Multiscaling in Wasserstein Spaces

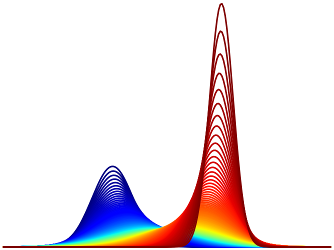
Wael Mattar
PhD student at Tel Aviv University
Joint work with *Nir Sharon*

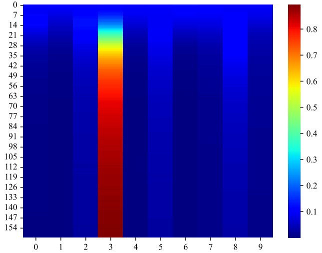
Curves & Surfaces conference
Saint-Malo, France

Outline

Outline

Motivation (curves of probability measures)





Absolutely continuous (Gaussian) measures.

Discrete measures supported on $\{0,\ldots,9\}$.

The main idea

Transforming a sequence $\boldsymbol{c}^{(1)}$ into $\{\boldsymbol{c}^{(0)}, \boldsymbol{d}^{(1)}\}$ where $\boldsymbol{c}^{(0)}$ is a coarse approximation and $\boldsymbol{d}^{(1)}$ encodes the detail coefficients.

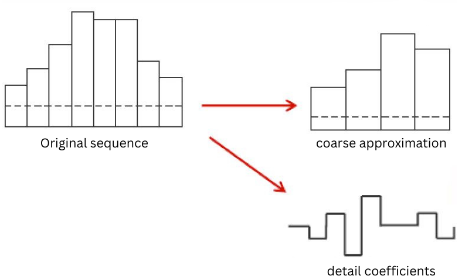
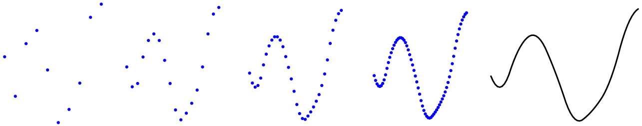


Figure: Sketch of decomposing a real-valued sequence.

Outline

Subdivision schemes

In multiscale transforms we use subdivision schemes as upsampling operators.



A subdivision scheme associated with a mask $\boldsymbol{\alpha} = \{\alpha_j\}_{j\in\mathbb{Z}} \subset \mathbb{R}$ is a refinement operator defined by

$$\mathcal{S}_{\boldsymbol{\alpha}}(\boldsymbol{c})_k = \sum_{j\in\mathbb{Z}} \alpha_{k-2j} c_j, \quad k \in \mathbb{Z},$$

for any sequence $\boldsymbol{c} = \{c_j\}_{j\in\mathbb{Z}} \subset \mathbb{R}$.

Subdivision schemes (cont.)

We employ subdivision scheme $\mathcal{S}_{\boldsymbol{\alpha}}$ that

- satisfy the constant-reproduction property $\sum_{j\in\mathbb{Z}} \alpha_{2j} = \sum_{j\in\mathbb{Z}} \alpha_{2j+1} = 1$.
- and *interpolating* in the sense that $\mathcal{S}_{\boldsymbol{\alpha}}(\boldsymbol{c})_{2k} = c_k$ for all $k \in \mathbb{Z}$.

The most elementary subdivision scheme is given by the mask $\boldsymbol{\alpha} = [1/2,\ 1,\ 1/2]$ supported on $\{-1, 0, 1\}$.

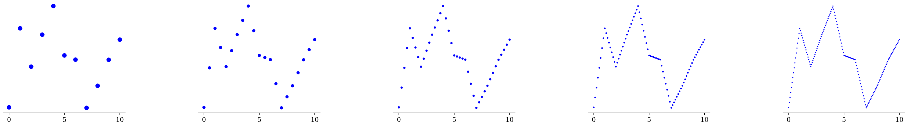


Figure: Depiction of refinements by the elementary subdivision scheme.

Interpolating decomposition

With an interpolating subdivision scheme $\mathcal{S}_\alpha$, a sequence $\boldsymbol{c}^{(1)}$ associated with $2^{-1}\mathbb{Z}$ can be decomposed into a coarse approximation $\boldsymbol{c}^{(0)}$ and detail coefficients $\boldsymbol{d}^{(1)}$ by

$$\boldsymbol{c}^{(0)} = \mathcal{D}\boldsymbol{c}^{(1)}, \quad \boldsymbol{d}^{(1)} = \boldsymbol{c}^{(1)} - \mathcal{S}_\alpha \boldsymbol{c}^{(0)},$$

where $(\mathcal{D}\boldsymbol{c})_k = c_{2k}$, $k \in \mathbb{Z}$ is the simple downsampling operator.

