Linear and Nonlinear Schemes for Forward Model Reduction and Inverse Problems

Olga Mula (TU Eindhoven)

Cemracs 2023

2023-07-17

Agenda

1. Elements of approximation theory
   - Linear and Nonlinear Approximation
   - Nonlinear approximation with Neural Networks
2. Forward Problem: Reduced Order Modelling of parametrized PDEs
   - Linear MOR
   - Nonlinear MOR
   - Role of geometry
3. Inverse Problems
   - Optimal linear and nonlinear algorithms for State Estimation
   - Role of Geometry
4. Hands-on session with Agustin Somacal

Part II.1

Reduced Order Modelling of Parametrized PDEs

Motivations and Linear Approximation

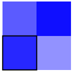
Example 1: Elliptic PDEs

Elliptic PDE:

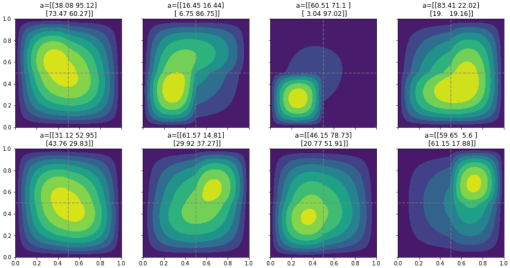
$$
\begin{aligned}
-\nabla \cdot (a(x)\nabla u(x)) + \sigma(x)u(x) &= f(x), \qquad && \forall x \in \Omega \\
u(x) &= 0, && \forall x \in \partial\Omega.
\end{aligned}
$$

Solution space: $u(\theta) \in V = H_0^1(\Omega)$ with $\Omega \subseteq \mathbb{R}^d$.

Parameters: $\theta = \{a, \sigma\} \subset L^\infty(\Omega) \times L^\infty(\Omega)$ or simply $\theta \in \mathbb{R}^p$.



$a(x) = \sum_{i=1}^4 a_i \mathbb{1}_{\Omega_i}(x)$

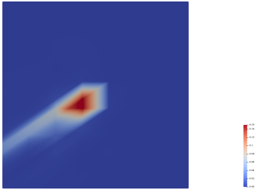


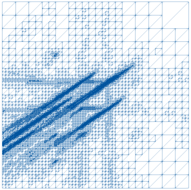
Pure transport PDEs:

$$
\begin{aligned}
\partial_t u(t,x) + a(t,x)\cdot\nabla_x u(t,x) &= f(t,x), && \forall (t,x)\in \mathbb{T}\times\Omega\\
u(t,x) &= g(t,x), && \forall (t,x)\in \mathbb{T}\times\partial\Omega_-\\
u(t=0,x) &= u_0(x), && \forall x\in\Omega.
\end{aligned}
$$

Solution space: $u(\theta)\in V = L^1((0,T),\mathbb{R}^d)$.

Parameters: $\theta = a \in \Theta$





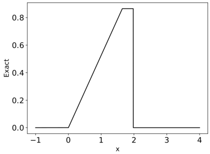
Conservation laws:

$$\partial_t u + f(u;\theta) = 0$$
$$u(t=0) = u_0$$

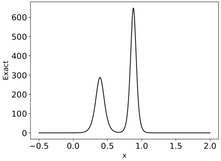
Solution space: $u(\theta) \in V = L^1((0,T),\mathbb{R}^d)$.

Structure: $\int_{\mathbb{R}^d} u(t,x)\mathrm{d}x = 1$.

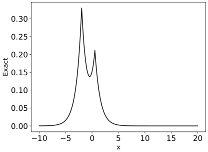
Parameters: $\theta \in \Theta$



Burgers



KdV



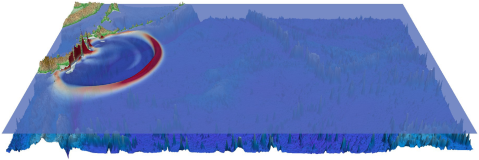
Camassa-Holm

Hamiltonian systems:

$$\begin{cases} \dot{u}(t,\theta) & = J_{2N} \nabla_u \mathcal{H}(u(t,\theta),\theta), \quad \forall t \in \mathbb{T} := (0, T] \\ u(0,\theta) & = u_0(\theta) \end{cases}$$

where:

- $u(t,x) \in \mathcal{C}^1(\mathbb{T};\mathbb{R}^{2N})$ is the state variable,
- $J_{2N} \in \mathbb{R}^{2N\times 2N}$ is skew-symmetric,
- $\mathcal{H}$ is the Hamiltonian,
- $\theta \in \Theta \subset \mathbb{R}^p$ is a vector of parameter.



Shallow-water

Hamiltonian structure:

- Preservation of the Hamiltonian along trajectories:

$$\frac{\mathrm{d}}{\mathrm{dt}}\mathcal{H}(u(t,\theta),\theta) = 0, \quad \forall t \in \mathbb{T},\ \forall \theta \in \Theta.$$

- The flow map $\varphi_t(y_0) = u(t)$ is a simplectic transformation:

$$\left(\frac{\partial \varphi_t}{\partial u_0}\right)^T J_{2N} \left(\frac{\partial \varphi_t}{\partial u_0}\right) = J_{2N}, \quad \forall t \in \mathbb{T}.$$

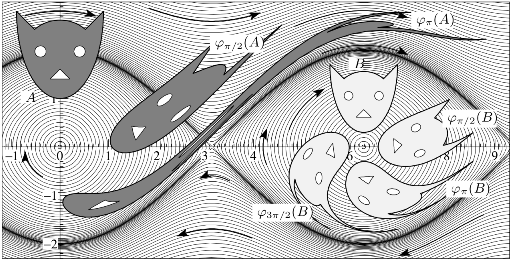


Figure: Area preservation of the flow [HWL10].

Starting point: Let (V, d) be a Banach space, and let

$$\mathcal{B}(u;\theta) = 0$$

be an operator equation where the solution

$$u = u(\theta) \in V$$

for parameters θ in a compact set $\Theta \subset \mathbb{R}^p$.

