GAMES
Geometric Deep Learning I

Qixing Huang

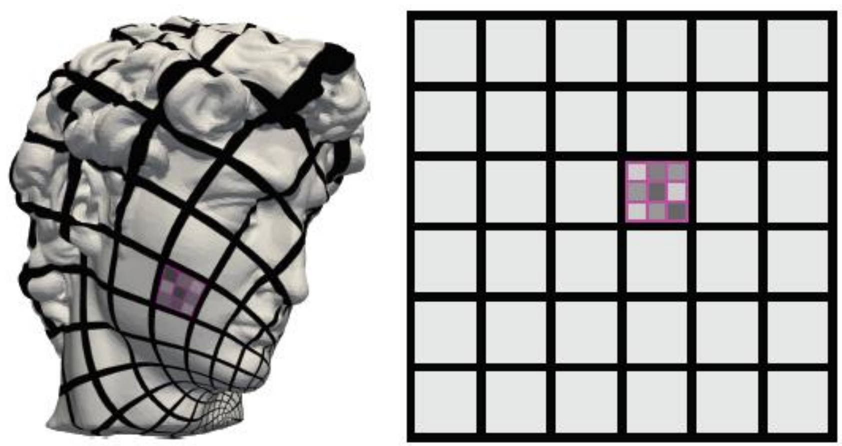
Sep. 16th 2021

Slide credit: Michael Bronstein

Different formulations of non-Euclidean CNNs

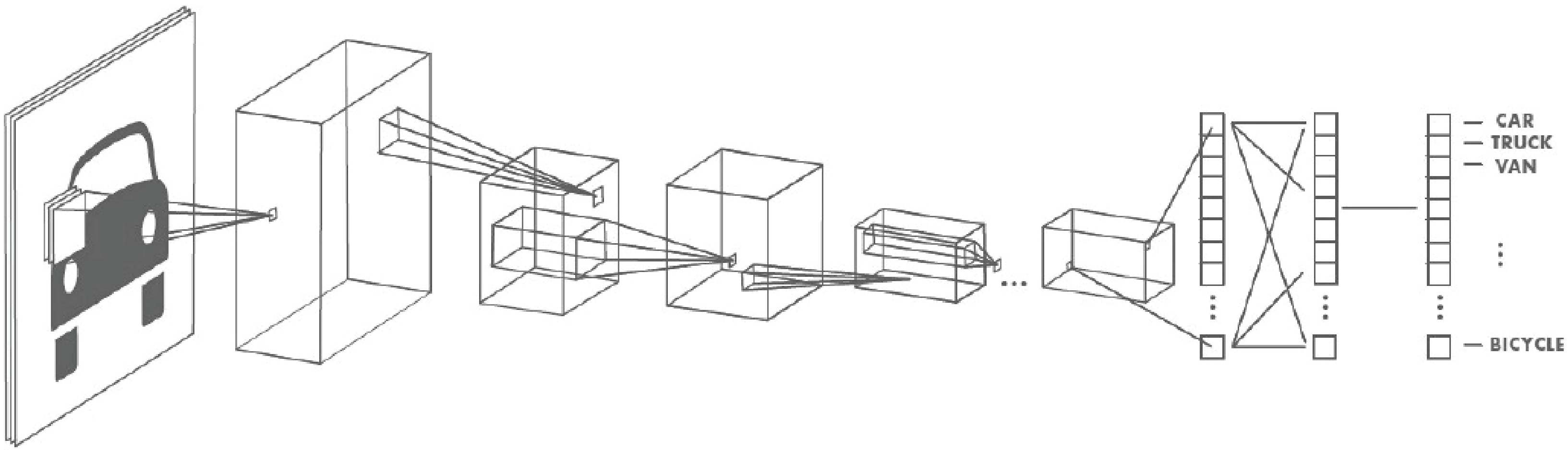
Spectral domain

Spatial domain



Parametric domain

Key properties of CNNs



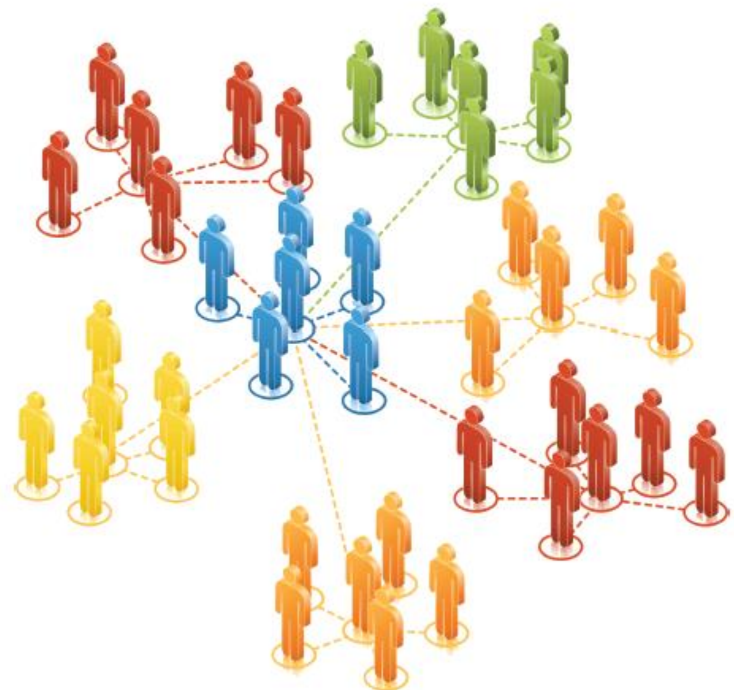
- ☺ Convolutional filters **(Translation invariance+Self-similarity)**
- ☺ Multiple layers **(Compositionality)**
- ☺ Filters localized in space **(Locality)**
- ☺ $\mathcal{O}(1)$ parameters per filter (independent of input image size n)
- ☺ $\mathcal{O}(n)$ complexity per layer (filtering done in the spatial domain)
- ☺ $\mathcal{O}(\log n)$ layers in classification tasks

LeCun et al. 1989

Going non-Euclidean

Prototypical non-Euclidean objects

Manifolds



Graphs

Challenges of geometric deep learning

- Extend neural network techniques to graph- or manifold-structured data
- Assumption: Non-Euclidean data are locally stationary and manifest hierarchical structures
- How to define compositionality? (convolution and pooling on graphs/manifolds)
- How to make them fast? (linear complexity)

Spectral analysis on graphs and manifolds

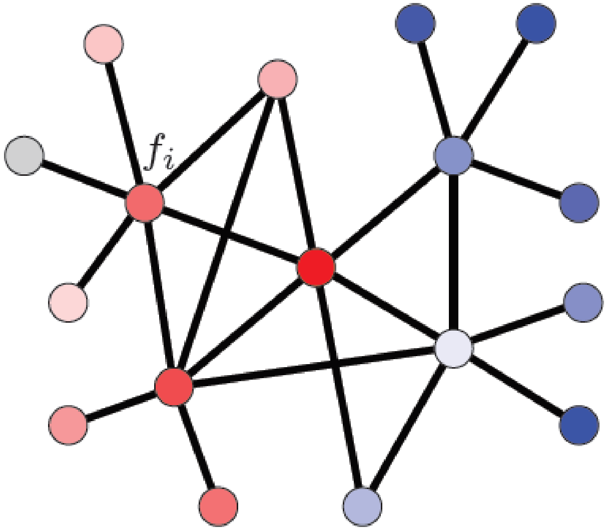
Graph theory in one minute

- Weighted undirected graph $\mathcal{G}$ with vertices $\mathcal{V} = \{1, \ldots, n\}$, edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ and edge weights $w_{ij} \geq 0$ for $(i, j) \in \mathcal{E}$

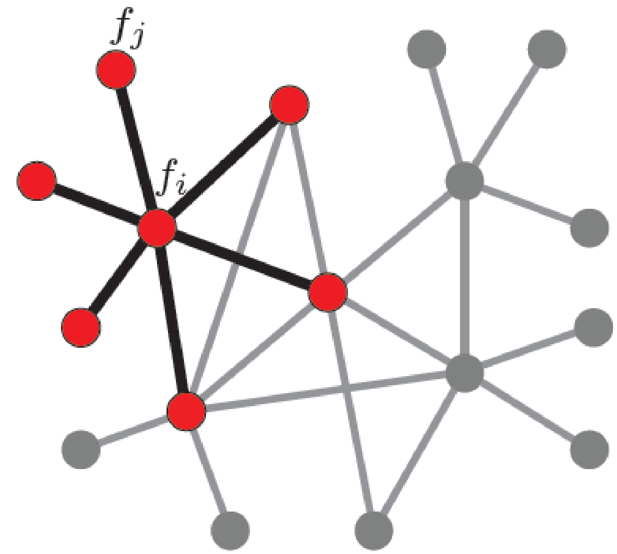
- Functions over the vertices $L^2(\mathcal{V}) = \{f : \mathcal{V} \to \mathbb{R}\}$ represented as vectors $\mathbf{f} = (f_1, \ldots, f_n)$

- Hilbert space with inner product

$$\langle f, g \rangle_{L^2(\mathcal{V})} = \sum_{i \in \mathcal{V}} f_i \, g_i = \mathbf{f}^\top \mathbf{g}$$



Graph Laplacian



- Unnormalized Laplacian $\Delta : L^2(\mathcal{V}) \to L^2(\mathcal{V})$

$$(\Delta f)_i \quad = \quad \sum_{j:(i,j)\in\mathcal{E}} w_{ij}(f_i - f_j)$$

(up to scale) difference between f and its local average

- Represented as a positive semi-definite $n \times n$ matrix $\mathbf{\Delta} = \mathbf{D} - \mathbf{W}$ where $\mathbf{W} = (w_{ij})$ and $\mathbf{D} = \mathrm{diag}(\sum_{j\neq i} w_{ij})$

- Dirichlet energy of f

$$\|f\|^2_{\mathcal{G}} = \frac{1}{2}\sum_{ij=1}^{n} w_{ij}(f_i - f_j)^2 = \mathbf{f}^\top \mathbf{\Delta} \mathbf{f}$$

measures the smoothness of f (how fast it changes locally)

