Organization of the Appendix

The appendix is organized as follows:

- Appendix A contains the proofs of the results from Section 2.
- Appendix B contains the proofs of existing results about FTRL for completeness.
- Appendix C contains the proofs of upper bounds on regret from Section 3.
- Appendix D contains the proofs of lower bounds on regret from Section 3.
- Appendix E contains the proofs of results about online linear control from Section 4.1.
- Appendix F contains the proofs of results about online performative prediction from Section 4.2.
- Appendix G discusses how to implement Algorithm 1 efficiently.
- Appendix H presents an algorithm for OCO with unbounded memory that provides the same upper bound on policy regret as Algorithm 1 while guaranteeting a small number of switches (Algorithm 2).
- Appendix I presents simulation experiments.

A Framework

In this section prove Theorem 2.1. But first we prove a lemma that we use for proofs involving linear sequence dynamics with the ξ-weighted p-norm (Definition 2.3). Recall that $\|\cdot\|_{\mathcal{U}}$ denotes the norm associated with a space $\mathcal{U}$ and the operator norm $\|L\|$ for a linear operator $L : \mathcal{U} \to \mathcal{V}$ is defined as $\|L\| = \max_{u:\|u\|_{\mathcal{U}} \le 1} \|Lu\|_{\mathcal{V}}$.

Lemma A.1. *Consider an online convex optimization with unbounded memory problem specified by $(\mathcal{X}, \mathcal{H}, A, B)$. If $(\mathcal{X}, \mathcal{H}, A, B)$ follows linear sequence dynamics with the ξ-weighted p-norm for $p \ge 1$, then for all $k \ge 1$*

$$\xi_k \|A_{k-1} \cdots A_0\| \le \|A^k\|.$$

Proof. Let $x \in \mathcal{X}$ with $\|x\|_{\mathcal{X}} = 1$. We have

$$\xi_k \|A_{k-1} \cdots A_0 x\|_{\mathcal{X}} = \|A^k(x, 0, \dots)\|_{\mathcal{H}} \le \|A^k\| \|(x, 0, \dots)\|_{\mathcal{H}} \le \|A^k\|,$$

where the last inequality follows because $\|(x, 0, \dots)\|_{\mathcal{H}} = \xi_0 \|x\|_{\mathcal{X}}$ and $\xi_0 = 1$ by Definition 2.3. Therefore, $\|A_{k-1} \cdots A_0\| \le \|A^k\|$. ■

Theorem 2.1. *Consider an online convex optimization with unbounded memory problem specified by $(\mathcal{X}, \mathcal{H}, A, B)$. If f_t is L-Lipschitz continuous, then $\tilde{f}_t$ is $\tilde{L}$-Lipschitz continuous for $\tilde{L} \le L \sum_{k=0}^{\infty} \|A^k\|$. If $(\mathcal{X}, \mathcal{H}, A, B)$ follows linear sequence dynamics with the ξ-weighted p-norm for $p \ge 1$, then $\tilde{L} \le L \left(\sum_{k=0}^{\infty} \|A^k\|^p\right)^{\frac{1}{p}}$.*

Proof. Let $x, \tilde{x} \in \mathcal{X}$. For the general case, we have

$$\begin{aligned}
\left|\tilde{f}_t(x) - \tilde{f}_t(\tilde{x})\right| &= \left|f_t\left(\sum_{k=0}^{t-1} A^k B x\right) - f_t\left(\sum_{k=0}^{t-1} A^k B \tilde{x}\right)\right| && \text{by Definition 2.1} \\
&\le L \left\|\sum_{k=0}^{t-1} A^k B (x - \tilde{x})\right\|_{\mathcal{H}} && f_t \text{ is } L\text{-Lipschitz continuous} \\
&\le L \sum_{k=0}^{t-1} \|A^k\| \|B\| \|x - \tilde{x}\|_{\mathcal{X}} \\
&\le L \sum_{k=0}^{t-1} \|A^k\| \|x - \tilde{x}\|_{\mathcal{X}} && \text{by Assumption A2} \\
&\le L \sum_{k=0}^{\infty} \|A^k\| \|x - \tilde{x}\|_{\mathcal{X}}.
\end{aligned}$$

If $(\mathcal{H}, \mathcal{X}, A, B)$ follows linear sequence dynamics with the ξ-weighted p-norm for $p \geq 1$, then we have

$$\begin{aligned}
\left|\tilde{f}_t(x) - \tilde{f}_t(\tilde{x})\right| &= \left| f_t\left(\sum_{k=0}^{t-1} A^k B x\right) - f_t\left(\sum_{k=0}^{t-1} A^k B \tilde{x}\right)\right| && \text{by Definition 2.1} \\
&\leq L \left\|\sum_{k=0}^{t-1} A^k B(x - \tilde{x})\right\|_{\mathcal{H}} && f_t \text{ is } L\text{-Lipschitz continuous} \\
&= L \left\|(0, A_0(x - \tilde{x}), A_1 A_0(x - \tilde{x}), \dots)\right\| && \text{by Definition 2.3} \\
&= L \left(\sum_{k=0}^{t-1} \xi_k^p \left\|A_{k-1} \cdots A_0(x - \tilde{x})\right\|^p\right)^{\frac{1}{p}} && \text{by Definition 2.3} \\
&\leq L \left(\sum_{k=0}^{t-1} \|A^k\|^p\right)^{\frac{1}{p}} \|x - \tilde{x}\|_{\mathcal{X}} && \text{by Lemma A.1} \\
&\leq L \left(\sum_{k=0}^{\infty} \|A^k\|^p\right)^{\frac{1}{p}} \|x - \tilde{x}\|_{\mathcal{X}}\,. && \blacksquare
\end{aligned}$$

B Standard Analysis of Follow-the-Regularized-Leader

In this section we state and prove some existing results about the follow-the-regularized-leader (FTRL) algorithm [Shalev-Shwartz and Singer, 2006, Abernethy et al., 2008]. These results are well known in the literature, but we prove them here for completeness and use them in the remainder of the paper. We use the below results for functions $\tilde{f}_t$ with Lipschitz constants $\tilde{L}$. However, in this section we use a more general notation, denoting functions by g_t and their Lipschitz constant by L_g.

Consider the following setup for an online convex optimization (OCO) problem. Let T denote the time horizon. Let the decision space $\mathcal{X}$ be a closed, convex subset of a Hilbert space and $g_t : \mathcal{X} \to \mathbb{R}$ be loss functions chosen by an oblivious adversary. The functions g_t are convex and L_g-Lipschitz continuous. The game between the learner and the adversary proceeds as follows. In each round $t \in [T]$, the learner chooses $x_t \in \mathcal{X}$ and the learner suffers loss $g_t(x_t)$. The goal of the learner is to minimize (static) regret,

$$R_T^{\text{static}} = \sum_{t=1}^{T} g_t(x_t) - \min_{x \in \mathcal{X}} \sum_{t=1}^{T} g_t(x). \tag{3}$$

Let $R : \mathcal{X} \to \mathbb{R}$ be an α-strongly convex regularizer satisfying $|R(x) - R(\tilde{x})| \leq D$ for all $x, \tilde{x} \in \mathcal{X}$. The FTRL algorithm chooses iterates x_t as

$$x_t \in \arg\min_{x \in \mathcal{X}} \sum_{s=1}^{t-1} g_s(x) + \frac{R(x)}{\eta}, \tag{4}$$

where η is a tunable parameter referred to as the step-size. In what follows, let $g_0 = \frac{R}{\eta}$. The analysis in this section closely follows Karlin [2017].

Lemma B.1. *For all $x \in \mathcal{X}$, FTRL (Eq. (4)) satisfies*

$$\sum_{t=0}^{T} g_t(x) \geq \sum_{t=0}^{T} g_t(x_{t+1}).$$

Proof. We use proof by induction on T. The base case is $T = 0$. By definition, $x_1 \in \arg\min_{x \in \mathcal{X}} R(x)$. Therefore, $R(x) \geq R(x_1)$ for all $x \in \mathcal{X}$. Recalling the notation $g_0 = \frac{R}{\eta}$ proves the base case. Now, assume that the lemma is true for $T - 1$. That is,

$$\sum_{t=0}^{T-1} g_t(x) \geq \sum_{t=0}^{T-1} g_t(x_{t+1}).$$

Let $x \in \mathcal{X}$ be arbitrary. Since $x_{T+1} \in \arg\min_{x\in\mathcal{X}} \sum_{t=0}^{T} g_t(x)$, we have

$$\begin{aligned}
\sum_{t=0}^{T} g_t(x) &\geq \sum_{t=0}^{T} g_t(x_{T+1}) \\
&= \sum_{t=0}^{T-1} g_t(x_{T+1}) + g_T(x_{T+1}) \\
&\geq \sum_{t=0}^{T-1} g_t(x_{t+1}) + g_T(x_{T+1}) && \text{by inductive hypothesis} \\
&= \sum_{t=0}^{T} g_t(x_{t+1}).
\end{aligned}$$

This completes the proof. ■

Lemma B.2. *For all $x \in \mathcal{X}$, FTRL (Eq. (4)) satisfies*

$$\sum_{t=1}^{T} g_t(x_t) - \sum_{t=1}^{T} g_t(x) \leq \frac{D}{\eta} + \sum_{t=1}^{T} g_t(x_t) - g_t(x_{t+1}).$$

Proof. Note that

$$\sum_{t=1}^{T} g_t(x_t) - \sum_{t=1}^{T} g_t(x) \leq \sum_{t=1}^{T} g_t(x_t) - \sum_{t=1}^{T} g_t(x) + g_0(x) - g_0(x_1)$$

because $x_1 \in \arg\min_{x\in\mathcal{X}} g_0(x)$. The proof of this lemma now follows by using the above inequality, Lemma B.1, the definition $g_0 = \frac{R}{\eta}$, and the definition of D. ■

Theorem B.1. *FTRL (Eq. (4)) satisfies*

$$\|x_{t+1} - x_t\|_{\mathcal{X}} \leq \eta \frac{L_g}{\alpha} \quad \text{and} \quad R_T^{\text{static}} \leq \frac{D}{\eta} + \eta \frac{T L_g^2}{\alpha}.$$

Choosing $\eta = \sqrt{\frac{\alpha D}{T L_g^2}}$ yields

$$R_T^{\text{static}} \leq O\left(\sqrt{\frac{D}{\alpha} T L_g^2}\right).$$

Proof. Let $x^* \in \arg\min_{x\in\mathcal{X}} \sum_{t=1}^{T} g_t(x)$. Using Lemma B.2 we have

$$\sum_{t=1}^{T} g_t(x_t) - \sum_{t=1}^{T} g_t(x^*) \leq \frac{D}{\eta} + \sum_{t=1}^{T} g_t(x_t) - g_t(x_{t+1}). \tag{5}$$

We can bound the summands in the sum above as follows. Define $G_t(x) = \sum_{s=0}^{t-1} g_s(x)$. Then, $x_t \in \arg\min_{x\in\mathcal{X}} G_t(x)$. and $x_{t+1} \in \arg\min_{x\in\mathcal{X}} G_{t+1}(x)$. Since $\{g_s\}_{s=1}^{T}$ are convex, R is α-strongly-convex, and $g_0 = \frac{R}{\eta}$, we have that G_t is $\frac{\alpha}{\eta}$-strongly-convex. So,

$$\begin{aligned}
G_t(x_{t+1}) &\geq G_t(x_t) + \frac{\alpha}{2\eta}\|x_{t+1} - x_t\|_{\mathcal{X}}^2, \\
G_{t+1}(x_t) &\geq G_{t+1}(x_{t+1}) + \frac{\alpha}{2\eta}\|x_{t+1} - x_t\|_{\mathcal{X}}^2.
\end{aligned}$$

Adding the above two inequalities yields

$$g_t(x_t) - g_t(x_{t+1}) \geq \frac{\alpha}{\eta}\|x_{t+1} - x_t\|_{\mathcal{X}}^2. \tag{6}$$

Since g_t are convex and L_g-Lipschitz continuous, we also have

$$g_t(x_t) - g_t(x_{t+1}) \leq L_g \|x_{t+1} - x_t\|_{\mathcal{X}}. \tag{7}$$

Combining Eqs. (6) and (7) we have

$$\|x_{t+1} - x_t\|_{\mathcal{X}} \leq \eta \frac{L_g}{\alpha}.$$

This proves the first part of the theorem. Now, using this in Eq. (7) we have

$$g_t(x_t) - g_t(x_{t+1}) \leq \eta \frac{L_g^2}{\alpha}. \tag{8}$$

Finally, substituting this in Eq. (5) proves the second part of the theorem. ■

C Regret Analysis: Upper Bounds

First we prove a lemma that bounds the difference in the value of f_t evaluated at the actual history h_t and an idealized history that would have been obtained by playing x_t in all prior rounds.

Lemma C.1. *Consider an online convex optimization with unbounded memory problem specified by $(\mathcal{X}, \mathcal{H}, A, B)$. If the decisions (x_t) are generated by Algorithm 1, then*

$$\left| f_t(h_t) - \tilde{f}_t(x_t) \right| \leq \eta \frac{L\tilde{L}H_1}{\alpha}$$

for all rounds t. When $(\mathcal{X}, \mathcal{H}, A, B)$ follows linear sequence dynamics with the ξ-weighted p-norm for $p \geq 1$, then

$$\left| f_t(h_t) - \tilde{f}_t(x_t) \right| \leq \eta \frac{L\tilde{L}H_p}{\alpha}$$

for all rounds t.

