3D Filament Plots from Optimally Smooth 2D Andrew's Plots

Nate Strawn

July 27th, 2021

Georgetown University

Table of Contents

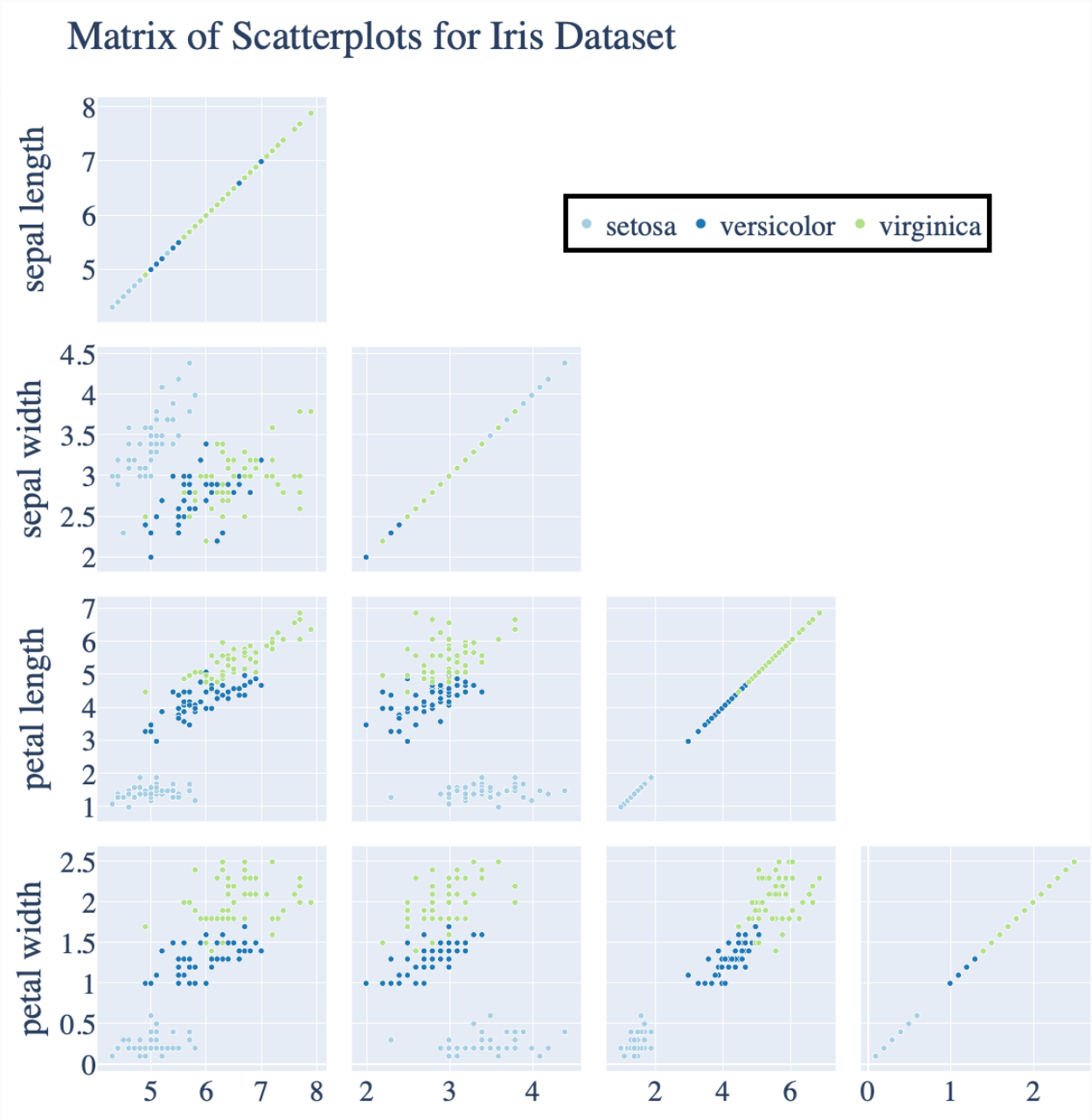
Introduction

Lossy Visualizations

1. Linear projection of data into 2D scatter plots is lossy
2. Non-linear methods for embeddings into 2D scatter plots attempt to mitigate distortion
   - 2.1 Multidimensional Scaling [7]
   - 2.2 t-SNE [9]
   - 2.3 UMAP [10]
3. Non-linear methods also involve more computation