Riemannian manifolds in one minute

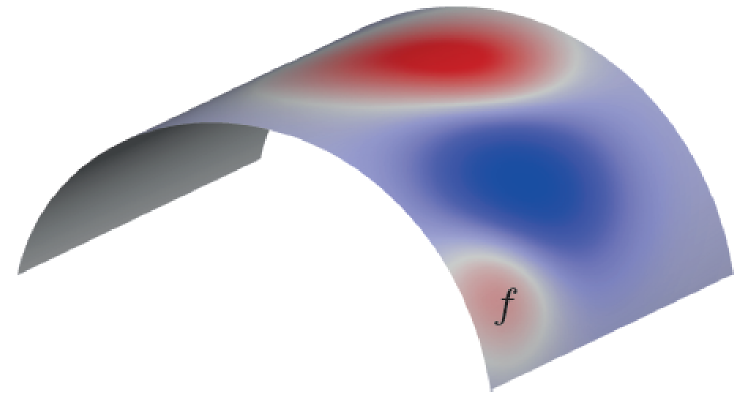
- Manifold $\mathcal{X}$ = topological space
- Tangent plane $T_x\mathcal{X}$ = local Euclidean representation of manifold $\mathcal{X}$ around x
- Riemannian metric describes the local intrinsic structure at x

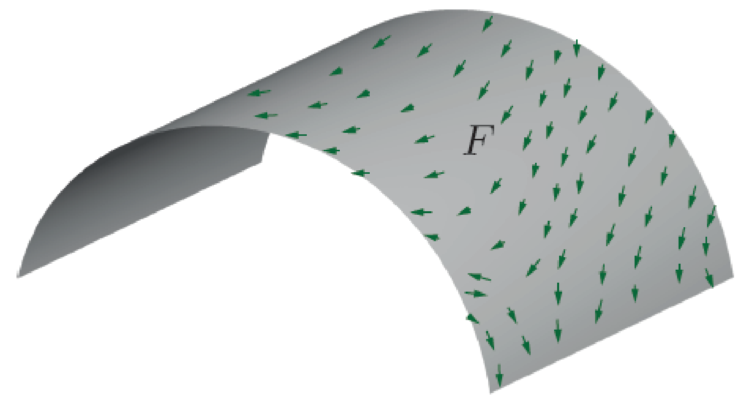
$$\langle \cdot, \cdot \rangle_{T_x\mathcal{X}} : T_x\mathcal{X} \times T_x\mathcal{X} \to \mathbb{R}$$

- Scalar fields $f : \mathcal{X} \to \mathbb{R}$ and vector fields $F : \mathcal{X} \to T\mathcal{X}$
- Hilbert spaces with inner products

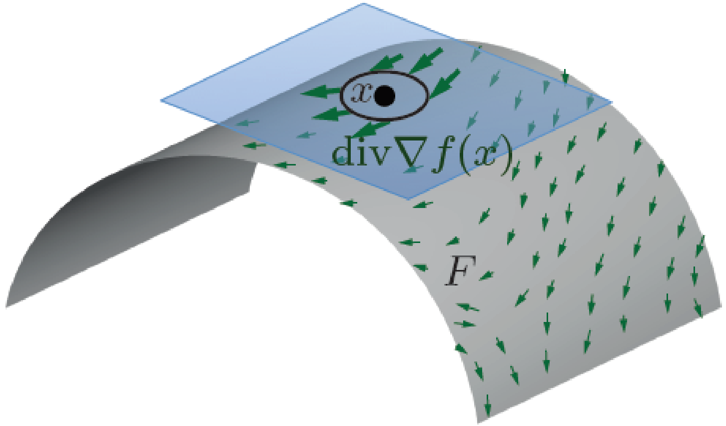
$$\langle f, g \rangle_{L^2(\mathcal{X})} = \int_{\mathcal{X}} f(x) g(x) dx$$

$$\langle F, G \rangle_{L^2(T\mathcal{X})} = \int_{\mathcal{X}} \langle F(x), G(x) \rangle_{T_x\mathcal{X}} dx$$





Manifold Laplacian



- Laplacian $\Delta : L^2(\mathcal{X}) \to L^2(\mathcal{X})$

$$\Delta f(x) = -\text{div}\, \nabla f(x)$$

where gradient $\nabla : L^2(\mathcal{X}) \to L^2(T\mathcal{X})$
and divergence $\text{div} : L^2(T\mathcal{X}) \to L^2(\mathcal{X})$
are adjoint operators

$$\langle F, \nabla f \rangle_{L^2(T\mathcal{X})} = \langle -\text{div} F, f \rangle_{L^2(\mathcal{X})}$$

- Laplacian is self-adjoint

$$\langle \Delta f, f \rangle_{L^2(\mathcal{X})} = \langle f, \Delta f \rangle_{L^2(\mathcal{X})}$$

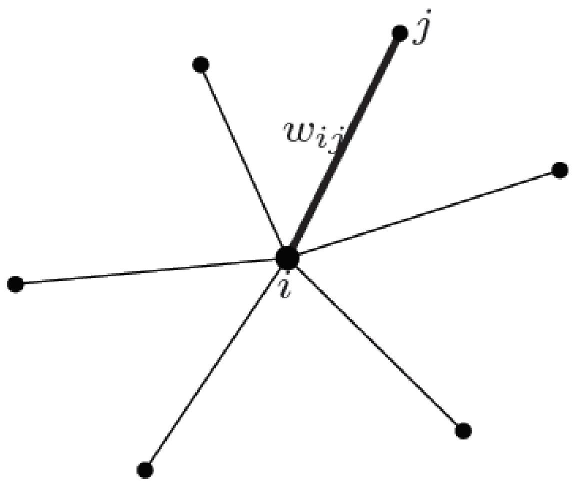
- Continuous limit of graph
Laplacian under some conditions

- Dirichlet energy of f

$$\langle \nabla f, \nabla f \rangle_{L^2(T\mathcal{X})} = \int_{\mathcal{X}} f(x) \Delta f(x) dx$$

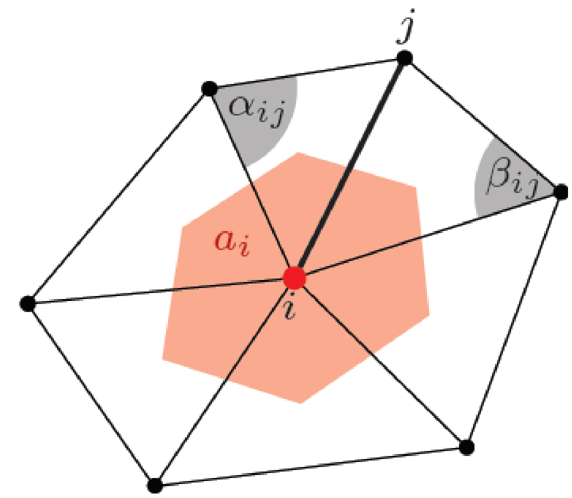
measures the smoothness of f (how fast it changes locally)

Discrete Laplacian



Undirected graph $(\mathcal{V},\mathcal{E})$

$$(\Delta f)_i \approx \sum_{(i,j)\in\mathcal{E}} w_{ij}(f_i - f_j)$$



Triangular mesh $(\mathcal{V},\mathcal{E},\mathcal{F})$

$$(\Delta f)_i \approx \frac{1}{a_i} \sum_{(i,j)\in\mathcal{E}} \frac{\cot\alpha_{ij}+\cot\beta_{ij}}{2}(f_i - f_j)$$

a_i = local area element

In matrix-vector notation

$$\mathbf{\Delta f} = \mathbf{A}^{-1}(\mathbf{D}-\mathbf{W})\mathbf{f}$$

where $\mathbf{f} = (f_1, \ldots, f_n)^\top$, $\mathbf{W}$ is the stiffness matrix, $\mathbf{A} = \mathrm{diag}(a_1, \ldots, a_n)$ is the mass matrix, and $\mathbf{D} = \mathrm{diag}(\sum_{j\neq 1} w_{1j}, \ldots, \sum_{j\neq n} w_{nj})$

Tutte 1963; MacNeal 1949; Duffin 1959; Pinkall, Polthier 1993

Orthogonal bases on graphs and manifolds

Find the smoothest orthogonal basis $\{\phi_1, \ldots, \phi_n\} \subseteq L^2(\mathcal{V})$

$$\min_{\mathbf{\Phi} \in \mathbb{R}^{n \times n}} \mathrm{trace}(\mathbf{\Phi}^\top \mathbf{\Delta} \mathbf{\Phi}) \quad \text{s.t.} \quad \mathbf{\Phi}^\top \mathbf{\Phi} = \mathbf{I}$$

Orthogonal bases on graphs and manifolds

Find the smoothest orthogonal basis $\{\phi_1, \ldots, \phi_n\} \subseteq L^2(\mathcal{V})$

$$\min_{\mathbf{\Phi} \in \mathbb{R}^{n \times n}} \mathrm{trace}(\mathbf{\Phi}^\top \mathbf{\Delta} \mathbf{\Phi}) \quad \text{s.t.} \quad \mathbf{\Phi}^\top \mathbf{\Phi} = \mathbf{I}$$

Solution: $\mathbf{\Phi}$ = Laplacian eigenvectors

Laplacian eigenvectors and eigenvalues

Eigendecomposition of a graph Laplacian

$$\mathbf{\Delta} = \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^\top$$

where $\mathbf{\Phi} = (\phi_1, \ldots, \phi_n)$ are orthogonal eigenvectors ($\mathbf{\Phi}^\top\mathbf{\Phi} = \mathbf{I}$) and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ the corresponding non-negative eigenvalues

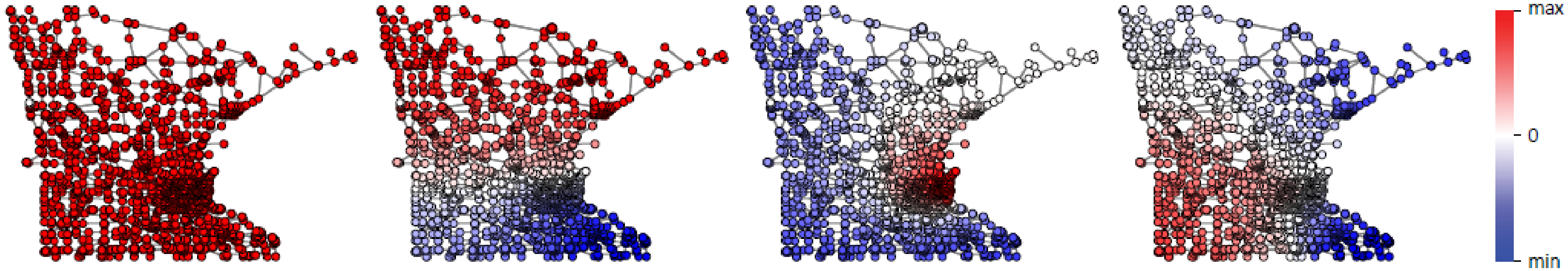
First eigenfunctions of 1D Euclidean Laplacian