By construction it can be easily seen that $d_{2k}^{(1)} = 0$ for all $k \in \mathbb{Z}$, and that is a vital property for applications in multiscaling.

Pyramid transform

$$\boxed{\text{Sequence } \boldsymbol{c}^{(J)}} \underset{\text{reconstruction}}{\overset{\text{decomposition}}{\rightleftharpoons}} \boxed{\text{Pyramid } \left\{ \boldsymbol{c}^{(0)}; \boldsymbol{d}^{(1)}, \boldsymbol{d}^{(2)}, \ldots, \boldsymbol{d}^{(J)} \right\}}$$

Definition (multiscale transform)

The multiscale transform of a sequence $\boldsymbol{c}^{(J)}$ associated with the grid $2^{-J}\mathbb{Z}$, $J \in \mathbb{N}$ is defined by

$$\boldsymbol{c}^{(\ell-1)} = \mathcal{D}\boldsymbol{c}^{(\ell)}, \quad \boldsymbol{d}^{(\ell)} = \boldsymbol{c}^{(\ell)} - \mathcal{S}_\alpha \boldsymbol{c}^{(\ell-1)}, \quad \ell = 1, \ldots, J,$$

while the inverse multiscale transform is given by

$$\boldsymbol{c}^{(\ell)} = \mathcal{S}_\alpha \boldsymbol{c}^{(\ell-1)} + \boldsymbol{d}^{(\ell)}, \quad \ell = 1, \ldots, J.$$

Outline

Multiscaling in Wasserstein spaces

We adapt the multiscaling to the space of probability measures

$$\mathcal{P}_p(\mathbb{R}^d) = \left\{ \mu \in \mathcal{P}(\mathbb{R}^d) \;\middle|\; \int_{\mathbb{R}^d} \|x\|^p d\mu(x) < \infty \right\}$$

endowed with the Wasserstein metric W_p defined by

$$W_p^p(\mu, \nu) = \min_{\sigma \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^p d\sigma(x, y)$$

where $\Pi(\mu, \nu)$ is the set of all probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν.

Challenges

For adapting the (elementary) multiscaling framework to Wasserstein spaces we are faced with two main challenges:

1. We need to find analogues to the Euclidean operators "$-$" and "$+$", which we will denote by $\ominus$ and $\oplus$, such that for any two admissible measures $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ we have the compatibility condition

$$\mu \oplus (\nu \ominus \mu) = \nu.$$

2. Defining a refinement operator in $\mathcal{P}_p(\mathbb{R}^d)$.

Solutions

Wasserstein spaces $\mathcal{P}_p(\mathbb{R}^d)$ exhibit a sort of infinite-dimensional Riemannian manifold structure. This is observed by McCann's interpolants and the continuity equation.

Definition (McCann's interpolants)

for $\mu_0, \mu_1 \in \mathcal{P}_p(\mathbb{R}^d)$ define

$$\mathfrak{M}(\mu_0, \mu_1; t) = (\pi^t)_{\#}\sigma, \quad t \in [0, 1],$$

where σ is an optimal transport plan pushing μ_0 onto μ_1, and the map $\pi^t : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is given by $\pi^t(x, y) = (1 - t)x + ty$.

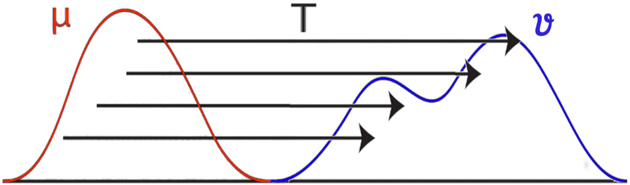
McCann's interpolants are the only constant-speed curves in the metric space $(\mathcal{P}_p(\mathbb{R}^d), W_p)$.

Tangent spaces

With the above notions, the tangent space at an absolutely continuous measure $\mu \in \mathcal{P}_p(\mathbb{R}^d)$ is realized as

$$\mathrm{Tan}_\mu = \mathrm{cl}_{L^p(\mathbb{R}^d;\mu)}\{s(T_\mu^\nu - I) \mid \nu \in \mathcal{P}_p(\mathbb{R}^d),\ s > 0\},$$

where $T_\mu^\nu : \mathbb{R}^d \to \mathbb{R}^d$ is an optimal transport mapping that pushes μ onto some arbitrary measure ν.