Lossless methods merge systems of projections to provide a complete picture of a dataset

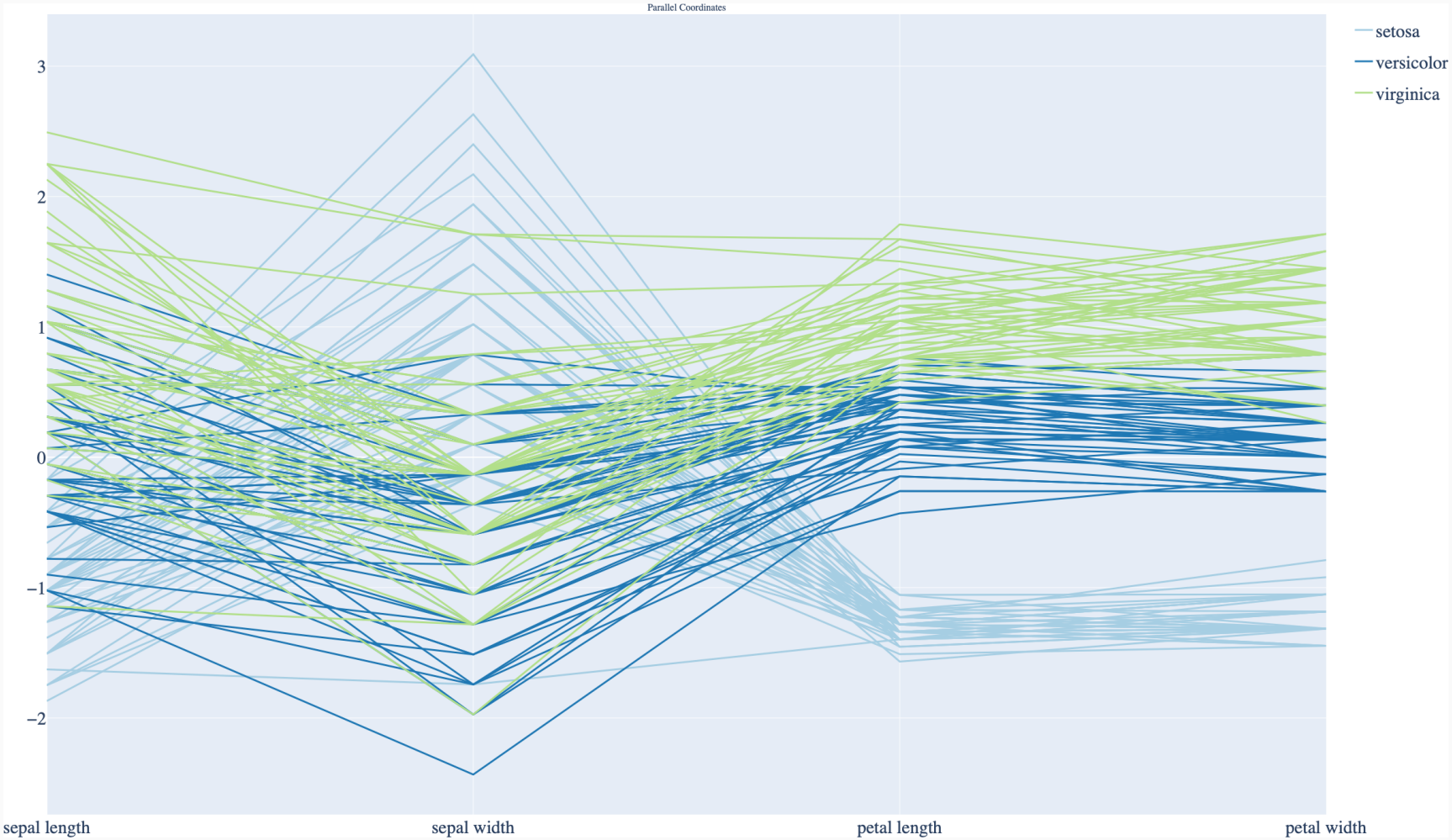
Lossless Visualizations: Matrix of Scatterplots