Laplacian eigenvectors and eigenvalues

Eigendecomposition of a graph Laplacian

$$\boldsymbol{\Delta} = \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^\top$$

where $\boldsymbol{\Phi} = (\phi_1, \ldots, \phi_n)$ are orthogonal eigenvectors ($\boldsymbol{\Phi}^\top\boldsymbol{\Phi} = \mathbf{I}$) and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ the corresponding non-negative eigenvalues



ϕ_1 ϕ_2 ϕ_3 ϕ_4

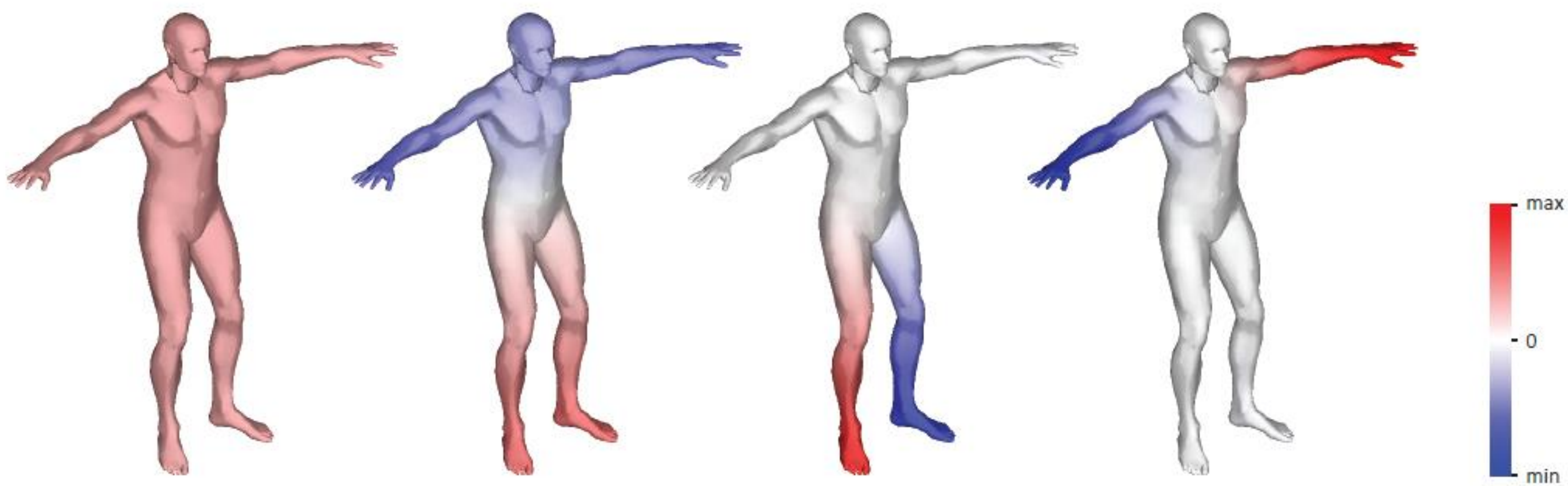
First eigenfunctions of a graph Laplacian

Laplacian eigenvectors and eigenvalues

Eigendecomposition of a graph Laplacian

$$\mathbf{\Delta} = \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^\top$$

where $\mathbf{\Phi} = (\phi_1, \ldots, \phi_n)$ are orthogonal eigenvectors ($\mathbf{\Phi}^\top\mathbf{\Phi} = \mathbf{I}$) and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ the corresponding non-negative eigenvalues



ϕ_1 ϕ_2 ϕ_3 ϕ_4

First eigenfunctions of a manifold Laplacian

Fourier analysis on Euclidean spaces

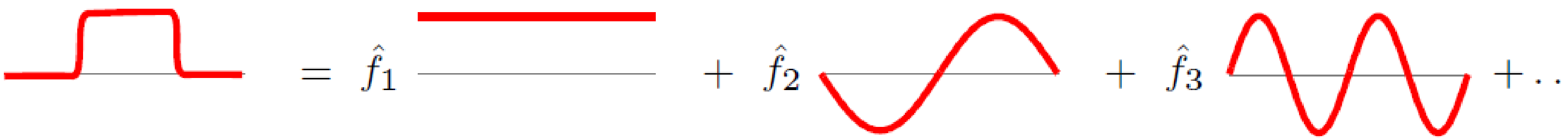
A function $f : [-\pi, \pi] \to \mathbb{R}$ can be written as a Fourier series

$$f(x) = \sum_{k \geq 0} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x') e^{-ikx'} dx' e^{ikx}$$

Fourier analysis on Euclidean spaces

A function $f : [-\pi, \pi] \to \mathbb{R}$ can be written as a Fourier series

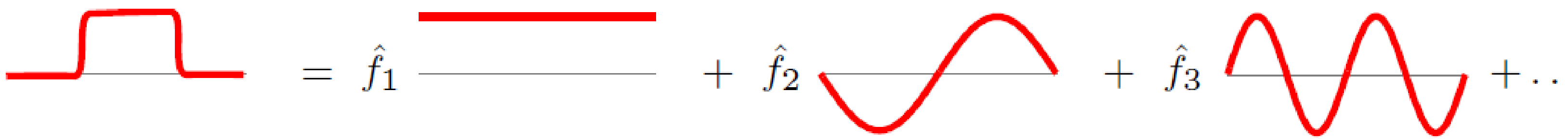
$$f(x) = \sum_{k \geq 0} \langle f, e^{ikx} \rangle_{L^2([-\pi,\pi])} \, e^{ikx}$$



Fourier analysis on Euclidean spaces

A function $f : [-\pi, \pi] \to \mathbb{R}$ can be written as a Fourier series

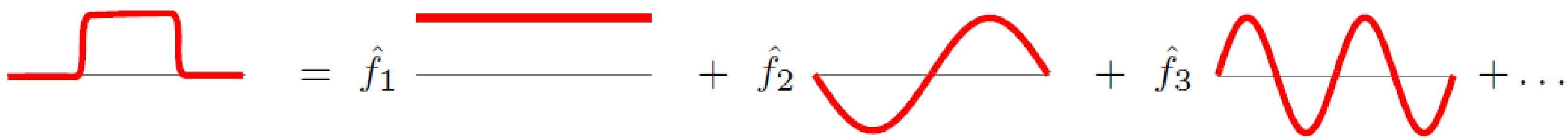
$$f(x) = \sum_{k \geq 0} \underbrace{\langle f, e^{ikx} \rangle_{L^2([-\pi,\pi])}}_{\hat{f}_k \text{ Fourier coefficient}} e^{ikx}$$



Fourier analysis on Euclidean spaces

A function $f : [-\pi, \pi] \to \mathbb{R}$ can be written as a Fourier series

$$f(x) = \sum_{k \geq 0} \underbrace{\langle f, e^{ikx} \rangle_{L^2([-\pi,\pi])}}_{\hat{f}_k \text{ Fourier coefficient}} e^{ikx}$$

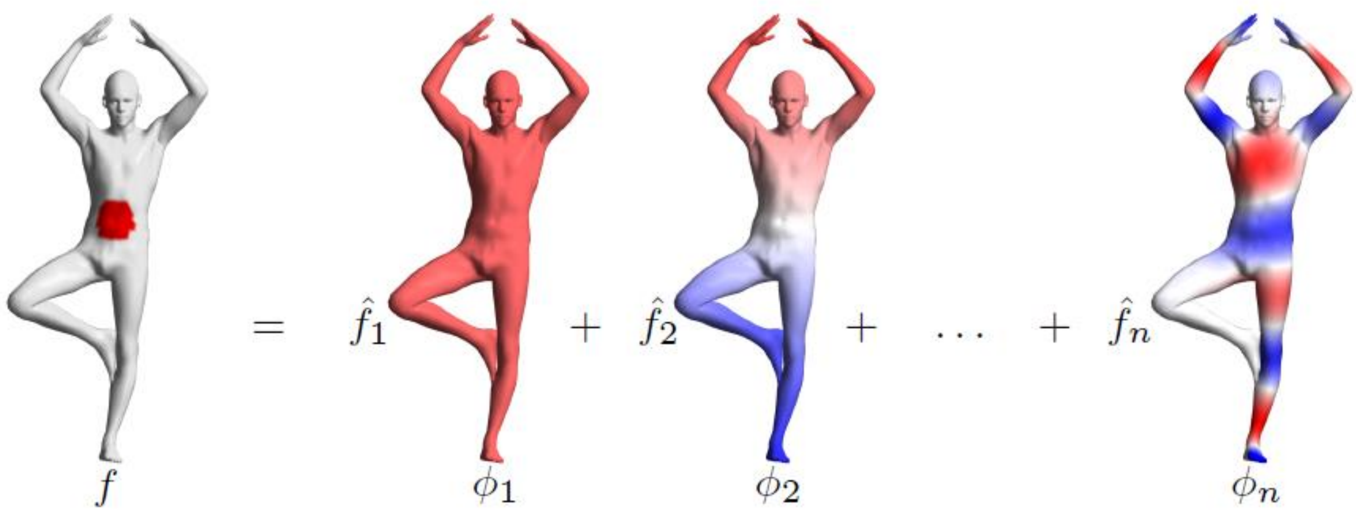


Fourier basis = Laplacian eigenfunctions: $-\frac{d^2}{dx^2} e^{ikx} = k^2 e^{ikx}$

