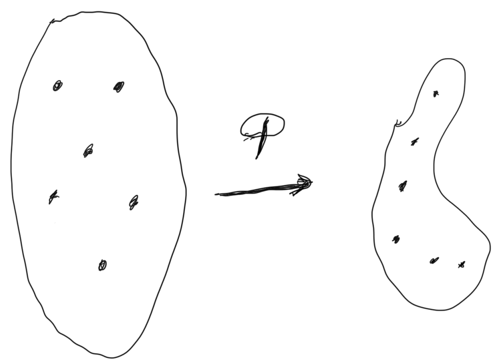
Control of Diffeomorphisms

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 Φ 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,

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

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

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

Continous 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, $f_1, \dots, f_r$ bounded smooth vector fields and

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

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 $u = (u_1(\cdot), \dots, u_r(\cdot))$.

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

Let $\ell > 0,\ K \Subset M$; we set:

$$\mathrm{Lie}^{\ell}_{K}\{f_1, \ldots, f_r\} = \left\{ g \in \mathrm{Lie}\{f_1, \ldots, f_r\} : \sup_{x \in K} (|g(x)| + \|\nabla_x g\|) < \ell \right\}.$$

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*

$$\inf \left\{ \sup_{x \in K} |g(x) - X(x)| : g \in \mathrm{Lie}^{\ell}_{K}\{f_1, \ldots, f_r\} \right\} = 0.$$

Theorem 3. *If* $\{f_1, \ldots, f_r\}$ *has property (A), then for any isotopic to the identity diffeomorphism* $\Phi : M \to M$*,* $K \Subset M$*, and* $\varepsilon > 0$*, there exists a control function* u *such that* $\sup_{x \in K} \delta\left(F_u(x), \Phi(x)\right) < \varepsilon$*, 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:

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

has property (A).

$M = \mathbb{S}^2 = \{x \in \mathbb{R}^3 : |x| = 1\}$. 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:

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

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 $\mathbb{S}^2$:

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

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 $F_u^t : x(0) \mapsto x(t),\ t \in [0,1]$, we consider the extended system:

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

and diffeomorphisms $G_v^t : y(0) \mapsto y(t)$, 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 $\|G_v^t - F_u^t\|_{k,K} < \varepsilon$ 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*

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

Then the flow generated by (ε) *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* $\|\cdot\|_{r,K}$, $r \geq 0, K \Subset M$.

The proof is based on a factorization of system (ε): the flow generated by $\varepsilon \dot{u}_\varepsilon(t) f(x)$ is taken out.

Chronological notations: let $f \in \mathrm{Vec}(M)$, we set $e^{tf} : x(0) \mapsto x(t)$ in virtue of $\dot{x} = f(x)$. Then:

$$e^{tf} : M \to M, \quad e_*^{tf} : \mathrm{Vec}(M) \to \mathrm{Vec}(M).$$

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

Given a time-varying vector field f_τ, we set $\overrightarrow{exp}\int\limits_0^t f_\tau d\tau : x(0) \mapsto x(t)$, in virtue of $\dot{x} = f_\tau(x)$. If $[f_\tau, f_s] = 0$ for all $0 \leq \tau, s \leq 1$, then $\overrightarrow{exp}\int\limits_0^t f_\tau d\tau = e^{\int_0^t f_\tau d\tau}$.

Variations formula:

$$\overrightarrow{exp}\int\limits_0^t f_\tau + g_\tau d\tau = \overrightarrow{exp}\int\limits_0^t f_\tau d\tau \circ \overrightarrow{exp}\int\limits_0^t (\overrightarrow{exp}\int\limits_0^\tau \mathrm{ad} f_s ds) g_\tau d\tau.$$

Let:

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

We have:

$$\overrightarrow{\exp} \int_0^t X_\tau + 1/\varepsilon \sin(\tau/\varepsilon^2) g + \varepsilon \dot{u}_\varepsilon(\tau) f \, d\tau =$$

$$e^{\varepsilon u_\varepsilon(\tau) f} \circ \overrightarrow{\exp} \int_0^t e^{\varepsilon u_\varepsilon(\tau) \mathrm{ad} f} (X_\tau + 1/\varepsilon \sin(\tau/\varepsilon^2) g) \, d\tau$$

$$= (I + O(\varepsilon)) \circ \overrightarrow{\exp} \int_0^t X_\tau + w(\tau)[f, g] + O(\varepsilon) \, d\tau.$$

We have a sample family of points $x_\alpha,\ \alpha \in \mathcal{A}$, and we wish $F_u(x_\alpha)$ to be close to y_α. A functional to minimize is:

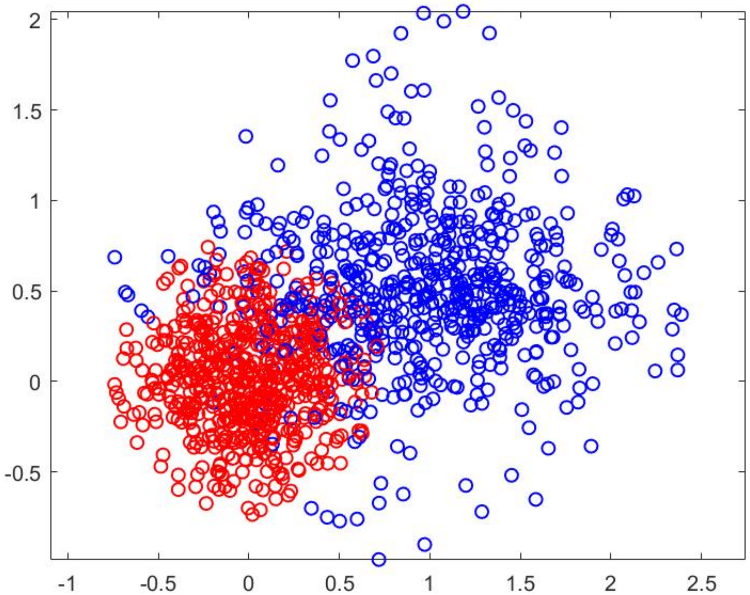
$$\varphi(u) = \sum_{\alpha\in\mathcal{A}} |F_u(x_\alpha) - y_\alpha|^2 + \nu \int_0^1 |u(t)|^2 \, dt.$$

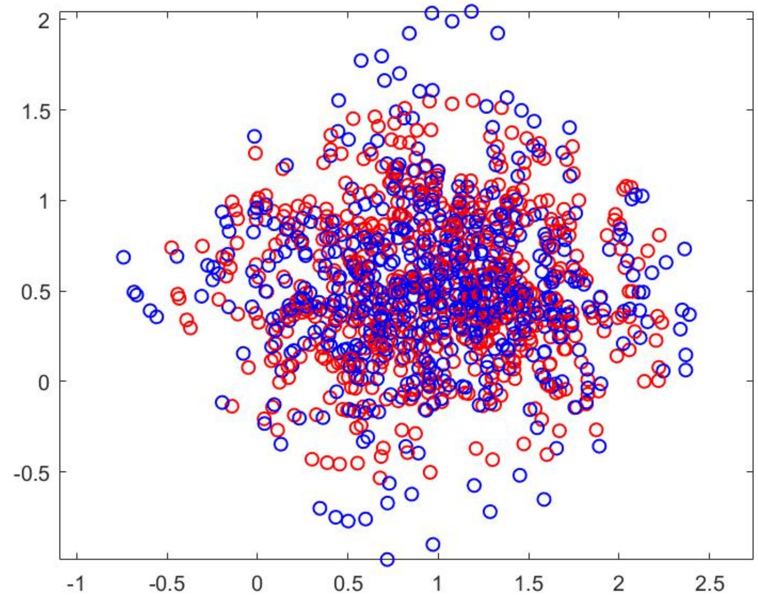
We have:

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

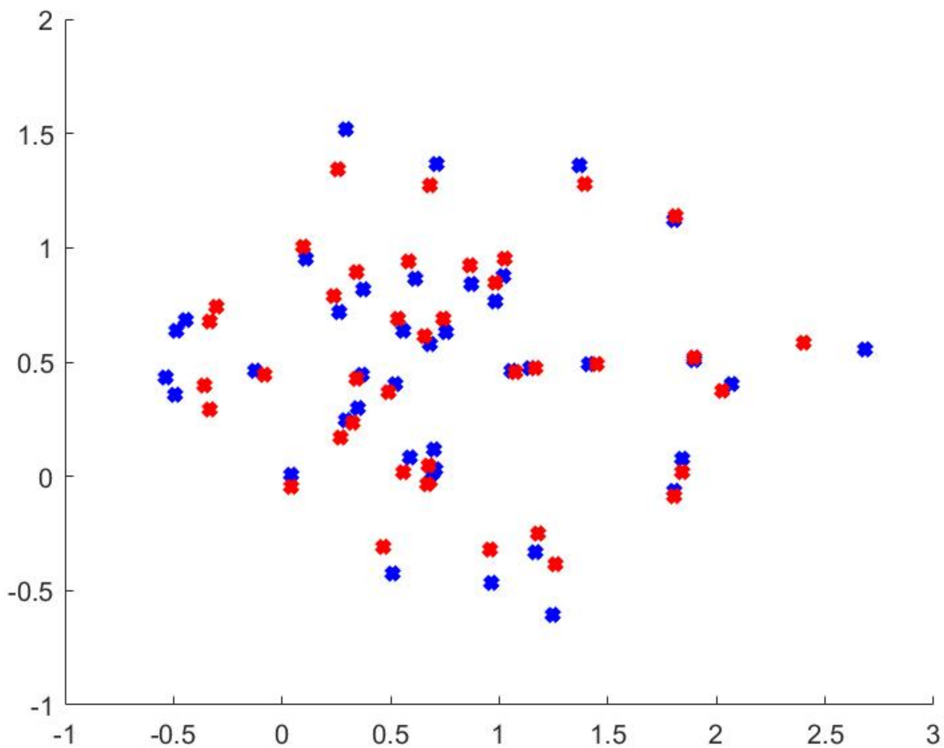
where $F_u^{\tau,s} : x(\tau) \mapsto x(s)$ in virtue of $\dot{x} = \sum_i u_i f_i(x)$; in particular, $F_u = F_u^{0,1}$.

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





. Transformation Approximation



. Test

Azat Miftakhov