Proof. We have

$$\begin{aligned}
\left| f_t(h_t) - \tilde{f}_t(x_t) \right| &= \left| f_t(h_t) - f_t\left(\sum_{k=0}^{t-1} A^k B x_t \right) \right| && \text{by Definition 2.1} \\
&\leq L \left\| h_t - \sum_{k=0}^{t-1} A^k B x_t \right\| && \text{by Assumption A4} \\
&= L \left\| \sum_{k=0}^{t-1} A^k B x_{t-k} - \sum_{k=0}^{t-1} A^k B x_t \right\| && \text{by definition of } h_t \\
&= L \underbrace{\left\| \sum_{k=0}^{t-1} A^k B (x_{t-k} - x_t) \right\|}_{(a)}. && (9)
\end{aligned}$$

First consider the general case where $(\mathcal{X}, \mathcal{H}, A, B)$ does not necessarily follow linear sequence dynamics. We can bound the term (a) as

$$\begin{aligned}
\left\| \sum_{k=0}^{t-1} A^k B (x_{t-k} - x_t) \right\| &\leq \sum_{k=0}^{t-1} \left\| A^k B \right\| \|x_t - x_{t-k}\| \\
&\leq \sum_{k=0}^{t-1} \left\| A^k B \right\| k\eta \frac{\tilde{L}}{\alpha} && \text{by Theorem B.1} \\
&\leq \sum_{k=0}^{t-1} \left\| A^k \right\| k\eta \frac{\tilde{L}}{\alpha} && \text{by Assumption A2} \\
&\leq \eta \frac{\tilde{L}}{\alpha} H_1.
\end{aligned}$$

Plugging this into Eq. (9) completes the proof for the general case. Now consider the case when $(\mathcal{X}, \mathcal{H}, A, B)$ follows linear sequence dynamcis with the ξ-weighted p-norm. We can bound the term (a) as

$$\begin{aligned}
\left\|\sum_{k=0}^{t-1} A^k B(x_{t-k} - x_t)\right\| &= \|(0, A_0(x_t - x_{t-1}), A_1 A_0(x_t - x_{t-2}), \dots)\| && \text{by Definition 2.3} \\
&= \left(\sum_{k=0}^{t-1} \xi_k^p \left\|A_{k-1} \cdots A_0 (x_t - x_{t-k})\right\|^p\right)^{\frac{1}{p}} && \text{by Definition 2.3} \\
&\le \left(\sum_{k=0}^{t-1} \xi_k^p \left\|A_{k-1} \cdots A_0\right\|^p \left\|x_t - x_{t-k}\right\|^p\right)^{\frac{1}{p}} \\
&\le \left(\sum_{k=0}^{t-1} \left\|A^k\right\|^p \left\|x_t - x_{t-k}\right\|^p\right)^{\frac{1}{p}} && \text{by Lemma A.1} \\
&\le \eta \frac{\tilde{L}}{\alpha} \left(\sum_{k=0}^{t-1} \left\|A^k\right\|^p k^p\right)^{\frac{1}{p}} && \text{by Theorem B.1} \\
&\le \eta \frac{\tilde{L}}{\alpha} H_p.
\end{aligned}$$

Plugging this into Eq. (9) completes the proof. ■

Now we restate and prove Theorem 3.1

Theorem 3.1. *Consider an online convex optimization with unbounded memory problem specified by $(\mathcal{X}, \mathcal{H}, A, B)$. Let the regularizer $R : \mathcal{X} \to \mathbb{R}$ be α-strongly-convex and satisfy $|R(x) - R(\tilde{x})| \le D$ for all $x, \tilde{x} \in \mathcal{X}$. Algorithm 1 with step-size η satisfies $R_T(\texttt{FTRL}) \le \frac{D}{\eta} + \eta \frac{T\tilde{L}^2}{\alpha} + \eta \frac{T L \tilde{L} H_1}{\alpha}$. If $\eta = \sqrt{\frac{\alpha D}{T\tilde{L}(LH_1 + \tilde{L})}}$, then*

$$R_T(\texttt{FTRL}) \le O\left(\sqrt{\frac{D}{\alpha} T L \tilde{L} H_1}\right).$$

When $(\mathcal{X}, \mathcal{H}, A, B)$ follows linear sequence dynamics with the ξ-weighted p-norm, then all of the above hold with H_p instead of H_1.

Proof. First consider the general case where $(\mathcal{X}, \mathcal{H}, A, B)$ does not necessarily follow linear sequence dynamics. Let $x^* \in \arg\min_{x \in \mathcal{X}} \sum_{t=1}^T \tilde{f}_t(x)$. Note that we can write the regret as

$$\begin{aligned}
R_T(\texttt{FTRL}) &= \sum_{t=1}^T f_t(h_t) - \min_{x \in \mathcal{X}} \sum_{t=1}^T \tilde{f}_t(x) \\
&= \underbrace{\sum_{t=1}^T f_t(h_t) - \tilde{f}_t(x_t)}_{(a)} + \underbrace{\sum_{t=1}^T \tilde{f}_t(x_t) - \tilde{f}_t(x^*)}_{(b)}.
\end{aligned}$$

We can bound term (a) using Lemma C.1 and term (b) using Theorem B.1. Therefore, we have

$$\begin{aligned}
R_T(\texttt{FTRL}) &= \underbrace{\sum_{t=1}^T f_t(h_t) - \tilde{f}_t(x_t)}_{(a)} + \underbrace{\sum_{t=1}^T \tilde{f}_t(x_t) - \tilde{f}_t(x^*)}_{(b)} \\
&\le \eta \frac{T L \tilde{L} H_1}{\alpha} + \frac{D}{\eta} + \eta \frac{T\tilde{L}^2}{\alpha}.
\end{aligned}$$

Choosing $\eta = \sqrt{\frac{\alpha D}{T\tilde{L}(LH_1+\tilde{L})}}$ yields

$$R_T(\texttt{FTRL}) \leq O\left(\sqrt{\frac{D}{\alpha}TL\tilde{L}H_1}\right),$$

where we used the definition of p-effective memory capacity (Definition 2.4) and the bound on $\tilde{L}$ (Theorem 2.1) to simplify the above expression. This completes the proof for the general case. The proof for when $(\mathcal{X}, \mathcal{H}, A, B)$ follows linear sequence dynamcis with the ξ-weighted p-norm is the same as above, except we bound the term (a) above using Lemma C.1 for linear sequence dynamics. ■

Now we restate and prove Theorem 3.3.

Theorem 3.3. *Consider an online convex optimization with finite memory problem with constant memory length m specified by $(\mathcal{X}, \mathcal{H} = \mathcal{X}^m, A_{\text{finite},m}, B_{\text{finite},m})$. Let the regularizer $R : \mathcal{X} \to \mathbb{R}$ be α-strongly-convex and satisfy $|R(x) - R(\tilde{x})| \leq D$ for all $x, \tilde{x} \in \mathcal{X}$. Algorithm 1 with step-size $\eta = \sqrt{\frac{\alpha D}{T\tilde{L}(Lm^{\frac{3}{2}}+\tilde{L})}}$ satisfies*

$$R_T(\texttt{FTRL}) \leq O\left(\sqrt{\frac{D}{\alpha}TL\tilde{L}m^{\frac{3}{2}}}\right) \leq O\left(m\sqrt{\frac{D}{\alpha}TL^2}\right).$$

The OCO with finite memory problem, as defined in the literature, follows linear sequence dynamics with the 2-norm. In this subsection we consider a more general version of the OCO with finite memory problem that follows linear sequence dynamics with the p-norm. We provide an upper bound on the policy regret for this more general formulation and the proof of Theorem 3.3 follows as a special case when $p = 2$.

Theorem C.1. *Consider an online convex optimization with finite memory problem with constant memory length m, $(\mathcal{X}, \mathcal{H} = \mathcal{X}^m, A_{\text{finite},m}, B_{\text{finite},m})$. Assume that the problem follows linear sequence dynamics with the p-norm for $p \geq 1$. Let the regularizer $R : \mathcal{X} \to \mathbb{R}$ be α-strongly-convex and satisfy $|R(x) - R(\tilde{x})| \leq D$ for all $x, \tilde{x} \in \mathcal{X}$. Algorithm 1 with step-size η satisfies*

$$R_T(\texttt{FTRL}) \leq O\left(\sqrt{\frac{D}{\alpha}TL\tilde{L}m^{\frac{p+1}{p}}}\right) \leq O\left(\sqrt{\frac{D}{\alpha}TL^2m^{\frac{p+2}{p}}}\right).$$

Proof. Using Theorem 3.1 it suffices to bound $\tilde{L}$ and H_p for this problem. Note that $\|A^k_{\text{finite}}\| = 1$ if $k \leq m$ and 0 otherwise. Using this we have

$$H_p = \left(\sum_{k=0}^{\infty}\left(k\|A^k_{\text{finite}}\|\right)^p\right)^{\frac{1}{p}} = \left(\sum_{k=0}^{m}k^p\right)^{\frac{1}{p}} \leq O\left(m^{\frac{p+1}{p}}\right).$$

This proves the first inequality in the statement of the theorem. The second inequality follows from the above and Theorem 2.1, which states that

$$\tilde{L} \leq L\left(\sum_{k=0}^{\infty}\|A^k_{\text{finite}}\|^p\right)^{\frac{1}{p}} = Lm^{\frac{1}{p}}.$$ ■

Finally, we provide an upper bound on the policy regret for the OCO with ρ-discounted infinite memory problem. For simplicity, we consider the case when the problem follows linear sequence dynamics with the 2-norm instead of a general p-norm.

Theorem C.2. *Consider an online convex optimization with ρ-discounted infinite memory problem $(\mathcal{X}, \mathcal{H}, A_{\text{infinite},\rho}, B_{\text{infinite}})$. Suppose that the problem follows linear sequence dynamics with the 2-norm. Let the regularizer $R : \mathcal{X} \to \mathbb{R}$ be α-strongly-convex and satisfy $|R(x) - R(\tilde{x})| \leq D$ for all $x, \tilde{x} \in \mathcal{X}$. Algorithm 1 with step-size η satisfies*

$$R_T(\texttt{FTRL}) \leq O\left(\sqrt{\frac{D}{\alpha}TL\tilde{L}(1-\rho^2)^{-\frac{3}{2}}}\right) \leq O\left(\sqrt{\frac{D}{\alpha}TL^2(1-\rho^2)^{-2}}\right) \leq O\left(\sqrt{\frac{D}{\alpha}TL^2(1-\rho)^{-2}}\right).$$

Proof. Using Theorem 3.1, it suffices to bound $\tilde{L}$ and H_p for this problem. Recall that $\|A^k_{\text{infinite},\rho}\| = \rho^k$. Using this we have

$$H_2 = \left(\sum_{k=0}^{\infty} \left(k\|A^k_{\text{finite}}\|\right)^2\right)^{\frac{1}{2}} = \left(\sum_{k=0}^{\infty} \left(k\rho^k\right)^2\right)^{\frac{1}{2}} \leq (1-\rho^2)^{-\frac{3}{2}}.$$

This proves the first inequality in the statement of the theorem. The second inequality follows from the above and Theorem 2.1, which states that

$$\tilde{L} \leq L\left(\sum_{k=0}^{\infty} \|A^k_{\text{infinite},\rho}\|^2\right)^{\frac{1}{2}} = L(1-\rho^2)^{-\frac{1}{2}}.$$

The last inequality follows because $1-\rho^2 = (1+\rho)(1-\rho)$, which implies that $1-\rho \leq 1-\rho^2 \leq 2(1-\rho)$ because $\rho \in (0,1)$. ■

C.1 Existing Regret Bound for OCO with Finite Memory

In this subsection we provide a detailed comparison of our upper bound on the policy regret for OCO with finite memory with that of Anava et al. [2015]. The material in this subsection comes from Appendix A.2 of their arXiv version or Appendix C.2 of their conference version.

The existing upper bound on regret is

$$O\left(\sqrt{DT\lambda m^{\frac{3}{2}}}\right),$$

where $D = \max_{x,\tilde{x}\in\mathcal{X}} |R(x) - R(\tilde{x})|$. Although the parameter λ is defined in terms of dual norms of the gradient of $\tilde{f}_t$, it is essentially the Lipschitz-continuity constant for $\tilde{f}_t$: for all $x, \tilde{x} \in \mathcal{X}$,

$$\left|\tilde{f}_t(x) - \tilde{f}_t(\tilde{x})\right| \leq \sqrt{\lambda\alpha}\|x - \tilde{x}\|,$$

where α is the strong-convexity parameter of the regularizer R (or σ in the notation of Anava et al. [2015]). Therefore, the existing regret bound can be rewritten as

$$O\left(\tilde{L}\sqrt{\frac{D}{\alpha}Tm^{\frac{3}{2}}}\right).$$

Our upper bound on the policy regret for OCO with finite memory Theorem 3.3 is

$$O\left(\sqrt{\frac{D}{\alpha}L\tilde{L}Tm^{\frac{3}{2}}}\right).$$

Since $\tilde{L} \leq \sqrt{m}L$ by Theorem 2.1, this leads to an improvement by a factor of $m^{\frac{1}{4}}$.

D Regret Analysis: Lower Bounds

We first restate Theorems 3.2 and 3.4.

Theorem 3.2. *There exists an instance of the online convex optimization with unbounded memory problem, $(\mathcal{X}, \mathcal{H}, A, B)$, that follows linear sequence dynamics with the ξ-weighted p-norm and there exist L-Lipschitz continuous loss functions $\{f_t : \mathcal{H} \to \mathbb{R}\}_{t=1}^T$ such that the regret of any algorithm $\mathcal{A}$ satisfies*

$$R_T(\mathcal{A}) \geq \Omega\left(\sqrt{TL\tilde{L}H_p}\right).$$

Theorem 3.4. *There exists an instance of the online convex optimization with finite memory problem with constant memory length m, $(\mathcal{X}, \mathcal{H} = \mathcal{X}^m, A_{\textit{finite},m}, B_{\textit{finite},m})$, and there exist L-Lipschitz continuous loss functions $\{f_t : \mathcal{H} \to \mathbb{R}\}_{t=1}^T$ such that the regret of any algorithm $\mathcal{A}$ satisfies*

$$R_T(\mathcal{A}) \geq \Omega\left(m\sqrt{TL^2}\right).$$

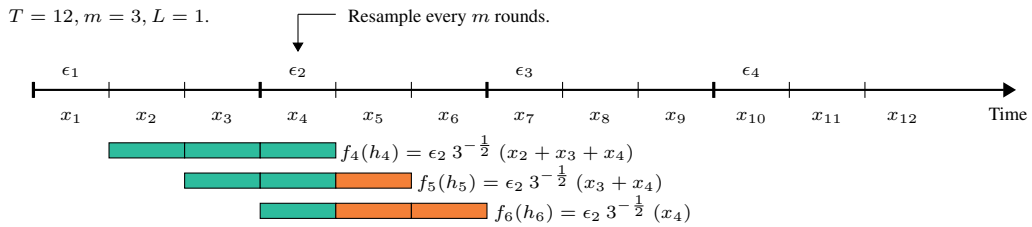


Figure 2: An illustration of the loss functions f_t for the OCO with finite memory lower bound. Suppose $T = 12, m = 3, L = 1$, and $p = 2$. Time is divided into blocks of size $m = 3$. Consider round $t = 5$. The history is $h_5 = (x_3, x_4, x_5)$. The loss function $f_5(h_5)$ is a product of three terms: a random sign ϵ_2 sampled for the block that round 5 belongs to, namely, block 2; a scaling factor of $m^{-\frac{1}{2}}$; a sum over the decisions in the history excluding those that were chosen after observing ϵ_2, i.e., a sum over x_3 and x_4, excluding x_5.