Lossless Visualizations: Matrix of Scatterplots

1. Also known by "pairs plots" or "draftsman plots" (Tukey [12])
2. Individual data points are not "connected" even in low dimension
3. Similar to "small multiples" plots, these plots are not useful when the data dimension is greater than a couple dozen

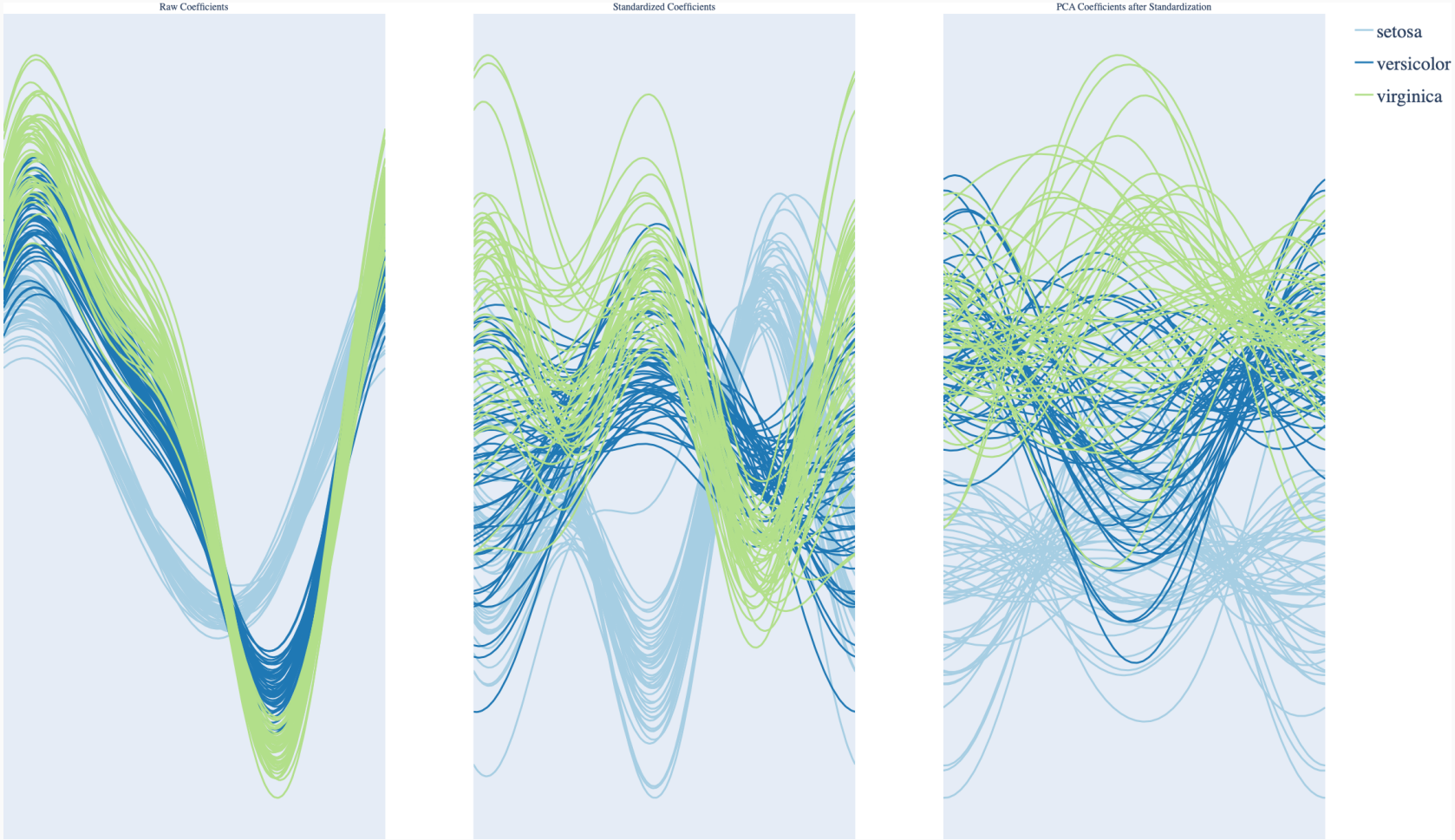
Lossless Visualizations: Parallel Plots



Lossless Visualizations: Parallel Plots

1. Introduced by Inselberg in 1985 [6]
