## Contents

## A Proofs for Section 3

**Proposition 3.1.** *Let the expert closed-loop system $f_{\mathsf{cl}}^{\pi_\star}$ be $\eta$-locally $\delta$-ISS for some $\eta > 0$. Fix an imitation gap bound $\varepsilon > 0$, initial condition $\xi$, and policy $\pi$. Then if*

$$\max_{0 \le t \le T-1} \sup_{\|\delta\| \le \varepsilon} \|\pi_\star(x_t^{\pi_\star}(\xi) + \delta) - \pi(x_t^{\pi_\star}(\xi) + \delta)\| \le \min\{\eta, \gamma^{-1}(\varepsilon)\}, \tag{5}$$

*we have that the imitation gap satisfies $\Gamma_T(\xi; \pi) \le \varepsilon$.*

*Proof.* We do a proof by induction.

**Base case $t = 0$:** We trivially have at $t = 0$:

$$\|x_t^{\pi_\star}(\xi) - x_t^{\pi}(\xi)\| = \|\xi - \xi\| = 0 \le \epsilon.$$

**Induction step:** Assume for some $k > 0$, we have $\max_{t \le k-1} \|x_t^{\pi_\star}(\xi) - x_t^{\pi}(\xi)\| \le \varepsilon$. We set $\delta := x_{k-1}^{\pi_\star}(\xi) - x_{k-1}^{\pi}(\xi)$ such that $\|\delta\| \le \varepsilon$. From Equation (5), we are guaranteed that

$$\begin{aligned}
\left\|\pi_\star(x_{k-1}^{\pi}(\xi)) - \pi(x_{k-1}^{\pi}(\xi))\right\| &= \left\|\pi_\star(x_{k-1}^{\pi_\star}(\xi) + \delta) - \pi(x_{k-1}^{\pi_\star}(\xi) + \delta)\right\| \\
&\le \max_{0 \le t \le k-1} \sup_{\|\delta\| \le \varepsilon} \|\pi_\star(x_t^{\pi_\star}(\xi) + \delta) - \pi(x_t^{\pi_\star}(\xi) + \delta)\| \\
&\le \min\{\eta, \gamma^{-1}(\varepsilon)\}.
\end{aligned}$$

Since $f_{\mathsf{cl}}^{\pi_\star}$ is $\eta$-locally $\delta$-ISS, we get from (3)

$$\begin{aligned}
\|x_k^{\pi}(\xi) - x_k^{\pi_\star}(\xi)\| &\le \gamma\left(\max_{0 \le s \le k-1} \|\pi_\star(x_s^{\pi}(\xi)) - \pi(x_s^{\pi}(\xi))\|\right) \\
&\le \gamma\left(\min\{\eta, \gamma^{-1}(\varepsilon)\}\right) \\
&\le \varepsilon,
\end{aligned}$$

and thus $\max_{t \le k} \|x_t^{\pi_\star}(\xi) - x_t^{\pi}(\xi)\| \le \varepsilon$, completing the induction step. □

**Theorem 3.1.** *Fix a test policy $\pi$ and initial condition $\xi \in \mathcal{X}$, and let Assumption 3.1 hold. Let $f_{\mathsf{cl}}^{\pi_\star}$ be $\eta$-locally $\delta$-ISS for some $\eta > 0$, and assume that the class $\mathcal{K}$ function $\gamma(\cdot)$ in (2) satisfies $\gamma(x) \le \mathcal{O}(x^{1+r})$ for some $r > 0$. Choose constants $\mu, \alpha > 0$ such that*

$$2L_\pi x + (x/\mu)^{\frac{1}{1+r}} \le \gamma^{-1}(x) \text{ for all } 0 \le x \le \alpha. \tag{6}$$

*Provided that the imitation error on the expert trajectory incurred by $\pi$ satisfies:*

$$\max_{0 \le t \le T-1} \mu \left\|\Delta_t^{\pi_\star}(\xi; \pi)\right\|^{1+r} \le \alpha, \ \max_{0 \le t \le T-1} 2L_\pi \mu \left\|\Delta_t^{\pi_\star}(\xi; \pi)\right\|^{1+r} + \|\Delta_t^{\pi_\star}(\xi; \pi)\| \le \eta, \tag{7}$$

*then for all $1 \le t \le T$ the instantaneous imitation gap is bounded as*

$$\|x_t^{\pi_\star}(\xi) - x_t^{\pi}(\xi)\| \le \max_{0 \le k \le t-1} \mu \left\|\Delta_k^{\pi_\star}(\xi; \pi)\right\|^{1+r}. \tag{8}$$

*Proof.* In order to leverage Proposition 3.1 we must first find a solution $\varepsilon$ to Equation (5). By Lipschitzness of the policy class,

$$\max_{0 \le t \le T-1} \sup_{\|\delta\| \le \varepsilon} \|\pi_\star(x_t^{\pi_\star}(\xi) + \delta) - \pi(x_t^{\pi_\star}(\xi) + \delta)\| \le 2L_\pi \varepsilon + \max_{0 \le t \le T-1} \|\Delta_t^{\pi_\star}(\xi; \pi)\|,$$

and using the lower bound in Equation (6) it is therefore sufficient to find a solution $\varepsilon \le \alpha$ to

$$\begin{aligned}
&2L_\pi \varepsilon + \max_{0 \le t \le T-1} \|\Delta_t^{\pi_\star}(\xi; \pi)\| \le 2L_\pi \varepsilon + (\varepsilon/\mu)^{\frac{1}{1+r}} \\
&\iff \max_{0 \le t \le T-1} \|\Delta_t^{\pi_\star}(\xi; \pi)\| \le (\varepsilon/\mu)^{\frac{1}{1+r}}.
\end{aligned}$$

Picking $\varepsilon = \max_{0 \le t \le T-1} \mu \|\Delta_t^{\pi_\star}(\xi; \pi)\|^{1+r}$ and adding the constraint $\varepsilon \le \alpha$ in order to ensure the solution is sufficiently small allows use to apply Proposition 3.1 and obtain the final result

$$\|x_t^{\pi_\star}(\xi) - x_t^{\pi}(\xi)\| \le \max_{0 \le t \le T-1} \mu \|\Delta_t^{\pi_\star}(\xi; \pi)\|^{1+r}.$$

Provided that

$$\max_{0 \le t \le T-1} \mu \|\Delta_t^{\pi_\star}(\xi; \pi)\|^{1+r} \le \alpha, \ \max_{0 \le t \le T-1} 2L_\pi \mu \|\Delta_t^{\pi_\star}(\xi; \pi)\|^{1+r} + \|\Delta_t^{\pi_\star}(\xi; \pi)\| \le \eta$$

Thus completing the proof. □

**Theorem 3.2.** *Let $f_{\mathsf{cl}}^{\pi_\star}$ be $\eta$-locally $\delta$-ISS for some $\eta > 0$, and assume that the class $\mathcal{K}$ function $\gamma(\cdot)$ in (2) satisfies $\gamma(x) \leq \mathcal{O}(x^{1/r})$ for some $r \geq 1$. Fix a test policy $\pi$ and initial condition $\xi \in \mathcal{X}$, and let Assumption 3.2 hold for $p \in \mathbb{N}$ satisfying $p + 1 - r > 0$. Choose $\mu, \alpha > 0$ such that*

$$2\frac{L_{\partial^p \pi}}{(p+1)!}x^{p+1} + (x/\mu)^r \leq \gamma^{-1}(x), \text{ for all } 0 \leq x \leq \alpha \leq \frac{1}{2}. \tag{10}$$

*Provided the $j$th total derivatives, $j = 0, \dots, p$, of the imitation error on the expert trajectory incurred by $\pi$ satisfy:*

$$\max_{0 \leq t \leq T-1} \max_{0 \leq j \leq p} \mu\left(\frac{2}{j!}\left\|\partial_x^j \Delta_t^{\pi_\star}(\xi;\pi)\right\|\right)^{1/r} \leq \alpha, \tag{11}$$

$$\max_{0 \leq t \leq T-1} \max_{0 \leq j \leq p} \frac{2L_{\partial^p \pi}\mu^{p+1}}{(p+1)!}\left(\frac{2}{j!}\left\|\partial_x^j \Delta_t^{\pi_\star}(\xi;\pi)\right\|\right)^{\frac{p+1}{r}} + \frac{2}{j!}\left\|\partial_x^j \Delta_t^{\pi_\star}(\xi;\pi)\right\| \leq \eta, \tag{12}$$

*then for all $1 \leq t \leq T$ the instantaneous imitation gap is bounded by*

$$\|x_t^{\pi_\star}(\xi) - x_t^{\pi}(\xi)\| \leq \max_{0 \leq k \leq t-1} \max_{0 \leq j \leq p} \mu\left(\frac{2}{j!}\left\|\partial_x^j \Delta_t^{\pi_\star}(\xi;\pi)\right\|\right)^{1/r}. \tag{13}$$

