Appendix

A Proofs

Proof of Proposition 1. Let π be parametrized by x. We prove the sufficient conditions by showing that $A_{\pi^*|t}(s,a)$ is strongly convex in a for all $s \in \mathbb{S}$, which by the linear policy assumption implies $f_n(\pi)$ is strongly convex in x.

For the first case, since $Q_{\pi^*|t}(s,a) = c_t(s,a) + \mathbb{E}_{s'|s,a}[V_{\pi^*|t+1}(s')]$, given the constant assumption, it follows that

$$A_{\pi^*|t}(s,a) = Q_{\pi^*|t}(s,a) - V_{\pi^*|t}(s) = c_t(s,a) + \text{const.}$$

is strongly convex in terms of a.

For the second case, consider a system $ds = (f(s) + g(s)a)\,dt + h(s)dw$, where f, g, h are some matrix functions and dw is a Wiener process. By Hamilton-Jacobi-Bellman equation (Bertsekas et al., 1995), the advantage function can be written as

$$A_{\pi^*|t}(s,a) = c_t(s,a) + \partial_s V_{\pi^*|t}(s)^T g(s)a + r(s)$$

where $r(s)$ is some function in s. Therefore, $A_{\pi^*|t}(s,a)$ is strongly convex in a. ∎

Proof of Theorem 1. The proof is based on a basic perturbation lemma in convex analysis (Lemma 4), which for example can be found in (McMahan, 2014), and a lemma for online learning (Lemma 5).

Lemma 4. *Let $\phi_1 : \mathbb{R}^d \mapsto \mathbb{R}\bigcup\{\infty\}$ be a convex function such that $x_1 = \arg\min_x \phi_t(x)$ exits. Let ψ be a function such that $\phi_2(x) = \phi_1(x) + \psi(x)$ is α-strongly convex with respect to $\|\cdot\|$. Let $x_2 = \arg\min_x \phi_2(x)$. Then, for any $g \in \partial\psi(x_1)$, we have*

$$\|x_1 - x_2\| \leq \frac{1}{\alpha}\|g\|_*$$

and for any x'

$$\phi_2(x_1) - \phi_2(x') \leq \frac{1}{2\alpha}\|g\|_*^2$$

When ϕ_1 and ψ are quadratics (with ψ possibly linear) the above holds with equality.

Lemma 5. *Let $l_t(x)$ be a sequence of functions. Denote $l_{1:t}(x) = \sum_{\tau=1}^t l_\tau(x)$. and let*

$$x_t^* = \arg\min_{x\in K} l_{1:t}(x)$$

Then for any sequence $\{x_1, \dots, x_T\}$, $\tau \geq 1$, and any $x^* \in K$, it holds

$$\sum_{t=\tau}^T l_t(x_t) \leq l_{1:T}(x_T^*) - l_{1:\tau-1}(x_{\tau-1}^*) + \sum_{t=\tau}^T l_{1:t}(x_t) - l_{1:t}(x_t^*)$$

Proof. Introduce a slack loss function $l_0(\cdot) = 0$ and define $x_0^* = 0$ for index convenience. This does not change the optimum, since $l_{0:t}(x) = l_{1:t}(x)$.

$$\begin{aligned}
\sum_{t=\tau}^T l_t(x_t) &= \sum_{t=\tau}^T l_{0:t}(x_t) - l_{0:t-1}(x_t) \\
&\leq \sum_{t=\tau}^T l_{0:t}(x_t) - l_{0:t-1}(x_{t-1}^*) \\
&= l_{0:T}(x_T^*) - l_{0:\tau-1}(x_{\tau-1}^*) \\
&\quad + \sum_{t=\tau}^T l_{0:t}(x_t) - l_{0:t}(x_t^*)
\end{aligned}$$ ∎

Note Lemma 5 does not require l_t to be convex and the minimum to be unique.

To prove Theorem 1, we first note that by definition of $\hat{x}_N$, it satisfies $F(\hat{x}_N, \hat{x}_N) \leq \frac{1}{N}\sum_{n=1}^N f_n(x_n)$. To bound the average performance, we use Lemma 5 and write

$$\sum_{n=1}^N f_n(x_n) \leq f_{1:N}(x_{N+1}) + \sum_{n=1}^N f_{1:n}(x_n) - f_{1:n}(x_{n+1})$$

since $x_n = \arg\min_{x\in\mathcal{X}} f_{1:n-1}(x)$. Then because $f_{1:k}$ is $k\alpha$-strongly convex, by Lemma 4,

$$\sum_{n=1}^N f_n(x_n) \leq f_{1:N}(x_n^*) + \sum_{n=1}^N \frac{\|\nabla f_n(x_n)\|_*^2}{2\alpha n}.$$

Finally, dividing the upper-bound by n and using the facts that $\sum_{k=1}^n \frac{1}{k} \leq \ln(n) + 1$ and $\min a_i \leq \frac{1}{n}\sum a_i$ for any scalar sequence $\{a_n\}$, we have the desired result. ■

Proof of Theorem 3. Consider the example in Section 4. For this problem, $T = 2$, $J(x^*) = 0$, and $\tilde{\epsilon}_{\Pi,\pi^*} = 0$, implying $F(x,x) = \frac{1}{2}J(x) = \frac{1}{2}(\theta-1)^2 x^2$. Therefore, to prove the theorem, we focus on the lower bound of x_N^2.

Since $x_n = \arg\min_{x\in\mathcal{X}} f_{1:n-1}(x)$ and the cost is quadratic, we can write

$$\begin{aligned} x_{n+1} &= \arg\min_{x\in\mathcal{X}} f_{1:n}(x) \\ &= \arg\min_{x\in\mathcal{X}} (n-1)(x-x_n)^2 + (x-\theta x_n)^2 \\ &= (1 - \frac{1-\theta}{n})x_n \end{aligned}$$

If $\theta = 1$, then $x_N = x_1$ and the bound holds trivially. For general cases, let $p_n = \ln(x_n^2)$.

$$\begin{aligned} p_N - p_2 &= 2\sum_{n=2}^{N-1} \ln\left(1 - \frac{1-\theta}{n}\right) \\ &\geq -2(1-\theta)\sum_{n=2}^{N-1} \frac{1}{n-(1-\theta)} \end{aligned}$$

where the inequality is due to the fact that $\ln(1-x) \geq \frac{-x}{1-x}$ for $x < 1$. We consider two scenarios. Suppose $\theta < 1$.

$$\begin{aligned} p_N - p_2 &\geq -2(1-\theta)\int_1^{N-1} \frac{1}{x-(1-\theta)} dx \\ &= -2(1-\theta)\ln(x-(1-\theta))|_1^{N-1} \\ &= -2(1-\theta)\left(\ln(N+\theta-2) - \ln(\theta)\right) \\ &\geq -2(1-\theta)\ln(N+\theta-2) \end{aligned}$$

Therefore, $x_N^2 \geq x_2^2 (N+\theta-2)^{2(\theta-1)} \geq \Omega(N^{2(\theta-1)})$.

