Learning and Control with Safety and Stability Guarantees for Nonlinear Systems

Part II: Applications

Stephen Tu
Google Brain Robotics
June 6th, 2021
F&MG Data-Driven Control Summer School

Google Research

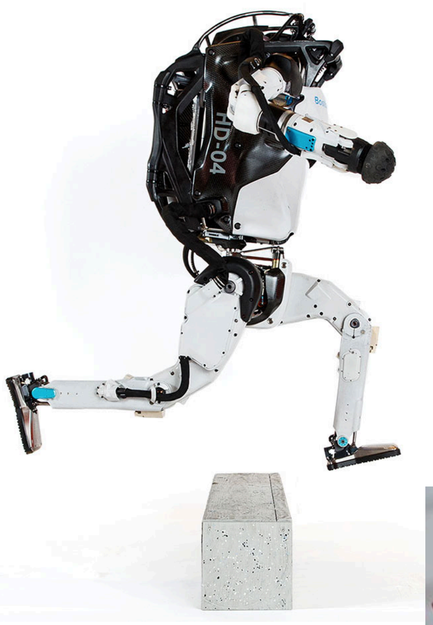
Overview

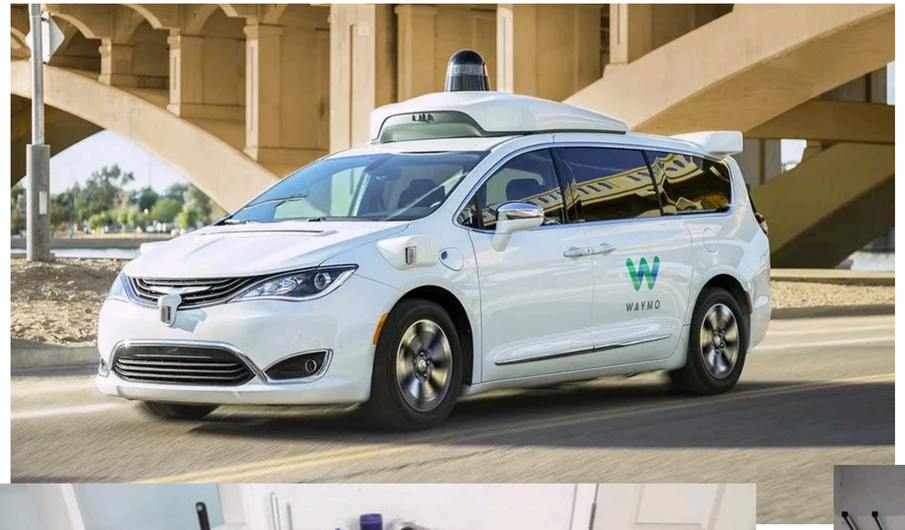
- In Part I, we covered a wide variety of foundational tools:
    - Concentration inequalities.
    - Uniform convergence / empirical process theory.
    - Nonlinear stability theory.
- In Part II, we will apply these tools to several applications:
    - Learning stability certificates from data.
    - Stability constrained imitation learning.
    - Regret bounds for adaptive nonlinear control.

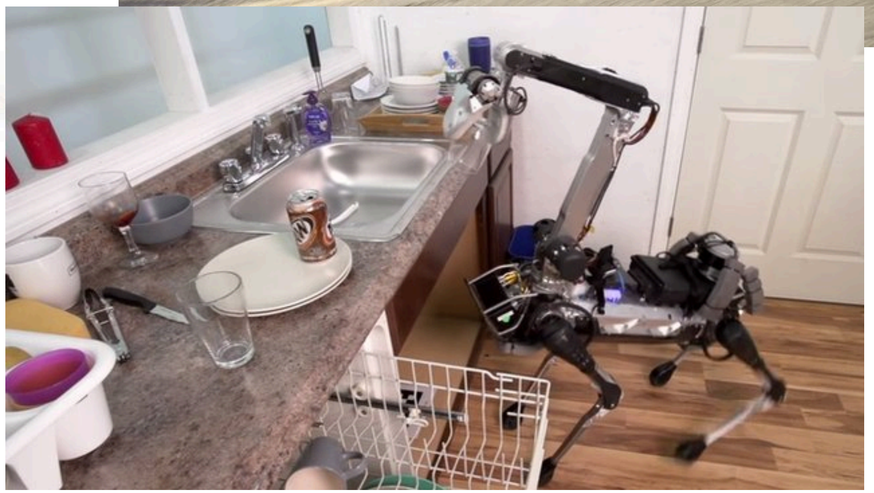
Learning Stability Certificates from Data

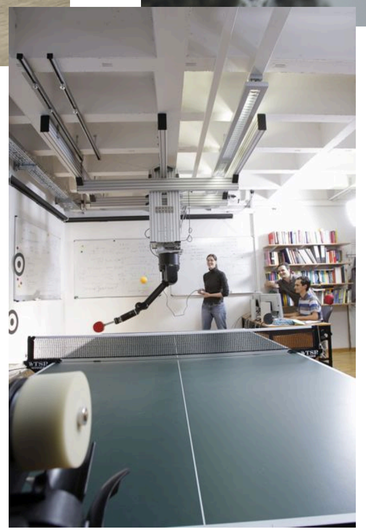
Nicholas M. Boffi, Stephen Tu, Nikolai Matni, Jean-Jacques E. Slotine, and Vikas Sindhwani
arxiv.org/abs/2008.05952

Lab-to-Reality Transfer?









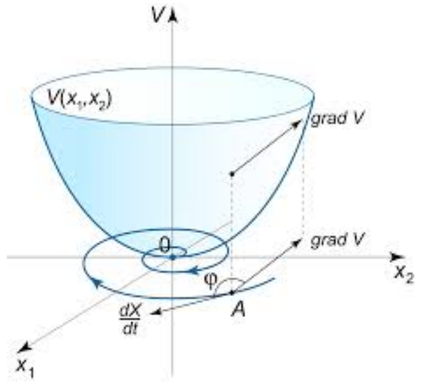
Lab-to-Reality Transfer?

- **How do we deploy these demos out into the real world at scale?**
- A lot of research has focused on:
    - Algorithmic improvements
    - Better models/simulators
- Less emphasis on **provable** stability and safety guarantees.
    - Understanding limits is key to real-world deployment.
- **This talk:** an algorithmic framework for provably certifying guarantees of a dynamical system **from trajectory data**.

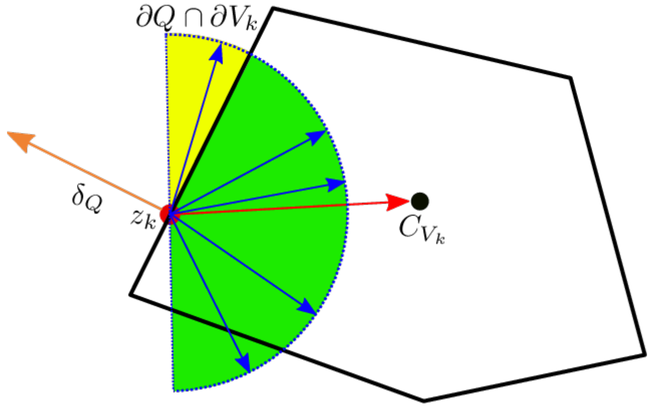
Problem formulation

- **Dynamical system:** $\dot{x} = f(x), x \in \mathbb{R}^n$.
- Let $\varphi_t(\xi)$ denote the **flow-map** at time $t \in T$ starting at $x(0) = \xi \in X$.
- **Certificate function space:** $\mathscr{V} \subseteq C^1(\mathbb{R}^n, \mathbb{R})$.
- **Evaluation function** $h \in C(\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^n, \mathbb{R})$.
- **Goal:** Find $V \in \mathscr{V}$ such that:
  $h(\varphi_t(\xi), \dot{\varphi}_t(\xi), V(\varphi_t(\xi)), \nabla V(\varphi_t(\xi))) \leqslant 0 \;\; \forall \xi \in X, t \in T.$

Why?



- V can be used to **certify** desirable system behavior.
- **Lyapunov stability:** Conditions (a) & (b) imply $x = 0$ is (locally) asymptotically stable:
  - (a) $V(x) > 0,\ V(0) = 0,$
  - (b) $\langle \nabla V(x), f(x) \rangle \leqslant -\alpha(x)\ \ \forall x \in \mathbb{B}_2^n(0,\varepsilon)$ [α is class $\mathscr{K}$].
- **Barrier function:** Conditions (a) & (b) imply $\mathscr{C} := \{x : V(x) \geqslant 0\}$ is invariant:
  - (a) $V(x) = 0 \Longrightarrow \nabla V(x) \neq 0,$
  - (b) $\langle \nabla V(x), f(x) \rangle \geqslant -\alpha(V(x))\ \ \forall x \in \mathscr{C}$ [α is class $\mathscr{K}$].



Why?

- Recall we want to find $V \in \mathscr{V}$ such that:
  $h(\varphi_t(\xi), \dot{\varphi}_t(\xi), V(\varphi_t(\xi)), \nabla V(\varphi_t(\xi))) \leqslant 0 \;\; \forall \xi \in X, t \in T.$
- **Lyapunov stability:** set $h(x, \dot{x}, V(x), \nabla V(x)) = \langle \nabla V(x), \dot{x} \rangle + \alpha(x).$
- **Barrier function:** set $h(x, \dot{x}, V(x), \nabla V(x)) = -\langle \nabla V(x), \dot{x} \rangle - \alpha(V(x)).$
- Note: can handle **incremental stability** by drawing **pairs** of initial conditions.

What about sum-of-squares programming?

- Given knowledge of dynamics $f(x)$, can search for V satisfying $h \leqslant 0$ using sum-of-squares programming.
    - Caveat: only if $f(x), V(x), h(x)$ are polynomials.
    - Caveat: only if degree of polynomials and state dimension is not too large (SDP does not scale well in # of decision variables).
- **Goal:** search for V using only trajectories of $\dot{x} = f(x)$.

Problem formulation

- Let $\mathscr{D}$ denote a distribution over X.
- **Observations:** $D = \{\{\varphi_t(\xi_i)\}_{t\in T}\}_{i=1,\dots,m}$, where $\xi_1, \dots, \xi_m$ are i.i.d. samples from $\mathscr{D}$.
- We study the following empirical estimate:

$$\text{find}_{V\in\mathscr{V}} \ \text{s.t.} \ h(\varphi_t(\xi_i), \dot{\varphi}_t(\xi_i), V(\varphi_t(\xi_i)), \nabla V(\varphi_t(\xi_i))) \leqslant -\gamma$$
$$i = 1,\dots,m, \ t \in T.$$

- Supposing problem is feasible, let $\hat{V}_m \in \mathscr{V}$ denote a solution. Define the **generalization error** of $\hat{V}_m$ as:

$$\text{err}(\hat{V}_m) := \mathbb{P}_{\xi\sim\mathscr{D}}\left\{\max_{t\in T} h\left(\varphi_t(\xi), \dot{\varphi}_t(\xi), \hat{V}_m(\varphi_t(\xi)), \nabla \hat{V}_m(\varphi_t(\xi))\right) > 0\right\}.$$

Problem formulation

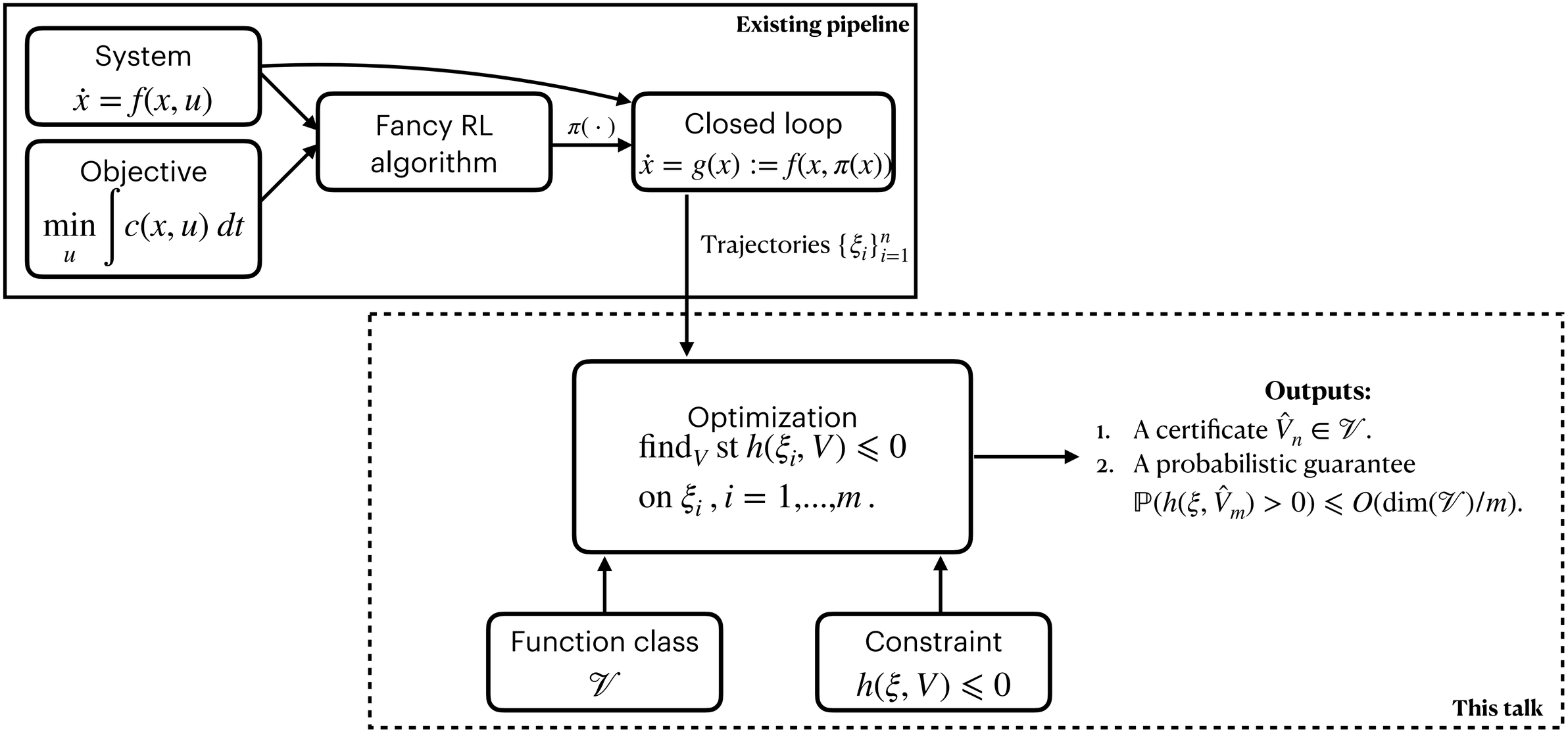
- Generalization error:

$$\mathrm{err}(\hat{V}_m) := \mathbb{P}_{\xi\sim\mathscr{D}}\left\{\max_{t\in T} h\left(\varphi_t(\xi), \dot{\varphi}_t(\xi), \hat{V}_m(\varphi_t(\xi)), \nabla\hat{V}_m(\varphi_t(\xi))\right) > 0\right\}.$$

- Measures the probability that $\hat{V}_m$ **fails to certify** a trajectory $\{\varphi_t(\xi)\}_{t\in T}$ when initialized at $\xi \sim \mathscr{D}$.

- **Question:** How many trajectories $m = m(\varepsilon, \delta)$ do we need such that $\mathrm{err}(\hat{V}_m) \leqslant \varepsilon$ with probability at least $1 - \delta$ (over $\xi_1, \ldots, \xi_m$)?

Algorithmic Framework



Supervised learning reduction

- Let $h(\xi, V)$ be shorthand for:

$$h(\xi, V) := \max_{t \in T} h\left(\varphi_t(\xi), \dot{\varphi}_t(\xi), V(\varphi_t(\xi)), \nabla V(\varphi_t(\xi))\right).$$

- For any $V \in \mathscr{V}$, define the empirical (margin) risk $\hat{R}_\gamma(V)$ and true risk $R(V)$ as:

$$\hat{R}_\gamma(V) := \frac{1}{m} \sum_{i=1}^{m} \mathbf{1}\{h(\xi_i, V) > -\gamma\}, \; R(V) := \mathbb{P}_{\xi \sim \mathscr{D}}\left(h(\xi, V) > 0\right).$$

- By definition of $\hat{V}_m$, we have $\hat{R}_\gamma(\hat{V}_m) = 0$ (all constraints are satisfied).

- To proceed, we want to upper bound $R(\hat{V}_m)$ by some expression containing $\hat{R}_\gamma(\hat{V}_m)$ (via a uniform convergence argument).

Supervised learning reduction

- We will use a result from [Srebro et al. 2010] which yields uniform convergence for H -smooth losses (H-Lipschitz gradients).
- Let us review the setup of supervised learning:
  - $\mathscr{H} \subseteq F(\mathscr{X}, \mathbb{R})$ is a hypothesis class.
  - $\phi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_{\geqslant 0}$ is a non-negative loss function.
  - $\mathscr{D}$ is a distribution over $\mathscr{X} \times \mathbb{R}$.
  - $L(h) := \mathbb{E}_{(x,y)\sim\mathscr{D}} \phi(h(x), y)$.
  - $\hat{L}(h) = \dfrac{1}{m} \displaystyle\sum_{i=1}^{m} \phi(h(x_i), y_i)$ with $(x_i, y_i) \sim_{\text{i.i.d.}} \mathscr{D}$.

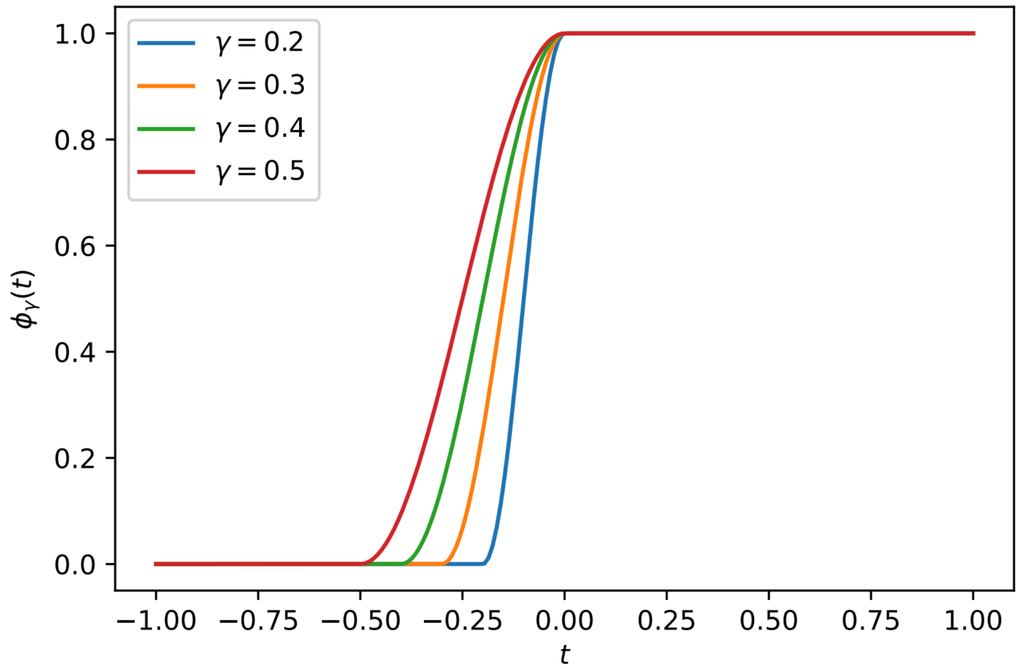
Supervised learning reduction

- **Theorem** [Srebro et al. 2010]:
  - Suppose $\sup_{x \in \mathcal{X}} \sup_{y \in \mathbb{R}} \sup_{h \in \mathcal{H}} |\phi(h(x), y)| \leqslant b$ and that $t \mapsto \phi(t, y)$ is H-smooth for every $y \in \mathbb{R}$.
  - Define $\mathscr{R}_m(\mathcal{H}) := \sup_{x_1, \ldots, x_m \in \mathcal{X}} \mathbb{E}_{\{\varepsilon_i\}} \sup_{h \in \mathcal{H}} \frac{1}{m} \sum_{i=1}^{m} \varepsilon_i h(x_i)$.
  - Define $\Gamma(\mathcal{H}, m, \delta) := H \log^3 m \cdot \mathscr{R}_m^2(\mathcal{H}) + \frac{b \log(1/\delta)}{m}$.
  - Then, with probability at least $1 - \delta$ (over $(x_1, y_1), \ldots, (x_m, y_m)$ drawn i.i.d. from $\mathscr{D}$), for any $h \in \mathcal{H}$,

    $L(h) \leqslant \hat{L}(h) + O(1)\sqrt{\hat{L}(h)\Gamma(\mathcal{H}, m, \delta)} + O(1)\Gamma(\mathcal{H}, m, \delta)$.
  - **Fast rate:** Therefore, if $\hat{h} \in \arg\min_{h \in \mathcal{H}} \hat{L}(h)$ satisfies $\hat{L}(\hat{h}) = 0$, then $L(\hat{h}) \leqslant O(1)\Gamma(\mathcal{H}, m, \delta)$.