The $\ominus$ and $\oplus$ operators

For A.C. measures $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ the difference is defined via

$$\nu \ominus \mu = T_\mu^\nu - I,$$

where T_μ^ν is the optimal transport map, and I is the identity map. With this definition, we have that $\mu \ominus \mu = 0$ the zero map, and

$$\|\nu \ominus \mu\|_{L^p(\mathbb{R}^d;\mu)}^p = \int_{\mathbb{R}^d} \|T_\mu^\nu(x) - x\|^p d\mu(x) = W_p^p(\mu, \nu).$$

Moreover, the compatible $\oplus$ operator is defined via

$$\mu \oplus \psi = (\psi + I)_\# \mu,$$

for any Borel measurable map $\psi : \mathbb{R}^d \to \mathbb{R}^d$.

The $\ominus$ and $\oplus$ operators (for discrete measures)

For discrete probability measures $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ given by

$$\mu = \sum_{i=1}^{m} p_i^\mu \delta_{x_i^\mu} \quad \text{and} \quad \nu = \sum_{j=1}^{n} p_j^\nu \delta_{x_j^\nu},$$

the difference is defined via

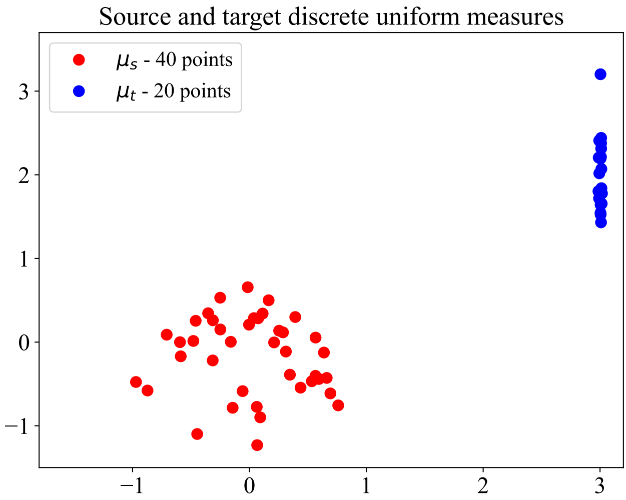
$$\nu \ominus \mu = \left(\left[x_j^\nu - x_i^\mu \right]_{i=1,\ldots,m}^{j=1,\ldots,n}, \ \Lambda_\mu^\nu \right),$$

where Λ_μ^ν is the coupling matrix between μ and ν. In this case, the compatible $\oplus$ operator is defined via

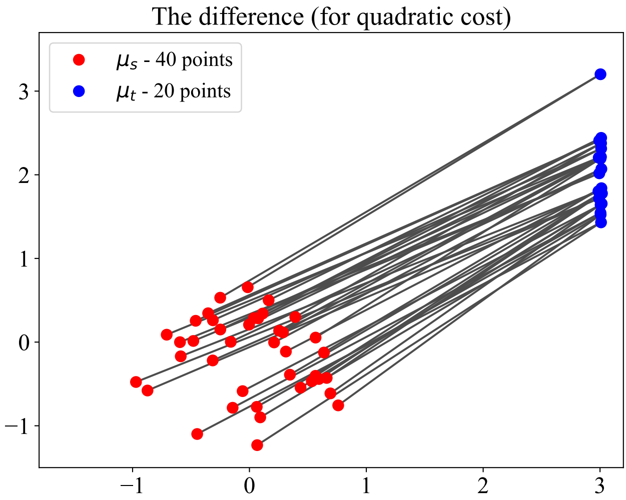
$$\mu \oplus \psi = \sum_{i=1}^{m} \sum_{j=1}^{k} \lambda_{i,j}^\psi \delta_{x_i^\mu + x_{i,j}^\psi},$$

for any tensor pair $\psi = (x^\psi, \Lambda^\psi)$ where $x \in \mathbb{R}^{m \times k \times d}$ and $\Lambda^\psi \in \mathbb{R}^{m \times k}$.

