Proceedings of Machine Learning Research vol xxx:1–6, 2020

Learning Dynamical Systems with Side Information (Proofs)

Amir Ali Ahmadi AAA@PRINCETON.EDU

Bachir El Khadir BKHADIR@PRINCETON.EDU

1. Preliminary: Measure of Similarity Between Vector Fields

Fix a set $\Omega \in \mathbb{R}^n$ and a time horizon T. In order the assess how close two vector fields $f, g : \Omega \to \mathbb{R}^n$ are, we define the following two metrics. The first one compares directly the values taken by f and g on the compact set Ω:

$$\|f - g\|_\Omega := \max_{x \in \Omega} \|f(x) - g(x)\|$$

The second metric of interest, which is more dynamics-focused, measures how different the trajectories of f and g starting from an invariant set Ω are after some time $T > 0$, i.e.

$$d_{\Omega,T}(f - g) := \sup_{x_0 \in \Omega, t \in [0,T]} \max\{\|x(t, x_0) - y(t, x_0)\|, \|\dot{x}(t, x_0) - \dot{y}(t, x_0)\|$$

where $x(t, x_0)$ (resp. $y(t, x_0)$) is the trajectory starting from $x_0 \in \Omega$ and following the dynamics of f (resp. g).

The metric $d_{\Omega,T}(\cdot)$ is "finer" than $\|\cdot\|_\Omega$. Indeed, by noting that $\dot{x}(0, x_0) = f(x_0)$ and $\dot{y}(0, x_0) = g(x_0)$, we obtain that $d_{\Omega,T}(f - g) \leq \|f - g\|_\Omega$. The next proposition shows that for vector fields that leave the set Ω invariant, these two metric become equivalent.

Proposition 1 *For any positive scalar $L > 0$, there exists a constant $C_{\Omega,T,L}$ such that*

$$\|f - g\|_\Omega \leq d_{\Omega,T}(f - g) \leq C_{\Omega,T,L} \|f - g\|_\Omega$$

for every two L-Lipchitz vector fields f and g that leave the set Ω invariant.

To present the proof of this proposition, we need to recall the classical lemma of *Gronwall*.

Lemma 2 (Gronwall) *Let $I = [a, b]$ denote a non-empty interval on the real line. Let α, β, and u be continuous, real-valued functions defined on I and satisfying*

$$u(t) \leq \alpha(t) + \int_a^t \beta(s) u(s) \, \mathrm{d}s \quad \forall t \in I.$$

If α is nondecreasing and β is nonnegative, then

$$u(t) \leq \alpha(t) \exp\left(\int_a^t \beta(s) \, \mathrm{d}s\right) \quad t \in I.$$

Proof (Proof of Proposition 1) Consider two trajectories $x(t)$ and $y(t)$ both starting from $x_0 \in \Omega$ and following f and g respectively. The proof is divided into two parts.

For the first part of the proof, we will bound $\|y(t) - x(t)\|$. By definition of y and x, for every $t \in [0, T]$ we have that

$$\begin{aligned} x(t) - y(t) &= \int_0^t f(x(s)) - g(y(s))\mathrm{d}s \\ &= \int_0^t f(x(s)) - g(x(s))\,\mathrm{d}s + \int_0^t g(x(s)) - g(y(s))\,\mathrm{d}s. \end{aligned}$$

Using the triangular inequality, we get

$$\|y(t) - x(t)\| \leq \int_0^t \|f(x(s)) - g(x(s))\|\,\mathrm{d}s + \int_0^t \|g(x(s)) - g(y(s))\|\,\mathrm{d}s. \tag{1}$$