On the other hand, suppose $\theta > 1$.

$$\begin{aligned} p_N - p_2 &\geq 2(\theta-1)\int_2^N \frac{1}{x-(1-\theta)} dx \\ &= 2(\theta-1)\ln(x-(1-\theta))|_2^N \\ &= 2(\theta-1)\left(\ln(N-1+\theta) - \ln(1+\theta)\right) \end{aligned}$$

Therefore, $x_N^2 \geq x_2^2 (N-1+\theta)^{2(\theta-1)}(1+\theta)^{-2(\theta-1)} \geq \Omega(N^{2(\theta-1)})$. Substituting the lower bound on x_N^2 into the definition of $F(x,x)$ concludes the proof. ■

Proof of Corollary 1. To prove the corollary, we introduce a basic lemma

Lemma 6. *(Lan, 2013, Lemma 1) Let $\gamma_k \in (0,1)$, $k = 1, 2, \ldots$ be given. If the sequence $\{\Delta_k\}_{k\geq 0}$ satisfies*

$$\Lambda_{k+1} \leq (1-\gamma_k)\Lambda_k + B_k,$$

then

$$\Lambda_k \leq \Gamma_k + \Gamma_k \sum_{i=1}^{k} \frac{B_i}{\Gamma_{i+1}}$$

where $\Gamma_1 = \Lambda_1$ and $\Gamma_{k+1} = (1-\gamma_k)\Gamma_k$.

To bound the sequence $S_{m:n+1}$, we first apply Lemma 2. Fixed m, for any $n \geq m+1$, we have

$$\begin{aligned} S_{m:n+1} &\leq \left(1 - \frac{1}{n-m+1}\right) S_{m:n} + \|x_{n+1} - x_n\| \\ &\leq \left(1 - \frac{1}{n-m+1}\right) S_{m:n} + \frac{\theta}{n} S_n \\ &\leq \left(1 - \frac{1}{n-m+1}\right) S_{m:n} + \frac{\theta c}{n^{2-\theta}} \end{aligned}$$

where $c = S_2 e^{1-\theta}$.

Then we apply Lemma 6. Let $k = n - m + 1$ and define $R_k = S_{m:m+k-1} = S_{m:n}$ for $k \geq 2$. Then we rewrite the above inequality as

$$R_{k+1} \leq \left(1 - \frac{1}{k}\right) R_k + \frac{\theta c}{(k+m-1)^{2-\theta}}$$

and define

$$\Gamma_k := \begin{cases} 1, & k = 1 \\ (1 - \frac{1}{k-1})\Gamma_{k-1}, & k \geq 2 \end{cases}$$

By Proposition 2, the above conversion implies for some positive constant c,

$$R_2 = S_{m:m+1} = \|x_{m+1} - x_m\| \leq \frac{\theta S_m}{m} \leq \frac{\theta c}{m^{2-\theta}}$$

and $\Gamma_k \leq O(1/k)$ and $\frac{\Gamma_k}{\Gamma_i} \leq O(\frac{i}{k})$. Thus, by Lemma 6, we can derive

$$\begin{aligned} R_k &\leq \frac{1}{k} R_2 + O\left(\theta c \sum_{i=1}^{k} \frac{i}{k} \frac{1}{(i+m-1)^{2-\theta}}\right) \\ &\leq \frac{1}{k} R_2 + O\left(\frac{\theta c}{k} \frac{k}{\theta} \frac{1}{(m+k-1)^{1-\theta}}\right) \\ &= \frac{1}{k} R_2 + O\left(\frac{1}{(m+k-1)^{1-\theta}}\right) \\ &\leq \frac{1}{k} \frac{\theta c}{m^{2-\theta}} + O\left(\frac{1}{(m+k-1)^{1-\theta}}\right) = O(\frac{1}{n^{1-\theta}}) \end{aligned}$$

where we use the following upper bound in the second inequality

$$\begin{aligned}
\sum_{i=1}^{k} \frac{i}{(i+m-1)^{2-\theta}} &\leq \int_0^k \frac{x}{(x+m-1)^{2-\theta}} dx \\
&= \left. \frac{m+(1-\theta)x-1}{\theta(1-\theta)(m+x-1)^{1-\theta}} \right|_0^k \\
&= \frac{(1-\theta)k+m-1}{\theta(1-\theta)(m+k-1)^{1-\theta}} - \frac{m-1}{\theta(1-\theta)(m-1)^{1-\theta}} \\
&= \frac{k}{\theta} \frac{1}{(m+k-1)^{1-\theta}} + \frac{m-1}{\theta(1-\theta)} \left(\frac{1}{(m+k-1)^{1-\theta}} - \frac{1}{(m-1)^{1-\theta}} \right) \\
&\leq \frac{k}{\theta} \frac{1}{(m+k-1)^{1-\theta}} \qquad \blacksquare
\end{aligned}$$

Proof of Lemma 3. Define $\delta_{\pi|t}$ such that $d_{\pi|t;q}(s) = (1-q^t)\delta_{\pi|t}(s) + q^t d_{\pi^*}(s)$, and define $g_{z|t}(s) = \nabla_z \mathbb{E}_\pi[Q_{\pi^*|t}](s)$, for π parametrized by z; then by assumption, $\|g_{z|t}\|_* < G_2$. Let π, π' be two policies parameterized by $x, y \in \mathcal{X}$, respectively. Then

$$\begin{aligned}
&\|\nabla_2 \hat{F}(x,z) - \nabla_2 \hat{F}(y,z)\|_* \\
&= \|\mathbb{E}_{d_{\tilde{\pi}}}[g_{z|t}] - \mathbb{E}_{d_{\tilde{\pi}'}}[g_{z|t}]\|_* \\
&= \|\frac{1}{T}\sum_{t=0}^{T-1} (1-q^t)(\mathbb{E}_{\delta_{\pi|t;q}}[g_{z|t}] - \mathbb{E}_{\delta_{\pi'|t;q}}[g_{z|t}])\|_* \\
&\leq (1-q^T)\frac{1}{T}\sum_{t=0}^{T-1} \|\mathbb{E}_{\delta_{\pi|t;q}}[g_{z|t}] - \mathbb{E}_{\delta_{\pi'|t;q}}[g_{z|t}]\|_* \\
&\leq (1-q^T)\frac{2G_2}{T}\sum_{t=0}^{T-1} \|\delta_{\pi|t;q} - \delta_{\pi'|t;q}\|_1 \\
&\leq (1-q^T)\frac{2G_2}{T}\sum_{t=0}^{T-1} \|d_{\pi|t} - d_{\pi'|t}\|_1 \\
&\leq (1-q^T)\beta\|x-y\|
\end{aligned}$$

in which the second to the last inequality is because the divergence between $d_{\pi|t}$ and $d_{\pi'|t}$ is the largest among all state distributions generated by the mixing policies. ■

Proof of Corollary 2. The proof is similar to Lemma 3 and the proof of (Ross et al., 2011, Theorem 4.1). ■

B Analysis of AggreVaTe in Stochastic Problems

Here we give the complete analysis of the convergence of AGGREVATE in stochastic problems using finite-sample approximation. For completeness, we restate the results below: Let $f(x;s) = \mathbb{E}_\pi[A_{\pi^*|t}]$ (i.e. $f_n(x) = \mathbb{E}_{d_{\pi_n}}[f(x;s)]$, where policy π is a policy parametrized by x. Instead of using $f_n(\cdot)$ as the per-round cost in the nth iteration, we use consider its finite samples approximation $g_n(\cdot) = \sum_{k=1}^{m_n} f(\cdot; s_{n,k})$, where m_n is the number of independent samples collected in the nth iteration.

Theorem 4. *In addition to Assumptions 5 and 6, assume $f(x;s)$ is α-strongly convex in x and $\|f(x;s)\|_* < G_2$ almost surely. Let $\theta = \frac{\beta}{\alpha}$ and suppose $m_n = m_0 n^r$ for some $r \geq 0$. For all $N > 0$, with probability at least $1-\delta$,*

$$\begin{aligned}
F(x_N, x_N) \leq \tilde{\epsilon}_{\Pi,\pi^*} &+ \tilde{O}\left(\frac{\theta^2}{c}\frac{\ln(1/\delta) + C_{\mathcal{X}}}{N^{\min\{r,2,2-2\theta\}}}\right) \\
&+ \tilde{O}\left(\frac{\ln(1/\delta) + C_{\mathcal{X}}}{cN^{\min\{2,1+r\}}}\right)
\end{aligned}$$

where $c = \frac{\alpha}{G_2^2 m_0}$ *and* $C_{\mathcal{X}}$ *is a constant*[9] *of the complexity of* Π.

