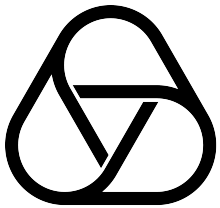


Mathematisches
Forschungsinstitut
Oberwolfach

Multiscale Analysis and Sparse Approximations of Manifold-Valued Data

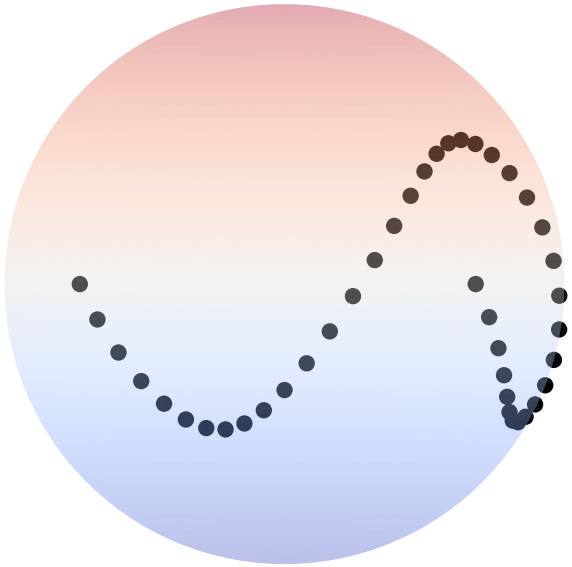
Wael Mattar
PhD student at Tel Aviv University

Workshop on *Approximation of Manifold-Valued Functions*
Oberwolfach, Germany

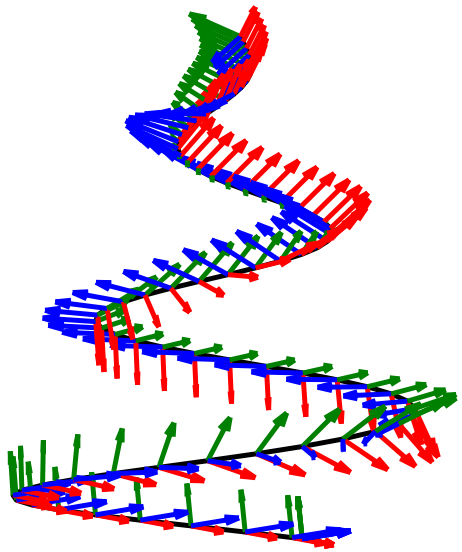
Outline

Outline

Motivation (univariate manifold data)



Points on the unit sphere $\mathbb{S}^2$.



Visualization of a trajectory in the special Euclidean group.

Motivation (bivariate manifold data)

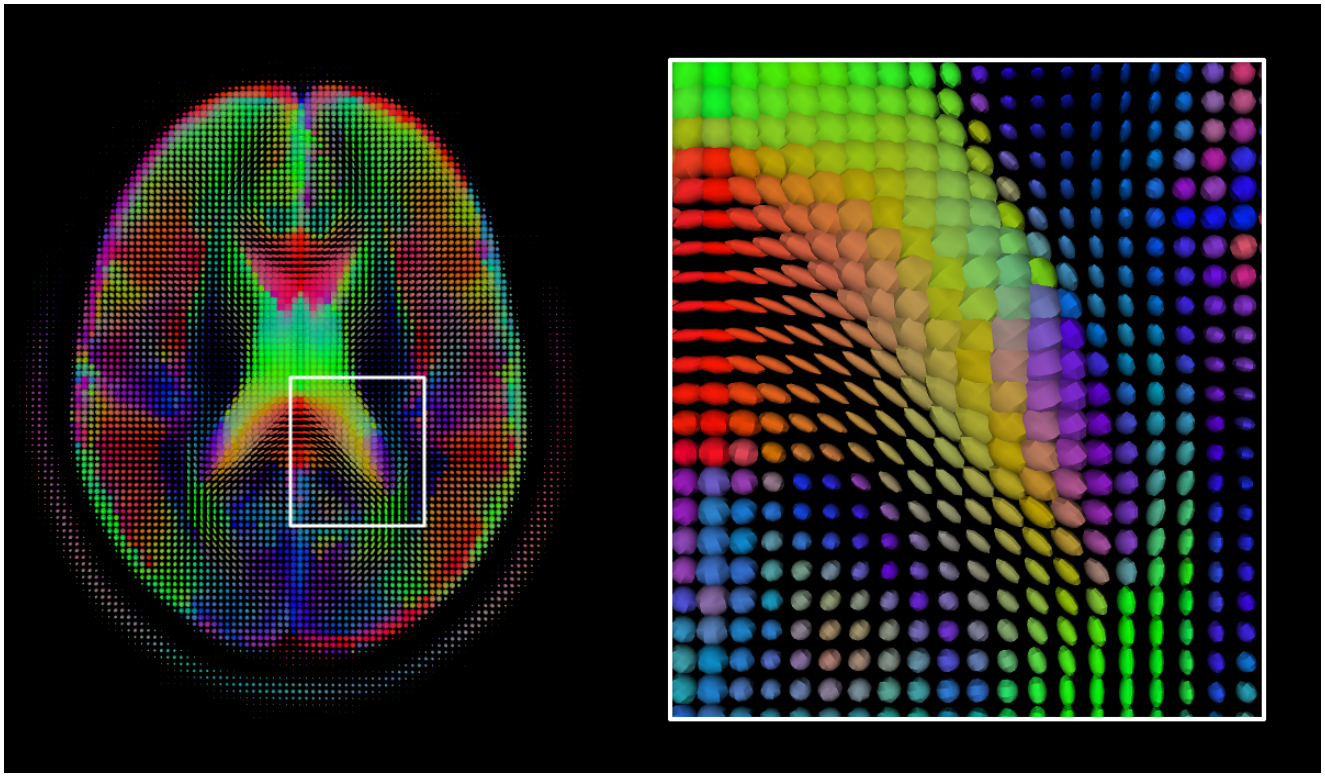


Figure: a 2D slice of Diffusion Tensor Imaging (DTI) consisting of 3×3 symmetric positive definite matrices. Source: Medical imaging, Wikipedia.

Multiscale decomposition

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

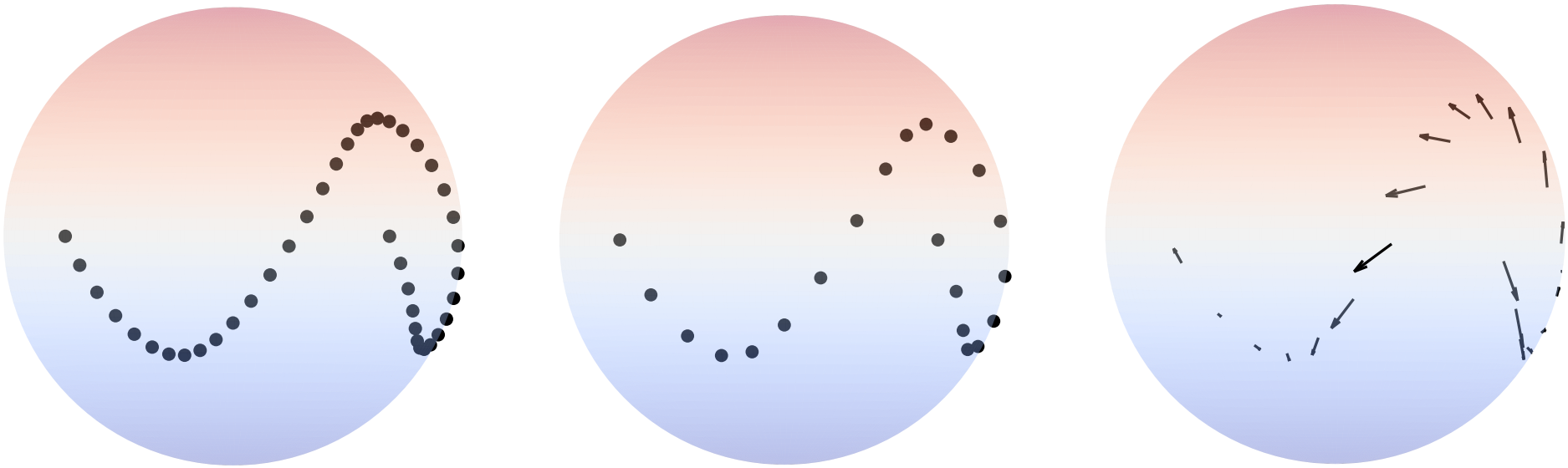
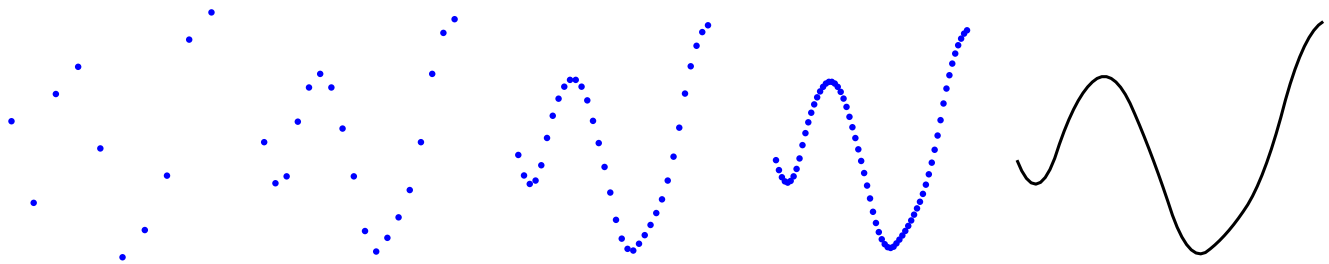


Figure: Sketch of multiscale decomposition of $\mathbb{S}^2$-valued sequence.