Since f leaves Ω invariant, we know that for all $s \in [0, t]$, $x(s) \in \Omega$ and therefore $\|f(x(s)) - g(x(s))\| \leq \|f - g\|_\Omega$. Furthermore, because the function g is L-Lipchiz, $\|g(x(s)) - g(y(s))\| \leq L\|x(s) - y(s)\|$. From (1) we conclude therefore that

$$\|y(t) - x(t)\| \leq t\|f - g\|_\Omega + L\int_0^t \|x(s) - y(s)\|\,\mathrm{d}s,$$

and by Gronwall's lemma,

$$\|y(t) - x(t)\| \leq t\exp(Lt)\|f - g\|_\Omega \;\forall t \in [0, T].$$

For the second part of the proof, we will bound $\|\frac{\partial}{\partial t}y(t) - \frac{\partial}{\partial t}x(t)\| = \|f(x(s)) - g(y(s))\|$. note that since f and g leave Ω invariant, we know that for all $s \in [0, t]$, $x(s), y(s) \in \Omega$. By triangular inequality again, we know that $\|f(x(s)) - g(y(s))\| \leq \|f(x(s)) - g(x(s))\| + \|g(x(s)) - g(y(s))\|$. Using the fact that g is L-Lipchiz on Ω, $\|g(x(s)) - g(y(s))\| \leq L\|y(t) - x(t)\|$. Therefore $\|f(x(s)) - g(y(s))\| \leq \|f - g\|_\Omega + L\|y(t) - x(t)\|$.

Putting the first and second part of the proof together, we have

$$\|y(t) - x(t)\| \leq T\exp(LT)\|f - g\|_\Omega$$

and

$$\begin{aligned} \|\frac{\partial}{\partial t}y(t) - \frac{\partial}{\partial t}x(t)\| &\leq \|f - g\|_\Omega + L\|y(t) - x(t)\| \\ &\leq \|f - g\|_\Omega + Lt\exp(Lt)\|f - g\|_\Omega. \end{aligned}$$

Therefore

$$d_{\Omega,T}(f - g) \leq C(\Omega, T, L)\|f - g\|_\Omega$$

with $C(\Omega, T, L) = \max\{T\exp(LT), 1 + LT\exp(LT)\}$. ■

2. Proof of Theorem 1

Fix a compact set $\Omega \subset R^n$, a time horizon $T > 0$, and a desired accuracy $\varepsilon > 0$. Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a continuous differentiable vector field that satisfies any one of the following constraints:

(i) equilibria at a given finite set of points,

(ii) symmetry.

(iii) nonnegativity,

(iv) directional monotonicity,

(v) invariance of a full-dimensional, star-shaped basic semialgebraic set,

(vi) gradient or Hamiltonian structure.

In this section, we prove that there exists a polynomial vector field $p : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$\|f - g\|_\Omega \leq \varepsilon,$$

(by proposition 1, this shows in particular that trajectories of f and p can be made arbitrarily close) and p satisfies the same side information as f. Before we give a case-by-case proof depending on which side information f satisfies, we present the following universal-approximation result that will be invoked frequently.

Theorem 3 *(Weirstrass approximation theorem) If $f : \Omega \to \mathbb{R}^n$ is continuously-differentiable, then for any $\varepsilon > 0$, there exists a polynomial $p : \mathbb{R}^n \to \mathbb{R}^n$ such that*

$$\|f(x) - p(x)\| \leq \varepsilon \quad \forall x \in \Omega.$$

2.1. If f satisfies the equilibrium conditions in (i)

Suppose f satisfies

$$f(x_i) = \alpha_i \quad i = 1, \dots, m. \tag{2}$$

Let p be a polynomial vector field uniformly approximating f as in Theorem 3. Let $v_i = (\alpha_i - (f(x_i) - p(x_i)))$, and note that $\|v\| \leq \varepsilon$.