Supervised learning reduction

- Define $\phi_\gamma(t) := \begin{cases} 0 & \text{if } t \leqslant -\gamma, \\ \dfrac{1+\cos(-\pi t/\gamma)}{2} & \text{if } t \in (-\gamma, 0), \\ 1 & \text{if } t \geqslant 0 \end{cases}$.



- Observe the relationship for all $t \in \mathbb{R}$:

$$\mathbf{1}\{t > 0\} \leqslant \phi_\gamma(t) \leqslant \mathbf{1}\{t > -\gamma\}.$$

- Therefore:
  - $\mathbb{E}\mathbf{1}\{h(\xi, \hat{V}_m) > 0\} \leqslant \mathbb{E}\phi_\gamma(h(\xi, \hat{V}_m))$
  - $\dfrac{1}{m}\sum_{i=1}^{m} \phi_\gamma(h(\xi_i, \hat{V}_m)) \leqslant \dfrac{1}{m}\sum_{i=1}^{m} \mathbf{1}\{h(\xi_i, \hat{V}_m) > -\gamma\} = 0.$

Supervised learning reduction

- We can apply [Srebro et al. 2010]'s result as follows.
- First, $\mathsf{err}(\hat{V}_m) = \mathbb{P}(h(\xi, \hat{V}_m) > 0) = \mathbb{E}\mathbf{1}\{h(\xi, \hat{V}_m) > 0\} \leqslant \mathbb{E}\phi_\gamma(h(\xi, \hat{V}_m))$.
- Next, one can check that ϕ_γ is $\dfrac{\pi^2}{4\gamma^2}$-smooth.
- Define $\mathscr{R}_m(\mathscr{V}) := \sup_{\xi_1,\ldots,\xi_m \in X} \mathbb{E}_{\{\varepsilon_i\}} \sup_{V \in \mathscr{V}} \dfrac{1}{m} \sum_{i=1}^{m} \varepsilon_i h(\xi_i, V)$.
- [Srebro et al. 2010]'s result implies $\mathsf{err}(\hat{V}_m) \leqslant O(1)\left(\dfrac{\log^3 m}{\gamma^2} \mathscr{R}_m^2(\mathscr{V}) + \dfrac{\log(1/\delta)}{m} \right)$.
- It remains to bound the Rademacher complexity $\mathscr{R}_m(\mathscr{V})$.

Rademacher complexity bounds

- We bound the Rademacher complexity $\mathscr{R}_m(\mathscr{V})$ by using Dudley's inequality.
- We need the following two assumptions.
- **Assumption** (stability i.s.L.): There exists a compact S such that $\bigcup_{t\in T} \varphi_t(X) \subseteq S$.
- **Assumption** (regularity of $\mathscr{V}$):
  - $\sup_{V\in\mathscr{V}} \sup_{x\in S} |V(x)| \leqslant B_V$
  - $\sup_{V\in\mathscr{V}} \sup_{x\in S} \|\nabla V(x)\|_2 \leqslant B_{\nabla V}$.

Rademacher complexity bounds

- Let L_h denote an upper bound such that $(V, \nabla V) \mapsto h(x, f(x), V, \nabla V)$ is L_h -Lipschitz for all $x \in S$.
- Define the following norm on $\mathscr{V}$: $\|V\|_{\mathscr{V}} := \sup_{x \in S} \left\| \begin{bmatrix} V(x) \\ \nabla V(X) \end{bmatrix} \right\|_2$.
- **Claim:** for any $\xi \in X$ and $V_1, V_2 \in \mathscr{V}$, we have:

$$|h(\xi, V_1) - h(\xi, V_2)| \leqslant L_h \|V_1 - V_2\|_{\mathscr{V}}.$$

- **Proof of claim** $|h(\xi, V_1) - h(\xi, V_2)| \leqslant L_h \|V_1 - V_2\|_{\mathscr{V}}$:
- Define $h_t(\xi, V) := h\left(\varphi_t(\xi), \dot{\varphi}_t(\xi), V(\varphi_t(\xi)), \nabla V(\varphi_t(\xi))\right)$.
- Let t_i be such that $h(\xi, V_i) = h_{t_i}(\xi, V_i)$ for $i \in \{1,2\}$.
- Observe that:

$$
\begin{aligned}
h(\xi, V_1) - h(\xi, V_2) &= h_{t_1}(\xi, V_1) - h_{t_2}(\xi, V_2) \\
&\leqslant h_{t_1}(\xi, V_1) - h_{t_1}(\xi, V_2) && [h_{t_1}(\xi, V_2) \leqslant h_{t_2}(\xi, V_2)] \\
&\leqslant L_h \left\| \begin{bmatrix} V_1(\varphi_{t_1}(\xi)) - V_2(\varphi_{t_1}(\xi)) \\ \nabla V_1(\varphi_{t_1}(\xi)) - \nabla V_2(\varphi_{t_1}(\xi)) \end{bmatrix} \right\|_2 && [\text{definition of } L_h] \\
&\leqslant L_h \sup_{x \in S} \left\| \begin{bmatrix} V_1(x) - V_2(x) \\ \nabla V_1(x) - \nabla V_2(x) \end{bmatrix} \right\|_2 && [\text{stability i.s.L.}] \\
&= L_h \|V_1 - V_2\|_{\mathscr{V}}.
\end{aligned}
$$

- An identical bound holds for $h(\xi, V_2) - h(\xi, V_1)$. ■

Rademacher complexity bounds

- Therefore by Dudley's inequality, we have:

$$\mathscr{R}_m(\mathscr{V}) \leqslant \frac{24L_h}{\sqrt{m}} \int_0^\infty \sqrt{\log N(\varepsilon; \mathscr{V}, \|\cdot\|_{\mathscr{V}})}\, d\varepsilon.$$

- Here, $N(\varepsilon; \mathscr{V}, \|\cdot\|_{\mathscr{V}})$ is the **covering number** of $\mathscr{V}$ in the $\|\cdot\|_{\mathscr{V}}$-norm at resolution ε.

- To further specialize this bound, we need to specialize the function class $\mathscr{V}$.

Rademacher complexity bounds

- Consider $\mathscr{V} = \{V_\theta(x) = g(x,\theta) : \theta \in \mathbb{R}^k,\ \|\theta\|_2 \leqslant B_\theta\}$ with $g \in C^1(\mathbb{R}^n \times \mathbb{R}^k, \mathbb{R})$.
- Let L_g be such that $\theta \mapsto g(x,\theta)$ is L_g-Lipschitz on $\mathbb{B}_2^k(0,B_\theta)$ for all $x \in S$.
- Let $L_{\nabla g}$ be such that $\theta \mapsto \nabla_x g(x,\theta)$ is $L_{\nabla g}$-Lipschitz on $\mathbb{B}_2^k(0,B_\theta)$ for all $x \in S$.
- **Claim:** $\mathscr{R}_m(\mathscr{V}) \leqslant 32.5 B_\theta L_h (L_g + L_{\nabla g})\sqrt{\frac{k}{m}}$.

Rademacher complexity bounds

- **Proof:** by Lipschitz assumptions, for all $\theta_1, \theta_2 \in \mathbb{B}_2^k(0, B_\theta)$:

$$\|V_1 - V_2\|_{\mathscr{V}} = \sup_{x \in S} \|g(x, \theta_1) - g(x, \theta_2)\|_2 \leqslant (L_g + L_{\nabla g}) \|\theta_1 - \theta_2\|_2.$$

- Therefore:

$$\begin{aligned}
\mathscr{R}_m(\mathscr{V}) &\leqslant \frac{24 L_h}{\sqrt{m}} \int_0^\infty \sqrt{\log N(\varepsilon; \mathscr{V}, \|\cdot\|_{\mathscr{V}})}\, d\varepsilon && \text{[Dudley's inequality]} \\
&\leqslant \frac{24 B_\theta L_h (L_g + L_{\nabla g})}{\sqrt{m}} \int_0^\infty \sqrt{\log N(\varepsilon; \mathbb{B}_2^k(0,1), \|\cdot\|_2)}\, d\varepsilon && \text{[Lipschitz inequality above]} \\
&\leqslant 24 B_\theta L_h (L_g + L_{\nabla g}) \sqrt{\frac{k}{m}} \int_0^1 \sqrt{\log(1 + 2/\varepsilon)}\, d\varepsilon && [N(\varepsilon; \mathbb{B}_2^k(0,1), \|\cdot\|_2) \leqslant (1 + 2/\varepsilon)^k] \\
&\leqslant 32.5 B_\theta L_h (L_g + L_{\nabla g}) \sqrt{\frac{k}{m}}. && \blacksquare
\end{aligned}$$

Rademacher complexity bounds

- Let us add some more structure to $\mathscr{V}$.
- A natural way to enforce non-negativity of $\mathscr{V}$ is to use the representation $\mathscr{V} = \{x \mapsto \psi(x)^T Q \psi(x) : Q \in \mathbb{R}^{p \times p},\ Q = Q^T \succcurlyeq 0,\ \|Q\|_F \leqslant B_Q\}$, where $\psi \in C^1(\mathbb{R}^n, \mathbb{R}^p)$ is a feature map.
- If we use the previous analysis, we have $\mathscr{R}_m(\mathscr{V}) \lesssim \sqrt{\frac{p^2}{m}}$ since there are $\Theta(p^2)$ parameters.
- **Claim:** $\mathscr{R}_m(\mathscr{V}) \lesssim \sqrt{\frac{p \log^2 p}{m}}$.

Rademacher complexity bounds

- **Proof:** Let $\sup_{x \in S} \|\psi(x)\|_2 \leqslant B_\psi$ and $\sup_{x \in S} \left\| \frac{\partial \psi}{\partial x}(x) \right\|_{\text{op}} \leqslant B_{D\psi}$.
- We have that $\nabla V(x) = 2 \frac{\partial \psi}{\partial x}^T (x) Q \psi(x)$.
- Not hard to check that $\|V_1 - V_2\|_{\mathscr{V}} \leqslant B_\psi (B_\psi + 2B_{D\psi}) \|Q_1 - Q_2\|_{\text{op}}$.
- Therefore, $\mathscr{R}_m(\mathscr{V}) \leqslant \frac{24 B_Q B_\psi (B_\psi + 2B_{D\psi}) L_h}{\sqrt{m}} \int_0^\infty \sqrt{\log N(\varepsilon; \mathbb{B}_2^{p \times p}(0,1), \|\cdot\|_{\text{op}})}\, d\varepsilon$.

Rademacher complexity bounds

- The mismatch in norms in the covering number $\log N(\varepsilon; \mathbb{B}_2^{p\times p}(0,1), \|\cdot\|_{\text{op}})$ allows for some wins.
- First, the trivial bound. Because $\|Q\|_{\text{op}} \leqslant \|Q\|_F$, we have:

  $$\log N(\varepsilon; \mathbb{B}_2^{p\times p}(0,1), \|\cdot\|_{\text{op}}) \leqslant \log N(\varepsilon; \mathbb{B}_2^{p\times p}(0,1), \|\cdot\|_F) \leqslant p^2 \log(1 + 2/\varepsilon).$$

- Next, by the dual Sudakov inequality [Vershynin 2010], we can improve the dependence on p at the expense of a worse dependence on ε:

  $$\log N(\varepsilon; \mathbb{B}_2^{p\times p}(0,1), \|\cdot\|_{\text{op}}) \leqslant cp/\varepsilon^2.$$

- Combining both inequalities:

  $$\log N(\varepsilon; \mathbb{B}_2^{p\times p}(0,1), \|\cdot\|_{\text{op}}) \leqslant \min\{p^2 \log(1 + 2/\varepsilon), cp/\varepsilon^2\}.$$

- Now for any $\tau \in (0,1)$:

$$\int_0^1 \sqrt{\log N(\varepsilon; \mathbb{B}_2^{p\times p}(0,1), \|\cdot\|_{\mathrm{op}})}\, d\varepsilon$$

$$\leqslant p\int_0^\tau \sqrt{\log(1+2/\varepsilon)}\, d\varepsilon + \sqrt{cp}\int_\tau^1 1/\varepsilon\, d\varepsilon \qquad [\log N(\varepsilon; \mathbb{B}_2^{p\times p}(0,1), \|\cdot\|_{\mathrm{op}}) \leqslant \min\{p^2\log(1+2/\varepsilon), cp/\varepsilon^2\}]$$

$$\leqslant p\int_0^\tau \sqrt{\log(3/\varepsilon)}\, d\varepsilon + \sqrt{cp}\log(1/\tau)\,.$$

- From Mathematica:

$$\int_0^\tau \sqrt{\log(3/\varepsilon)}\, d\varepsilon = \tau\sqrt{\log(3/\tau)} + \frac{3\sqrt{\pi}}{2}\mathrm{erfc}\left(\sqrt{\log(3/\tau)}\right)$$

$$\leqslant \tau\sqrt{\log(3/\tau)} + \frac{3\sqrt{\pi}}{2}\exp(-\log(3/\tau)) \qquad [\mathrm{erfc}(x) \leqslant \exp(-x^2) \text{ for } x > 0]$$

$$= \tau\sqrt{\log(3/\tau)} + \tau\frac{\sqrt{\pi}}{2}\,.$$

- Therefore, setting $\tau = 1/\sqrt{p}$, we have $\int_0^1 \sqrt{\log N(\varepsilon; \mathbb{B}_2^{p\times p}(0,1), \|\cdot\|_{\mathrm{op}})}\, d\varepsilon \leqslant O(1)\sqrt{p}\log p$.

Rademacher complexity bounds

- Therefore, we have the bound:

$$\mathscr{R}_m(\mathscr{V}) \leqslant O(1) B_Q B_\psi (B_\psi + B_{D\psi}) L_h \sqrt{\frac{p \log^2 p}{m}}. \quad \blacksquare$$

Generalization error bounds

- With bounds on $\mathscr{R}_m(\mathscr{V})$, we can obtain bounds on $\mathrm{err}(\hat{V}_m)$.
- When $\mathscr{V} = \{V_\theta(x) = g(x,\theta) : \theta \in \mathbb{R}^k,\ \|\theta\|_2 \leqslant B_\theta\}$, we have with probability at least $1 - \exp(-k)$,

  $$\mathrm{err}(\hat{V}_m) \leqslant O(1) B_\theta^2 L_h^2 (L_g^2 + L_{\nabla g}^2) \frac{k \log^3 m}{m}.$$

- When $\mathscr{V} = \{x \mapsto \psi(x)^T Q \psi(x) : Q \in \mathbb{R}^{p \times p},\ Q = Q^T \succcurlyeq 0,\ \|Q\|_F \leqslant B_Q\}$, we have with probability at least $1 - \exp(-p)$,

  $$\mathrm{err}(\hat{V}_m) \leqslant O(1) B_Q^2 B_\psi^2 (B_\psi^2 + B_{D\psi}^2) L_h^2 \frac{p \log^2 p \log^3 m}{m}.$$

Random convex programs (RCP)

- In certain special cases, we can actually obtain sharper bounds on $\mathsf{err}(\hat{V}_m)$ using results on random convex programs [Calafiore 2010].
- Suppose again that $\mathscr{V} = \{V_\theta(x) = g(x,\theta) : \theta \in \mathbb{R}^k\}$. Suppose furthermore that $\theta \mapsto h(\xi, V_\theta)$ is convex for every $\xi \in X$.
- Then the feasibility problem find_θ s.t. $h(\xi_i, V_\theta) \leqslant 0,\ i = 1,...,m$ is a random convex program.
- **Theorem** [Calafiore 2010]: For any $\varepsilon \in (0,1)$, define $\beta(\varepsilon) := \sum_{i=0}^{k-1} \binom{m}{i} \varepsilon^i (1-\varepsilon)^{m-i}$.

  With probability at least $1 - \beta(\varepsilon)$ (over $\xi_1, \ldots, \xi_m$), we have $\mathsf{err}(\hat{V}_m) \leqslant \varepsilon$.

Random convex programs (RCP)

- To compare this bound with ours, we need to invert $\beta(\varepsilon) = \delta$.
- While this can be done numerically, there is no closed form formula for this.
- We can get an upper bound using the Chernoff inequality, which yields that with probability at least $1 - \delta$:

$$\mathsf{err}(\hat{V}_m) \leqslant O(1)\frac{k - 1 + \log(1/\delta)}{m}.$$

- Order-wise, this matches our parametric bound. However, the RCP constants are **much sharper.**

Random convex programs (RCP)

- Let's see an example where $\theta \mapsto h(\xi, V_\theta)$ is convex.
- Suppose $h(x, \dot{x}, V(x), \nabla V(x)) = \langle \nabla V(x), \dot{x} \rangle + \rho V(x)$.
- Now suppose $V(x) = \psi(x)^T Q \psi(x)$. Then, $h(x, \dot{x}, V(x), \nabla V(x)) = \left\langle 2\frac{\partial \psi}{\partial x}(x)^T Q \psi(x), \dot{x} \right\rangle + \rho \psi(x)^T Q \psi(x)$, which is convex (linear) in the parameters Q.

An aside: Towards practical bounds

- Uniform convergence laws typically give the right scaling, but the **constants are not practical.**

- **Randomized convex programs** (cf. [Calafiore 2010], [Campi and Garatti 2008]) give sharp generalization bounds, but only apply to convex optimization problems.

- **PAC-Bayes bounds** [McAllester 1999] have been shown to give non-vacuous bounds (cf. [Dziugaite and Roy 2017], [Majumdar and Goldstein 2018]) for both deep learning and control problems, but require delicate/expensive joint training of empirical risk + bound.

An aside: Towards practical bounds

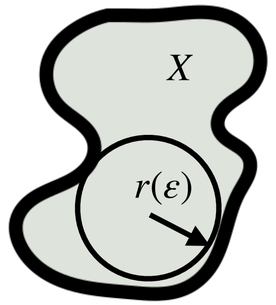
- A basic **holdout set bound** [Langford 2005] can be very powerful in practical.
- Procedure:
  - Split available data into train and holdout set.
  - Compute $\hat{V}$ on the training set.
  - Compute $\hat{p} = \frac{1}{m_h} \sum_{i=1}^{m_h} \mathbf{1}\{h(\xi_i, \hat{V}) > 0\}$ from holdout set.
  - Compute UCB on $\mathrm{err}(\hat{V})$: $\sup\{p \in (0,1) : \mathrm{KL}(\hat{p}, p) \leq m_h^{-1} \log(1/\delta)\}$.
  - (Set $m_h = 0.1m$ (say) for the UCB to tend to zero.)

Deterministic guarantees

- We have given bounds on $\mathbf{err}(\hat{V}_m)$, which is the probability that $\hat{V}_m$ will **fail to certify** a trajectory randomly initialized at $\xi \sim \mathscr{D}$.
- Can we convert this probabilistic bound into a **deterministic bound** regarding the “size” of the state space for which the certificate condition is violated?

Deterministic guarantees

- We first study the following technical question.
- Let μ_{Leb} denote the Lebesgue measure and $X \subseteq \mathbb{R}^n$ be a compact set with $\mu_{\text{Leb}}(X) > 0$.
- Let μ_X denote the uniform measure on X, i.e., $\mu_X(A) = \dfrac{\mu_{\text{Leb}}(A)}{\mu_{\text{Leb}}(X)}$ for all measurable $A \subseteq X$.
- **Question:** Fix an $\varepsilon \in (0,1)$. What is the largest radius $r(\varepsilon)$ such that there exists a $U \subseteq X$ with $\mu_X(U) \leqslant \varepsilon$ which contains a ℓ_2^n-ball of radius $r(\varepsilon)$?
- Mathematically, $r(\varepsilon) = \sup_{U \subseteq X : \mu_X(U) \leqslant \varepsilon} \sup\{r > 0 : \exists x \in U \text{ s.t. } \mathbb{B}_2^n(x, r) \subseteq U\}$.