Parameter-to-solution map:
$$\begin{aligned} u : \Theta &\to V \\ \theta &\mapsto u(\theta) \end{aligned}$$

Solution manifold: $\mathcal{M} := \mathrm{Im}(u) = u(\Theta) = \{u(\theta) \ : \ \theta \in \Theta\} \subset V$

Note that we are working with a particular decoder map:
$\mathcal{M}$ is a nonlinear set of the form

$$V_p = \{\mathrm{D}(c) \ : \ c \in \Theta \subset \mathbb{R}^p\} \subset V$$

where $\mathrm{D} = u$, and $c = \theta$

Relevant problem classes need to evaluate $\theta \mapsto u(\theta)$ many-times:

- Parameter optimization
- Inverse problems
- Uncertainty quantification:
  if $\theta \sim \rho \in \mathcal{P}(\Theta) \Rightarrow u(\theta) \sim u\#\rho \in \mathcal{P}(V)$?

MOR develops methods to approximate

$$\theta \mapsto u(\theta) \quad \text{and} \quad \mathcal{M} := u(\Theta)$$

with small complexity.

Our goal

We want to build a decoder, an algorithm $A : \Theta \to V_n$ such that

$$A(\theta) \approx u(\theta), \quad \forall \theta \in \Theta.$$

To reduce complexity:

- Computing $A(\theta)$ must be must faster compared to $u(\theta)$...
- ... so $A \neq u$, and the dimension of $V_n = \mathrm{Im}(A)$ should be small.
- We have the freedom to choose V_n.

Performance benchmarks

Performance of a given decoder map $A : \Theta \to V_n$:

- In the average sense:

$$\mathcal{E}^{\mathrm{av}}(A) := \mathbb{E}^{1/2}_{\theta\sim\rho_\Theta}\left[d^2(A(\theta), u(\theta))\right].$$

- Worst case:

$$\mathcal{E}^{\max}(A) := \max_{\theta\in\Theta} d(A(\theta), u(\theta)).$$

Ideally, we want to work with the best mapping, namely:

$$A^* \in \operatorname*{arg\,min}_{A:\Theta\mapsto V_n} \mathcal{E}^{\star}(A), \quad \star \in \{\max, \mathrm{av}\}.$$

where the min. runs over all decoders $A : \Theta \to V_n$ with $\dim(V_n) = n$.

Optimal performance of linear approximation

If we search only among linear spaces $V_n \subset V$,

$$\min_{\substack{A:\Theta\mapsto V_n \\ V_n \text{ linear}}} \mathcal{E}^\star(A), \quad \star \in \{\max, \text{av}\},$$

is reached by

$$A^*(\theta) = P_{V_n^{\text{opt},\star}}(u(\theta))$$

for some optimal space $V_n^{\text{opt},\star}$, and

$$d_n(\mathcal{M}) = \min_{\substack{A:\Theta\mapsto V_n \\ V_n \text{ linear}}} \mathcal{E}^{\max}(A) \qquad \text{Kolmogorov } n\text{-width}$$

$$d_n^{(2,\rho_\Theta)}(\mathcal{M}) = \min_{\substack{A:\Theta\mapsto V_n \\ V_n \text{ linear}}} \mathcal{E}^{\text{av}}(A) \qquad \text{Weighted Kolm. width (SVD)}$$

Bounds for Kolmogorov n-width

For $\mathcal{M}$ the solution manifold of a parametric PDE:

- Elliptic/Parabolic Problems ([CD16]):

$$d_n(\mathcal{M}) \lesssim e^{-\alpha n^\beta}$$

- Pure transport, wave propagation ([BCOW17, GU19]):

$$d_n(\mathcal{M}) \geq C n^{-1/2}$$

Need for nonlinear approximation beyond the elliptic case, but let us discuss linear approximation a bit further.

Approximation of V_n^{opt}

We can compute a sequence of $(V_n)_n$ that gives the same decay rate as $(V_n^{\text{opt}})_n$. For this, we sample

$$\mathcal{M} \approx \widetilde{\mathcal{M}} = \{u(\theta_1), \dots, u(\theta_K)\}$$

and then we run:

- a greedy algorithm (worst case).
- an SVD (average case).

Greedy algorithm:

- $n = 1$: Choose u_1 randomly or pick

$$u_1 = \arg\max_{u \in \widetilde{\mathcal{M}}} \|u\|$$
$$\mathrm{U}_1 = \{u_1\}$$
$$V_1 := \mathrm{span}\{\mathrm{U}_1\}$$

- $n > 1$: Given U_{n-1} and V_{n-1},

$$u_n = \arg\max_{u \in \widetilde{\mathcal{M}}} \|u - P_{V_{n-1}} u\|$$
$$\mathrm{U}_n = \mathrm{U}_{n-1} \cup \{u_n\}$$
$$V_n = \mathrm{span}\{\mathrm{U}_n\}$$

Theorem ([BCD$^+$11, DPW13]):

$$\begin{cases} d_n(\mathcal{M}) & = \mathcal{O}(n^{-\alpha}) \\ d_n(\mathcal{M}) & = \mathcal{O}(e^{cn^{-\beta}}) \end{cases} \Longrightarrow \begin{cases} \max_{u \in \mathcal{M}} \|u - P_{V_n} u\| & = \mathcal{O}(n^{-\alpha}) \\ \max_{u \in \mathcal{M}} \|u - P_{V_n} u\| & = \mathcal{O}(e^{\tilde{c}n^{-\beta}}) \end{cases}$$

Sampling: Quality of V_n from the greedy algorithm depends on $\widetilde{\mathcal{M}}$. Impact is difficult to quantify (see [CDDN20]).

Galerkin Projection:
The mapping

$$A(\theta) = P_{V_n} u(\theta) = \sum_{i=1}^{n} \langle u(\theta), \varphi_i \rangle u_i$$

requires computing $u(\theta)$ so this A is not admissible.