Fourier analysis on graphs and manifolds

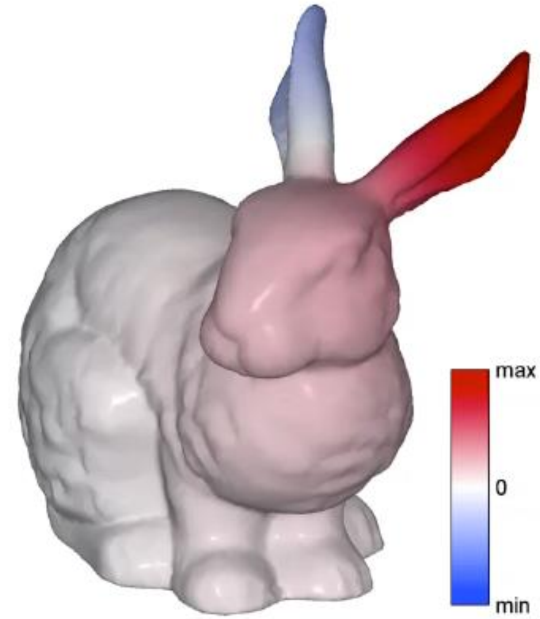
A function $f : \mathcal{X} \to \mathbb{R}$ can be written as Fourier series

$$f = \sum_{k=1}^{n} \underbrace{\langle f, \phi_k \rangle_{L^2(\mathcal{X})}}_{\hat{f}_k} \; \phi_k$$



Fourier basis = Laplacian eigenfunctions: $\Delta \phi_k = \lambda_k \phi_k$

Heat diffusion on manifolds



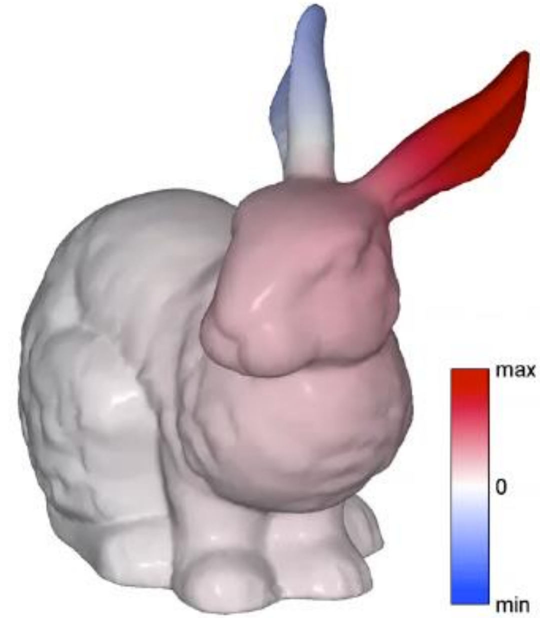
$$\begin{cases} f_t(x,t) = -\Delta f(x,t) \\ f(x,0) = f_0(x) \end{cases}$$

- $f(x,t)$ = amount of heat at point x at time t
- $f_0(x)$ = initial heat distribution

Solution of the heat equation expressed through the heat operator

$$f(x,t) \quad = \quad e^{-t\Delta} f_0(x)$$

Heat diffusion on manifolds



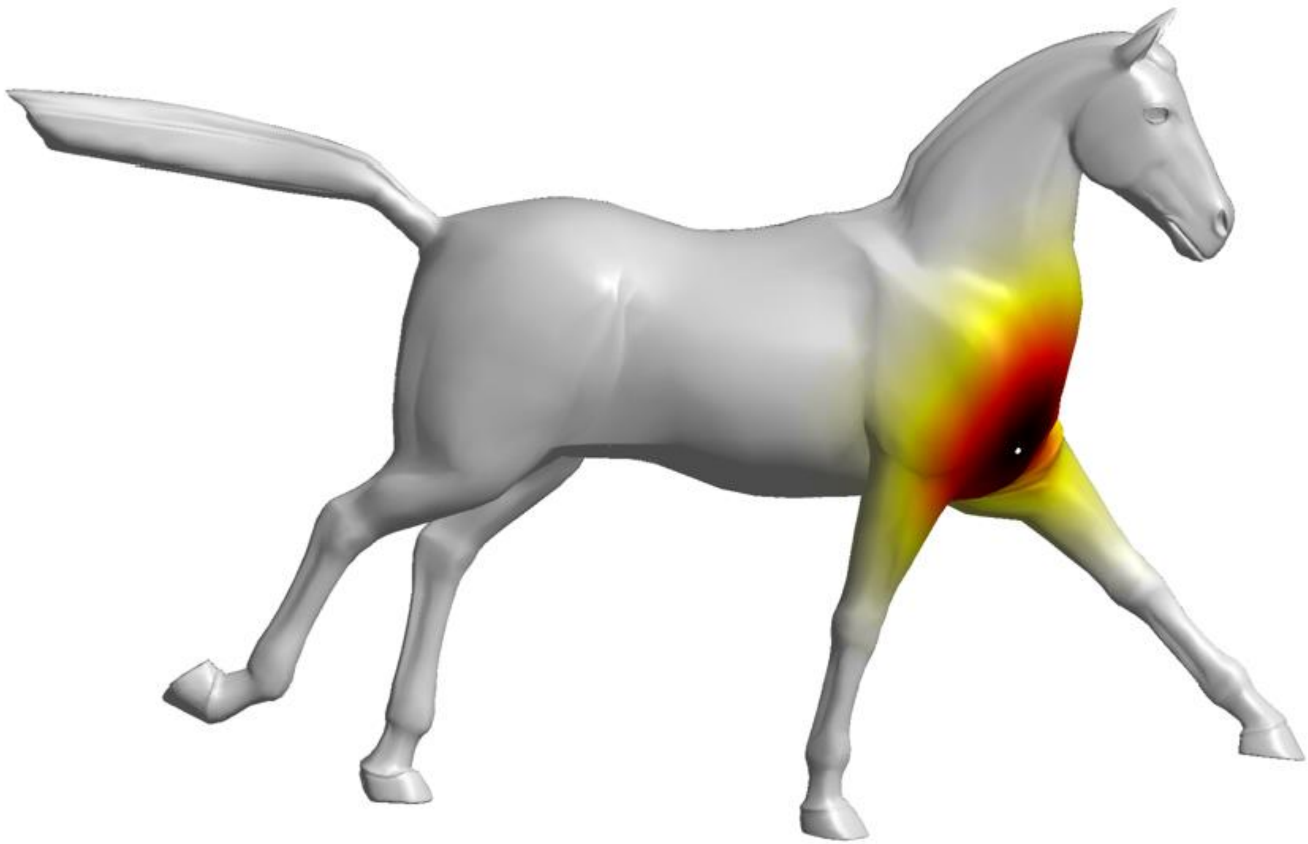
$$\begin{cases} f_t(x,t) = -\Delta f(x,t) \\ f(x,0) = f_0(x) \end{cases}$$

- $f(x,t)$ = amount of heat at point x at time t
- $f_0(x)$ = initial heat distribution

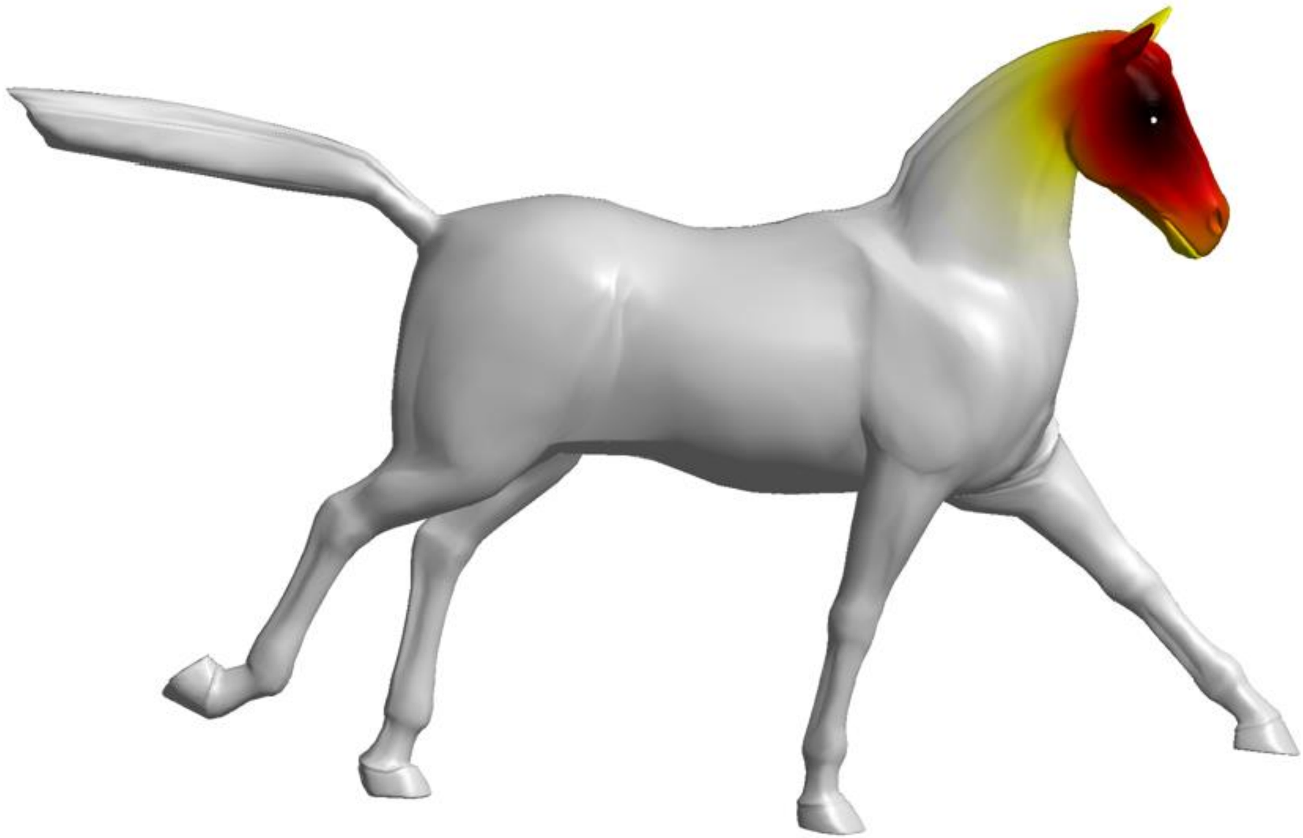
Solution of the heat equation expressed through the heat operator

$$\begin{aligned} f(x,t) &= e^{-t\Delta} f_0(x) = \sum_{k\geq 1} \langle f_0, \phi_k \rangle_{L^2(\mathcal{X})} e^{-t\lambda_k} \phi_k(x) \\ &= \int_{\mathcal{X}} f_0(x') \underbrace{\sum_{k\geq 1} e^{-t\lambda_k} \phi_k(x) \phi_k(x')}_{\text{heat kernel } h_t(x,x')} \, dx' \end{aligned}$$

