A Bandit OCO with memory

The proof consists of four parts: In A.1 we cover notation for functions and sets relevant to our analysis. In A.2, we cover some properties of the exploration noises u_t. In A.3, we prove a few important lemmas about the gradient estimator g_t. Finally, in A.4 we combine our lemmas from above with a reduction of the main theorem to obtain our main result.

A.1 Notation and Basic Results

Denote the ball and sphere of dimension d with radius r respectively as

$$\mathbb{B}_r^d \doteq \{x \in \mathbb{R}^d \,:\, \|x\| \,\leq\, r\}\,,\; \mathbb{S}_r^d \doteq \{x \in \mathbb{R}^d \,:\, \|x\| = r\}.$$

Consider a convex set $\mathcal{K} \subset \mathbb{R}^d$ bounded with diameter D and containing the unit ball $\mathbb{B}$.[1] For $0 < \delta < 1$, consider the Minkowski subset:

$$\mathcal{K}_\delta \doteq \{x \in \mathcal{K} \,:\, \frac{1}{1-\delta} x \in \mathcal{K}\},$$

and observe that $\mathcal{K}_\delta$ is convex and $\forall u \in B_1^d, x \in \mathcal{K}_\delta$ we have $x + \delta u \in \mathcal{K}$ because $\mathcal{K}$ contains the unit ball.

Next, we define the δ-smoothed version of a function $f : \mathbb{R}^d \to \mathbb{R}$ to be:

$$\hat{f}_\delta(x) \doteq \mathop{\mathbb{E}}_{v \sim \mathbb{B}} [f(x + \delta v)] \tag{A.1}$$

The following standard facts about the gradient of a smoothed function can be found in the literature, e.g. [15] Chapter 6 Lemma 6.7:

Fact A.1. *Let f be G-Lipschitz, and $\hat{f}_\delta$ as defined in eq. A.1. We then have:*

1. $\mathop{\mathbb{E}}_{u \sim \mathbb{S}} [f(x + \delta u)u] = \frac{\delta}{d} \nabla \hat{f}_\delta(x)$
2. $|\hat{f}_\delta(x) - f(x)| \,\leq\, \delta G,\; \forall x \in \mathcal{K}$

We additionally introduce the function $\tilde{f}_t : \mathcal{K} \to \mathbb{R}$ for loss functions with memory defined as:

$$\tilde{f}_t(x) \doteq f_t(\overbrace{x, \ldots, x}^{\times H})$$

Throughout our analysis, it will be helpful to denote the collection of vectors $(v_{t-n}, \ldots, v_t)$ by $v_{t-n:t}$. Using this notation, addition and scalar multiplication will also be compactly expressed as $v_{t-n:t} + \alpha\, w_{t-n:t} \doteq (v_{t-n} + \alpha w_{t-n}, \ldots, v_t + \alpha w_t)$. Because we are interested in loss functions with H inputs, we will mostly be interested in collections of the form $v_{t-H+1:t}$. To avoid the excessive use of $H \pm 1$ throughout the rest of the paper, we will introduce the notation $\bar{H} \doteq H - 1$.

We now introduce the index-wise gradients $\nabla_i f_t$ to be the derivative of f_t with respect to the i'th input vector, namely:

$$\nabla_i f_t(x_{t-\bar{H}:t}) = \frac{\partial f_t(x_{t-\bar{H}}, \ldots, x_t)}{\partial x_{t-\bar{H}+i}}$$

such that $\nabla f_t = (\nabla_0 f_t, \ldots, \nabla_{\bar{H}} f_t)$. We make the following observation about the gradients $\nabla_i f_t$.

Lemma A.2. *The gradient $\nabla \tilde{f}_t(x) = \frac{\partial \tilde{f}_t(x)}{\partial x}$ is related to the gradient of f_t by*

$$\nabla \tilde{f}_t(x) = \sum_{i=0}^{\bar{H}} \nabla_i f_t(x_{t-\bar{H}}, \ldots, x_t)\Big|_{x_{t-\bar{H}} = \ldots = x_t = x}$$

which we denote as $\nabla \tilde{f}_t(x) = \sum_{i=0}^{\bar{H}} \nabla_i \tilde{f}_t(x)$.

[1] We suppress the radius and dimensionality indices for $\mathbb{S}_1^d$ and $\mathbb{B}_1^d$ for the sake of presentation.

Proof. Applying chain rule over $f_t(x_{t-\bar{H}:t})$ with $x_{t-i}(x) = x,\ i = 0, \ldots, \bar{H}$ yields the product of the dH dimensional gradient $\frac{\partial f_t}{\partial x_{t-\bar{H}:t}}$ and the $dH \times d$ dimensional Jacobian $\frac{\partial x_{t-\bar{H}:t}}{\partial x}$, which is equal to H copies of the $d \times d$ identity matrix. Specifically,