If the PDE is uniformly inf-sup stable (coercive), we can replace the orthogonal projection by a computable Galerkin projection:

$$P_{V_n} u(\theta) \rightsquigarrow u_n(\theta) \in V_n.$$

Example: Suppose $0 < \theta_{\min} \le \theta \le \theta_{\max}$, and consider

$$-\theta\Delta u = f \text{ in } \Omega$$
$$u = 0, \text{ on } \partial\Omega$$

Weak formulation: Find $u(\theta) \in V = H_0^1(\Omega)$ s.t.

$$\underbrace{\theta \int_\Omega \nabla u(\theta) \cdot \nabla v}_{:=a(u,v;\theta)} = \underbrace{\int_\Omega fv}_{:=f(v)}, \quad \forall v \in H_0^1(\Omega).$$

Well-posed and stable iif there exist $C, c > 0$ s.t.

$$a(v,v;\theta) \ge c\|v\|_V^2, \quad a(v,v;\theta) \le C\|v\|_V^2, \quad \forall v \in V, \forall\theta \in \Theta.$$

Galerkin projection in reduced space: Find $u_n(\theta) \in V_n$ s.t.

$$\underbrace{\theta \int_\Omega \nabla u_n(\theta) \cdot \nabla v}_{:=a(u,v;\theta)} = \underbrace{\int_\Omega fv}_{:=f(v)}, \quad \forall v \in V_n.$$

We then have

$$\|u(\theta) - P_{V_n} u(\theta)\|_V \sim \|u(\theta) - u_n(\theta)\|_V \sim \mathcal{R}(\theta) := \|a(u_n(\theta), \cdot, \theta) - f(\cdot)\|_{V'}.$$

Approximation of $V_n^{\text{opt},\star}$ (updated)

Greedy algorithm:

- $n = 1$: Choose u_1 randomly and set

$$\mathrm{U}_1 = \{u_1\}$$
$$V_1 := \text{span}\{\mathrm{U}_1\}$$

- $n > 1$: Given U_{n-1} and V_{n-1},

$$u_n = \arg\max_{u \in \widetilde{\mathcal{M}}} \|u - P_{V_{n-1}} u\| \quad \rightsquigarrow \theta_n \in \arg\max_{\theta \in \widetilde{\Theta}} \mathcal{R}(\theta) \rightsquigarrow u_n(\theta_n)$$
$$\mathrm{U}_n = \mathrm{U}_{n-1} \cup \{u_n\}$$
$$V_n = \text{span}\{\mathrm{U}_n\}$$

Theorem ([BCD+11, DPW13]):

$$\begin{cases} d_n(\mathcal{M}) & = \mathcal{O}(n^{-\alpha}) \\ d_n(\mathcal{M}) & = \mathcal{O}(e^{cn^{-\beta}}) \end{cases} \Longrightarrow \begin{cases} \max_{\theta \in \Theta} \|u(\theta) - u_n(\theta)\| & = \mathcal{O}(n^{-\alpha}) \\ \max_{\theta \in \Theta} \|u(\theta) - u_n(\theta)\| & = \mathcal{O}(e^{\tilde{c}n^{-\beta}}) \end{cases}$$

Linear approximation is a very solid approach for MOR of elliptic problems.

Part II.2

Reduced Order Modelling of Parametrized PDEs

Nonlinear Approximation

The main idea is:

- V Hilbert space.
- Compute SVD for a small dimension n:

$$V_n = \mathrm{span}\{\varphi_i\}_{i=1}^n \quad (\mathrm{ONB}), \qquad V = V_n \oplus V_n^\perp.$$

- For every $\theta \in \Theta$,

$$\begin{aligned} u(\theta) &= P_{V_n} u(\theta) + P_{V_n^\perp} u(\theta) \\ &= \sum_{i=1}^n a_i(\theta)\varphi_i + P_{V_n^\perp} u(\theta), \quad \forall \theta \in \Theta. \end{aligned}$$

- We want to use

$$a(\theta) := (a_1(\theta), \dots, a_n(\theta))$$

to approximate $P_{V_n^\perp} u(\theta)$. So we want to learn a decoder

$$\mathrm{D} : \mathbb{R}^n \to V_n^\perp$$

such that

$$a \mapsto \mathrm{D}(a(\theta)) \approx P_{V_n^\perp} u(\theta), \quad \forall \theta \in \Theta.$$

- How to parametrize $V_n^\perp$?

Nonlinear compressive reduced basis [CFSM23]

- Compute SVD for a large dimension $N \gg n \geq 1$:

$$V_N = \text{span}\{\varphi_i\}_{i=1}^N \quad \text{(ONB)}$$

Take

$$V_n \approx \text{span}\{\varphi_1, \dots, \varphi_n\}$$
$$V_n^\perp \approx \text{span}\{\varphi_{n+1}, \dots, \varphi_N\}$$

- Approximate

$$u(\theta) \approx u_N(\theta) := \sum_{i=1}^n a_i(\theta)\varphi_i + \sum_{j>n}^N b_j(\theta)\varphi_j, \quad \forall \theta \in \Theta,$$

where the ideal a_i and b_j are

$$a_i(\theta) := \langle u(\theta), \varphi_i \rangle_V, \quad \text{and} \quad b_j(\theta) := \langle u(\theta), \varphi_j \rangle_V.$$

- Compute SVD for a large dimension $N \gg n \geq 1$:

$$V_N = \text{span}\{\varphi_i\}_{i=1}^N \quad \text{(ONB)}$$

Take

$$V_n \approx \text{span}\{\varphi_1, \dots, \varphi_n\}$$
$$V_n^\perp \approx \text{span}\{\varphi_{n+1}, \dots, \varphi_N\}$$

- Approximate

$$u(\theta) \approx u_N(\theta) := \sum_{i=1}^{n} a_i(\theta)\varphi_i + \sum_{j>n}^{N} b_j(\theta)\varphi_j, \quad \forall \theta \in \Theta,$$

where the ideal a_i and b_j are

$$a_i(\theta) := \langle u(\theta), \varphi_i \rangle_V, \quad \text{and} \quad b_j(\theta) := \langle u(\theta), \varphi_j \rangle_V.$$

- Build mappings

$$b_j : \Theta \to \mathbb{R}$$
$$\theta \mapsto b_j(\theta) = \psi_j(\underbrace{a_1(\theta), \dots, a_n(\theta)}_{:=a(\theta)}), \quad n < j \leq N$$

for a well chosen $\psi_j : \mathbb{R}^n \to \mathbb{R}$.