Theorem 3.2 follows from Theorem 3.4. However, the lower bound is true for a much broader class of problems as we show in this section. We first provide a lower bound for a more general formulation of the OCO with finite memory problem (Theorem D.1). The proof of Theorem 3.4 follows as a special case when $p = 2$. Then, we provide a lower bound for the OCO with ρ-discounted infinite memory problem (Theorem D.2).

The OCO with finite memory problem, as defined in the literature, follows linear sequence dynamics with the 2-norm. In this section we consider a more general version of the OCO with finite memory problem that follows linear sequence dynamics with the p-norm. We provide a lower bound on the policy regret for this more general formulation and the proof of Theorem 3.4 follows as a special case when $p = 2$.

Theorem D.1. *For all $p \geq 1$, there exists an instance of the online convex optimization with finite memory problem with constant memory length m, $(\mathcal{X}, \mathcal{H} = \mathcal{X}^m, A_{\text{finite},m}, B_{\text{finite},m})$, that follows linear sequence dynamics with the p-norm, and there exist L-Lipschitz continuous loss functions $\{f_t : \mathcal{H} \to \mathbb{R}\}_{t=1}^T$ such that the regret of any algorithm $\mathcal{A}$ satisfies*

$$R_T(\mathcal{A}) \geq \Omega\left(\sqrt{TL^2 m^{\frac{p+2}{p}}}\right).$$

Proof. Let $\mathcal{X} = [-1, 1]$ and consider an OCO with finite memory problem with constant memory length m, $(\mathcal{X}, \mathcal{H} = \mathcal{X}^m, A_{\text{finite},m}, B_{\text{finite},m})$, that follows linear sequence dynamics with the p-norm. For simplicity, assume that T is a multiple of m (otherwise, the same proof works but with slightly more tedious bookkeeping) and that $L = 1$ (otherwise, multiply the functions f_t defined below by L).

Divide the T rounds into $N = \frac{T}{m}$ blocks of m rounds each. Sample N independent Rademacher random variables $\{\epsilon_1, \ldots, \epsilon_N\}$, where each ϵ_i is equal to ± 1 with probability $\frac{1}{2}$. Recall that $h_t = (x_t, \ldots, x_{t-m+1})$. Define the loss functions $\{f_t\}_{t=1}^T$ as follows. (See Fig. 2 for an illustration.) If $t \leq m$, let $f_t = 0$. Otherwise, let

$$\begin{aligned} f_t(h_t) &= \epsilon_{\lceil \frac{t}{m} \rceil} m^{\frac{1-p}{p}} \sum_{k=0}^{m-1-(t-m\lfloor \frac{t}{m} \rfloor - 1)} x_{m\lfloor \frac{t}{m} \rfloor + 1 - k} \\ &= \epsilon_{\lceil \frac{t}{m} \rceil} m^{\frac{1-p}{p}} \left(x_{t-m+1} + \cdots + x_{m\lfloor \frac{t}{m} \rfloor + 1} \right). \end{aligned}$$

In words, the loss in the first m rounds is equal to 0. Thereafter, in round t the loss is equal to a random sign $\epsilon_{\lceil \frac{t}{m} \rceil}$, which is *fixed for that block*, times a scaling factor, which is chosen according to the p-norm to ensure that the Lipschitz constant L is at most 1, times a sum of a *subset* of past decisions in the history $h_t = (x_t, \ldots, x_{t-m+1})$. This subset consists of all past decisions until and including the first decision of the current block, which is the decision in round $m\lfloor \frac{t}{m} \rfloor + 1$.

The functions f_t are linear, so they are convex. In order to show that they satisfy Assumptions **A3** and **A4**, it remains to show that they are 1-Lipschitz continuous. Let $h = (x^{(1)}, \ldots, x^{(m)})$ and

$\tilde{h} = (\tilde{x}^{(1)}, \ldots, \tilde{x}^{(m)})$ be arbitrary elements of $\mathcal{H} = \mathcal{X}^m$. We have

$$\begin{aligned}
&\left|f_t(h) - f_t(\tilde{h})\right| \\
&\leq \left|\epsilon_{\lceil \frac{t}{m} \rceil} m^{\frac{1-p}{p}} \left((x^{(1)} - \tilde{x}^{(1)}) + \cdots + (x^{(m)} - \tilde{x}^{(m)})\right)\right| \\
&\leq m^{\frac{1-p}{p}} \left|(x^{(1)} - \tilde{x}^{(1)}) + \cdots + (x^{(m)} - \tilde{x}^{(m)})\right| && \text{because } \epsilon_{\lceil \frac{t}{m} \rceil} \in \{-1, +1\} \\
&\leq m^{\frac{1-p}{p}} m^{1-\frac{1}{p}} \left(\sum_{k=1}^{m} \left|x^{(k)} - \tilde{x}^{(k)}\right|^p\right)^{\frac{1}{p}} && \text{by Hölder's inequality} \\
&= \|h - \tilde{h}\|_{\mathcal{H}},
\end{aligned}$$

where the last equality follows because of our assumption that the problem that follows linear sequence dynamics with the p-norm.

First we will show that the total expected loss of any algorithm is 0, where the expectation is with respect to the randomness in the choice of $\{\epsilon_1, \ldots, \epsilon_N\}$. The total loss in the first block is 0 because $f_t = 0$ for $t \in [m]$. For each subsequent block $n \in \{2, \ldots, N\}$, the total loss in block n depends on the algorithm's choices made *before* observing ϵ_n, namely, $\{x_{(n-2)m+2}, \ldots, x_{(n-1)m+1}\}$. Since ϵ_n is equal to ± 1 with probability $\frac{1}{2}$, the expected loss of any algorithm in a block is equal to 0 and the total expected loss is also equal to 0.

Now we will show that the expected loss of the benchmark is at most

$$-O\left(\sqrt{T m^{\frac{p+2}{p}}}\right),$$

where the expectation is with respect to the randomness in the choice of $\{\epsilon_1, \ldots, \epsilon_N\}$. We have

$$\begin{aligned}
\mathbb{E}\left[\min_{x \in \mathcal{X}} \sum_{t=1}^{T} \tilde{f}_t(x)\right] &= \mathbb{E}\left[\min_{x \in \mathcal{X}} \sum_{n=2}^{N} \sum_{t=(n-1)m+1}^{nm} \tilde{f}_t(x)\right] \\
&= \mathbb{E}\left[\min_{x \in \mathcal{X}} \sum_{n=2}^{N} \sum_{t=(n-1)m+1}^{nm} \epsilon_n m^{\frac{1-p}{p}} \times x \times (m - (t - (n-1)m - 1))\right].
\end{aligned}$$

The first equality follows from first summing over blocks and then summing over the rounds in that block. The second equality follows from the definitions of f_t above and of $\tilde{f}_t$ (Definition 2.1). By the defintion of $\tilde{f}_t$, the history h_t consists of m copies of x for $t \geq m$.. By the definition of f_t, which sums over all past decisions until the first round of the current block, we have that within a block the sum first extends over m copies of x (in the first round of the block), then $m - 1$ copies of x (in the second round of the block), and so on until the last round of the block. So, we have

$$\begin{aligned}
\mathbb{E}\left[\min_{x \in \mathcal{X}} \sum_{t=1}^{T} \tilde{f}_t(x)\right] &= \mathbb{E}\left[\min_{x \in \mathcal{X}} \sum_{n=2}^{N} \sum_{t=(n-1)m+1}^{nm} \epsilon_n m^{\frac{1-p}{p}} \times x \times (m - (t - (n-1)m - 1))\right] \\
&= \mathbb{E}\left[\min_{x \in \mathcal{X}} \sum_{n=2}^{N} \sum_{k=0}^{m-1} \epsilon_n m^{\frac{1-p}{p}} \times x \times (m - k)\right] \\
&= m^{\frac{1-p}{p}} \frac{m^2 + m}{2} \mathbb{E}\left[\min_{x \in \mathcal{X}} \sum_{n=2}^{N} \epsilon_n x\right] \\
&= m^{\frac{1-p}{p}} \frac{m^2 + m}{2} \mathbb{E}\left[\min_{x \in \{-1,1\}} \sum_{n=2}^{N} \epsilon_n x\right] \\
&= m^{\frac{1-p}{p}} \frac{m^2 + m}{2} \mathbb{E}\left[\frac{1}{2} \sum_{n=2}^{N} \epsilon_n(-1+1) - \frac{1}{2}\left|\sum_{n=2}^{N} \epsilon_n(-1-1)\right|\right],
\end{aligned}$$

where the second-last equality follows because the minima of a linear function over an interval is at one of the endpoints and the last equality follows because $\min\{x, y\} = \frac{1}{2}(x + y) - \frac{1}{2}|x - y|$. Since ϵ_n are Rademacher random variables equal to ± 1 with probability $\frac{1}{2}$, we can simplify the above as

$$\begin{aligned}
\mathbb{E}\left[\min_{x\in\mathcal{X}} \sum_{t=1}^{T} \tilde{f}_t(x)\right] &= m^{\frac{1-p}{p}} \frac{m^2+m}{2} \mathbb{E}\left[-\frac{1}{2}\left|\sum_{n=2}^{N} -2\epsilon_n\right|\right] \\
&= m^{\frac{1-p}{p}} \frac{m^2+m}{2} \mathbb{E}\left[-\left|\sum_{n=2}^{N} \epsilon_n\right|\right] \\
&= -m^{\frac{1-p}{p}} \frac{m^2+m}{2} \mathbb{E}\left[\left|\sum_{n=2}^{N} \epsilon_n\right|\right] \\
&\leq -m^{\frac{1-p}{p}} \frac{m^2+m}{2} \sqrt{N},
\end{aligned}$$

where the last inequality follows from Khintchine's inequality. Using the definition $N = \frac{T}{m}$, we have

$$\begin{aligned}
\mathbb{E}\left[\min_{x\in\mathcal{X}} \sum_{t=1}^{T} \tilde{f}_t(x)\right] &\leq -m^{\frac{1-p}{p}} \frac{m^2+m}{2} \sqrt{\frac{T}{m}} \\
&= -\frac{1}{2}\sqrt{T}\left(m^{\frac{3}{2}+\frac{1-p}{p}} + m^{\frac{1}{2}+\frac{1-p}{p}}\right) \\
&\leq -O\left(\sqrt{T} m^{\frac{3}{2}+\frac{1-p}{p}}\right) \\
&= -O\left(\sqrt{T} m^{\frac{p+2}{2p}}\right) \\
&= -O\left(\sqrt{T m^{\frac{p+2}{p}}}\right).
\end{aligned}$$

Therefore, we have

$$\mathbb{E}_{\epsilon_1,\ldots,\epsilon_N}\left[R_T(\texttt{FTRL})\right] = \mathbb{E}\left[\sum_{t=1}^{T} f_t(h_t)\right] - \mathbb{E}\left[\min_{x\in\mathcal{X}} \sum_{t=1}^{T} \tilde{f}_t(x)\right] \geq \Omega\left(\sqrt{T m^{\frac{p+2}{p}}}\right).$$

This completes the proof. ■

Now we provide a lower bound for the OCO with ρ-discounted infinite memory problem. For simplicity, we consider the case when the problem follows linear sequence dynamics with the 2-norm instead of a general p-norm.

Theorem D.2. *Let $\rho \in [\frac{1}{2}, 1)$. There exists an instance of the online convex optimization with ρ-discounted infinite memory problem, $(\mathcal{X}, \mathcal{H}, A_{\text{infinite},\rho}, B_{\text{infinite}})$, that follows linear sequence dynamics with the 2-norm and there exist L-Lipschitz continuous loss functions $\{f_t : \mathcal{H} \to \mathbb{R}\}_{t=1}^{T}$ such that the regret of any algorithm $\mathcal{A}$ satisfies*

$$R_T(\mathcal{A}) \geq \Omega\left(\sqrt{TL^2(1-\rho)^{-2}}\right).$$

The proof is very similar to that of Theorem D.1 with slight adjustments to account for a ρ-discounted infinite memory instead of a finite memory of constant size m.

Proof. Let $\mathcal{X} = [-1, 1]$ and consider an OCO with infinite memory problem with discount factor ρ, $(\mathcal{X}, \mathcal{H}, A_{\text{infinite},\rho}, B_{\text{infinite}})$, that follows linear sequence dynamics with the 2-norm. For simplicity, assume that T is a multiple of $(1-\rho)^{-1}$ (otherwise, the same proof works but with slightly more tedious bookkeeping) and that $L = 1$ (otherwise, multiply the functions f_t defined below by L).

Define $m = (1-\rho)^{-1}$. Divide the T rounds into $N = \frac{T}{m}$ blocks of m rounds each. Sample N independent Rademacher random variables $\{\epsilon_1, \ldots, \epsilon_N\}$, where each ϵ_i is equal to ± 1 with

probability $\frac{1}{2}$. Recall that $h_t = (x_t, \rho x_{t-1}, \dots, \rho^{t-1} x_1, 0, \dots)$. Define the loss functions $\{f_t\}_{t=1}^T$ as follows. If $t \le m$, let $f_t = 0$. Otherwise, let

$$f_t(h_t) = \epsilon_{\lceil \frac{t}{m} \rceil} m^{-\frac{1}{2}} \sum_{k=0}^{m-1} \rho^{k+t-m\lfloor \frac{t}{m} \rfloor - 1} x_{m\lfloor \frac{t}{m} \rfloor + 1 - k}.$$

The functions f_t are linear, so they are convex. In order to show that they satisfy Assumptions **A3** and **A4**, it remains to show that they are 1-Lipschitz continuous. Let $h = (x^{(1)}, \rho x^{(2)}, \dots)$ and $\tilde{h} = (\tilde{x}^{(1)}, \rho \tilde{x}^{(2)}, \dots)$ be arbitrary elements of $\mathcal{H}$. We have

$$\begin{aligned}
&\left| f_t(h) - f_t(\tilde{h}) \right| \\
&\le \left| \epsilon_{\lceil \frac{t}{m} \rceil} m^{-\frac{1}{2}} \sum_{k=1}^{m} \rho^{k-1} \left(x^{(k)} - \tilde{x}^{(k)} \right) \right| \\
&\le m^{-\frac{1}{2}} \left| \sum_{k=1}^{m} \rho^{k-1} \left(x^{(k)} - \tilde{x}^{(k)} \right) \right| && \text{because } \epsilon_{\lceil \frac{t}{m} \rceil} \in \{-1, +1\} \\
&\le m^{-\frac{1}{2}} m^{\frac{1}{2}} \left(\sum_{k=1}^{m} \rho^{2(k-1)} \left| x^{(k)} - \tilde{x}^{(k)} \right|^2 \right)^{\frac{1}{2}} && \text{by Hölder's inequality} \\
&\le \|h - \tilde{h}\|_{\mathcal{H}},
\end{aligned}$$

where the last equality follows because the follows linear sequence dynamics with the 2-norm.