$$\begin{aligned}
\nabla \tilde{f}_t(x) = \frac{\partial \tilde{f}_t(x)}{\partial x} &= \frac{\partial f_t(x_{t-\bar{H}:t})}{\partial x_{t-\bar{H}:t}}^\top \cdot \frac{\partial x_{t-\bar{H}:t}}{\partial x} \\
&= \begin{bmatrix} \frac{\partial f_t(x_{t-\bar{H}:t})}{\partial x_{t-\bar{H}}} \\ \vdots \\ \frac{\partial f_t(x_{t-\bar{H}:t})}{\partial x_t} \end{bmatrix}^\top \cdot \begin{bmatrix} I_d \\ \vdots \\ I_d \end{bmatrix} \\
&= \sum_{i=0}^{\bar{H}} \frac{\partial f_t(x_{t-\bar{H}:t})}{\partial x_{t-i}} = \sum_{i=0}^{\bar{H}} \nabla_i f_t(x_{t-\bar{H}}, \ldots, x_t)\Big|_{x_{t-\bar{H}}=\ldots=x_t=x}
\end{aligned}$$

where the derivatives $\frac{\partial f_t(x_{t-\bar{H}:t})}{\partial x_{t-\bar{H}:t}}$ are evaluated at $x_{t-\bar{H}} = \ldots = x_t = x$ implicitly on lines 2 through 4 for clarity. □

Finally, we denote the optimizer over $\mathcal{K}$ with respect to all observed loss functions as $x^\star = \arg\min_{x \in \mathcal{K}} \sum_{t=H}^{T} f_t(x, \ldots, x)$.

A.2 Properties of the random exploration noise

Claim A.3. (Independence) x_t is independent of $u_{t-\bar{H}}, \ldots, u_t$.

Proof. Base case: for $t \leq H$ all x_t's are set arbitrarily to be equal and so the conclusion is immediate. For $t \geq H$: Assume this holds for x_t and observe that $x_{t+1} = x_t + \eta_t g_{t-\bar{H}}$ is uniquely defined by x_t and $g_{t-\bar{H}}$, for which the latter satisfies

$$g_{t-\bar{H}} = \frac{d}{\delta} f_{t-\bar{H}}(x_{t-2\bar{H}:t-\bar{H}} + u_{t-2\bar{H}:t-\bar{H}}) \sum_{i=0}^{\bar{H}} u_{t-\bar{H}-i}.$$

Now, since $f_{t-\bar{H}}$ and $u_{t-2\bar{H}:t-\bar{H}}$ are sampled before $u_{t-\bar{H}+1:t+1}$, the random variables that uniquely determine g_t are independent from $u_{t-\bar{H}+1:t+1}$. Furthermore, by induction hypothesis x_t is independent of $u_{t-\bar{H}}, \ldots, u_t$ and clearly also of u_{t+1}. Thus, the components that uniquely define x_{t+1} are independent of $u_{t-\bar{H}+1:t+1}$, which means that x_{t+1} is independent of $u_{t-\bar{H}+1:t+1}$ as well, as desired. □

Remark. Claim A.3 above allows us to conclude that $u_{t-\bar{H}:t}$ is independent of $x_{t-\bar{H}:t}$, which crucially allows us to apply fact A.1 to our gradient estimator g_t.

Lemma A.4. *The sum of $u_{t-\bar{H}}, \ldots, u_t$ for all t has expected squared norm less than or equal to H.*

Proof. Since $u_t \in_R \mathbb{S}\ \forall t$, we have $\mathbb{E}[u_i \cdot u_j] = 0$ whenever $i \neq j$, hence

$$\begin{aligned}
\mathbb{E}\left[\left\|\sum_{i=0}^{\bar{H}} u_{t-i}\right\|^2\right] &= \mathbb{E}\left[\left(\sum_{i=0}^{\bar{H}} u_{t-i}\right) \cdot \left(\sum_{i=0}^{\bar{H}} u_{t-i}\right)\right] \\
&= \mathbb{E}\left[\sum_{i=0}^{\bar{H}} \|u_{t-i}\|^2\right] + \mathbb{E}\left[\sum_{i \neq j} u_{t-i} \cdot u_{t-j}\right] \\
&= H
\end{aligned}$$

□

A.3 Properties of the gradient estimator

The goal of this section is to prove a lemma showing that our gradient estimator g_t is a valid estimator of $\nabla \tilde{f}_t(x_t)$ by bounding the difference in expectation between the two, as well as bounding the norm of g_t itself. We recall our previous assumptions that the loss functions are bounded by one, have gradients bounded by $||\nabla f_t|| \leq G$ (which is equivalent to f_t being G-Lipschitz), and have hessians bounded by $||\nabla^2 f_t|| \leq \beta$ (which is equivalent to f_t being β-smooth).

We start by bounding the expected square norm of our gradient estimator.

Lemma A.5. *The gradient estimator g_t satisfies $\mathbb{E}\left[\|g_t\|^2\right] \leq \frac{d^2 H}{\delta^2}$.*

Proof. Combining lemma A.4 with $f_t(y_{t-\bar{H}:t}) \leq 1$ and the definition of g_t, it follows that

$$\begin{aligned}
\mathbb{E}\left[\|g_t\|^2\right] &= \mathbb{E}_{u_{t-\bar{H}:t}}\left[\left\|\frac{d}{\delta} f_t\left(x_{t-\bar{H}:t} + \delta u_{t-\bar{H}:t}\right) \cdot \sum_{i=0}^{\bar{H}} u_{t-i}\right\|^2\right] \\
&= \mathbb{E}\left[\frac{d^2}{\delta^2} f_t(y_{t-\bar{H}:t})^2 \left\|\sum_{i=0}^{\bar{H}} u_{t-i}\right\|^2\right] \\
&\leq \frac{d^2}{\delta^2}\mathbb{E}\left[\left\|\sum_{i=0}^{\bar{H}} u_{t-i}\right\|^2\right] \\
&\leq \frac{d^2 H}{\delta^2}.
\end{aligned}$$