Choose ψ_j among a class $\mathcal{F}$ of (parametrized) decoder functions from $\mathbb{F}(\mathbb{R}^n, \mathbb{R})$ and do empiral risk minimization:

$$\psi_j := \arg\min_{f \in \mathcal{F}} \left\{ \sum_{i=1}^{K} \left| \langle u(\theta_i), \varphi_j \rangle - f(\underbrace{a_1(\theta_i), \dots, a_n(\theta_i)}_{:= a(\theta_i)}) \right| \right\}$$

In [CFSM23] they work with **neural networks**:

$$\mathcal{F} = \{ \mathcal{N}_c : \mathbb{R}^n \to \mathbb{R} \ : \ c \in \mathbb{R}^q \}$$

Therefore

$$c_j^* \in \arg\min_{c \in \mathbb{R}^q} \left\{ \sum_{i=1}^{K} \left| \langle u(\theta_i), \varphi_j \rangle - \mathcal{N}_c(a(\theta_i)) \right| \right\}$$
$$\psi_j(a) = \mathcal{N}_{c_j^*}(a).$$
$$b_j(\theta) = \mathcal{N}_{c_j^*}(a(\theta)).$$

Quadratic manifold approximation [BF22, GWW23]

An alternative strategy is to build

$$\mathrm{D}(a(\theta)) \approx P_{V_n^\perp} u(\theta)$$

by introducing the tensor product of the coefficients

$$a \otimes a = (a_1 a_1, a_1 a_2, \dots, a_1 a_n, a_2 a_1, \dots, a_n a_n) \in \mathbb{R}^{n^2}$$

and then we search for the best basis vectors

$$\{\widetilde{\varphi}_{i,j}\}_{1 \leq i,j \leq n} \subset V_n^\perp$$

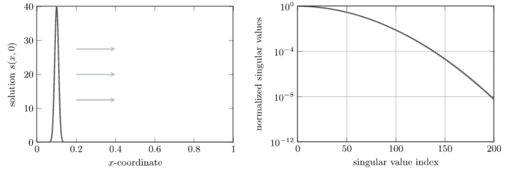
to approximate

$$u(\theta) \approx \sum_{i=1}^{n} a_i(\theta)\varphi_i + \sum_{1 \leq i,j \leq n} a_i(\theta) a_j(\theta) \widetilde{\varphi}_{i,j}.$$

Compared to the previous approach:

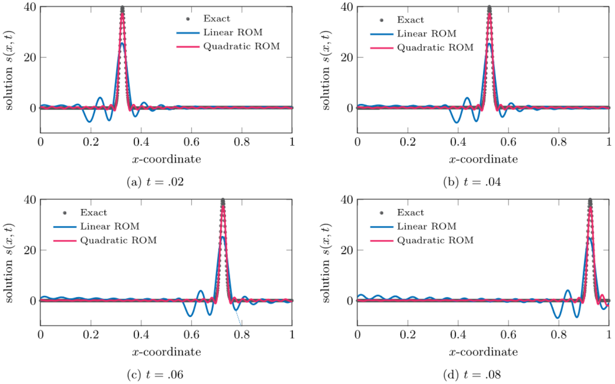
- The rule for the coefs is much simpler (quadratic VS neural network)
- Finding the $\{\widetilde{\varphi}_{i,j}\}_{1 \leq i,j \leq n}$ is more involved.

Pure transport $\partial_t u + v\nabla_x u = 0$



(a) Initial condition at $\mu = 0.10$ and transport direction.

(b) Singular values.



(a) $t = .02$

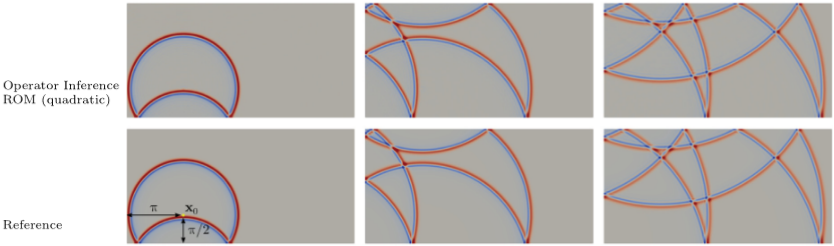
(b) $t = .04$

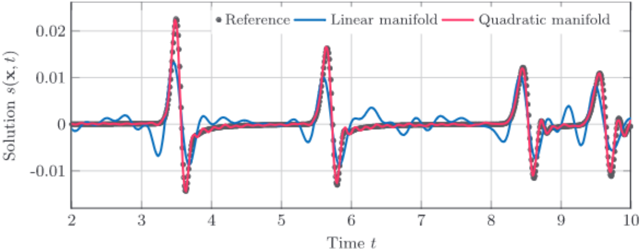
(c) $t = .06$

(d) $t = .08$

Some results with quadratic approximation (from [GWW23])

Wave equation $\partial_{tt} u - \Delta u = 0$



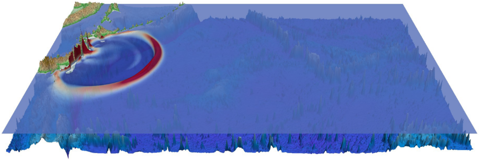


Part II.3

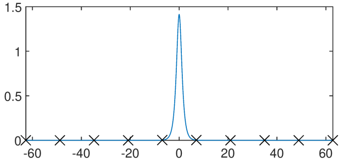
Reduced Order Modelling of Parametrized PDEs

The role of geometry

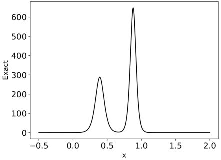
Hamiltonian Problems



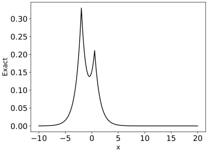
Shallow-water



Schrödinger



KdV



Camassa-Holm

Hamiltonian systems:

$$\begin{cases} \dot{u}(t,\theta) & = J_{2N} \nabla_u \mathcal{H}(u(t,\theta),\theta), \quad \forall t \in \mathbb{T} := (0,T] \\ u(0,\theta) & = u_0(\theta) \end{cases}$$

where:

- $u \in \mathcal{C}^1(\mathbb{T};\mathbb{R}^{2N})$ is the state variable. Here: $V = \mathbb{R}^{2N}$
- $J_{2N} \in \mathbb{R}^{2N \times 2N}$ is skew-symmetric,
- $\mathcal{H}$ is the Hamiltonian,
- $\theta \in \Theta \subset \mathbb{R}^p$ is a vector of parameters.

Hamiltonian structure:

- Preservation of the Hamiltonian along trajectories:

$$\frac{\mathrm{d}}{\mathrm{dt}}\mathcal{H}(u(t,\theta),\theta) = 0, \quad \forall t \in \mathbb{T},\ \forall \theta \in \Theta.$$

- The flow map $\varphi_t(y_0) = u(t)$ is a simplectic transformation:

$$\left(\frac{\partial \varphi_t}{\partial u_0}\right)^T J_{2N} \left(\frac{\partial \varphi_t}{\partial u_0}\right) = J_{2N}, \quad \forall t \in \mathbb{T}.$$

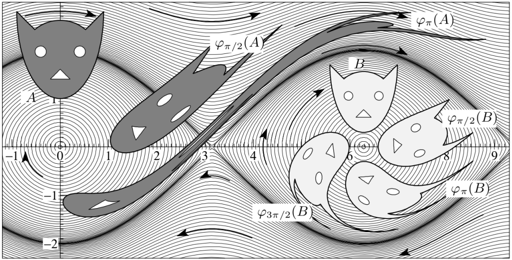


Figure: Area preservation of the flow [HWL10].