The proposition below, shows that there exists a polynomial q such that

$$(p + q)(x_i) = \alpha_i \quad i = 1, \dots, m,$$

and $\|q\|_\Omega \leq C\varepsilon$, where C is a constant depending only on the x_i. The last inequality implies in particular that

$$\|f - (p + q)\|_\Omega \leq \varepsilon(1 + C).$$

Proposition 4 *For any positive integers d and n, for any d points $x^1, \dots, x^d$ in a compact set Ω, there exists a constant $C(x_1, \dots, x_d)$ such that the following holds:*
For any vector $v \in \mathbb{R}^d$, there exists a polynomial $p \in \mathbb{R}_d[x]$ such that

$$p(x_i) = v_i, \quad i = 1, \dots, d \text{ and } \max_{x \in \Omega} |p(x)| \leq C(x_1, \dots, x_d)\|v\|_2$$

Lemma 5 (Proposition 4.3 in Comon et al.) ***(Multivariate Polynomial Interpolation)*** *If $x^1, \ldots, x^d$ are d different points of $\mathbb{R}^n$, then the vectors $m_d(x^1), \ldots, m_d(x^d)$ are linearly independent. (Here $m_d(x^i)$ is the vector of monomials in x_i up to degree d.)*

Proof [Proof of Proposition 4] By Lemma 5, the system of linear equations

$$p(x_i) = v_i \quad i = 1, \ldots, n \tag{3}$$

in the variable $p \in \mathbb{R}_d[x]$ are independent. If we identify p by its coefficients in the monomial basis, then we can rewrite the system in Equation (3) as

$$Ap = v,$$

where $A = (m_d(x^1), \ldots, m_d(x^d))^T$ is a matrix depending only on $x_1, \ldots, x_d$ that has independent rows. The proof follows by taking $C(x_1, \ldots, x_d)$ to be the operator norm of the matrix $M := (A^T A)^{-1} A^T$ and $p = Mv$. ■

2.2. If f satisfies the symmetry in (ii)

Suppose f satisfies a symmetry

$$B^{-1} f(Ax) = f(x). \tag{4}$$

For a vector field h, define

$$[\psi(h)](x) := \frac{B^{-1} h(Ax) + h(x)}{2}$$

Note that ψ is a linear function and that the equality in Equation (4) translates to $\psi f = f$. Moreover, ψ is an involution (i.e., $\psi^2 = \psi$) and a contraction (i.e. $\|\psi(h)\|_\Omega \leq \|h\|_\Omega$.)

Let p be a polynomial vector field uniformly approximating f as in Theorem 3, i.e., $\|p - f\|_\Omega \leq \varepsilon$, and let $q = \psi p$. Then, $\psi(q) = q$, and $\|q - f\|_\Omega \leq \|q - f\|_\Omega \leq \varepsilon$.

2.3. If f satisfies the nonnegativity condition in (iii)

Suppose f satisfies $f_i(x) \trianglerighteq_i 0 \; \forall x \in B_i$ for $i, j \in \{1, \ldots, n\}$. Let $f_\varepsilon(x)$ the vector field defined component wise by

$$f_{\varepsilon,j}(x) = f_j(x) + \varepsilon \alpha_i,$$

where $\alpha_i = 1$ if $\trianglerighteq_i = \geq$, and $\alpha_i = -1$ otherwise. Approximating f_ε uniformly by a polynomial p gives the desired construction.

2.4. If f satisfies the monotonicity condition in (iv)

Suppose f satisfies $\frac{\partial f_i}{x_j}(x) \trianglerighteq_{i,j} 0 \; \forall x \in B_{i,j}$ for $i, j \in \{1, \ldots, n\}$. Let $f_\varepsilon(x)$ the vector field defined component wise by

$$f_{\varepsilon,j}(x) = f_j(x) + \varepsilon \alpha_{ij} x_i,$$

where $\alpha_{ij} = 1$ if $\trianglerighteq_{i,j} = \geq$, and $\alpha_{ij} = -1$ otherwise. Approximating f_ε uniformly by a polynomial p gives the desired construction.