2. Individual data points have connected representations
3. High-frequency linear splines introduce visual noise in high dimension
4. Plotting many data points decreases legibility quickly

Lossless Visualizations: Andrew's Plots



Lossless Visualizations: Andrew's Plots

$\Phi : \mathbb{R}^d \to L^2([0,1])$ given by

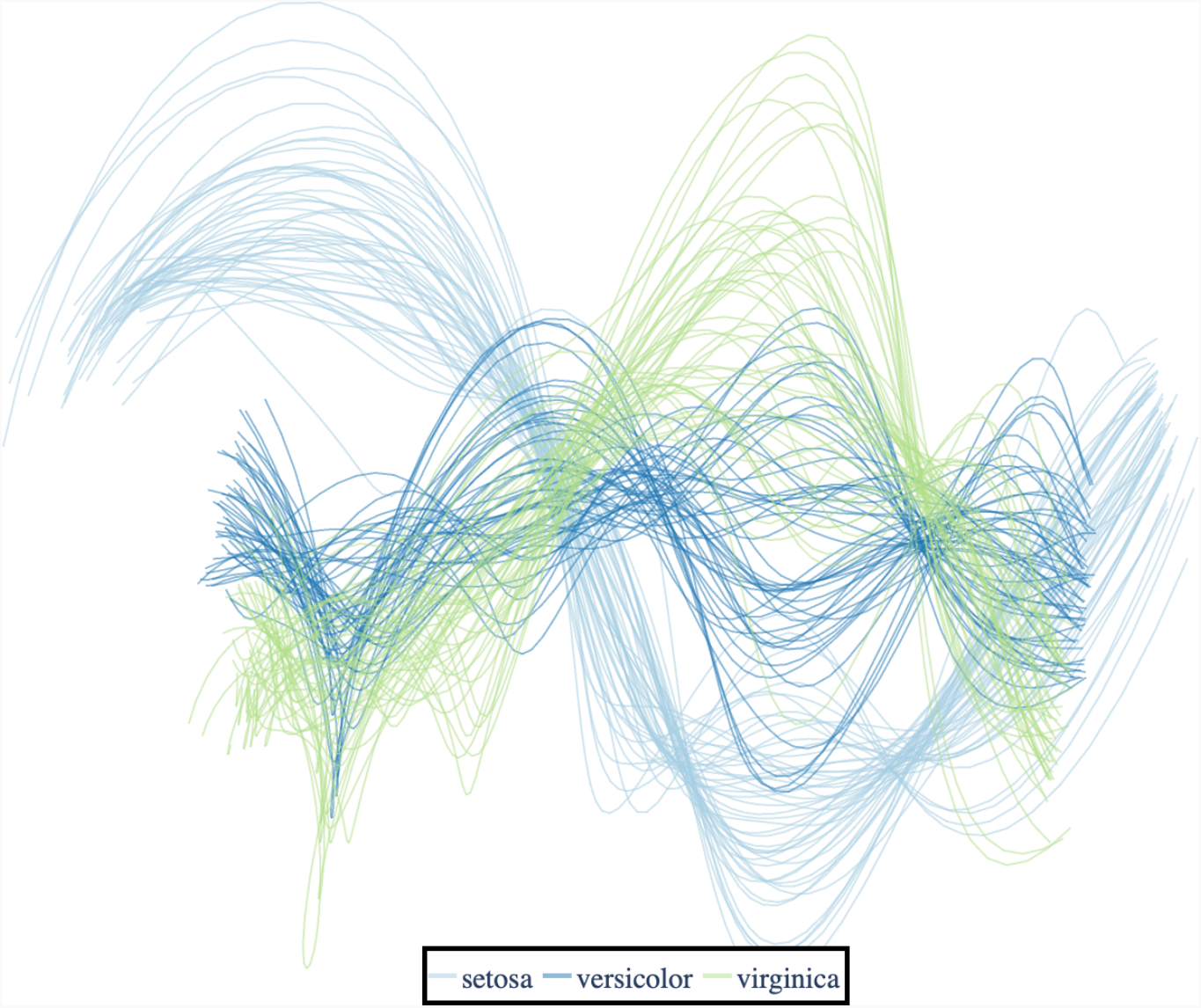
$$\Phi[x](t) = x_1 + x_2\sqrt{2}\cos(2\pi t) + x_3\sqrt{2}\sin(2\pi t) + x_4\sqrt{2}\cos(4\pi t) + x_5\sqrt{2}\sin(4\pi t) + \cdots$$