B.1 Uniform Convergence of Vector-Valued Martingales

To prove Theorem 4, we first introduces several concentration inequalities of vector-valued martingales by (Hayes, 2005) in Section B.1.1. Then we prove some basic lemmas regarding the convergence the stochastic dynamical systems of $\nabla g_n(x)$ specified by AggreVaTe in Section B.1.2 and B.1.3. Finally, the lemmas in these two sections are extended to provide uniform bounds, which are required to prove Theorem 4. In this section, we will state the results generally without limiting ourselves to the specific functions used in AggreVaTe.

B.1.1 Generalization of Azuma-Hoeffding Lemma

First we introduce two theorems by Hayes (2005) which extend Azuma-Hoeffding lemma to vector-valued martingales but without dependency on dimension.

Theorem 5. *(Hayes, 2005, Theorem 1.8) Let* $\{X_n\}$ *be a (very-weak) vector-valued martingale such that* $X_0 = 0$ *and for every* n, $\|X_n - X_{n-1}\| \leq 1$ *almost surely. Then, for every* $a > 0$, *it holds*

$$\Pr(\|X_n\| \geq a) < 2e \exp\left(\frac{-(a-1)^2}{2n}\right)$$

Theorem 6. *(Hayes, 2005, Theorem 7.4) Let* $\{X_n\}$ *be a (very-weak) vector-valued martingale such that* $X_0 = 0$ *and for every* n, $\|X_n - X_{n-1}\| \leq c_n$ *almost surely. Then, for every* $a > 0$, *it holds*

$$\Pr(\|X_n\| \geq a) < 2 \exp\left(\frac{-(a-Y_0)^2}{2\sum_{i=1}^n c_i^2}\right)$$

where $Y_0 = \max\{1 + \max c_i, 2\max c_i\}$.

B.1.2 Concentration of i.i.d. Vector-Valued Functions

Theorem 5 immediately implies the concentration of approximating vector-valued functions with finite samples.

Lemma 7. *Let* $x \in \mathcal{X}$ *and let* $f(x) = \mathbb{E}_\omega[f(x;\omega)]$, *where* $f : \mathcal{X} \to E$ *and* E *is equipped with norm* $\|\cdot\|$. *Assume* $\|f(x;\omega)\| \leq G$ *almost surely. Let* $g(x) = \frac{1}{M}\sum_{m=1}^M f(x;\omega_k)$ *be its finite sample approximation. Then, for all* $\epsilon > 0$,

$$\Pr(\|g(x) - f(x)\| \geq \epsilon) < 2e \exp\left(-\frac{(\frac{M\epsilon}{2G} - 1)^2}{2M}\right)$$

In particular, for $0 < \epsilon \leq 2G$,

$$\Pr(\|g(x) - f(x)\| \geq \epsilon) < 2e^2 \exp\left(-\frac{M\epsilon^2}{8G^2}\right)$$

Proof. Define $X_m = \frac{1}{2G}\sum_{k=1}^m f(x;\omega_k) - f(x)$. Then X_m is vector-value martingale and $\|X_m - X_{m-1}\| \leq 1$. By Theorem 5,

$$\Pr(\|g(x) - f(x)\| \geq \epsilon) = \Pr(\|X_M\| \geq \frac{M\epsilon}{2G}) < 2e \exp\left(-\frac{(\frac{M\epsilon}{2G} - 1)^2}{2M}\right)$$

Suppose $\frac{\epsilon}{2G} < 1$. Then $\Pr(\|X_M\| \geq \epsilon) < 2e^2 \exp\left(-\frac{M\epsilon^2}{8G^2}\right)$. ■

[9] The constant $C_{\mathcal{X}}$ can be thought as $\ln|\mathcal{X}|$, where $|\mathcal{X}|$ measures the size of $\mathcal{X}$ in e.g. Rademacher complexity or covering number (Mohri et al., 2012). For example, $\ln|\mathcal{X}|$ can be linear in $\dim \mathcal{X}$.

B.1.3 Concentration of the Stochastic Process of AggreVaTe

Here we consider a stochastic process that shares the same characteristics of the dynamics of $\frac{1}{n}\nabla g_{1:n}(x)$ in AggreVaTe and provide a lemma about its concentration.

Lemma 8. *Let $n = 1 \dots N$ and $\{m_i\}$ be a non-decreasing sequence of positive integers. Given $x \in \mathcal{X}$, let $Y_n := \{f_n(x;\omega_{n,k})\}_{k=1}^{m_n}$ be a set of random vectors in some normed space with norm $\|\cdot\|$ defined as follows: Let $Y_{1:n} := \{Y_k\}_{k=1}^{n}$. Given $Y_{1:n-1}$, $\{f_n(x;\omega_{n,k})\}_{k=1}^{m_n}$ are m_n independent random vectors such that $f_n(x) := \mathbb{E}_\omega[f_n(x;\omega)|Y_{1:n-1}]$ and $\|f_n(x;\omega)\| \leq G$ almost surely. Define $g_n(x) := \frac{1}{m_n}\sum_{k=1}^{m_n} f_n(x;\omega_{n,k})$, and let $\bar{g}_n = \frac{1}{n}g_{1:n}$ and $\bar{f}_n = \frac{1}{n}f_{1:n}$. Then for all $\epsilon > 0$,*

$$\Pr(\|\bar{g}_n(x) - \bar{f}_n(x)\| \geq \epsilon) < 2\exp\left(\frac{-(nM^*\epsilon - Y_0)^2}{8G^2 M^{*2}\sum_{i=1}^{n}\frac{1}{m_i}}\right)$$

in which $M^* = \prod_{i=1}^{n} m_i$ and $Y_0 = \max\{1 + \frac{2M^*G}{m_0}, 2\frac{2M^*G}{m_0}\}$.

In particular, if $\frac{2M^*G}{m_0} > 1$, for $0 < \epsilon \leq \frac{Gm_0}{n}\sum_{i=1}^{n}\frac{1}{m_i}$,

$$\Pr(\|\bar{g}_n(x) - \bar{f}_n(x)\| \geq \epsilon) < 2e\exp\left(\frac{-n^2\epsilon^2}{8G^2\sum_{i=1}^{n}\frac{1}{m_i}}\right)$$

Proof. Let $M = \sum_{i=1}^{n} m_i$. Consider a martingale, for $m = l + \sum_{i=1}^{k-1} m_i$,

$$X_m = \frac{M^*}{m_k}\sum_{i=1}^{l} f_k(x;\omega_{k,i}) - f_k(x) + \sum_{i=1}^{k-1}\frac{M^*}{m_i}\sum_{j=1}^{m_i} f_i(x;\omega_{i,j}) - f_i(x).$$

That is, $X_M = nM^*(\bar{g}_n - \bar{f}_n)$ and $\|X_m - X_{m-1}\| \leq \frac{2M^*G}{m_i}$ for some appropriate m_i. Applying Theorem 6, we have

$$\Pr(\|\bar{g}_n - \bar{f}_n\| \geq \epsilon) = \Pr(\|X_M\| \geq nM^*\epsilon) < 2\exp\left(\frac{-(nM^*\epsilon - Y_0)^2}{2\sum_{m=1}^{M} c_m^2}\right)$$

where

$$\sum_{m=1}^{M} c_m^2 = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(\frac{2GM^*}{m_i}\right)^2 = 4G^2M^{*2}\sum_{i=1}^{n}\frac{1}{m_i}.$$

In addition, by assumption $m_i \leq m_{i-1}$, $Y_0 = \max\{1 + \frac{2M^*G}{m_0}, 2\frac{2M^*G}{m_0}\}$. This gives the first inequality.