We consider for every $t \in \mathbb{T}$,

$$\mathcal{M}(t) := \{u(t,\theta) \ : \ \theta \in \Theta\} \subset \mathbb{R}^{2N}, \qquad \mathcal{M} = \cup_{t\in\mathbb{T}} \mathcal{M}(t),$$

and approximate

$$\mathcal{M}(t) \approx V_n(t).$$

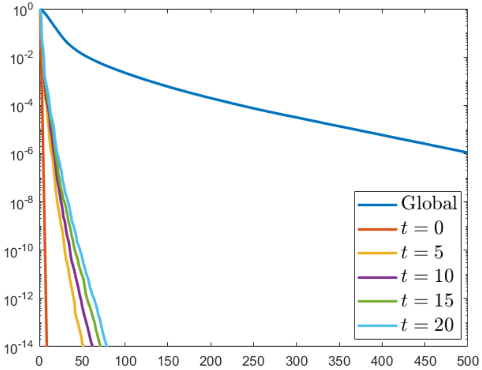
We then work with the **time-dependent linear ansatz**

$$u(t,\theta) \approx u_n(t,\theta) = \sum_{i=1}^{2n} c_i(t,\theta) v_i(t) \ \in V_n(t) = \operatorname{span}\{v_i(t)\}_{i=1}^{2n} \subset \mathbb{R}^{2N}.$$

Such a strategy is called dynamical low rank.

Very efficient when

$$d_n(\mathcal{M}(t)) \ll d_n(\mathcal{M}), \quad \text{or} \quad d_n^{(2,\rho_\Theta)}(\mathcal{M}(t)) \ll d_n^{(2,\rho_\Theta)}(\mathcal{M})$$



SVD of $\mathcal{M}(t)$ and $\mathcal{M} = \cup_{t\in\mathbb{T}}\mathcal{M}(t)$

Starting from

$$\begin{cases} \dot{u}(t,\theta) & = J_{2N}\nabla_u \mathcal{H}(u(t,\theta),\theta), \quad \forall t \in \mathbb{T} := (0,T] \\ u(0,\theta) & = u_0(\theta) \end{cases}$$

with $u(t,\theta) \in \mathbb{R}^{2N}$, we approximate

$$u(t,\theta) \approx u_n(t,\theta) = \sum_{i=1}^{2n} c_i(t,\theta) v_i(t) = \mathbf{V}(t) c(t,\theta),$$

where

$$V_n(t) := \{v_i(t)\}_{i=1}^{2n} \subset V = \mathbb{R}^{2N} \quad \leftrightsquigarrow \quad \mathbf{V}(t) \in \mathbb{R}^{2N\times 2n}.$$

Hamiltonian simplectic preservation requires:

$$\mathbf{V}(t) \text{ orthosymplectic} \quad \Longleftrightarrow \quad \begin{cases} \mathbf{V}(t)^T J_{2N} \mathbf{V}(t) = J_{2n}, \\ \mathbf{V}(t)^T \mathbf{V}(t) = \mathrm{I}_{2n}. \end{cases}$$

Method to build $V_n(t)$ (from [HP21, Pag21, HPR22])

Starting from a good $V_n(0)$, and $c(0,\theta)$, how to do the time integration?

- Consider the training set

$$\mathcal{U}(t) = [u(t,\theta_1),\dots,u(t,\theta_K)] \approx \mathrm{U}_{2n}(t) = \mathbf{V}(t)C(t)$$

with

$$\mathbf{V}(t) \in \mathbb{R}^{2N\times 2n} \qquad \text{(basis)}$$

$$C(t) = (c_i(t,\theta_j))_{\substack{1\le i\le 2n\\ 1\le j\le K}} \in \mathbb{R}^{2n\times K}, \qquad \text{(coefs)}$$

- To have a symplectic low-rank integration, we require that

$$\mathrm{U}_{2n}(t) \in \mathcal{S} := \{U \in \mathbb{R}^{2N\times 2n} \;:\; U = VC \text{ with } V \in \mathcal{V}_{2n}, C \in \mathcal{C}_{2n}\}$$

and

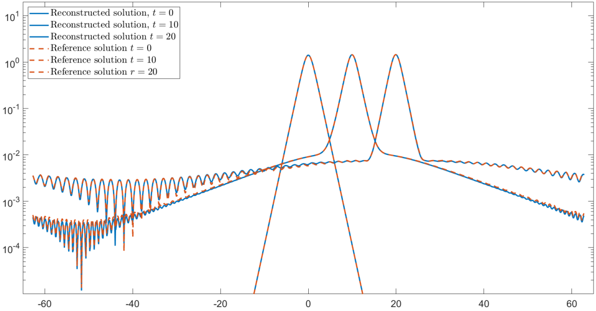
$$\mathcal{V}_{2n} := \{V \in \mathbb{R}^{2N\times 2n} \;:\; V^TV = I_{2n},\; V^TJ_{2N}V = J_{2N}\} \qquad \text{(orthosymplectic)}$$

$$\mathcal{C}_{2n} := \{C \in \mathbb{R}^{2n\times K} \;:\; \mathrm{rank}(C^TC + J_{2n}^TCC^TJ_{2n}) = 2n\} \qquad \text{(full rank)}$$

- We then search for $\mathrm{U} \in \mathcal{C}^1(\mathbb{T},\mathcal{S})$ such that

$$\dot{\mathrm{U}}(t) = P_{T_{\mathcal{S}}U(t)}J_{2N}\nabla H(\mathrm{U}(t)) \quad \Rightarrow \begin{cases} \dot{V}(t) & = \dots \\ \dot{C}(t) & = \dots \end{cases}$$

$$i\frac{\partial u}{\partial t} + \Delta u + |u|^2 u = 0 \quad \text{in } \mathbb{T} \times \Omega$$
$$u(t=0, x, \theta) = (1 + \alpha \sin x)(2 + \beta \sin y)$$



2D: See video.

