha_j} + \frac{(1-\alpha_j)^{\frac{1-\alpha_j}{\alpha_j}}(1+\alpha_j)^{\frac{1}{\alpha_j}}}{\alpha_j^{\frac{1-\alpha_j}{\alpha_j}}} C_j^{\frac{1}{\alpha_j}}\,T_{\mathrm{bc}}\eta_j^{\frac{1-\alpha_j}{\alpha_j}}$$
$$+ 1152(m_j+1)^{2\gamma_j^+}\log(T_{\mathrm{bc}}^3 N/\delta)T_{\mathrm{bc}}\eta_j + \sum_{k\neq j}\frac{1}{\eta_k}.$$

Note that the form of the target regret R_j slightly differs from the one presented in (Cutkosky et al., 2020, Corollary 2) due to the different range of the rewards. The algorithm, which is an adaptation of Bandit Combiner in Cutkosky et al. (2020), is summarized in Algorithm 2.

C Additional Experiments

We provide an additional experiment comparing the regrets of `O3M` and `OM-Block`. In order to be able to plot the regret, we must know OPT which is hard to compute in general. Since in the rising scenario with an isotropic initialization OPT is oracle greedy, which is easy to compute, we present this experiment in a rising setting with $m = 1$ and $\gamma = 2$. We plot the regret of `O3M` and `OM-Block` against the number of time steps, measuring the performance at different time horizons and for different sizes of L (where L depends on T, see at the end of Section 3.2). Specifically, we instantiated `O3M` and `OM-Block` for increasing values of L, setting the horizon of each instance based on the equations in Theorem 1 and Proposition 4. Figure 3 shows how the dimension of $\widehat{\theta}$, which is d for `O3M` and $d \times L$ for `OM-Block`, has an actual impact on the performance since `O3M` outperforms `OM-Block`. The code is written in Python and it is publicly available at the following GitHub repository: Linear Bandits with Memory.

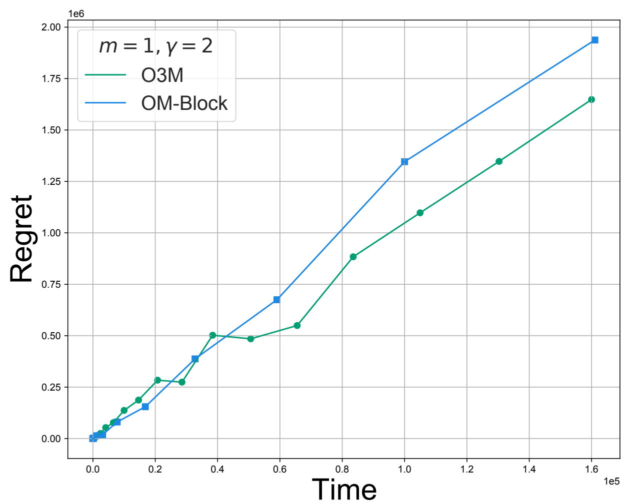

Figure 3: The regret of `O3M` and `OM-Block`. Each dot is a separate run where the value of L is tuned to the corresponding horizon.