For the special case, the following holds

$$\frac{-(nM^*\epsilon - Y_0)^2}{2\sum_{m=1}^{M} c_m^2} = \frac{-n^2M^{*2}\epsilon^2}{8G^2M^{*2}\sum_{i=1}^{n}\frac{1}{m_i}} + \frac{2nM^*\epsilon Y_0 - Y_0^2}{8G^2M^{*2}\sum_{i=1}^{n}\frac{1}{m_i}} \leq \frac{-n^2\epsilon^2}{4G^2\sum_{i=1}^{n}\frac{1}{m_i}} + 1$$

if ϵ satisfies

$$2nM^*\epsilon Y_0 < 8G^2M^{*2}\sum_{i=1}^{n}\frac{1}{m_i} \implies \epsilon < \frac{4G^2M^*}{Y_0 n}\sum_{i=1}^{n}\frac{1}{m_i}$$

Substituting the condition that $Y_0 = \frac{4M^*G}{m_0}$ when $\frac{2M^*G}{m_0} > 1$, a sufficient range of ϵ can be obtained as

$$\frac{4G^2M^*}{Y_0 n}\sum_{i=1}^{n}\frac{1}{m_i} = \frac{Gm_0}{n}\sum_{i=1}^{n}\frac{1}{m_i} \geq \epsilon.$$

■

B.1.4 Uniform Convergence

The above inequality holds for a particular $x \in \mathcal{X}$. Here we use the concept of covering number to derive uniform bounds that holds for all $x \in \mathcal{X}$. (Similar (and tighter) uniform bounds can also be derived using Rademacher complexity.)

Definition 1. Let S be a metric space and $\eta > 0$. The covering number $\mathcal{N}(S, \eta)$ is the minimal $l \in \mathcal{N}$ such that S is is covered by l balls of radius η. When S is compact, $\mathcal{N}(S, \eta)$ is finite.

As we are concerned with vector-valued functions, let E be a normed space with norm $\|\cdot\|$. Consider a mapping $f : \mathcal{X} \to \mathcal{B}$ defined as $f : x \mapsto f(x, \cdot)$, where $\mathcal{B} = \{g : \Omega \to E\}$ is a Banach space of vector-valued functions with norm $\|g\|_{\mathcal{B}} = \sup_{\omega \in \Omega} \|g(\omega)\|$. Assume $\mathcal{B}_{\mathcal{X}} = \{f(x, \cdot) : x \in \mathcal{X}\}$ is a compact subset in $\mathcal{B}$. Then the covering number of $\mathcal{H}$ is finite and given as $\mathcal{N}(\mathcal{B}_{\mathcal{X}}, \eta)$. That is, there exists a finite set $\mathcal{C}_{\mathcal{X}} = \{x_i \in \mathcal{X}\}_{i=1}^{\mathcal{N}(\mathcal{B}_{\mathcal{X}}, \eta)}$ such that $\forall x \in \mathcal{X}$, $\min_{y \in \mathcal{C}_{\mathcal{X}}} \|f(x, \cdot) - f(y, \cdot)\|_{\mathcal{B}} < \eta$.

Usually, the covering is a polynomial function of η. For example, suppose $\mathcal{X}$ is a ball of radius R in a d-dimensional Euclidean space, and f is L-Lipschitz in x (i.e. $\|f(x, \cdot) - f(y, \cdot)\|_{\mathcal{B}} \leq L\|x - y\|$). Then (Cucker and Zhou, 2007) $\mathcal{N}(\mathcal{B}_{\mathcal{X}}, \eta) \leq \mathcal{N}(\mathcal{X}, \frac{\eta}{L}) \leq \left(\frac{2RL}{\eta} + 1\right)^d$. Therefore, henceforth we will assume

$$\ln \mathcal{N}(\mathcal{B}_X X, \eta) \leq C_{\mathcal{X}} \ln(\frac{1}{\eta}) < \infty \tag{15}$$

for some constant $C_{\mathcal{X}}$ independent of η, which characterizes the complexity of $\mathcal{X}$.

Using covering number, we derive uniform bounds for the lemmas in Section B.1.2 and B.1.3.

Lemma 9. *Under the assumptions in Lemma 7, for* $0 < \epsilon \leq 2G$*,*

$$\Pr(\sup_{x \in \mathcal{X}} \|g(x) - f(x)\| \geq \epsilon) < 2e^2 \mathcal{N}(\mathcal{B}_{\mathcal{X}}, \frac{\epsilon}{4}) \exp\left(-\frac{M\epsilon^2}{32G^2}\right)$$

Proof. Choose $\mathcal{C}_{\mathcal{X}}$ be the set of the centers of the covering balls such that $\forall x \in \mathcal{X}$, $\min_{y \in \mathcal{C}_{\mathcal{X}}} \|f(x, \cdot) - f(y, \cdot)\|_{\mathcal{B}} < \eta$. Since $f(x) = \mathbb{E}_{\omega}[f(x, \omega)]$, it also holds $\min_{y \in \mathcal{C}_{\mathcal{X}}} \|f(x) - f(y)\| < \eta$. Let B_y be the η-ball centered for $y \in \mathcal{C}_{\mathcal{X}}$. Then

$$\begin{aligned}
\sup_{y \in \mathcal{X}} \|g(x) - f(x)\| &\leq \max_{y \in \mathcal{C}_{\mathcal{X}}} \sup_{x \in B_y} \|g(x) - g(y)\| + \|g(y) - f(y)\| + \|f(y) - f(x)\| \\
&\leq \max_{y \in \mathcal{C}_{\mathcal{X}}} \|g(y) - f(y)\| + 2\eta
\end{aligned}$$

Choose $\eta = \frac{\epsilon}{4}$ and then it follows that

$$\sup_{x \in \mathcal{X}} \|g(x) - f(x)\| \geq \epsilon \implies \max_{y \in \mathcal{C}_{\mathcal{X}}} \|g(y) - f(y)\| \geq \frac{\epsilon}{2}$$

The final result can be obtained by first for each $y \in \mathcal{C}_{\mathcal{X}}$ applying the concentration inequality with $\epsilon/2$ and then a uniform bound over $\mathcal{C}_{\mathcal{X}}$. ■

Similarly, we can give a uniform version of Lemma 8.

Lemma 10. *Under the assumptions in Lemma 8, if* $\frac{2M^* G}{m_0} > 1$*, for* $0 < \epsilon \leq \frac{Gm_0}{n} \sum_{i=1}^{n} \frac{1}{m_i}$ *and for a fixed* $n \geq 0$*,*

$$\Pr(\sup_{x \in \mathcal{X}} \|\bar{g}_n(x) - \bar{f}_n(x)\| \geq \epsilon) < 2e\mathcal{N}\left(\mathcal{B}_{\mathcal{X}}, \frac{\epsilon}{4}\right) \exp\left(\frac{-n^2\epsilon^2}{32G^2 \sum_{i=1}^{n} \frac{1}{m_i}}\right)$$

B.2 Proof of Theorem 4

We now refine Lemma 2 and Proposition 2 to prove the convergence of AggreVaTe in stochastic problems. We use $\bar{\cdot}$ to denote the average (e.g. $\bar{f}_n = \frac{1}{n} f_{1:n}$.)

B.2.1 Bound on $\|x_{n+1} - x_n\|$

First, we show the error due to finite-sample approximation.