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.)

A necessary condition for the convergence of a subdivision scheme $\mathcal{S}_{\boldsymbol{\alpha}}$ is the constant-reproduction property

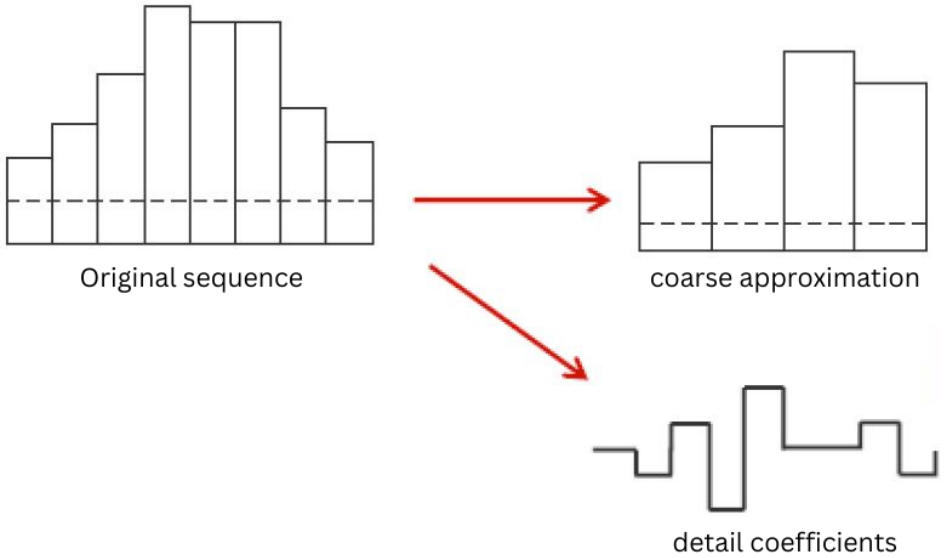
$$\sum_{j\in\mathbb{Z}} \alpha_{2j} = \sum_{j\in\mathbb{Z}} \alpha_{2j+1} = 1.$$

The refinement is called *interpolating* if $\alpha_0 = 1$. An essential tool for analyzing the convergence of $\mathcal{S}_{\boldsymbol{\alpha}}$ is the z-transform of $\boldsymbol{\alpha}$, that is defined by

$$\boldsymbol{\alpha}(z) = \sum_{j\in\mathbb{Z}} \alpha_j z^j, \quad z \in \mathbb{C},$$

and termed the *symbol* – becoming a complex Fourier series on the unit circle $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$.

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.

Multiscaling via interpolating subdivision scheme

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.

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

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

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

while the inverse multiscale transform is given by

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

The non-interpolating case

Iterating the multiscale transform with a non-interpolating subdivision scheme $\mathcal{S}_\alpha$ and the elementary downsampling operator $\mathcal{D}$ does not(!) necessarily give

$$d_{2k}^{(\ell)} = 0, \quad k \in \mathbb{Z}, \quad \ell = 1, \ldots, J.$$

The non-interpolating case

Iterating the multiscale transform with a non-interpolating subdivision scheme $\mathcal{S}_{\boldsymbol{\alpha}}$ and the elementary downsampling operator $\mathcal{D}$ does not(!) necessarily give

$$d_{2k}^{(\ell)} = 0, \quad k \in \mathbb{Z}, \quad \ell = 1, \ldots, J.$$

We solve this problem by replacing $\mathcal{D}$ with the more general decimation operator. For a sequence $\boldsymbol{\gamma} \in \ell_1(\mathbb{Z})$, the operator

$$(\mathcal{D}_{\boldsymbol{\gamma}} \boldsymbol{c})_k = \sum_{j \in \mathbb{Z}} \gamma_{k-j} c_{2j}$$

gives zero even detail coefficients in multiscaling if

$$\boldsymbol{\gamma} * (\mathcal{D}\boldsymbol{\alpha}) = \boldsymbol{\delta}$$

where $\boldsymbol{\delta}$ is the Kronecker delta sequence.

Reversing

Under the z-transform, the convolution equation becomes

$$\gamma(z)(\mathcal{D}\boldsymbol{\alpha})(z) = 1, \quad z \in \mathbb{C}.$$

Wiener's lemma

In the Banach algebra of continuous functions

$$\mathcal{A}(\mathbb{T}) = \left\{ f(t) = \sum_{j\in\mathbb{Z}} a_j e^{int},\ t \in [0, 2\pi) \mid \boldsymbol{a} \in \ell_1(\mathbb{Z}) \right\},$$

if $f \in \mathcal{A}(\mathbb{T})$ does not vanish on $\mathbb{T}$, then $1/f(t)$ is also in $\mathcal{A}(\mathbb{T})$.

Reversing (cont.)

In case a reverse $\gamma \in \ell_1(\mathbb{Z})$ exists, then there exist constants $C(\kappa) > 0$ and $\lambda(\kappa) \in (0, 1)$ such that

$$|\gamma_k| \leq C\lambda^{|k|}, \quad k \in \mathbb{Z},$$

where κ is the reversibility condition number given by

$$\kappa = \frac{\sup_{z\in\mathbb{T}} |(\mathcal{D}\alpha)(z)|}{\inf_{z\in\mathbb{T}} |(\mathcal{D}\alpha)(z)|} \in [1, \infty].$$

Reversing (cont.)

In case a reverse $\boldsymbol{\gamma} \in \ell_1(\mathbb{Z})$ exists, then there exist constants $C(\kappa) > 0$ and $\lambda(\kappa) \in (0, 1)$ such that

$$|\gamma_k| \leq C\lambda^{|k|}, \quad k \in \mathbb{Z},$$

where κ is the reversibility condition number given by

$$\kappa = \frac{\sup_{z\in\mathbb{T}} |(\mathcal{D}\boldsymbol{\alpha})(z)|}{\inf_{z\in\mathbb{T}} |(\mathcal{D}\boldsymbol{\alpha})(z)|} \in [1, \infty].$$

However, a reverse does not always exist! For example, least-squares subdivision schemes are usually **irreversible**!

Pseudo-reversing

Let $p(z)$ be a polynomial of degree n satisfying $p(1) = 1$, and denote by Λ the set of its roots including multiplicities. By the complete factorization theorem we rewrite p as

$$p(z) = C(p) \prod_{r \in \Lambda} (z - r).$$

Definition (pseudo-reversing)

The pseudo-reverse of p is defined by

$$p_\xi^\dagger(z) = \left(C(p_\xi^\dagger) \prod_{r \in \Lambda \setminus \mathbb{T}} (z - r) \prod_{r \in \Lambda \cap \mathbb{T}} \left(z - (1 + \xi) r\right) \right)^{-1}$$

for some $\xi > 0$ where $C(p_\xi^\dagger)$ is chosen such that $p_\xi^\dagger(1) = 1$.

Example & properties

Example: Consider $p(z) = (z^2 + z + 1)/3$ which vanishes for $z = 1/2 \pm i\sqrt{3}/2 \in \mathbb{T}$. The pseudo-reverse of p is

$$p_\xi^\dagger(z) = \frac{3 + 3\xi + \xi^2}{z^2 + z(1+\xi) + (1+\xi)^2}, \quad \xi > 0.$$

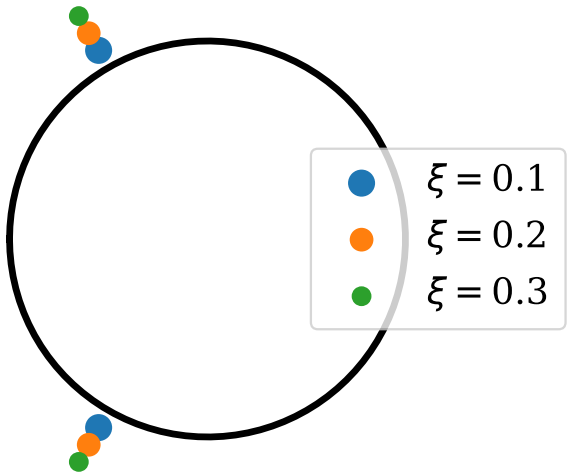
Some properties:

1. The product $p(z)p_\xi^\dagger(z)$ converges in $\mathcal{A}$-norm (the ℓ_1 norm of coefficients) to the constant 1 as $\xi \to 0^+$.
2. The function $p_\xi^\dagger(z)$ converges to 1 as $\xi \to \infty$ on every compact subset of $\mathbb{C}$, provided $\Lambda \subset \mathbb{T}$. Moreover,
3. the reversibility condition number κ of $p_\xi^{-\dagger}(z)$ satisfies