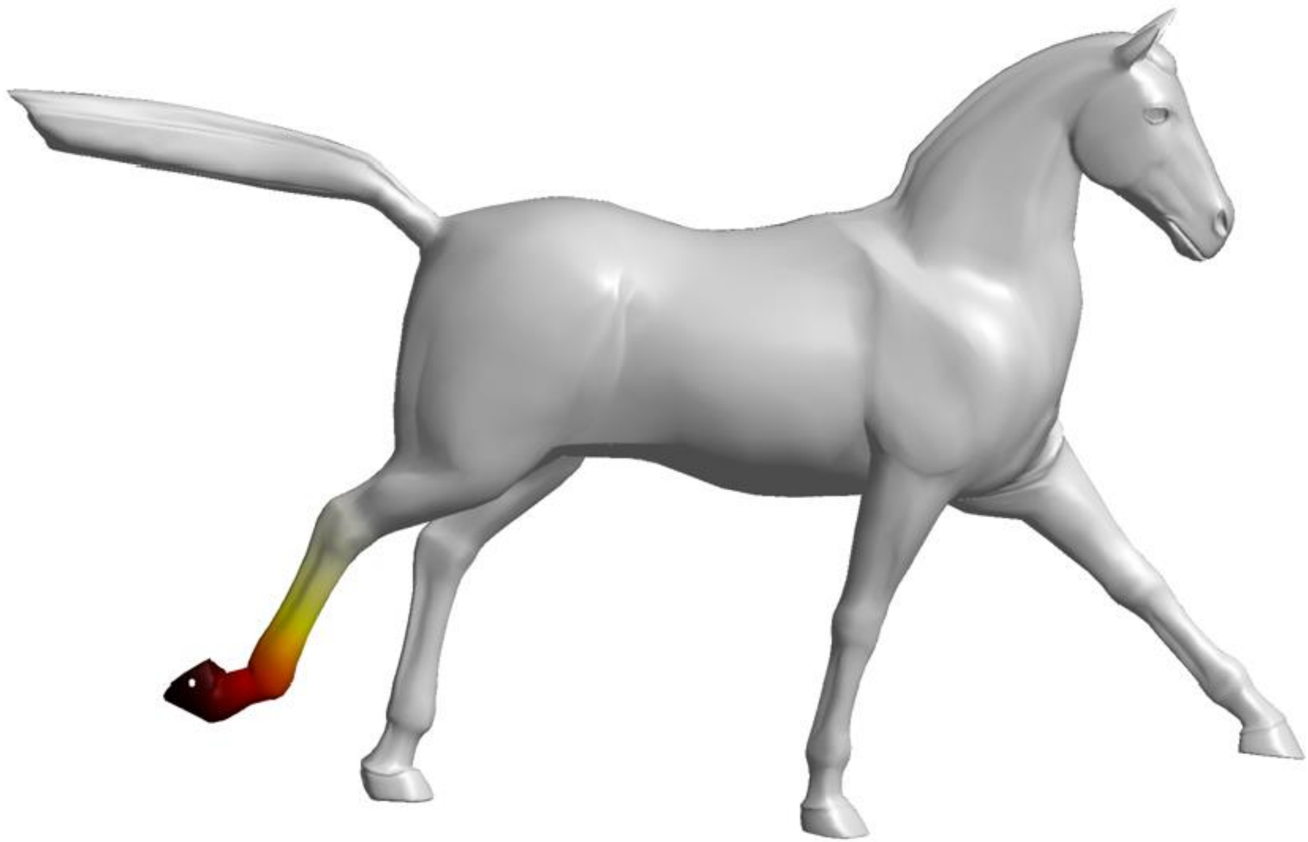
Heat kernels



Heat kernels



Heat kernels



Convolution: Euclidean space

Given two functions $f, g : [-\pi, \pi] \to \mathbb{R}$ their convolution is a function

$$(f \star g)(x) = \int_{-\pi}^{\pi} f(x')g(x - x')dx'$$

- Shift-invariance: $f(x - x_0) \star g(x) = (f \star g)(x - x_0)$
- Convolution theorem: Fourier transform diagonalizes the convolution operator $\Rightarrow$ convolution can be computed in the Fourier domain as

$$\widehat{(f \star g)} = \hat{f} \cdot \hat{g}$$

Convolution Theorem

Convolution of two vectors $\mathbf{f} = (f_1, \dots, f_n)^\top$ and $\mathbf{g} = (g_1, \dots, g_n)^\top$

$$\mathbf{f} \star \mathbf{g} \quad = \quad \underbrace{\begin{bmatrix} g_1 & g_2 & \cdots & \cdots & g_n \\ g_n & g_1 & g_2 & \cdots & g_{n-1} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ g_3 & g_4 & \cdots & g_1 & g_2 \\ g_2 & g_3 & \cdots & \cdots & g_1 \end{bmatrix}}_{\text{circulant matrix}} \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix}$$

Convolution Theorem

Convolution of two vectors $\mathbf{f} = (f_1, \dots, f_n)^\top$ and $\mathbf{g} = (g_1, \dots, g_n)^\top$

$$\mathbf{f} \star \mathbf{g} \quad = \quad \underbrace{\begin{bmatrix} g_1 & g_2 & \cdots & \cdots & g_n \\ g_n & g_1 & g_2 & \cdots & g_{n-1} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ g_3 & g_4 & \cdots & g_1 & g_2 \\ g_2 & g_3 & \cdots & \cdots & g_1 \end{bmatrix}}_{\text{diagonalized by Fourier basis}} \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix}$$

Convolution Theorem

Convolution of two vectors $\mathbf{f} = (f_1, \dots, f_n)^\top$ and $\mathbf{g} = (g_1, \dots, g_n)^\top$

$$\mathbf{f} \star \mathbf{g} = \begin{bmatrix} g_1 & g_2 & \cdots & \cdots & g_n \\ g_n & g_1 & g_2 & \cdots & g_{n-1} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ g_3 & g_4 & \cdots & g_1 & g_2 \\ g_2 & g_3 & \cdots & \cdots & g_1 \end{bmatrix} \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix}$$

$$= \boldsymbol{\Phi} \begin{bmatrix} \hat{g}_1 & & \\ & \ddots & \\ & & \hat{g}_n \end{bmatrix} \boldsymbol{\Phi}^\top \mathbf{f}$$

Convolution Theorem

Convolution of two vectors $\mathbf{f} = (f_1, \dots, f_n)^\top$ and $\mathbf{g} = (g_1, \dots, g_n)^\top$

$$\mathbf{f} \star \mathbf{g} = \begin{bmatrix} g_1 & g_2 & \cdots & \cdots & g_n \\ g_n & g_1 & g_2 & \cdots & g_{n-1} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ g_3 & g_4 & \cdots & g_1 & g_2 \\ g_2 & g_3 & \cdots & \cdots & g_1 \end{bmatrix} \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix}$$

$$= \mathbf{\Phi} \begin{bmatrix} \hat{g}_1 & & \\ & \ddots & \\ & & \hat{g}_n \end{bmatrix} \begin{bmatrix} \hat{f}_1 \\ \vdots \\ \hat{f}_n \end{bmatrix}$$

Convolution Theorem

Convolution of two vectors $\mathbf{f} = (f_1, \ldots, f_n)^\top$ and $\mathbf{g} = (g_1, \ldots, g_n)^\top$

$$\mathbf{f} \star \mathbf{g} = \begin{bmatrix} g_1 & g_2 & \cdots & \cdots & g_n \\ g_n & g_1 & g_2 & \cdots & g_{n-1} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ g_3 & g_4 & \cdots & g_1 & g_2 \\ g_2 & g_3 & \cdots & \cdots & g_1 \end{bmatrix} \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix}$$

$$= \boldsymbol{\Phi} \begin{bmatrix} \hat{f}_1 \cdot \hat{g}_1 \\ \vdots \\ \hat{f}_n \cdot \hat{g}_n \end{bmatrix}$$

Spectral convolution

Spectral convolution of $f, g \in L^2(\mathcal{X})$ can be defined by analogy

$$f \star g = \underbrace{\sum_{k \geq 1} \underbrace{\langle f, \phi_k \rangle_{L^2(\mathcal{X})} \langle g, \phi_k \rangle_{L^2(\mathcal{X})}}_{\text{product in the Fourier domain}} \phi_k}_{\text{inverse Fourier transform}}$$

Spectral convolution

Spectral convolution of $f, g \in L^2(\mathcal{X})$ can be defined by analogy

$$f \star g \;=\; \sum_{k \geq 1} \langle f, \phi_k \rangle_{L^2(\mathcal{X})} \langle g, \phi_k \rangle_{L^2(\mathcal{X})} \, \phi_k$$

In matrix-vector notation

$$\mathbf{f} \star \mathbf{g} = \underbrace{\boldsymbol{\Phi} \, \mathrm{diag}(\hat{g}_1, \ldots, \hat{g}_n) \boldsymbol{\Phi}^\top}_{\mathbf{G}} \, \mathbf{f}$$

- Not shift-invariant! ($\mathbf{G}$ has no circulant structure)
- Filter coefficients depend on basis $\phi_1, \ldots, \phi_n$