Lemma 11. *Let $\xi_n = \nabla f_n - \nabla g_n$. Running* AGGREVATE *with $g_n(\cdot)$ as per-round cost gives, for $n \geq 2$,*

$$\|x_{n+1} - x_n\| \leq \frac{\theta S_n}{n} + \frac{1}{n\alpha}\left(\|\xi_n(x_n)\|_* + \|\bar{\xi}_{n-1}(x_n)\|_*\right)$$

Proof. Because $g_{1:n}(x)$ is $n\alpha$-strongly convex in x, we have

$$\begin{aligned}
n\alpha\|x_{n+1} - x_n\|^2 &\leq \langle \nabla g_n(x_n), x_n - x_{n+1}\rangle \\
&\leq \langle \nabla g_n(x_n) - \nabla \bar{g}_{n-1}(x_n), x_n - x_{n+1}\rangle && \because x_n = \arg\min_{x\in\mathcal{X}} g_{1:n-1}(x) \\
&\leq \|\nabla f_n(x_n) - \nabla \bar{f}_{n-1}(x_n)\|_* \|x_n - x_{n+1}\| \\
&\quad + \|\nabla f_n(x_n) - \nabla g_n(x_n) - \nabla \bar{f}_{n-1}(x_n) + \nabla \bar{g}_{n-1}(x_n)\|_* \|x_n - x_{n+1}\|
\end{aligned}$$

Now we use the fact that the smoothness applies to f (not necessarily to g) and derive the statement

$$\begin{aligned}
\|x_{n+1} - x_n\| &\leq \frac{\theta S_n}{n} + \frac{1}{n\alpha}\|\nabla f_n(x_n) - \nabla g_n(x_n) - \nabla \bar{f}_{n-1}(x_n) + \nabla \bar{g}_{n-1}(x_n)\|_* \\
&\leq \frac{\theta S_n}{n} + \frac{1}{n\alpha}\left(\|\xi_n(x_n)\|_* + \|\bar{\xi}_{n-1}(x_n)\|_*\right)
\end{aligned}$$ ■

Given the intermediate step in Lemma 11, we apply Lemma 5 to bound the norm of ξ_k and give the refinement of Lemma 2 for stochastic problems.

Lemma 12. *Suppose $m_n = m_0 n^r$ for some $r \geq 0$. Under previous assumptions, running* AGGREVATE *with $g_n(\cdot)$ as per-round cost, the following holds with probability at least $1-\delta$: For a fixed $n \geq 2$,*

$$\|x_{n+1} - x_n\| \leq \frac{\theta S_n}{n} + O\left(\frac{G_2}{n\alpha\sqrt{m_0}}\left(\sqrt{\frac{\ln(1/\delta)}{n^{\min\{r,2\}}}} + \sqrt{\frac{C_{\mathcal{X}}/n}{n^{\min\{r,1\}}}}\right)\right)$$

where $C_{\mathcal{X}}$ is a constant depending on the complexity of $\mathcal{X}$ and the constant term in big-O is some universal constant.

Proof. To show the statement, we bound $\|\xi_n(x_n)\|_*$ and $\|\bar{\xi}_{1:n-1}(x_n)\|_*$ in Lemma 11 using the concentration lemmas derived in Section B.1.4.

The First Term: To bound $\|\xi_n(x_n)\|_*$, because the sampling of ξ_n is independent of x_n, bounding $\|\xi_n(x_n)\|_*$ does not require a uniform bound. Here we use Lemma 7 and consider ϵ_1 such that

$$2e^2 \exp\left(-\frac{m_n \epsilon_1^2}{8G_2^2}\right) = \frac{\delta}{2} \implies \epsilon_1 = \sqrt{\frac{8G_2^2}{m_n}\ln\left(\frac{4e^2}{\delta}\right)} = O\left(\sqrt{\frac{G_2^2}{m_n}\ln\left(\frac{1}{\delta}\right)}\right) \tag{16}$$

Note we we used the particular range of ϵ in Lemma 7 for convenience, which is valid if we choose $m_0 > 2G_2 \ln\left(\frac{4e^2}{\delta}\right)$. This condition is not necessary; it is only used to simplify the derivation, and using a different range of ϵ would simply lead to a different constant.

The Second Term: To bound $\|\bar{\xi}_{n-1}(x_n)\|_*$, we apply a uniform bound using Lemma 10. For simplicity, we use the particular range $0 < \epsilon \leq \frac{G_2 m_0}{n}\sum_{i=1}^{n}\frac{1}{m_i}$ and assume $\frac{2M^* G_2}{m_0} > 1$ (which implies $Y_0 = \frac{4M^* G_2}{m_0}$) (again this is not necessary). We choose ϵ_2 such that

$$2e\mathcal{N}(\mathcal{B}_{\mathcal{X}}, \frac{\epsilon_2}{4}) \exp\left(\frac{-(n-1)^2\epsilon_2^2}{32G_2^2\sum_{i=1}^{n-1}\frac{1}{m_i}}\right) \leq \frac{\delta}{2} \implies \ln(2e) + \ln\mathcal{N}(\mathcal{B}_{\mathcal{X}}, \frac{\epsilon_2}{4}) + \frac{-(n-1)^2\epsilon_2^2}{32G_2^2\sum_{i=1}^{n-1}\frac{1}{m_i}} \leq -\ln(\frac{2}{\delta})$$

Since $\ln \mathcal{N}(\mathcal{B}_{\mathcal{X}}, \frac{\epsilon_2}{4}) = C_{\mathcal{X}} \ln\left(\frac{4}{\epsilon_2}\right) \leq c_s C_{\mathcal{X}} \epsilon_2^{-s}$ for arbitrary $s > 0$ and some c_s, a sufficient condition can be obtained by solving for ϵ_2 such that

$$\frac{c_0}{\epsilon_2^s} - c_2 \epsilon_2^2 = -c_1 \implies c_2 \epsilon_2^{2+s} - c_1 \epsilon_2^s - c_0 = 0$$

where $c_0 = c_s C_{\mathcal{X}}$, $c_2 = \frac{(n-1)^2}{32 G_2^2 \sum_{i=1}^{n-1} \frac{1}{m_i}}$, and $c_1 = \ln(\frac{4e}{\delta})$. To this end, we use a basic lemma of polynomials.

Lemma 13. *(Cucker and Zhou, 2007, Lemma 7.2) Let $c_1, c_2, \ldots, c_l > 0$ and $s > q_1 > q_2 > \cdots > q_{l-1} > 0$. Then the equation*

$$x^s - c_1 x^{q_1} - c_2 x^{q_2} - \cdots - c_{l-1} x^{q_{l-1}} - c_l = 0$$

has a unique solution x^*. In addition,

$$x^* \leq \max\left\{ (l c_1)^{1/(s-q_1)}, (l c_2)^{1/(s-q_2)}, \ldots, (l c_{l-1})^{1/(s-q_{l-1})}, (l c_1)^{1/s} \right\}$$

Therefore, we can choose an ϵ_2 which satisfies

$$\epsilon_2 \leq \max\left\{ \left(\frac{2c_1}{c_2}\right)^{1/2}, \left(\frac{2c_0}{c_2}\right)^{1/(2+s)} \right\} = \max\left\{ \left(\frac{64 \ln(\frac{4e}{\delta}) G_2^2 \sum_{i=1}^{n-1} \frac{1}{m_i}}{(n-1)^2}\right)^{1/2}, \left(\frac{64 c_s C_{\mathcal{X}} G_2^2 \sum_{i=1}^{n-1} \frac{1}{m_i}}{(n-1)^2}\right)^{1/(2+s)} \right\}$$
$$\leq O\left(\sqrt{\left(C_{\mathcal{X}} + \ln\left(\frac{1}{\delta}\right)\right) \frac{G_2^2}{n^2} \sum_{i=1}^{n} \frac{1}{m_i}} \right)$$

Error Bound Suppose $m_n = m_0 n^r$, for $r \geq 0$. Now we combine the two bounds above: fix $n \geq 2$, with probability at least $1 - \delta$,

$$\|\xi_n(x_n)\|_* + \|\bar{\xi}_{n-1}(x_n)\|_* \leq O\left(\sqrt{\frac{G_2^2}{m_0 n^r} \ln\left(\frac{1}{\delta}\right)} + \sqrt{\left(C_{\mathcal{X}} + \ln\left(\frac{1}{\delta}\right)\right) \frac{G_2^2}{m_0 n^2} \sum_{i=1}^{n} \frac{1}{i^r}} \right)$$

Due to the nature of harmonic series, we consider two scenarios.