2.5. If f satisfies invariance in (v)

Let B be the set defined by the inequalities $h_1(x) \geq 0, \ldots, h_m(x) \geq 0$, where each of the h_i is concave, and suppose there exists a point $x_0 \in B$ such that $h_i(x) > 0$ for all i. Suppose f leaves B invariant, i.e.,

$$h_j(x) \geq 0 \; \forall j \text{ and } h_i(x) = 0 \implies \langle f(x), \nabla h_i(x) \rangle \geq 0 \quad \forall x \in \mathbb{R}^n, i = 1, \ldots, m. \tag{5}$$

Let $f_\varepsilon(x) = f(x) - \varepsilon(x - x_0)$, and notice that $\|f_\varepsilon - f\| \leq |\Omega|\varepsilon$. (Here $|\Omega| := \sup_{x,y \in \Omega} \|x - y\|$). Moreover, if $h_j(x) \geq 0 \; \forall j$ and $h_i(x) = 0$, then

$$\langle f_\varepsilon(x), \nabla h_i(x) \rangle = \langle f(x), \nabla h_i(x) \rangle + \varepsilon \langle x - x_0, \nabla h_i(x) \rangle \geq \varepsilon h_i(x_0).$$

Let p be a polynomial vector field such that $\|p - f_\varepsilon\|_\Omega \leq \varepsilon$, then by continuity, p satisfies Equation (5). Moreover,

$$\|f - p\| \leq \|f - f_\varepsilon\| + \|f_\varepsilon - p\| \leq \varepsilon(|\Omega| + 1).$$

2.6. If f satisfies the gradient/Hamiltonian condition in (vi)

Suppose f is a gradient, i.e., $f(x) = -\nabla V(x)$ for some scalar-valued function V. Let W be polynomial approximation of V, i.e.

$$\|V(x) - W(x)\| \leq \varepsilon \quad \text{and} \quad \|\nabla V - \nabla W\| \leq \varepsilon.$$

Then, the polynomial vector field $p := -\nabla W$ is a gradient and approximates f uniformly on Ω.

3. Proof of Theorem 2

Let C denote the set of continuously differentiable vector fields. The table below shows for every side information S, there exists a continuous functional

$$L_S : C \to \mathbb{R}$$

that is continuous with respect to the $\|\cdot\|_\Omega$ norm, and such that

$$\forall f \in C, \; f \text{ satisfies } S \iff L_S(f) = 0,$$

and

$$\forall f \in C, \; f \; \delta\text{-satisfies } S \iff L_S(f) \leq \delta.$$

Side Information S	Linear functional L_S
$\mathbf{Interp}(\{x_i, y_i)_{i=1}^m\}$	$L_S(f) := \sum_{i=1}^m \|f(x_i) - y_i\|$
$\mathbf{Sym}(A, B)$	$L_S(f) := \|f(Ax) - Bf(x)\|$
$\mathbf{Pos}(\{\trianglerighteq_i, B_i\}_{i=1}^n)$	$L_S(f) := \min_i \min_{B_i} f$
$\mathbf{Mon}(\{\trianglerighteq_{i,j}, B_{i,j}\}_{i,j=1}^n)$	$L_S(f) := \min_{i,j} \min_{B_{ij}} \frac{\partial f_j}{\partial x_j}$
$\mathbf{Inv}(B)$	$L_S(f) := \min_i \min_{x \in B, h_i(x)=0} \langle f(x), \nabla h_i(x) \rangle$
$\mathbf{Grad}$	$L_S(f) := \min_{V:\mathbb{R}^n \to \mathbb{R}} \|f - \nabla V\|_\Omega$

Now let f be a vector field in C satisfying a side information S. Then, for all $\delta > 0$, there exists $\varepsilon > 0$ such that

$$\forall p \in C, \quad \|p - f\| \leq \varepsilon \implies L_S(p) - L_S(f) \leq \delta.$$

By Theorem 3, there exists a polynomial vector field p such that $\|p - f\|_\Omega \leq \varepsilon$. In particular, p δ-satisfies the side information S. By Putinar's approximation theorem, the following polynomial inequality admits an sos certificate.

$$L_S(p) < 2\delta.$$

References

Pierre Comon, Gene Golub, Lek-Heng Lim, and Bernard Mourrain. Symmetric tensors and symmetric tensor rank. URL http://arxiv.org/abs/0802.1681.