Spectral CNN

Convolution expressed in the spectral domain

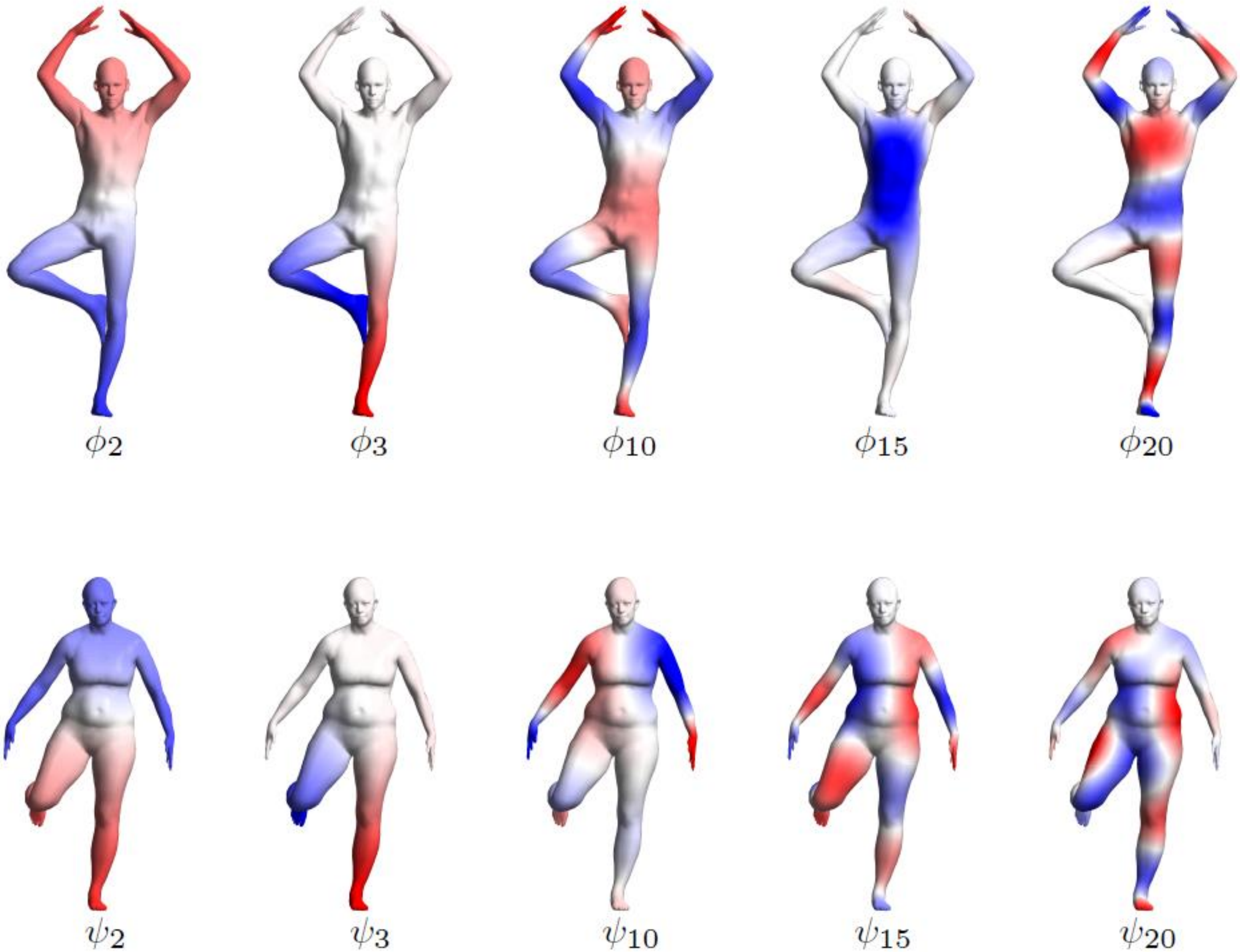
$$\mathbf{g} = \mathbf{\Phi} \mathbf{W} \mathbf{\Phi}^\top \mathbf{f}$$

where $\mathbf{W}$ is $n \times n$ diagonal matrix of learnable spectral filter coefficients

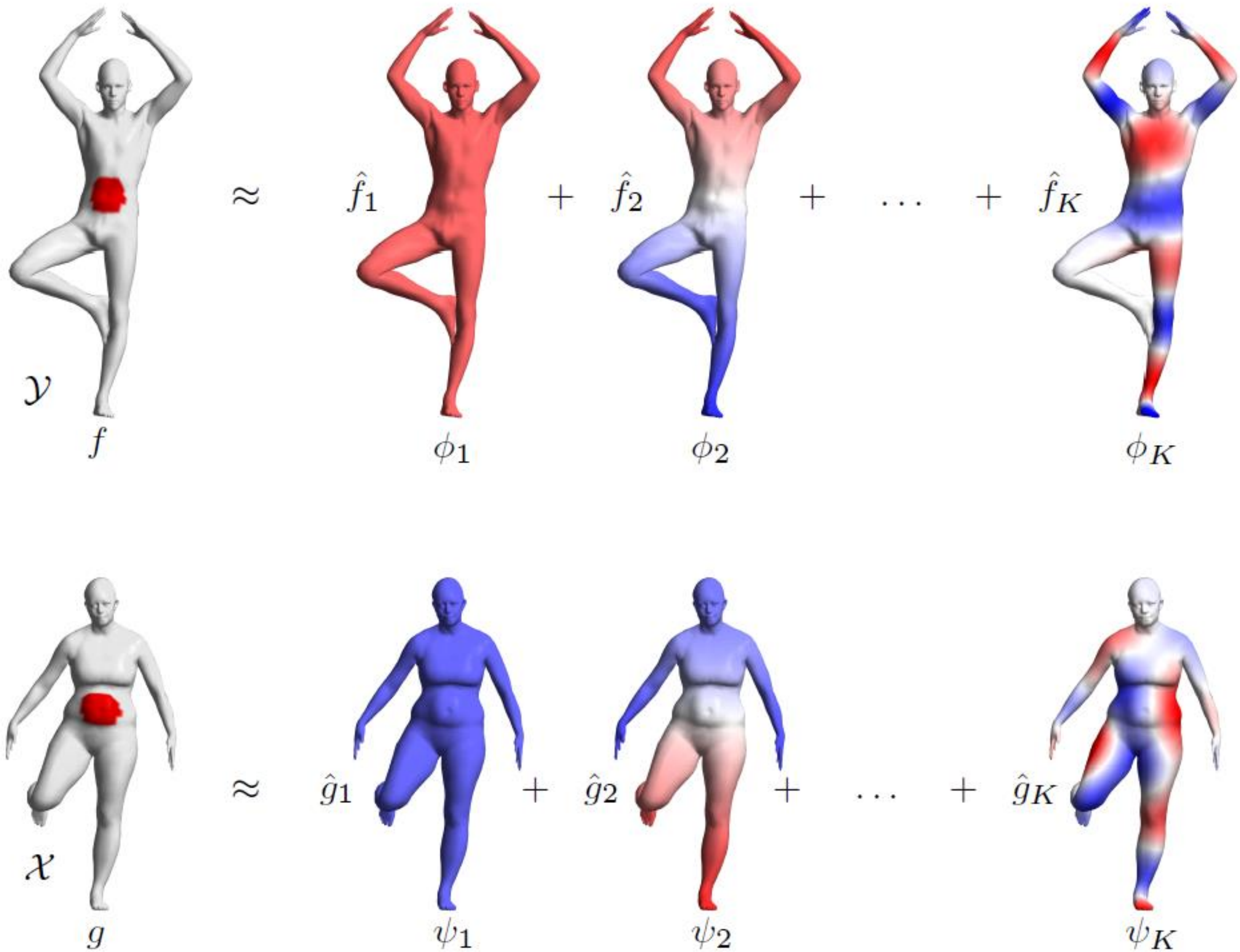
- ☹ Filters are basis-dependent $\Rightarrow$ do not generalize across domains
- ☹ $\mathcal{O}(n)$ parameters per layer
- ☹ $\mathcal{O}(n^2)$ computation of forward and inverse Fourier transforms $\mathbf{\Phi}^\top, \mathbf{\Phi}$ (no FFT on graphs)
- ☹ No guarantee of spatial localization of filters