*Proof.* We proceed similarly as in the proof of Theorem 3.1. From Proposition 3.1, we can take the $p$th Taylor expansion of the left hand side of Equation (5) and apply the triangle inequality a few times to yield:

$$\begin{aligned}
&\max_{0 \leq t \leq T-1} \sup_{\|\delta\| \leq \varepsilon} \|\pi_\star(x_t^{\pi_\star}(\xi) + \delta) - \pi(x_t^{\pi_\star}(\xi) + \delta)\| \\
\leq &\max_{0 \leq t \leq T-1} \sup_{\|\delta\| \leq \varepsilon} \|\pi_\star(x_t^{\pi_\star}(\xi)) - \pi(x_t^{\pi_\star}(\xi))\| \\
&\qquad + \|\pi_\star(x_t^{\pi_\star}(\xi) + \delta) - \pi_\star(x_t^{\pi_\star}(\xi)) - (\pi(x_t^{\pi_\star}(\xi) + \delta) - \pi(x_t^{\pi_\star}(\xi)))\| \\
\leq &\max_{0 \leq t \leq T-1} \sup_{\|\delta\| \leq \varepsilon} \|\pi_\star(x_t^{\pi_\star}(\xi)) - \pi(x_t^{\pi_\star}(\xi))\| \\
&\qquad + \left\|\sum_{j=1}^{p} \frac{1}{j!}\partial_x^j \pi_\star(x_t^{\pi_\star}(\xi)) \cdot \delta^{\otimes j} - \sum_{j=1}^{p} \frac{1}{j!}\partial_x^j \pi(x_t^{\pi_\star}(\xi)) \cdot \delta^{\otimes j}\right\| + 2\frac{L_{\partial^p \pi}}{(p+1)!}\|\delta\|^{p+1} \\
\leq &\max_{0 \leq t \leq T-1} \sup_{\|\delta\| \leq \varepsilon} \sum_{j=0}^{p} \frac{1}{j!}\left\|\partial_x^j \Delta_t^{\pi_\star}(\xi;\pi) \cdot \delta^{\otimes j}\right\| + 2\frac{L_{\partial^p \pi}}{(p+1)!}\|\delta\|^{p+1} \\
\leq &\max_{0 \leq t \leq T-1} \sum_{j=0}^{p} \frac{1}{j!}\left\|\partial_x^j \Delta_t^{\pi_\star}(\xi;\pi)\right\|\varepsilon^j + 2\frac{L_{\partial^p \pi}}{(p+1)!}\varepsilon^{p+1}.
\end{aligned}$$

Therefore, it suffices to find an $\varepsilon$ small enough such that

$$\max_{0 \leq t \leq T-1} 2\frac{L_{\partial^p \pi}}{(p+1)!}\varepsilon^{p+1} + \sum_{j=0}^{p} \frac{1}{j!}\left\|\partial_x^j \Delta_t^{\pi_\star}(\xi;\pi)\right\|\varepsilon^j \leq \gamma^{-1}(\varepsilon).$$

Since we are given $\gamma(x) \leq \mathcal{O}(x^{1/r})$, we have $\gamma^{-1}(x) \geq \Omega(x^r)$. This motivates finding a large enough $\mu$ and small enough neighborhood $\alpha$ such that

$$\max_{0 \leq t \leq T-1} 2\frac{L_{\partial^p \pi}}{(p+1)!}\varepsilon^{p+1} + \left(\frac{\varepsilon}{\mu}\right)^r \leq \gamma^{-1}(\varepsilon),$$

for all $0 < \varepsilon \leq \alpha \leq 1/2$. In essence, we want to find a sufficiently small neighborhood $\alpha$ such that the $\varepsilon^{p+1}$ term is dominated by the $\varepsilon^r$ term, while also selecting a $\mu$ such that the total sum is still upper bounded by $\gamma^{-1}(x) \geq \Omega(x^r)$ in this neighborhood. The choice of raising $\varepsilon/\mu$ to the $r$-th power arises from the fact that $r$ is the smallest exponent–thus affecting the imitation gap in Equation (13) downstream least severely–that ensures $\mu, \alpha$ will always exist. Having found such $\mu, \alpha$, we now simply have to find $\left\|\partial_x^j \Delta_t^{\pi_\star}(\xi;\pi)\right\|$ small enough such that

$$
\begin{aligned}
\sum_{j=0}^{p} \frac{1}{j!} \left\|\partial_x^j \Delta_t^{\pi_\star}(\xi;\pi)\right\| \varepsilon^j &\leq \max_{j\leq p} \frac{1}{j!} \left\|\partial_x^j \Delta_t^{\pi_\star}(\xi;\pi)\right\| \sum_{j=0}^{p} \varepsilon^j \\
&\leq \max_{j\leq p} \frac{2}{j!} \left\|\partial_x^j \Delta_t^{\pi_\star}(\xi;\pi)\right\| \qquad\qquad \varepsilon \leq \alpha \leq 1/2 \\
&= \left(\frac{\varepsilon}{\mu}\right)^r.
\end{aligned} \tag{21}
$$

Solving this for $\varepsilon$, we get

$$
\varepsilon = \max_{j\leq p} \mu \left(\frac{2}{j!} \left\|\partial_x^j \Delta_t^{\pi_\star}(\xi;\pi)\right\|\right)^{1/r},
$$

as long as $\varepsilon \leq \alpha$, the neighborhood condition, and $2\frac{L_{\partial^p \pi}}{(p+1)!}\varepsilon^{p+1} + \left(\frac{\varepsilon}{\mu}\right)^r \leq \eta$, the locality for $\delta$-ISS. These correspond to the conditions (11) and (12), respectively. This completes the proof. $\square$

For completeness we present here a stronger variant of Theorem 3.2 for the special case where $p = r \in \mathbb{N}$. In this scenario we are able to remove the dependency of the imitation gap bounds on the $p$th order derivative provided it can be made sufficiently small.

**Theorem A.1.** *Let $f_{\mathsf{cl}}^{\pi_\star}$ be $\eta$-locally $\delta$-ISS for some $\eta > 0$, and assume that the class $\mathcal{K}$ function $\gamma(\cdot)$ in (2) satisfies $\gamma(x) \leq \mathcal{O}(x^{1/r})$ for some $r \geq 1$. Fix a test policy $\pi$ and initial condition $\xi \in \mathcal{X}$, and let Assumption 3.2 hold with $p = r \in \mathbb{N}$. Choose $\mu, \alpha > 0$ such that*

$$
2\frac{L_{\partial^p \pi}}{(p+1)!} x^{p+1} + (x/\mu)^p \leq \gamma^{-1}(x), \text{ for all } 0 \leq x \leq \alpha \leq \frac{1}{2}. \tag{22}
$$

*Provided the $j$th total derivatives, $j = 0, \dots, p$, of the imitation error on the expert trajectory incurred by $\pi$ satisfy:*

$$
\max_{0\leq t\leq T-1} \max_{0\leq j\leq p-1} \mu \left(\frac{4}{j!} \left\|\partial_x^j \Delta_t^{\pi_\star}(\xi;\pi)\right\|\right)^{1/p} \leq \alpha, \tag{23}
$$

$$
\max_{0\leq t\leq T-1} \max_{0\leq j\leq p-1} \frac{2L_{\partial^p \pi}\mu^{p+1}}{(p+1)!} \left(\frac{4}{j!} \left\|\partial_x^j \Delta_t^{\pi_\star}(\xi;\pi)\right\|\right)^{\frac{p+1}{p}} + \frac{4}{j!} \left\|\partial_x^j \Delta_t^{\pi_\star}(\xi;\pi)\right\| \leq \eta, \tag{24}
$$

$$
\|\partial_x^p \Delta_t^{\pi_\star}(\xi;\pi)\| \leq \frac{p!}{2\mu^p} \tag{25}
$$

*then for all $1 \leq t \leq T$ the instantaneous imitation gap is bounded by*

$$
\|x_t^{\pi_\star}(\xi) - x_t^{\pi}(\xi)\| \leq \max_{0\leq k\leq t-1} \max_{0\leq j\leq p-1} \mu \left(\frac{4}{j!}\right)^{1/r} \left\|\partial_x^j \Delta_t^{\pi_\star}(\xi;\pi)\right\|^{1/r}. \tag{26}
$$

*Proof.* We follow the proof of Theorem 3.2 until Equation (21). We then wish to solve

$$
\sum_{j=0}^{p} \frac{1}{j!} \|\partial_x^j \Delta_t^{\pi_\star}(\xi;\pi)\| \varepsilon^j \leq \left(\frac{\varepsilon}{\mu}\right)^p.
$$

Since the order of the RHS is $p$, provided that $\frac{1}{p!}\|\partial_x^p \Delta_t^{\pi_\star}(\xi;\pi)\| \leq \frac{1}{2}\frac{1}{\mu^p}$ we can write

$$
\sum_{j=0}^{p-1} \frac{1}{j!} \|\partial_x^j \Delta_t^{\pi_\star}(\xi;\pi)\| \varepsilon^j \leq \frac{1}{2}\left(\frac{\varepsilon}{\mu}\right)^p.
$$

Upper-bounding the polynomial on the LHS using a geometric series and solving for $\varepsilon$ we get

$$
\varepsilon = \max_{j\leq p-1} \mu \left(\frac{4}{j!} \left\|\partial_x^j \Delta_t^{\pi_\star}(\xi;\pi)\right\|\right)^{1/p},
$$

provided that $\varepsilon \leq \alpha$, $2\frac{L_{\partial^p \pi}}{(p+1)!}\varepsilon^{p+1} + \left(\frac{\varepsilon}{\mu}\right)^p \leq \eta$, and $\|\partial_x^p \Delta_t^{\pi_\star}(\xi;\pi)\| < \frac{p!}{2\mu^p}$. These conditions correspond to that of the theorem, completing the proof. $\square$

**Corollary A.1.** *Consider a $\delta$-ISS $f_{\mathsf{cl}}^{\pi_\star}$ system with $\gamma(x) := \gamma x, \gamma > 0$ and $\eta = \infty$. Let Assumption 3.2 hold with $p = 1$ and assume without loss of generality $\gamma L_{\partial\pi} \geq 1$. Provided*

$$\max_{0 \leq t \leq T-1} \|\partial_x \Delta_t^{\pi_\star}(\xi; \pi)\| \leq \frac{1}{4\gamma}, \quad \max_{0 \leq t \leq T-1} \|\Delta_t^{\pi_\star}(\xi; \pi)\| \leq \frac{1}{16\gamma^2 L_{\partial\pi}}$$