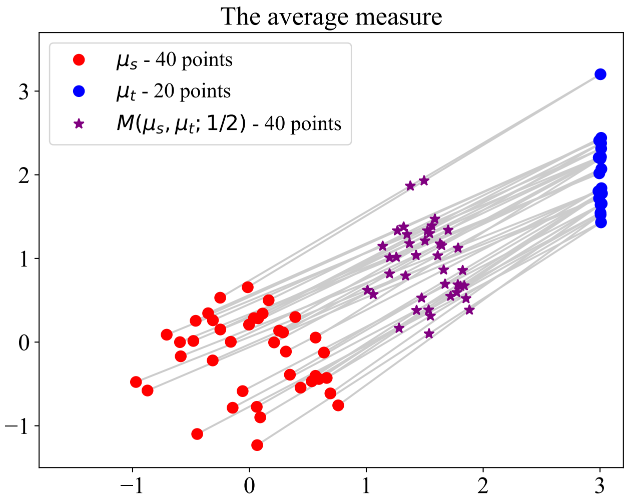
Operators on discrete measures (illustrations)



Operators on discrete measures (illustrations)



Operators on discrete measures (illustrations)



Refinements in Wasserstein spaces

We focus on multiscaling measures with the most elementary subdivision scheme $\mathcal{S}$ that is given by the rules

$$\begin{cases} (\mathcal{S}\boldsymbol{c})_{2k} = c_k, \\ (\mathcal{S}\boldsymbol{c})_{2k+1} = \frac{1}{2}c_k + \frac{1}{2}c_{k+1}, \end{cases} \qquad k \in \mathbb{Z},$$

for a given $\mathbb{R}$-valued sequence $\boldsymbol{c}$. Therefore, for a $\mathcal{P}_p(\mathbb{R}^d)$-valued sequence $\boldsymbol{\mu}$, the adaptation of $\mathcal{S}$ becomes

$$\begin{cases} (\mathcal{S}\boldsymbol{\mu})_{2k} = \mu_k, \\ (\mathcal{S}\boldsymbol{\mu})_{2k+1} = \mathfrak{M}(\mu_k, \mu_{k+1}; 1/2), \end{cases} \qquad k \in \mathbb{Z}.$$

Outline

Theoretical results

Multiscaling a sequence $\boldsymbol{\mu}^{(J)}$ in $\mathcal{P}_p(\mathbb{R}^d)$ with the above notations via the interpolating elementary $\mathcal{S}$ becomes straightforward.

Theorem (decay of detail coefficients)

Assume $\boldsymbol{\mu}^{(J)}$ is sampled over the dyadic grid parametrization $2^{-J}\mathbb{Z}$ from an absolutely continuous curve $\boldsymbol{\mu}$ in $\mathcal{P}_p(\mathbb{R}^d)$ with a finite metric derivative $\Gamma = \sup_{t\in\mathbb{R}} |\mu'|_t < \infty$. Then, the resulting detail coefficients $\psi^{(\ell)}$ satisfy

$$\|\boldsymbol{\psi}^{(\ell)}\|_\infty \leq \Gamma 2^{1-\ell}, \quad \ell = 1, \ldots, J,$$

where

$$\|\boldsymbol{\psi}^{(\ell)}\|_\infty = \sup_{k\in\mathbb{Z}} \|\psi_k^{(\ell)}\|_{L^p(\mu_k^{(\ell)})}.$$

Theoretical results

The refinement $\mathcal{S}$ is called stable if there exists $K > 0$ such that

$$\mathcal{W}_p(\mathcal{S}\boldsymbol{\mu}, \mathcal{S}\boldsymbol{\nu}) \leq K\, \mathcal{W}_p(\boldsymbol{\mu}, \boldsymbol{\nu})$$

for any sequences $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ where $\mathcal{W}_p(\boldsymbol{\mu}, \boldsymbol{\nu}) = \sup_{k\in\mathbb{Z}} W_p(\mu_k, \nu_k)$.

Theorem (stability of reconstruction)

Let $\{\boldsymbol{\mu}^{(0)}; \boldsymbol{\psi}^{(1)}, \dots, \boldsymbol{\psi}^{(J)}\}$ and $\{\widetilde{\boldsymbol{\mu}}^{(0)}; \widetilde{\boldsymbol{\psi}}^{(1)}, \dots, \widetilde{\boldsymbol{\psi}}^{(J)}\}$ be two pyramid representations of $\boldsymbol{\mu}^{(J)}$ and $\widetilde{\boldsymbol{\mu}}^{(J)}$, respectively. Assume that $\|\psi_k^{(\ell)}\|_{\text{Lip}} \leq C$ for all $\ell = 1, \dots, J$ and $k \in \mathbb{Z}$. If $\mathcal{S}$ is stable with constant K, then for $L = \max\{1, (KC)^J\}$ we have

$$\mathcal{W}_p(\boldsymbol{\mu}^{(J)}, \widetilde{\boldsymbol{\mu}}^{(J)}) \leq L\left(\mathcal{W}_p(\boldsymbol{\mu}^{(0)}, \widetilde{\boldsymbol{\mu}}^{(0)}) + \sum_{\ell=1}^{J} \|\boldsymbol{\psi}^{(\ell)} - \widetilde{\boldsymbol{\psi}}^{(\ell)}\|_\infty\right).$$

Outline

Denoising

Multiscaling A.C. measures in the Wasserstein space $\mathcal{P}_2(\mathbb{R})$.