Remark: Even if the losses f_t are bounded by some constant $M > 1$, the results for our algorithm and proofs still hold if one scales down the gradient estimator to $\frac{1}{M} g_t$, only adding a factor M. □

Using the lemma above, we can now bound the distance between our predictions as follows:

Lemma A.6. *For $x_0, \ldots, x_T$ selected according to Algorithm 1, we have that:*

$$\mathbb{E}\left[\left\|x_{t-\bar{H}:t} - (x_{t+\bar{H}}, \ldots, x_{t+\bar{H}})\right\|^2\right] \leq 4\eta_{t-\bar{H}}^2 \frac{d^2 H^4}{\delta^2}.$$

Proof. Starting with the first inequality, since $x_{t+1} = \prod_{\mathcal{K}_\delta}[x_t - \eta_t g_{t-\bar{H}}]$, we have that:

$$\begin{aligned}
\mathbb{E}\left[\left\|(x_{t-\bar{H}}, \ldots, x_t) - (x_{t+\bar{H}}, \ldots, x_{t+\bar{H}})\right\|^2\right] &= \mathbb{E}\left[\sum_{i=0}^{\bar{H}} \left\|x_{t+\bar{H}} - x_{t-i}\right\|^2\right] \\
&\leq \mathbb{E}\left[\sum_{i=0}^{\bar{H}} \left(\sum_{j=1}^{i+\bar{H}} \left\|x_{t+\bar{H}-j+1} - x_{t+\bar{H}-j}\right\|\right)^2\right] && (\triangle\text{-ineq.}) \\
&\leq \mathbb{E}\left[\sum_{i=0}^{\bar{H}} \left(\sum_{j=1}^{i+\bar{H}} \eta_{t+\bar{H}-j} \left\|g_{t-j}\right\|\right)^2\right] && (\text{projection property}) \\
&\leq \mathbb{E}\left[\eta_{t-\bar{H}}^2 \sum_{i=0}^{\bar{H}} \left(\sum_{j=1}^{i+\bar{H}} \|g_{t-j}\|\right)^2\right] && (\eta_t \text{ decreasing}) \\
&\leq \eta_{t-\bar{H}}^2 \sum_{i=0}^{\bar{H}} (2\bar{H})^2 \cdot \frac{d^2 H}{\delta^2} && (\text{C-S \& lemma A.5}) \\
&\leq 4\eta_{t-\bar{H}}^2 \frac{d^2 H^4}{\delta^2}
\end{aligned}$$

□

Corollary A.7. *We also have that*

$$\mathbb{E}\left[\left\|x_{t-\bar{H}:t} - (x_{t+\bar{H}}, \ldots, x_{t+\bar{H}})\right\|\right] \;\leq\; 2\eta_{t-\bar{H}}\frac{dH^2}{\delta}.$$

Proof. This is an immediate consequence of lemma A.6 since $\mathbb{E}[\|X\|]^2 \;\leq\; \mathbb{E}[\|X\|^2]$. □

We continue by proving our desired properties about the estimator g_t. We first observe the following property for *linear* δ-smoothed functions.

Lemma A.8. *For f **linear** and satisfying our assumptions, we have that*

$$\mathop{\mathbb{E}}_{u_{t-\bar{H}:t}\sim\bigoplus_{t=1}^{H}\mathbb{S}}\left[\frac{d}{\delta}f(x_{t-\bar{H}:t}+\delta u_{t-\bar{H}:t})u_{t-\bar{H}:t}\right] = \nabla f(x_{t-\bar{H}:t})$$

Proof. By the independence of $x_{t-\bar{H}:t}$ and $u_{t-\bar{H}:t}$ (Claim A.3), we can apply Fact A.1 to each index $i = 0, \ldots, \bar{H}$ and obtain

$$\begin{aligned}
\mathop{\mathbb{E}}_{u_{t-\bar{H}:t}\sim\bigoplus_{t=1}^{H}\mathbb{S}}\left[f(x_{t-\bar{H}:t}+\delta u_{t-\bar{H}:t})u_{t-i}\right] &= \mathop{\mathbb{E}}_{u_{t-i}\sim\mathbb{S}}\left[\mathop{\mathbb{E}}_{u_{[t-\bar{H}:t]\setminus\{t-i\}}\sim\bigoplus_{t=1}^{H-1}\mathbb{S}}\left[f(x_{t-\bar{H}:t}+\delta u_{t-\bar{H}:t})u_{t-i}\right]\right] \\
&= \mathop{\mathbb{E}}_{u_{t-i}\sim\mathbb{S}}\left[f(x_{t-\bar{H}:t}+\delta(\mathbf{0},\ldots,u_{t-i},\ldots,\mathbf{0}))u_{t-i}\right] \\
&= \frac{\delta}{d}\nabla_{\bar{H}-i}\hat{f}_\delta(x_{t-\bar{H}:t}) \\
&= \frac{\delta}{d}\nabla_{\bar{H}-i}f(x_{t-\bar{H}:t})
\end{aligned}$$

where the second and last lines follow by the linearity of f, the symmetry of the sphere and the fact that expectation commutes with linear operators. Since $\nabla f(x_{t-\bar{H}:t}) = (\nabla_0 f(x_{t-\bar{H}:t}), \ldots, \nabla_{\bar{H}} f(x_{t-\bar{H}:t}))$, the lemma then follows. □