*then for all $0 \leq t \leq T$*

$$\|x_t^{\pi_\star}(\xi) - x_t^{\pi}(\xi)\| \leq \max_{0 \leq k \leq t-1} 8\gamma \|\Delta_k^{\pi_\star}(\xi; \pi)\|.$$

*Proof.* Choose $\alpha := \frac{1}{2\gamma L_{\partial\pi}}$ and $\mu := 2\gamma$. Assume $\gamma L_{\partial\pi} \geq 1$. Since $\gamma^{-1}(x) = \frac{x}{\gamma}$, for $x \leq \alpha$ it holds that

$$L_{\partial^p\pi} x^2 + (x/\mu) \leq \frac{x}{2\gamma} + \frac{x}{2\gamma} \leq \gamma^{-1}(x) := \frac{x}{\gamma}.$$

and we can directly apply the $p = r = 1$ special case of Theorem A.1. Then, if the constraints described by Equations (23) and (25) are satisfied:

$$\max_{0 \leq t \leq T-1} \|\partial_x \Delta_t^{\pi_\star}(\xi; \pi)\| \leq \frac{p!}{2\mu^p} = \frac{1}{4\gamma}, \quad \max_{0 \leq t \leq T-1} \|\Delta_t^{\pi_\star}(\xi; \pi)\| \leq \frac{0!}{4}\left(\frac{\alpha}{\mu}\right)^p = \frac{1}{16\gamma^2 L_{\partial\pi}},$$

it holds for all $1 \leq t \leq T$

$$\|x_t^{\pi_\star}(\xi) - x_t^{\pi}(\xi)\| \leq \max_{0 \leq k \leq t-1} 8\gamma \|\Delta_k^{\pi_\star}(\xi; \pi)\|.$$

□

# B Proofs for Section 4

## B.1 Preliminaries

Let $\mathcal{G} \subset \mathbb{R}^{\mathcal{X}}$ be a set of functions, and let $x_1, \ldots, x_n \in \mathcal{X}$ be a fixed set of points. We will endow $\mathcal{G}$ with the following empirical $L^2$ pseudo-metric space structure:

$$d(f, g) := \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f(x_i) - g(x_i))^2}, \;\; f, g \in \mathcal{G}.$$

The empirical Rademacher complexity of $\mathcal{G}$ is defined as:

$$\mathcal{R}_n(\mathcal{G}) := \mathbb{E}_\varepsilon\left[\sup_{g \in \mathcal{G}} \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i g(x_i)\right],$$

where the $\{\varepsilon_i\}_{i=1}^n$ are independent Rademacher random variables. Dudley's inequality yields a bound on $\mathcal{R}_n(\mathcal{G})$ using the metric space structure of $(\mathcal{G}, d)$.

**Lemma B.1** (Dudley's inequality [cf. 31, Lemma A.3])**.** *Let $R := \sup_{f \in \mathcal{G}} d(f, 0)$ be the radius of the set $\mathcal{G}$. We have that:*

$$\mathcal{R}_n(\mathcal{G}) \leq \inf_{\alpha \in [0,R]} \left\{4\alpha + \frac{12}{\sqrt{n}}\int_\alpha^R \sqrt{\log N(\mathcal{G}; d, \varepsilon)}d\varepsilon\right\}.$$

*Here, $N(\mathcal{G}; d, \varepsilon)$ denotes the covering number of $\mathcal{G}$ in the metric $d$ at resolution $\varepsilon$.*

## B.2 Generalization bound for the non-realizable setting

We use standard techniques to derive a generalization bound for the *non-realizable setting*, i.e., where $\pi_\star$ may not necessarily be contained in the hypothesis class $\Pi$. Let $\mathcal{G} \subset [0, 1]^{\mathcal{X}}$ be a given function class. We have the following standard uniform convergence generalization bound [cf. 32, Theorem 4.10]: with probability greater than $1 - \delta$ over $x_1, \ldots, x_n \overset{\text{i.i.d.}}{\sim} \mathcal{D}$, we have

$$\sup_{g \in \mathcal{G}} |\mathbb{E}_x[g] - \mathbb{E}_n[g]| \leq 2\mathbb{E}_{x_{1:n}}[\mathcal{R}_n(\mathcal{G})] + \sqrt{\frac{\log(2/\delta)}{n}}, \tag{27}$$

where $\mathbb{E}_{x_{1:n}}$ denotes expectation over the randomness of $x_1, \ldots, x_n$. To establish an upper bound on $\mathbb{E}_{x_{1:n}}[\mathcal{R}_n(\mathcal{G})]$, we focus on the Lipschitz parametric case, though we note many analogous bounds can be computed for a plethora of other function classes [32].

**Theorem B.1.** *Let $\mathcal{G} \subset [0,1]^{\mathcal{X}}$ be a $(B_\theta, L_\theta, q)$-Lipschitz parametric function class. Given $\delta \in (0,1)$, with probability at least $1-\delta$ over the i.i.d. draws $x_1, \dots, x_n \sim \mathcal{D}$, the following bound holds:*

$$\sup_{g \in \mathcal{G}} |\mathbb{E}_x[g] - \mathbb{E}_n[g]| \le 48\sqrt{\frac{q\log(3B_\theta L_\theta)}{n}} + \sqrt{\frac{\log(2/\delta)}{n}}. \tag{28}$$

*Proof.* This argument is fairly standard. Fix a set of points $x_1, \dots, x_n \in \mathcal{X}$. Since $\mathcal{G}$ contains only functions with range $[0,1]$, the radius of the set $\mathcal{G}$ in the empirical $L^2$ metric is:

$$\sup_{f \in \mathcal{G}} d(f, 0) \le 1.$$

Therefore, Dudley's inequality (Lemma B.1) yields:

$$\mathcal{R}_n(\mathcal{G}) \le \frac{12}{\sqrt{n}} \int_0^1 \sqrt{\log N(\mathcal{G}; d, \varepsilon)}d\varepsilon.$$

Now using the fact that $\mathcal{G}$ is a $(B_\theta, L_\theta, q)$-Lipschitz parametric function class, it is not hard to see that for any $\varepsilon > 0$, an $\varepsilon/(B_\theta L_\theta)$-cover of $\mathbb{B}_2^q(1)$ in the Euclidean metric yields an $\varepsilon$-cover of $\mathcal{G}$ in the $d$-metric. Hence, for any $\varepsilon \in (0,1)$, by a standard volume comparison argument:

$$\begin{aligned}\log N(\mathcal{G}; d, \varepsilon) &\le \log N\left(\mathbb{B}_2^q(1); \|\cdot\|, \frac{\varepsilon}{B_\theta L_\theta}\right) \\ &\le q \log\left(1 + \frac{2B_\theta L_\theta}{\varepsilon}\right) \\ &\le q \log\left(\frac{3B_\theta L_\theta}{\varepsilon}\right).\end{aligned}$$

Therefore, we have:

$$\begin{aligned}\int_0^1 \sqrt{\log N(\mathcal{G}; d, \varepsilon)}d\varepsilon &\le \sqrt{q}\int_0^1 \sqrt{\log\left(\frac{3B_\theta L_\theta}{\varepsilon}\right)}d\varepsilon \\ &\le \sqrt{q\log(3B_\theta L_\theta)} + \sqrt{q}\int_0^1 \sqrt{\log(1/\varepsilon)}d\varepsilon && \text{using } \sqrt{a+b} \le \sqrt{a} + \sqrt{b} \\ &\le \sqrt{q\log(3B_\theta L_\theta)} + \sqrt{q} && \text{using } \int_0^1 \sqrt{\log\left(\frac{1}{\varepsilon}\right)}d\varepsilon \le 1 \\ &\le 2\sqrt{q\log(3B_\theta L_\theta)}.\end{aligned}$$

Plugging this back into Dudley's inequality:

$$\mathcal{R}_n(\mathcal{G}) \le 24\sqrt{\frac{q}{n}}\sqrt{\log(3B_\theta L_\theta)}.$$

The claim now follows from the standard uniform convergence inequality (27). $\square$

Applying this generalization bound to the $(B_\theta, B_{\ell,p}^{-1}L_{\ell,p}, q)$-Lipschitz parametric function class $B_{\ell,p}^{-1}(\ell_p^{\pi_\star} \circ \Pi_{\theta,p})$, we get the non-realizable analogue to Corollary 4.1.

**Corollary B.1.** *Let the policy class $\Pi_{\theta,p}$ be defined as in (18). Let the function class $\ell_p^{\pi_\star} \circ \Pi_{\theta,p}$ be defined as in (19), and constants $B_{\ell,p}$, $L_{\ell,p}$ be defined as above. Let $\hat{\pi}_{\mathsf{TaSIL},p}$ be any empirical risk minimizer (15). Then with probability at least $1-\delta$ over the initial conditions $\{\xi_i\}_{i=1}^n \overset{\text{i.i.d.}}{\sim} \mathcal{D}^n$,*

$$\mathbb{E}_\xi\left[\ell_p^{\pi_\star}(\xi; \hat{\pi}_{\mathsf{TaSIL},p})\right] \le \mathbb{E}_n[\ell_p^{\pi_\star}(\cdot\,; \hat{\pi}_{\mathsf{TaSIL},p})] + 48B_{\ell,p}\sqrt{\frac{q\log\left(3B_\theta B_{\ell,p}^{-1}L_{\ell,p}\right)}{n}} + B_{\ell,p}\sqrt{\frac{\log(2/\delta)}{n}}. \tag{29}$$

Inserting the generalization bound in Corollary B.1 in lieu of Corollary 4.1 for the rest of the bounds seen in Section 4 yields the sample complexity bounds relevant to our problem in the non-realizable setting. However, we note an important subtlety that manifests in the non-realizable regime. We note that in Corollary 4.1, due to realizability, the generalization bound monotonically decreases to 0 with $n$, whereas in Corollary B.1, we have an additive factor of $\mathbb{E}_n[\ell_p^{\pi_\star}(\cdot\ ; \hat{\pi}_{\mathsf{TaSIL},p})]$. It is therefore possible for either small enough $n$ or insufficiently expressive function classes $\Pi_{\theta,p}$ that the non-zero empirical risk automatically violates the imitation error requirements in Theorems 3.1 and 3.2. Thus, a necessary assumption must be made in the non-realizable setting for the function class to be expressive enough such that the empirical risk it incurs on sufficiently large datasets satisfies the imitation error requirements with high probability.