1. If $r \in [0, 1]$, then the bound can be simplified as

$$O\left(\sqrt{\frac{G_2^2}{m_0 n^r} \ln\left(\frac{1}{\delta}\right)} + \sqrt{\left(C_{\mathcal{X}} + \ln\left(\frac{1}{\delta}\right)\right) \frac{G_2^2}{m_0 n^2} \sum_{i=1}^{n} \frac{1}{i^r}} \right)$$
$$= O\left(\sqrt{\frac{G_2^2}{m_0 n^r} \ln\left(\frac{1}{\delta}\right)} + \sqrt{\left(C_{\mathcal{X}} + \ln\left(\frac{1}{\delta}\right)\right) \frac{G_2^2 n^{1-r}}{m_0 n^2}} \right) = O\left(G_2 \sqrt{\frac{\ln(1/\delta)}{m_0 n^r}} + G_2 \sqrt{\frac{C_{\mathcal{X}}}{m_0 n^{1+r}}} \right)$$

2. If $r > 1$, then the bound can be simplified as

$$O\left(\sqrt{\frac{G_2^2}{m_0 n^r} \ln\left(\frac{1}{\delta}\right)} + \sqrt{\left(C_{\mathcal{X}} + \ln\left(\frac{1}{\delta}\right)\right) \frac{G_2^2}{m_0 n^2} \sum_{i=1}^{n} \frac{1}{i^r}} \right)$$
$$= O\left(\sqrt{\frac{G_2^2}{m_0 n^r} \ln\left(\frac{1}{\delta}\right)} + \sqrt{\left(C_{\mathcal{X}} + \ln\left(\frac{1}{\delta}\right)\right) \frac{G_2^2}{m_0 n^2}} \right) = O\left(G_2 \sqrt{\frac{\ln(1/\delta)}{m_0 n^{\min\{r,2\}}}} \right) + O\left(G_2 \sqrt{\frac{C_{\mathcal{X}}}{m_0 n^2}} \right)$$

Therefore, we conclude for $r \geq 0$,

$$\|\xi_n(x_n)\|_* + \|\bar{\xi}_{n-1}(x_n)\|_* = O\left(\sqrt{\frac{G_2^2 \ln(1/\delta)}{m_0 n^{\min\{r,2\}}}} + \sqrt{\frac{G_2^2 C_{\mathcal{X}}}{m_0 n^{1+\min\{r,1\}}}} \right)$$

Combining this inequality with Lemma 11 gives the final statement. ∎

B.2.2 Bound on S_n

Now we use Lemma 12 to refine Proposition 2 for stochastic problems.

Proposition 3. *Under the assumptions Proposition 2, suppose $m_n = m_0 n^r$. For a fixed $n \geq 2$, the following holds with probability at least $1-\delta$.*

$$S_n \leq \tilde{O}\left(\frac{G_2}{\alpha\sqrt{m_0}}\left(\frac{\sqrt{\ln(1/\delta)}}{n^{\min\{r/2,1,1-\theta\}}} + \frac{\sqrt{C_\mathcal{X}}}{n^{\min\{(1+r)/2,1,1-\theta\}}}\right)\right)$$

Proof. The proof is similar to that of Proposition 2, but we use the results from Lemma 12. Note Lemma 12 holds for a particular n. Here need the bound to apply for all $n = 1 \dots N$ so we can apply the bound for each S_n. This will add an additional $\sqrt{\ln N}$ factor to the bounds in Lemma 12.

First, we recall that

$$S_{n+1} \leq \left(1 - \frac{1}{n}\right) S_n + \|x_{n+1} - x_n\|$$

By Lemma 12, let $c_1 = \frac{G_2\sqrt{\ln(1/\delta)}}{n\alpha\sqrt{m_0}}$ and $c_2 = \frac{G_2\sqrt{C_\mathcal{X}}}{n\alpha\sqrt{m_0}}$, and it holds that

$$\|x_{n+1} - x_n\| \leq \frac{\theta S_n}{n} + O\left(\frac{G_2}{n\alpha\sqrt{m_0}}\left(\sqrt{\frac{\ln(1/\delta)}{n^{\min\{r,2\}}}} + \sqrt{\frac{C_\mathcal{X}}{n^{1+\min\{r,1\}}}}\right)\right) = \frac{\theta S_n}{n} + O(\frac{c_1}{n^{1+\min\{r,2\}/2}} + \frac{c_2}{n^{3/2+\min\{r,1\}/2}})$$

which implies

$$S_{n+1} \leq \left(1 - \frac{1}{n}\right) S_n + \|x_{n+1} - x_n\| \leq \left(1 - \frac{1-\theta}{n}\right) S_n + O(\frac{c_1}{n^{1+\min\{r,2\}/2}} + \frac{c_2}{n^{3/2+\min\{r,1\}/2}}).$$

Recall

Lemma 6. *(Lan, 2013, Lemma 1) Let $\gamma_k \in (0,1)$, $k = 1, 2, \dots$ be given. If the sequence $\{\Delta_k\}_{k\geq 0}$ satisfies*

$$\Lambda_{k+1} \leq (1-\gamma_k)\Lambda_k + B_k,$$

then

$$\Lambda_k \leq \Gamma_k + \Gamma_k \sum_{i=1}^{k} \frac{B_i}{\Gamma_{i+1}}$$

where $\Gamma_1 = \Lambda_1$ and $\Gamma_{k+1} = (1-\gamma_k)\Gamma_k$.

From Proposition 2, we know the unperturbed dynamics is bounded by $e^{1-\theta}n^{\theta-1}S_2$ (and can be shown in $\Theta(n^{\theta-1})$ as in the proof of Theorem 3). To consider the effect of the perturbations, due to linearity we can treat each perturbation separately and combine the results by superposition. Suppose a particular perturbation is of the form $O(\frac{C_2}{n^{1+s}})$ for some C_2 and $s > 0$. By Lemma 6, suppose $\theta + s < 1$,

$$S_n \leq O(n^{\theta-1}) + O\left(n^{\theta-1}\sum_{k=1}^{n} k^{1-\theta}\frac{C_2}{k^{1+s}}\right) \leq O(n^{\theta-1}) + O\left(C_2 n^{\theta-1} n^{1-s-\theta}\right) = O(n^{\theta-1}) + O\left(C_2 n^{-s}\right)$$

For $\theta - s = 1$, $S_n \leq O(n^{\theta-1}) + O(C_2 n^{\theta-1}\ln(n))$; for $\theta + s > 1$, $S_n \leq O(n^{\theta-1}) + O(C_2 n^{\theta-1})$. Therefore, we can conclude $S_n \leq C_1 n^{\theta-1} + \tilde{O}(C_2 n^{-\min\{s,1-\theta\}})$, where the constant $C_1 = e^{1-\theta}S_2$. Finally, using $S_2 \leq \frac{G_2}{\alpha}$ and setting C_2 as c_1 or c_2 gives the final result

$$S_n \leq \tilde{O}\left(\frac{G_2}{\alpha\sqrt{m_0}}\left(\frac{\sqrt{\ln(1/\delta)}}{n^{\min\{r/2,1,1-\theta\}}} + \frac{\sqrt{C_\mathcal{X}}}{n^{\min\{(1+r)/2,1,1-\theta\}}}\right)\right)$$

■

B.2.3 Performance Guarantee

Given Proposition 3, now we can prove the performance of the last iterate.