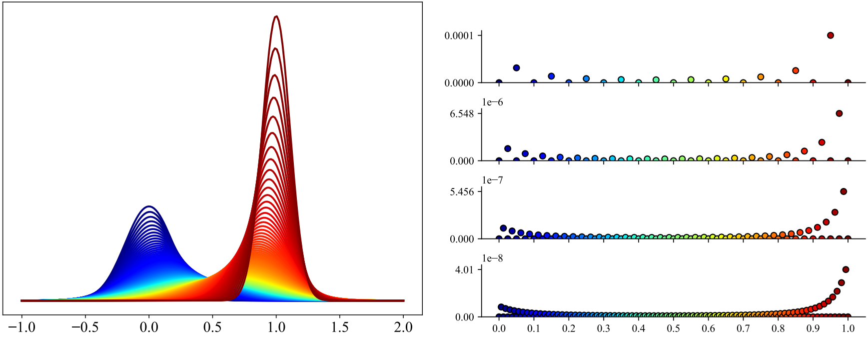


Figure: Sequence of Gaussian measures and its multiscale transform.

Denoising

Multiscaling A.C. measures in the Wasserstein space $\mathcal{P}_2(\mathbb{R})$.

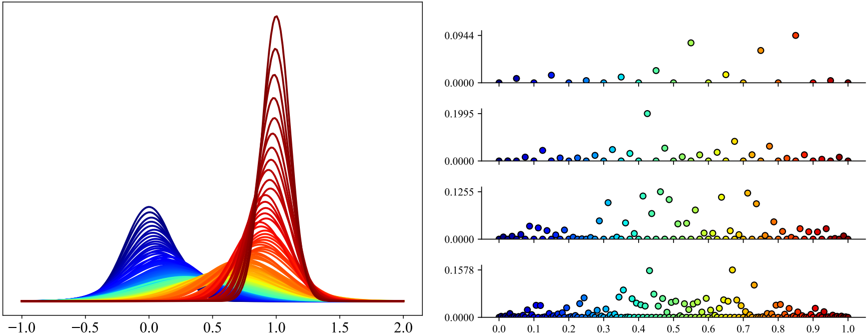


Figure: Contamination with noise.

Denoising

Multiscaling A.C. measures in the Wasserstein space $\mathcal{P}_2(\mathbb{R})$.

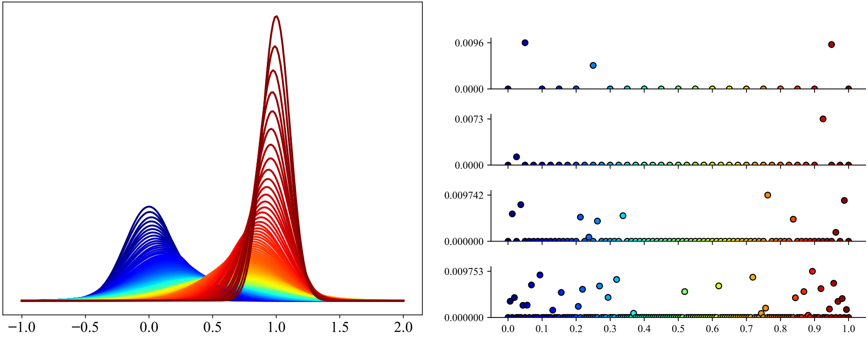


Figure: Denoising result obtained by thresholding with 0.01.

Anomaly detection

Multiscaling A.C. measures in the Wasserstein space $\mathcal{P}_2(\mathbb{R})$.