First we will show that the total expected loss of any algorithm is 0, where the expectation is with respect to the randomness in the choice of $\{\epsilon_1, \dots, \epsilon_N\}$. The total loss in the first block is 0 because $f_t = 0$ for $t \in [m]$. For each subsequent block $n \in \{2, \dots, N\}$, the total loss in block n depends on the algorithm's choices made *before* observing ϵ_n, namely, $\{x_{(n-2)m+2}, \dots, x_{(n-1)m+1}\}$. Since ϵ_n is equal to ± 1 with probability $\frac{1}{2}$, the expected loss of any algorithm in a block is equal to 0 and the total expected loss is also equal to 0.

Now we will show that the expected loss of the benchmark is at most

$$-O\left(\sqrt{T(1-\rho)^{-2}}\right),$$

where the expectation is with respect to the randomness in the choice of $\{\epsilon_1, \dots, \epsilon_N\}$. We have

$$\begin{aligned}
\mathbb{E}\left[\min_{x \in \mathcal{X}} \sum_{t=1}^{T} \tilde{f}_t(x) \right] &= \mathbb{E}\left[\min_{x \in \mathcal{X}} \sum_{n=2}^{N} \sum_{t=(n-1)m+1}^{nm} \tilde{f}_t(x) \right] \\
&= \mathbb{E}\left[\min_{x \in \mathcal{X}} \sum_{n=2}^{N} \sum_{t=(n-1)m+1}^{nm} \epsilon_n m^{-\frac{1}{2}} \sum_{k=0}^{m-1} \rho^{k+t-(n-1)m-1} x \right] \\
&= m^{-\frac{1}{2}} \mathbb{E}\left[\min_{x \in \mathcal{X}} \sum_{n=2}^{N} \epsilon_n x \sum_{t=(n-1)m+1}^{nm} \rho^{t-(n-1)m-1} \sum_{k=0}^{m-1} \rho^k \right] \\
&= m^{-\frac{1}{2}} \frac{1-\rho^m}{1-\rho} \mathbb{E}\left[\min_{x \in \mathcal{X}} \sum_{n=2}^{N} \epsilon_n x \sum_{t=(n-1)m+1}^{nm} \rho^{t-(n-1)m-1} \right] \\
&= m^{-\frac{1}{2}} \left(\frac{1-\rho^m}{1-\rho} \right)^2 \underbrace{\mathbb{E}\left[\min_{x \in \mathcal{X}} \sum_{n=2}^{N} \epsilon_n x \right]}_{(a)}.
\end{aligned}$$

The term (a) above can be bounded above by $-\sqrt{N}$ as in the proof of Theorem D.1 using Khintchine's inequality. Therefore, using that $N = \frac{T}{m}$ and $m = (1-\rho)^{-1}$ we have

$$\begin{aligned}
\mathbb{E}\left[\min_{x\in\mathcal{X}} \sum_{t=1}^{T} \tilde{f}_t(x)\right] &\leq -m^{-\frac{1}{2}} \left(\frac{1-\rho^m}{1-\rho}\right)^2 \sqrt{N} \\
&\leq -(1-\rho)^{\frac{1}{2}} \left(\frac{1-\rho^m}{1-\rho}\right)^2 \sqrt{T(1-\rho)} \\
&= -\sqrt{T}\frac{(1-\rho^m)^2}{1-\rho} \\
&= -\sqrt{T(1-\rho)^{-2}}(1-\rho^m)^2 \\
&\leq -O\left(\sqrt{T(1-\rho)^{-2}}\right),
\end{aligned}$$

where the last inequality follows from the assumption that $\rho \in [\frac{1}{2}, 1)$ and the following argument:

$$\begin{aligned}
\rho^m = (1-(1-\rho))^m = (1-(1-\rho))^{\frac{1}{1-\rho}} \leq \frac{1}{e} \\
\Rightarrow (1-\rho^m) \geq 1 - \frac{1}{e} \\
\Rightarrow (1-\rho^m)^2 \geq \left(1-\frac{1}{e}\right)^2 \\
\Rightarrow -(1-\rho^m)^2 \leq -\left(1-\frac{1}{e}\right)^2
\end{aligned}$$

This completes the proof. ■

E Online Linear Control

E.1 Formulation as OCO with Unbounded Memory

Now we formulate the online linear control problem in our framework by defining the decision space $\mathcal{X}$, the history space $\mathcal{H}$, and the linear operators $A : \mathcal{H} \to \mathcal{H}$ and $B : \mathcal{W} \to \mathcal{H}$. Then, we define the functions $f_t : \mathcal{H} \to \mathbb{R}$ in terms of c_t and finally, prove an upper bound on the policy regret. For notational convenience, let $(M^{[s]})$ and (Y_k) denote the sequences $(M^{[1]}, M^{[2]}, \dots)$ and $(Y_0, Y_1, \dots)$ respectively.

Recall that we fix $K \in \mathcal{K}$ to be an arbitrary (κ, ρ)-strongly stable linear controller and consider the disturbance-action controller policy class $\mathcal{M}_K$ (Definition 4.1). For the rest of this paper let $\widetilde{F} = F - GK$. The first step is a change of variables with respect to the control inputs from linear controllers to DACs and the second is a corresponding change of variables for the state. Define the decision space $\mathcal{X}$ as

$$\mathcal{X} = \{M = (M^{[s]}) : M^{[s]} \in \mathbb{R}^{d\times d}, \|M^{[s]}\|_2 \leq \kappa^4 \rho^s\} \tag{10}$$

with

$$\|M\|_{\mathcal{X}} = \sqrt{\sum_{s=1}^{\infty} \rho^{-s} \|M^{[s]}\|_F^2}. \tag{11}$$

Define the history space $\mathcal{H}$ to be the set consisting of sequences $h = (Y_k)$, where $Y_0 \in \mathcal{X}$ and $Y_k = \widetilde{F}^{k-1} G X_k$ for $X_k \in \mathcal{X}, k \geq 1$ with

$$\|h\|_{\mathcal{H}} = \sqrt{\sum_{k=0}^{\infty} \xi_k^2 \|Y_k\|_{\mathcal{X}}^2}, \tag{12}$$

where the weights (ξ_k) are nonnegative real numbers defined as

$$\xi = (1, 1, 1, \rho^{-\frac{1}{2}}, \rho^{-1}, \rho^{-\frac{3}{2}}, \dots). \tag{13}$$

Define the linear operators $A : \mathcal{H} \to \mathcal{H}$ and $B : \mathcal{W} \to \mathcal{H}$ as

$$A((Y_0, Y_1, \dots)) = (0, GY_0, \widetilde{F}Y_1, \widetilde{F}Y_2, \dots) \quad \text{and} \quad B(M) = (M, 0, 0, \dots).$$

Note that the problem follows linear sequence dynamics with the ξ-weighted 2-norm (Definition 2.3), where ξ is defined above in Eq. (13). The weights in the weighted norms on $\mathcal{X}$ and $\mathcal{H}$ increase exponentially. However, the norms $\|M^{[s]}\|_F^2$ and $\|\widetilde{F}^{k-1}G\|_F^2$ decrease exponentially as well: by definition of $M^{[s]}$ in Eq. (11) and the assumption on $\widetilde{F} = F - GK$ for $K \in \mathcal{K}$. Leveraging this exponential decrease in $\|M^{[s]}\|_F^2$ and $\|\widetilde{F}^{k-1}G\|_F^2$ to define exponentially increasing weights turns out to be crucial for deriving our regret bounds that are stronger than existing results. Furthermore, the choice to have $\xi_p = 1$ for $p \in \{1, 2\}$ in addition to $p = 0$ (as required by Definition 2.3) might seem like a small detail, but this also turns out to be crucial for avoiding unnecessary factors of ρ^{-1} in the regret bounds.

Recall that the loss functions in the online linear control problem are $c_t(s_t, u_t)$, where s_t and u_t are the state and control at round t. Now we will show how to construct the functions $f_t : \mathcal{H} \to \mathbb{R}$ that correspond to $c_t(s_t, u_t)$. By definition, given a sequence of decisions $(M_0, \dots, M_t)$, the history at the end of round t is given by

$$h_t = (M_t, GM_{t-1}, \widetilde{F}GM_{t-2}, \dots, \widetilde{F}^{t-1}GM_0, 0, \dots).$$

A simple inductive argument shows that the state and control in round t can be written as

$$s_t = \widetilde{F}^t s_0 + \sum_{k=0}^{t-1} \sum_{s=1}^{k+1} \widetilde{F}^{t-k-1} G M_k^{[s]} w_{k-s} + w_{t-1}, \tag{14}$$

$$u_t = -Ks_t + \sum_{s=1}^{t+1} M_t^{[s]} w_{t-s}. \tag{15}$$

Define the functions $f_t : \mathcal{H} \to \mathbb{R}$ by $f_t(h) = c_t(s, u)$, where s and u are the state and control determined by the history as above. Note that f_t is parameterized by the past disturbances. Since the state and control are linear functions of the history and c_t is convex, this implies that f_t is convex.

With the above formulation and the fact that the class of disturbance-action controllers is a superset of the class of (κ, ρ)-strongly-stable linear controllers, we have that the policy regret for the online linear control problem is at most

$$\sum_{t=0}^{T-1} f_t(h_t) - \min_{M \in \mathcal{X}} \sum_{t=0}^{T-1} \tilde{f}_t(M).$$

This completes the specification of the online convex optimization with unbounded memory problem, $(\mathcal{X}, \mathcal{H}, A, B)$, corresponding to the online linear control problem. Using Algorithm 1 and Theorem 3.1 we can upper bound the above by

$$O\left(\sqrt{\frac{D}{\alpha} T L \tilde{L} H_2}\right),$$

where L is the Lipschitz constant of f_t, $\tilde{L}$ is the Lipschitz constant of $\tilde{f}_t$, H_2 is the 2-effective memory capacity, and $D = \max_{x, \tilde{x} \in \mathcal{X}} |R(x) - R(\tilde{x})|$ for an α-strongly-convex regularizer $R : \mathcal{X} \to \mathbb{R}$. In the next subsection we bound these quantities in terms of the problem parameters of the online linear control problem. We use $O(\cdot)$ to hide absolute constants.

E.2 Regret Analysis

We use the following standard facts about matrix norms.

Lemma E.1. *Let $M, N \in \mathbb{R}^{d \times d}$. Then,*

1. $\|M\|_2 \le \|M\|_F \le \sqrt{d}\|M\|_2$.
2. $\|MN\|_F \le \|M\|_2\|N\|_F$.

Proof. Part 1 can be found in, for example, Golub and Loan [1996, Section 2.3.2]. Letting N_j denote the j-th column of N, part 2 follows from

$$\|MN\|_F^2 = \sum_{j=1}^{d} \|MN_j\|_2^2 \le \|M\|_2^2 \sum_{j=1}^{d} \|N_j\|_2^2 = \|M\|_2^2\|N\|_F^2.$$

This completes the proof. ■

Lemma E.2. *For $s \ge 2$, the operator norm $\|A^s\|$ is bounded above as*

$$\|A^s\| \le O\left(\kappa^4 \rho^{\frac{s}{2}}\right).$$

Proof. Recall the definition of $\mathcal{H}$ and $\|\cdot\|_{\mathcal{H}}$ (Eq. (12)). Let

$$(Y_0, Y_1, \dots) = (Y_0, GX_1, \widetilde{F}GX_2, \widetilde{F}^2GX_3, \dots)$$

be an element of $\mathcal{H}$ with unit norm, i.e.,

$$\sqrt{\sum_{k=0}^{\infty} \xi_k^2 \|Y_k\|_{\mathcal{X}}^2} = 1,$$

where the weights (ξ_k) are defined in Eq. (13). Note that $\xi_p = 1$ for $p = 0, 1$ and $\xi_p^2 = \rho^{-p+2}$ for $p = 2, 3, \dots$. From the definition of the operator A and for $s \ge 2$, we have

$$A^s((Y_0, Y_1, \dots)) = (0, \dots, 0, \widetilde{F}^{s-1}GY_0, \widetilde{F}^sGX_1, \widetilde{F}^{s+1}GX_2, \dots).$$

Now we bound $\|A^s\|$ as follows. By definition of A^s and $\|\cdot\|_{\mathcal{H}}$ (Eq. (12)), and part 2 of Lemma E.1, we have

$$\begin{aligned}
\|A^s((Y_0, Y_1, \dots))\| &= \sqrt{\rho^{-s+2}\|\widetilde{F}^{s-1}GY_0\|_{\mathcal{X}}^2 + \sum_{k=1}^{\infty} \rho^{-s-k+2}\|\widetilde{F}^{s+k-1}GX_k\|_{\mathcal{X}}^2} \\
&\le \sqrt{\rho^{-s+2}\|\widetilde{F}^{s-1}G\|_2^2\|Y_0\|_{\mathcal{X}}^2 + \sum_{k=1}^{\infty} \rho^{-s-k+2}\|\widetilde{F}^{s-1}\|_2^2\|\widetilde{F}\|_2^2\|\widetilde{F}^{k-1}GX_k\|_{\mathcal{X}}^2} \\
&\le \rho^{-\frac{s}{2}}\|\widetilde{F}^{s-1}\|_2 \sqrt{\rho^2\|G\|_2^2\|Y_0\|_{\mathcal{X}}^2 + \sum_{k=1}^{\infty} \rho^{-k+2}\|\widetilde{F}\|_2^2\|\widetilde{F}^{k-1}GX_k\|_{\mathcal{X}}^2} \\
&= \rho^{-\frac{s}{2}}\|\widetilde{F}^{s-1}\|_2 \sqrt{\rho^2\|G\|_2^2\|Y_0\|_{\mathcal{X}}^2 + \sum_{k=1}^{\infty} \rho^{-k+2}\|\widetilde{F}\|_2^2\|Y_k\|_{\mathcal{X}}^2}.
\end{aligned}$$