1. Introduced by Andrews in 1972 [2]
2. Similar benefits to parallel plots, but individual features are obscured
3. Lower frequency components lead to less visual "noise"
4. This map is a linear isometry from $\mathbb{R}^d$ to $L^2([0,1])$, so the bounds

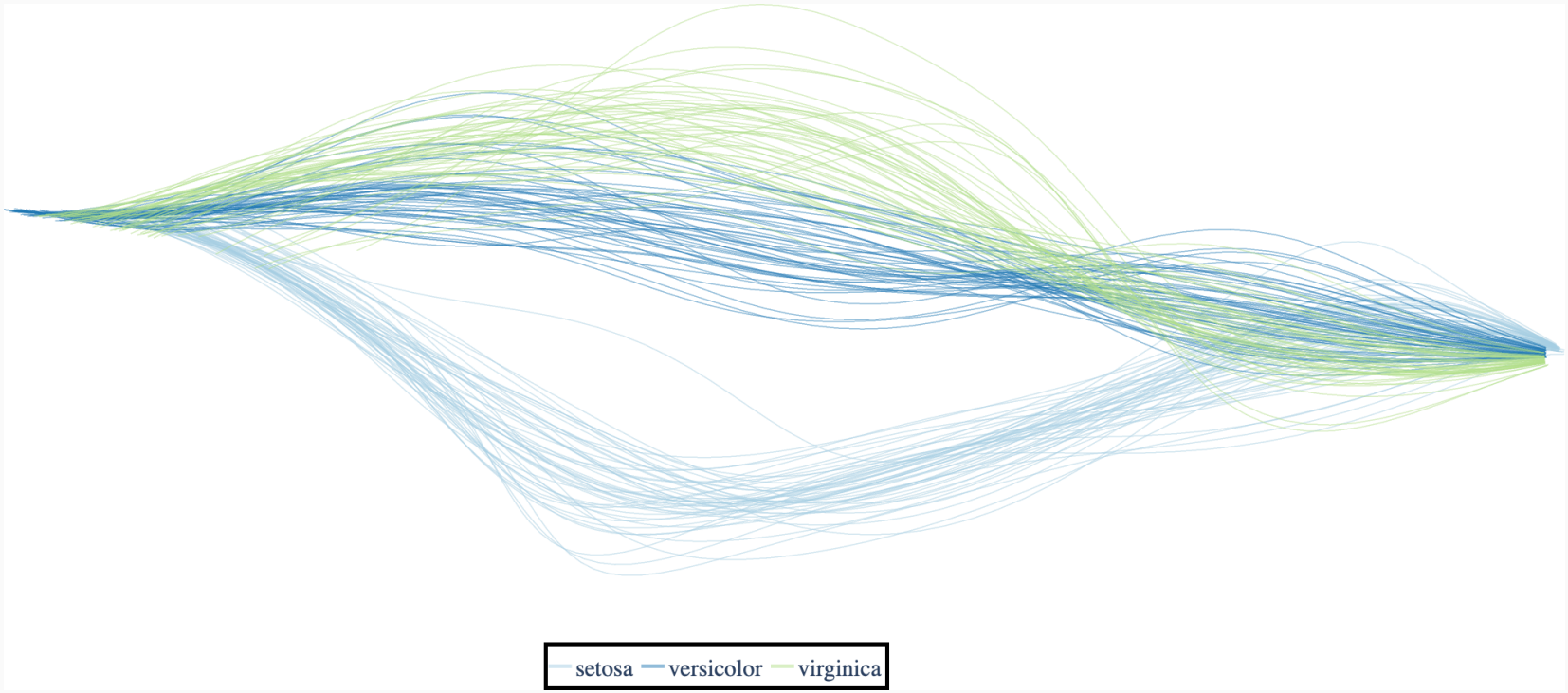
   $$\|\Phi[x] - \Phi[y]\|_{L^1} \leq \|\Phi[x] - \Phi[y]\|_{L^2} \leq \|\Phi[x] - \Phi[y]\|_{L^\infty}$$