$\mu(X) \leqslant \varepsilon$

- **Claim:** $r(\varepsilon) \leqslant \left(\dfrac{\varepsilon \mu_{\text{Leb}}(X)}{\mu_{\text{Leb}}(\mathbb{B}_2^n(0,1))} \right)^{1/n}$.
- **Proof:** for any x, r such that $\mathbb{B}_2^n(x, r) \subseteq U$, then by translation invariance:

  $\mu_{\text{Leb}}(U) \geqslant \mu_{\text{Leb}}(\mathbb{B}_2^n(x, r)) = \mu_{\text{Leb}}(\mathbb{B}_2^n(0,r)) = r^n \mu_{\text{Leb}}(\mathbb{B}_2^n(0,1))$.
- But since $\mu_X(U) \leqslant \varepsilon$, we have $\mu_{\text{Leb}}(U) \leqslant \varepsilon \mu_{\text{Leb}}(X)$.
- Therefore, $r^n \mu_{\text{Leb}}(\mathbb{B}_2^n(0,1)) \leqslant \varepsilon \mu_{\text{Leb}}(X)$, which implies the claim. ■

Deterministic guarantees

- Let us consider certifying exponential Lyapunov stability.
- Let $V \in C^1(\mathbb{R}^n, \mathbb{R})$ satisfy $V(x) \geqslant \nu \|x\|_2^2$. Define the "bad" set X_b as:

$$X_b := \left\{ \xi \in X : \max_{t \in T} \langle \nabla V(\varphi_t(\xi)), f(\varphi_t(\xi)) \rangle > -\lambda V(\varphi_t(\xi)) \right\}.$$

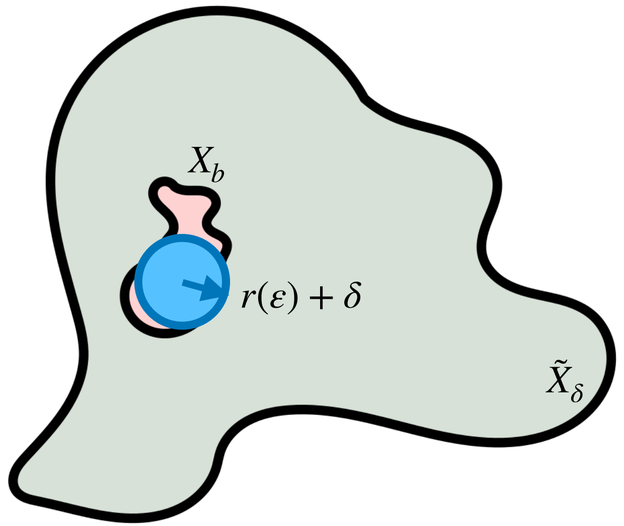
- Assume that $\mu_X(X_b) \leqslant \varepsilon$ (if $\mathscr{D}$ is uniform over X, then $\mu_X(X_b) = \mathsf{err}(\hat{V}_m)$).
- For which $x \in S$ can we assert that $\max_{t \in T} \langle \nabla V(x), f(x) \rangle \leqslant -\lambda V(x)$?
- **Assumption (incremental stability):** There exists a class $\mathscr{KL}$ function β such that for all $\xi_1, \xi_2 \in X, t \in T$: $\|\varphi_t(\xi_1) - \varphi_t(\xi_2)\|_2 \leqslant \beta(\|\xi_1 - \xi_2\|_2, t)$.

Deterministic guarantees

- For $\delta > 0$, define:
  - $\tilde{X}_\delta := \{\xi \in X : \mathbb{B}_2^n(\xi, r(\varepsilon) + \delta) \subseteq X\}$,
  - $\tilde{S}_\delta := \bigcup_{t \in T} \varphi_t(\tilde{X}_\delta)$ and $\tilde{S} := \bigcap_{\delta > 0} \tilde{S}_\delta$.
- Let L_V denote the Lipschitz constant of V over S.
- Define $q(x) := \langle \nabla V(x), f(x) \rangle$ and let L_q denote its Lipschitz constant over S.
- **Theorem:** For all $\eta \in (0,1)$:

$$\langle \nabla V(x), f(x) \rangle \leqslant -(1-\eta)\lambda V(x) \;\; \forall x \in \tilde{S} \setminus \mathbb{B}_2^n \left(0, \sqrt{\frac{(L_q + \lambda L_V)\beta(r(\varepsilon), 0)}{\eta \lambda \nu}}\right).$$

- **Proof:** Let $x \in \tilde{S}_\delta$. By definition of $\tilde{S}_\delta$, there exists a $\xi \in \tilde{X}_\delta, t \in T$ such that $\varphi_t(\xi) = x$.
- Define the “good” set as $X_g := X \backslash X_b$.
- **Claim:** there must exist a $\xi' \in X_g$ satisfying $\|\xi - \xi'\|_2 \leqslant r(\varepsilon) + \delta$.
- If $\xi \in X_g$, there is nothing to prove, so suppose that $\xi \in X_b$.
- We know that $\mathbb{B}_2^n(\xi, r(\varepsilon) + \delta) \nsubseteq X_b$ since $\mu_X(X_b) \leqslant \varepsilon$.
- Furthermore, since $\xi \in \tilde{X}_\delta$, we have $\mathbb{B}_2^n(\xi, r(\varepsilon) + \delta) \subseteq X$.
- Therefore, $\mathbb{B}_2^n(\xi, r(\varepsilon) + \delta) \cap X_g$ is non-empty, proving the claim.



- Therefore:
$$
\begin{aligned}
q(\varphi_t(\xi)) &\leqslant q(\varphi_t(\xi')) + L_q\|\varphi_t(\xi) - \varphi_t(\xi')\|_2 \\
&\leqslant -\lambda V(\varphi_t(\xi')) + L_q\|\varphi_t(\xi) - \varphi_t(\xi')\|_2 && [\xi' \in X_g] \\
&\leqslant -\lambda V(\varphi_t(\xi)) + (L_q + \lambda L_V)\|\varphi_t(\xi) - \varphi_t(\xi')\|_2 \\
&\leqslant -\lambda V(\varphi_t(\xi)) + (L_q + \lambda L_V)\beta(\|\xi - \xi'\|_2, t) && \text{[incremental stability]} \\
&\leqslant -\lambda V(\varphi_t(\xi)) + (L_q + \lambda L_V)\beta(r(\varepsilon) + \delta, 0)\,. && \text{[properties of class KL function]}
\end{aligned}
$$

- This shows for any $x \in \tilde{S}_\delta$, we have $q(x) \leqslant -\lambda V(x) + (L_q + \lambda L_V)\beta(r(\varepsilon) + \delta, 0)$.

- Taking $\delta \to 0$, by continuity of β in its first argument, for any $x \in \tilde{S}$:
$q(x) \leqslant -\lambda V(x) + (L_q + \lambda L_V)\beta(r(\varepsilon), 0)$.

- Therefore, for any $\eta \in (0,1)$, since $V(x) \geqslant \nu\|x\|_2^2$:
$q(x) \leqslant -(1-\eta)\lambda V(x) - \eta\lambda\nu\|x\|_2^2 + (L_q + \lambda L_V)\beta(r(\varepsilon), 0)$. ■

Deterministic guarantees

- We have shown that the Lyapunov condition holds for all $x \in \tilde{S}$ except in a radius around the origin of size $r_b \lesssim \sqrt{\beta(r(\varepsilon),0)}$. If $\beta(s,0) \lesssim s$, then $r_b \lesssim \varepsilon^{\frac{1}{2n}}$.
- We know that $\mathbf{err}(\hat{V}_m) \lesssim \frac{k}{m}$. Hence if we want $r_b \leqslant \zeta$, we need $m \gtrsim k\zeta^{-2n}$ trajectories.
- This exponential dependence on ζ in state dimension is (probably) unavoidable.

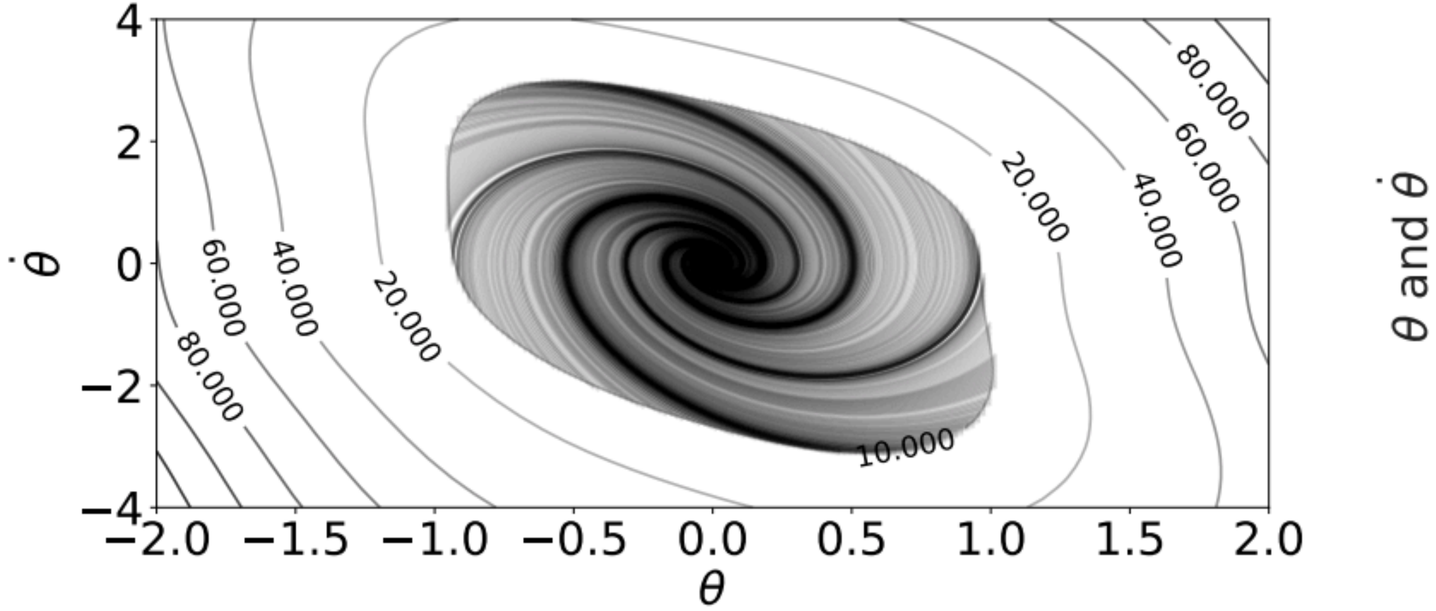
Damped pendulum

- We rollout $m = 1000$ trajectories for $T = 8$ seconds ($dt = 0.02$) initialized from $X = [-2,2] \times [-2,2]$.
- We fit a neural network Lyapunov function of the form $V_\theta(x) = x^T(L_\theta(x)L_\theta(x) + I)x$ where $L_\theta(x)$ is a reshaped fully connected NN.
- We use a soft loss

$$\ell(\theta) = \sum_{i=1}^{1000} \sum_{k=1}^{400} \text{ReLU}(\langle \nabla V_\theta(x_i(k)), \dot{x}_i(k) \rangle + \gamma V_\theta(x_i(k))) + \lambda \|\theta\|_2^2.$$

- Time derivatives $\dot{x}_i(k)$ are computed by finite differencing (savgol_filter in scipy).

Damped pendulum



Level sets of the learned Lyapunov function.

Damped pendulum

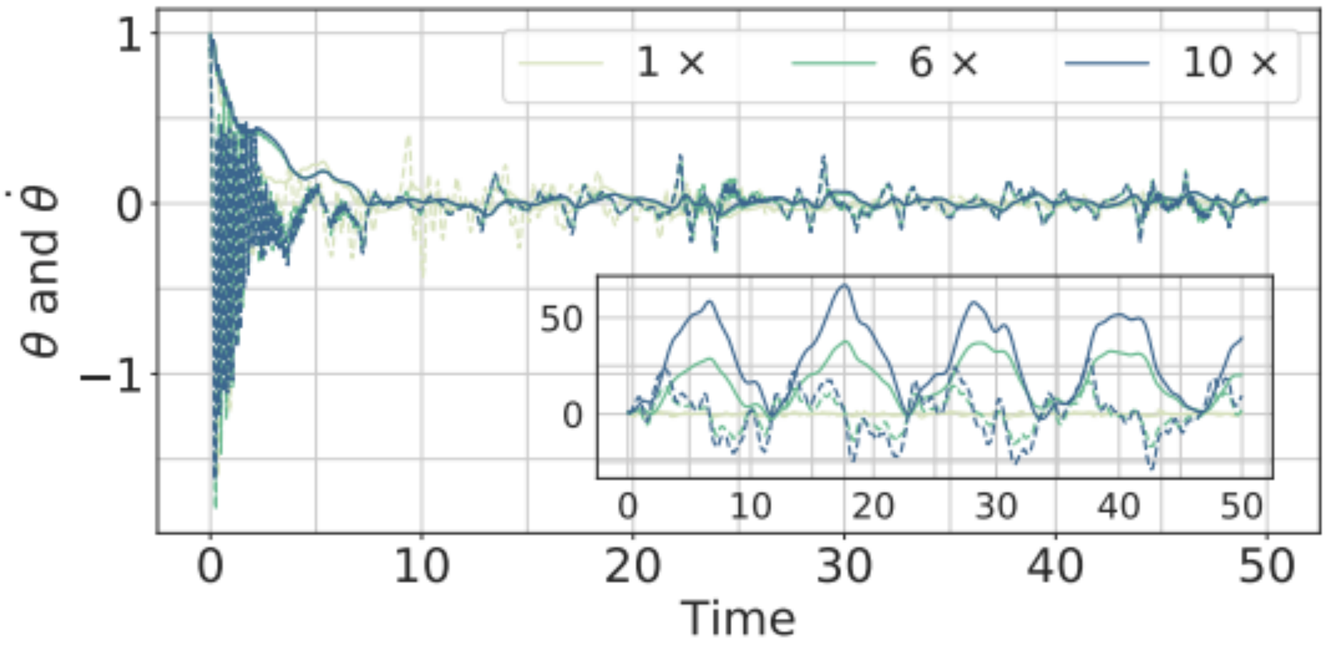
- To show that the learned Lyapunov function can perform useful downstream tasks, we add a disturbance to the pendulum dynamics (here a is unknown):
$$m\ell^2\ddot{\theta} + b\dot{\theta} + mg\ell \sin\theta + \langle a, \kappa\phi(t)\rangle = u.$$

- We use a “speed-gradient” (cf. [Fradkov et al.1999]) adaptive controller to reject the disturbance:

$$u(t) = \langle \hat{a}(t), \kappa\phi(t)\rangle ,$$

$$\dot{\hat{a}}(t) = -\gamma\phi(t)\langle \nabla_x V_\theta(x(t)), e_2\rangle .$$

Damped pendulum



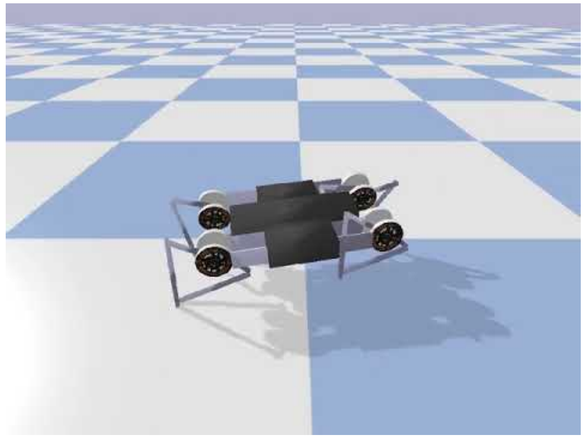
Different colors correspond to various values of κ.
Inset shows performance without adaptation.

Standing minitaur

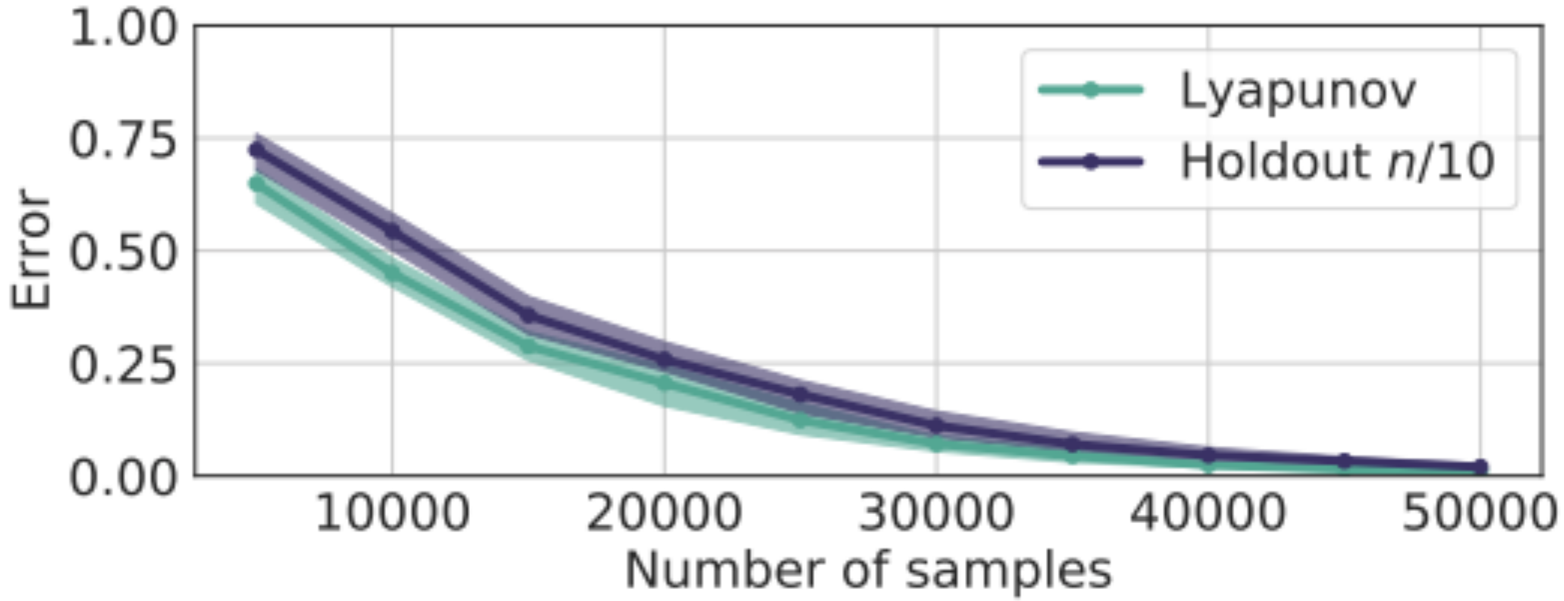
- We use the minitaur environment in PyBullet. The state dimension is 16 (excluding the base).

- A random impulse force (kick) is applied to the minitaur at time $t = 0$, and a simple PD controller is used to return the minitaur to a desired standing position.

- We learn a discrete-time Lyapunov function to satisfy $V_\theta(e_i(k+1)) \leq \rho V_\theta(e_i(k)) + \gamma$, where $e_i(k)$ is the error of the i-th trajectory at timestep k.



Standing minitaur



References

- G. C. Calafiore. Random convex programs. SIAM Journal on Optimization, 2010.
- M. C. Campi and S. Garatti. The exact feasibility of randomized solutions of uncertain convex programs. SIAM Journal on Optimization, 2008.
- G. K. Dziugaite and D. M. Roy. Computing Nonvacuous Generalization Bounds for Deep (Stochastic) Neural Networks with Many More Parameters than Training Data. UAI, 2016.
- A. L. Frakdov, I. V. Miroshnik, and V. O. Nikiforov. Nonlinear and Adaptive Control of Complex Systems, 1999.
- J. Langford. Tutorial on practical prediction theory for classification. Journal of Machine Learning Research, 2005.
- A. Majumdar and M. Goldstein. PAC-Bayes Control: Synthesizing Controllers that Provably Generalize to Novel Environments. Conference on Robot Learning, 2018.
- D. A. McAllester. Some PAC-Bayesian Theorems. Machine Learning, 1999.
- N. Srebro, K. Sridharan, and A. Tewari. Smoothness, low-noise, and fast rates. Neural Information Processing Systems, 2010.
- R. Vershynin. Lectures in geometric functional analysis, 2019.

On the Sample Complexity of Stability Constrained Imitation Learning

Stephen Tu, Alexander Robey, Tingnan Zhang, and Nikolai Matni.

arxiv.org/abs/2102.09161

Imitation learning

- In many tasks, there exists an expert policy which solves the task.
- However, relying on this expert may be undesirable:
    - Expert could be a human demonstrator.
    - Expert could require lots of computation (e.g. solution to non-convex optimization problem).
    - Expert could be a black-box policy.
- **Goal:** imitate (via supervised learning) a policy from demonstrations, so that the learned policy can solve the desired task.

Problem notation

- Discrete-time control-affine dynamical system:
  $x_{t+1} = f(x_t) + g(x_t)u_t,\ f(0) = 0,\ x_t \in \mathbb{R}^n,\ u_t \in \mathbb{R}^d.$
- $\mathscr{D}$ is a distribution over initial states $\xi \in X$ (can also generalize to include environment stochasticity).
- For policy $\pi : \mathbb{R}^n \to \mathbb{R}^d$, let $\varphi_t^\pi(\xi)$ denote the state x_t with $u_t = \pi(x_t)$.
- **Expert policy** $\pi_\star : \mathbb{R}^n \to \mathbb{R}^d$.
- (Weighted) **Policy deviation** $\Delta_{\pi_1,\pi_2}(x) := g(x)(\pi_1(x) - \pi_2(x))$.
- **Imitation loss** $\ell_{\pi'}(\xi; \pi_1, \pi_2) := \sum_{t=0}^{T-1} \|\Delta_{\pi_1,\pi_2}(\varphi_t^{\pi'}(\xi))\|_2$.
- The **generalization error** of a learned policy π is $\mathrm{err}(\pi) := \mathbb{E}_{\xi\sim\mathscr{D}}\ell_\pi(\xi; \pi, \pi_\star)$.

