Control of Diffeomorphisms

Andrei Agrachev
SISSA, Trieste

Joint work with Andrey Sarychev (Florence) motivated by the deep learning of artificial neural networks treated as an interpolation problem.
Maps interpolation:

Given a class of “good” maps $\mathcal{F}$ we look for $F \in \mathcal{F}$ that is close to $\Phi$ at the marked points.
In neural networks, the class of “good maps” $\mathcal{F}$ consists of the “input – output” transformations of discrete time control systems of the form:

$$x(t + 1) = \bar{\sigma}(U(t)x(t) + v(t)), \quad x \in \mathbb{R}^n, \ t = 0, 1, \ldots, k,$$

where the matrix $U$ and vector $v$ are control parameters,

$$\bar{\sigma}(x_1, \ldots, x_n) = (\sigma(x_1), \ldots, \sigma(x_n)),$$

$\sigma$ is a monotone nonlinear function with a bounded derivative, and $F : x(0) \mapsto x(k)$. Some samples:

$$\sigma(s) = \max\{0, s\}, \quad \sigma(s) = \frac{1}{1 + e^{-s}}, \quad \sigma(s) = \int_{-\infty}^{s} e^{-\tau^2} \, d\tau.$$
Continuous time:

\[ \dot{x} = f(x, u(t)), \quad F_{u(\cdot)} : x(0) \mapsto x(1), \quad \mathcal{F} = \{ F_{u(\cdot)} \}. \]

The goal is to uniformly approximate given transformation \( \Phi : \mathbb{R}^n \to \mathbb{R}^n \) on a compact \( K \subset \mathbb{R}^n \).

**Example:** \( u = (v, w), \)

\[ f(x, u) = (v_1 e^{-|x|^2} + w_1, \ldots, v_n e^{-|x|^2} + w_n). \]

**Theorem 1.** Let \( \Phi : \mathbb{R}^n \to \mathbb{R}^n \) be an isotopic to the identity diffeomorphism, \( K \subset \mathbb{R}^n \), and \( \varepsilon > 0 \). Then there exists \( u(\cdot) \) such that

\[ \sup_{x \in K} |F_{u(\cdot)}(x) - \Phi(x)| < \varepsilon. \]
General result:

Let $M$ be a complete Riemannian manifold, $f_1, \ldots, f_r$ bounded smooth vector fields and

$$\text{Lie}\{f_1, \ldots, f_r\} = \text{span} \left\{ [f_{i_1}, \ldots, f_{i_k}] \cdots : k \in \mathbb{Z}_+ \right\}.$$

We consider a system:

$$\dot{x} = u_1 f_1(x) + \cdots + u_r f_r(x), \quad x \in M, \; u_i \in \mathbb{R};$$

$F_u : x(0) \mapsto x(1)$, where $u = (u_1(\cdot), \ldots, u_r(\cdot))$.

**Theorem 2** (Rashevskij–Chow). If $\text{Lie}\{f_1, \ldots, f_r\}|_q = T_q M$, $\forall q \in M$, then, for any $q_0, q_1 \in M$, $\exists u$ such that $F_u(q_0) = q_1$. 
Corollary 1. Let \( \dim M > 1 \) and \( \text{Lie}\{f_1, \ldots, f_r\} \) is everywhere dense in \( \text{Vec}(M) \) in the \( C_0 \)-topology. Then for any finite families of points \( x_\alpha, y_\alpha \in M, \alpha \in A, \# A < \infty \), there exists \( u \) such that \( F_u(x_\alpha) = y_\alpha, \forall \alpha \in A \).

Let \( \ell > 0, K \subset M \); we set:

\[
\text{Lie}^\ell_K\{f_1, \ldots, f_r\} = \left\{ g \in \text{Lie}\{f_1, \ldots, f_r\} : \sup_{x \in K} (|g(x)| + \|\nabla_x g\|) < \ell \right\}.
\]

Definition 1. We say that \{\( f_1, \ldots, f_r \)\} has property (A) if for any smooth vector field \( X \) and any \( K \subset M \) there exists \( \ell > 0 \) such that

\[
\inf \left\{ \sup_{x \in K} |g(x) - X(x)| : g \in \text{Lie}^\ell_K\{f_1, \ldots, f_r\} \right\} = 0.
\]
Theorem 3. If \( \{f_1, \ldots, f_r\} \) has property (A), then for any isotopic to the identity diffeomorphism \( \Phi : \mathbb{R}^n \rightarrow \mathbb{R}^n \), \( K \subseteq \mathbb{R}^n \), and \( \varepsilon > 0 \), there exists a control function \( u \) such that \( \sup_{x \in K} \delta(F_u(x), \Phi(x)) < \varepsilon \), where \( \delta(\cdot, \cdot) \) is the Riemannian distance in \( M \).

Examples:

\( M = \mathbb{R}^n \); the family of vector fields:

\[
\frac{\partial}{\partial x_i}, \quad e^{-|x|^2} \frac{\partial}{\partial x_i}, \quad i = 1, \ldots, n,
\]

has property (A). The iterated commutators of these vector fields produce Hermit polynomials.
$M = \mathbb{T}^n = \{(\theta_1, \ldots, \theta_n) : \theta_i \in \mathbb{R}/2\pi\mathbb{Z}\}$. The family of vector fields:

$$\frac{\partial}{\partial \theta_i}, \sin(\theta_i) \frac{\partial}{\partial \theta_i}, \sin(2\theta_i) \frac{\partial}{\partial \theta_i}, \sum_{j=1}^n \sin(\theta_j) \frac{\partial}{\partial \theta_i}, \quad i = 1, \ldots, n,$$

has property (A).