Part II.3

Reduced Order Modelling of Parametrized PDEs

The role of geometry

Conservation Laws, Measured-Valued problems, and the role of Optimal Transport

Sparse Interpolation from a Dictionary in W_2

(a) H. Do (Dauphine)

(b) J. Feydy (Inria)

(c) V. Ehrlacher (Ecole Ponts)

(d) D. Lombardi (Inria)

(e) F.X. Vialard (U. Gustave Eiffel)

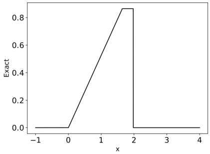
References [ELMV20, DFM23]: Approximation and Structured Prediction with Sparse Wasserstein Barycenters. arXiv:2302.05356

Measure-valued problems:

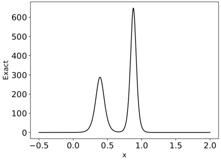
- Conservation laws (Burgers, Camassa-Holm, KdV)
- Fokker-Planck equations
- Wasserstein gradient flows (heat eq., porous media, Keller-Segel...)

If viewed in classical Banach spaces (e.g., $L^1(\Omega), L^2(\Omega)$), **slow decay of the Kolomogorov n-width** due to:

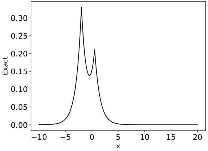
- Transport of shocks and discontinuities
- Non-smooth parameter dependence



Burgers



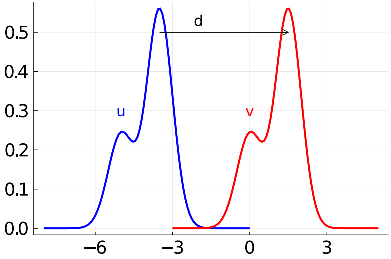
KdV



Camassa-Holm

Viewing solutions in $(\mathcal{P}_2(\Omega), W_2)$ allows us:

- To **preserve mass**.
- To **penalize translations** through the metric W_2. This helps to locate shocks in MOR approximations.



$$W_2(u, v) = d,$$

$$L_1(u, v) = 2 \quad \text{as soon as } \text{supp}(u) \cap \text{supp}(v) = \emptyset$$

In the Hilbert setting, a linear approximation reads

$$u(\theta) \approx u_n(\theta) := \sum_{i=1}^{n} c_i(\theta) u_i \quad \in V_n = \mathrm{span}\{u_1, \dots, u_n\}$$

where

$$\mathrm{U}_n = \{u_i\}_{i=1}^{n}$$

are n solution snapshots.

The analogue in the Wasserstein space is to work with barycenters

$$u(\theta) \approx \mathrm{Bar}(\Lambda_n(\theta), \mathrm{U}_n) = \operatorname*{arg\,min}_{v \in \mathcal{P}_2(\Omega)} \sum_{i=1}^{n} \lambda_i(\theta) W_2^2(v, u_i)$$

where

$$\Lambda_n(\theta) \in \Sigma_n := \{z \in \mathbb{R}^n \ : \ \sum_{i=1}^{n} z_i = 1, \ z_i \geq 0\}$$

$$\mathrm{Bar}(\Lambda_n, \mathrm{U}_n) = \underset{v \in \mathcal{P}_2(\Omega)}{\arg\min} \sum_{i=1}^{n} \lambda_i W_2^2(v, u_i).$$

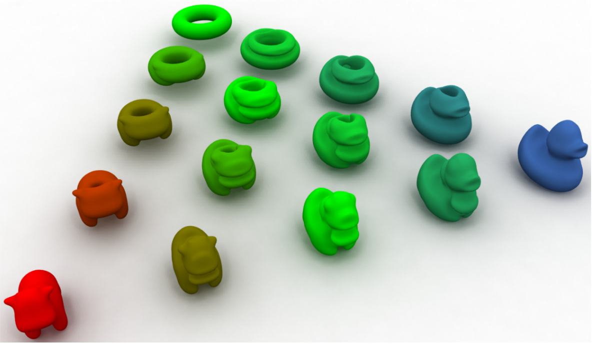


Figure: Image from [SDGP+15]

Snapshot data/dictionary:

$$\Theta_N := \{\theta_i\}_{i=1}^N, \quad \mathrm{U}_N := \{u_i = u(\theta_i)\}_{i=1}^N, \quad N \gg 1.$$

Linear approximation:

- Hilbert spaces: Find V_N^n with greedy algorithm, POD, etc.
- W_2 space: Find U_N^n with a greedy barycenter algorithm ([ELMV20, BBE+22])

Nonlinear version: Given $\theta \in \Theta$,

- Hilbert: Find $V_N^n(\theta)$.
- W_2: Find n snapshots $\mathrm{U}_N^n(\theta)$ among U_N $\rightarrow$ sparse barycenters.

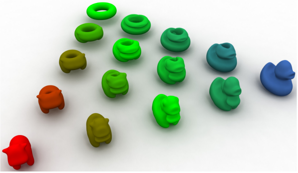
Our contribution: We give an algorithm to approximate the optimal $\mathrm{U}_N^n(\theta)$ and weights $\Lambda_N^n(\theta)$.

The class of n-sparse barycenters:

$$\mathcal{F} := \{\mathrm{Bar}(\Lambda_N^n, \mathrm{U}_N) \ : \ \Lambda_N^n \in \Sigma_N^n\} \subset \mathcal{P}_2(\Omega)$$

where

$$\Sigma_N^n := \{\Lambda_N^n \in \Sigma_N \ : \ \#\mathrm{supp}(\Lambda_N) = n\}.$$



Example: Suppose

$$\Lambda_N^n = (0, 0, \lambda_{i_1}, 0, \dots, \lambda_{i_2}, 0, \dots, \lambda_{i_n}, 0, \dots, 0) \ \in \Sigma_N^n$$

then

$$\begin{aligned}\mathrm{Bar}(\Lambda_N^n, \mathrm{U}_N) &= \underset{v \in \mathcal{P}_2(\Omega)}{\arg\min} \sum_{i=1}^{N} \lambda_i W_2^2(v, u_i) \\ &= \underset{v \in \mathcal{P}_2(\Omega)}{\arg\min}\ \lambda_{i_1} W_2^2(v, u_{i_1}) + \dots + \lambda_{i_n} W_2^2(v, u_{i_n})\end{aligned}$$