Behavior cloning

- Most basic imitation learning algorithm.
- Observe m **expert trajectories** $\{\{\varphi_t^{\pi_\star}(\xi_i)\}_{t=0}^{T}\}_{i=1}^{m}$, with $\xi_1, \ldots, \xi_m$ i.i.d. from $\mathscr{D}$.
- The behavior cloning policy is $\pi_{\mathrm{bc}} \in \arg\min_{\pi \in \Pi} \frac{1}{m} \sum_{i=1}^{m} \ell_{\pi_\star}(\xi_i; \pi, \pi_\star)$.
- **Note:** A bound on the generalization error $\mathsf{err}(\pi_{\mathrm{bc}})$ is not immediate from supervised learning theory, which yields bounds on the related (but not the same) quantity: $\mathbb{E}_{\xi \sim \mathscr{D}} \ell_{\pi_\star}(\xi; \pi_{\mathrm{bc}}, \pi_\star)$.
- This **distribution shift** $\mathbb{E}_{\xi \sim \mathscr{D}} \ell_{\pi_\star}(\xi; \pi_{\mathrm{bc}}, \pi_\star)$ vs $\mathbb{E}_{\xi \sim \mathscr{D}} \ell_{\pi_{\mathrm{bc}}}(\xi; \pi_{\mathrm{bc}}, \pi_\star)$ is what makes imitation learning trickier to analyze.

Distribution shift

- The key technical question in analyzing imitation learning is the following.
- Given two policies π_1, π_2, define $\mathrm{disc}_T(\xi; \pi_1, \pi_2) := \sum_{t=0}^{T} \|\varphi_t^{\pi_1}(\xi) - \varphi_t^{\pi_2}(\xi)\|_2$.
- Can we upper bound $\mathrm{disc}_T(\xi; \pi_1, \pi_2)$ by some function of the imitation loss $\ell_{\pi_1}(\xi; \pi_1, \pi_2)$ under π_1?
- **Claim** (Gronwall bound): Suppose f, g, π_1, π_2 are B-bounded and L-Lipschitz. Then: $\mathrm{disc}_T(\xi; \pi_1, \pi_2) \leqslant \frac{(L(1+2B))^T - 1}{L(1+2B) - 1} \ell_{\pi_1}(\xi; \pi_1, \pi_2)$.

- **Proof:** let us consider two trajectories:
$$x_{t+1} = f(x_t) + g(x_t)\pi_1(x_t),\ x_0 = \xi,$$
$$y_{t+1} = f(y_t) + g(y_t)\pi_2(y_t),\ y_0 = \xi\,.$$
- We have:
$$\begin{aligned} x_{t+1} - y_{t+1} &= f(x_t) - f(y_t) + g(x_t)\pi_1(x_t) - g(y_t)\pi_2(y_t) \\ &= f(x_t) - f(y_t) + g(x_t)\pi_1(x_t) - g(x_t)\pi_2(x_t) + g(x_t)\pi_2(x_t) - g(y_t)\pi_2(y_t)\,. \end{aligned}$$
- Because f, g, π_1, π_2 are B-bounded and L-Lipschitz, by triangle inequality:
$$\begin{aligned} \|x_{t+1} - y_{t+1}\|_2 &\leqslant L(1+2B)\|x_t - y_t\|_2 + \|g(x_t)(\pi_1(x_t) - \pi_2(x_t))\|_2 \\ &= L(1+2B)\|x_t - y_t\|_2 + \|\Delta_{\pi_1,\pi_2}(x_t)\|_2. \end{aligned}$$

- Previously we showed:
$$\|x_{t+1} - y_{t+1}\|_2 \leqslant L(1 + 2B)\|x_t - y_t\|_2 + \|\Delta_{\pi_1,\pi_2}(x_t)\|_2. .$$

- Unrolling this recursion, we obtain:
$$\|x_t - y_t\|_2 \leqslant \sum_{i=0}^{t-1} (L(1 + 2B))^{t-1-i}\|\Delta_{\pi_1,\pi_2}(x_i)\|_2.$$

- Therefore:
$$\sum_{t=0}^{T} \|x_t - y_t\|_2 \leqslant \sum_{t=0}^{T-1} \left(\frac{(L(1 + 2B))^{T-t} - 1}{L(1 + 2B) - 1} \right) \|\Delta_{\pi_1,\pi_2}(x_t)\|_2$$
$$\leqslant \frac{(L(1 + 2B))^{T} - 1}{L(1 + 2B) - 1} \sum_{t=0}^{T-1} \|\Delta_{\pi_1,\pi_2}(x_t)\|_2.$$ ■

Incremental ISS

- Gronwall bound is **exponential** in horizon length T!
- To improve the dependence on the horizon length, we need to use some stability theory.
- We start with the following definition of incremental input-to-state stability [Tran et al. 2016].
- **Notation:** for dynamics $x_{t+1} = f(x_t, u_t)$, let $\varphi_t(\xi; \{u_t\}_{t\geq 0})$ denote the state x_t initialized at $x_0 = \xi$ with input signal $\{u_t\}_{t\geqslant 0}$.
- **Definition:** the dynamics $x_{t+1} = f(x_t, u_t)$ is δISS if there exists class $\mathscr{KL}$ function ζ and class $\mathscr{K}_\infty$ function γ such that for every $\xi_1, \xi_2 \in X$, $\{u_t\}_{t\geqslant 0} \subseteq U$, and $t \in \mathbb{N}_{\geqslant 0}$:

$$\|\varphi_t(\xi_1, \{u_t\}_{t\geqslant 0}) - \varphi_t(\xi_2, \{0\}_{t\geqslant 0})\|_X \leqslant \zeta(\|\xi_1 - \xi_2\|_X, t) + \gamma\left(\max_{0\leqslant k\leqslant t-1} \|u_k\|_U\right).$$

Incremental ISS

- Now suppose that $\tilde{f}(x, u) = f(x) + g(x)\pi_2(x) + u$ is δISS.
- Write $x_{t+1} = f(x_t) + g(x_t)\pi_1(x_t) = f(x_t) + g(x_t)\pi_2(x_t) + g(x_t)(\pi_1(x_t) - \pi_2(x_t))$.
- We can now treat $\Delta_{\pi_1,\pi_2}(x_t) = g(x_t)(\pi_1(x_t) - \pi_2(x_t))$ as an input signal to $\tilde{f}(x, u)$:
  - $\varphi_t^{\pi_1}(\xi) = \varphi_t(\xi, \{\Delta_{\pi_1,\pi_2}(\varphi_t^{\pi_1}(\xi))\}_{t\geqslant 0})$ and $\varphi_t^{\pi_2}(\xi) = \varphi_t(\xi, \{0\}_{t\geqslant 0})$.
- Therefore, by δISS:

$$\|\varphi_t^{\pi_1}(\xi) - \varphi_t^{\pi_2}(\xi)\|_2 \leqslant \gamma\left(\max_{0\leqslant k\leqslant t-1} \|\Delta_{\pi_1,\pi_2}(\varphi_k^{\pi_1}(\xi))\|_2\right) \leqslant \gamma\left(\sum_{k=0}^{t-1} \|\Delta_{\pi_1,\pi_2}(\varphi_k^{\pi_1}(\xi))\|_2\right).$$

- Hence:

$$\mathrm{disc}_T(\xi; \pi_1, \pi_2) = \sum_{t=0}^{T} \|\varphi_t^{\pi_1}(\xi) - \varphi_t^{\pi_2}(\xi)\|_2 \leqslant \sum_{t=0}^{T-1} \gamma\left(\sum_{k=0}^{t-1} \|\Delta_{\pi_1,\pi_2}(\varphi_k^{\pi_1}(\xi))\|_2\right) \leqslant T\gamma\left(\ell_{\pi_1}(\xi; \pi_1, \pi_2)\right).$$

Incremental ISS

- The bound $\mathsf{disc}_T(\xi;\pi_1,\pi_2) \leqslant T\gamma\left(\ell_{\pi_1}(\xi;\pi_1,\pi_2)\right)$ improves the dependence on T when compared to the Gronwall bound.
- However, this bound is not sharp: consider the Gronwall bound when $L(1+2B)<1$, in which case $\mathsf{disc}_T(\xi;\pi_1,\pi_2) \leqslant O(1)\ell_{\pi_1}(\xi;\pi_1,\pi_2)$ is independent of T.
- This motivates a more **quantitative** definition of δISS.

Incremental gain stability (IGS)

- **Definition:** Let $\Psi = (a, a_0, a_1, b_0, b_1, \zeta, \gamma)$ with $a, a_0, a_1, b_0, b_1 \in [1, \infty)$, $a_0 \leqslant a_1$, $b_0 \leqslant b_1$, and ζ, γ positive. The dynamics $x_{t+1} = f(x_t, u_t)$ is Ψ-IGS if for every $\xi_1, \xi_2 \in X$, $\{u_t\}_{t \geqslant 0} \subseteq U$, $T \in \mathbb{N}_{\geqslant 0}$, letting $\Delta_t := \varphi_t(\xi_1, \{u_t\}_{t \geqslant 0}) - \varphi_t(\xi_2, \{0\}_{t \geqslant 0})$:

$$\sum_{t=0}^{T} \min\{\|\Delta_t\|_X^{a \wedge a_0}, \|\Delta_t\|_X^{a \vee a_1}\} \leqslant \zeta \|\xi_1 - \xi_2\|_X^a + \gamma \sum_{t=0}^{T-1} \max\{\|u_t\|_U^{b_0}, \|u_t\|_U^{b_1}\}.$$

- **Claim:** If $\tilde{f}(x, u) = f(x) + g(x)\pi_2(x) + u$ is Ψ-IGS, then:

$$\mathrm{disc}_T(\xi; \pi_1, \pi_2) \leqslant 4(\gamma \vee 1)^{\frac{1}{a \wedge a_0}} T^{1 - \frac{1}{a \vee a_1}} \max\left\{ \ell_{\pi_1}(\xi; \pi_1, \pi_2)^{\frac{b_0}{a \vee a_1}}, \ell_{\pi_1}(\xi; \pi_1, \pi_2)^{\frac{b_1}{a \vee a_0}} \right\}.$$

- **Proof** (of simple case when $a = a_0$ and $b_0 = b_1$).
- We have $\varphi_t^{\pi_1}(\xi) = \varphi_t(\xi, \{\Delta_{\pi_1,\pi_2}(\varphi_t^{\pi_1}(\xi))\}_{t\geqslant 0})$ and $\varphi_t^{\pi_2}(\xi) = \varphi_t(\xi, \{0\}_{t\geqslant 0})$.
- Since $f(x) + g(x)\pi_2(x) + u$ is Ψ-IGS:
$$\sum_{t=0}^{T} \min\{\|\Delta_t\|_X^{a_0}, \|\Delta_t\|_X^{a_1}\} \leqslant \gamma \sum_{t=0}^{T-1} \|u_t\|_U^{b_0}.$$

- Let $I := \{t \in \{0,...,T\} : \|\Delta_t\|_2^{a_0} \leqslant \|\Delta_t\|_2^{a_1}\}$. Then,
  $\mathrm{disc}_T(\xi;\pi_1,\pi_2) = \sum_{t\in I} \|\Delta_t\|_2 + \sum_{t\in I^c} \|\Delta_t\|_2$

$$\leqslant |I|^{1-1/a_0}\left(\sum_{t\in I}\|\Delta_t\|_2^{a_0}\right)^{1/a_0} + |I^c|^{1-1/a_1}\left(\sum_{t\in I}\|\Delta_t\|_2^{a_1}\right)^{1/a_1} \qquad \text{[Holder's inequality]}$$

$$\leqslant T^{1-1/a_0}\left(\sum_{t\in I}\min\{\|\Delta_t\|_2^{a_0},\|\Delta_t\|_2^{a_1}\}\right)^{1/a_0} + T^{1-1/a_1}\left(\sum_{t\in I^c}\min\{\|\Delta_t\|_2^{a_0},\|\Delta_t\|_2^{a_1}\}\right)^{1/a_1} \qquad [|I|,|I^c| \leqslant T]$$

$$\leqslant T^{1-1/a_0}\left(\sum_{t=0}^{T}\min\{\|\Delta_t\|_2^{a_0},\|\Delta_t\|_2^{a_1}\}\right)^{1/a_0} + T^{1-1/a_1}\left(\sum_{t=0}^{T}\min\{\|\Delta_t\|_2^{a_0},\|\Delta_t\|_2^{a_1}\}\right)^{1/a_1} \qquad [I \subseteq \{0,...,T\}]$$

$$\leqslant T^{1-1/a_0}\left(\gamma\sum_{t=0}^{T-1}\|u_t\|_2^{b_0}\right)^{1/a_0} + T^{1-1/a_1}\left(\gamma\sum_{t=0}^{T-1}\|u_t\|_2^{b_0}\right)^{1/a_1} \qquad [\Psi\text{-IGS with } u_t = \Delta_{\pi_1,\pi_2}(\varphi_t^{\pi_1}(\xi))]$$

$$\leqslant T^{1-1/a_0}\gamma^{1/a_0}\left(\sum_{t=0}^{T-1}\|u_t\|_2\right)^{b_0/a_0} + T^{1-1/a_1}\gamma^{1/a_1}\left(\sum_{t=0}^{T-1}\|u_t\|_2\right)^{b_1/a_1} \qquad \left[\sum_t \|u_t\|_2^{b_0} \leqslant \left(\sum_t \|u_t\|_2\right)^{b_0}\right]$$

$$\leqslant 2(\gamma\vee 1)^{\frac{1}{a_0}} T^{1-\frac{1}{a_1}} \max\left\{\ell_{\pi_1}(\xi;\pi_1,\pi_2)^{\frac{b_0}{a_0}}, \ell_{\pi_1}(\xi;\pi_1,\pi_2)^{\frac{b_1}{a_1}}\right\}. \qquad [a+b \leqslant 2\max\{a,b\}, 1 \leqslant a_0 \leqslant a_1]$$

■

Incremental gain stability

- **Claim (IGS implies bounded state):** if f is Ψ-IGS, then for all $\xi_1, \xi_2 \in X$, defining $\Delta_t := \varphi_t(\xi_1, \{0\}_{t\geqslant 0}) - \varphi_t(\xi_1, \{0\}_{t\geqslant 0})$, we have:

$$\sum_{t=0}^{T-1} \|\Delta_t\|_X \leqslant 2(\zeta \vee 1)^{\frac{1}{a \wedge a_0}} T^{1-\frac{1}{a \vee a_1}} \max\left\{ \|\xi_1 - \xi_2\|_X^{\frac{a}{a \wedge a_0}}, \|\xi_1 - \xi_2\|_X^{\frac{a}{a \vee a_1}} \right\}.$$

- Similar proof to bound on $\mathsf{disc}_T(\xi; \pi_1, \pi_2)$.

Incremental gain stability

- **Claim (Lyapunov characterization of IGS):** Suppose there exists $V : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_{\geqslant 0}$ satisfying:
  - $\underline{\alpha}\|x-y\|_X^a \leqslant V(x,y) \leqslant \overline{\alpha}\|x-y\|_X^a,$
  - $V(f(x,u),f(y,0)) - V(x,y) \leqslant -\mathfrak{a}\min\{\|x-y\|_X^{a_0},\|x-y\|_X^{a_1}\} + \mathfrak{b}\max\{\|u\|_U^{b_0},\|u\|_U^{b_1}\}.$
- Then f is Ψ-IGS with $\Psi = \left(a, a_0, a_1, b_0, b_1, \dfrac{\overline{\alpha}}{\underline{\alpha}\wedge\mathfrak{a}}, \dfrac{\mathfrak{b}}{\underline{a}\wedge\mathfrak{a}}\right).$

- **Proof:** Define two dynamics:
$$x_{t+1} = f(x_t, u_t),\ x_0 = \xi_1,$$
$$y_{t+1} = f(y_t, 0),\ y_0 = \xi_2\,.$$

- With $V_t := V(x_t, y_t)$, we have:
$$\begin{aligned} V_{t+1} &= V(x_{t+1}, y_{t+1}) = V(f(x_t, u_t), f(y_t, 0)) \\ &\leqslant V(x_t, y_t) - \mathfrak{a} \min\{\|x_t - y_t\|_X^{a_0}, \|x_t - y_t\|_X^{a_1}\} + \mathfrak{b} \max\{\|u_t\|_U^{b_0}, \|u_t\|_U^{b_1}\} \\ &= V_t - \mathfrak{a} \min\{\|x_t - y_t\|_X^{a_0}, \|x_t - y_t\|_X^{a_1}\} + \mathfrak{b} \max\{\|u_t\|_U^{b_0}, \|u_t\|_U^{b_1}\}\,. \end{aligned}$$

- Unrolling, we obtain:
$$\begin{aligned} V_T + \mathfrak{a} \sum_{t=0}^{T-1} \min\{\|x_t - y_t\|_X^{a_0}, \|x_t - y_t\|_X^{a_1}\} &\leqslant V_0 + \mathfrak{b} \sum_{t=0}^{T-1} \max\{\|u_t\|_U^{b_0}, \|u_t\|_U^{b_1}\} \\ &\leqslant \overline{\alpha} \|\xi_1 - \xi_2\|_X^a + \mathfrak{b} \sum_{t=0}^{T-1} \max\{\|u_t\|_U^{b_0}, \|u_t\|_U^{b_1}\}\,. \end{aligned}$$

- Previously, we showed:

$$V_T + \mathfrak{a} \sum_{t=0}^{T-1} \min\{ \|x_t - y_t\|_X^{a_0}, \|x_t - y_t\|_X^{a_1} \} \leqslant \overline{\alpha} \|\xi_1 - \xi_2\|_X^a + \mathfrak{b} \sum_{t=0}^{T-1} \max\{ \|u_t\|_U^{b_0}, \|u_t\|_U^{b_1} \} .$$

- We can lower bound the LHS by:

$$\begin{aligned} V_T + \mathfrak{a} \sum_{t=0}^{T-1} \min\{ \|x_t - y_t\|_X^{a_0}, \|x_t - y_t\|_X^{a_1} \} &\geqslant \underline{\alpha} \|x_T - y_T\|_X^a + \mathfrak{a} \sum_{t=0}^{T-1} \min\{ \|x_t - y_t\|_X^{a_0}, \|x_t - y_t\|_X^{a_1} \} \\ &\geqslant (\underline{\alpha} \wedge \mathfrak{a}) \sum_{t=0}^{T} \min\{ \|x_t - y_t\|_X^{a_0}, \|x_t - y_t\|_X^{a}, \|x_t - y_t\|_X^{a_1} \} \\ &= (\underline{\alpha} \wedge \mathfrak{a}) \sum_{t=0}^{T} \min\{ \|x_t - y_t\|_X^{a_0 \wedge a}, \|x_t - y_t\|_X^{a_1 \vee a} \} . \end{aligned}$$

- Therefore:

$$\sum_{t=0}^{T} \min\{ \|x_t - y_t\|_X^{a_0 \wedge a}, \|x_t - y_t\|_X^{a_1 \vee a} \} \leqslant \frac{\overline{\alpha}}{\underline{\alpha} \wedge a} \|\xi_1 - \xi_2\|_X^a + \frac{\mathfrak{b}}{\underline{\alpha} \wedge a} \sum_{t=0}^{T-1} \max\{ \|u_t\|_U^{b_0}, \|u_t\|_U^{b_1} \}.$$ ■

Contraction implies exponential IGS

- Suppose there exists $\mu I \preccurlyeq M \preccurlyeq LI,$ and $\rho \in (0,1)$ such that:
  - $\|f(x,0) - f(y,0)\|_M \leqslant \rho \|x - y\|_M \;\; \forall x, y \in \mathbb{R}^n.$
  - $\|f(x,u) - f(x,0)\|_M \leqslant L_u \|u\|_2 \;\; \forall x \in \mathbb{R}^n, u \in \mathbb{R}^d.$
- Note that the first condition is equivalent to the contraction condition $\frac{\partial f}{\partial x}(x,0)^T M \frac{\partial f}{\partial x}(x,0) \preccurlyeq \rho^2 M \;\; \forall x \in \mathbb{R}^n.$
- **Claim:** f is Ψ-IGS with $\Psi = \left(1,1,1,1,1, \frac{L}{(1-\rho)\mu}, \frac{L_u}{(1-\rho)\mu}\right)$, or equivalently:

$$\sum_{t=0}^{T} \|\Delta_t\|_2 \leqslant \frac{L}{(1-\rho)\mu} \|\xi_1 - \xi_2\|_2 + \frac{L_u}{(1-\rho)\mu} \sum_{t=0}^{T-1} \|u_t\|_2.$$

- **Proof:** Letting $V(x,y) = \|x-y\|_M$,

$$\begin{aligned} V(f(x,u),f(y,0)) &= \|f(x,u)-f(y,0)\|_M \\ &\leqslant \|f(x,u)-f(x,0)\|_M + \|f(x,0)-f(y,0)\|_M \\ &\leqslant \rho\|x-y\|_M + L_u\|u\|_2. \end{aligned}$$

- Therefore the IGS Lyapunov condition is verified:

$$V(f(x,u),f(y,0)) - V(x,y) \leqslant -(1-\rho)\mu\|x-y\|_2 + L_u\|u\|_2. \quad \blacksquare$$

Contraction implies exponential IGS

- The same conclusion applies if the metric $M(x)$ is allowed to depend on the state (although the proof is a bit more technical):
  $\frac{\partial f}{\partial x}(x,0)^T M(x) \frac{\partial f}{\partial x}(x,0) \preccurlyeq \rho^2 M(x) \;\; \forall x \in \mathbb{R}^n.$
- Examples of contracting systems:
  - $f(x,u) = \sum_{i=1}^{K} A_i \mathbf{1}\{x \in \mathscr{C}_i\} x + B(x)u$ for $\{A_i\}$ with common quadratic Lyapunov function P and $B(x)$ bounded (metric is $M(x) = P$).
  - $f(x,u) = \log(1+x^2) + u$ with metric $M(x) = 2[1+\exp(-|x|)]^{-1}$,
  - $f(x,u) = x - \eta[\nabla V(x) + u]$ with $V \in C^2(\mathbb{R}^n, \mathbb{R})$ satisfying $\mu I \preccurlyeq \nabla^2 V(x) \preccurlyeq LI$ and $\eta \in (0, 1/L]$.