$M = S^2 = \{x \in \mathbb{R}^3 : |x| = 1\}$. Given a smooth function $a : \mathbb{R}^3 \to \mathbb{R}$, we define spherical gradient field $\nabla^s a$ and Hamiltonian field $\vec{a}$ by the formulas:

$$\nabla^s a = \nabla x a - \langle x, \nabla x a \rangle x, \quad \vec{a}(x) = x \times \nabla x a.$$

Let linear functions $e_1, e_2, e_3$ form a basis of $\mathbb{R}^3^*$, $p : \mathbb{R}^3 \to \mathbb{R}$ be a quadratic harmonic polynomial and $q : \mathbb{R}^3 \to \mathbb{R}$ be a cubic harmonic polynomial. The family of vector fields on $S^2$:

$$\nabla^s p, \quad \vec{p}, \quad \vec{q}, \quad \nabla^s e_i, \quad \vec{e}_i, \quad i = 1, 2, 3,$$

has property (A).
Sketch of proof.

Together with the system $\dot{x} = \sum_i u_i f_i(x)$ and generated by this system diffeomorphisms $F^t_{u} : x(0) \mapsto x(t), ~ t \in [0, 1]$, we consider the extended system:

$$\dot{y} = \sum_i u_i f_i(y) + \sum_{i<j} u_{ij} [f_i, f_j](y), ~ y \in M, ~ u_i, u_{ij} \in \mathbb{R},$$

and diffeomorphisms $G^t_v : y(0) \mapsto y(t)$, where $v = \{u_i(\cdot), u_{ij}(\cdot)\}$.

**Theorem 4.** For any extended control $v$, any $\varepsilon > 0$, $k \geq 0$, and $K \subset M$ there exists an appropriate control $u = \{u_i(\cdot)\}$ such that $\|G^t_v - F^t_u\|_{k,K} < \varepsilon$ for any $t \in [0, 1]$, where $\| \cdot \|_{k,K}$ is a $C^k$ norm for maps defined on $K$. 


Lemma 1. Let $X_t$, $t \in [0, 1]$, be a time-dependent vector field, $w : [0, 1] \to \mathbb{R}$ a smooth function, and $\varepsilon > 0$. We set $u_\varepsilon(t) = 2\sin(t/\varepsilon^2)w(t)$ and consider systems

$$\dot{x} = X_t(x) + 1/\varepsilon \sin(t/\varepsilon^2)g(x) + \varepsilon u_\varepsilon(t)f(x).$$

Then the flow generated by $(\varepsilon)$ converges uniformly to the flow generated by the system

$$\dot{x} = X_t(x) + w(t)[f, g](x)$$

as $\varepsilon \to 0$, in any norm $\| \cdot \|_{r,K}$, $r \geq 0, K \in M$.

The proof is based on a factorization of system $\varepsilon u_\varepsilon(t)f(x)$ is taken out.
Chronological notations: let \( f \in \text{Vec}(M) \), we set \( e^{tf} : x(0) \mapsto x(t) \) in virtue of \( \dot{x} = f(x) \). Then:

\[
e^{tf} : M \to M, \quad e^{tf}_* : \text{Vec}(M) \to \text{Vec}(M).
\]

Moreover, \( e^{tf}_* = e^{-t \text{ad} f} \), where \( (\text{ad} f)g = [f, g] \).

Given a time-varying vector field \( f_\tau \), we set \( \exp^t \int_0^t f_\tau d\tau : x(0) \mapsto x(t) \), in virtue of \( \dot{x} = f_\tau(x) \). If \( [f_\tau, f_s] = 0 \) for all \( 0 \leq \tau, s \leq 1 \), then

\[
\exp^t \int_0^t f_\tau d\tau = e^{\int_0^t f_\tau d\tau}.
\]

Variations formula:

\[
\exp^t \int_0^t f_\tau + g_\tau d\tau = \exp^t \int_0^t f_\tau d\tau \circ \exp^t \int_0^t \exp^\tau \int_0^\tau \text{ad} f_s ds g_\tau d\tau.
\]
Let:

\[ f_\tau = \varepsilon \dot{u}_\varepsilon(\tau) f(x), \quad g_\tau = X_\tau(x) + 1/\varepsilon \sin(\tau/\varepsilon^2)g(x). \]

We have:

\[
\exp \int_0^t X_\tau + 1/\varepsilon \sin(\tau/\varepsilon^2)g + \varepsilon \dot{u}_\varepsilon(\tau) f \, d\tau = \\
\exp \int_0^t e^{\varepsilon u_\varepsilon(\tau)} \text{ad}_f(X_\tau + 1/\varepsilon \sin(\tau/\varepsilon^2)g) \, d\tau \\
= (I + O(\varepsilon)) \circ \exp \int_0^t X_\tau + w(\tau)[f, g] + O(\varepsilon) \, d\tau.
\]
We have a sample family of points $x_{\alpha}$, $\alpha \in A$, and we wish $F_u(x_{\alpha})$ to be close to $y_{\alpha}$. A functional to minimize is:

$$\varphi(u) = \sum_{\alpha \in A} |F_u(x_{\alpha}) - y_{\alpha}|^2 + \nu \int_0^1 |u(t)|^2 \, dt.$$ 

We have:

$$\frac{\partial \varphi}{\partial u_i}(t) = \sum_{\alpha} \langle f_i, \nabla |F_u^{t,1} - y_{\alpha}|^2 \rangle |F_u^{0,t}(x_{\alpha}) + 2\nu u(t),$$

where $F_u^{t,s} : x(\tau) \mapsto x(s)$ in virtue of $\dot{x} = \sum_i u_i f_i(x)$; in particular, $F_u = F_u^{0,1}$. 

13
Simulations (Alessandro Scagliotti, SISSA). Gradient descent for the discretized system, $\nu = 0$. 

- Transformation
- Approximation
Test
Azat Miftakhov