   allow visual assessment of $\|x - y\| = \|\Phi[x] - \Phi[y]\|_{L^2}$!
5. Still suffers from the legibility issue

A New Alternative: Optimal 2D Andrews Plots



A New Alternative: Filament Plots



Summary of Contributions

We let $\mathcal{H} = \mathcal{H}^1([0,1])$ denote the Sobolev space of $L^2([0,1])$ curves with $L^2([0,1])$ derivatives.

1. Given a data matrix $X \subset \mathbb{R}^{d \times N}$, we parameterize the linear isometries $\Phi : \mathbb{R}^d \to \mathcal{H}^2$ that produce (on average) optimally smooth curves over the dataset
2. This parameterization reveals many degrees of freedom, and we exhibit a particular set of parameters so that the maps $x \mapsto \Phi[x](t)$ are uniformly close to projections for all $t \in [0,1]$.
3. Using equations like the Frenet-Serret system, we construct unit-length 3D curves from these optimally smooth 2D curves. The isometry then entails that distances in the data space are bounded by "total relative curvature" and "maximum relative curvature" in the curve space.

Code: https://github.com/n8epi/filaments
Preprint: https://arxiv.org/abs/2107.10869

Aesthetic Optimization

Aesthetic Conditions on Linear Isometries to Curves

We let $X \in \mathbb{R}^{d \times N}$ and $\Phi \in \mathcal{L}(\mathbb{R}^d, \mathcal{H}^2)$. We identify several desirable conditions for Φ:

1. **Global non-degeneracy:** to avoid collapsing the graphs, we enforce a isotropic isometry condition so that
$$\|u^T \Phi[x]\|_{L^2} = \|x\| \text{ for all } x \in \mathbb{R}^d, u \in \mathbb{R}^2, \|u\| = 1.$$
2. **Local non-degeneracy:** to avoid local collapsing graphs, we enforce a projective tour condition so that $\Phi[\cdot](t)$ is a (scaled) projection matrix for all $t \in [0, 1]$.
3. **Visual interpretability:** to ensure unbiased, visually accessible interpretations of L^1 and L^∞ bounds, we require that Φ maps to derivatives of periodic curves
4. **Smoothness:** Φ has minimum mean quadratic variation:
$$\text{MQV}(\Phi; X) = \frac{1}{N} \sum_{n=1}^{N} \left\| \frac{d\Phi[x_n]}{dt} \right\|_{L^2}^2.$$