Tunable Ψ-IGS system

- **Claim:** consider the scalar dynamics $f(x_t, u_t) = x_t - \eta x_t \frac{|x_t|^p}{1+|x_t|^p} + \eta u_t$ for $p \in (0,\infty)$ and $0 < \eta < \frac{4}{5+p}$. Then f is Ψ-IGS with $\Psi = \left(1,1,1+p,1,1,\frac{2^{2+p}}{\eta},2^{2+p}\right)$, i.e.,
$$\sum_{t=0}^{T} \min\{|\Delta_t|, |\Delta_t|^{1+p}\} \leqslant \frac{2^{2+p}}{\eta}|\xi_1 - \xi_2| + 2^{2+p}\sum_{t=0}^{T-1}|u_t|.$$
- **Proof idea:** Show $V(x,y) = |x-y|$ is an IGS Lyapunov function.

Behavior cloning

- Let us now use IGS to analyze **stability constrained** behavior cloning.
- Recall we observe m **expert trajectories** $\{\{\varphi_t^{\pi_\star}(\xi_i)\}_{t=0}^{T}\}_{i=1}^{m}$, with $\xi_1, \ldots, \xi_m$ i.i.d. from $\mathscr{D}$.
- The standard behavior cloning policy is $\hat{\pi}_{\mathrm{bc}} \in \arg\min_{\pi \in \Pi} \frac{1}{m} \sum_{i=1}^{m} \ell_{\pi_\star}(\xi_i; \pi, \pi_\star)$.
- **Stability constrained** BC policy is $\hat{\pi}_{\mathrm{bc}} \in \arg\min_{\pi \in \Pi_\Psi} \frac{1}{m} \sum_{i=1}^{m} \ell_{\pi_\star}(\xi_i; \pi, \pi_\star)$, where Π_Ψ is the set of policies such that $\tilde{f}(x,u) = f(x) + g(x)\pi(x) + u$ is Ψ-IGS.

Behavior cloning

- **Main assumptions for analysis:**
  - **Policy preserves fixed point:** $\pi(0) = 0$ for $\pi \in \Pi$.
  - **Realizability+stability:** $\pi_\star \in \Pi_\Psi$ with $\Psi = (a, a_0, a_1, b_0, b_1, \zeta, \gamma)$ satisfying $a = a_0$, $b_0 = b_1$, $a_1 \leqslant b_0$, $\zeta \geqslant 1$, $\gamma \geqslant 1$ (simplifications made for clarity, not actually necessary).
  - **Lipschitz+bounded:** Δ_{π_1,π_2} is L_Δ-Lipschitz for all $\pi_1, \pi_2 \in \Pi$ and $\sup_{x \in \mathbb{R}^n} \|g(x)\|_{\mathrm{op}} \leqslant B_g$.

Behavior cloning

- **Basic inequality:**
  $\mathsf{err}(\hat{\pi}_{\mathrm{bc}}) = \mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\hat{\pi}_{\mathrm{bc}}}(\xi; \hat{\pi}_{\mathrm{bc}}, \pi_\star)$

$$\leqslant L_\Delta \mathbb{E}_{\xi\sim\mathscr{D}} \sum_{t=0}^{T-1} \|\varphi_t^{\hat{\pi}_{\mathrm{bc}}}(\xi) - \varphi_t^{\pi_\star}(\xi)\|_2 = L_\Delta \mathbb{E}_{\xi\sim\mathscr{D}} \mathsf{disc}_T(\xi; \hat{\pi}_{\mathrm{bc}}, \pi_\star)\,.$$

- Because $\hat{\pi}_{\mathrm{bc}} \in \Pi_\Psi$, by Ψ-IGS + Jensen's inequality, assuming $\mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_\star}(\xi; \hat{\pi}_{\mathrm{bc}}, \pi_\star) \leqslant 1$:

  $\mathbb{E}_{\xi\sim\mathscr{D}} \mathsf{disc}_T(\xi; \hat{\pi}_{\mathrm{bc}}, \pi_\star) \leqslant 8\gamma^{\frac{1}{a_0}} T^{1-\frac{1}{a_1}} \left(\mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_\star}(\xi; \hat{\pi}_{\mathrm{bc}}, \pi_\star)\right)^{\frac{b_0}{a_1}}\,.$

- We now use uniform convergence to obtain a bound on $\mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_\star}(\xi; \hat{\pi}_{\mathrm{bc}}, \pi_\star)$.

Uniform convergence of imitation loss

- In stability constrained BC, we minimized $\hat{\pi}_{\mathrm{bc}} \in \arg\min_{\pi \in \Pi_\Psi} \frac{1}{m} \sum_{i=1}^{m} \ell_{\pi_\star}(\xi_i; \pi, \pi_\star)$, but now we want a bound on $\mathbb{E}_{\xi \sim \mathscr{D}} \ell_{\pi_\star}(\xi; \hat{\pi}_{\mathrm{bc}}, \pi_\star)$.
- Because $\hat{\pi}_{\mathrm{bc}}$ is a function of $\xi_1, \ldots, \xi_m$, $\mathbb{E}_{\{\xi_i\}} \frac{1}{m} \sum_{i=1}^{m} \ell_{\pi_\star}(\xi_i; \hat{\pi}_{\mathrm{bc}}, \pi_\star) \neq \mathbb{E}_{\xi \sim \mathscr{D}} \ell_{\pi_\star}(\xi; \hat{\pi}_{\mathrm{bc}}, \pi_\star)$, and hence a standard Hoeffding inequality is insufficient.
- We need a uniform law of the form:

$$\sup_{\pi_d \in \Pi_\Psi, \pi_g \in \Pi} \mathbb{P}_{\{\xi_i\}} \left(\sup_{\pi \in \Pi} \left| \mathbb{E}_{\xi \sim \mathscr{D}} \ell_{\pi_d}(\xi; \pi, \pi_g) - \frac{1}{m} \sum_{i=1}^{m} \ell_{\pi_d}(\xi_i; \pi, \pi_g) \right| \geqslant f(m, \delta, \Pi) \right) \leqslant \delta$$

Uniform convergence of imitation loss

- Define the following:
  - **[Uniform bound]** $B_\ell := \sup_{\pi_d \in \Pi_\Psi} \sup_{\pi_1, \pi_2 \in \Pi} \sup_{\xi \in X} \ell_{\pi_d}(\xi; \pi_1, \pi_2).$
  - **[Rademacher complexity]** $\mathscr{R}_m(\Pi) := \sup_{\pi_d \in \Pi_\Psi} \sup_{\pi_g \in \Pi} \mathbb{E}_{\{\xi_i\}} \mathbb{E}_{\{\varepsilon_i\}} \sup_{\pi \in \Pi} \frac{1}{m} \sum_{i=1}^{m} \varepsilon_i \ell_{\pi_d}(\xi_i; \pi, \pi_g).$
- From uniform convergence results (cf. lecture this morning), for any fixed $\pi_d \in \Pi_\Psi, \pi_g \in \Pi$, with probability at least $1 - \delta$:

$$\sup_{\pi \in \Pi} \left| \mathbb{E}_{\xi \sim \mathscr{D}} \ell_{\pi_d}(\xi; \pi, \pi_g) - \frac{1}{m} \sum_{i=1}^{m} \ell_{\pi_d}(\xi_i; \pi, \pi_g) \right| \leqslant 2\mathscr{R}_m(\Pi) + B_\ell \sqrt{\frac{\log(2/\delta)}{m}}.$$

Uniform convergence of imitation loss

- **Claim:** $B_\ell \leqslant 2\zeta^{\frac{1}{a_0}} B_0 L_\Delta T^{1-\frac{1}{a_1}}$, with $B_0 = \sup_{\xi \in X} \|\xi\|_2$.
- Proof is simple and uses IGS => boundedness of state.

Uniform convergence of imitation loss

- Rademacher complexity bound is more involved.
- Need to fix policy class Π to obtain concrete bounds.
- For simplicity, we consider a parametric policy class $\Pi = \{x \mapsto \pi(x, \theta) : \theta \in \mathbb{R}^q,\ \|\theta\|_2 \leqslant B_\theta\}$ with $\pi \in C^2(\mathbb{R}^n \times \mathbb{R}^q, \mathbb{R}^d)$.
- Define $L_{\partial^2\pi} := \sup_{\|x\|_2 \leqslant \zeta^{\frac{1}{a_0}} B_0,\ \|\theta\|_2 \leqslant B_\theta} \left\| \frac{\partial^2 \pi}{\partial \theta \partial x}(x, \theta) \right\|_{\ell_2^q \to M(\mathbb{R}^{d \times n})}$.
- **Claim:** $\mathscr{R}_m(\Pi) \leqslant 65 \zeta^{\frac{1}{a_0}} B_0 B_g B_\theta L_{\partial^2\pi} T^{1-\frac{1}{a_1}} \sqrt{\frac{q}{m}}$.

- **Proof:** Again, our main tool will be Dudley's inequality.
- By Taylor's theorem, not hard to check that:
  $\|\pi(x,\theta_1)-\pi(x,\theta_2)\|_2 \leqslant L_{\partial^2\pi}\|x\|_2\|\theta_1-\theta_2\|_2 \;\; \forall x \in \mathbb{B}_2^n(0,\zeta^{\frac{1}{a_0}}B_0),\;\; \theta_1,\theta_2 \in \mathbb{B}_2^q(0,B_\theta).$
- Therefore, for two policies $\pi_1, \pi_2 \in \Pi$:

$$
\begin{aligned}
|\ell_{\pi_d}(\xi;\pi_1,\pi_g)-\ell_{\pi_d}(\xi;\pi_2,\pi_g)| &\leqslant \sum_{t=0}^{T-1}\|\Delta_{\pi_1,\pi_g}(\varphi_t^{\pi_d}(\xi))-\Delta_{\pi_2,\pi_g}(\varphi_t^{\pi_d}(\xi))\|_2 && \text{[reverse triangle inequality]}\\
&= \sum_{t=0}^{T-1}\|\Delta_{\pi_1,\pi_2}(\varphi_t^{\pi_d}(\xi))\|_2 && [\Delta_{\pi_1,\pi_g}(x)-\Delta_{\pi_2,\pi_g}(x)=\Delta_{\pi_1,\pi_2}(x)]\\
&\leqslant B_g\sum_{t=0}^{T-1}\|\pi(\varphi_t^{\pi_d}(\xi),\theta_1)-\pi(\varphi_t^{\pi_d}(\xi),\theta_2)\|_2 && [\|g(x)\|_{\mathrm{op}}\leqslant B_g]\\
&\leqslant B_g L_{\partial^2\pi}\left(\sum_{t=0}^{T-1}\|\varphi_t^{\pi_d}(\xi)\|_2\right)\|\theta_1-\theta_2\|_2 && [\|\varphi_t^{\pi_d}(\xi)\|_2\leqslant\zeta^{\frac{1}{a_0}}B_0]\\
&\leqslant 2\zeta^{\frac{1}{a_0}}B_0B_gL_{\partial^2\pi}T^{1-\frac{1}{a_1}}\|\theta_1-\theta_2\|_2 && \text{[IGS bounded state]}\,.
\end{aligned}
$$

- Therefore by Dudley's inequality:

$$\mathscr{R}_m(\Pi) \leqslant 48\zeta^{\frac{1}{a_0}} B_0 B_g L_{\partial^2\pi} T^{1-\frac{1}{a_1}} \frac{1}{\sqrt{m}} \int_0^\infty \sqrt{\log N(\varepsilon; \mathbb{B}_2^q(0, B_\theta), \|\cdot\|_2)}\, d\varepsilon$$

$$\leqslant 48\zeta^{\frac{1}{a_0}} B_\theta B_0 B_g L_{\partial^2\pi} T^{1-\frac{1}{a_1}} \sqrt{\frac{q}{m}} \int_0^1 \sqrt{\log(1 + 2/\varepsilon)}\, d\varepsilon \qquad [N(\varepsilon; \mathbb{B}_2^q(0,1), \|\cdot\|_2) \leqslant (1 + 2/\varepsilon)^q]$$

$$\leqslant 65\zeta^{\frac{1}{a_0}} B_\theta B_0 B_g L_{\partial^2\pi} T^{1-\frac{1}{a_1}} \sqrt{\frac{q}{m}}\,. \qquad \blacksquare$$

Behavior cloning

- **Theorem:** With probability at least $1 - 2e^{-q}$,

  $$\mathsf{err}(\hat{\pi}_{\mathrm{bc}}) \leqslant O(1)\gamma^{\frac{1}{a_0}}\zeta^{\frac{b_0}{a_0 a_1}}B_0^{\frac{b_0}{a_1}}L_\Delta \max\{(B_\theta B_g L_{\partial^2\pi})^{\frac{b_0}{a_1}}, L_\Delta^{\frac{b_0}{a_1}}\}T^{(1-\frac{1}{a_1})(1+\frac{b_0}{a_1})}\left(\frac{q}{m}\right)^{\frac{b_0}{2a_1}}.$$

- **Proof:** Recall the basic inequality $\mathsf{err}(\hat{\pi}_{\mathrm{bc}}) \leqslant 4\gamma^{\frac{1}{a_0}}L_\Delta T^{1-\frac{1}{a_1}}\left(\mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_\star}(\xi;\hat{\pi}_{\mathrm{bc}},\pi_\star)\right)^{\frac{b_0}{a_1}}$.

- From uniform convergence, with probability at least $1-\delta$:

  $$\mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_\star}(\xi;\hat{\pi}_{\mathrm{bc}},\pi_\star) \leqslant \frac{1}{m}\sum_{i=1}^m \ell_{\pi_\star}(\xi;\hat{\pi}_{\mathrm{bc}},\pi_\star) + 2\mathscr{R}_m(\Pi) + B_\ell\sqrt{\frac{\log(2/\delta)}{m}}.$$

- Since $\pi_\star \in \Pi_\psi$, then $\pi_\star$ is feasible for the optimization defining $\hat{\pi}_{\mathrm{bc}}$, and therefore

  $$\frac{1}{m}\sum_{i=1}^m \ell_{\pi_\star}(\xi;\hat{\pi}_{\mathrm{bc}},\pi_\star) \leqslant \frac{1}{m}\sum_{i=1}^m \ell_{\pi_\star}(\xi;\pi_\star,\pi_\star) = 0.$$

- Therefore, with probability at least $1 - 2e^{-q}$:

$$\mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_\star}(\xi;\hat{\pi}_{\mathrm{bc}},\pi_\star) \leqslant 2\mathscr{R}_m(\Pi) + B_\ell\sqrt{\frac{q}{m}}$$

$$\leqslant O(1)\zeta^{\frac{1}{a_0}}B_\theta B_0 B_g L_{\partial^2\pi} T^{1-\frac{1}{a_1}}\sqrt{\frac{q}{m}} + O(1)\zeta^{\frac{1}{a_0}}B_0 L_\Delta T^{1-\frac{1}{a_1}}\sqrt{\frac{q}{m}}$$

$$\leqslant O(1)\zeta^{\frac{1}{a_0}}B_0 \max\{B_\theta B_g L_{\partial^2\pi}, L_\Delta\} T^{1-\frac{1}{a_1}}\sqrt{\frac{q}{m}}\,.$$

- Claim follows by plugging this bound into the basic inequality

$$\mathsf{err}(\hat{\pi}_{\mathrm{bc}}) \leqslant 8\gamma^{\frac{1}{a_0}} L_\Delta T^{1-\frac{1}{a_1}} \left(\mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_\star}(\xi;\hat{\pi}_{\mathrm{bc}},\pi_\star)\right)^{\frac{b_0}{a_1}}.$$ ■

Behavior cloning

- Only focusing on T, q, m: $\mathsf{err}(\hat{\pi}_{\mathrm{bc}}) \leqslant O(1) T^{(1-\frac{1}{a_1})(1+\frac{b_0}{a_1})} \left(\frac{q}{m}\right)^{\frac{b_0}{2a_1}}$.
- For contraction, $b_0 = a_1 = 1$ which yields $\mathsf{err}(\hat{\pi}_{\mathrm{bc}}) \leqslant O(1) \left(\frac{q}{m}\right)^{1/2}$.
  - Implies $m \geqslant \Omega(1) q/\varepsilon^2$ trajectories suffice for $\mathsf{err}(\hat{\pi}_{\mathrm{bc}}) \leqslant \varepsilon$.
- For tunable Ψ-IGS system, $b_0 = 1$ and $a_1 = 1 + p$ for $p \in (0,1)$, which yields $\mathsf{err}(\hat{\pi}_{\mathrm{bc}}) \leqslant O(1) T^{1-\frac{1}{(1+p)^2}} \left(\frac{q}{m}\right)^{\frac{1}{2(1+p)}}$.
  - Implies $m \geq \Omega(1) q \frac{T^{\frac{p(p+2)}{1+p}}}{\varepsilon^{2(1+p)}}$ trajectories suffice for $\mathsf{err}(\hat{\pi}_{\mathrm{bc}}) \leqslant \varepsilon$.
  - Note that for $p \in (0,(\sqrt{5}-1)/2) \approx (0,0.618)$, $\frac{p(p+2)}{1+p} < 1$ and hence sublinear in T trajectories suffice.

Epoch based algorithms

- In practice, behavior cloning is not desirable due to compounding errors.
- Two more practical algorithms are SMILe [Ross and Bagnell 2010] and DAgger [Ross et al. 2011].

SMILe

- Fix number of epochs E, mixing weight $\alpha \in (0,1]$.
- Set $\pi_0 = \pi_\star$.
- For $k \in \{0,\ldots,E-2\}$:
  - Collect rollouts $\{\{\varphi_t^{\pi_k}(\xi_i^k)\}_{t=0}^T\}_{i=1}^{m/E}$ and set $\hat{\pi}_k \in \arg\min_{\pi\in\Pi} \frac{1}{m/E} \sum_{i=1}^{m/E} \ell_{\pi_k}(\xi_i^k; \pi, \pi_\star)$.
  - $\pi_{k+1} = (1-\alpha)\pi_k + \alpha\hat{\pi}_k$.
- Collect rollouts $\{\{\varphi_t^{\pi_{E-1}}(\xi_i^{E-1})\}_{t=0}^T\}_{i=1}^{m/E}$ and set $\hat{\pi}_{E-1} \in \arg\min_{\pi\in\Pi} \frac{1}{m/E} \sum_{i=1}^{m/E} \ell_{\pi_{E-1}}(\xi_i^{E-1}; \pi, \pi_\star)$.
- Return $\pi_E = \frac{1}{1-(1-\alpha)^E}\left[(1-\alpha)\pi_{E-1} + \alpha\hat{\pi}_{E-1} - (1-\alpha)^E\pi_\star\right]$.

DAgger

- Fix number of epochs E, mixing weight $\alpha \in (0,1]$.
- Set $\hat{\pi}_0 \in \Pi$ arbitrarily.
- For $k \in \{0,\ldots,E-1\}$:
  - Set $\pi_k = \alpha^k \pi_\star + (1-\alpha^k)\hat{\pi}_k$.
  - Rollout policy π_k: $\{\{\varphi_t^{\pi_k}(\xi_i^k)\}_{t=0}^{T}\}_{i=1}^{m/E}$.
  - $\hat{\pi}_{k+1} \in \arg\min_{\pi\in\Pi} \sum_{j=0}^{k} \frac{1}{m/E} \sum_{i=1}^{m} \ell_{\pi_j}(\xi_i^j; \pi, \pi_\star)$.
- Return best of $\hat{\pi}_1, \ldots, \hat{\pi}_E$.