Using the theorem above, we can generalize Fact A.1 in the following manner:

Theorem A.9. *For general convex f satisfying our assumptions, we have:*

$$\left\|\mathop{\mathbb{E}}_{u_{t-\bar{H}:t}\sim\bigoplus_{t=1}^{H}\mathbb{S}}\left[\frac{d}{\delta}f(x_{t-\bar{H}:t}+\delta u_{t-\bar{H}:t})u_{t-\bar{H}:t}\right] - \nabla f(x_{t-\bar{H}:t})\right\| \;\leq\; \frac{d\delta}{2}H^{3/2}$$

Proof. Consider the linear function $\bar{f}_{x_{t-\bar{H}:t}}(z_{t-\bar{H}:t}) = f(x_{t-\bar{H}:t}) + \nabla f(x_{t-\bar{H}:t})(z_{t-\bar{H}:t} - x_{t-\bar{H}:t})$. By lemma A.8 above,

$$\begin{aligned}
\mathop{\mathbb{E}}_{u_{t-\bar{H}:t}\sim\bigoplus_{t=1}^{H}\mathbb{S}}\left[\frac{d}{\delta}\bar{f}_{x_{t-\bar{H}:t}}(x_{t-\bar{H}:t}+\delta u_{t-\bar{H}:t})u_{t-\bar{H}:t}\right] &= \nabla\bar{f}_{x_{t-\bar{H}:t}}(x_{t-\bar{H}:t}) \\
&= \nabla f(x_{t-\bar{H}:t}).
\end{aligned}$$

The lemma then follows when we bound the difference between f and $\bar{f}_{x_{t-\bar{H}:t}}$ such that:

$$\begin{aligned}
&\left\|\mathbb{E}\left[\frac{d}{\delta}f(x_{t-\bar{H}:t}+\delta u_{t-\bar{H}:t})u_{t-\bar{H}:t}\right]-\nabla f(x_{t-\bar{H}:t})\right\| \\
\leq\ &\left\|\mathbb{E}\left[\frac{d}{\delta}f(x_{t-\bar{H}:t}+\delta u_{t-\bar{H}:t})u_{t-\bar{H}:t}\right]-\mathbb{E}\left[\frac{d}{\delta}\bar{f}_{x_{t-\bar{H}:t}}(x_{t-\bar{H}:t}+\delta u_{t-\bar{H}:t})u_{t-\bar{H}:t}\right]\right\| \\
\leq\ &\mathbb{E}\left[\frac{d}{\delta}|f(x_{t-\bar{H}:t}+\delta u_{t-\bar{H}:t})-\bar{f}_{x_{t-\bar{H}:t}}(x_{t-\bar{H}:t}+\delta u_{t-\bar{H}:t})|\left\|u_{t-\bar{H}:t}\right\|\right] && \text{(Jensen)} \\
\leq\ &\mathbb{E}\left[\frac{d}{\delta}\frac{\beta}{2}\left\|\delta u_{t-\bar{H}:t}\right\|^2\left\|u_{t-\bar{H}:t}\right\|\right] && \text{(Taylor \& } \left\|\nabla^2\right\| \text{ bound)} \\
\leq\ &\frac{d\delta}{2}H^{3/2}
\end{aligned}$$

where the expectations are taken over $u_{t-\bar{H}:t}\sim\bigoplus_{t=1}^{H}\mathbb{S}$. □

Corollary A.10. *g_t satisfies:*

$$\left\|\mathbb{E}[g_t]-\sum_{i=0}^{\bar{H}}\nabla_i f_t(x_{t-\bar{H}:t})\right\|\ \leq\ \frac{d\delta}{2}H^2$$

where the expectation is over all randomness in the algorithm.

Proof. By the definition of g_t (line 10 of Algorithm 1), we have

$$\begin{aligned}
&\left\|\mathbb{E}[g_t]-\sum_{i=0}^{\bar{H}}\nabla_i f_t(x_{t-\bar{H}:t})\right\| \\
\leq\ &\left\|\mathbb{E}\left[\frac{d}{\delta}f(x_{t-\bar{H}:t}+\delta u_{t-\bar{H}:t})\sum_{i=0}^{\bar{H}}u_{t-i}\right]-\sum_{i=0}^{\bar{H}}\nabla_i f_t(x_{t-\bar{H}:t})\right\| \\
\leq\ &\sqrt{H}\left\|\mathbb{E}\left[\frac{d}{\delta}f(x_{t-\bar{H}:t}+\delta u_{t-\bar{H}:t})u_{t-\bar{H}:t}\right]-\nabla f(x_{t-\bar{H}:t})\right\| && (\triangle\text{-ineq. \& C-S}) \\
\leq\ &\frac{d\delta}{2}H^2. && \text{(lemma A.9)}
\end{aligned}$$