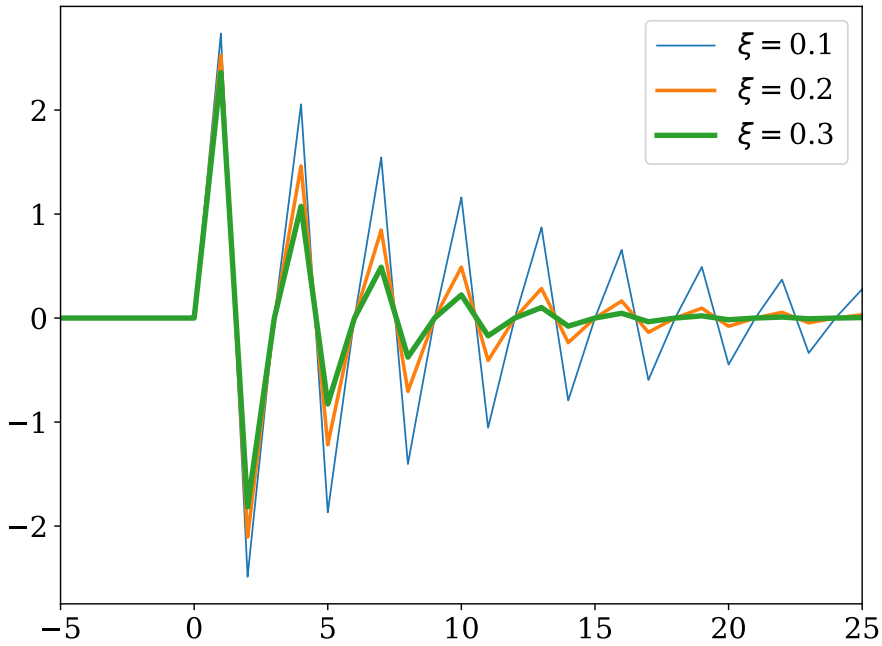
$$\kappa \leq (1 + 2/\xi)^n.$$

Illustration

Pseudo-reversing a least-squares-based subdivision scheme with $\boldsymbol{\alpha} = 1/12\ [3,\ 4,\ 3,\ 4,\ 3,\ 4,\ 3]$ supported on $[-3, 3] \cap \mathbb{Z}$.



(a) Roots displacement



(b) Pseudo-reverse coefficients

Outline

Adaptations to Riemannian manifolds

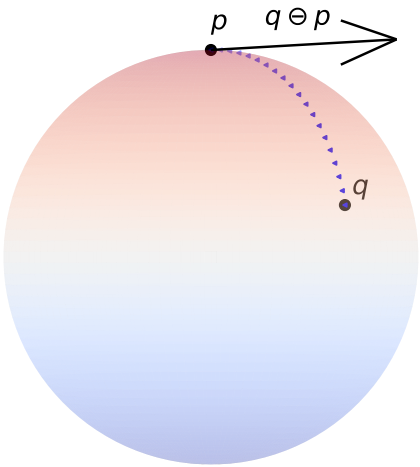
Let $\mathcal{M}$ be a Riemannian manifold associated with the geodesic distance metric ρ defined by

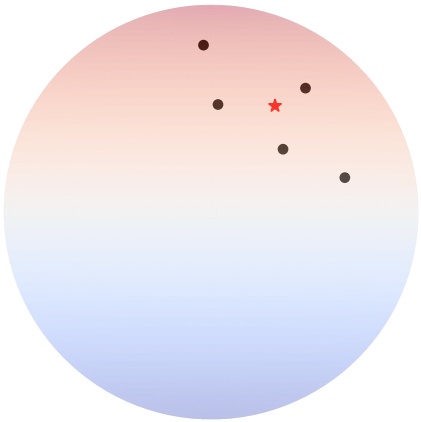
$$\rho(x, y) = \inf_{\Gamma} \int_a^b \|\dot{\Gamma}(t)\| dt,$$

where $\Gamma : [a, b] \to \mathcal{M}$ is a piecewise $\mathcal{C}^1$ curve connecting the points $x = \Gamma(a)$ and $y = \Gamma(b)$ with the induced norm $\|\cdot\|^2 = \langle \cdot, \cdot \rangle$.

Adaptations to Riemannian manifolds

Let $(\mathcal{M}, \rho)$ be a Riemannian manifold.





Adaptations of "$-$" and "$+$":

$$q \ominus p = \mathrm{Log}_p(q) \in T_p\mathcal{M}$$

$$p \oplus v = \mathrm{Exp}_p(v) \in \mathcal{M}$$

$$\boxed{p \oplus (q \ominus p) = q}$$

Euclidean CoM $x^* = \sum_{j=1}^n \beta_j x_j$ is generalized via:

$$x^* \in \mathrm{argmin}_{x \in \mathcal{M}} \sum_{j=1}^n \beta_j \rho^2(x, x_j)$$

Representing a curve on $\mathbb{S}^2$

Multiscaling with the cubic B-spline subdivision scheme.

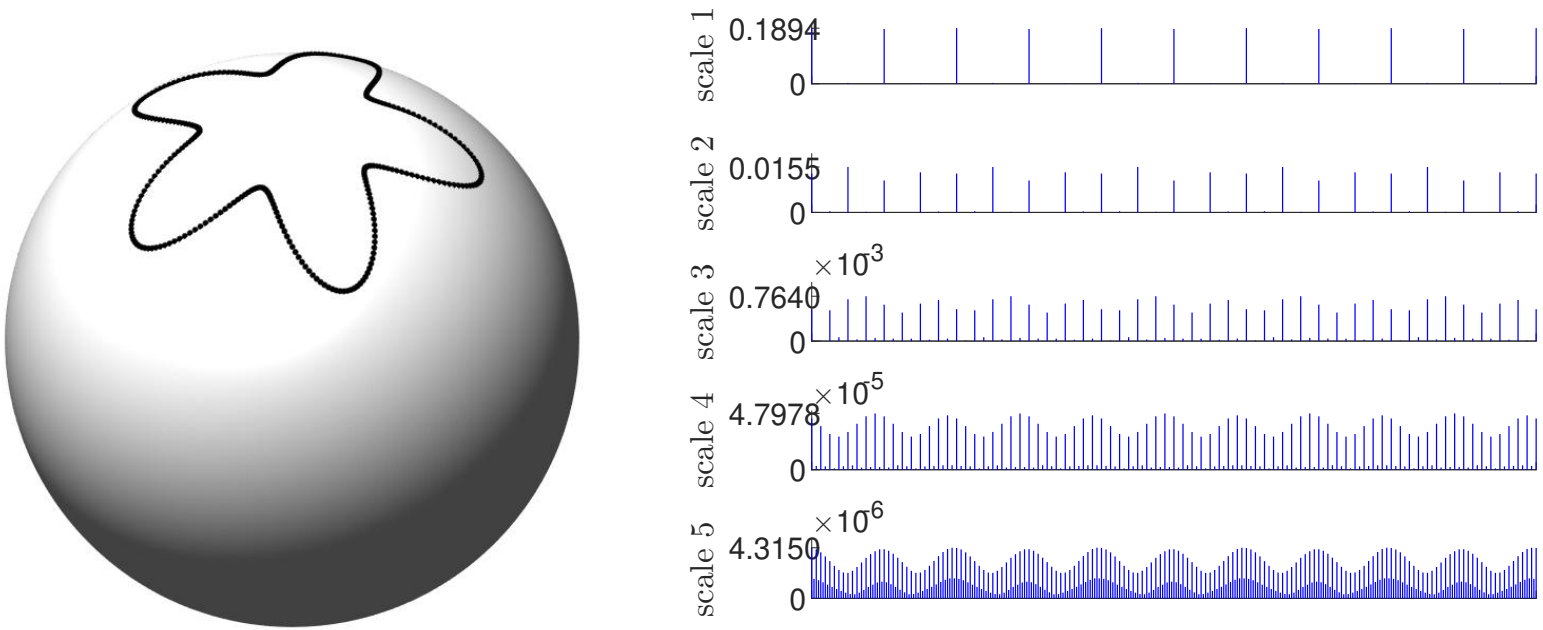


Figure: A smooth $\mathbb{S}^2$-valued curve with its corresponding multiscale representation. The decay of details indicate the smoothness of the curve.

Representing a curve on $\mathbb{S}^2$

Contaminating the sequence with noise.