Theorem 4. *In addition to Assumptions 5 and 6, assume $f(x;s)$ is α-strongly convex in x and $\|f(x;s)\|_* < G_2$ almost surely. Let $\theta = \frac{\beta}{\alpha}$ and suppose $m_n = m_0 n^r$ for some $r \geq 0$. For all $N > 0$, with probability at least $1-\delta$,*

$$F(x_N, x_N) \leq \tilde{\epsilon}_{\Pi,\pi^*} + \tilde{O}\left(\frac{\theta^2}{c}\frac{\ln(1/\delta)+C_{\mathcal{X}}}{N^{\min\{r,2,2-2\theta\}}}\right) + \tilde{O}\left(\frac{\ln(1/\delta)+C_{\mathcal{X}}}{cN^{\min\{2,1+r\}}}\right)$$

where $c = \frac{\alpha}{G_2^2 m_0}$ and $C_{\mathcal{X}}$ is a constant[10] *of the complexity of* Π.

Proof. The proof is similar to the proof of Theorem 2. Let $x_n^* \coloneqq \arg\min_{x\in\mathcal{X}} f_n(x)$. Then

$$\begin{aligned}
f_n(x_n) - \min_{x\in\mathcal{X}} f_n(x) &\leq \langle \nabla f_n(x_n), x_n - x_n^*\rangle - \frac{\alpha}{2}\|x_n - x_n^*\|^2 \\
&\leq \langle \nabla f_n(x_n) - \nabla \bar{f}_{n-1}(x_n), x_n - x_n^*\rangle + \langle \nabla \bar{f}_{n-1}(x_n) - \nabla \bar{g}_{n-1}(x_n), x_n - x_n^*\rangle - \frac{\alpha}{2}\|x_n - x_n^*\|^2 \\
&\leq \|\nabla f_n(x_n) - \nabla \bar{f}_{n-1}(x_n)\|_*\|x_n - x_n^*\| + \|\nabla \bar{g}_{n-1}(x_n) - \nabla \bar{f}_{n-1}(x_n)\|_*\|x_n - x_n^*\| - \frac{\alpha}{2}\|x_n - x_n^*\|^2 \\
&\leq \frac{(\|\nabla f_n(x_n) - \nabla \bar{f}_{n-1}(x_n)\|_* + \|\nabla \bar{g}_{n-1}(x_n) - \nabla \bar{f}_{n-1}(x_n)\|_*)^2}{2\alpha} \\
&\leq \frac{\|\nabla f_n(x_n) - \nabla \bar{f}_{n-1}(x_n)\|_*^2 + \|\nabla \bar{g}_{n-1}(x_n) - \nabla \bar{f}_{n-1}(x_n)\|_*^2}{\alpha}
\end{aligned}$$

where the second inequality is due to $x_n = \arg\min_{x\in\mathcal{X}} \bar{g}_{n-1}(x)$. To bound the first term, recall the fact that $\|\nabla f_n(x_n) - \nabla \bar{f}_{n-1}(x_n)\|_* < \beta S_n$ and recall by Proposition 3 that

$$S_n \leq \tilde{O}\left(\frac{G_2}{\alpha\sqrt{m_0}}\left(\frac{\sqrt{\ln(1/\delta)}}{n^{\min\{r/2,1,1-\theta\}}} + \frac{\sqrt{C_{\mathcal{X}}}}{n^{\min\{(1+r)/2,1,1-\theta\}}}\right)\right)$$

For the second term, we use the proof in Lemma 12 with an additional $\ln(N)$ factor, i.e.

$$\|\nabla \bar{g}_{n-1}(x_n) - \nabla \bar{f}_{n-1}(x_n)\|_* = \tilde{O}\left(\frac{G_2}{\sqrt{m_0}}\sqrt{\frac{\ln(1/\delta)+C_{\mathcal{X}}}{n^{1+\min\{r,1\}}}}\right)$$

Let $c = \frac{\alpha m_0}{G_2^2}$. Therefore, combining all the results, we have the following with probability at least $1-\delta$:

$$\begin{aligned}
f_n(x_n) - \min_{x\in\mathcal{X}} f_n(x) &\leq \frac{\|\nabla f_n(x_n) - \nabla \bar{f}_{n-1}(x_n)\|_*^2 + \|\nabla \bar{g}_{n-1}(x_n) - \nabla \bar{f}_{n-1}(x_n)\|_*^2}{\alpha} \\
&\leq \frac{\beta^2 S_n^2}{\alpha} + \frac{\|\nabla \bar{g}_{n-1}(x_n) - \nabla \bar{f}_{n-1}(x_n)\|_*^2}{\alpha} \\
&\leq \tilde{O}\left(\frac{\theta^2 G_2^2}{\alpha m_0}\frac{\ln(1/\delta)}{n^{2\min\{r/2,1,1-\theta\}}}\right) + \tilde{O}\left(\frac{\theta^2 G_2^2}{\alpha m_0}\frac{C_{\mathcal{X}}}{n^{2\min\{(r+1)/2,1,1-\theta\}}}\right) + \tilde{O}\left(\frac{G_2^2}{\alpha m_0}\frac{\ln(1/\delta)+C_{\mathcal{X}}}{n^{1+\min\{r,1\}}}\right) \\
&= \tilde{O}\left(\frac{\theta^2}{c}\frac{\ln(1/\delta)}{n^{2\min\{r/2,1,1-\theta\}}}\right) + \tilde{O}\left(\frac{\theta^2}{c}\frac{C_{\mathcal{X}}}{n^{2\min\{(r+1)/2,1,1-\theta\}}}\right) + \tilde{O}\left(\frac{1}{c}\frac{\ln(1/\delta)+C_{\mathcal{X}}}{n^{1+\min\{r,1\}}}\right) \\
&\leq \tilde{O}\left(\frac{\theta^2}{c}\frac{\ln(1/\delta)+C_{\mathcal{X}}}{n^{2\min\{r/2,1,1-\theta\}}}\right) + \tilde{O}\left(\frac{\ln(1/\delta)+C_{\mathcal{X}}}{cn^{1+\min\{r,1\}}}\right)
\end{aligned}$$

Note the last inequality is unnecessary and is used to simplify the result. It can be seen that the upper bound originally has a weaker dependency on $C_{\mathcal{X}}$.

∎

[10] The constant $C_{\mathcal{X}}$ can be thought as $\ln|\mathcal{X}|$, where $|\mathcal{X}|$ measures the size of $\mathcal{X}$ in e.g. Rademacher complexity or covering number (Mohri et al., 2012). For example, $\ln|\mathcal{X}|$ can be linear in $\dim\mathcal{X}$.

C AggreVaTe with Function Approximations

Here we give a sketch of applying the techniques used in Theorem 4 to problems where a function approximator is used to learn $f(\cdot; s)$, as in the case considered by Ross et al. (2011) for learning the Q-function.

We consider a meta learning scenario where a linear function approximator $\hat{f}(x,s) = \phi(x,s)^T w$ is used to approximate $f(x;s)$. We assume $\phi(x,s)^T w$ satisfies Assumption 3 and Assumption 5 with some appropriate constants.