□

Lemma A.11. *We have that:*

$$\mathbb{E}\left[\left\|\sum_{i=0}^{\bar{H}}\nabla_i f_t(x_{t-\bar{H}:t})-\nabla\tilde{f}_t(x_{t+\bar{H}})\right\|\right]\ \leq\ 2\frac{\eta_{t-\bar{H}}\beta d H^{5/2}}{\delta}$$

Proof. Using the results derived thus far, we obtain:

$$\begin{aligned}
&\mathbb{E}\left[\left\|\sum_{i=0}^{\bar{H}}\nabla_i f_t(x_{t-\bar{H}:t}) - \nabla\tilde{f}_t(x_{t+\bar{H}})\right\|^2\right] \\
&= \mathbb{E}\left[\left\|\sum_{i=0}^{\bar{H}}\nabla_i f_t(x_{t-\bar{H}:t}) - \sum_{i=0}^{\bar{H}}\nabla_i f_t(x_{t+\bar{H}},\ldots,x_{t+\bar{H}})\right\|^2\right] && \text{(lemma A.2)} \\
&\leq H\mathbb{E}\left[\sum_{i=0}^{\bar{H}}\left\|\nabla_i f_t(x_{t-\bar{H}:t}) - \nabla_i f_t(x_{t+\bar{H}},\ldots,x_{t+\bar{H}})\right\|^2\right] && \text{(Cauchy-Schwarz)} \\
&= H\mathbb{E}\left[\left\|\nabla f_t(x_{t-\bar{H}},\ldots,x_t) - \nabla f_t(x_{t+\bar{H}},\ldots,x_{t+\bar{H}})\right\|^2\right] \\
&\leq H\beta^2\mathbb{E}\left[\left\|(x_{t-\bar{H}},\ldots,x_t) - (x_{t+\bar{H}},\ldots,x_{t+\bar{H}})\right\|^2\right] && (\beta\text{-smoothness}) \\
&\leq H\beta^2\frac{4\eta_{t-\bar{H}}^2 d^2 H^4}{\delta^2} && \text{(lemma A.6)} \\
&= 4\frac{\eta_{t-\bar{H}}^2\beta^2 d^2 H^5}{\delta^2}.
\end{aligned}$$

after which our result follows by $\mathbb{E}[\|X\|]^2 \leq \mathbb{E}[\|X\|^2]$. □

The lemmas above allow us to obtain our desired result regarding the gradient estimator g_t, presented below.

Corollary A.12. *The gradient estimator g_t satisfies:*

$$\mathbb{E}\left[\left\|\mathbb{E}[g_t] - \nabla\tilde{f}_t(x_{t+\bar{H}})\right\|\right] \leq \frac{d\delta}{2}H^2 + 2\frac{\eta_{t-\bar{H}}\beta d H^{5/2}}{\delta}$$

Proof. This follows from Corollary A.10 and Lemma A.11 due to the triangle inequality. □

A.4 Proof of Theorem 3.1

We start by performing a reduction from bounding the regret over $f_t(y_{t-\bar{H}:t})$ to that over $\tilde{f}_{t-\bar{H}}(x_t)$ against $x_\delta^\star = \Pi_{\mathcal{K}_\delta}(x^\star)$.

Lemma A.13. *We have that:*

$$\mathbb{E}\left[\sum_{t=H}^{T}\left(f_t(y_{t-\bar{H}:t}) - \tilde{f}_t(x^\star)\right)\right] - \mathbb{E}\left[\sum_{t=H}^{T}\left(\tilde{f}_{t-\bar{H}}(x_t) - \tilde{f}_{t-\bar{H}}(x_\delta^\star)\right)\right] \leq 3\delta G D H^{1/2}T + \frac{dGH^2}{\delta}\sum_{t=1}^{T-\bar{H}}\eta_t$$

Proof. First we look at $t = \overline{H, T-\bar{H}}$. By properties of projection, we have that $\|x^\star - x_\delta^\star\| \leq \delta D$ and hence G-Lipschitzness guarantees that $f_t(x^\star) - f_t(x_\delta^\star) \leq GD\delta$. Further,

$$\begin{aligned}
\mathbb{E}\left[\left(f_t(y_{t-\bar{H}:t}) - \tilde{f}_t(x_{t+\bar{H}})\right)\right] &= \mathbb{E}\left[\left(f_t(x_{t-\bar{H}:t} + \delta u_{t-\bar{H}:t}) - \tilde{f}_t(x_{t+\bar{H}})\right)\right] \\
&\leq G\,\mathbb{E}[||(x_{t-\bar{H}:t} + \delta u_{t-\bar{H}:t}) - (x_{t+\bar{H}},\ldots,x_{t+\bar{H}})||] \\
&\leq \eta_{t-\bar{H}}\frac{dGH^2}{\delta} + \delta G H^{1/2} && \text{(cor. A.7)}
\end{aligned}$$

summing over $t = \overline{H, T-\bar{H}}$ concludes the proof by the conveniency-motivated assumptions $D \geq 1, H \leq T, \delta \geq \frac{1}{G\sqrt{T}}$ (which are satisfied by our ultimate choice of parameters). □

We now move on to bounding $\tilde{f}_{t-\bar{H}}(x_t) - \tilde{f}_{t-\bar{H}}(x_\delta^\star)$.