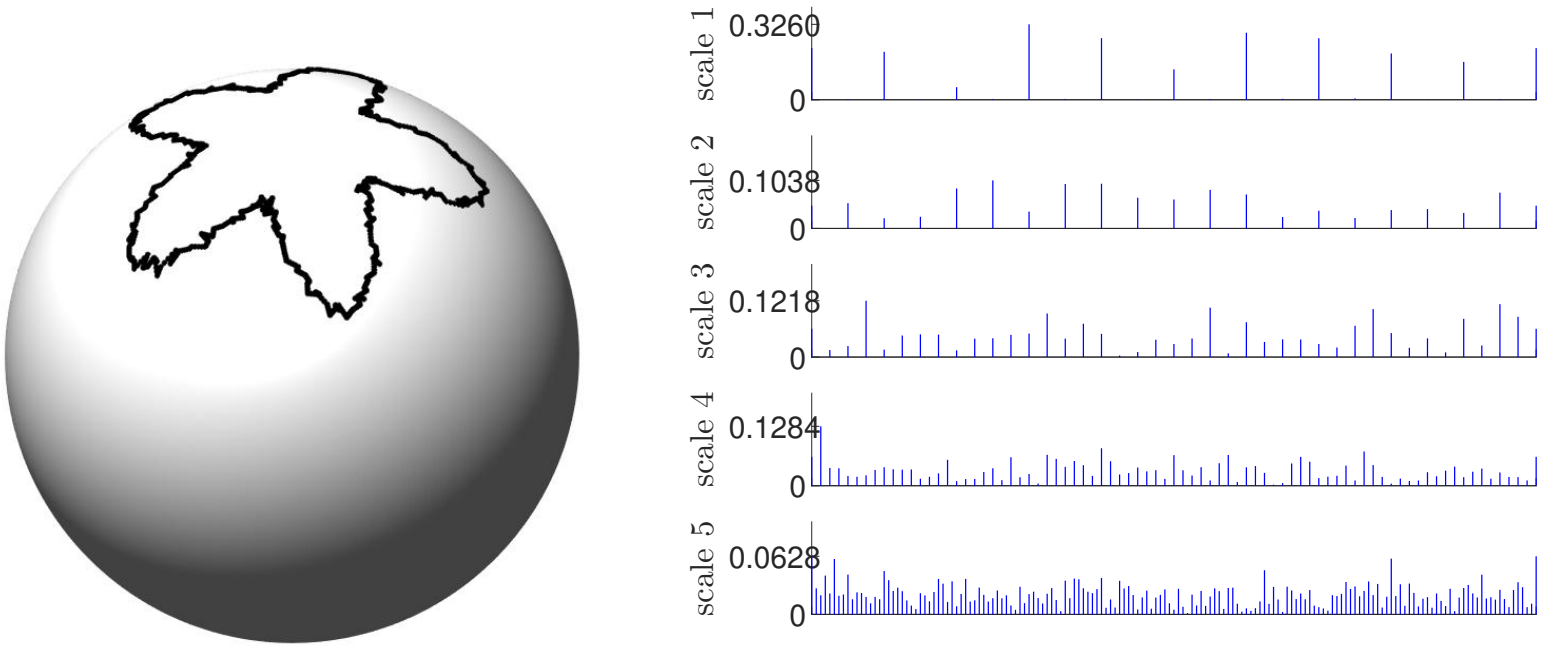


Figure: The curve contaminated with additive noise and its corresponding representation. Large norms on high scales indicating non-smoothness.

Representing a curve on $\mathbb{S}^2$

Denoising effect by thresholding.

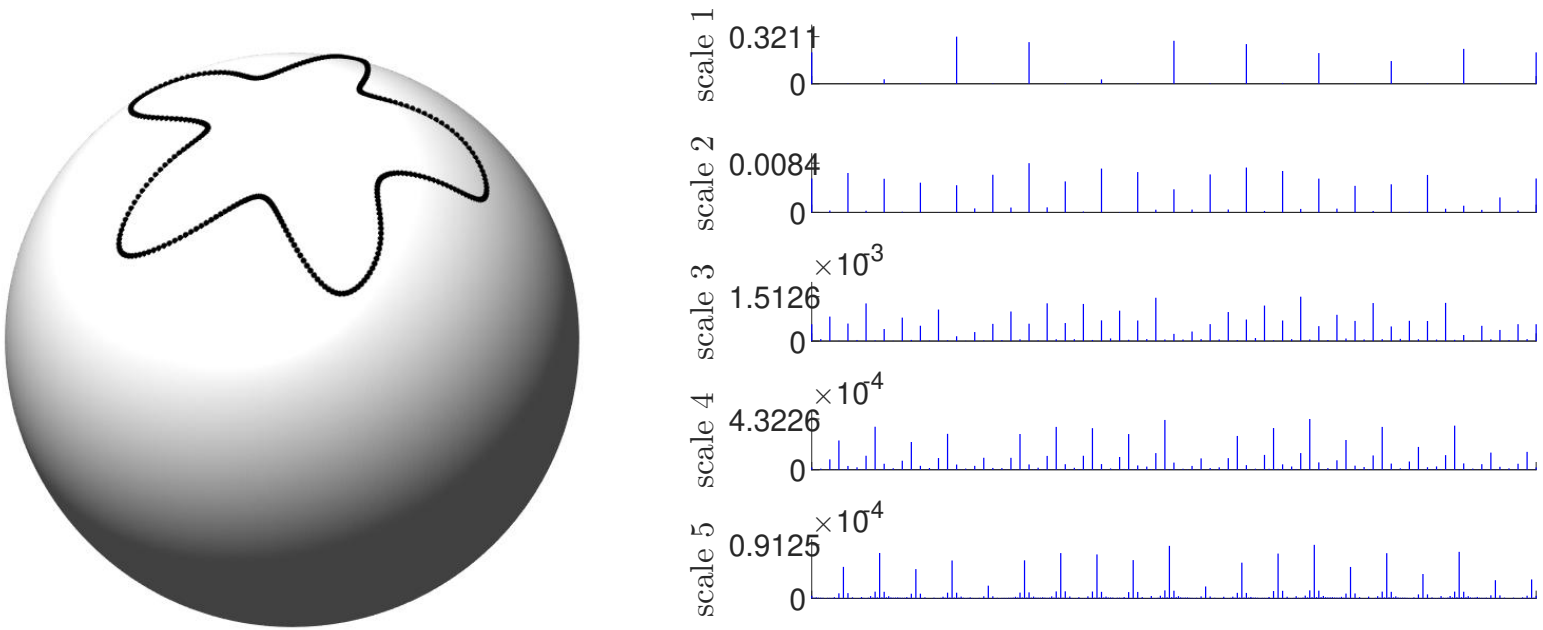


Figure: The synthesized curve obtained by the resulted sparse pyramid representation.

Anomaly Detection in a time series

Let $\mathcal{M} = \mathcal{SPD}(3)$ be the cone of symmetric positive definite matrices. The data is visualized by centric ellipsoids:

The first $\mathcal{SPD}(3)$-valued sequence

The second $\mathcal{SPD}(3)$-valued sequence – with two “jump” points

Anomaly Detection in a time series

Multiscaling with the corner-cutting subdivision scheme.

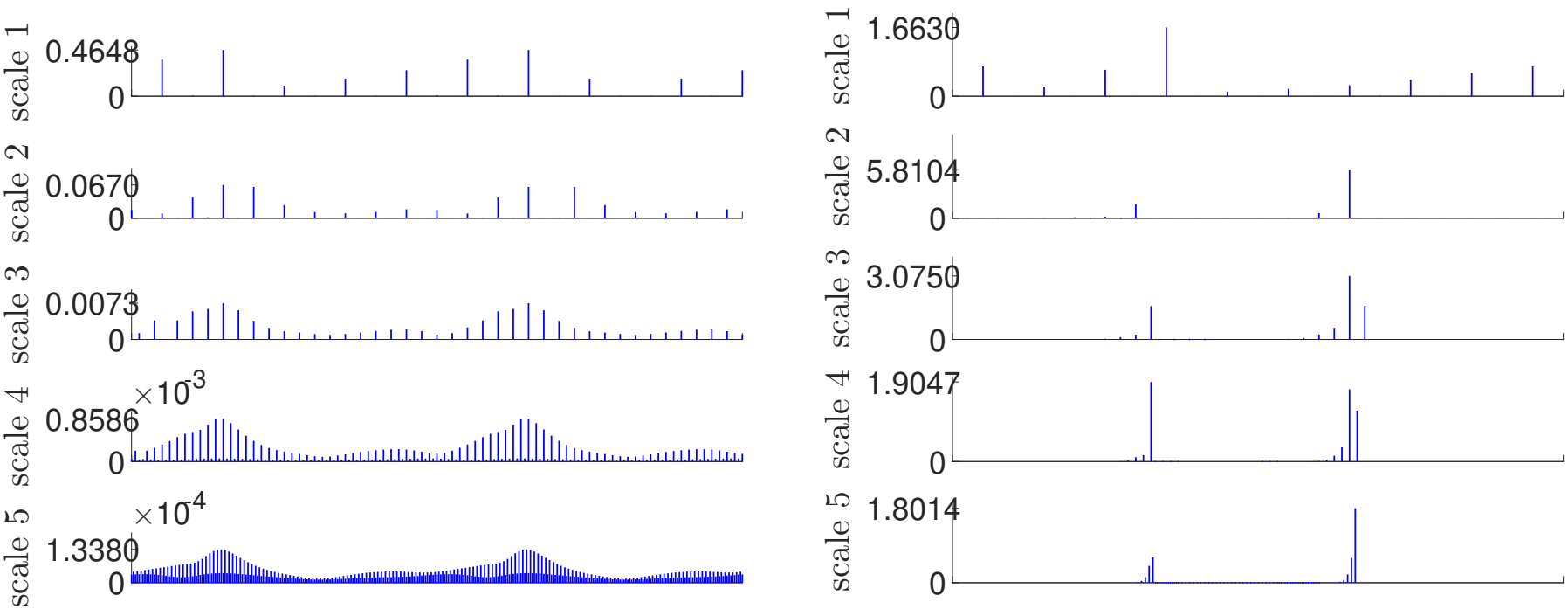


Figure: Frobenius norms of the details coefficients. The jump locations are clearly seen via the large magnitude detail coefficients.

Contrast enhancement

The manifold of interest is $\mathcal{M} = \mathcal{SO}(3)$.