Isotropic Isometry

If we only retain the isometry condition on Φ, then

$$\Phi[x](t) = \begin{pmatrix} x_1 + x_2\sqrt{2}\cos(2\pi t) + x_3\sqrt{2}\sin(2\pi t) + \cdots \\ 0 \end{pmatrix}$$

is permissible, but then the $3D$ graphs of these curves are confined to a single plane.

Projective Tour Property

In the extreme case where the rank of $\Phi(t)$ is always 1, then there is a $\phi \in \mathcal{H}^2$ such that $\|\phi(t)\| = 1$ for all $t \in [0,1]$, and for all $x \in \mathbb{R}^d$ there is a $\gamma_x \in \mathcal{H}$ satisfying

$$\Phi[x](t) = \gamma_x(t)\phi(t).$$

If we enforce a rank 2 condition on $\Phi(t)$, this becomes

$$\Phi[x](t) = \gamma_{x,1}(t)\phi_1(t) + \gamma_{x,2}(t)\phi_2(t)$$

where $\phi_1(t), \phi_2(t)$ are an o.n.b. for all $t \in [0,1]$.

Visual Interpretations

If $\gamma_x' = \Phi[x]$ and $\gamma_y' = \Phi[y]$ are derivatives of curves and Φ is an isometry, then

$$\int_0^1 \|\gamma_x'(t) - \gamma_y'(t)\| \, dt \leq \|x - y\| \leq \sup_{t \in [0,1]} \|\gamma_x'(t) - \gamma_y'(t)\|$$

so the length of the difference curve $\gamma_x - \gamma_y$ is a lower bound for Euclidean distance and the maximum velocity of the difference curve $\gamma_x - \gamma_y$ is an upper bound for the Euclidean distance.