Using our assumptions that $\|G\|_2 \le \kappa$ and $\|\widetilde{F}\|_2 \le \kappa^2\rho$, we have

$$\begin{aligned}
\|A^s((Y_0, Y_1, \dots))\| &\le \rho^{-\frac{s}{2}}\|\widetilde{F}^{s-1}\|_2 \sqrt{\rho^2\kappa^2\|Y_0\|_{\mathcal{X}}^2 + \sum_{k=1}^{\infty} \rho^{-k+2}\kappa^4\rho^2\|Y_k\|_{\mathcal{X}}^2} \\
&\le \rho^{-\frac{s}{2}}\rho\kappa^2\|\widetilde{F}^{s-1}\|_2 \sqrt{\|Y_0\|_{\mathcal{X}}^2 + \sum_{k=1}^{\infty} \rho^{-k+2}\|Y_k\|_{\mathcal{X}}^2} \\
&\le \rho^{-\frac{s}{2}}\rho\kappa^2\kappa^2\rho^{s-1} \sqrt{\|Y_0\|_{\mathcal{X}}^2 + \sum_{k=1}^{\infty} \rho^{-k+2}\|Y_k\|_{\mathcal{X}}^2} \\
&= \kappa^4\rho^{\frac{s}{2}} \sqrt{\|Y_0\|_{\mathcal{X}}^2 + \sum_{k=1}^{\infty} \rho^{-k+2}\|Y_k\|_{\mathcal{X}}^2}.
\end{aligned}$$

Using $\rho^{-1+2} = \rho < 1$ for $k = 1$ in the above sum, the definition of (ξ_k), and our assumption that $(Y_0, Y_1 \dots)$ has unit norm, we have

$$\|A^s((Y_0, Y_1, \dots))\| \le \kappa^4 \rho^{\frac{s}{2}} \sqrt{\xi_0^2 \|Y_0\|_{\mathcal{X}}^2 + \sum_{k=1}^{\infty} \xi_k^2 \|Y_k\|_{\mathcal{X}}^2} = \kappa^4 \rho^{\frac{s}{2}}.$$

This completes the proof. ■

Lemma E.3. *The 2-effective memory capacity is bounded above as*

$$H_2 \le O\left(\kappa^4 (1-\rho)^{-\frac{3}{2}}\right).$$

Proof. Using Lemma E.2 to bound $\|A^k\|$ for $k \ge 2$, we have

$$H_2 = \sqrt{\sum_{k=0}^{\infty} k^2 \|A^k\|^2} \le O\left(\sqrt{\sum_{k=2}^{\infty} k^2 \kappa^8 \rho^k}\right) \le O\left(\kappa^4 (1-\rho)^{-\frac{3}{2}}\right). \qquad \blacksquare$$

Lemma E.4. *Suppose $R : \mathcal{X} \to \mathbb{R}$ is defined by $R(M) = \frac{1}{2}\|M\|_{\mathcal{X}}^2$. Then, it is 1-strongly-convex and $D = \max_{M, \widetilde{M} \in \mathcal{X}} |R(M) - R(\widetilde{M})| \le d\kappa^8 (1-\rho)^{-1}$.*

Proof. Note that R is 1-strongly-convex by definition. Using part 1 of Lemma E.1 and the definition of $\mathcal{X}$ (Eq. (10)), we have for all $M, \widetilde{M} \in \mathcal{X}$,

$$\begin{aligned}
D &= \max_{M, \widetilde{M} \in \mathcal{X}} |R(M) - R(\widetilde{M})| \\
&= \max_{M, \widetilde{M} \in \mathcal{X}} \left| \frac{1}{2}\|M\|_{\mathcal{X}}^2 - \frac{1}{2}\|\widetilde{M}\|_{\mathcal{X}}^2 \right| \\
&\le \max_{M \in \mathcal{X}} \|M\|_{\mathcal{X}}^2 \\
&= \max_{M \in \mathcal{X}} \sum_{s=1}^{\infty} \rho^{-s} \|M^{[s]}\|_F^2 && \text{by Eq. (11)} \\
&\le \max_{M \in \mathcal{X}} \sum_{s=1}^{\infty} \rho^{-s} d \|M^{[s]}\|_2^2 && \text{by Lemma E.1} \\
&\le \sum_{s=1}^{\infty} \rho^{-s} d\kappa^8 \rho^{2s} && \text{by Eq. (10)} \\
&\le d\kappa^8 (1-\rho)^{-1}.
\end{aligned}$$

This completes the proof. ■

Lemma E.5. *We can bound the norm of the state and control at time t as*

$$\max\{\|s_t\|_2, \|u_t\|_2\} \le D_{\mathcal{X}} = O\left(W\kappa^8 (1-\rho)^{-2}\right).$$

Proof. We can bound the norm of s_t and u_t using Eqs. (14) and (15) as

$$\begin{aligned}
\|s_t\|_2 &\le \left\| \widetilde{F}^t s_0 + \sum_{k=0}^{t-1} \sum_{s=1}^{k+1} \widetilde{F}^{t-k-1} G M_k^{[s]} w_{k-s} + w_{t-1} \right\|_2 \\
&\le \kappa^2 \rho^t + W + \sum_{k=0}^{t-1} \sum_{s=1}^{k+1} \kappa^2 \rho^{t-k-1} \kappa \kappa^4 \rho^s W \\
&\le \kappa^2 + W + W \kappa^7 (1-\rho)^{-2} \\
&\le O\left(W \kappa^7 (1-\rho)^{-2}\right). \\
\|u_t\|_2 &\le \left\| K s_t + \sum_{s=1}^{t+1} M_t^{[s]} w_{t-s} \right\|_2 \\
&\le O\left(W \kappa^8 (1-\rho)^{-2}\right) + \sum_{s=1}^{t+1} W \kappa^4 \rho^s \\
&\le O\left(W \kappa^8 (1-\rho)^{-2}\right).
\end{aligned}$$

Above, we used the assumptions that $\kappa, W \ge 1$. This completes the proof. ■

Lemma E.6. *The Lipschitz constant of f_t can be bounded above as*

$$L \le O\left(L_0 D_{\mathcal{X}} W \kappa (1-\rho)^{-1}\right),$$

where $D_{\mathcal{X}}$ is defined in Lemma E.5.

Proof. Let $(M_0, \ldots, M_t)$ and $(\widetilde{M}_0, \ldots, \widetilde{M}_t)$ be two sequences of decisions, where M_k and $\widetilde{M}_k \in \mathcal{X}$. Let h_t and $\tilde{h}_t$ be the corresponding histories, and (s_t, u_t) and $(\tilde{s}_t, \tilde{u}_t)$ be the corresponding state-control pairs at the end of round t. We have

$$\begin{aligned}
\left| f_t(h_t) - f_t(\tilde{h}_t) \right| &= |c_t(s_t, u_t) - c_t(\tilde{s}_t, \tilde{u}_t)| \\
&\le L_0 D_{\mathcal{X}} \max\{\|s_t - \tilde{s}_t\|_2, \|u_t - \tilde{u}_t\|_2\},
\end{aligned}$$

where the last inequality follows from our assumptions about the functions c_t and Lemma E.5. It suffices to bound the two norms on the right-hand side in terms of $\|h_t - \tilde{h}_t\|_{\mathcal{H}}$. For $k = 0, \ldots, t-1$,

define $Z_k^{[s]} = \widetilde{F}^{t-k-1} G(M_k^{[s]} - \widetilde{M}_k^{[s]})$. Using Eq. (14), we have

$$\begin{aligned}
\|s_t - \tilde{s}_t\|_2 &= \left\| \sum_{k=0}^{t-1} \sum_{s=1}^{k+1} Z_k^{[s]} w_{k-s} \right\|_2 \\
&\le \sum_{k=0}^{t-1} \sum_{s=1}^{k+1} \left\| Z_k^{[s]} w_{k-s} \right\|_2 \\
&= \sum_{k=0}^{t-1} \sum_{s=1}^{k+1} \left\| \rho^{-\frac{s}{2}} Z_k^{[s]} \rho^{\frac{s}{2}} w_{k-s} \right\|_2 \\
&\le \sum_{k=0}^{t-1} \sum_{s=1}^{k+1} \left\| \rho^{-\frac{s}{2}} Z_k^{[s]} \right\|_2 \left\| \rho^{\frac{s}{2}} w_{k-s} \right\|_2 \\
&= \sum_{k=0}^{t-1} \sum_{s=1}^{k+1} \xi_{1+t-1-k} \left\| \rho^{-\frac{s}{2}} Z_k^{[s]} \right\|_2 \xi_{1+t-1-k}^{-1} \left\| \rho^{\frac{s}{2}} w_{k-s} \right\|_2 \\
&\le \sqrt{\sum_{k=0}^{t-1} \sum_{s=1}^{k+1} \xi_{t-k}^2 \left\| \rho^{-\frac{s}{2}} Z_k^{[s]} \right\|_2^2} \sqrt{\sum_{k=0}^{t-1} \sum_{s=1}^{k+1} \xi_{t-k}^{-2} \left\| \rho^{\frac{s}{2}} w_{k-s} \right\|_2^2} \qquad (16) \\
&= \underbrace{\sqrt{\sum_{k=0}^{t-1} \xi_{t-k}^2 \sum_{s=1}^{k+1} \left\| \rho^{-\frac{s}{2}} Z_k^{[s]} \right\|_2^2}}_{(a)} \underbrace{\sqrt{\sum_{k=0}^{t-1} \xi_{t-k}^{-2} \sum_{s=1}^{k+1} \left\| \rho^{\frac{s}{2}} w_{k-s} \right\|_2^2}}_{(b)},
\end{aligned}$$

where Eq. (16) follows from the Cauchy-Schwarz inequality. The specific choice of weighted norms on $\mathcal{X}$ and $\mathcal{H}$ allow us to bound the terms (a) and (b) in terms of $\|h_t - \tilde{h}_t\|_{\mathcal{H}}$. We can bound the term (a) using the definition of $Z_k^{[s]}$, $\|\cdot\|_{\mathcal{X}}$, and $\|\cdot\|_{\mathcal{H}}$ as

$$\begin{aligned}
\sqrt{\sum_{k=0}^{t-1} \xi_{t-k}^2 \sum_{s=1}^{k+1} \left\| \rho^{-\frac{s}{2}} Z_k^{[s]} \right\|_2^2} &= \sqrt{\sum_{k=0}^{t-1} \xi_{t-k}^2 \sum_{s=1}^{k+1} \rho^{-s} \left\| \widetilde{F}^{t-k-1} G(M_k^{[s]} - \widetilde{M}_k^{[s]}) \right\|_2^2} \\
&\le \sqrt{\sum_{k=0}^{t-1} \xi_{t-k}^2 \sum_{s=1}^{k+1} \rho^{-s} \left\| \widetilde{F}^{t-k-1} G(M_k^{[s]} - \widetilde{M}_k^{[s]}) \right\|_F^2} \qquad (17) \\
&\le \|h_t - \tilde{h}_t\|_{\mathcal{H}}, \qquad (18)
\end{aligned}$$

where Eq. (17) follows from part 1 of Lemma E.1 and Eq. (18) follows from the definitions of $\|\cdot\|_{\mathcal{X}}$ and $\|\cdot\|_{\mathcal{H}}$. Using $\|w_t\|_2 \le W$ for all rounds t, we can bound the term (b) as

$$\begin{aligned}
\sqrt{\sum_{k=0}^{t-1} \xi_{t-k}^{-2} \sum_{s=1}^{k+1} \left\| \rho^{\frac{s}{2}} w_{k-s} \right\|_2^2} &\le W \sqrt{\sum_{k=0}^{t-1} \xi_{t-k}^{-2} \sum_{s=1}^{k+1} \rho^s} \\
&\le W \sqrt{\sum_{k=0}^{t-1} \xi_{t-k}^{-2} \frac{\rho(1-\rho^{k+1})}{1-\rho}} \\
&\le W(1-\rho)^{-1}, \qquad (19)
\end{aligned}$$

where Eq. (19) follows from the definition of (ξ_k) (Eq. (13)). Substituting Eqs. (18) and (19) in Eq. (16), we have

$$\|s_t - \tilde{s}_t\|_2 \le W(1-\rho)^{-1} \|h_t - \tilde{h}_t\|_{\mathcal{H}}.$$

Similarly,

$$\begin{aligned}\|u_t - \tilde{u}_t\| &= \left\| K(s_t - \tilde{s}_t) + \sum_{s=1}^{t+1} (M_t^{[s]} - \widetilde{M}_t^{[s]}) w_{t-s} \right\|_2 \\ &\leq O\left(W\kappa(1-\rho)^{-1} \left\| h_t - \tilde{h}_t \right\|_{\mathcal{H}} \right),\end{aligned}$$

where the last inequality follows from our assumption that $\|K\|_2 \leq \kappa$ and the above inequality for $\|s_t - \tilde{s}_t\|_2$. This completes the proof. ■

Lemma E.7. *The Lipschitz constant of $\tilde{f}_t$ can be bounded above as*

$$\tilde{L} \leq O\left(L_0 D_{\mathcal{X}} W \kappa^5 (1-\rho)^{-\frac{3}{2}} \right),$$

where $D_{\mathcal{X}}$ is defined in Lemma E.5.

Proof. Using Lemma E.2 that bounds $\|A^k\|$, we have

$$\sqrt{\sum_{k=0}^{\infty} \|A^k\|^2} \leq O\left(\kappa^4 (1-\rho)^{-\frac{1}{2}} \right).$$

Using Theorem 2.1 that bounds $\tilde{L}$ in terms of L and the above, we have

$$\tilde{L} \leq O\left(L\kappa^4 (1-\rho)^{-\frac{1}{2}} \right) \leq O\left(L_0 D_{\mathcal{X}} W \kappa^5 (1-\rho)^{-\frac{3}{2}} \right),$$

where the last inequality follows from Lemma E.6. ■

Now we restate and prove Theorem 4.1.

Theorem 4.1. *Consider the online linear control problem as defined in Section 4.1. Suppose the decisions in round t are chosen using Algorithm 1. Then, the upper bound on the policy regret is*

$$O\left(L_0 W^2 \sqrt{T} d^{\frac{1}{2}} \kappa^{17} (1-\rho)^{-4.5} \right). \tag{2}$$

Proof. Using Theorem 3.1 and the above lemmas, we can upper bound the policy regret of Algorithm 1 for the online linear control problem by

$$\begin{aligned}&O\left(\sqrt{\frac{D}{\alpha} T L \tilde{L} H_2} \right) \\ &= O\left(\sqrt{d\kappa^8 (1-\rho)^{-1}\, T\, \left(L_0 W^2 \kappa^9 (1-\rho)^{-3}\right)^2 \kappa^4 (1-\rho)^{-\frac{1}{2}}\, \kappa^4 (1-\rho)^{-\frac{3}{2}}} \right) \\ &= O\left(L_0 W^2 \sqrt{T} d^{\frac{1}{2}} \kappa^{17} (1-\rho)^{-4.5} \right).\end{aligned}$$

This completes the proof. ■

E.3 Existing Regret Bound

The upper bound on policy regret for the online linear control problem in existing work is given in Agarwal et al. [2019b, Theorem 5.1]. The theorem statement only shows the dependence on $\tilde{L}, W$, and T. The dependence on d, κ, and ρ can be found in the details of the proof. Below we give a detailed accounting of all of these terms in their regret bound.