Observation A.14. *If we denote by $\mathbb{E}$ the expectation over the u_t's and apply the law of total expectation, we have that:*

$$\mathbb{E}\left[(\mathbb{E}[g_{t-\bar{H}}] - g_{t-\bar{H}}) \cdot (x_t - x_\delta^\star)\right] = \mathbb{E}\left[\mathbb{E}[(\mathbb{E}[g_{t-\bar{H}}] - g_{t-\bar{H}}) \cdot (x_t - x_\delta^\star) \mid (u_0, \dots, u_{t-\bar{H}})]\right] = 0$$

Observation A.15. *By convexity of $\tilde{f}_{t-\bar{H}}$, we have that:*

$$\tilde{f}_{t-\bar{H}}(x_t) - \tilde{f}_{t-\bar{H}}(x_\delta^\star) \;\leq\; \nabla\tilde{f}_{t-\bar{H}}(x_t)^\top (x_t - x_\delta^\star)$$

Lemma A.16. *The delayed regret against $x_\delta^\star$ in terms of $\tilde{f}$ satisfies:*

$$\mathbb{E}\left[\sum_{t=H}^{T} \tilde{f}_{t-\bar{H}}(x_t)\right] - \sum_{t=H}^{T} \tilde{f}_{t-\bar{H}}(x_\delta^\star) \;\leq\; \frac{D^2}{2\eta_T} + \left(\frac{d^2 H}{2\delta^2} + \frac{2d\beta D H^{5/2}}{\delta}\right)\sum_{t=1}^{T}\eta_t + \frac{d\delta}{2}H^2 DT + HGD$$

Proof. Observe that:

$$\begin{aligned}
\|x_{t+1} - x_\delta^\star\|^2 &= \left\|\Pi_{\mathcal{K}_\delta}[x_t - \eta_t g_{t-\bar{H}}] - x_\delta^\star\right\|^2 \\
&\leq \left\|x_t - \eta_t g_{t-\bar{H}} - x_\delta^\star\right\|^2 && \text{(Pythagoras)} \\
&= \|x_t - x_\delta^\star\|^2 + \left\|\eta_t g_{t-\bar{H}}\right\|^2 - 2\eta_t g_{t-\bar{H}}^\top \cdot (x_t - x_\delta^\star) \\
\Rightarrow \qquad 2g_{t-\bar{H}}^\top \cdot (x_t - x_\delta^\star) &\leq \frac{\|x_t - x_\delta^\star\|^2 - \|x_{t+1} - x_\delta^\star\|^2}{\eta_t} + \eta_t \left\|g_{t-\bar{H}}\right\|^2 && \text{(A.2)}
\end{aligned}$$

Therefore, we get:

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=H}^{T} \tilde{f}_{t-\bar{H}}(x_t)\right] - \sum_{t=H}^{T} \tilde{f}_{t-\bar{H}}(x_\delta^\star) &= \mathbb{E}\left[\sum_{t=H}^{T}\left(\tilde{f}_{t-\bar{H}}(x_t) - \tilde{f}_{t-\bar{H}}(x_\delta^\star)\right)\right] \\
&\leq \mathbb{E}\left[\sum_{t=H}^{T} \nabla\tilde{f}_{t-\bar{H}}(x_t)^\top (x_t - x_\delta^\star)\right] \\
&= \mathbb{E}\left[\sum_{t=H}^{T}\left(g_{t-\bar{H}} + (\mathbb{E}[g_{t-\bar{H}}] - g_{t-\bar{H}}) + (\nabla\tilde{f}_{t-\bar{H}}(x_t) - \mathbb{E}[g_{t-\bar{H}}])\right)^\top (x_t - x_\delta^\star)\right]
\end{aligned}$$

By equation (A.2), observation A.14 and Cauchy-Schwarz, we have:

$$\begin{aligned}
&\leq \frac{1}{2}\mathbb{E}\left[\sum_{t=H}^{T}\left(\frac{\|x_t - x_\delta^\star\|^2 - \|x_{t+1} - x_\delta^\star\|^2}{\eta_t} + \eta_t \left\|g_{t-\bar{H}}\right\|^2\right)\right] + 0 \\
&\quad + \mathbb{E}\left[\sum_{t=H}^{T}\left\|\nabla\tilde{f}_{t-\bar{H}}(x_t) - \mathbb{E}[g_{t-\bar{H}}]\right\| \cdot \|x_t - x_\delta^\star\|\right] \\
&\leq \frac{1}{2}\mathbb{E}\left[\sum_{t=H+1}^{T}\|x_t - x_\delta^\star\|^2\left(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}}\right) + \frac{||x_H - x_\delta^\star||^2}{\eta_H}\right] + \frac{d^2 H}{2\delta^2}\cdot\sum_{t=H}^{T}\eta_t && \text{(lem. A.5)} \\
&\quad + \sum_{t=H}^{T}\left(\frac{d\delta}{2}H^2 + 2\frac{\eta_{t-\bar{H}}\beta d H^{5/2}}{\delta}\right)\cdot D + HGD && \text{(lem. A.12)}
\end{aligned}$$

where we used $g_t = 0$ for all $t < H$ and $\left\|\nabla\tilde{f}\right\| \leq G$. Since η_t is a decreasing sequence we have:

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=H}^{T} \tilde{f}_{t-\bar{H}}(x_t)\right] - \sum_{t=H}^{T} \tilde{f}_{t-\bar{H}}(x_\delta^\star) \;\leq\; &\frac{D^2}{2\eta_T} + \frac{d^2 H}{2\delta^2}\cdot\sum_{t=H}^{T}\eta_t + \frac{d\delta}{2}H^2 DT \\
&+ \frac{2d\beta D H^{5/2}}{\delta}\sum_{t=1}^{T}\eta_t + HGD
\end{aligned}$$

□

We are now able to conclude our main proof.