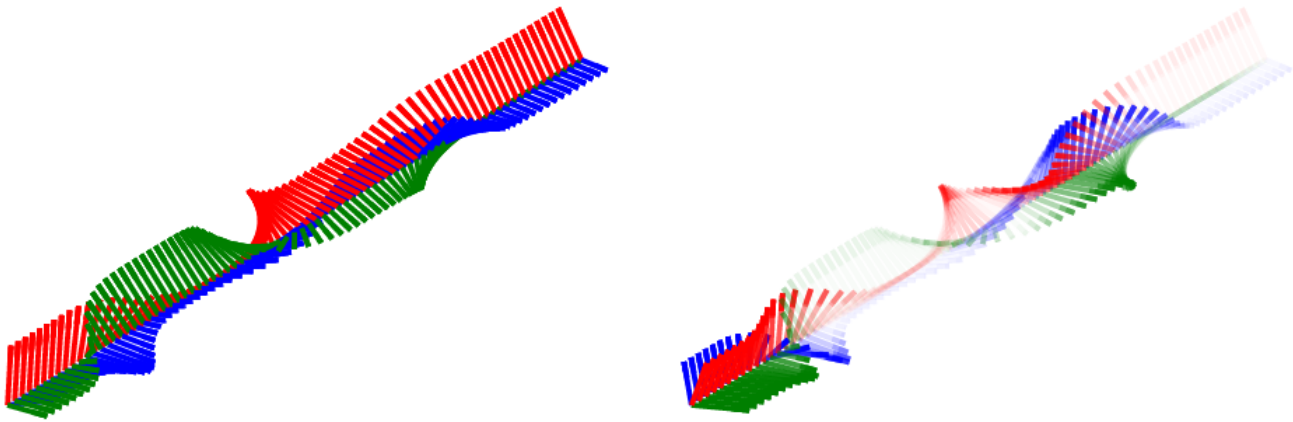


Figure: Contrast enhancement of a sequence of rotation matrices. On the left, illustration of the original sequence applied to the standard basis of $\mathbb{R}^3$. On the right, the enhanced sequence obtained by scaling the top 20% of detail coefficients with a factor of 40%.

Data compression

The manifold of interest is $\mathcal{M} = \mathcal{SO}(3) \ltimes \mathbb{R}^3$.

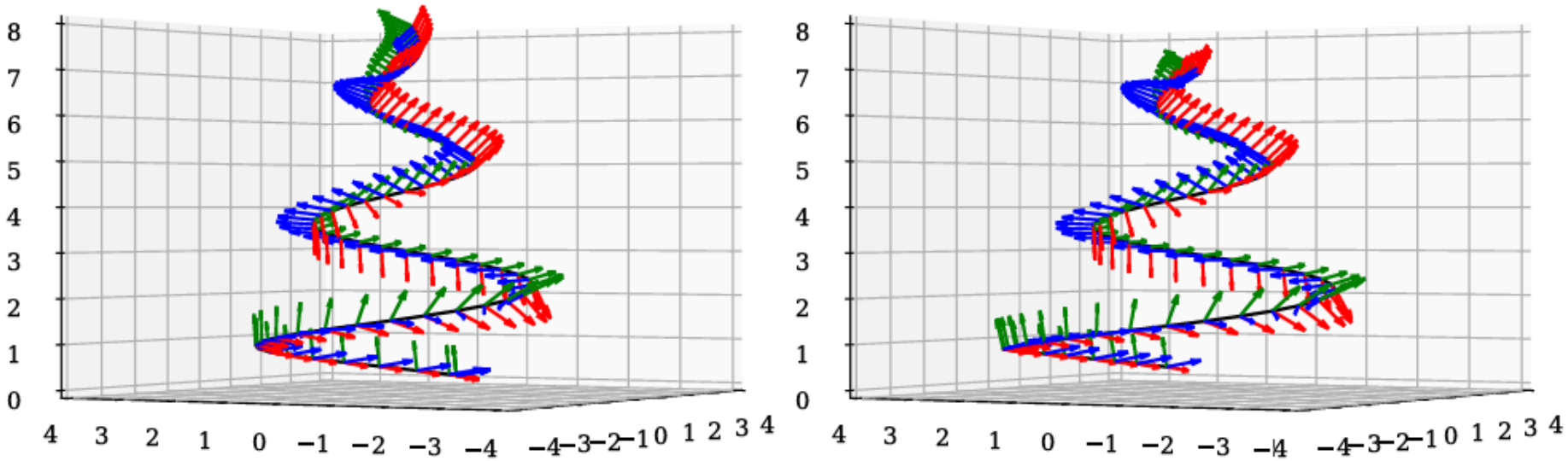


Figure: Data compression of a rigid body motion. On the left, illustration of 641 matrices representing the special Euclidean sequence. On the right, the decompressed trajectory that is obtained by 41 matrices in addition to 12 detail coefficients.

Outline

Bivariate multiscale transform

Multivariate multiscale transforms can be constructed by applying tensor products to α and γ.

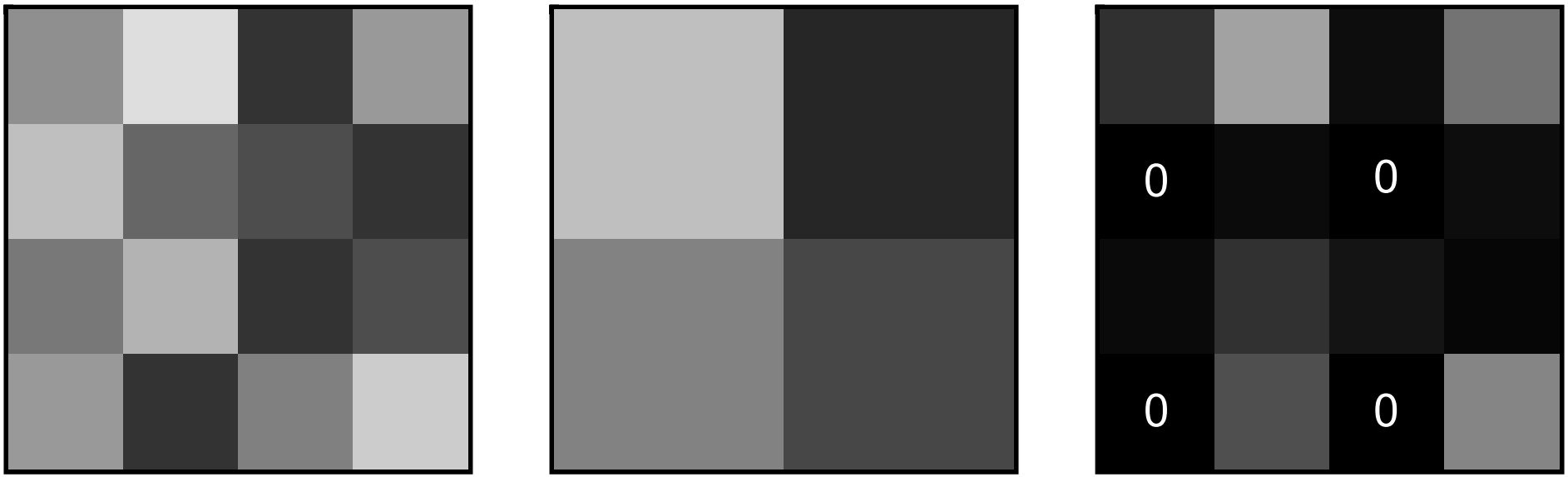


Figure: Sketch for decomposing a 4×4 grayscale image.

Multiscaling an image

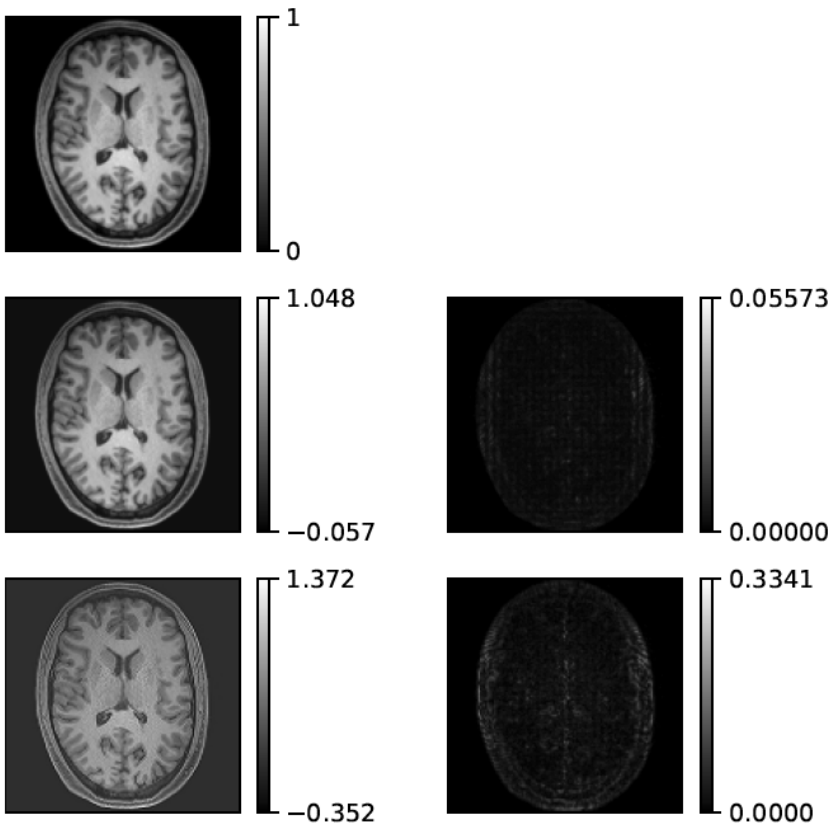


Figure: Multiscaling an image of the author's own brain (MRI).