SMILe

- To analyze SMILe using our IGS machinery, we need to make a few modifications.
- We make the following modifications:
    - The learned policies are constrained to be IGS stable.
    - The learned policies are constrained to not change too much from epoch to epoch (trust-region constraint).
- We call these modifications CMILe (Constrained SMILe).

CMILe

- Fix number of epochs E, mixing weight $\alpha \in (0,1]$, trust-region weights $\{c_k\}$.
- Set $\pi_0 = \pi_\star$.
- For $k \in \{0,\ldots,E-2\}$:
    - Collect rollouts $\{\{\varphi_t^{\pi_k}(\xi_i^k)\}_{t=0}^{T}\}_{i=1}^{m/E}$ and set:

$$\hat{\pi}_k \in \arg\min_{\pi\in\Pi} \frac{1}{m/E} \sum_{i=1}^{m/E} \ell_{\pi_k}(\xi_i^k; \pi, \pi_\star)$$

$$\text{s.t. } (1-\alpha)\pi_k + \alpha\pi \in \Pi_\Psi,\ \frac{1}{m/E} \sum_{i=1}^{m} \ell_{\pi_k}(\xi_i^k; \pi, \pi_k) \leqslant c_k\,.$$

    - $\pi_{k+1} = (1-\alpha)\pi_k + \alpha\hat{\pi}_k$.
- Collect rollouts $\{\{\varphi_t^{\pi_{E-1}}(\xi_i^{E-1})\}_{t=0}^{T}\}_{i=1}^{m/E}$ and set :

$$\hat{\pi}_{E-1} \in \arg\min_{\pi\in\Pi} \frac{1}{m/E} \sum_{i=1}^{m/E} \ell_{\pi_{E-1}}(\xi_i^{E-1}; \pi, \pi_\star)$$

$$\text{s.t. } \frac{1}{1-(1-\alpha)^E}\left[(1-\alpha)\pi_{E-1} + \alpha\pi - (1-\alpha)^E\pi_\star\right] \in \Pi_\Psi,\ \frac{1}{m/E} \sum_{i=1}^{M} \ell_{\pi_{E-1}}(\xi_i^{E-1}; \pi, \pi_{E-1}) \leqslant c_{E-1}\,.$$

- Return $\pi_E = \dfrac{1}{1-(1-\alpha)^E}\left[(1-\alpha)\pi_{E-1} + \alpha\hat{\pi}_{E-1} - (1-\alpha)^E\pi_\star\right]$.

Analysis of CMILe

- We will analyze CMILe similar to how we analyzed BC.
- The key idea is to bound $\mathsf{err}(\pi_{k+1})$ by some function of $\mathsf{err}(\pi_k)$.
- We start with the basic algebraic identity:

$$\begin{aligned}\Delta_{\pi_{k+1},\pi_\star}(x) &= g(x)(\pi_{k+1}(x) - \pi_\star(x)) \\ &= (1-\alpha)g(x)(\pi_k(x) - \pi_\star(x)) + \alpha g(x)(\hat{\pi}_k(x) - \pi_\star(x)) \\ &= (1-\alpha)\Delta_{\pi_k,\pi_\star}(x) + \alpha\Delta_{\hat{\pi}_k,\pi_\star}(x)\,.\end{aligned}$$

- With this identity, we derive the following CMILe **basic inequality**:

$$\begin{aligned}\mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_{k+1}}(\xi;\pi_{k+1},\pi_\star) &= \mathbb{E}_{\xi\sim\mathscr{D}}\sum_{t=0}^{T-1}\|\Delta_{\pi_{k+1},\pi_\star}(\varphi_t^{\pi_{k+1}}(\xi))\|_2 \\ &\leqslant \mathbb{E}_{\xi\sim\mathscr{D}}\sum_{t=0}^{T-1}\|\Delta_{\pi_{k+1},\pi_\star}(\varphi_t^{\pi_k}(\xi))\|_2 + L_\Delta\mathbb{E}_{\xi\sim\mathscr{D}}\mathsf{disc}_T(\xi;\pi_{k+1},\pi_k) \\ &\leqslant (1-\alpha)\mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_k}(\xi;\pi_k,\pi_\star) + \alpha\mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_k}(\xi;\hat{\pi}_k,\pi_\star) + L_\Delta\mathbb{E}_{\xi\sim\mathscr{D}}\mathsf{disc}_T(\xi;\pi_{k+1},\pi_k)\,.\end{aligned}$$

- We first focus on bounding $\mathbb{E}_{\xi\sim\mathscr{D}}\mathrm{disc}_T(\xi;\pi_{k+1},\pi_k)$.
- Due to the update $\pi_{k+1}=(1-\alpha)\pi_k+\alpha\hat{\pi}_k$, we have $\pi_{k+1}-\pi_k=\alpha(\hat{\pi}_k-\pi_k)$. Therefore $\mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_k}(\xi;\pi_{k+1},\pi_k)=\alpha\mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_k}(\xi;\hat{\pi}_k,\pi_k)$.
- Because π_{k+1} is Ψ-IGS, as long as $\mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_k}(\xi;\hat{\pi}_k,\pi_k)\leqslant 1$:

  $$\mathbb{E}_{\xi\sim\mathscr{D}}\mathrm{disc}_T(\xi;\pi_{k+1},\pi_k)\leqslant 8\gamma^{\frac{1}{a_0}}T^{1-\frac{1}{a_1}}\left(\alpha\mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_k}(\xi;\hat{\pi}_k,\pi_k)\right)^{\frac{b_0}{a_1}}.$$

- By uniform convergence:

  $$\mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_k}(\xi;\hat{\pi}_k,\pi_k)\leqslant\frac{1}{m/E}\sum_{i=1}^{m/E}\ell_{\pi_k}(\xi_i^k;\hat{\pi}_k,\pi_k)+2\mathscr{R}_{m/E}(\Pi)+B_\ell\sqrt{\frac{\log(2/\delta)}{m/E}}$$

  $$\leqslant c_k+2\mathscr{R}_{m/E}(\Pi)+B_\ell\sqrt{\frac{\log(2/\delta)}{m/E}}$$ [trust region constraint] .

- Next, we turn to bounding $\mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_k}(\xi;\hat{\pi}_k,\pi_\star)$.
- By uniform convergence,

$$\begin{aligned}
\mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_k}(\xi;\hat{\pi}_k,\pi_\star) &\leqslant \frac{1}{m/E}\sum_{i=1}^{m/E}\ell_{\pi_k}(\xi_i^k;\hat{\pi}_k,\pi_\star) + 2\mathscr{R}_{m/E}(\Pi) + B_\ell\sqrt{\frac{\log(2/\delta)}{m/E}} \\
&\leqslant \frac{1}{m/E}\sum_{i=1}^{m/E}\ell_{\pi_k}(\xi_i^k;\pi_k,\pi_\star) + 2\mathscr{R}_{m/E}(\Pi) + B_\ell\sqrt{\frac{\log(2/\delta)}{m/E}} && [\pi_k \text{ is feasible}] \\
&\leqslant \mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_k}(\xi;\pi_k,\pi_\star) + 4\mathscr{R}_{m/E}(\Pi) + 2B_\ell\sqrt{\frac{\log(2/\delta)}{m/E}} && [\text{uniform convergence}]\,.
\end{aligned}$$

- Recall our basic inequality:
$\mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_{k+1}}(\xi;\pi_{k+1},\pi_\star) \leqslant (1-\alpha)\mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_k}(\xi;\pi_k,\pi_\star) + \alpha\mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_k}(\xi;\hat{\pi}_k,\pi_\star) + L_\Delta\mathbb{E}_{\xi\sim\mathscr{D}}\mathsf{disc}_T(\xi;\pi_{k+1},\pi_k)\,.$

- Combining bounds on $\mathbb{E}_{\xi\sim\mathscr{D}}\mathsf{disc}_T(\xi;\pi_{k+1},\pi_k)$ and $\mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_k}(\xi;\hat{\pi}_k,\pi_\star)$, with probability at least $1-3\delta$,

$$\mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_{k+1}}(\xi;\pi_{k+1},\pi_\star) \leqslant \mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_k}(\xi;\pi_k,\pi_\star) + 4\alpha\mathscr{R}_{m/E}(\Pi) + 2\alpha B_\ell\sqrt{\frac{\log(2/\delta)}{m/E}}$$

$$+8\gamma^{\frac{1}{a_0}}T^{1-\frac{1}{a_1}}L_\Delta\left[\alpha c_k + 2\alpha\mathscr{R}_{m/E}(\Pi) + \alpha B_\ell\sqrt{\frac{\log(2/\delta)}{m/E}}\right]^{\frac{b_0}{a_1}}.$$

- This recursion yields a bound on $\mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_{E-1}}(\xi;\pi_{E-1},\pi_\star)$.

- Obtaining a bound on the final policy (with no expert) $\mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_E}(\xi;\pi_E,\pi_\star)$ is more involved but follows similar ideas.

CMILe

- **Theorem:** Suppose that:
  - $m \geqslant \Omega(1) E(q \vee \log E) \zeta^{\frac{2}{a_0}} B_0^2 T^{2(1-\frac{1}{a_1})} \max\{B_g B_\theta L_{\partial^2 \pi}, L_\Delta\}^2$,
  - $c_k \leqslant O(1) \zeta^{\frac{1}{a_0}} B_0 T^{1-\frac{1}{a_1}} \max\{B_g B_\theta L_{\partial^2 \pi}, L_\Delta\} \sqrt{\frac{E(q \vee \log E)}{m}}$,
  - $E \geqslant \frac{1}{\alpha} \log(1/\alpha)$ and $\alpha \leqslant \min\left\{\frac{1}{2}, \frac{1}{L_\Delta \gamma^{\frac{1}{a_0}} T^{1-\frac{1}{a_1}}}\right\}$.
- Then with probability at least $1 - e^{-q}$:

  $\mathrm{err}(\pi_E) \leqslant O(1) \zeta^{\frac{b_0}{a_0 a_1}} \gamma \left(1 - \frac{b_0^2}{a_1^2}\right)^{\frac{1}{a_0}} B_0^{\frac{b_0}{a_1}} T^{1-\frac{1}{a_1}} L^{1-\frac{b_0^2}{a_1^2}} \max\{B_g B_\theta L_{\partial^2 \pi}, L_\Delta\}^{\frac{b_0}{a_1}} E^{1+\frac{b_0}{2a_1}} \left(\frac{q \vee \log E}{m}\right)^{\frac{b_0^2}{2a_1^2}}$.

CMILe

- Only focusing on the dependence on T, q, m: $\mathsf{err}(\pi_E) \leqslant \tilde{O}(1) T^{\left(1-\frac{1}{a_1}\right)\left(2+\frac{b_0}{2a_1}\right)} \left(\frac{q}{m}\right)^{\frac{b_0^2}{2a_1^2}}$.
- For contraction, we have $\mathsf{err}(\pi_E) \leqslant \tilde{O}(1)\sqrt{\frac{q}{m}}$.
  - $m \geqslant \tilde{\Omega}(1) q/\varepsilon^2$ trajectories suffice for $\mathsf{err}(\pi_E) \leqslant \varepsilon$.
- For the tunable Ψ-IGS system, $\mathsf{err}(\pi_E) \leqslant O(1) T^{\left(1-\frac{1}{1+p}\right)\left(2+\frac{1}{2(1+p)}\right)} \left(\frac{q}{m}\right)^{\frac{1}{2(1+p)^2}}$,
  - $m \geq \tilde{\Omega}(1) \frac{q}{\varepsilon^{2(1+p)^2}} T^{\frac{2p}{1+p}\left(2(1+p)^2+\frac{1}{2(1+p)}\right)}$ trajectories suffice for $\mathsf{err}(\pi_E) \leqslant \varepsilon$.
  - For $p \in (0, 0.183)$, we have that $\frac{2p}{1+p}\left(2(1+p)^2+\frac{1}{2(1+p)}\right) < 1$, and hence sublinear in T trajectories suffice.

Why do we care about imitation loss?

- Suppose that $h : \mathbb{R}^{n\times(T+1)} \to \mathbb{R}^s$ is an L_h-Lipschitz function of the state trajectory.
- Define $h_\pi(\xi) := h(\{\varphi_t^\pi(\xi)\}_{t=0}^T)$.
- Then because $\pi_\star \in \Pi_\Psi$, we have:

  $$\mathbb{E}_{\xi\sim\mathscr{D}}\|h_{\pi_\star}(\xi) - h_{\hat{\pi}}(\xi)\|_2 \leqslant L_h\mathbb{E}_{\xi\sim\mathscr{D}}\mathsf{disc}_T(\xi;\hat{\pi},\pi_\star) \leqslant 4L_h\gamma^{\frac{1}{a_0}}T^{1-\frac{1}{a_1}}\left(\mathsf{err}(\hat{\pi})\right)^{\frac{b_0}{a_1}}.$$

- Therefore, small $\mathsf{err}(\hat{\pi})$ implies that the performance of the observable h under the learned policy mimics that of the observable h under the expert policy.
    - This observable can encode things like trajectory tracking error and safety constraints.

Practical implementation

- **Trust region policy constraint:** the policy trust region constraints are implemented by constraining the parameters $\hat{\theta}_k$ for $\hat{\pi}_k$ to lie close in Euclidean norm to the parameters $\hat{\theta}_{k-1}$ for $\hat{\pi}_{k-1}$: $\|\hat{\theta}_k - \hat{\theta}_{k-1}\|_2 \leqslant \kappa$.

- **Ψ-IGS constraint:** while one could use the Lyapunov characterization, it is challenging to ensure that the Lyapunov equality holds for all $x \in \mathbb{R}^n$. In our implementation, we simply drop this constraint and note that it only affects the low-data regime in practice.

    - **Open question:** can we analyze BC/CMILe by arguing that the learned policies are already stable without explicit constraints.

Tunable Ψ-IGS system

- Consider the dynamical system in $\mathbb{R}^{10}$:

  $$x_{t+1} = x_t - 0.5x_t \frac{|x_t|^p}{1 + 0.5|x_t|^p} + \frac{1}{1 + |x_t|^p}(h(x_t) + u_t).$$

- All arithmetic operations are element-wise.

- $h : \mathbb{R}^{10} \to \mathbb{R}^{10}$ is a randomly initialized two-layer MLP with zero biases, hidden width 32, and $\tanh$ activations.

- Expert policy is $\pi_\star = -h$, so expert's closed-loop dynamics are:

  $x_{t+1} = x_t - 0.5x_t \frac{|x_t|^p}{1 + 0.5|x_t|^p}$, which is Ψ-IGS with $a_1 = 1 + p$.

Tunable Ψ-IGS system

Table of final $\|x_T^{\text{expert}} - x_t^{\text{IL}}\|_2$ for all IL algorithms

p	**BC**	**CMILe**	**CMILe+Agg**	**DAgger**
1	0.615 ± 0.154	0.247 ± 0.071	0.239 ± 0.038	0.474 ± 0.131
2	1.194 ± 0.130	0.737 ± 0.059	0.578 ± 0.095	0.865 ± 0.142
3	1.637 ± 0.220	1.115 ± 0.066	0.868 ± 0.065	1.199 ± 0.130
4	1.976 ± 0.106	1.409 ± 0.080	1.111 ± 0.080	1.441 ± 0.126
5	2.107 ± 0.079	1.570 ± 0.055	1.240 ± 0.091	1.594 ± 0.146

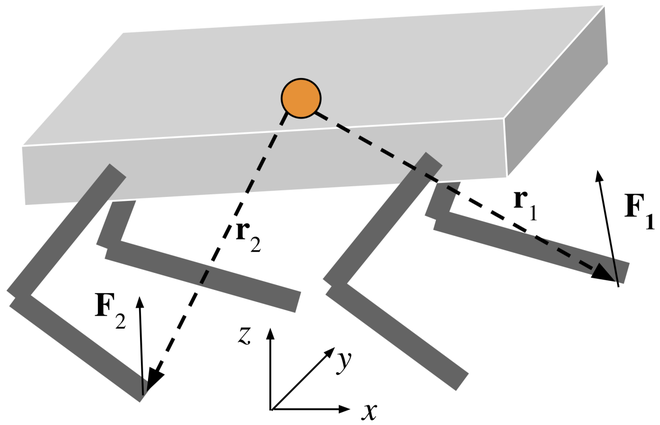
Tunable Ψ-IGS system

Table of final average closed-loop imitation loss $\frac{1}{T}\mathbb{E}_{\xi\sim\mathscr{D}}\ell_{\pi_E}(\xi;\pi_E,\pi_\star)$ for all IL algorithms

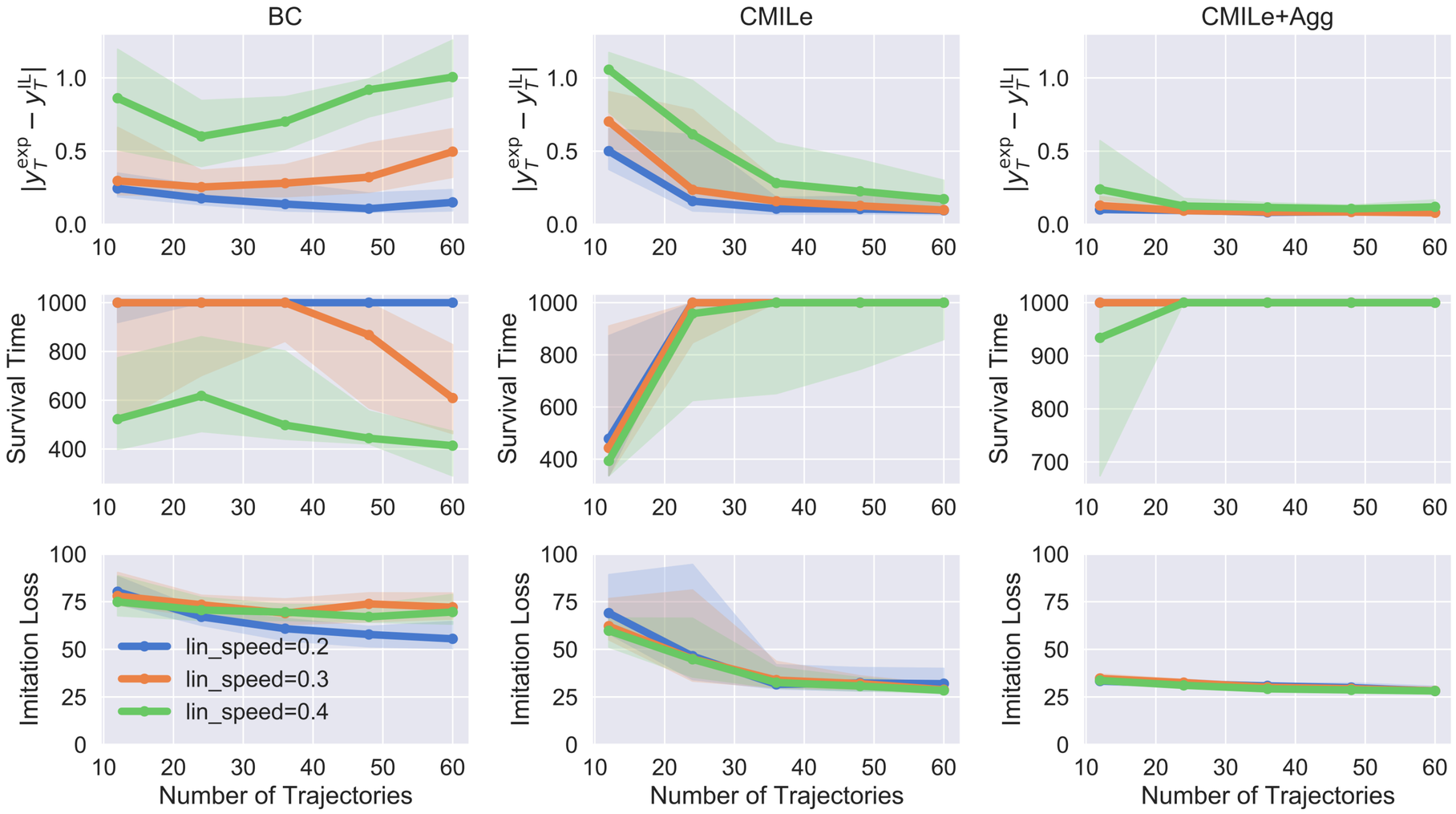
p	**BC**	**CMILe**	**CMILe+Agg**	**DAgger**
1	0.161 ± 0.088	0.113 ± 0.019	0.021 ± 0.005	0.050 ± 0.017
2	0.324 ± 0.097	0.128 ± 0.013	0.021 ± 0.003	0.052 ± 0.021
3	0.498 ± 0.120	0.151 ± 0.023	0.028 ± 0.007	0.064 ± 0.018
4	0.672 ± 0.139	0.163 ± 0.019	0.031 ± 0.007	0.061 ± 0.023
5	0.905 ± 0.154	0.154 ± 0.030	0.035 ± 0.007	0.058 ± 0.015

Unitree Laikago

- We consider IL for the Unitree Laikago in simulation.
- We imitate an expert MPC controller which allows the Laikago to **walk sideways** at varying fixed speeds:
    - Increasing the linear (sideways) velocity decreases task stability.
- MPC controller is based on center-of-mass dynamics as described in [Di Carlo et al. 2018].