We want to build $A : \Theta \to \mathcal{F}$ such that

$$A(\theta) \approx u(\theta), \quad \forall x \in \Theta.$$

Performance of a map $A : \Theta \mapsto \mathcal{F}$:

- In the average sense:

$$\mathcal{E}^{\mathrm{av}}(A) := \mathbb{E}_{\theta \sim \rho_\Theta} \left[W_2^2(A(\theta), u(\theta)) \right].$$

- Worst case:

$$\mathcal{E}^{\max}(A) := \max_{\theta \in \Theta} W_2(A(\theta), u(\theta)).$$

We want to work with the best mapping, namely:

$$A^* \in \operatorname*{arg\,min}_{A:\Theta \mapsto \mathcal{F}} \mathcal{E}^\star(A), \quad \star \in \{\max, \mathrm{av}\}.$$

For both performance benchmarks, the optimal map is to choose

$$A^*(\theta) \in \arg\min_{b \in \mathcal{F}} W_2(u(\theta), b),$$

that is,

$$A^*(\theta) = \mathrm{Bar}(\Lambda_N^n(\theta), \mathrm{U}_N), \quad \text{s.t.} \quad \Lambda_N^n(\theta) \in \arg\min_{\Lambda_N^n \in \Sigma_N^n} W_2^2(u(\theta), \mathrm{Bar}(\Lambda_N^n, \mathrm{U}_N)).$$

$\Longrightarrow$ Best n-term barycenter for $u(\theta)$.
$\Longrightarrow$ Implementable only if we know $u(\theta)$.

As an alternative, consider

$$\min_{\Lambda_N^n \in \Sigma_N^n} \sum_{i=1}^{N} |W_2^2(u(\theta), u(\theta_i)) - W_2^2(\mathrm{Bar}(\Lambda_N^n, \mathrm{U}_N), u(\theta_i))|^2.$$

$u(\theta)$ is still present, BUT...

We can build a local Euclidean metric around each training point $\theta_i \in \Theta_N$ in order to approximate

$$W_2^2(u(\theta), u(\theta_i)) \approx (\theta - \theta_i)^T M(\theta_i)(\theta - \theta_i), \quad \forall i \in \{1, \dots, N\}.$$

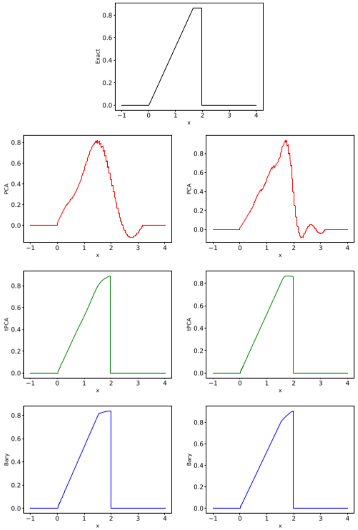
This yields the computable problem

$$\Lambda_N^n(\theta) \in \min_{\Lambda_N^n \in \Sigma_N^n} \sum_{i=1}^{N} |(\theta - \theta_i)^T M(\theta_i)(\theta - \theta_i) - W_2^2(\mathrm{Bar}(\Lambda_N^n, \mathrm{U}_N), u(\theta_i))|^2.$$

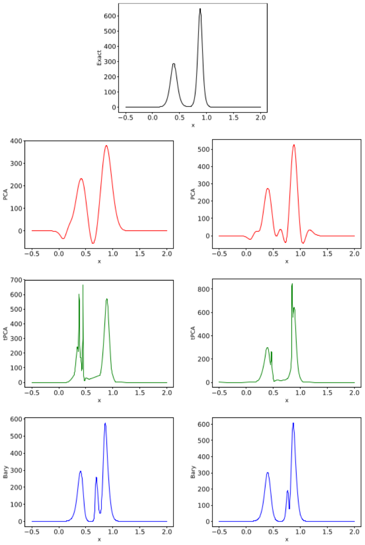
Why is this a good construction?

- Gives optimal map in simple cases (Diracs, translated Gaussians).
- Interpolation: if $\theta_i \in \Theta_N$, $\Lambda_N^n(\theta) = e_i$.
- Invariant under affine reparametrizations in Θ
- Full adaptivity of the support w.r.t. θ and without any extra heuristic.

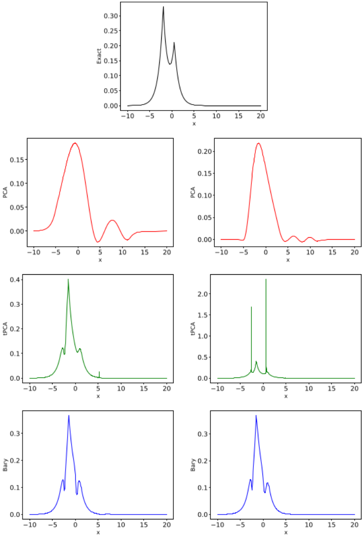
1D Burgers (n=5) [ELMV20]



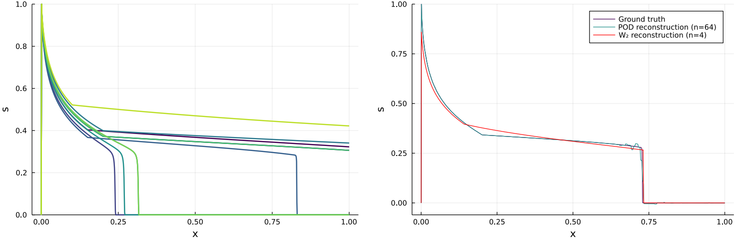
1D Camassa-Holm (n=5) [ELMV20]



1D KdV (n=5) [ELMV20]



Two phase-flow in porous media [BBE+22]



See [BBE+22]: Wasserstein model reduction approach for parametrized flow problems in porous media. arXiv:2205.02721

A Burgers' equation in 2D: Let $\Omega = [0,1]^2$. We want to solve $\forall (t,x) \in [0,T] \times \Omega$,

$$\partial_t u + \frac{1}{2} \nabla_x (u^2) = \beta \Delta_r u$$

with a parametrized initial condition u_0.