Sparse approximations

The variational formulation[1] of the sparsity problem is as follows. We consider data $\boldsymbol{c}^{(J)} \in \mathcal{M}^K$ where $K = (K_1, \ldots, K_s) \in \mathbb{N}^s$ and its multiscale representation $\{\boldsymbol{c}^{(0)}; \boldsymbol{d}^{(1)}, \ldots, \boldsymbol{d}^{(J)}\}$. We solve for

$$\text{argmin}_{\boldsymbol{u}^{(J)} \in \mathcal{M}^K} \left(\text{dist}\left(\boldsymbol{u}^{(J)}, \boldsymbol{c}^{(J)}\right)^p + \mathcal{W}_\lambda^{\beta,q}(\boldsymbol{u}^{(J)}) \right)$$

where $p, q \geq 1$ and $\beta \geq 0$ is a smoothness parameter,

Sparse approximations

The variational formulation[1] of the sparsity problem is as follows. We consider data $\boldsymbol{c}^{(J)} \in \mathcal{M}^K$ where $K = (K_1, \ldots, K_s) \in \mathbb{N}^s$ and its multiscale representation $\{\boldsymbol{c}^{(0)}; \boldsymbol{d}^{(1)}, \ldots, \boldsymbol{d}^{(J)}\}$. We solve for

$$\operatorname{argmin}_{\boldsymbol{u}^{(J)} \in \mathcal{M}^K} \left(\operatorname{dist}\left(\boldsymbol{u}^{(J)}, \boldsymbol{c}^{(J)}\right)^p + \mathcal{W}_\lambda^{\beta,q}(\boldsymbol{u}^{(J)}) \right)$$

where $p, q \geq 1$ and $\beta \geq 0$ is a smoothness parameter, while

$$\operatorname{dist}\left(\boldsymbol{u}^{(J)}, \boldsymbol{c}^{(J)}\right)^p = \sum_{i_1=0}^{K_1-1} \cdots \sum_{i_s=0}^{K_s-1} \rho(u^{(J)}_{i_1,\ldots,i_s}, c^{(J)}_{i_1,\ldots,i_s})^p,$$

and the sparsity regularization term

$$\mathcal{W}_\lambda^{\beta,q}(\boldsymbol{u}^{(J)}) = \lambda_1 \sum_{n,\ell} 2^{\ell q \left(\beta + \frac{s}{2} - \frac{s}{q}\right)} \|d_n^{(\ell)}\|^q + \lambda_2 \sum_{n,i} \rho(u^{(0)}_{n+e_i}, u^{(0)}_n)^q.$$

Sparse approximations

Few comments:

1. For $i = 1, \ldots, s$, the notation e_i in the regularization term is the ith unit vector $(\delta_{ij})_{1 \leq j \leq s}$.
2. For $q = 1$ the problem becomes the manifold analogue of the LASSO (or ℓ_1) regularization problem.
3. The ℓ_0 regularization problem is formulated with

$$\mathcal{W}_\lambda^0(\boldsymbol{u}^{(J)}) = \lambda_1 \#\{(n, \ell) \mid d_n^{(\ell)} \neq 0\} + \lambda_2 \#\{(n, i) \mid u_n^{(0)} \neq u_{n+e_i}^{(0)}\}.$$

4. The problem has been studied[1] exclusively for interpolating multiscale transforms.

[1] **Wavelet Sparse Regularization for Manifold-Valued Data.**

Martin Storath and Andreas Weinmann. Multiscale Modeling & Simulation, 2020.

Numerical solutions

We borrow two results from [1] to illustrate the solution of the sparsity optimization problem. Denoising on $\mathcal{M} = \mathbb{S}^1$.

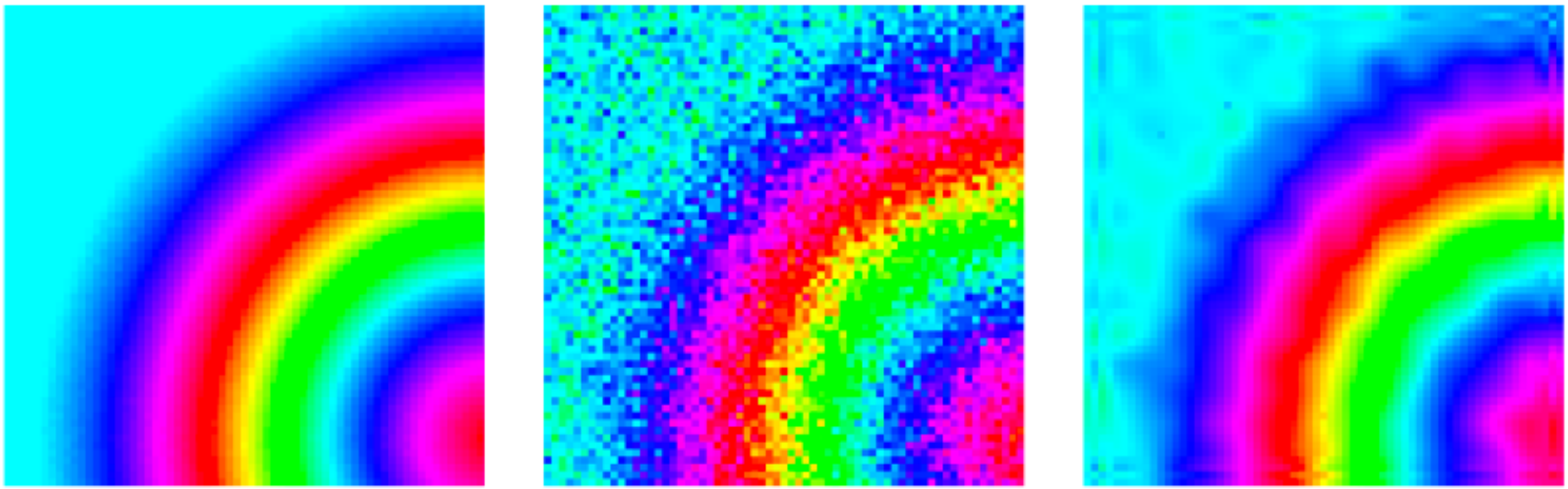


Figure 5: Denoising of a synthetic S^1-valued image. *Left:* Original, *Center:* Noisy data, *Right:* Denoising using ℓ^1 regularization and the DD wavelet (ΔSNR: 10.9 dB).

Numerical solutions (cont.)

Denoising on $\mathcal{M} = \mathcal{SPD}(3)$.

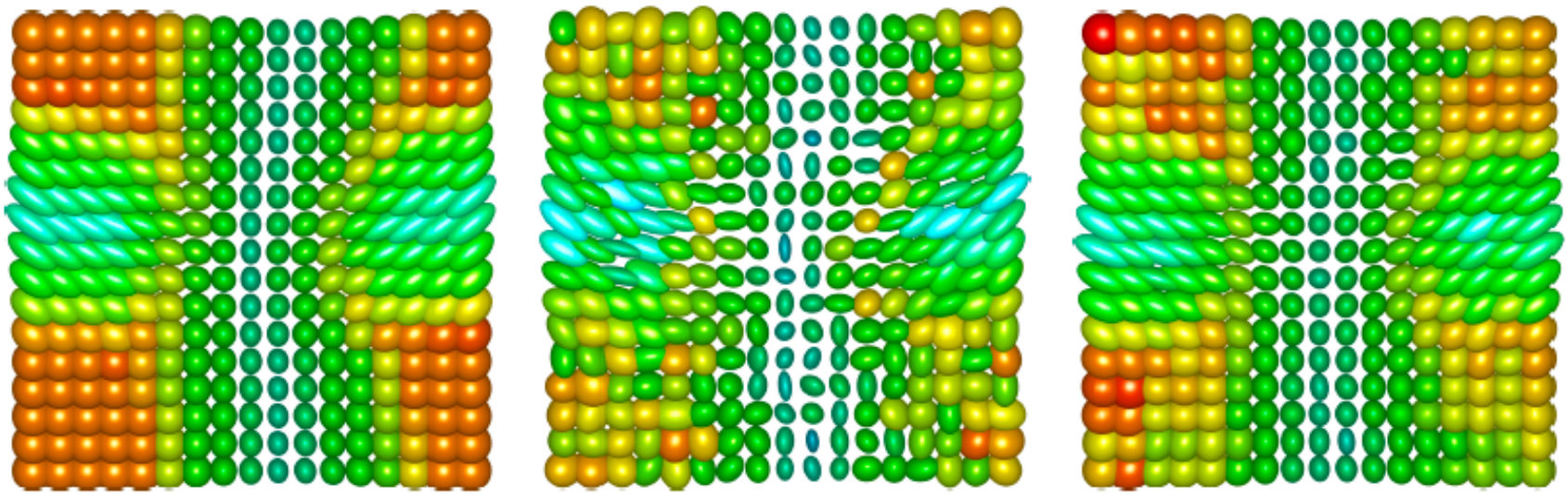


Figure 6: Denoising of a synthetic Pos_3-valued image. *Left:* Original, *Center:* Noisy data, *Right:* Denoising using ℓ^1 regularization and the DD wavelet (ΔSNR: 4.5 dB).

Outline

Multiscaling in Wasserstein spaces

We adapt the multiscaling to the space of probability measures

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

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.

Formal Riemannian structure

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 geodesics in the metric space $(\mathcal{P}_p(\mathbb{R}^d), W_p)$.