### B.3 Proof of Theorem 4.1

Before turning to the proof of Theorem 4.1, we introduce some notation and tools from the local Rademacher complexity literature [33, 34].

**Definition B.1** (Sub-root function)**.** *A function $\phi : [0,\infty) \to \mathbb{R}$ is said to be a* sub-root function *if:*

*a) $\phi$ is non-negative.*

*b) $\phi$ is not the zero function.*

*c) $\phi$ is non-decreasing.*

*d) $r \mapsto \phi(r)/\sqrt{r}$ is non-increasing.*

For any non-negative function class $\mathcal{G}$, scalar $r \geq 0$, and $n$ points $x_1, \dots, x_n \in \mathcal{X}$, define:

$$\mathcal{H}_n(r; x_{1:n}) := \{g \in \mathcal{G} \mid \mathbb{E}_n[g] \leq r\}.$$

The following is from Bousquet [34].

**Theorem B.2** (Bousquet [34, Theorem 6.1])**.** *Let $\mathcal{G} \subset [0,1]^{\mathcal{X}}$, and fix a $\delta \in (0,1)$. With probability at least $1-\delta$ over the i.i.d. draws of $x_1, \dots, x_n$, the following holds. Let $\phi_n$ be any sub-root function (cf. Definition B.1) satisfying:*

$$\mathcal{R}_n(\mathcal{H}_n(r; x_{1:n})) \leq \phi_n(r), \;\; \forall\, r > 0.$$

*Let $r_n^*$ denote the largest solution to the equation $\phi_n(r) = r$. Then, for all $g \in \mathcal{G}$:*

$$\mathbb{E}_x[g] \leq 2\mathbb{E}_n[g] + 106 r_n^* + \frac{48(\log(1/\delta) + 6\log\log n)}{n}.$$

With these definitions and preliminary results in place, we turn to the proof of Theorem 4.1.

**Theorem 4.1.** *Let $\mathcal{G} \subset [0,1]^{\mathcal{X}}$ be a $(B_\theta, L_\theta, q)$-Lipschitz parametric function class. There exists a universal positive constant $K < 10^6$ such that the following holds. Given $\delta \in (0,1)$, with probability at least $1-\delta$ over the i.i.d. draws $x_1, \dots, x_n \sim \mathcal{D}$, for all $g \in \mathcal{G}$, the following bound holds:*

$$\mathbb{E}_x[g] \leq 2\mathbb{E}_n[g] + K\left(\frac{q\log(B_\theta L_\theta n) + \log(1/\delta)}{n}\right). \tag{17}$$

*Proof.* Fix a set of points $x_1, \dots, x_n \in \mathcal{X}$. Define $\mathcal{G}_n(r; x_{1:n})$ as:

$$\mathcal{G}_n(r; x_{1:n}) := \{g \in \mathcal{G} \mid \mathbb{E}_n[g^2] \leq r\}.$$

For what follows, we often suppress the explicit dependence on $x_{1:n}$ in the notation for $\mathcal{H}_n$ and $\mathcal{G}_n$. Observe that since $\mathcal{G} \subset [0,1]^{\mathcal{X}}$, we have $\mathbb{E}_n[g^2] \leq \mathbb{E}_n[g]$ for every $g \in \mathcal{G}$, and therefore:

$$\mathcal{H}_n(r) \subseteq \mathcal{G}_n(r), \;\; \forall\, r \geq 0.$$

Hence $\mathcal{R}_n(\mathcal{H}_n(r)) \leq \mathcal{R}_n(\mathcal{G}_n(r))$, and it suffices for us to prove an upper bound on the latter.

**Proposition B.1.** *Let $\mathcal{G} \subset [0,1]^{\mathcal{X}}$ be a $(B_\theta, L_\theta, q)$-Lipschitz parametric function class. Fix a set of points $x_1, \dots, x_n \in \mathcal{X}$. We have that:*

$$\mathcal{R}_n(\mathcal{G}_n(r; x_{1:n})) \leq 24\sqrt{2}\sqrt{\frac{q}{n}}\min\{\sqrt{r}, 1\}\sqrt{\log\left(\frac{6B_\theta L_\theta}{\min\{\sqrt{r},1\}}\right)}.$$

*Proof of Proposition B.1.* The radius of the set $\mathcal{G}_n(r)$ in the empirical $L^2$ metric $d$ is upper bounded by $\sqrt{r}$ by definition. Furthermore, the radius of $\mathcal{G}$ in the metric $d$ is upper bounded by one. Hence, since $\mathcal{G}_n(r) \subseteq \mathcal{G}$, the radius of $\mathcal{G}_n(r)$ is upper bounded by $\min\{\sqrt{r}, 1\}$.

Dudley's inequality (Lemma B.1) yields:

$$\mathcal{R}_n(\mathcal{G}_n(r)) \leq \inf_{\alpha \in [0, \min\{\sqrt{r},1\}]} \left\{ 4\alpha + \frac{12}{\sqrt{n}} \int_\alpha^{\min\{\sqrt{r},1\}} \sqrt{\log N(\mathcal{G}; d, \varepsilon/2)} d\varepsilon \right\}. \tag{30}$$

Here, we have used the fact that the inclusion $\mathcal{G}_n(r) \subseteq \mathcal{G}$ implies $N(\mathcal{G}_n(r); d, \varepsilon) \leq N(\mathcal{G}; d, \varepsilon/2)$ by Vershynin [35, Exercise 4.2.10].

Since $\mathcal{G}$ is $(B_\theta, L_\theta, q)$-Lipschitz, for any $\varepsilon > 0$, an $\varepsilon$-covering of $\mathcal{G}$ in the $d$-metric can be constructed from an $\varepsilon/(B_\theta L_\theta)$-covering of $\mathbb{B}_2^q(1)$ in the Euclidean metric. Therefore, for any $\varepsilon \in (0,1)$, by the standard volume comparison bound:

$$\begin{aligned}
\log N(\mathcal{G}; d, \varepsilon) &\leq \log N\left(\mathbb{B}_2^q(1); \|\cdot\|, \frac{\varepsilon}{B_\theta L_\theta}\right) \\
&\leq q \log\left(1 + \frac{2B_\theta L_\theta}{\varepsilon}\right) \\
&\leq q \log\left(\frac{3B_\theta L_\theta}{\varepsilon}\right).
\end{aligned}$$

Putting $R := \min\{\sqrt{r}, 1\}$,

$$\begin{aligned}
&\int_0^R \sqrt{\log N(\mathcal{G}; d, \varepsilon/2)} d\varepsilon \\
&\leq \sqrt{q} \left[ R\sqrt{\log(6B_\theta L_\theta)} + \int_0^R \sqrt{\log(1/\varepsilon)} d\varepsilon \right] && \text{using } \sqrt{a+b} \leq \sqrt{a} + \sqrt{b} \\
&= \sqrt{q} \left[ R\sqrt{\log(6B_\theta L_\theta)} + R\int_0^1 \sqrt{\log\left(\frac{1}{R\varepsilon}\right)} d\varepsilon \right] && \text{change of variables } \varepsilon \leftarrow \varepsilon/R \\
&\leq \sqrt{q} \left[ R\sqrt{\log(6B_\theta L_\theta)} + R\sqrt{\log\left(\frac{1}{R}\right)} + R \right] && \text{using } \int_0^1 \sqrt{\log\left(\frac{1}{\varepsilon}\right)} d\varepsilon \leq 1 \\
&\leq R\sqrt{q} \left[ \sqrt{\log(6B_\theta L_\theta)} + 2\sqrt{\log\left(\frac{1}{R}\right)} \right] \\
&\leq 2\sqrt{2}R\sqrt{q}\sqrt{\log\left(\frac{6B_\theta L_\theta}{R}\right)} && \text{using } \sqrt{a} + \sqrt{b} \leq \sqrt{2}\sqrt{a+b}.
\end{aligned}$$

The claim now follows. □

We complete the proof by upper bounding $r_n^*$ and invoking Theorem B.2. First, observe that by Cauchy-Schwarz, the inequality $\mathbb{E}_n[g^2] \leq \mathbb{E}_n[g]$ for $g \in \mathcal{G}$, and Jensen's inequality:

$$\mathcal{R}_n(\mathcal{H}_n(r)) \leq \sup_{g \in \mathcal{H}_n(r)} \sqrt{\mathbb{E}_n[g^2]} \mathbb{E}_\varepsilon \sqrt{\frac{1}{n} \sum_{i=1}^n \varepsilon_i^2} \leq \sqrt{r}.$$

This bound holds for any $r \geq 0$. Hence, when $r \leq 1/n^2$:

$$\mathcal{R}_n(\mathcal{H}_n(r)) \leq 1/n.$$

On the other hand, when $r > 1/n^2$, by $\mathcal{R}_n(\mathcal{H}_n(r)) \leq \mathcal{R}_n(\mathcal{G}_n(r))$, Proposition B.1, and the inequalities $1/n < \min\{\sqrt{r}, 1\} \leq \sqrt{r}$:

$$\mathcal{R}_n(\mathcal{H}_n(r)) \leq 24\sqrt{2}\sqrt{\frac{q}{n}}\sqrt{r}\sqrt{\log(6B_\theta L_\theta n)}.$$