Theorem 3.1. Setting step sizes $\eta_t = \Theta(t^{-3/4}H^{-3/2}d^{-1}D^{2/3}G^{-2/3}\beta^{-1/2})$ and perturbation constant $\delta = \Theta(T^{-1/4}H^{-1/2}D^{1/3}G^{-1/3})$, Algorithm 1 produces a sequence $\{y_t\}_{t=0}^{T}$ that satisfies:

$$\text{Regret} \;\leq\; \mathcal{O}\left(T^{3/4}H^{3/2}dD^{4/3}G^{2/3}\beta^{1/2}\right)$$

Proof. Putting A.13 and A.16 together, we get:

$$\begin{aligned}\text{Regret} &= \mathbb{E}\left[\sum_{t=H}^{T}\left(f_t(y_{t-\bar{H}:t}) - \tilde{f}_t(x^\star)\right)\right] \\ &\leq \frac{3D^2}{2\eta_T} + \left(\frac{d^2H}{2\delta^2} + \frac{3d\beta DGH^{5/2}}{\delta}\right)\sum_{t=1}^{T}\eta_t + d\delta H^2DT + 3\delta GDH^{1/2}T + HGD\end{aligned}$$

Noting that $\sum_{t=1}^{T}\frac{1}{t^{3/4}} \leq 4T^{1/4}+1$, setting the parameters as specified yields the desired result, concluding the proof of Theorem 3.1. □

B Regret Analysis for Known Systems

Proof. Observe that, if we fix $x_{t-\bar{H}}$ (the state starting $\bar{H}$ time steps back) and the observed disturbances $w_{t-2\bar{H}-1}, \ldots, w_t$, then the state x_t and action u_t at $\bar{H}$ time steps later are uniquely determined by the sequence of H policies $M_{t-\bar{H}}, \ldots M_t$, which means that $c_t(x_t, u_t)$ can be considered as an implicit functions of the past H policies played. It then follows that $\forall c_t$, $\exists$ unique f_t such that:

$$f_t(M_{t-\bar{H}}, \ldots M_t) \equiv c_t\left(x_t(M_{t-\bar{H}:t}), u_t(M_{t-\bar{H}:t})|x_{t-H}, w_{t-2\bar{H}-1:t}\right).$$

Due to the analysis by [4], sections 4.3 and 4.4, we know that f_t is convex with respect to $M_{t-\bar{H}}, \ldots, M_t$ when $x_{t-\bar{H}}$, K, and the perturbations w_t are fixed. Furthermore, because c_t is Lipschitz and smooth, f_t is G'-Lipschitz and β'-smooth as well, for some G', β'. This means we can successfully apply the approach in Algorithm 1 to our current setting. Therefore, by Theorem 3.1 we get that for any fixed initial (κ, γ)-stable K, if we denote the actions taken by Algorithm 2 as $u_0^K, \ldots, u_T^K$, and $M^* = \arg\min_{M\in\mathcal{M}}\sum_{t=H}^{T} c_t(x_t^K(M), u_t^K(M))$ the best DAC policy in hindsight, then we have that:

$$\begin{aligned}&\mathbb{E}\left[\sum_{t=0}^{T} c_t(x_t^K, u_t^K)\right] - \sum_{t=0}^{T} c_t(x_t^K(M^\star), u_t^K(M^\star)) \\ &\leq H + \frac{D^2}{2\eta_T} + \left(\frac{d^2H}{2\delta^2} + \frac{3d\beta' DG'H^{5/2}}{\delta}\right)\sum_{t=1}^{T}\eta_t + d\delta H^2DT + 3\delta G'DH^{1/2}T + HG'D\end{aligned}$$

where $d = Hmn$ because each policy M_t consists of H matrices of dimension $m \times n$. Setting $H = \Theta(\log T)$ and the other parameters as in 3.1, we get $J_T(BPC) - J_T(M^*) \leq \mathcal{O}(T^{3/4}\log^{5/2}T)$, where the factor $\log^{5/2}T$ follows from $d = \Theta(H)$ and $H = \Theta(\log T)$. Due to the exponential decay of the component norms of elements in $\mathcal{M}$, we can treat all other quantities as constants.