Continuity equation

For any A.C. curve μ_t there exists a Borel vector field v_t such that

$$\frac{\partial}{\partial t}\mu_t + \nabla \cdot (\mu_t v_t) = 0.$$

Benamou-Brenier formula

For any $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$, the Benamou-Brenier formulation of the Wasserstein metric is given by

$$W_p^p(\mu, \nu) = \min_{\mu_t} \int_0^1 |\mu_t'|^p \, dt$$

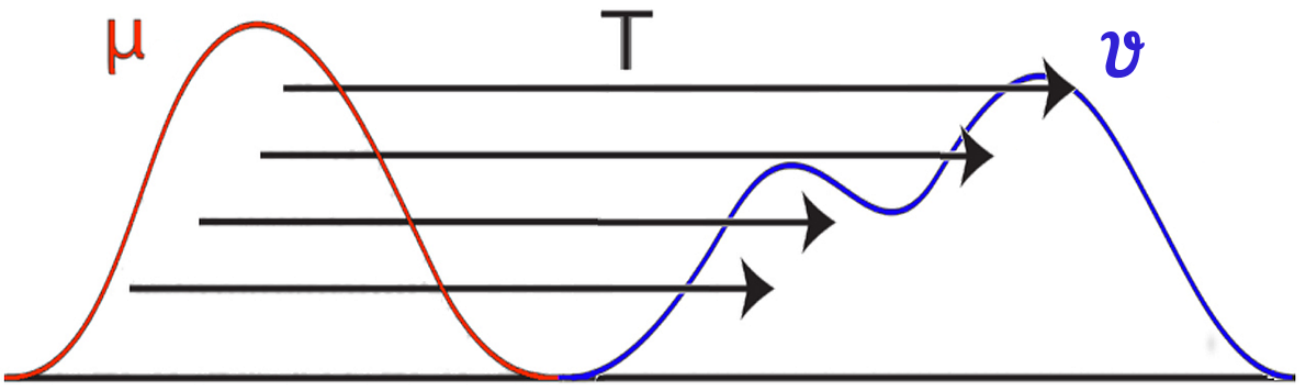
where $|\mu_t'|$ is the metric derivative of μ_t and the minimization is done over A.C. curves with $\mu_0 = \mu$ and $\mu_1 = \nu$.

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

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

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

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

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

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

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

Denoising

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

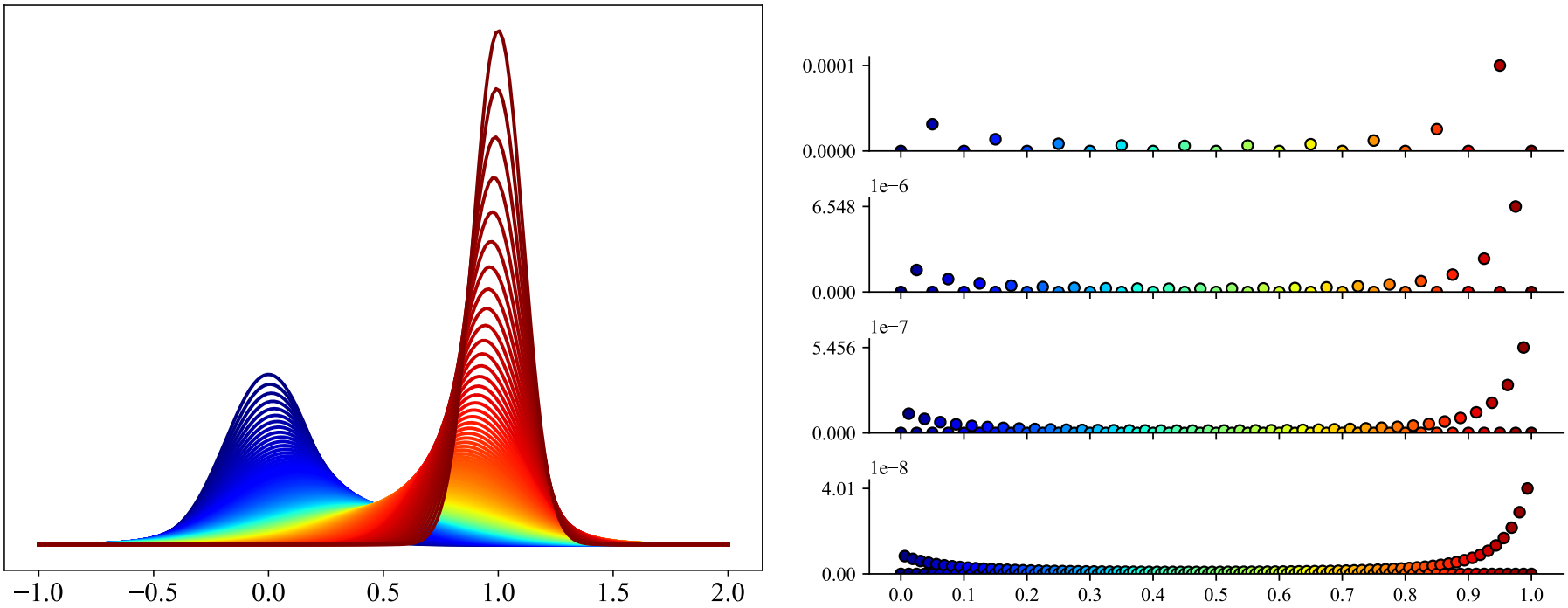


Figure: Sequence of Gaussian measures and its multiscale transform.

Denoising

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

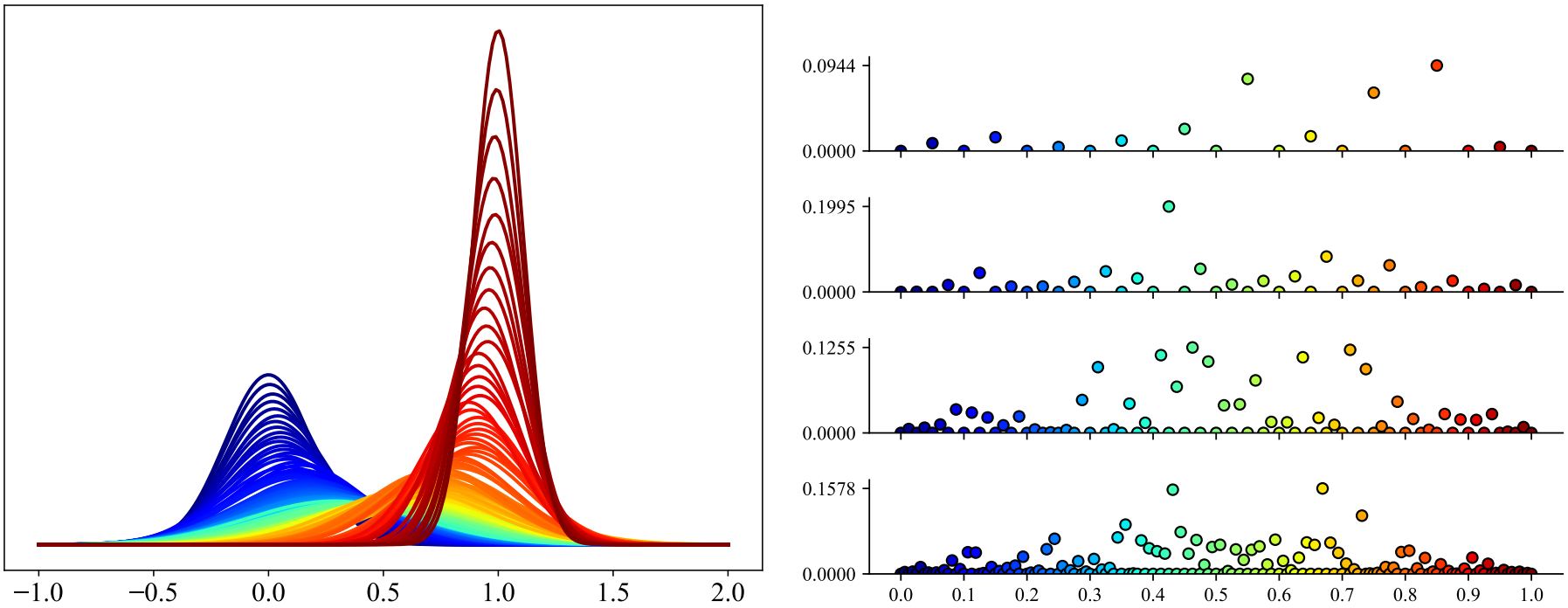
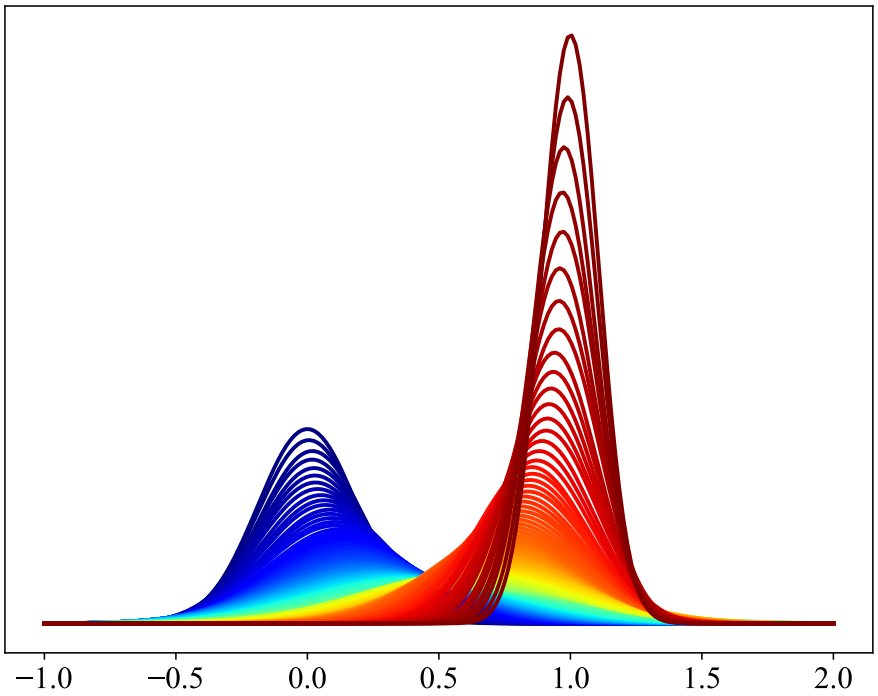


Figure: Contamination with noise.