Hence, the function $\phi_n$ defined as:

$$\phi_n(r) := \max\left\{24\sqrt{2}\sqrt{\frac{q\log(6B_\theta L_\theta n)}{n}}\sqrt{r}, \frac{1}{n}\right\},$$

satisfies $\mathcal{R}_n(\mathcal{H}_n(r)) \leq \phi_n(r)$ for all $r \geq 0$. It is also not hard to see that $\phi_n$ is a sub-root function (cf. Definition B.1). Therefore, there is a unique solution $r_n^*$ satisfying $\phi_n(r_n^*) = r_n^*$. Now, for any positive constants $A, B$, the root of $r = \max\{A\sqrt{r}, B\}$ is upper bounded by $\max\{A^2, B\}$. Hence,

$$r_n^* \leq 1152\frac{q\log(6B_\theta L_\theta n)}{n}.$$

Theorem 4.1 now follows by Theorem B.2. $\square$

### B.4 Proof of Corollary 4.1

**Lemma B.2.** *Let $B_{\ell,p} := \frac{2}{p+1}\sum_{j=0}^p B_j$ and $L_{\ell,p} := \frac{B_X}{p+1}\sum_{j=0}^p L_j$. Then $B_{\ell,p}^{-1}(\ell_p^{\pi_\star} \circ \Pi_{\theta,p})$ is a $(B_\theta, B_{\ell,p}^{-1}L_{\ell,p}, q)$-Lipschitz parametric function class*

*Proof.* It suffices to show that

$$\max_{0\leq t\leq T-1}\left\|\partial_x^j \Delta_t^{\pi_\star}(\xi;\pi)\right\|$$

is $2B_j$-bounded and $B_X L_j$ Lipschitz with respect to $\Theta$. By definition, we immediately get

$$\begin{aligned}\left\|\partial_x^j \Delta_t^{\pi_\star}(\xi;\pi)\right\| &= \left\|\partial_x^j \pi_\star(x_t^{\pi_\star}(\xi)) - \partial_x^j \pi(x_t^{\pi_\star}(\xi))\right\| \\ &\leq 2\sup_{\|x\|\leq B_X, \|\theta\|\leq B_\theta}\left\|\partial_x^j \pi(x,\theta)\right\| \\ &= 2B_j.\end{aligned}$$

To bound the Lipschitz constant, we iteratively apply the Fundamental Theorem of Line Integrals:

$$\begin{aligned}\partial_x^j\pi(x;\theta_1) - \partial_x^j\pi(x;\theta_2) &= \int_{\theta_2}^{\theta_1}\int_0^x \frac{\partial^{j+2}\pi}{\partial x^{j+1}\partial\theta}(z\otimes\omega)\, dz d\omega \\ &= \int_{\theta_2}^{\theta_1}\left(\int_0^1 \frac{\partial^{j+2}\pi}{\partial x^{j+1}\partial\theta}(\alpha x\otimes\omega)\, d\alpha\right)x\, d\omega \\ &= \left(\int_0^1\int_0^1 \frac{\partial^{j+2}\pi}{\partial x^{j+1}\partial\theta}(\alpha x\otimes(\theta_2 + \beta(\theta_1-\theta_2)))\, d\alpha d\beta\right)x\otimes(\theta_1-\theta_2).\end{aligned}$$

Taking norms on both sides, we get

$$\begin{aligned}\left\|\partial_x^j\pi(x;\theta_1) - \partial_x^j\pi(x;\theta_2)\right\| &\leq \sup_{\|x\|\leq B_X, \|\theta\|\leq B_\theta}\left\|\frac{\partial^{j+2}\pi}{\partial x^{j+1}\partial\theta}\right\|\|x\|\,\|\theta_1-\theta_2\| \\ &\leq B_X L_j\,\|\theta_1-\theta_2\|,\end{aligned}$$

which establishes that $\left\|\partial_x^j \Delta_t^{\pi_\star}(\xi;\pi)\right\|$ is $B_X L_j$-Lipschitz. Recalling that

$$\ell_p^{\pi_\star}(\xi;\pi) := \tfrac{1}{p+1}\textstyle\sum_{j=0}^p \max_{0\leq t\leq T-1}\left\|\partial_x^j \Delta_t^{\pi_\star}(\xi;\pi)\right\|,$$

it follows that $\ell_p^{\pi_\star}(\xi;\pi)$ is $\frac{2}{p+1}\sum_{j=0}^p B_j$-bounded and $\frac{B_X}{p+1}\sum_{j=0}^p L_j$-Lipschitz. $\square$

**Corollary 4.1.** *Let the policy class $\Pi_{\theta,p}$ be defined as in (18), and assume that $\pi_\star \in \Pi_{\theta,p}$. Let the function class $\ell_p^{\pi_\star}\circ\Pi_{\theta,p}$ be defined as in (19), and constants $B_{\ell,p}$, $L_{\ell,p}$ be defined as above. Let $\hat{\pi}_{\mathsf{TaSIL},p}$ be any empirical risk minimizer (15). Then with probability at least $1-\delta$ over the initial conditions $\{\xi_i\}_{i=1}^n \overset{\text{i.i.d.}}{\sim} \mathcal{D}^n$,*

$$\mathbb{E}_\xi\left[\ell_p^{\pi_\star}(\xi;\hat{\pi}_{\mathsf{TaSIL},p})\right] \leq \mathcal{O}(1)B_{\ell,p}\frac{q\log\left(B_\theta B_{\ell,p}^{-1}L_{\ell,p}n\right) + \log(1/\delta)}{n}. \tag{20}$$

*Proof.* This follows by directly applying the constants derived in Lemma B.2 to Theorem 4.1, and using the assumption that $\pi_\star \in \Pi_{\theta,p}$ such that $\mathbb{E}_n\left[\ell_p^{\pi_\star}(\cdot\,;\hat{\pi}_{\mathsf{TaSIL},p})\right] = 0$. $\square$

### B.5 Proofs of Theorem 4.2 and Theorem 4.3

Before proceeding to the proofs of the main sample complexity bounds, we introduce the following lemma for inverting functions of the form $\log n/n$, adapted from Simchowitz et al. [36, Lemma A.4].

**Lemma B.3.** *Given $n \in \mathbb{N}$, $n \geq b\log(cn)$ as long as $n \geq 2b\log(2bc)$, where we assume $b, c \geq 1$.*

*Proof.* We observe by derivatives that $n - b\log(cn)$ is strictly increasing for $n \geq b$. Therefore, it suffices to show $b\log(cn) \leq n$ when $n = 2b\log(2bc)$.

$$\begin{aligned} b\log(2bc\log(2bc)) &= b\log(2\log(2)bc + 2bc\log(bc)) \\ &\leq b\log\big((2\log(2)+2)(bc)^2\big) && bc \geq 1 \\ &= 2b\log\Big(\sqrt{2\log(2)+2}\,bc\Big) \\ &< 2b\log(2bc). \end{aligned}$$

□

**Theorem B.3** (Full version of Theorem 4.2)**.** *Assume that $\pi_\star \in \Pi_{\theta,0}$ and let the assumptions of Theorem 3.1 hold for all $\pi \in \Pi_{\theta,0}$. Let Equation (6) hold with constants $\mu, \alpha > 0$, and assume without loss of generality that $\alpha/\mu \leq 1$, $L_\pi \mu \geq 1/2$. Let $\hat{\pi}_{\mathsf{TaSIL},0}$ be an empirical risk minimizer of $\ell_0^{\pi_\star}$ over the policy class $\Pi_{\theta,0}$ for initial conditions $\{\xi_i\} \overset{\text{i.i.d.}}{\sim} \mathcal{D}^n$. Fix a failure probability $\delta \in (0,1)$, and assume that*

$$n \geq \mathcal{O}(1)\max\left\{ B_{\ell,0}\frac{\kappa_\alpha}{\delta}\log\left(\frac{\kappa_\alpha B_\theta B_{\ell,0}^{-1}L_{\ell,0}}{\delta}\right),\ B_{\ell,0}\frac{\kappa_\eta}{\delta}\log\left(\frac{\kappa_\eta B_\theta B_{\ell,0}^{-1}L_{\ell,0}}{\delta}\right)\right\},$$

*where $\kappa_\alpha := q(\mu/\alpha)^{\frac{1}{1+r}}$, $\kappa_\eta := qL_\pi\mu/\eta$. Then with probability at least $1-\delta$, the imitation gap evaluated on $\xi \sim \mathcal{D}$ (drawn independently from $\{\xi_i\}_{i=1}^n$) satisfies*

$$\Gamma_T(\xi;\hat{\pi}_{\mathsf{TaSIL},0}) \leq \mathcal{O}(1)\mu\left(\frac{1}{\delta}\frac{B_{\ell,0}q\log\Big(B_\theta B_{\ell,0}^{-1}L_{\ell,0}n\Big)}{n}\right)^{1+r}.$$

*Proof.* Applying Corollary 4.1 to the $(B_\theta, B_{\ell,0}, q)$-Lipschitz parametric function class $B_{\ell,0}^{-1}(\ell_0^{\pi_\star} \circ \Pi_{\theta,0})$, we get that with probability at least $1-\delta/2$ over i.i.d. initial conditions $\xi_i \sim \mathcal{D}^n$,

$$\mathbb{E}_\xi\left[\max_{0\leq t\leq T-1}\|\Delta_t^{\pi_\star}(\xi;\hat{\pi}_{\mathsf{TaSIL},0})\|\right] \leq \mathcal{O}(1)B_{\ell,0}\frac{q\log\Big(B_\theta B_{\ell,0}^{-1}L_{\ell,0}n\Big)+\log(1/\delta)}{n}.$$