Parameters, and associated solution:

$$\theta = (t, \beta, u_0), \qquad u(\theta)(x) = u(t, \beta, u_0; x) \in \mathcal{P}_2(\Omega)$$

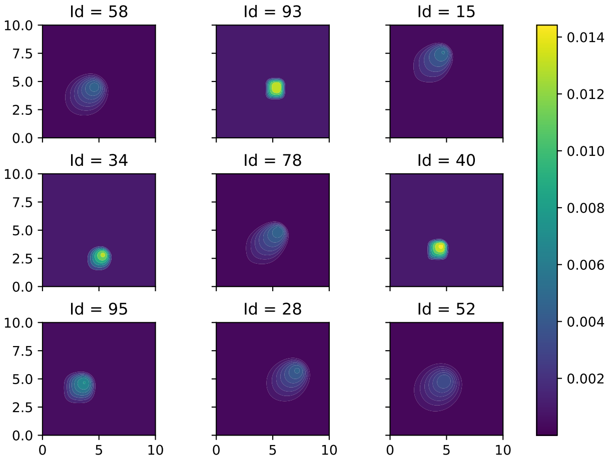


Figure: Some measures from the training set U_N.

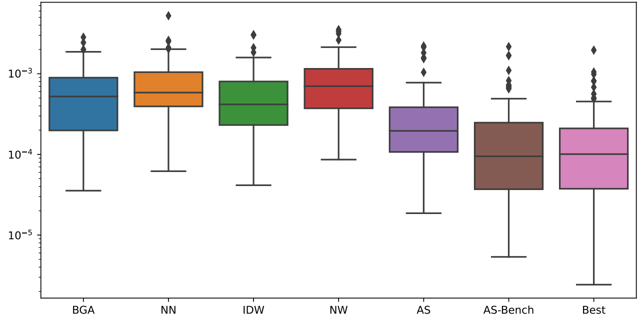


Figure: Approximation errors in the validation set.

Comparison of different approaches

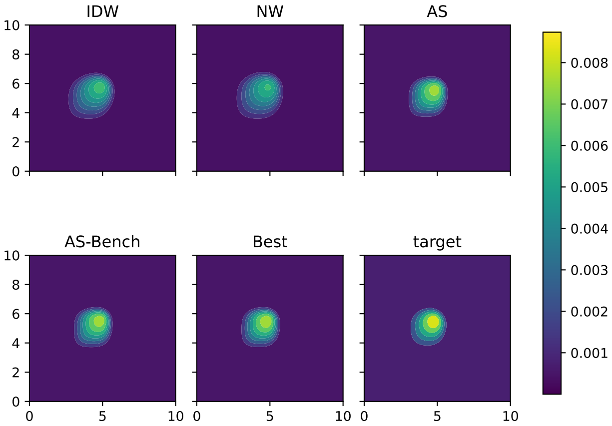


Figure: Approximation of a sample from the validation set.

Conlusions

1) Landscape for linear approximation is very complete nowadays.

2) Vibrant developments in nonlinear approximation.

3) Each PDE requires its own method:

- Elliptic and parabolic problems: Linear Approximation.
- Nonlinear methods for other PDEs:
  - Nonlinear compressive MOR/Quadratic Manifold Learning
  - Exploiting geometry is a promising approach
    - Dynamical low rank
    - Nonlinear, Metric spaces: Tools from OT for measure-valued solutions.

B. Battisti, T. Blickhan, G. Enchery, V. Ehrlacher, D. Lombardi, and O. Mula, *Wasserstein model reduction approach for parametrized flow problems in porous media*, working paper or preprint, May 2022.

P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk, *Convergence rates for greedy algorithms in reduced basis methods*, SIAM Journal on Mathematical Analysis **43** (2011), no. 3, 1457–1472.

P. Benner, A. Cohen, M. Ohlberger, and K. Willcox, *Model reduction and approximation: Theory and algorithms*, vol. 15, SIAM, 2017.

J. Barnett and C. Farhat, *Quadratic approximation manifold for mitigating the kolmogorov barrier in nonlinear projection-based model order reduction*, Journal of Computational Physics **464** (2022), 111348.

A. Cohen and R. DeVore, *Kolmogorov widths under holomorphic mappings*, IMA Journal of Numerical Analysis **36** (2016), no. 1, 1–12.

A. Cohen, W. Dahmen, R. DeVore, and J. Nichols, *Reduced basis greedy selection using random training sets*, ESAIM: Mathematical Modelling and Numerical Analysis **54** (2020), no. 5, 1509–1524.

A. Cohen, C. Farhat, A. Somacal, and Y. Maday, *Nonlinear compressive reduced basis approximation for pdes*.

H. Do, J. Feydy, and O. Mula, *Approximation and Structured Prediction with Sparse Wasserstein Barycenters*, 2022.

M.-H. Do, J. Feydy, and O. Mula, *Approximation and structured prediction with sparse wasserstein barycenters*, arXiv preprint arXiv:2302.05356 (2023).

R. DeVore, G. Petrova, and P. Wojtaszczyk, *Greedy algorithms for reduced bases in Banach spaces*, Constructive Approximation **37** (2013), no. 3, 455–466.

V. Ehrlacher, D. Lombardi, O. Mula, and F.-X. Vialard, *Nonlinear model reduction on metric spaces. application to one-dimensional conservative pdes in wasserstein spaces*, ESAIM M2AN **54** (2020), no. 6, 2159–2197.

C. Greif and K. Urban, *Decay of the kolmogorov n-width for wave problems*, Applied Mathematics Letters **96** (2019), 216–222.

- R. Geelen, S. Wright, and K. Willcox, *Operator inference for non-intrusive model reduction with quadratic manifolds*, Computer Methods in Applied Mechanics and Engineering **403** (2023), 115717.
- J. S. Hesthaven and C. Pagliantini, *Structure-preserving reduced basis methods for poisson systems*, Mathematics of Computation **90** (2021), no. 330, 1701–1740.
- J. S. Hesthaven, C. Pagliantini, and N. Ripamonti, *Rank-adaptive structure-preserving model order reduction of hamiltonian systems*, ESAIM: Mathematical Modelling and Numerical Analysis **56** (2022), no. 2, 617–650.
- E. Hairer, G. Wanner, and C. Lubich, *Geometric numerical integration*, 2010.
- C. Pagliantini, *Dynamical reduced basis methods for Hamiltonian systems*, Numerische Mathematik **148** (2021), no. 2, 409–448.
- Justin Solomon, Fernando De Goes, Gabriel Peyré, Marco Cuturi, Adrian Butscher, Andy Nguyen, Tao Du, and Leonidas Guibas, *Convolutional wasserstein distances: Efficient optimal transportation on geometric domains*, ACM Transactions on Graphics (ToG) **34** (2015), no. 4, 1–11.