Denoising

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



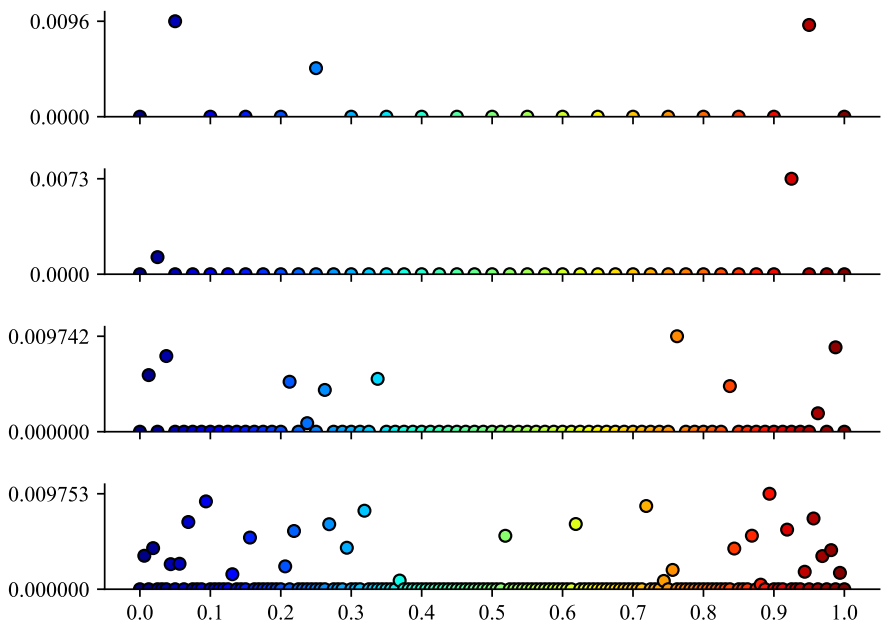


Figure: Denoising result obtained by thresholding with 0.01.

Anomaly detection

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

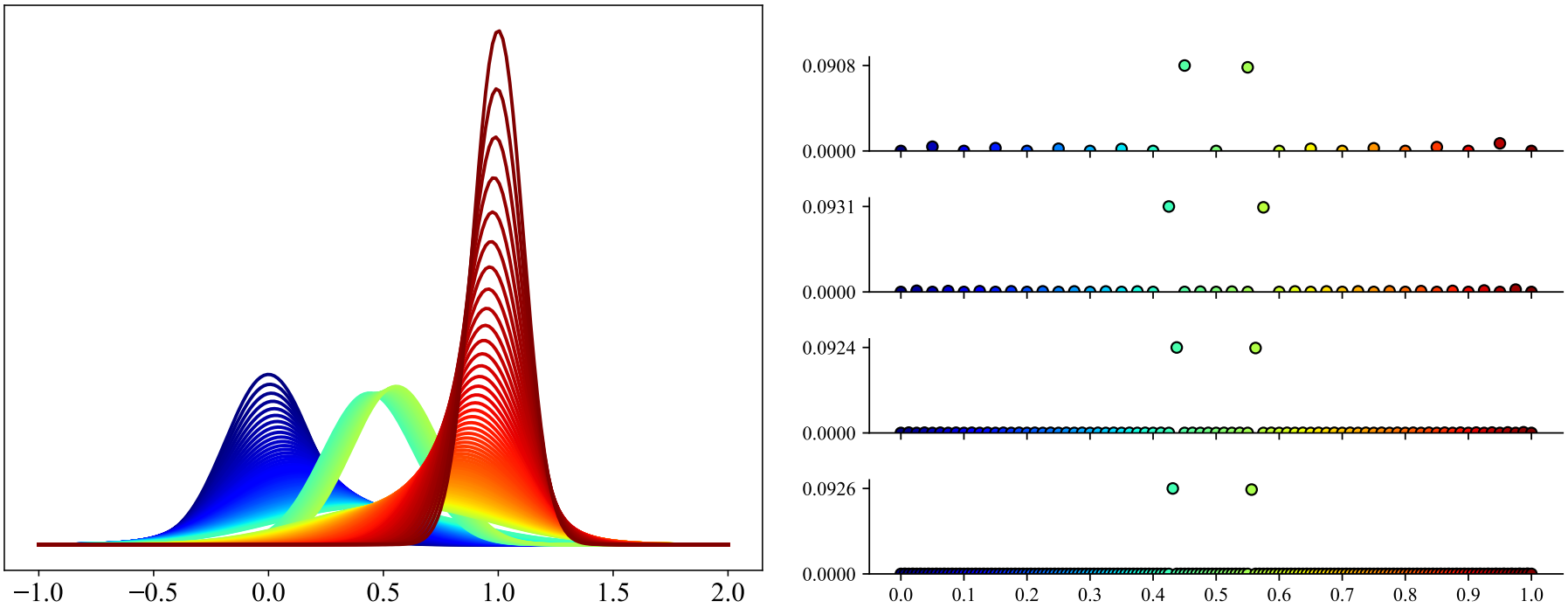


Figure: Detecting jump discontinuities via multiscaling.

Analyzing NN learning dynamics

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

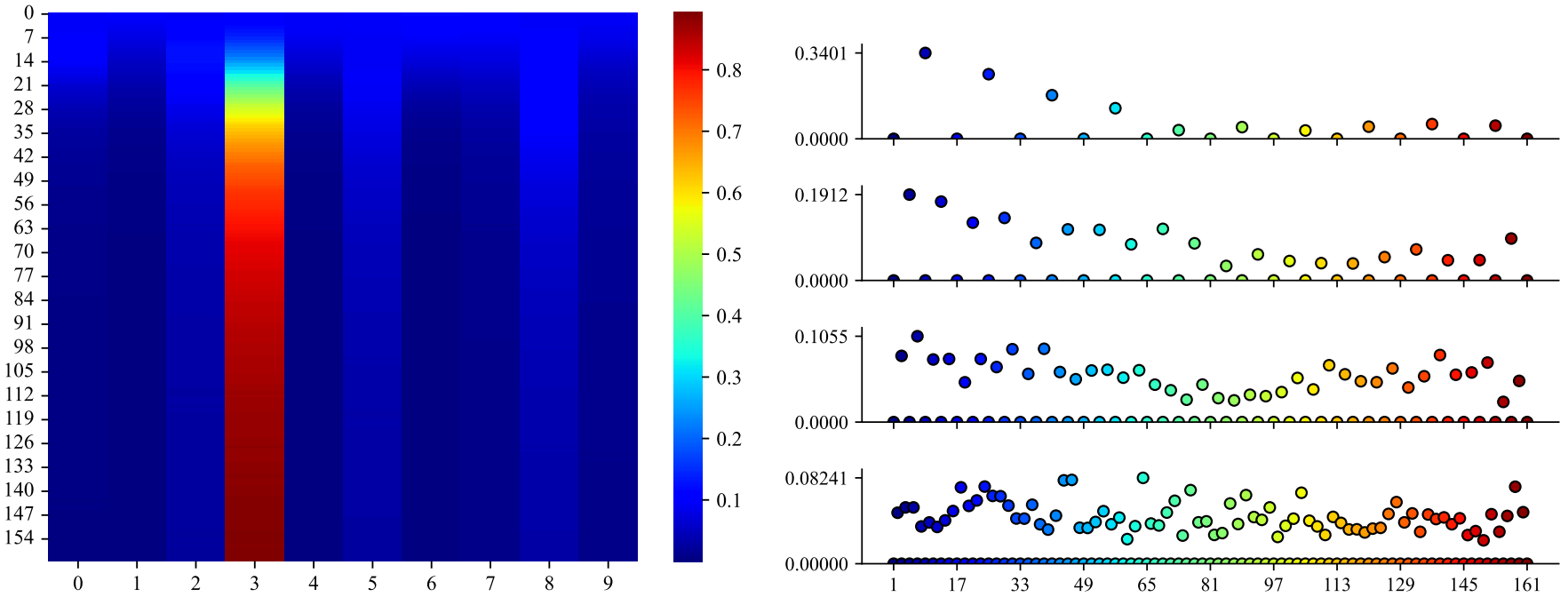


Figure: Analyzing the learning dynamics of a simple neural network on MNIST dataset. On the left, the prediction of the digit "3" across epoch iterations. On the right, the multiscale transform of the resulted measure sequence. The convergence is clear on coarse scales.

Thank you for listening!
Questions?

art by Robert Delaunay, Rhythm n°1, 1938