Applying Markov's inequality to $\max_t\|\Delta_t^{\pi_\star}(\xi;\hat{\pi}_{\mathsf{TaSIL},0})\|$, for a new draw $\xi \sim \mathcal{D}$, with probability greater than $1-\delta/2$,

$$\max_{0\leq t\leq T-1}\|\Delta_t^{\pi_\star}(\xi;\hat{\pi}_{\mathsf{TaSIL},0})\| \leq \frac{2}{\delta}\mathbb{E}_\xi\left[\max_{0\leq t\leq T-1}\|\Delta_t^{\pi_\star}(\xi;\hat{\pi}_{\mathsf{TaSIL},0})\|\right].$$

Thus applying a union bound over the two events, we have with probability greater than $1-\delta$ that

$$\max_{0\leq t\leq T-1}\|\Delta_t^{\pi_\star}(\xi;\hat{\pi}_{\mathsf{TaSIL},0})\| \leq \mathcal{O}(1)B_{\ell,0}q\frac{1}{\delta}\frac{\log\Big(B_\theta B_{\ell,0}^{-1}L_{\ell,0}n\Big)+\log(1/\delta)}{n}, \tag{31}$$

where we absorb numerical constants into $\mathcal{O}(1)$. We want $\max_t\|\Delta_t^{\pi_\star}(\xi;\hat{\pi}_{\mathsf{TaSIL},0})\|$ to satisfy the conditions in (7); that is,

$$\begin{aligned} &\max_{0\leq t\leq T-1}\mu\|\Delta_t^{\pi_\star}(\xi;\hat{\pi}_{\mathsf{TaSIL},0})\|^{1+r} \leq \alpha, \\ &\max_{0\leq t\leq T-1}2L_\pi\mu\|\Delta_t^{\pi_\star}(\xi;\hat{\pi}_{\mathsf{TaSIL},0})\|^{1+r} + \|\Delta_t^{\pi_\star}(\xi;\hat{\pi}_{\mathsf{TaSIL},0})\| \leq \eta. \end{aligned}$$

For notational convenience, we further require $\max_t \|\Delta_t^{\pi_\star}(\xi; \hat{\pi}_{\mathsf{TaSIL},0})\| \le 1$, so that

$$\max_t \|\Delta_t^{\pi_\star}(\xi; \hat{\pi}_{\mathsf{TaSIL},0})\|^{1+r} \le \max_t \|\Delta_t^{\pi_\star}(\xi; \hat{\pi}_{\mathsf{TaSIL},0})\| .$$

By assumption, since $\alpha/\mu \le 1$, satisfying the first condition above implies $\max_t \|\Delta_t^{\pi_\star}(\xi; \hat{\pi}_{\mathsf{TaSIL},0})\| \le 1$. We observe that for $n \ge \delta^{-1}\log(1/\delta)$ we have $\log n \ge 2\log(1/\delta)$, thus it suffices to absorb the $\log(1/\delta)$ term into $\log n$. Inserting the generalization bound (31) and shifting $n$ to the right-hand side of the above conditions, we have the following requirements on $n$:

$$\begin{aligned} n &\ge \mathcal{O}(1)\max\left\{ \left(\frac{\mu}{\alpha}\right)^{1/1+r} B_{\ell,0} q \frac{1}{\delta}\log\left(B_\theta B_{\ell,0}^{-1} L_{\ell,0} n\right),\right. \\ &\qquad \left.\left(\frac{L_\pi \mu}{\eta}\right) B_{\ell,0} q \frac{1}{\delta}\log\left(B_\theta B_{\ell,0}^{-1} L_{\ell,0} n\right)\right\} \\ &=: \mathcal{O}(1)\max\left\{ B_{\ell,0}\kappa_\alpha \frac{1}{\delta}\log\left(B_\theta B_{\ell,0}^{-1} L_{\ell,0} n\right),\right. \\ &\qquad \left. B_{\ell,0}\kappa_\eta \frac{1}{\delta}\log\left(B_\theta B_{\ell,0}^{-1} L_{\ell,0} n\right)\right\}, \end{aligned}$$

where we define $\kappa_\alpha = q(\mu/\alpha)^{\frac{1}{1+r}}$ and $\kappa_\eta = qL_\pi\mu/\eta$. Therefore, applying Lemma B.3 on each of the arguments of the maximum, setting $b = B_{\ell,0}\kappa_\alpha q/\delta$ (respectively $b = B_{\ell,0}\kappa_\eta q/\delta$) and $c = B_\theta B_{\ell,0}^{-1} L_{\ell,0}$, we get the following sample complexity bounds. For $n$ satisfying

$$n \ge \mathcal{O}(1)\max\left\{ B_{\ell,0}\frac{\kappa_\alpha}{\delta}\log\left(\frac{\kappa_\alpha B_\theta L_{\ell,0}}{\delta}\right),\ B_{\ell,0}\frac{\kappa_\eta}{\delta}\log\left(\frac{\kappa_\eta B_\theta L_{\ell,0}}{\delta}\right)\right\},$$

we have with probability greater than $1-\delta$

$$\Gamma_T(\xi; \hat{\pi}_{\mathsf{TaSIL},0}) \le \mathcal{O}(1)\ \mu\left(B_{\ell,0} q \frac{1}{\delta}\frac{\log\left(B_\theta B_{\ell,0}^{-1} L_{\ell,0} n\right)}{n}\right)^{1+r}.$$

This completes the proof. $\square$

**Theorem B.4** (Full version of Theorem 4.3)**.** *Assume that $\pi_\star \in \Pi_{\theta,p}$, and let the assumptions of Theorem 3.2 hold for all $\pi \in \Pi_{\theta,p}$. Let Equation (10) hold with constants $\mu, \alpha > 0$, and without loss of generality let $\left(\frac{\alpha}{\mu}\right)^r p! \le 2$. Let $\hat{\pi}_{\mathsf{TaSIL},p}$ be an empirical risk minimizer of $\ell_p^{\pi_\star}$ over the policy class $\Pi_{\theta,p}$ for initial conditions $\{\xi_i\} \overset{\text{i.i.d.}}{\sim} \mathcal{D}^n$. Fix a failure probability $\delta \in (0,1)$, and assume*

$$n \ge \mathcal{O}(1)\max_{j\le p}\max\left\{ B_j\frac{\kappa_{\alpha,j}}{\delta}\log\left(\frac{\kappa_{\alpha,j}B_\theta B_j^{-1}B_X L_j}{\delta}\right),\ B_j\frac{\kappa_{\eta,j}}{\delta}\log\left(\frac{\kappa_{\eta,j}B_\theta B_j^{-1}B_X L_j}{\delta}\right)\right\},$$

*where $\kappa_{\alpha,j} := \left(\frac{\mu}{\alpha}\right)^r \frac{pq}{j!}$ and $\kappa_{\eta,j} := \left(\frac{L_{\partial^p\pi}}{(p+1)!}\frac{\mu^{p+1}}{(j!)^{\frac{p+1}{r}}} + \frac{1}{j!}\right)\frac{pq}{\eta\delta}$. Then with probability at least $1-\delta$, the imitation gap evaluated on $\xi \sim \mathcal{D}$ (drawn independently from $\{\xi_i\}_{i=1}^n$) satisfies*

$$\Gamma_T(\xi; \hat{\pi}_{\mathsf{TaSIL},p}) \le \mathcal{O}(1)\mu\max_{j\le p}\left(\frac{p}{j!\delta}\frac{B_j q\log\left(B_\theta B_j^{-1}B_X L_j n\right)}{n}\right)^{1/r}.$$

*Proof.* Let us first define the following losses on a specific partial:

$$h_j^{\pi_\star}(\xi;\pi) := \max_{0\le t\le T-1}\left\|\partial_x^j \Delta_t^{\pi_\star}(\xi;\pi)\right\|.$$

We observe that by definition, $h_j^{\pi_\star}\circ\Pi_{\theta,p}$ is $2B_j$ bounded, and $h_j^{\pi_\star}$ is $B_X L_j$-Lipschitz with respect to $\Theta$ for $j \le p$, such that $0.5B_j^{-1}\left(h_j^{\pi_\star}\circ\Pi_{\theta,p}\right)$ is a $(B_\theta, 0.5B_j^{-1}B_X L_j, q)$-Lipschitz loss class. We note that since $\pi_\star \in \Pi_{\theta,p}$, we have for any dataset $\{\xi_i\} \subset \mathcal{X}$

$$\mathbb{E}_n\left[\ell_p^{\pi_\star}(\cdot\,; \hat{\pi}_{\mathsf{TaSIL},p})\right] =: \frac{1}{p+1}\sum_{j=0}^{p}\max_{0\le t\le T-1}\left\|\partial_x^j\Delta_t^{\pi_\star}(\xi;\pi)\right\| = 0,$$

which therefore implies $\mathbb{E}_n\left[h_j^{\pi^\star}(\cdot\,;\hat{\pi}_{\mathsf{TaSIL},p})\right] = 0$ for $j \leq p$. We now apply the same proof structure in Theorem 4.2 to each $0.5B_j^{-1}\left(h_j^{\pi^\star} \circ \Pi_{\theta,p}\right)$, where we have with probability greater than $1 - \frac{\delta}{2(p+1)}$ that

$$\mathbb{E}_\xi\left[\max_{0\leq t\leq T-1}\left\|\partial_x^j \Delta_t^{\pi^\star}(\xi;\hat{\pi}_{\mathsf{TaSIL},p})\right\|\right] \leq \mathcal{O}(1)B_j\frac{q\log\left(B_\theta B_j^{-1}B_X L_j n\right) + \log\left(\frac{2(p+1)}{\delta}\right)}{n}.$$