To simplify notation let $\gamma = 1 - \rho$. Agarwal et al. [2019b] define

$$H = \frac{\kappa^2}{\gamma} \log(T) \quad \text{and} \quad C = \frac{W(\kappa^2 + H\kappa_B \kappa^2 a)}{\gamma(1 - \kappa^2(1-\gamma)^{H+1})} + \frac{\kappa_B \kappa^3 W}{\gamma}.$$

The value of a is not specified in Theorem 5.1. However, from Theorem 5.3 and the definition of $\mathcal{M}$ in Algorithm 1 their paper, we can infer that $a = \kappa_B \kappa^3$.

The final regret bound is obtained by summing Equations 5.1, 5.3, and 5.4. Given the definition of H above, we have that

$$(1-\gamma)^{H+1} \leq \exp(-\kappa^2 \log T) = T^{-\kappa^2}.$$

So, the dominant term in the regret bound is Equation 5.4, which is

$$O\left(L_0 W C d^{\frac{3}{2}} \kappa_B^2 \kappa^6 H^{2.5} \gamma^{-1} \sqrt{T}\right).$$

Substituting the values of H and C from above and collecting terms, we have that the upper bound on policy regret in existing work [Agarwal et al., 2019b, Theorem 5.1] is

$$\begin{aligned}
&O\left(L_0 W d^{\frac{3}{2}} \sqrt{T} \log(T)^{2.5} \kappa_B^2 \kappa^{11} \gamma^{-3.5} C\right) \\
&= O\left(L_0 W d^{\frac{3}{2}} \sqrt{T} \log(T)^{2.5} \kappa_B^2 \kappa^{11} \gamma^{-3.5} \left(\frac{W(\kappa^2 + H\kappa_B \kappa^2 a)}{\gamma(1-\kappa^2(1-\gamma)^{H+1})} + \frac{\kappa_B \kappa^3 W}{\gamma}\right)\right) \\
&= O\left(L_0 W d^{\frac{3}{2}} \sqrt{T} \log(T)^{2.5} \kappa_B^2 \kappa^{11} \gamma^{-3.5} \left(\frac{W\kappa^2}{\gamma(1-\kappa^2(1-\gamma)^{H+1})} + \frac{W\kappa_B^2 \kappa^7 \log(T)}{\gamma^2(1-\kappa^2(1-\gamma)^{H+1})} + \frac{\kappa_B \kappa^3 W}{\gamma}\right)\right) \\
&= O\left(L_0 W^2 d^{\frac{3}{2}} \sqrt{T} \log(T)^{2.5} \kappa_B^2 \kappa^{13} \gamma^{-4.5} (1-\kappa^2(1-\gamma)^{H+1})^{-1}\right) \\
&\quad + O\left(L_0 W^2 d^{\frac{3}{2}} \sqrt{T} \log(T)^{3.5} \kappa_B^4 \kappa^{18} \gamma^{-5.5} (1-\kappa^2(1-\gamma)^{H+1})^{-1}\right) \\
&\quad + O\left(L_0 W^2 d^{\frac{3}{2}} \sqrt{T} \log(T)^{2.5} \kappa_B^3 \kappa^{14} \gamma^{-4.5}\right) \\
&= O\left(L_0 W^2 d^{\frac{3}{2}} \sqrt{T} \log(T)^{3.5} \kappa_B^4 \kappa^{18} \gamma^{-5.5}\right).
\end{aligned}$$

Above we used that $\lim_{T\to\infty}(1-\kappa^2(1-\gamma)^{H+1})^{-1} = 1$ to simplify the expressions. Therefore, the upper bound on policy regret for the online linear control problem in existing work is

$$O\left(L_0 W^2 d^{\frac{3}{2}} \sqrt{T} \log(T)^{3.5} \kappa_B^4 \kappa^{18} \gamma^{-5.5}\right). \tag{20}$$

F Online Performative Prediction

Before formulating the online performative prediction problem in our OCO with unbounded memory framework, we state the definition of 1-Wasserstein distance that we use in our regret analysis. Informally, the 1-Wasserstein distance is a measure of the distance between two probability measures.

Definition F.1 (1-Wasserstein Distance)**.** Let $(\mathcal{Z}, d)$ be a metric space. Let $\mathbb{P}(\mathcal{Z})$ denote the set of Radon probability measures ν on $\mathcal{Z}$ with finite first moment. That is, there exists $z' \in \mathcal{Z}$ such that $\mathbb{E}_{z\sim\nu}[d(z, z')] < \infty$. The 1-Wasserstein distance between two probability measures $\nu, \nu' \in \mathbb{P}(\mathcal{Z})$ is defined as

$$W_1(\nu, \nu') = \sup\{\mathbb{E}_{z\sim\nu}[f(z)] - \mathbb{E}_{z\sim\nu'}[f(z)]\},$$

where the supremum is taken over all 1-Lipschitz continuous functions $f : \mathcal{Z} \to \mathbb{R}$.

F.1 Formulation as OCO with Unbounded Memory

Now we formulate the online performative prediction problem in our framework by defining the decision space $\mathcal{X}$, the history space $\mathcal{H}$, and the linear operators $A : \mathcal{H} \to \mathcal{H}$ and $B : \mathcal{W} \to \mathcal{H}$. Then, we define the functions $f_t : \mathcal{H} \to \mathbb{R}$ in terms of l_t and finally, prove an upper bound on the policy regret. For notational convenience, let (y_k) denote the sequence $(y_0, y_1, \dots)$.

Let $\rho \in (0, 1)$. Let the decision space $\mathcal{X} \subseteq \mathbb{R}^d$ be closed and convex with $\|\cdot\|_{\mathcal{X}} = \|\cdot\|_2$. Let the history space $\mathcal{H}$ be the ℓ^1-direct sum of countably infinte number of copies of $\mathcal{X}$. Define the linear operators $A : \mathcal{H} \to \mathcal{H}$ and $B : \mathcal{X} \to \mathcal{H}$ as

$$A((y_0, y_1, \dots)) = (0, \rho y_0, \rho y_1, \dots) \quad \text{and} \quad B(x) = (x, 0, \dots).$$

Note that the problem is an OCO with ρ-discounted infinite memory problem and follows linear sequence dynamics with the 1-norm (Definition 2.3).

Given a sequence of decisions $(x_k)_{k=1}^t$, the history is $h_t = (x_t, \rho x_{t-1}, \dots, \rho^{t-1} x_1, 0, \dots)$ and the data distribution $p_t = p_t(h_t)$ satisfies:

$$z \sim p_t \text{ iff } z \sim \sum_{k=1}^{t-1} (1-\rho)\rho^{k-1}(\xi + F x_{t-k}) + \rho^t p_1. \tag{21}$$

This follows from the recursive definition of p_t and parametric assumption about $\mathcal{D}(x)$. Define the functions $f_t : \mathcal{H} \to [0,1]$ by

$$f_t(h_t) = \mathbb{E}_{z \sim p_t}[l_t(x_t, z)].$$

With the above formulation and definition of f_t, the original goal of minimizing the difference between the algorithm's total loss and the total loss of the best fixed decision is equivalent to minimizing the policy regret,

$$\sum_{t=1}^{T} f_t(h_t) - \min_{x \in \mathcal{X}} \sum_{t=1}^{T} \tilde{f}_t(x).$$

F.2 Regret Analysis

Lemma F.1. *The operator norm $\|A^s\|$ is bounded above as*

$$\|A^s\| \le O\left(\rho^s\right).$$

Proof. Recall the definition of $\mathcal{H}$ and $\|\cdot\|_{\mathcal{H}}$. Let

$$(y_0, y_1, \dots) = (x_0, \rho x_1, \rho^2 x_2, \dots)$$

be an element of $\mathcal{H}$ with unit norm, i.e.,

$$\sum_{k=0}^{\infty} \|y_k\| = 1.$$

From the definition of the operator A, we have

$$A^s((y_0, y_1, \dots)) = (0, \dots, 0, \rho^s x_0, \rho^{s+1} x_1, \dots).$$

Now we bound $\|A^s\|$ as follows. By definition of A^s and $\|\cdot\|_{\mathcal{H}}$, we have

$$\|A^s((y_0, y_1, \dots))\| = \sum_{k=0}^{\infty} \rho^{s+k}\|x_k\| = \rho^s \sum_{k=0}^{\infty} \rho^k \|x_k\| = \rho^s \sum_{k=0}^{\infty} \|y_k\| = \rho^s. \qquad \blacksquare$$

Lemma F.2. *The 1-effective memory capacity is bounded above as*

$$H_2 \le O\left((1-\rho)^{-2}\right).$$

Proof. Using Lemma F.1 to bound $\|A^k\|$, we have

$$H_1 = \sum_{k=0}^{\infty} k\|A^k\| = \sum_{k=0}^{\infty} k\rho^k \le O\left((1-\rho)^{-2}\right). \qquad \blacksquare$$

Lemma F.3. *Suppose $R : \mathcal{X} \to \mathbb{R}$ is defined by $R(x) = \frac{1}{2}\|x\|_{\mathcal{X}}^2$. Then, it is 1-strongly-convex and $D = \max_{x,\tilde{x} \in \mathcal{X}} |R(x) - R(\tilde{x})| \le D_{\mathcal{X}}^2$.*

Proof. Note that R is 1-strongly-convex by definition. By the assumption that $\|x\|_{\mathcal{X}} \le D_{\mathcal{X}}$ for all $x \in \mathcal{X}$, we have that $D \le D_{\mathcal{X}}^2$. $\blacksquare$

Lemma F.4. *The Lipschitz constant of f_t can be bounded above as*

$$L \le O\left(L_0 \frac{1-\rho}{\rho}\|F\|_2\right).$$

Proof. Let $(x_1, \ldots, x_t)$ and $(\tilde{x}_1, \ldots, \tilde{x}_t)$ be two sequences of decisions, where $x_k, \tilde{x}_k \in \mathcal{X}$. Let h_t and $\tilde{h}_t$ be the corresponding histories, and p_t and $\tilde{p}_t$ be the corresponding distributions at the end of round t. We have

$$\begin{aligned}
&\left|f_t(h_t) - f_t(\tilde{h}_t)\right| \\
&= \left|\mathbb{E}_{z\sim p_t}\left[l_t(x_t, z)\right] - \mathbb{E}_{z\sim \tilde{p}_t}\left[l_t(\tilde{x}_t, z)\right]\right| \\
&= \left|\mathbb{E}_{z\sim p_t}\left[l_t(x_t, z)\right] - \mathbb{E}_{z\sim p_t}\left[l_t(\tilde{x}_t, z)\right] + \mathbb{E}_{z\sim p_t}\left[l_t(\tilde{x}_t, z)\right] - \mathbb{E}_{z\sim \tilde{p}_t}\left[l_t(\tilde{x}_t, z)\right]\right| \\
&\leq L_0\|x_t - \tilde{x}_t\|_2 + L_0 W_1(p_t, \tilde{p}_t),
\end{aligned}$$

where the last inequality follows from the assumptions about the functions l_t and the definition of the Wasserstein distance W_1. By definition of p_t (Eq. (21)), we have

$$\begin{aligned}
W_1(p_t, \tilde{p}_t) &\leq \sum_{k=1}^{t-1} \frac{1-\rho}{\rho}\rho^k \|F\|_2 \|x_{t-k} - \tilde{x}_{t-k}\|_2 \\
&\leq \frac{1-\rho}{\rho}\|F\|_2 \|h_t - \tilde{h}_t\|_{\mathcal{H}},
\end{aligned}$$

where the last inequality follows from the definition of $\|\cdot\|_{\mathcal{H}}$. Therefore, $L \leq L_0 \frac{1-\rho}{\rho}\|F\|_2$. ■

Lemma F.5. *The Lipschitz constant of f_t can be bounded above as*

$$\tilde{L} \leq O\left(L_0 \frac{1}{\rho}\|F\|_2\right).$$

Proof. Using Lemma F.1 that bounds $\|A^k\|$, we have

$$\sum_{k=0}^{\infty} \|A^k\| = (1-\rho)^{-1}.$$

Using Theorem 2.1 that bounds $\tilde{L}$ in terms of L and the above, we have

$$\tilde{L} \leq O\left(L(1-\rho)^{-1}\right) = O\left(L_0 \frac{1}{\rho}\|F\|_2\right),$$

where the last equality follows from Lemma F.4. ■

Now we restate and prove Theorem 4.2.

Theorem 4.2. *Consider the online performative prediction problem as defined in Section 4.2. Suppose the decisions in round t are chosen using Algorithm 1. Then, the upper bound on the policy regret is*

$$O\left(D_{\mathcal{X}} L_0 \sqrt{T}\|F\|_2 (1-\rho)^{-\frac{1}{2}}\rho^{-1}\right).$$

Proof. Using Theorem 3.1 and the above lemmas, we can upper bound the policy regret of Algorithm 1 for the online performative prediction problem by

$$O\left(\sqrt{\frac{D}{\alpha} T L \tilde{L} H_1}\right) = O\left(D_{\mathcal{X}} L_0 \|F\|_2 (1-\rho)^{-\frac{1}{2}}\rho^{-1}\sqrt{T}\right).$$

This completes the proof. ■

We note that the upper bound can be improved by defining a weighted norm on $\mathcal{H}$ similar to the approach in Appendix E. However, here we present the looser anaysis for simplicity of exposition.

G Implementation Details for Algorithm 1

In this section we discuss how to implement Algorithm 1 efficiently.

Dimensionality of $\mathcal{X}$. First, note that the decisions $x \in \mathcal{X}$ could be high-dimensional, e.g., an unbounded sequence of matrices as in the online linear control problem, but this is external to our framework and is application dependent. Our framework can be applied to $\mathcal{X}$ or to a lower-dimensional decision space $\mathcal{X}'$. However, the choice of $\mathcal{X}'$ and analyzing the difference

$$\min_{x' \in \mathcal{X}'} \sum_{t=1}^{T} \tilde{f}_t(x') - \min_{x \in \mathcal{X}} \sum_{t=1}^{T} \tilde{f}_t(x)$$

is application dependent. For example, for the online linear control problem one could consider a restricted class of disturbance-action controllers that operate on a constant number of past disturbances as opposed to all the past disturbances, and then analyze the difference between these two policy classes. See, for example, Agarwal et al. [2019b, Lemma 5.2].