Bruna et al. 2014

Laplacian eigenbases on non-isometric domains

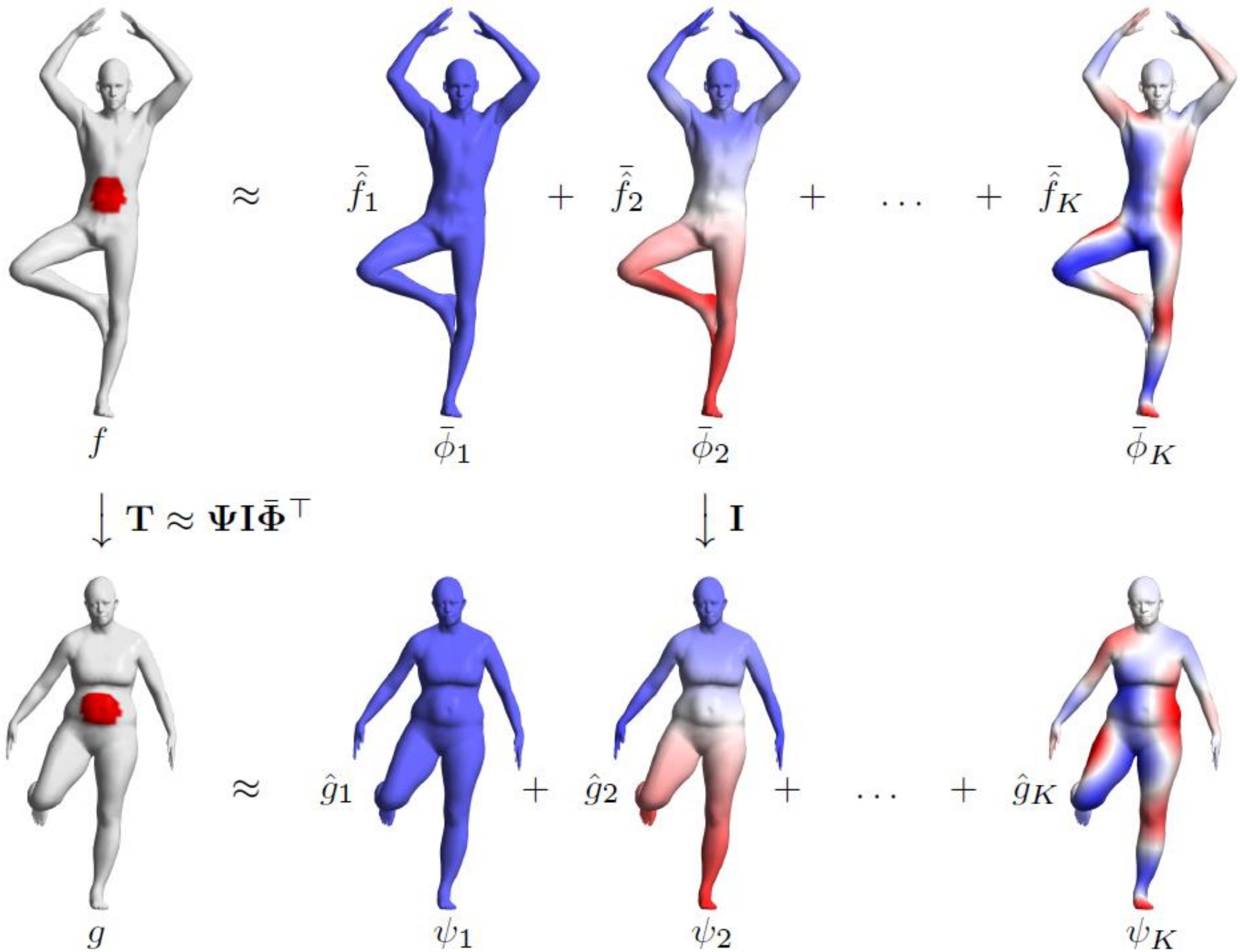


Functional maps



Ovsjanikov et al. 2012; Eynard et al. 2012; Kovnatsky et al. 2013

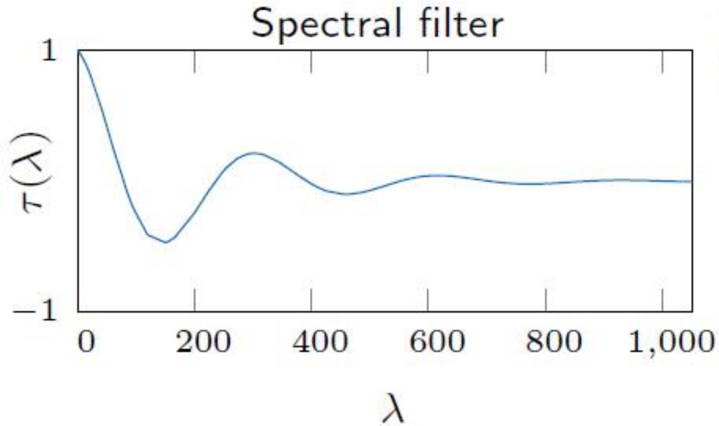
Basis synchronization with functional maps

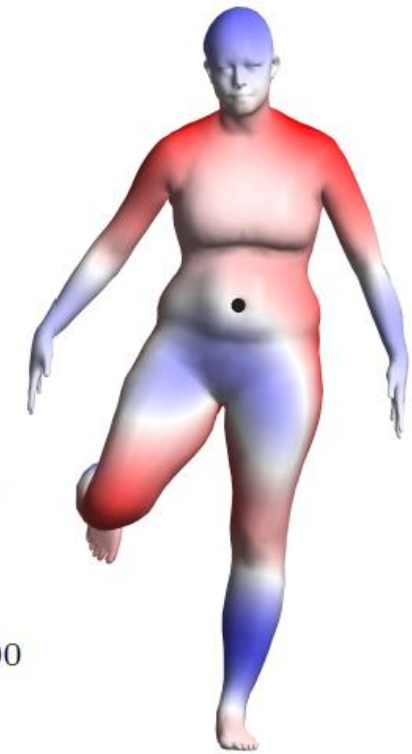


Ovsjanikov et al. 2012; Eynard et al. 2012; Kovnatsky et al. 2013

Filtering in different bases

$\mathbf{\Phi}\tau(\mathbf{\Lambda_\Phi})\mathbf{\Phi}^\top\boldsymbol{\delta}_0$





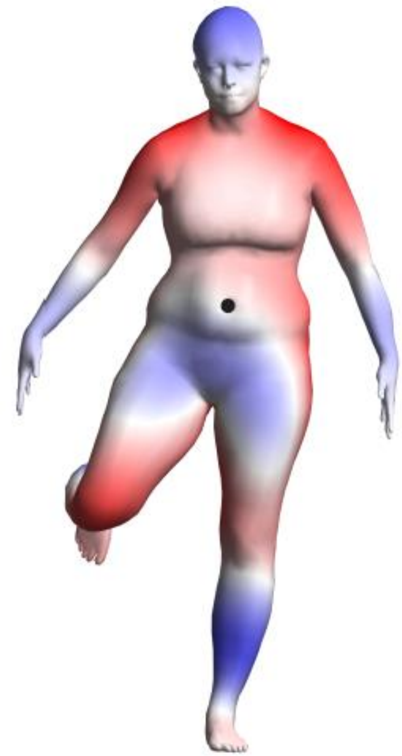
$\mathbf{\Psi}\tau(\mathbf{\Lambda_\Psi})\mathbf{\Psi}^\top\boldsymbol{\delta}_0$

Apply spectral filter $\tau(\lambda)$ in different bases $\mathbf{\Phi}$ and $\mathbf{\Psi}$
$\Rightarrow$ **different results!**

Filtering in different bases

$\mathbf{\Phi}\tau(\mathbf{\Lambda_\Phi})\mathbf{\Phi}^\top\boldsymbol{\delta}_0$

Canonical shape
with basis $\mathbf{\Sigma}, \mathbf{\Lambda}$

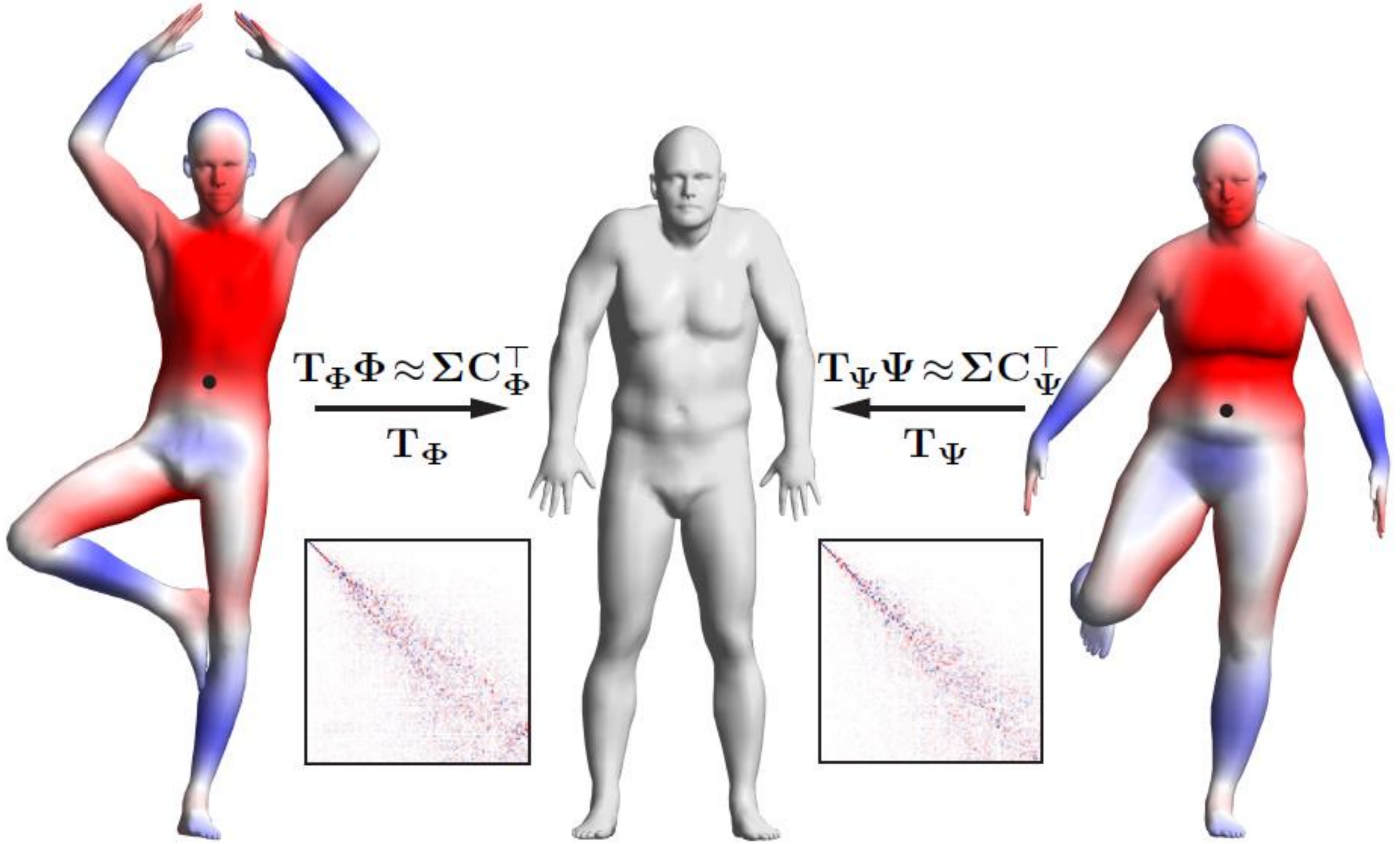


$\mathbf{\Psi}\tau(\mathbf{\Lambda_\Psi})\mathbf{\Psi}^\top\boldsymbol{\delta}_0$

Apply spectral filter $\tau(\lambda)$ in different bases $\mathbf{\Phi}$ and $\mathbf{\Psi}$

$\Rightarrow$ **different results!**

Filtering in synchronized bases



$\Phi \mathbf{C}_\Phi \tau(\Lambda) \mathbf{C}_\Phi^\top \Phi^\top \delta_0$

Canonical shape
with basis Σ, Λ

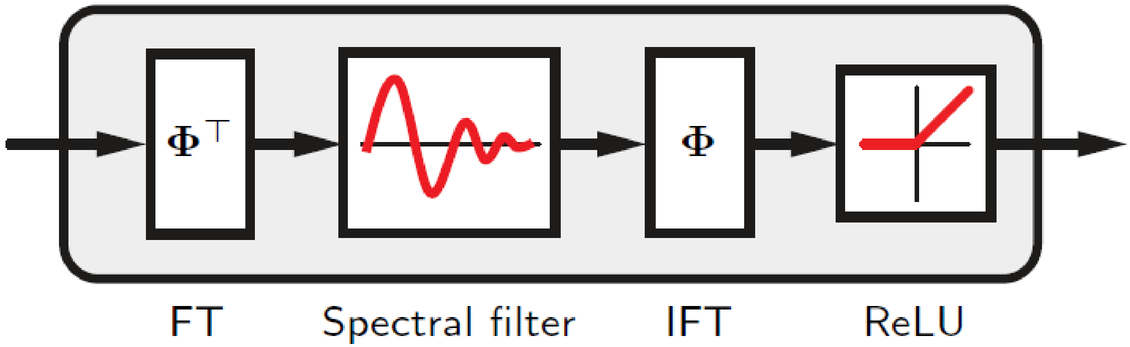
$\Psi \mathbf{C}_\Psi \tau(\Lambda) \mathbf{C}_\Psi^\top \Psi^\top \delta_0$