Applying Markov's inequality at level $\frac{\delta}{2(p+1)}$, we get with total probability greater than $1 - \frac{\delta}{p+1}$ over a new initial condition $\xi \sim \mathcal{D}$ that $\hat{\pi}_{\mathsf{TaSIL},p}$ satisfies the generalization bound

$$\max_{0\leq t\leq T-1}\left\|\partial_x^j \Delta_t^{\pi^\star}(\xi;\hat{\pi}_{\mathsf{TaSIL},p})\right\| \leq \mathcal{O}(1)B_j\frac{p+1}{\delta}\frac{q\log\left(B_\theta B_j^{-1}B_X L_j n\right) + \log\left(\frac{p+1}{\delta}\right)}{n}. \tag{32}$$

For each partial, we want to satisfy the constraints outlined in (11):

$$\begin{aligned}
&\max_{0\leq t\leq T-1}\max_{0\leq j\leq p}\mu\left(\frac{2}{j!}\left\|\partial_x^j\Delta_t^{\pi^\star}(\xi;\pi)\right\|\right)^{1/r} \leq \alpha, \\
&\max_{0\leq t\leq T-1}\max_{0\leq j\leq p}\frac{2L_{\partial^p\pi}\mu^{p+1}}{(p+1)!}\left(\frac{2}{j!}\left\|\partial_x^j\Delta_t^{\pi^\star}(\xi;\pi)\right\|\right)^{\frac{p+1}{r}} + \frac{2}{j!}\left\|\partial_x^j\Delta_t^{\pi^\star}(\xi;\pi)\right\| \leq \eta,
\end{aligned} \tag{33}$$

By assumption, we have $\left(\frac{\alpha}{\mu}\right)^r p! \leq 2$, and thus the first condition implies $\max_t\left\|\partial_x^j\Delta_t^{\pi^\star}(\xi;\pi)\right\| \leq 1$ for all $j \leq p$; in particular, this conveniently ensures $\left\|\partial_x^j\Delta_t^{\pi^\star}(\xi;\pi)\right\|^{\frac{p+1}{r}} \leq \left\|\partial_x^j\Delta_t^{\pi^\star}(\xi;\pi)\right\|$. Plugging the earlier generalization bound (32) into the above constraints and shifting $n$ to the RHS, and observing like earlier we may absorb the $\log(1/\delta)$ term into the $\log n$ term, we get:

$$\begin{aligned}
n &\geq \mathcal{O}(1)\max\left\{\left(\frac{\mu}{\alpha}\right)^r\frac{1}{j!}B_j\frac{p}{\delta}q\log\left(B_\theta B_j^{-1}L_j n\right),\right. \\
&\qquad\left.\left(\frac{L_{\partial^p\pi}}{(p+1)!}\frac{\mu^{p+1}}{(j!)^{\frac{p+1}{r}}} + \frac{1}{j!}\right)B_j\frac{p}{\delta}q\log\left(B_\theta B_j^{-1}L_j n\right)\right\} \\
&=: \mathcal{O}(1)\max\left\{B_j\frac{\kappa_{\alpha,j}}{\delta}\log\left(B_\theta B_j^{-1}L_j n\right), B_j\frac{\kappa_{\eta,j}}{\delta}q\log\left(B_\theta B_j^{-1}L_j n\right)\right\},
\end{aligned}$$

where we define $\kappa_{\alpha,j} = \left(\frac{\mu}{\alpha}\right)^r\frac{pq}{j!}$, $\kappa_{\eta,j} = \left(\frac{L_{\partial^p\pi}}{(p+1)!}\frac{\mu^{p+1}}{(j!)^{\frac{p+1}{r}}} + \frac{1}{j!}\right)\frac{pq}{\eta\delta}$. Therefore applying Lemma B.3, setting $b = B_j\frac{\kappa_{\alpha,j}}{\delta}$ (respectively $b = B_j\frac{\kappa_{\eta,j}}{\delta}$) and $c = B_\theta B_j^{-1}L_j$, for $n$ satisfying:

$$n \geq \mathcal{O}(1)\max\left\{B_j\frac{\kappa_{\alpha,j}}{\delta}\log\left(\frac{\kappa_{\alpha,j}B_\theta L_j}{\delta}\right), B_j\frac{\kappa_{\eta,j}}{\delta}\log\left(\frac{\kappa_{\eta,j}B_\theta L_j}{\delta}\right)\right\},$$

we have with probability greater than $1 - \frac{\delta}{p+1}$ that the conditions (33) are satisfied. To finish the proof, since we have with probability $1 - \frac{\delta}{p+1}$ that each $j$th partial difference satisfies the necessary conditions, we union bound over $0 \leq j \leq p$, such that we take a maximum over $j$ for the sample complexity and the resulting imitation gap. This gets us with probability greater than $1 - \delta$, for $n$ satisfying

$$n \geq \mathcal{O}(1)\max_{j\leq p}\max\left\{B_j\frac{\kappa_{\alpha,j}}{\delta}\log\left(\frac{\kappa_{\alpha,j}B_\theta L_j}{\delta}\right), B_j\frac{\kappa_{\eta,j}}{\delta}\log\left(\frac{\kappa_{\eta,j}B_\theta L_j}{\delta}\right)\right\},$$

that the following bound on the imitation gap holds

$$\Gamma_T(\xi;\hat{\pi}_{\mathsf{TaSIL},p}) \leq \mathcal{O}(1)\max_{j\leq p}\mu\left(\frac{p}{j!\delta}\frac{B_j q\log\left(B_\theta B_j^{-1}B_X L_j n\right)}{n}\right)^{1/r}.$$

This completes the proof. □

# C Using finite-differencing to approximate derivatives

## C.1 Satisfying Conditions (11) and (12) with approximate derivatives

We recall the closeness conditions on the partials along expert trajectories that guarantee bounds on the imitation gap:

$$\max_{0\le t\le T-1}\max_{0\le j\le p}\mu\left(\frac{2}{j!}\left\|\partial_x^j\Delta_t^{\pi_\star}(\xi;\pi)\right\|\right)^{1/r}\le\alpha, \tag{11}$$

$$\max_{0\le t\le T-1}\max_{0\le j\le p}\frac{2L_{\partial^p\pi}\mu^{p+1}}{(p+1)!}\left(\frac{2}{j!}\left\|\partial_x^j\Delta_t^{\pi_\star}(\xi;\pi)\right\|\right)^{\frac{p+1}{r}}+\frac{2}{j!}\left\|\partial_x^j\Delta_t^{\pi_\star}(\xi;\pi)\right\|\le\eta. \tag{12}$$

If we have access to approximate derivatives of the expert $\widehat{\partial_x^j\pi_\star}(x)$ such that

$$\left\|\widehat{\partial_x^j\pi_\star}(x)-\partial_x^j\pi_\star(x)\right\|\le b<1$$

for all $x\in\mathbb{R}^d$, then it suffices to tighten the constraints by some function of $b$ such that minimizing with respect to the approximate partial derivatives will still result in the deviation from the true derivatives satisfying the requisite bounds. Let us define

$$\widehat{\partial_x^j\Delta_t^{\pi_\star}}(\xi;\pi):=\partial_x^j\pi(x_t^{\pi_\star}(\xi))-\widehat{\partial_x^j\pi_\star}(x_t^{\pi_\star}(\xi)),$$

such that

$$\begin{aligned}\left\|\partial_x^j\Delta_t^{\pi_\star}(\xi;\pi)\right\|&\le\left\|\widehat{\partial_x^j\Delta_t^{\pi_\star}}(\xi;\pi)\right\|+\left\|\widehat{\partial_x^j\pi_\star}(x_t^{\pi_\star}(\xi))-\partial_x^j\pi_\star(x_t^{\pi_\star}(\xi))\right\|\\&\le\left\|\widehat{\partial_x^j\Delta_t^{\pi_\star}}(\xi;\pi)\right\|+b.\end{aligned}$$

Therefore, it suffices to match the approximate partial derivatives such that

$$\max_{0\le t\le T-1}\max_{0\le j\le p}\mu\left(\frac{2}{j!}\left\|\widehat{\partial_x^j\Delta_t^{\pi_\star}}(\xi;\pi)\right\|\right)^{1/r}\le\hat\alpha,$$

$$\max_{0\le t\le T-1}\max_{0\le j\le p}\frac{2L_{\partial^p\pi}\mu^{p+1}}{(p+1)!}\left(\frac{2}{j!}\left\|\widehat{\partial_x^j\Delta_t^{\pi_\star}}(\xi;\pi)\right\|\right)^{\frac{p+1}{r}}+\frac{2}{j!}\left\|\widehat{\partial_x^j\Delta_t^{\pi_\star}}(\xi;\pi)\right\|\le\hat\eta,$$

where, provided $\left\|\widehat{\partial_x^j\Delta_t^{\pi_\star}}(\xi;\pi)\right\|<1$:

$$\hat\alpha:=\left(\alpha^r-\frac{2\mu^r}{j!}b\right)^{1/r},\ \ \hat\eta:=\eta-\left(\frac{2L_{\partial^p\pi}\mu^{p+1}}{(p+1)!}\left(\frac{2}{j!}\right)^{\frac{p+1}{r}}+\frac{2}{j!}\right)b.$$

A similar bound holds if we also do not have access to the exact derivatives of the learned policy. In practice, these bounds tell us qualitatively that if a sufficiently precise estimate of the derivatives is used, such as through finite differencing, then the imitation gap bounds in Theorem 3.2 still hold.