Computational cost of each iteration of Algorithm 1. Now we discuss how to implement each iteration of Algorithm 1 efficiently. We are interested in the computational cost of computing the decision x_{t+1} as a function of t. (Given the above discussion about the dimensionality of $\mathcal{X}$, we ignore the fact that the dimensionality of the decisions themselves could depend on t.) Therefore, for the purposes of this section we (i) use $O(\cdot)$ notation to hide absolute constants and problem parameters excluding t and T; (ii) invoke the operators A and B by calling oracles $\mathcal{O}_A(\cdot)$ and $\mathcal{O}_B(\cdot)$; and (iii) evaluate the functions f_t by calling oracles $\mathcal{O}_f(t, \cdot)$. Recall from Assumption **A1** that we assume the learner knows the operators A and B, and observes f_t at the end of each round t. So, the oracles $\mathcal{O}_A, \mathcal{O}_B$, and $\mathcal{O}_f$ are readily available.

Algorithm 1 chooses the decision x_{t+1} as

$$x_{t+1} \in \arg\min_{x \in \mathcal{X}} \sum_{s=1}^{t} \tilde{f}_s(x) + \frac{R(x)}{\eta} = \arg\min_{x \in \mathcal{X}} \underbrace{\sum_{s=1}^{t} f_s\left(\sum_{k=0}^{s-1} A^k B x\right)}_{=F_t(x)} + \frac{R(x)}{\eta}.$$

Since $F_t(x)$ is a sum of $f_1, \ldots, f_t$, evaluating $F_t(x)$ requires $\Theta(t)$ oracle calls to $\mathcal{O}_f$. However, this issue is present in FTRL for OCO and OCO with finite memory as well and is not specific to our framework. To deal with this issue, one could consider mini-batching algorithms [Dekel et al., 2012, Altschuler and Talwar, 2018, Chen et al., 2020] such as Algorithm 2.

A naïve implementation to evaluate $F_t(x)$ could require $O(t^3)$ oracle calls to $\mathcal{O}_A$: for each $s \in [t]$, constructing the argument $\sum_{k=0}^{s-1} A^k B x$ for f_s could require k oracle calls to $\mathcal{O}_A$ to compute $A^k B x$, for a total of $O(s^2)$ oracle calls. However, $F_t(x)$ can be evaluated with just $O(t)$ oracle calls to $\mathcal{O}_A$ by constructing the arguments incrementally. For $t \geq 0$, define $\Gamma_t : \mathcal{X} \to \mathcal{H}$ as

$$\begin{aligned} \Gamma_0(x) &= Bx \\ \Gamma_t(x) &= A\left(\Gamma_{t-1}(x)\right) \quad \text{for } t \geq 1. \end{aligned}$$

Note that $\Gamma_t(Bx) = A^t B x$. Also, for $t \geq 1$, define $\Phi_t : \mathcal{X} \to \mathcal{H}$ as

$$\begin{aligned} \Phi_1(x) &= \Gamma_0(x) \\ \Phi_t(x) &= \Phi_{t-1}(x) + \Gamma_{t-1}(x) \quad \text{for } t \geq 2. \end{aligned}$$

Note that $\Phi_s(x) = \sum_{k=0}^{s-1} A^k B x$ is the argument for f_s. These can be constructed incrementally as follows.

1. Construct $\Gamma_0(x)$ using one oracle call to $\mathcal{O}_B$.
2. For $s = 1$,
   (a) Construct $\Phi_1(x) = \Gamma_0(x)$.
   (b) Construct $\Gamma_1(x)$ from $\Gamma_0(x)$ using one oracle call to $\mathcal{O}_A$.
3. For $s \geq 2$,
   (a) Construct $\Phi_s(x)$ by adding $\Phi_{s-1}(x)$ and $\Gamma_{s-1}(x)$. This can be done in $O(1)$ time. Recall from our earlier discussion that $O(\cdot)$ hides absolute constants and problem parameters excluding t and T.
   (b) Construct $\Gamma_s(x)$ from $\Gamma_{s-1}(x)$ using one oracle call to $\mathcal{O}_A$.

By incrementally constructing $\Phi_s(x)$ as above, we can evaluate $F_t(x)$ in $O(t)$ time with $O(1)$ oracle calls to $\mathcal{O}_B$, $O(t)$ oracle calls to $\mathcal{O}_A$, and $O(t)$ oracle calls to $\mathcal{O}_f$.

Memory usage of Algorithm 1. We end with a brief discussion of the memory usage of Algorithm 1. We are interested in the memory usage of computing the decision x_{t+1} as a function of t. (Given the discussion about the dimensionality of $\mathcal{X}$ at the start of this section, we ignore the fact that the dimensionality of the decisions themselves could depend on t.) For each $t \in [T]$, the memory usage could be as low as $O(1)$ (if, for example, $\mathcal{X} \subseteq \mathbb{R}^d$, and $A, B \in \mathbb{R}^{d \times d}$, which implies that $\Phi_t(x)$ is a d-dimensional vector) or as high as $O(t)$ (if, for example, $\Phi_t(x)$ is a t-length sequence of d-dimensional vectors). However, the memory usage is already $\Omega(t)$ to store the functions $f_1, \ldots, f_t$. Therefore, Algorithm 1 only incurs a constant factor overhead.

H An Algorithm with A Low Number of Switches: Mini-Batch FTRL

In this section we present an algorithm (Algorithm 2) for OCO with unbounded memory that provides the same upper bound on policy regret as Algorithm 1 while guaranteeing a small number of switches. Algorithm 2 combines FTRL on the functions $\tilde{f}_t$ with a mini-batching approach. First, it divides rounds into batches of size S, where S is a parameter. Second, at the start of batch $b \in \{1, \ldots, \lceil T/S \rceil\}$, it performs FTRL on the functions $\{g_1, \ldots, g_b\}$, where g_i is the average of the functions $\tilde{f}_t$ in batch i. Then, it uses this decision for the entirety of the current batch. By design, Algorithm 2 switches decisions at most $O(T/S)$ times. This algorithm is insipired by similar algorithms for online learning and OCO [Dekel et al., 2012, Altschuler and Talwar, 2018, Chen et al., 2020].

Algorithm 2: `Mini-Batch FTRL`

Input : Time horizon T, step size η, α-strongly-convex regularizer $R : \mathcal{X} \to \mathbb{R}$, batch size S.

Initialize history $h_0 = 0$.
for $t = 1, 2, \ldots, T$ **do**
  if $t \mod S = 1$ **then**
    Let $N_t = \{1, \ldots, \lceil \frac{t}{S} \rceil\}$ denote the number of batches so far.
    For $b \in N_t$, let $T_b = \{(b-1)S + 1, \ldots, bS\}$ denote the rounds in batch b.
    For $b \in N_t$, let $g_b = \frac{1}{S} \sum_{s \in T_b} \tilde{f}_s$. denote the average of the functions in batch b.
    Learner chooses $x_t \in \arg\min_{x \in \mathcal{X}} \sum_{b \in N_t} g_b(x) + \frac{R(x)}{\eta}$.
  end
  else
    Learner chooses $x_t = x_{t-1}$.
  end
  Set $h_t = Ah_{t-1} + Bx_t$.
  Learner suffers loss $f_t(h_t)$ and observes f_t.
end

Theorem H.1. *Consider an online convex optimization with unbounded memory problem specified by $(\mathcal{X}, \mathcal{H}, A, B)$. Let the regularizer $R : \mathcal{X} \to \mathbb{R}$ be α-strongly-convex and satisfy $|R(x) - R(\tilde{x})| \le D$ for all $x, \tilde{x} \in \mathcal{X}$. Algorithm 2 with batch size S and step-size η satisfies*

$$R_T(\texttt{Mini-Batch FTRL}) \le \frac{SD}{\eta} + \eta \frac{T\tilde{L}^2}{\alpha} + \eta \frac{T L \tilde{L} H_1}{S\alpha}.$$

If $\eta = \sqrt{\frac{\alpha S D}{T\tilde{L}\left(\frac{LH_1}{S} + \tilde{L}\right)}}$, then

$$R_T(\texttt{Mini-Batch FTRL}) \le O\left(\sqrt{\frac{D}{\alpha} T \left(L\tilde{L}H_1 + S\tilde{L}^2\right)}\right).$$

Setting the batch size to be $S = LH_1/\tilde{L}$ we obtain the same upper bound on policy regret as Algorithm 1 while guaranteeing that the decisions x_t switch at most $T\tilde{L}/LH_1$ times.

Corollary H.1. *Consider an online convex optimization with unbounded memory problem specified by $(\mathcal{X}, \mathcal{H}, A, B)$. Let the regularizer $R : \mathcal{X} \to \mathbb{R}$ be α-strongly-convex and satisfy $|R(x) - R(\tilde{x})| \le D$*

for all $x, \tilde{x} \in \mathcal{X}$. Algorithm 2 with batch size $S = \frac{LH_1}{\tilde{L}}$ and step-size $\eta = \sqrt{\frac{\alpha SD}{T\tilde{L}\left(\frac{LH_1}{S}+\tilde{L}\right)}}$ satisfies

$$R_T(\texttt{Mini-Batch FTRL}) \leq O\left(\sqrt{\frac{D}{\alpha}TL\tilde{L}H_1}\right).$$

Furthermore, the decisions x_t switch at most $\frac{T\tilde{L}}{LH_1}$ times.

Intuitively, in the OCO with unbounded memory framework each decision x_t is penalized not just in round t but in future rounds as well. Therefore, instead of immediately changing the decision, it is prudent to stick to it for a while, collect more data, and then switch decisions. For the OCO with finite memory problem, the constant memory length m provides a natural measure of how long decisions penalized for and when one should switch decisions. In the general case, this is measured by the quantity $LH_1/\tilde{L}$. Note that this simplifies to m for OCO with finite memory for all p-norms.

Proof of Theorem H.1. For simplicity, assume that T is a multiple of S. Otherwise, the same proof works after replacing $\frac{T}{S}$ with $\lceil \frac{T}{S} \rceil$. Let $x^* \in \arg\min_{x\in\mathcal{X}} \sum_{t=1}^{T} \tilde{f}_t(x)$. Note that we can write the regret as

$$\begin{aligned} R_T(\texttt{Mini-Batch FTRL}) &= \sum_{t=1}^{T} f_t(h_t) - \min_{x\in\mathcal{X}} \sum_{t=1}^{T} \tilde{f}_t(x) \\ &= \underbrace{\sum_{t=1}^{T} f_t(h_t) - \tilde{f}_t(x_t)}_{(a)} + \underbrace{\sum_{t=1}^{T} \tilde{f}_t(x_t) - \tilde{f}_t(x^*)}_{(b)}. \end{aligned}$$

We can bound the term (b) using Theorem B.1 for mini-batches [Dekel et al., 2012, Altschuler and Talwar, 2018, Chen et al., 2020] by

$$\frac{SD}{\eta} + \eta\frac{T\tilde{L}^2}{\alpha}.$$

It remains to bound term (a). Let $N = T/S$ denote the number of batches and $T_n = \{(n-1)S + 1, \ldots, nS\}$ denote the rounds in batch $n \in [N]$. We can write

$$\begin{aligned} \sum_{t=1}^{T} f_t(h_t) - \tilde{f}_t(x_t) &= \sum_{t=1}^{T} f_t\left(\sum_{k=0}^{t-1} A^k B x_{t-k}\right) - f_t\left(\sum_{k=0}^{t-1} A^k B x_t\right) && \text{by Definition 2.1} \\ &\leq L\sum_{t=1}^{T}\left\|\sum_{k=0}^{t-1} A^k B x_{t-k} - \sum_{k=0}^{t-1} A^k B x_t\right\| && \text{by Assumption A4} \\ &\leq \frac{T}{S}L\underbrace{\sum_{t\in T_N}\left\|\sum_{k=0}^{t-1} A^k B x_{t-k} - \sum_{k=0}^{t-1} A^k B x_t\right\|}_{(c)}, \end{aligned}$$

where the last inequality follows because of the following. Consider rounds $t_1 = b_1 S + r$ and $t_2 = b_2 S + r$ for $b_1 < b_2$ and $r \in [S]$. Then, $\|h_{t_1} - \sum_{k=0}^{t_1-1} A^k B x_{t_1}\| \leq \|h_{t_2} - \sum_{k=0}^{t_2-1} A^k B x_{t_2}\|$ because the latter sums over more terms in its history and decisions in consecutive batches have distance bounded above by $\eta\tilde{L}/\alpha$ (Theorem B.1). Therefore, it suffices to show that term (c) is upper bounded by $\eta\tilde{L}H_1/\alpha$. We have

$$\begin{aligned} \sum_{t\in T_N}\left\|\sum_{k=0}^{t-1} A^k B x_{t-k} - \sum_{k=0}^{t-1} A^k B x_t\right\| &\leq \sum_{t\in T_N}\sum_{k=0}^{t-1}\left\|A^k B x_{t-k} - A^k B x_t\right\| \\ &\leq \sum_{t\in T_N}\sum_{k=0}^{t-1}\|A^k\|\|B\|\|x_{t-k} - x_t\| \\ &\leq \sum_{t\in T_N}\sum_{k=0}^{t-1}\|A^k\|\|x_{t-k} - x_t\| && \text{by Assumption A2.} \end{aligned}$$

Since the same decision x_n is chosen in all rounds of batch n, we can reindex and rewrite

$$\begin{aligned}\sum_{t\in T_N}\left\|\sum_{k=0}^{t-1}A^kBx_{t-k}-\sum_{k=0}^{t-1}A^kBx_t\right\| &\le \sum_{t\in T_N}\sum_{k=0}^{t-1}\|A^k\|\|x_{t-k}-x_t\| \\ &\le \sum_{o=0}^{S-1}\sum_{n=1}^{N-1}\sum_{s=1}^{S}\|A^{(N-n-1)S+s+o}\|\|x_N-x_n\| \\ &\le \eta\frac{\tilde{L}}{\alpha}\sum_{o=0}^{S-1}\sum_{n=1}^{N-1}\sum_{s=1}^{S}(N-n)\|A^{(N-n-1)S+s+o}\| \\ &= \eta\frac{\tilde{L}}{\alpha}\sum_{o=0}^{S-1}\sum_{n=1}^{N-1}\sum_{s=1}^{S}n\|A^{(n-1)S+s+o}\|,\end{aligned}$$

where the last inequality follows from bounding the distance between decision in consecutive batches Theorem B.1 and the triangle inequality. Expanding the triple sum yields

$$\begin{aligned}&\sum_{o=0}^{S-1}\sum_{n=1}^{N-1}\sum_{s=1}^{S}n\|A^{(n-1)S+s+o}\| \\ &\le \|A\|+\cdots+\|A^S\|+2\|A^{S+1}\|+\cdots+2\|A^{2S}\|+3\|A^{2S+1}\|+\cdots+3\|A^{3S}\|+\dots \\ &\quad+\|A^2\|+\cdots+\|A^{S+1}\|+2\|A^{S+2}\|+\cdots+2\|A^{2S+1}\|+3\|A^{2S+2}\|+\cdots+3\|A^{3S+1}\|+\dots \\ &\quad\vdots \\ &\quad+\|A^S\|+\cdots+\|A^{2S-1}\|+2\|A^{2S}\|+\cdots+2\|A^{3S-1}\|+3\|A^{3S}\|+\cdots+3\|A^{4S-1}\|+\dots,\end{aligned}$$

where each line above corresponds to a value of $o\in\{0,\dots,S-1\}$. Adding up these terms yields H_1. This completes the proof. ■