Now we analyze the case where $\sum_{i=1}^{m_n} \hat{f}(\cdot, s_{n,i})$ is used as the per-round cost in AggreVaTe. Specifically, in the nth iteration of AggreVaTe, m_n samples$\{f(x_n; s_{n,k})\}_{k=1}^{m_n}$ are first collected, and then w_n is updated by

$$w_n = \arg\min_{w \in \mathcal{W}} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \left(f(x_i; s_{i,j}) - \phi(x_i, s_{i,j})^T w \right)^2 \tag{17}$$

where $\mathcal{W}$ is the domain of w. Given the new w_n, the policy is updated by

$$x_{n+1} = \arg\min_{x \in \mathcal{X}} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \phi(x_i, s_{i,j})^T w_n \tag{18}$$

To prove the performance, we focus on the inequality used in the proof of performance in Theorem 4.

$$f_n(x_n) - \min_{x \in \mathcal{X}} f_n(x) \le \langle \nabla f_n(x_n), x_n - x_n^* \rangle - \frac{\alpha}{2} \|x_n - x_n^*\|^2$$

And we expand the inner product term:

$$\langle \nabla f_n(x_n), x_n - x_n^* \rangle = \langle \nabla \bar{g}_{n;w_{n-1}}(x_n), x_n - x_n^* \rangle + \langle \nabla \bar{g}_{n;w_n} - \nabla \bar{g}_{n;w_{n-1}}, x_n - x_n^* \rangle + \langle \nabla f_n(x_n) - \nabla \bar{g}_{n;w_n}, x_n - x_n^* \rangle$$

where $\bar{g}_{n;w_n}$ is the finite-sample approximation using w_n . By (18), $x_n = \arg\min_{x \in \mathcal{X}} \bar{g}_{n;w_{n-1}}(x)$, and therefore

$$\langle \nabla f_n(x_n), x_n - x_n^* \rangle \le \langle \nabla \bar{g}_{n;w_n} - \nabla \bar{g}_{n;w_{n-1}}, x_n - x_n^* \rangle + \langle \nabla f_n(x_n) - \nabla \bar{g}_{n;w_n}, x_n - x_n^* \rangle$$

In the first term, $\|\nabla \bar{g}_{n;w_n} - \nabla \bar{g}_{n;w_{n-1}}\|_* \le O(\|w_n - w_{n-1}\|)$. As w_n is updated by another value aggregation algorithm, this term can be further bounded similarly as in Lemma 2, by assuming a similar condition like Assumption 5 but on the change of the gradient in the objective function in (17). In the second term, $\|\nabla f_n(x_n) - \nabla \bar{g}_{n;w_n}\|_*$ can be bounded by the uniform bound of vector-valued martingale in Lemma 10. Given these two bounds, it follows that

$$f_n(x_n) - \min_{x \in \mathcal{X}} f_n(x) \le \frac{\|\nabla \bar{g}_{n;w_n} - \nabla \bar{g}_{n;w_{n-1}}\|_*^2 + \|\nabla f_n(x_n) - \nabla \bar{g}_{n;w_n}\|_*^2}{\alpha}$$

Compared with Theorem 4, since here additional Lipschitz constant is introduced to bound the change $\|\nabla \bar{g}_{n;w_n} - \nabla \bar{g}_{n;w_{n-1}}\|_*$, one can expect that the stability constant θ for this meta-learning problem will increase.

D Weighted Regularization

Here we discuss the case where $R(x) = F(\pi^*, x)$ regardless the condition $R(x) \ge 0$.

Corollary 4. *Let $\tilde{F}(x,x) = F(x,x) + \lambda F(\pi^*, x)$. Suppose $\forall x \in \mathcal{X}$, $\min_{x \in \mathcal{X}} \tilde{F}(x,x) \le (1+\lambda)\tilde{\epsilon}_{\Pi,\pi^*}$. Define $\Delta_N = (1+\lambda)\frac{(\tilde{\theta} e^{1-\tilde{\theta}} G_2)^2}{2\alpha} N^{2(\tilde{\theta}-1)}$. Running* AggreVaTe *with $\tilde{F}$ in* (14) *as the per-round cost has performance satisfies: for all $N > 0$,*

$$\begin{aligned} F(x_N, x_N) &\le (1+\lambda)\tilde{\epsilon}_{\Pi,\pi^*} - \lambda F(x^*, x_N) + \Delta_N \\ &\le \Delta_N + \tilde{\epsilon}_{\Pi,\pi^*} + \lambda G_2 \left(\frac{2\lambda G_2}{\alpha} + \sqrt{\frac{2\Delta_N}{\alpha}} \right) \end{aligned}$$

Proof. The first inequality can be seen by the definition $F(x_N, x_N) = \tilde{F}(x_N, x_N) - \lambda F(x^*, x_N)$ and then by applying Theorem 2 to $\tilde{F}(x_N, x_N)$.

The second inequality shows that $-F(x^*, x_N)$ cannot be too large. Let $f_*(x) = F(x^*, x)$ and $x_N^* = \arg\min_{x \in \mathcal{X}} f_N(x)$. Then

$$
\begin{aligned}
f_N(x_N) &= f_N(x_N) + \lambda f_*(x_N) - \lambda f_*(x_N) \\
&\leq \Delta_N - \lambda f_*(x_N) + \min_{x \in \mathcal{X}} f_N(x) + \lambda f_*(x) \\
&\leq \Delta_N + f_N(x_N^*) + \lambda(f_*(x_N^*) - f_*(x_N)) \\
&\leq \Delta_N + f_N(x_N^*) + \lambda G_2 \|x_N^* - x_N\|
\end{aligned}
$$

where the first inequality is due to Theorem 2 and the third inequality is due to f_* is G_2-Lipschitz continuous. Further, since f_N is α-strongly convex,

$$
\begin{aligned}
\frac{\alpha}{2}\|x_N^* - x_N\|^2 &\leq f_N(x_N) - f_N(x_N^*) \\
&\leq \Delta_N + \lambda G_2 \|x_N^* - x_N\|
\end{aligned}
$$

which implies

$$
\begin{aligned}
\|x_N^* - x_N\| &\leq \frac{\lambda G_2 + \sqrt{\lambda^2 G^2 + 2\alpha\Delta_N}}{\alpha} \\
&\leq \frac{2\lambda G_2 + \sqrt{2\alpha\Delta_N}}{\alpha}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
f_N(x_N) &\leq \Delta_N + f_N(x_N^*) + \lambda G_2 \|x_N^* - x_N\| \\
&\leq \Delta_N + \tilde{\epsilon}_{\Pi,\pi^*} + \lambda G_2 \left(\frac{2\lambda G_2}{\alpha} + \sqrt{\frac{2\Delta_N}{\alpha}} \right)
\end{aligned}
$$

■

Corollary 4 indicates that when π^* is better than all policies under the distribution of π^* (i.e. $F(x^*, x) \geq 0, \forall x \in \mathcal{X}$), then using AGGREVATE with the weighted problem such that $\tilde{\theta} < 1$ generates a convergent sequence and then the performance on the last iterate is bounded by $(1+\lambda)\tilde{\epsilon}_{\Pi,\pi^*} + \Delta_N$. That is, it only introduces a multiplicative constant on $\tilde{\epsilon}_{\Pi,\pi^*}$. Therefore, the bias due to regularization can be ignored by choosing a larger policy class. This suggests for applications like DAGGER introducing additional weighted cost $\lambda F(x^*, x)$ (i.e. demonstration samples collected under the expert policy's distribution) does not hurt.

However, in generally, $F(x^*, x_N)$ can be negative, when there is a better policy in Π than π^* in sense of the state distribution $d_{\pi^*}(s)$ generated by the expert policy π^*. Corollary 4 also shows this additional bias introduced by AGGREVATE is bounded at most $O(\frac{\lambda^2 G_2^2}{\alpha})$.