Apply spectral filter $\tau(\lambda)$ in synchronized bases $\Phi \mathbf{C}_\Phi$ and $\Psi \mathbf{C}_\Psi$

$\Rightarrow$ **similar results!**

Yi et al. 2017

Spectral CNN



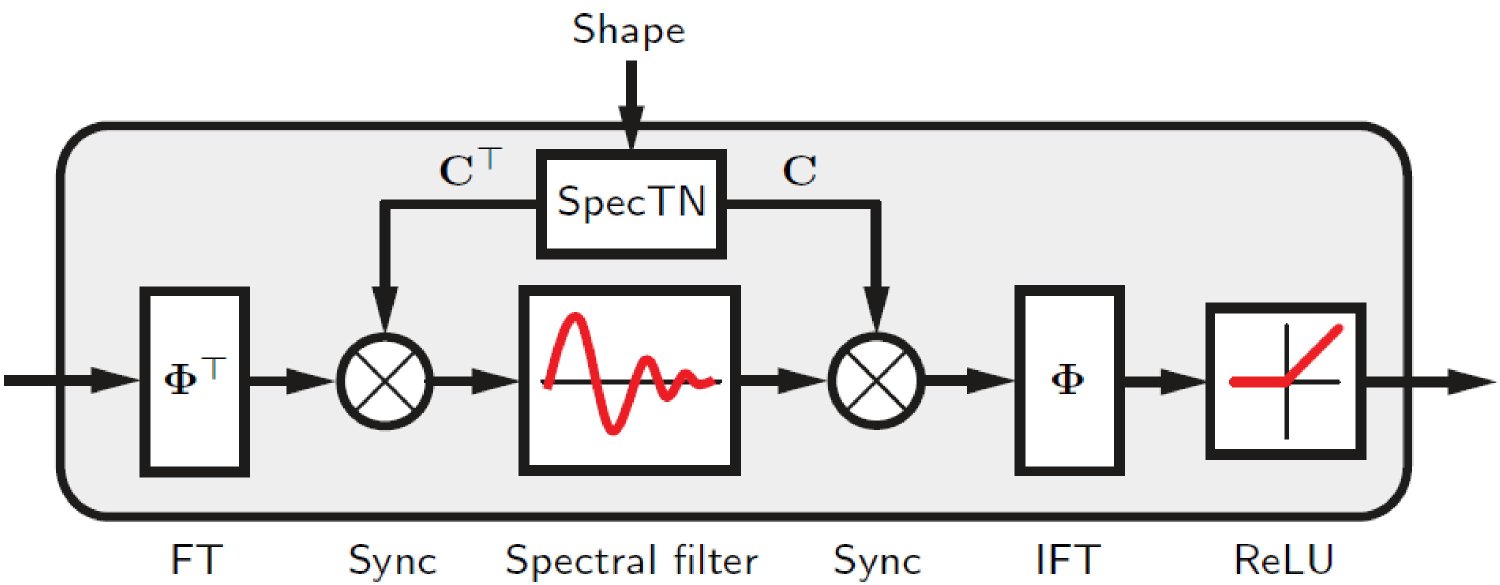
Convolutional filter of a Spectral CNN

☹ Fixed basis ⇒ Does not generalize across domains

☺ Possible $\mathcal{O}(n)$ complexity avoiding explicit FT and IFT

Bruna et al. 2014

Spectral Transformer Network



Convolutional filter of a Spectral Transformer Network

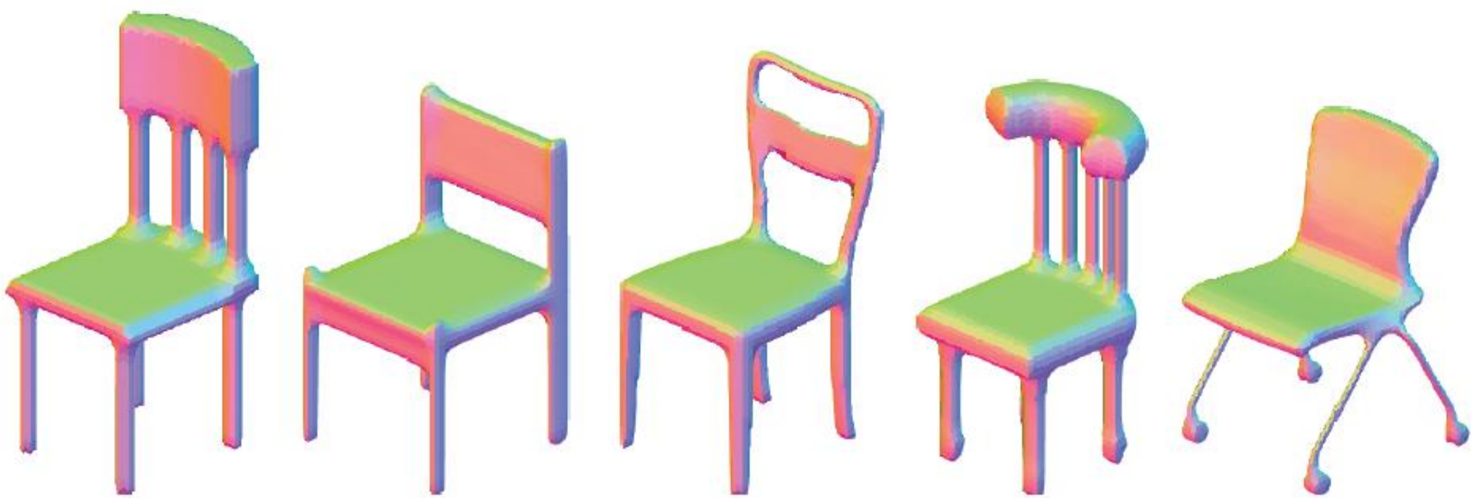
☺ Basis synchronization allows generalization across domains

☹ Explicit FT and IFT

(Jaderberg et al. 2015); Yi et al. 2017

Example: normal prediction with SpecTN

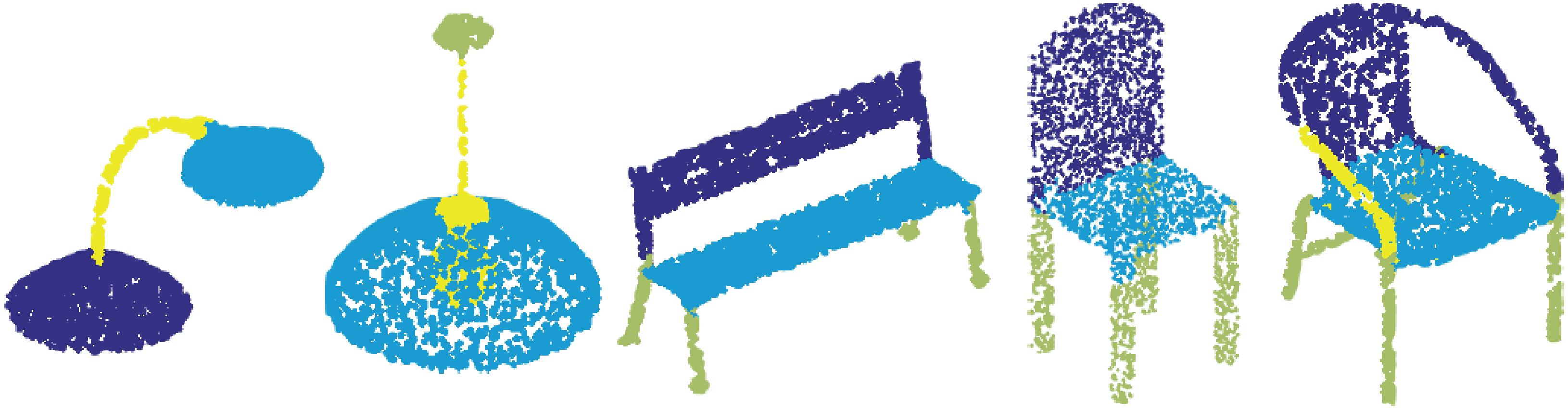
Predicted



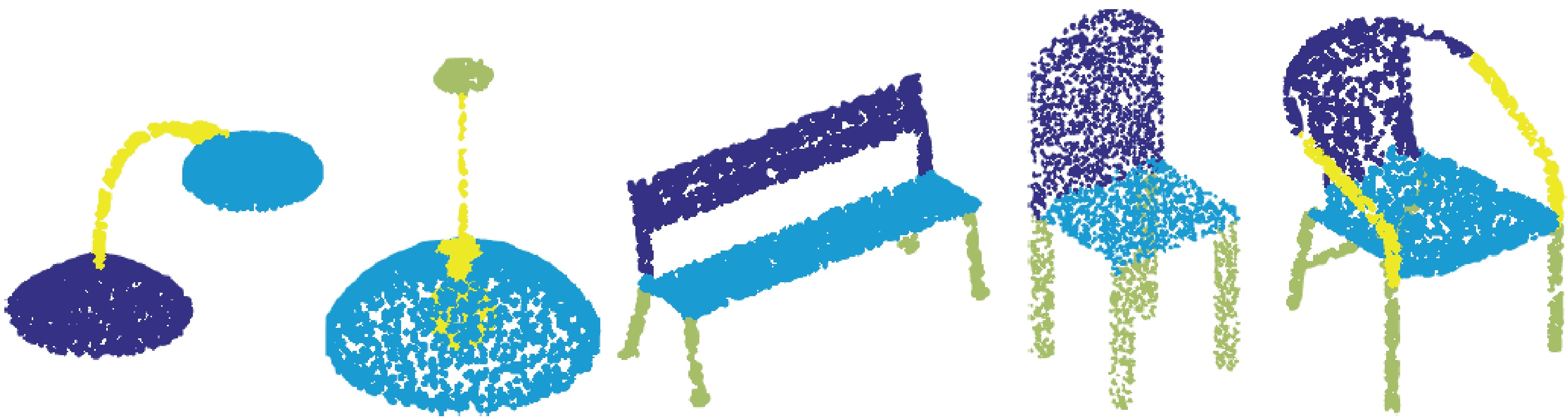
Groundtruth

Yi et al. 2017

Example: shape segmentation with SpecTN



Predicted



Groundtruth

Yi et al. 2017