Note that Theorem H.1 only provides an upper bound on the policy regret for the general case. Unlike Algorithm 1, it is unclear how to obtain a stronger bound depending on H_p for the case of linear sequence dynamics with the ξ-weighted p-norm for $p>1$. The above proof can be specialized for this special case, similar to the proofs of Theorem 2.1 and Lemma C.1, to obtain

$$\sum_{t\in T_N}\left\|\sum_{k=0}^{t-1}A^kBx_{t-k}-\sum_{k=0}^{t-1}A^kBx_t\right\|\le\eta\frac{\tilde{L}}{\alpha}\sum_{o=0}^{S-1}\left(\sum_{n=1}^{N-1}\sum_{s=1}^{S}\left(n\|A^{(n-1)S+s+o}\|\right)^p\right)^{\frac{1}{p}}$$

and

$$\begin{aligned}&\sum_{o=0}^{S-1}\left(\sum_{n=1}^{N-1}\sum_{s=1}^{S}\left(n\|A^{(n-1)S+s+o}\|\right)^p\right)^{\frac{1}{p}} \\ &\le\left(\|A\|^p+\cdots+\|A^S\|^p+2^p\|A^{S+1}\|^p+\dots 2^p\|A^{2S}\|^p+3^p\|A^{2S+1}\|^p+\dots\right)^{\frac{1}{p}} \\ &\quad+\left(\|A^2\|^p+\cdots+\|A^{S+1}\|^p+2^p\|A^{S+2}\|^p+\dots 2^p\|A^{2S+1}\|^p+3^p\|A^{2S+2}\|^p+\dots\right)^{\frac{1}{p}} \\ &\quad\vdots \\ &\quad\left(\|A^S\|^p+\cdots+\|A^{2S-1}\|^p+2^p\|A^{2S}\|^p+\dots 2^p\|A^{3S-1}\|^p+3^p\|A^{3S}\|^p+\dots\right)^{\frac{1}{p}}.\end{aligned}$$

The above expression cannot be easily simplified to $O(H_p)$. However, for the special case of OCO with finite memory, which follows linear sequence dynamics with the 2-norm, we can do so by leveraging the special structure of the linear operator $A_{\text{finite},m}$.

Theorem H.2. *Consider an online convex optimization with finite memory problem with constant memory length m specified by $(\mathcal{X},\mathcal{H}=\mathcal{X}^m,A_{\text{finite},m},B_{\text{finite},m})$. Let the regularizer $R:\mathcal{X}\to\mathbb{R}$ be α-strongly-convex and satisfy $|R(x)-R(\tilde{x})|\le D$ for all $x,\tilde{x}\in\mathcal{X}$. Algorithm 2 with batch size m*

and step-size $\eta = \sqrt{\frac{\alpha m D}{T\tilde{L}\left(Lm^{\frac{1}{2}}+\tilde{L}\right)}}$ *satisfies*

$$R_T(\texttt{Mini-Batch FTRL}) \leq O\left(\sqrt{\frac{D}{\alpha}TL\tilde{L}m^{\frac{3}{2}}}\right) \leq O\left(m\sqrt{\frac{D}{\alpha}TL^2}\right).$$

Furthermore, the decisions x_t *switch at most* $\frac{T}{m}$ *times.*

Proof. Given the proof of Theorem H.1 and the above discussion, it suffices to show that

$$\sum_{o=0}^{S-1}\left(\sum_{n=1}^{N-1}\sum_{s=1}^{S}\left(n\|A^{(n-1)S+s+o}\|\right)^2\right)^{\frac{1}{2}} \leq H_2 = m^{\frac{3}{2}}.$$

Recall that $\|A^k_{\text{finite}}\| = 1$ if $k \leq m$ and 0 otherwise. Using this and $S = m$, we have that the above sum is at most $\sqrt{m} + \sqrt{m-1} + \cdots + \sqrt{1} = O\left(m^{\frac{3}{2}}\right)$. This completes the proof. ■

I Experiments

In this section we present some simple simulation experiments.[2]

Problem Setup. We consider the problem of online linear control with a constant input controller class $\Pi = \{\pi_u : \pi(s) = u \in \mathcal{U}\}$. Let T denote the time horizon. Let $\mathcal{S} = \mathbb{R}^d$ and $\mathcal{U} = \{u \in \mathbb{R}^d : \|u\|_2 \leq 1\}$ denote the state and control spaces. Let s_t and u_t denote the state and control at time t with s_0 being the initial state. The system evolves according to linear dynamics $s_{t+1} = Fs_t + Gu_t + w_t$, where $F, G \in \mathbb{R}^{d\times d}$ are system matrices and $w_t \in \mathbb{R}^d$ is a disturbance. The loss function in round t is simply $c_t(s_t, u_t) = c_t(s_t) = \sum_{j=1}^{d} s_{t,j}$, where $s_{t,j}$ denotes the j-th coordinate of s_t. The goal is to choose a sequence of control inputs $u_0, \ldots, u_{T-1} \in \mathcal{U}$ to minimize the regret

$$\sum_{t=0}^{T-1} c_t(s_t, u_t) - \min_{u\in\mathcal{U}} \sum_{t=0}^{T-1} c_t(s_t^u, u),$$

where s_t^u denotes the state in round t upon choosing control input u in each round. Note that the state in round t can be written as

$$s_t = \sum_{k=1}^{t} F^k G u_{t-k} + \sum_{k=1}^{t} F^k w_{t-k}.$$

Therefore, we can formulate this problem as an OCO with unbounded memory problem by setting $\mathcal{X} = \mathcal{U}, \mathcal{H} = \{y \in \mathbb{R}^d : y = \sum_{k=0}^{t} F^k G u \text{ for some } u \in \mathcal{U} \text{ and } t \in \mathbb{N}\}, A(h) = Fh, B(x) = Gx$, and $f_t(h_t) = c_t(\sum_{k=1}^{t} F^k G u_{t-k} + \sum_{k=1}^{t} F^k w_{t-k})$. Note that $\mathcal{H}, A$, and B are all finite-dimensional.

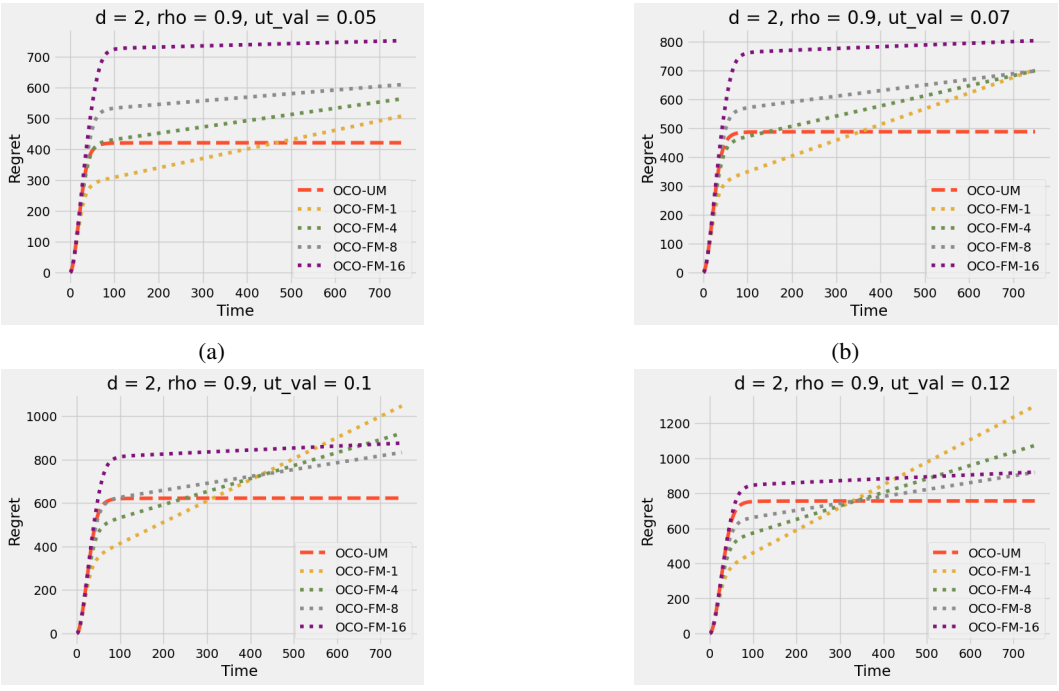
Data. We set the time horizon $T = 750$ and dimension $d = 2$. We sample the disturbances $\{w_t\}$ from a standard normal distribution. We set the system matrix G to be the identity and the system matrix F to be a diagonal plus upper triangular matrix with the diagonal entries equal to ρ and the upper triangular entries equal to α. We run simulations with various values of ρ and α.

Implementation. We use the `cvxpy` library [Diamond and Boyd, 2016, Agrawal et al., 2018] for implementing Algorithm 1. We use step-sizes according to Theorems 3.1 and 3.3. We run the experiments on a standard laptop.

[2] https://github.com/raunakkmr/oco-with-memory-code.

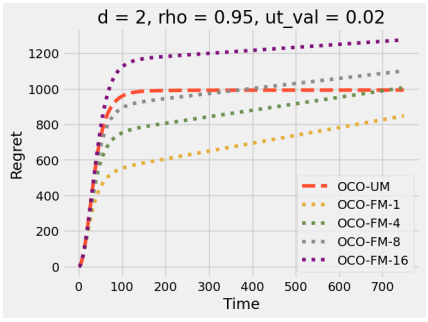
Results. We compare the regret with respect to the optimal control input of OCO with unbounded memory and OCO with finite memory for various memory lengths m in Fig. 3 for $\rho = 0.90$ and Fig. 4 for $\rho = 0.95$. There are a few important takeaways.

1. OCO with unbounded memory either performs as well as or better than OCO with finite memory, and it does so at comparable computational cost (Appendix G). In fact, the regret curve for OCO with unbounded memory reaches an asymptote whereas this is not the case for OCO with finite memory for a variety of memory lengths.
2. Knowledge of the spectral radius of F, ρ, is not sufficient to tune the memory length m for OCO with finite memory. This is illustrated by comparing Figs. 3a to 3d. Even though small memory lengths perform well when the upper triangular value is small, they perform poorly when the upper triangular value is large. In contrast, OCO with unbounded memory performs well in all cases.
3. For a fixed memory length, OCO with unbounded memory eventually performs better than OCO with finite memory. This is illustrated by comparing Figs. 3a to 3d.
4. As we increase the memory length, the performance of OCO with finite memory eventually approaches that of OCO with unbounded memory. However, an advantage of OCO with unbounded memory is that it does not require tuning the memory length. For example, when $\rho = 0.90$ and the upper triangular entry of $F = 0.10$, OCO with finite memory with $m = 4$ performs comparably to $m = 8$ and $m = 16$ (Fig. 3c). However, when the upper triangular entry of $F = 0.12$, then it performs much worse (Fig. 3d). However, OCO with unbounded memory performs well in all cases without the need for tuning an additional hyperparameter in the form of memory length.

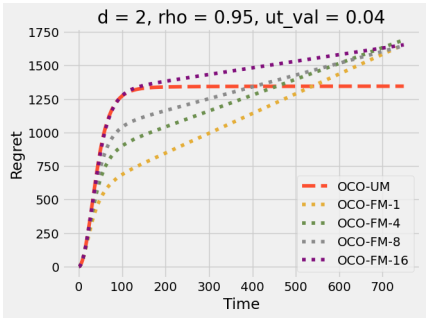


(a) (b) (c) (d)

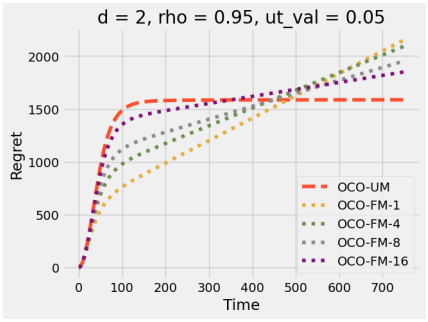
Figure 3: Regret plot for $\rho = 0.90$. The label `OCO-UM` refers to formulating the problem as an OCO with unbounded memory problem. The `OCO-FM-m` refers to formulating the problem as an OCO with finite memory problem with constant memory length m. The titles of the plots indicate the values of the dimension, the diagonal entries of F, and the upper triangular entries of F.



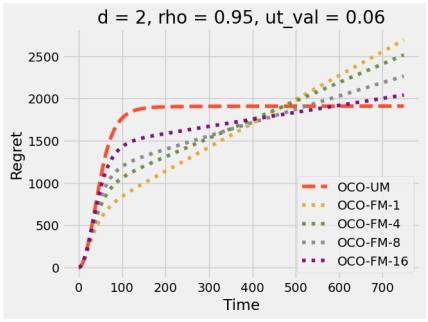
(a)



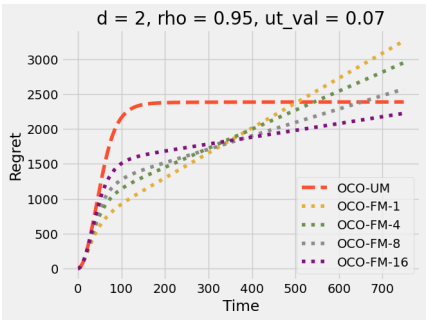
(b)



(c)



(d)



(e)

Figure 4: Regret plot for $\rho = 0.95$. The label `OCO-UM` refers to formulating the problem as an OCO with unbounded memory problem. The `OCO-FM-m` refers to formulating the problem as an OCO with finite memory problem with constant memory length m. The titles of the plots indicate the values of the dimension, the diagonal entries of F, and the upper triangular entries of F.