□

C Regret Analysis for Unknown Systems

Proof. We split the regret incurred by Algorithm 4, which we will denote by $\mathcal{A}$, into:

$$\text{Regret} \;= \text{Regret}_1 + \text{Regret}_2 + \text{Regret}_3$$

where the first term corresponds to the regret of the system identification phase, the second term to the regret of algorithm 2 relative to the optimal DAC policy $M^\star$, and the final term to the difference between the performance of $M^\star$ on the estimated and true dynamics. Specifically, for $M^\star \doteq \arg\min_{M\in\mathcal{M}}[J(M|A, B, w)]$ we have:

$$\text{Regret}_1 = J_{T_0}(\text{System identification}) \tag{C.1}$$

$$\text{Regret}_2 = J_{T-T_0}(\mathcal{A}|\hat{A}, \hat{B}, \hat{w}) - J_{T-T_0}(M^\star|\hat{A}, \hat{B}, \hat{w}) \tag{C.2}$$

$$\text{Regret}_3 = J_{T-T_0}(M^\star|\hat{A}, \hat{B}, \hat{w}) - J_{T-T_0}(M^\star|A, B, w). \tag{C.3}$$

By Lemma 20 in [16], the cost incurred during the system identification phase adds up to $\text{Regret}_1 = \mathcal{O}(T_0) = \mathcal{O}(T^{2/3}\log\hat{\delta}^{-1}) = \mathcal{O}(T^{2/3}\log T)$, and since the regret incurred by the second phase of the algorithm has an $\mathcal{O}(T^{3/4}\log^{5/2}T)$ bound, Regret_1 is insignificant to our final result.

Next, since $J(M^\star|\hat{A},\hat{B},\hat{w}) \geq \min_{M\in\mathcal{M}} J(M|\hat{A},\hat{B},\hat{w})$ and phase 2 corresponds to running Algorithm 2 on $\hat{A},\hat{B}$ by the Simulation Lemma, Theorem 5.1 implies:

$$\text{Regret}_2 \leq \mathcal{O}\left(T^{3/4}\log^{5/2}T\right)$$

We now move on to Regret_3. Let A, B denote the true, unknown dynamics and let $\hat{A},\hat{B}$ be output of Phase 1 after T_0 exploration rounds. By Theorem 19 in [16], with probability $1-\hat{\delta}$, we have that:

$$\left\|A-\hat{A}\right\|_F, \left\|B-\hat{B}\right\|_F \leq \varepsilon_{A,B} \tag{C.4}$$

where $T_0 = \Theta\left(\varepsilon_{A,B}^{-2}\log\hat{\delta}^{-1}\right)$. Our choice of T_0 therefore implies that $\varepsilon_{A,B} = \Theta\left(T^{-1/3}\log^{-1/2}\hat{\delta}^{-1}\right)$. Now, by our assumptions on the bound on the perturbations there exists a constant ε_w such that $\|w_t-\hat{w}_t\| \leq \varepsilon_w$. Observe that if $\hat{A},\hat{B}$ satisfy C.4, then:

$$\begin{aligned}
\|w_t-\hat{w}_t\| &= \left\|(x_{t+1}-Ax_t-Bu_t)-(x_{t+1}-\hat{A}x_t-\hat{B}u_t)\right\| \\
&\leq \left\|A-\hat{A}\right\|\cdot\|x_t\| + \left\|B-\hat{B}\right\|\cdot\|u_t\| && (\triangle\text{-inequality}) \\
&= \mathcal{O}(\varepsilon_{A,B})
\end{aligned}$$

since by assumption x_t and u_t are bounded, which means that the smallest value for ε_w satisfies $\varepsilon_w = \mathcal{O}(\varepsilon_{A,B})$. By Lemma 17 in [16] and the formula of state evolution, it follows that for any $M\in\mathcal{M}$:

$$\begin{aligned}
|J(M|\hat{A},\hat{B},\hat{w}) - J(M|A,B,w)| &\leq |J(M|\hat{A},\hat{B},\hat{w}) - J(M|A,B,\hat{w})| + |J(M|A,B,\hat{w}) - J(M|A,B,w)| \\
&\leq \mathcal{O}(T(\varepsilon_w+\varepsilon_{A,B})) \\
&\leq \mathcal{O}(T^{2/3}\log^{-1/2}\hat{\delta}^{-1})
\end{aligned}$$

with probability $1-\hat{\delta}$, and hence $\text{Regret}_3 = \mathcal{O}(T^{2/3})$ with probability $1-\hat{\delta}$ as well.

Adding up everything we get that with probability $1-\hat{\delta}$:

$$\text{Regret} \leq \mathcal{O}\left(T^{2/3}\log\hat{\delta}^{-1} + T^{3/4}\log^{5/2}T + T^{2/3}\log^{-1/2}\hat{\delta}^{-1}\right).$$

With at most probability $\hat{\delta}$ we obtain worst-case regret of $\mathcal{O}(T)$ since our costs are bounded. Thus we can set $\hat{\delta} = \Theta(T^{-1})$ and obtain our final regret bound:

$$\begin{aligned}
\text{Regret} &\leq \mathcal{O}\left(T^{2/3}\log\hat{\delta}^{-1} + T^{3/4}\log^{5/2}T + +T^{2/3}\log^{-1/2}\hat{\delta}^{-1} + \hat{\delta}T\right) \\
&\leq \mathcal{O}(T^{3/4}\log^{5/2}T).
\end{aligned}$$

□

Remark C.1. *We see that Algorithm 4 enjoys the same regret bound as Algorithm 2 despite acting in an unknown system. This is because both the regret incurred during exploration and the difference in performance between the $\hat{A},\hat{B}$-optimal DAC and the true optimal DAC are of lower order than the regret incurred by Algorithm 2.*

Remark C.2. *Our general results from Section 3 are also suitable for the policy parametrization of [32]. Under this alternate parametrization, one can overcome the need for controllability for the case of unknown systems (and require stabilizability and detectability only instead). We leave the precise implementation of this remark to future work.*