## C.2 Practical approaches for approximating derivatives

Minimizing $\sum_{j=1}^k\|\partial_x^j\Delta_t^{\pi_\star}(\xi;\pi)\|$ can be approximated provided $\pi_\star$ can be evaluated at points $\{x_t(\xi)+\delta_i\}_{i=1}^N$ by minimizing the finite difference loss:

$$\ell_{p,\mathsf{FD}}(\xi;\pi,\{\delta_i\}_{i=1}^N):=\max_{1\le i\le N}\|\pi_\star(x_t(\xi)+\delta_i)-\pi_\star(x_t(\xi))-(\pi(x_t(\xi)+\delta_i)-\pi(x_t(\xi))\|,$$

where the $\{\delta_i\}$ are chosen such that the Taylor expansion

$$\begin{aligned}\sum_{j=1}^p\frac{1}{p!}\partial_x^j\Delta_t^{\pi_\star}(\xi;\pi)\cdot\delta_i^{\otimes j}=\pi_\star(x_t(\xi)+\delta_i)-\pi_\star(x_t(\xi))&\\-(\pi(x_t(\xi)+\delta_i)-\pi(x_t(\xi)))-\frac{R_{p+1}(\delta_i)}{(p+1)!},\ \forall\,1\le i\le N,&\end{aligned}$$

forms a linearly independent system of equations in the derivative parameters. Here, $R_{p+1}(\delta_i)$ denotes the Taylor remainder, which satisfies the inequality $\|R_{p+1}(\delta_i)\| \leq 2L_{\partial^p \pi}\|\delta_i\|^{p+1}$ by Assumption 3.2.

For the case $p = 1$, we can stack the $\{\delta_i\}$ into a matrix $S$, the finite differences into a matrix $M$ and the remainders into a matrix $R$ to write

$$\partial_x^j \Delta_t^{\pi_\star}(\xi; \pi) S = M - R.$$

Provided the $\{\delta_i\}$ are chosen such that $S$ is invertible, the operator norm of $\partial_x^j \Delta_t^{\pi_\star}(\xi; \pi)$ can be upper bounded

$$\begin{aligned} \|\partial_x^j \Delta_t^{\pi_\star}(\xi; \pi)\| &= \|MS^{-1} - RS^{-1}\| \\ &\leq \|S^{-1}\|(\|M\| + \|R\|) \\ &\leq \|S^{-1}\|(\|M\| + L_{\partial^p \pi}\|S\|^2). \end{aligned}$$

For instance, using a standard basis $S = \varepsilon I$ as the finite difference perturbations yields the following bound on the operator norm:

$$\|\partial_x^j \Delta_t^{\pi_\star}(\xi; \pi)\| \leq \frac{1}{\varepsilon}\|M\| + \varepsilon L_{\partial^2 \pi},$$

where $M$ is the stacked error matrix at the finite differences. Therefore by ensuring sufficiently small $\varepsilon$ and finite difference loss, the bound on the Jacobian error can be made arbitrarily small.

Alternatively, if the finite differences $\delta_i$ are sampled from a uniform distribution on a sphere of radius $\varepsilon$ for each evaluation of $\ell_{1,\mathsf{FD}}$ (i.e, the expert can be cheaply queried during training), Woolfe et al. [37, Theorem 3.15] shows that

$$\|\partial_x^i \Delta_t^{\pi_\star}(\xi; \pi)\| \leq \frac{0.8\sqrt{d}}{\zeta^{1/N}} \left( \frac{1}{\varepsilon} \ell_{p,\mathsf{FD}}(\xi; \pi, \{\delta_i\}_{i=1}^N) + \varepsilon L_{\partial^2 \pi} \right),$$

with probability $1 - \zeta$, where $d$ is the dimensionality of the state space. This suggests that provided the expert can be requeried each iteration, $N \ll d$ finite differencing terms can be used.

### C.3 Experimental results

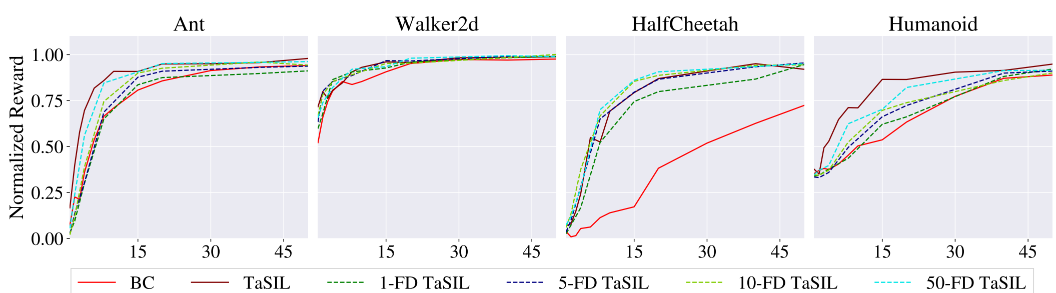


**Figure 3:** Mean normalized reward for vanilla Behavior Cloning, Behavior Cloning with TaSIL loss, and Behavior Cloning with finite-differencing based TaSIL. The average across 5 random seeds is shown.

We perform several experiments using finite differencing to approximate minimizing the higher order derivatives. Figure 3 shows configurations with 1, 5, 10, and 50 difference vectors across different MuJoCo environments. The different vectors drawn from a uniform distribution over a sphere of radius 0.01. The difference vectors were drawn once for each state-action pair and did not change during training. This was done to simulate the effect of getting progressively closer to using the full standard basis with additional finite differencing terms.

For Walker2d and HalfCheetah with a state dimension of 17, finite difference with a single random vector is sufficient to achieve performance on par with TaSIL using the explicit Jacobians. Humanoid and Ant with higher dimensional observation spaces (376 and 111 dimensions respectively) also show significant improvements the more finite differences are used.

## D Additional information for stability experiments

**Theorem D.1.** *For $\eta \in [0,1)$, the system*

$$x_{t+1} = \eta x_t + (1-\eta) \cdot \frac{\gamma(\|h(x_t) + u_t\|)}{\|h(x_t) + u_t\|}(h(x_t) + u_t), \tag{34}$$

*is $\delta$-ISS around $\pi_\star(x) = -h(x)$ with class $\mathcal{K}$ function $\gamma$.*

*Proof.* We use the shorthand $x_t(\xi_1) := x_t(\xi_1, \{u_k\}_{k=0}^{t-1})$ and $x_t(\xi_2) := x_t(\xi_2, \{0\}_{k=0}^{t-1})$. We can prove this directly using

$$\begin{aligned}
&\|x_{t+1}(\xi_1) - x_{t+1}(\xi_2)\| \\
&= \left\|\eta(x_t(\xi_1) - x_t(\xi_2)) + (1-\eta) \cdot \frac{\gamma(\|h(x_t) + u_t\|)}{\|h(x_t) + u_t\|}(h(x_t) + u_t)\right\| \\
&\leq \eta\|x_t(\xi_1) - x_t(\xi_2)\| + (1-\eta)\gamma(\|h(x_t) + u_t\|).
\end{aligned}$$

Since $x_0(\xi_1) = \xi_1$ and $x_1(\xi_2) = \xi_2$, repeated composition of this upper bound yields

$$\begin{aligned}
\|x_t(\xi_1) - x_t(\xi_2)\| &\leq \eta^t\|\xi_1 - \xi_2\| + \sum_{k=0}^{t-1}\eta^{t-1-k}(1-\eta)\gamma(\|h(x_k) + u_k\|) \\
&\leq \eta^t\|\xi_1 - \xi_2\| + \max_{0\leq k\leq t-1}\gamma(\|h(x_k) + u_k\|) \\
&= \eta^t\|\xi_1 - \xi_2\| + \gamma\left(\max_{0\leq k\leq t-1}\|h(x_k) + u_k\|\right).
\end{aligned}$$

□

**Experiment details** The expert MLP has two hidden layers of 32 units each with GELU activations while the learned policy has three hidden layers of 64 units and GELU activations. A tanh nonlinearity was applied to obtain the final policy output. Expert weights were initialized using Lecun Normal initialization LeCun et al. [38] for the kernels and drawn form a normal distribution with $\Sigma = 0.1I$ for the biases. The learned policy weights are initialized using orthogonal intialization for the kernels and zeros for the bias.

For all stability experiments we train on 20 trajectories of length $T = 100$. Initial states were sampled from a standard normal distribution. The state-action pairs are shuffled independently into batches of size 100 and weight updates were performed using the Adam optimizer with $\beta_1 = 0.9, \beta_2 = 0.999$, and $\varepsilon = 1 \times 10^{-4}$. The training rate was decayed with a cosine learning rate decay using an initial rate of $\alpha = 1 \times 10^{-3}$. We additionally employed $\ell^2$ weight regularization with $\lambda = 0.01$. All training is run for 4500 iterations on our internal cluster.

To weight the various derivative terms for the different TaSIL losses we use $\lambda_0 = 1, \lambda_1 = 1$, and $\lambda_2 = 10$.

## E Additional information for MuJoCo experiments

We use a $\beta$-decay-rate of $p = 0.5$ for DAgger and $\alpha = T\mathrm{Tr}[\Sigma_k]$ for DART, the same parameters used by Laskey et al. [7] for their Mujoco experiments. For DART, we use an independent sample of 5 trajectories to update the noise statistics. The same optimization setup from the stability experiments was used, with a batch size of 100, Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 1 \times 10^{-4}$, cosine learning rate scheduling with an initial learning rate of $1 \times 10^{-3}$ decaying over the entire training duration of 4500 epochs, and $\ell^2$ weight regularization with $\lambda = 0.01$.

We train over 4500 epochs for all experiments with a training and test trajectory length of $T = 300$. All TaSIL losses use $\lambda_0 = 1$. $\lambda_1 = 0.01$ is used for the jacobian term in the 1-TaSIL loss.

Similar to Laskey et al. [7], DAgger rollout policies and DART noise statistics were updated sparsely rather than after every trajectory. We performed updates after $1, 5, 20$, and $30$ trajectories.