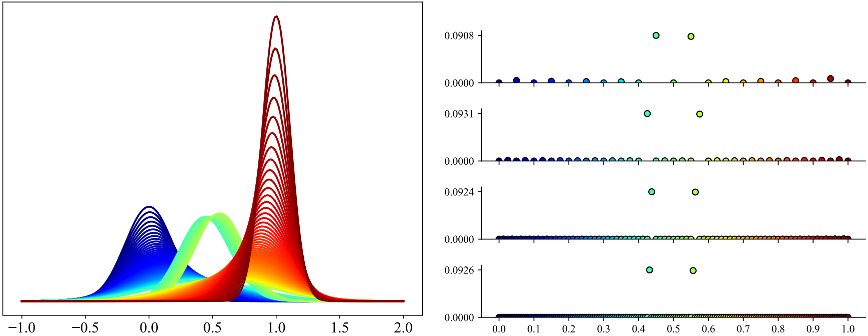


Figure: Detecting jump discontinuities via multiscaling.

Analyzing point cloud dynamics

Multiscaling discrete measures in the Wasserstein space $\mathcal{P}_2(\mathbb{R}^2)$.

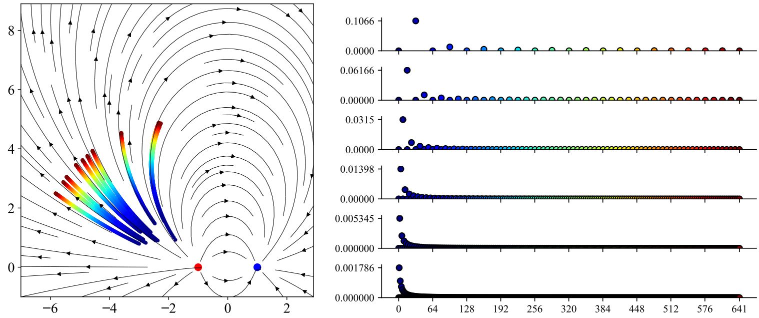


Figure: Multiscaling the evolution of 10 atoms with uniform distribution according to the electric field induced by a positive charge located at $(-1, 0)$ and a negative one at $(1, 0)$. Large detail coefficients correspond to regions with large vector fields. Decay in details imply smoothness.

Analyzing point cloud dynamics

Multiscaling discrete measures in the Wasserstein space $\mathcal{P}_2(\mathbb{R}^2)$.

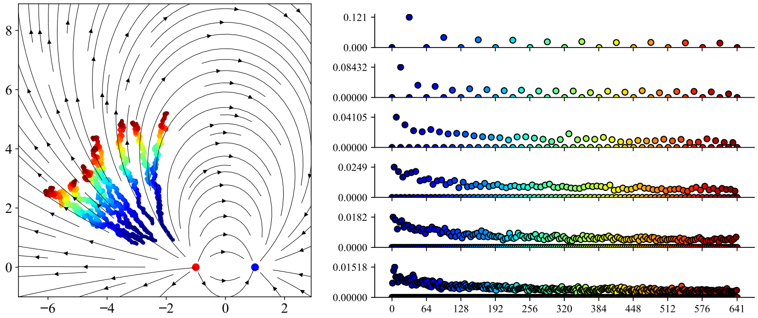


Figure: Adding synthetic noise to the evolution of the point cloud and unveiling the noise values via multiscaling. The large magnitude of detail norms on high scales indicate that the analyzed sequence does not vary smoothly.

Analyzing NN learning dynamics

Multiscaling discrete measures in the Wasserstein space $\mathcal{P}_2(\mathbb{R})$.

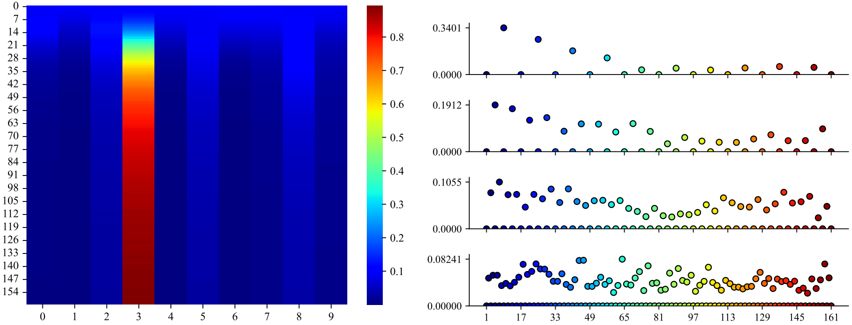


Figure: Analyzing the learning dynamics of a simple neural network on MNIST dataset. On the left, the prediction of the digit “3” across epoch iterations. On the right, the multiscale transform of the resulted measure sequence. The convergence is clear on coarse scales.

Thank you for listening!
Questions?