Unitree Laikago



References

- J. Di Carlo, P. M. Wensing, B. Katz, G. Bledt, and S. Kim. Dynamics Locomotion in the MIT Cheetah 3 Through Convex Model-Predictive Control. IROS, 2018.

- S. Ross and J. A. Bagnell. Efficient Reductions for Imitation Learning. AISTATS, 2010.

- S. Ross, G. J. Gordon, and J. A. Bagnell. A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning. AISTATS, 2011.

- D. N. Tran and B. S. Rüffer and C. M. Kellett. Incremental Stability Properties for Discrete-Time Systems. CDC, 2016.

Regret Bounds for Adaptive Nonlinear Control

Nicholas M. Boffi, Stephen Tu, and Jean-Jacques E. Slotine

arxiv.org/abs/2011.13101

Motivation

- Recent success in **reinforcement learning** (RL) has sparked much interest in analyzing RL for learning to control unknown systems in continuous settings.
- **Adaptive control** from the 90s also studies learning to control unknown systems.
- **Question:** how do these two fields/communities relate to each other?

Comparing RL to adaptive control

	RL	Adaptive Control
Dynamics	Some structure may be known (e.g., linear MDPs, linear Q-functions, low Bellman rank, etc.), but otherwise very general.	Very structured: nominal dynamics known, known basis functions for un-modeled dynamics.
Discrete-time or continuous-time	Discrete-time.	Continuous-time.
Guarantees	Regret bounds, PAC-style guarantees.	Asymptotic convergence, no rates.
Algorithms	Typically optimism/UCB-based, can be intractable.	Typically certainty-equivalence based, computationally efficient.

Goal

- Port over several nice properties of RL to adaptive control:
    - Handle stochasticity.
    - Discrete-time evolution.
    - Finite-time regret bounds.

Problem setup

- Consider the following non-linear system evolving in $\mathbb{R}^n$:
  $x_{t+1} = f(x_t, t) + B(x_t, t)(u_t - Y(x_t, t)\alpha) + w_t$.
- $f(x, t)$ and $B(x, t)$ are known **nominal** dynamics. $f(0,t) = 0$ for all t.
- $Y(x, t)$ are known **basis** functions.
- $\alpha \in \mathbb{R}^p$ are **unknown** parameters.
- w_t is zero-mean Gaussian noise.
- Known as **matched uncertainty** setup in adaptive control.

When does this model arise?

- Consider a fully-actuated system: $x_{t+1} = f(x_t) + u_t$.
- We want to drive x_t to the origin, but $f(x_t)$ is unknown.
- Picking any $\rho \in (-1,1)$, and playing $u_t = \rho x_t - \tilde{u}_t$ yields the closed-loop: $x_{t+1} = \rho x_t + f(x_t) - \tilde{u}_t$.
- If $f(x_t)$ is well approximated (e.g., with physical or random basis functions) as $f(x_t) \approx Y(x_t)\alpha$ for some α, then this follows in our model.

When does this model arise?

- Suppose a controller u_{sim} is designed on model $f_{\text{sim}}(x) + g_{\text{sim}}(x)u$.
- Suppose that u_{real} is optimal for $f_{\text{real}}(x) + g_{\text{real}}(x)u$.
- Then playing u_{sim} on the real system yields the closed-loop dynamics:
  $$x_{t+1} = f_{\text{real}}(x) + g_{\text{real}}(x)u_{\text{sim}}(x)$$
  $$= f^{\text{cl}}_{\text{real}}(x) + g_{\text{real}}(x)(u_{\text{sim}}(x) - u_{\text{real}}(x)) \qquad [f^{\text{cl}}_{\text{real}}(x) := f_{\text{real}}(x) + g_{\text{real}}(x)u_{\text{real}}(x)]\,.$$
- Now suppose that $u_{\text{real}}(x) \approx Y(x)\alpha$ for α.
- This falls within our model*.

Problem setup

- Model: $x_{t+1} = f(x_t, t) + B(x_t, t)(u_t - Y(x_t, t)\alpha) + w_t$.
- Comparator dynamics: $x^c_{t+1} = f(x^c_t, t) + w_t$, i.e., $u_t = Y(x^c_t, t)\alpha$.
- The regret of a control strategy $\{u^a_t\}$ is:

  $$\mathrm{Reg}_T^{\mathrm{Ctrl}} := \mathbb{E}\left[\sum_{t=0}^{T} \|x^a_t\|_2^2 - \sum_{t=0}^{T} \|x^c_t\|_2^2\right].$$

- Here, $x^a_{t+1} = f(x^a_t, t) + B(x^a_t, t)(u^a_t - Y(x^a_t, t)) + w_t$.
- Regret measures the cost of dealing with the uncertain (unmodelled) dynamics.

Online convex optimization primer

- A game played between an adversary and a learner.
- For $t = 1,2,...$:
  - Adversary chooses convex loss function $\ell_t : \mathbb{R}^n \to \mathbb{R}$.
  - Learner chooses $w_t \in W$, suffers loss $\ell_t(w_t)$, and observes $(\ell_t(w_t), \nabla \ell_t(w_t))$.
- The performance of the learner is measured via the regret, comparing the performance to the best fixed predictor in hindsight:

  $$\mathrm{Reg}_T^{\mathrm{Pred}} := \sum_{t=1}^{T} \ell_t(w_t) - \min_{w \in W} \sum_{t=1}^{T} \ell_t(w).$$

- A learner that achieves $\mathrm{Reg}_T^{\mathrm{Pred}} \leqslant o(T)$ is said to be **“no-regret”**.
- See [Hazan 2016] for a thorough treatment of the subject.

Online convex optimization primer

- **Online gradient descent:**
  - Learner updates $w_{t+1} = \Pi_W(w_t - \eta_t \nabla \ell_t(w_t))$, where $\Pi_W(x) = \arg\min_{y \in W} \|x - y\|_2$.
  - Achieves $\mathrm{Reg}_T^{\mathrm{Pred}} \leqslant O(\sqrt{T})$ with $\eta_t = 1/\sqrt{T}$.
- **Online Newton:**
  - Learner updates $w_{t+1} = \Pi_{t,W}(w_t - \eta A_t^{-1} \nabla \ell_t(w_t))$, with $A_t = \lambda I + \sum_{i=1}^{t} \nabla \ell_i(w_i) \nabla \ell_i(w_i)^T$ and $\Pi_{t,W}(x) = \arg\min_{y \in W} \|x - y\|_{A_t}$.
  - Achieves $\mathrm{Reg}_T^{\mathrm{Pred}} \leqslant O(\log T)$ with $\eta \asymp 1$ for exp-concave functions, i.e., $\nabla^2 \ell_t(w) \succcurlyeq \mu \nabla \ell_t(w) \nabla \ell_t(w)^T$.

Adaptive control with OCO

- We use OCO to estimate $\hat{\alpha}_t$.
- We play the certainty-equivalence controller: $u_t = Y(x_t, t)\hat{\alpha}_t$.
- Closed-loop: $x_{t+1} = f(x_t, t) + G(x_t, t)(\hat{\alpha}_t - \alpha) + w_t$, $G(x, t) := B(x, t)Y(x, t)$.
- After observing x_{t+1}, can obtain necessary signal to update $\hat{\alpha}_t$:
  - **Recursive least-squares:** $y_t := f(x_t, t) + G(x_t, t)\hat{\alpha}_t - x_{t+1} = G(x_t, t)\alpha - w_t$.
  - **Gradient-based:** $\ell_t(\hat{\alpha}) = \frac{1}{2}\|G(x_t, t)(\hat{\alpha} - \alpha) + w_t\|_2^2$,
    $\nabla \ell_t(\hat{\alpha}_t) = G(x_t, t)^T(x_{t+1} - f(x_t, t))$.

Adaptive control with OCO

- Focus on online gradient based algorithms for now.
- Recall $\ell_t(\hat{\alpha}) = \frac{1}{2}\|G_t(\hat{\alpha} - \alpha) + w_t\|_2^2$.
- Online convex optimization bounds give us:
  $$\frac{1}{2}\sum_{t=0}^{T-1} \mathbb{E}\|G_t\tilde{\alpha}_t\|_2^2 = \mathbb{E}\left[\sum_{t=0}^{T-1} \ell_t(\hat{\alpha}_t) - \ell_t(\alpha)\right] \leqslant \mathsf{Reg}_T^{\mathrm{Pred}}.$$
- If $f(x,t) \equiv 0$, then:
  $\mathsf{Reg}_T^{\mathrm{Ctrl}} = \mathbb{E}\left[\sum_{t=0}^{T} \|x_t^a\|_2^2 - \sum_{t=0}^{T} \|x_t^c\|_2^2\right] = \sum_{t=0}^{T-1} \mathbb{E}\|G_t\tilde{\alpha}_t\|_2^2$, so we are done.
- In general though, how do we handle non-zero dynamics?

Adaptive control with OCO

- **Theorem (informal):** under suitable stability assumptions on $f(x, t)$, we have: $\mathsf{Reg}_T^{\mathrm{Ctrl}} \lesssim \sqrt{T \cdot \mathsf{Reg}_T^{\mathrm{Pred}}}$.
- Consequently, if you use either online Newton or online recursive LS, we have $\mathsf{Reg}_T^{\mathrm{Ctrl}} \lesssim \sqrt{T}\mathrm{polylog(T)}$.
- If you use vanilla gradient descent, then $\mathsf{Reg}_T^{\mathrm{Ctrl}} \lesssim T^{3/4}$.
- **Note:** a similar $\sqrt{T \cdot \mathsf{Reg}_T^{\mathrm{Pred}}}$ regret bound found in [Foster and Rakhlin 2020] for contextual bandits.

Main stochastic perturbation result

- The main technical ingredient to show $\mathsf{Reg}_T^{\mathrm{Ctrl}} \lesssim \sqrt{T \cdot \mathsf{Reg}_T^{\mathrm{Pred}}}$ is a stochastic perturbation result.
- Consider two stochastic processes $\{x_t\}$, $\{y_t\}$ with $\{w_t\}$, $\{v_t\}$ iid $N(0,\sigma_w^2 I)$.
  $x_{t+1} = f(x_t, t) + d_t(x_0, \ldots, x_t) + w_t,$
  $y_{t+1} = f(y_t, t) + v_t\,.$
- **Question:** what assumptions on $f(x,t)$ allow us to bound $\mathbb{E}\left[\sum_{t=0}^{T} \|x_t\|_2^2 - \sum_{t=0}^{T} \|y_t\|_2^2\right]$ by a function of $\mathbb{E}\sum_{t=0}^{T-1} \|d_t\|_2^2$?
- Note that in the context of our adaptive control problem, $\mathbb{E}\sum_{t=0}^{T-1} \|d_t\|_2^2 \asymp \mathsf{Reg}_T^{\mathrm{Pred}}$.

Main stochastic perturbation result

- Recall that $\{y_t\}$ is the **unperturbed** process.
- **Assumption (L2 geometric ergodicity):** there exists a $\rho \in (0,1)$ and positive $R(x)$ such that for all non-negative t, k:
  $$|\,\mathbb{E}[\|y_{t+k}\|_2^2 | \, y_t = x] - \mathbb{E}[\|y_{t+k}\|_2^2 | \, y_t = 0]\,| \leqslant R(x)\rho^k.$$
- **Theorem:** Under L2 geometric ergodicity, we have:
  $$\mathbb{E}\left[\sum_{t=0}^{T} \|x_t\|_2^2 - \sum_{t=0}^{T} \|y_t\|_2^2\right]$$
  $$\leqslant \frac{\sqrt{2}}{1-\rho}\sqrt{\mathbb{E}\sum_{t=0}^{T-1} \min\left\{\frac{\|d_t\|_2^2}{\sigma_w^2}, 1\right\}}\sqrt{\mathbb{E}\sum_{t=1}^{T} R(x_t)^2 + \mathbb{E}\sum_{t=0}^{T-1} R(f(x_t,t)+w_t)^2}\,.$$

- **Proof:** We build on ideas from [Kakade et al. 2020] and [Yu et al. 2020].
- We start by defining a “cost-to-go” function for the unperturbed dynamics:

$$V_{s:t}(x) := \sum_{k=s}^{t} \mathbb{E}[\|y_k\|_2^2 | \, y_s = x].$$

- Now define $h_k(x) := V_{k:T}(x) - V_{k:T}(0)$. We have:

$$\begin{aligned}
|\, h_k(x) \,| &= \left| \sum_{t=k}^{T} \mathbb{E}[\|y_t\|_2^2 | \, y_k = x] - \sum_{t=k}^{T} \mathbb{E}[\|y_t\|_2^2 | \, y_k = 0] \right| \\
&\leqslant \sum_{t=k}^{T} |\, \mathbb{E}[\|y_t\|_2^2 | \, y_k = x] - \mathbb{E}[\|y_t\|_2^2 | \, y_k = 0] \,| && \text{[triangle inequality]} \\
&\leqslant R(x) \sum_{t=k}^{T} \rho^{t-k} && \text{[L2 geometric ergodicity]} \\
&\leqslant \frac{R(x)}{1-\rho}.
\end{aligned}$$

- Now define $W_k, k \in \{0,\ldots,T\}$ as:

$$W_k = \mathbb{E}\sum_{t=0}^{T} \|\varphi_t\|_2^2,\ \varphi_{t+1} = \begin{cases} f(\varphi_t, t) + d_t(\varphi_0, \ldots, \varphi_t) + w_t & \text{if } t < k \\ f(\varphi_t, t) + w_t & \text{if } t \geqslant k \end{cases},\ \varphi_0 = x_0\,.$$

- With this notation, we have:

$$\mathbb{E}\sum_{t=0}^{T} \|x_t\|_2^2 = W_T,\ \mathbb{E}\sum_{t=0}^{T} \|y_t\|_2^2 = W_0.$$

- Therefore, we can write the telescoping sum:

$$\mathbb{E}\left[\sum_{t=0}^{T} \|x_t\|_2^2 - \sum_{t=0}^{T} \|y_t\|_2^2\right] = W_T - W_0 = \sum_{k=0}^{T-1} (W_{k+1} - W_k).$$

- Now observe that for $k \in \{0,\ldots,T-1\}$, with $\tilde{p}_{0:k}$ denoting the joint distribution of $(x_0, \ldots, x_k)$:

$$W_k = \mathbb{E}\sum_{t=0}^{k} \|x_t\|_2^2 + \mathbb{E}_{\tilde{p}_{0:k}}\mathbb{E}_{x\sim N(f(x_k,k),\sigma_w^2 I)} V_{k+1:T}(x),$$

$$W_{k+1} = \mathbb{E}\sum_{t=0}^{k} \|x_t\|_2^2 + \mathbb{E}_{\tilde{p}_{0:k}}\mathbb{E}_{x\sim N(f(x_k,k)+d_k(x_0,\ldots,x_k),\sigma_w^2 I)} V_{k+1:T}(x)\,.$$

- With this notation, we have the identity:

$$\mathbb{E}\left[\sum_{t=0}^{T}\|x_t\|_2^2-\sum_{t=0}^{T}\|y_t\|_2^2\right]=\sum_{k=0}^{T-1}(W_{k+1}-W_k)$$

$$=\sum_{k=0}^{T-1}\mathbb{E}_{\tilde{p}_{0:k}}[\mathbb{E}_{x\sim N(f(x_k,k)+d_k,\sigma_w^2 I)}V_{k+1:T}(x)-\mathbb{E}_{x\sim N(f(x_k,k),\sigma_w^2 I)}V_{k+1:T}(x)]$$

$$=\sum_{k=0}^{T-1}\mathbb{E}_{\tilde{p}_{0:k}}[\mathbb{E}_{x\sim N(f(x_k,k)+d_k,\sigma_w^2 I)}V_{k+1:T}(x)-V_{k+1:T}(0)+V_{k+1:T}(0)-\mathbb{E}_{x\sim N(f(x_k,k),\sigma_w^2 I)}V_{k+1:T}(x)]$$

$$=\sum_{k=0}^{T-1}\mathbb{E}_{\tilde{p}_{0:k}}[\mathbb{E}_{x\sim N(f(x_k,k)+d_k,\sigma_w^2 I)}h_{k+1}(x)-\mathbb{E}_{x\sim N(f(x_k,k),\sigma_w^2 I)}h_{k+1}(x)]\,.$$

- We now state a technical lemma from [Kakade et al. 2020].
- **Lemma:** for any measurable g and two Gaussians $N_i = N(\mu_i, \sigma^2 I)$, we have:
$$\mathbb{E}_{N_1} g - \mathbb{E}_{N_2} g \leqslant \min\left\{\frac{\|\mu_1 - \mu_2\|_2}{\sigma}, 1\right\}\left[\sqrt{\mathbb{E}_{N_1} g^2} + \sqrt{\mathbb{E}_{N_2} g^2}\right].$$
- **Note:** this is similar to a classic result which states that for any $g : X \to [-B, B]$ and any two distributions μ_1, μ_2, we have:
$$|\mathbb{E}_{\mu_1} g - \mathbb{E}_{\mu_2} g| \leqslant 2B\|\mu_1 - \mu_2\|_{\text{tv}}.$$
    - But this result requires that g is bounded almost surely, while the lemma above only requires bounded second moments.

- Define $\tilde{N}_k = N(f(x_k, k)+d_k, \sigma_w^2 I)$ and $N_k = N(f(x_k, k), \sigma_w^2 I)$.
- With the [Kakade et al. 2020] lemma, we have:

$$\mathbb{E}_{x\sim N(f(x_k,k)+d_k,\sigma_w^2 I)} h_{k+1}(x) - \mathbb{E}_{x\sim N(f(x_k,k),\sigma_w^2 I)} h_{k+1}(x)$$

$$\leqslant \min\left\{\frac{\|d_k\|_2}{\sigma_w}, 1\right\}\left[\sqrt{\mathbb{E}_{\tilde{N}_k} h_{k+1}^2} + \sqrt{\mathbb{E}_{N_k} h_{k+1}^2}\right]$$

$$\leqslant \frac{1}{1-\rho}\min\left\{\frac{\|d_k\|_2}{\sigma_w}, 1\right\}\left[\sqrt{\mathbb{E}_{\tilde{N}_k} R^2} + \sqrt{\mathbb{E}_{N_k} R^2}\right]. \qquad \left[\text{since } |h_k(x)| \leqslant \frac{R(x)}{1-\rho}\right]$$

- Therefore:

$$\mathbb{E}\left[\sum_{t=0}^{T}\|x_t\|_2^2-\sum_{t=0}^{T}\|y_t\|_2^2\right]\leqslant\frac{1}{1-\rho}\sum_{k=0}^{T-1}\mathbb{E}_{\tilde{p}_{0:k}}\left[\min\left\{\frac{\|d_k\|_2}{\sigma_w},1\right\}\left[\sqrt{\mathbb{E}_{\tilde{N}_k}R^2}+\sqrt{\mathbb{E}_{N_k}R^2}\right]\right]$$

$$\leqslant\frac{1}{1-\rho}\sum_{k=0}^{T-1}\sqrt{\mathbb{E}_{\tilde{p}_{0:k}}\min\left\{\frac{\|d_k\|_2^2}{\sigma_w^2},1\right\}}\sqrt{\mathbb{E}_{\tilde{p}_{0:k}}\left(\sqrt{\mathbb{E}_{\tilde{N}_k}R^2}+\sqrt{\mathbb{E}_{N_k}R^2}\right)^2}\qquad[\mathbb{E}AB\leqslant\sqrt{\mathbb{E}A^2}\sqrt{\mathbb{E}B^2}]$$

$$\leqslant\frac{1}{1-\rho}\sqrt{\sum_{k=0}^{T-1}\mathbb{E}_{\tilde{p}_{0:k}}\min\left\{\frac{\|d_k\|_2^2}{\sigma_w^2},1\right\}}\sqrt{\sum_{k=0}^{T-1}\mathbb{E}_{\tilde{p}_{0:k}}\left(\sqrt{\mathbb{E}_{\tilde{N}_k}R^2}+\sqrt{\mathbb{E}_{N_k}R^2}\right)^2}\qquad\text{[Cauchy-Schwarz]}$$

$$\leqslant\frac{\sqrt{2}}{1-\rho}\sqrt{\sum_{k=0}^{T-1}\mathbb{E}_{\tilde{p}_{0:k}}\min\left\{\frac{\|d_k\|_2^2}{\sigma_w^2},1\right\}}\sqrt{\sum_{k=0}^{T-1}\mathbb{E}_{\tilde{p}_{0:k}}\mathbb{E}_{\tilde{N}_k}R^2+\mathbb{E}_{\tilde{p}_{0:k}}\mathbb{E}_{N_k}R^2}\qquad[(a+b)^2\leqslant2(a^2+b^2)]$$

$$=\frac{\sqrt{2}}{1-\rho}\sqrt{\sum_{t=0}^{T-1}\mathbb{E}\min\left\{\frac{\|d_t\|_2^2}{\sigma_w^2},1\right\}}\sqrt{\mathbb{E}\sum_{t=1}^{T}R(x_t)^2+\mathbb{E}\sum_{t=0}^{T-1}R(f(x_t,t)+w_t)^2}\ .$$

■

Wasserstein contraction implies L2 geometric ergodicity

- We show a relationship between contraction in 2-Wasserstein metric and L2 geometric ergodicity.

- Recall that the 2-Wasserstein distance between μ, ν is:

$$W_2(\mu, \nu) := \left(\inf_{(x,y) \in \Gamma(\mu,\nu)} \mathbb{E}\|x - y\|_2^2 \right)^{1/2}.$$

- $\Gamma(\mu, \nu)$ is the set of **couplings** between μ, ν, i.e., distributions (X, Y) such that the marginal of X is μ and the marginal of Y is ν.

Wasserstein contraction implies L2 geometric ergodicity

- **Claim:** $|\mathbb{E}_\mu\|x\|_2^2 - \mathbb{E}_\nu\|x\|_2^2| \leqslant \sqrt{2}\sqrt{\mathbb{E}_\mu\|x\|_2^2 + \mathbb{E}_\nu\|x\|_2^2}W_2(\mu,\nu)\,.$
- **Proof:** Let $(x,y) \in \Gamma(\mu,\nu)$. We have:

$$
\begin{aligned}
|\mathbb{E}_\mu\|x\|_2^2 - \mathbb{E}_\nu\|x\|_2^2| &= |\mathbb{E}(\|x\|_2 + \|y\|_2)(\|x\|_2 - \|y\|_2)\,| \\
&\leqslant \mathbb{E}(\|x\|_2 + \|y\|_2)\|x-y\|_2 && \text{[reverse triangle]} \\
&\leqslant \sqrt{\mathbb{E}(\|x\|_2 + \|y\|_2)^2}\sqrt{\mathbb{E}\|x-y\|_2^2} && [\mathbb{E}AB \leqslant \sqrt{\mathbb{E}A^2}\sqrt{\mathbb{E}B^2}] \\
&\leqslant \sqrt{2}\sqrt{\mathbb{E}_\mu\|x\|_2^2 + \mathbb{E}_\nu\|x\|_2^2}\sqrt{\mathbb{E}\|x-y\|_2^2} && [(a+b)^2 \leqslant 2(a^2+b^2)]\,.
\end{aligned}
$$

- Now take the infimum over $(x,y) \in \Gamma(\mu,\nu)$ on RHS which yields the claim. ■

Wasserstein contraction implies L2 geometric ergodicity

- Let $P_t(x, A)$ denote the Markov kernel at time t for the nominal dynamics.
- For a measure μ, let μP_t denote the measure $[\mu P_t](A) = \int P_t(x, A)\, \mu(dx)$, and let $P_{s:t} = P_s P_{s+1} \dots P_t$.
- **Claim:** Suppose that $W_2(\mu P_{t:t+k-1}, \nu P_{t:t+k-1}) \leqslant C\gamma^k W_2(\mu, \nu)$ for all $t \in \mathbb{N}_{\geqslant 0}$. Then the L2 geometric ergodicity condition holds.

- **Proof:**

$$
\begin{aligned}
&|\,\mathbb{E}[\|y_{t+k}\|_2^2|\,y_t = x] - \mathbb{E}[\|y_{t+k}\|_2^2|\,y_t = 0]\,| \\
&= |\,\mathbb{E}_{y_{t+k}\sim\delta_x P_{t:t+k-1}}\|y_{t+k}\|_2^2 - \mathbb{E}_{y_{t+k}\sim\delta_0 P_{t:t+k-1}}\|y_{t+k}\|_2^2| \\
&\leqslant \sqrt{2}\sqrt{\mathbb{E}_{y_{t+k}\sim\delta_x P_{t:t+k-1}}\|y_{t+k}\|_2^2 + \mathbb{E}_{y_{t+k}\sim\delta_0 P_{t:t+k-1}}\|y_{t+k}\|_2^2}\,W_2(\delta_x P_{t:t+k-1}, \delta_0 P_{t:t+k-1}) && \text{[Wasserstein inequality]} \\
&\leqslant \sqrt{2}\sqrt{\mathbb{E}[\|y_{t+k}\|_2^2|\,y_t = x] + \mathbb{E}[\|y_{t+k}\|_2^2|\,y_t = 0]}\,C\gamma^k W_2(\delta_x, \delta_0) && \text{[Wasserstein contraction assumption]} \\
&\leqslant \sqrt{2}\sqrt{\mathbb{E}[\|y_{t+k}\|_2^2|\,y_t = x] + \mathbb{E}[\|y_{t+k}\|_2^2|\,y_t = 0]}\,C\gamma^k\|x\|_2 && [W_2(\delta_x, \delta_0) \leqslant \|x\|_2] \\
&= R(x)\gamma^k\,.
\end{aligned}
$$

■

Contraction implies Wasserstein contraction

- Suppose there exists a sequence $\{M_t\}$ of positive definite matrices such that:
  - $\|f(x,t) - f(y,t)\|_{M_{t+1}} \leqslant \gamma \|x - y\|_{M_t}$ for all x, y, t.
  - $\mu I \preccurlyeq M_t \preccurlyeq LI$ for all t.
- **Claim:** $W_2(\mu P_{t:t+k-1}, \nu P_{t:t+k-1}) \leqslant \sqrt{\frac{L}{\mu}} \gamma^k W_2(\mu, \nu)$.

- **Proof:** Let $\zeta \in \Gamma(\mu, \nu)$ be a coupling. Let $w_t, w_{t+1}, \ldots, w_{t+k}$ be iid from $N(0, \sigma_w^2 I)$.
- We construct the following coupled dynamical systems:
$$x_{t+1} = f(x_t, t) + w_t,$$
$$y_{t+1} = f(y_t, t) + w_t,$$
$$(x_t, y_t) \sim \zeta\,.$$
- Define $V_t := \mathbb{E}\|x_t - y_t\|_{M_t}^2$. Observe that:
$$V_{t+1} = \mathbb{E}\|x_{t+1} - y_{t+1}\|_{M_{t+1}}^2 = \mathbb{E}\|f(x_t, t) - f(y_t, t)\|_{M_{t+1}}^2 \leqslant \gamma^2 \mathbb{E}\|x_t - y_t\|_{M_t}^2 = \gamma^2 V_t.$$
- Therefore: $V_{t+k} \leqslant \gamma^{2k} V_t \implies \|x_{t+k} - y_{t+k}\|_2^2 \leqslant \dfrac{L}{\mu}\gamma^{2k}\|x_t - y_t\|_2^2.$
- Now taking the infimum over $\zeta \in \Gamma(\mu, \nu)$ yields that:
$$\|x_{t+k} - y_{t+k}\|_2^2 \leqslant \frac{L}{\mu}\gamma^{2k} W_2^2(\mu, \nu).$$
- The claim now follows since we have constructed a coupling for (x_{t+k}, y_{t+k}). ■

Contraction implies Wasserstein contraction

- Now we consider the more general form of contraction:
  - $\frac{\partial f}{\partial x}(x,t)^T M(f(x,t),t+1)\frac{\partial f}{\partial x}(x,t) \preccurlyeq \gamma^2 M(x,t) \;\; \forall x \in \mathbb{R}^n, t \in \mathbb{N}_{\geqslant 0}$,
  - $\mu I \preccurlyeq M(x,t) \preccurlyeq LI \;\; \forall x \in \mathbb{R}^n, t \in \mathbb{N}_{\geqslant 0}$.
- **Claim:**
  - Define $\Psi := \sup_{x \in \mathbb{R}^n, t \in \mathbb{N}_{\geqslant 0}} \lambda_{\max}(\mathbb{E}_w[M(x,t)^{-1/2}(M(x+w,t)-M(x,t))M(x,t)^{-1/2}]) \vee 0.$
  - Suppose that $\gamma^2(1+\Psi) < 1$.
  - Then: $W_2(\mu P_{t:t+k-1}, \nu P_{t:t+k-1}) \leqslant \sqrt{\frac{L}{\mu}}\left[\gamma\sqrt{1+\Psi}\right]^k W_2(\mu,\nu)$.
- Proof is similar to the state independent metric case, but more technical so we skip it.

Lyapunov stability implies L2 geometric ergodicity

- **Claim:** Suppose there exists a differentiable $Q(x, t)$ such that:
  - $Q(f(x, t), t + 1) \leqslant \gamma Q(x, t)$
  - $\mu \|x\|_2^2 \leqslant Q(x, t) \leqslant \psi \|x\|_2^2,$
  - $x \mapsto \nabla Q(x, t)$ is L-Lipschitz.
- Then we have:
  $$|\,\mathbb{E}[\|y_{t+k}\|_2^2 | \, y_t = x] - \mathbb{E}[\|y_{t+k}\|_2^2 | \, y_t = 0\,| \lesssim \frac{\psi}{\mu}(1 + \|x\|_2^2)\overline{\alpha}^k.$$
- (Unfortunately, $\overline{\alpha} \approx 1 - \exp(-n)$!)

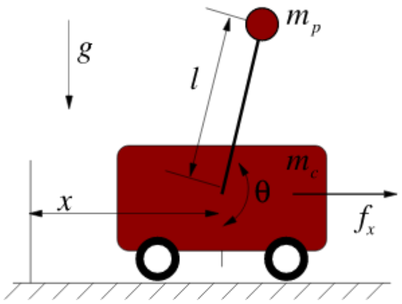
Lyapunov stability implies L2 geometric ergodicity

- Proof is based on ideas used to prove ergodicity of Markov chains [Meyn and Tweedie 1993], [Hairer and Mattingly 2008].
- The two key conditions needed are:
  - (Drift condition): $\mathbb{E}[Q(y_{t+1}, t+1) \,|\, y_t] \leqslant \gamma Q(y_t, t) + R$
  - ("Small-set" minorization): $\inf_{x : Q(x,t) \leqslant R} P_t(x, A) \geqslant \alpha(R) \cdot \nu(A).$
- The issue is that $R = O(n)$ and $\alpha(R) = \Omega(\exp(-R))$, yielding the $(1 - \exp(-n))$ convergence rate.

Open question

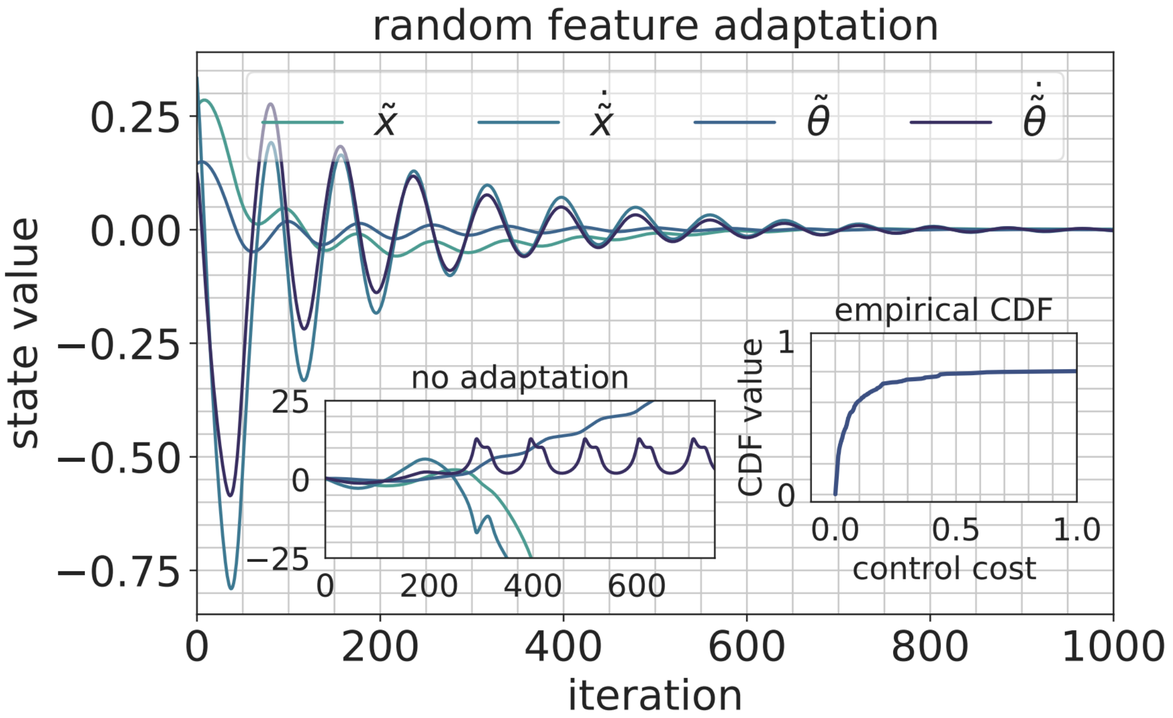
- If we only assume Lyapunov stability, are $1 - \exp(-n)$ rates unavoidable?
    - See https://mathoverflow.net/questions/389180/convergence-rate-for-ergodic-markov-chains-induced-by-stable-dynamical-systems.
- Is there some assumption in between Lyapunov stability and smooth contraction metrics that admit rates that do not depend exponentially in n?

Cartpole



- State is $q = (x, \dot{x}, \theta, \dot{\theta})$.
- We design an LQR controller K using the linearized (RK4 discretized) dynamics about the unstable equilibrium: $q_{\text{eq}} = (0,0,\pi,0)$.
  - We use the **wrong** mass and length parameters.
  - LQR gives us an (incorrect) quadratic Lyapunov function:
    $Q(q) = \frac{1}{2}(q - q_{\text{eq}})^T P (q - q_{\text{eq}})$.
- We then adapt to the model-misspecification using velocity-gradients with $Q(q)$ and with 400 random Fourier features [Rahimi and Recht 2007] as the basis functions.

Cartpole



- Notice the system without adaption is unstable, but adaption manages to correct the model mis-specification.

Semi-contracting system

- We now consider the SDE:

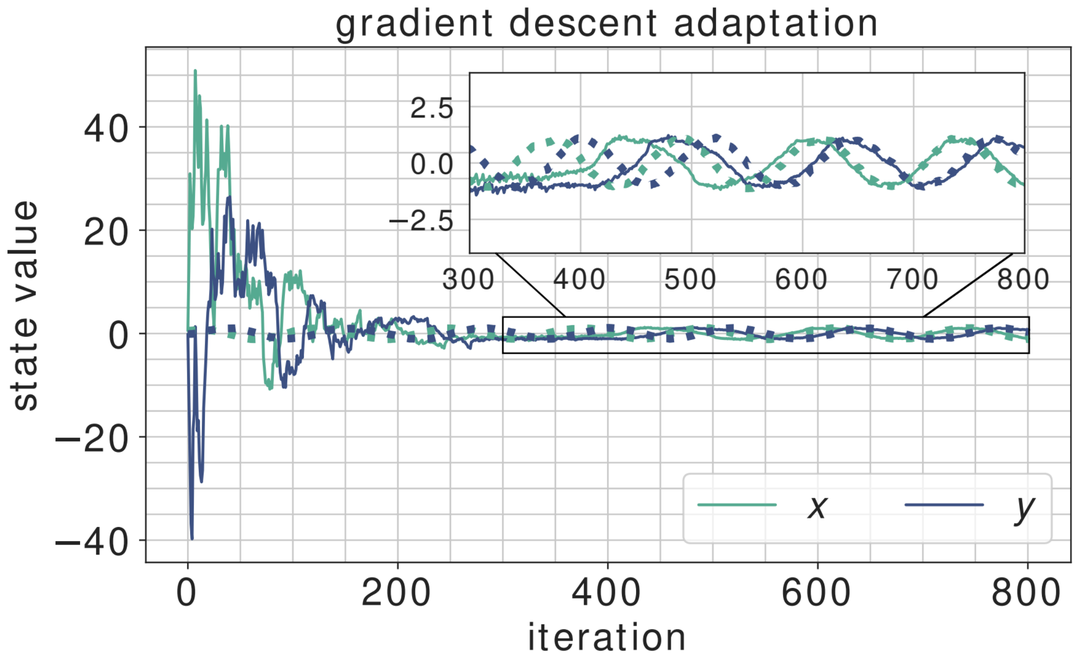
$$dx = \left(-y + \frac{x}{\sqrt{x^2 + y^2}} - x + u_1 - Y_x(x,t)^T \alpha \right) dt + \sigma dw_1,$$

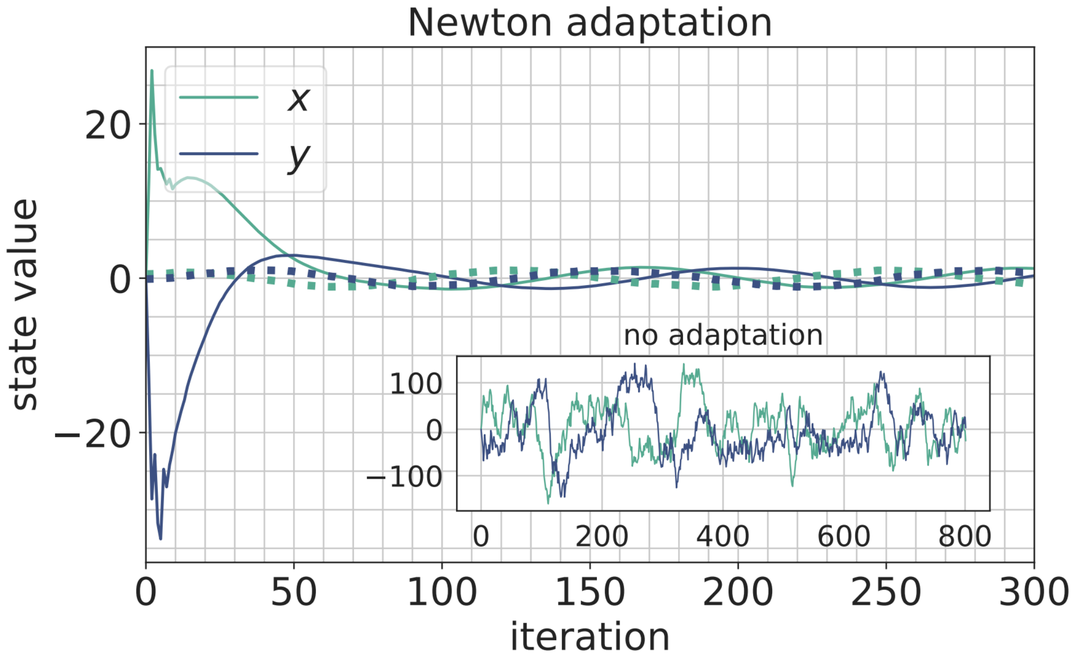
$$dy = \left(x + \frac{y}{\sqrt{x^2 + y^2}} - y + u_2 - Y_y(y,t)^T \alpha \right) dt + \sigma dw_2 \,.$$

- Note the nominal system in polar coordinates is: $\dot{r} = -(r-1)$, $\dot{\theta} = 1$, which is contracting towards the unit angular velocity limit cycle on the unit circle.

Semi-contracting system

- We apply both GD and online Newton to the discretized SDE:





References

- D. Foster and A. Rakhlin. Beyond UCB: Optimal and Efficient Contextual Bandits with Regression Oracles. International Conference on Machine Learning, 2020.
- M. Hairer and J. C. Mattingly. Yet Another Look at Harris' Ergodic Theorem for Markov Chains. Seminar on Stochastic Analysis, Random Fields, and Applications VI, 2011.
- E. Hazan. Introduction to Online Convex Optimization. Foundations and Trends in Optimization, 2016.
- S. M. Kakade, A. Krishnamurthy, K. Lowry, M. Ohnishi, and W. Sun. Information Theoretic Regret Bounds for Online Nonlinear Control. Neural Information Processing Systems, 2020.
- S. Meyn and R. L. Tweedie. Markov Chains and Stochastic Stability, 1993.
- A. Rahimi and B. Recht. Random Features for Large-Scale Kernel Machines. Neural Information Processing Systems, 2007.
- T. Yu, G. Thomas, L. Yu, S. Ermon, J. Zou, S. Levine, C. Finn, and T. Ma. MOPO: Model-based Offline Policy Optimization. Neural Information Processing Systems, 2020.