Main Results

"Tensor" Characterization of Linear Maps

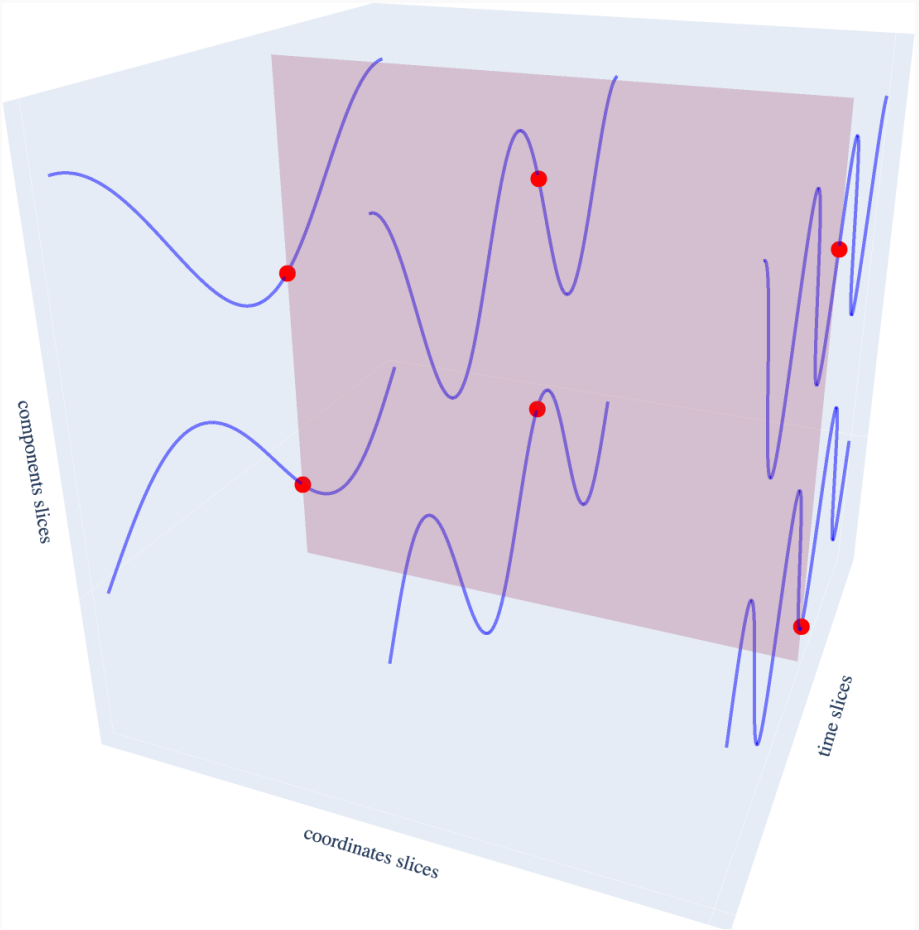
Proposition

$\Omega \in \mathcal{L}(\mathbb{R}^d, \mathcal{H}^2)$ is linear if and only if there is a $\Phi \in \mathcal{H}^{2\times d}$ such that $\Omega[x](t) = \Phi(t)x$ for all $x \in \mathbb{R}^d$ and almost all $t \in [0,1]$.

For $\Phi \in \mathcal{H}^{2\times d}$, we abuse notation and write $\Phi[x]$ to denote the linear transformation of $x \in \mathbb{R}^d$ to $\mathcal{H}^2$.

Mapping "Tensors"

For $\Phi \in \mathcal{H}^{2\times d}$, we call the $\phi^T_{j,\cdot} \in \mathcal{H}^d$ functions the *component slices* of Φ, we call the 2 by d matrices $\Phi(t)$ the *time slices* of Φ, and we call the functions $\phi_{\cdot,k} \in \mathcal{H}^2$ the *coordinate slices* of Φ.



Constraints on the Tensors

1. The isotropic isometry condition becomes the L^2 "quadratic" constraints

$$\int_0^1 \phi_{j,k}(t)\phi_{j',k'}(t)\,dt = \delta_{(j,k),(j',k')} \text{ for all } (j,k),(j',k') \in [2] \times [d].$$

2. The projective tour condition becomes the quadratic constraints

$$\phi_{j,\cdot}(t)\phi_{j',\cdot}(t)^T = d\delta_{j,j'} \text{ for all } j,j' \in [2], t \in [0,1].$$

3. The "derivative of a periodic function" condition is encoded by the linear constraints

$$\phi_{j,k}(0) = \phi_{j,k}(1) \text{ and } \int_0^1 \phi_{j,k}(t)\,dt = 0 \text{ for all } (j,k) \in [2] \times [d].$$

Full MMQV Program

$$\min_{\Phi \in \mathcal{H}^{2\times d}} \mathrm{MQV}(\Phi; X)$$

subject to

$$\int_0^1 \phi_{j,k}(t)\phi_{j',k'}(t)\,dt = \delta_{(j,k),(j',k')} \text{ for all } (j,k),(j',k') \in [2]\times[d],$$

$$\phi_{j,\cdot}(t)\phi_{j',\cdot}(t)^T = d\delta_{j,j'} \text{ for all } j,j' \in [2], t \in [0,1],$$

$$\phi_{j,k}(0) = \phi_{j,k}(1), \text{ and } \int_0^1 \phi_{j,k}(t)\,dt = 0 \text{ for all } (j,k) \in [2]\times[d].$$

Partial MMQV Program

These combined constraints produce a difficult optimization problem. However, if we only enforce the isotropic isometry and periodic derivative constraint, we get

$$\min_{\Phi \in \mathcal{H}^{2\times d}} \mathrm{MQV}(\Phi; X)$$

subject to

$$\int_0^1 \phi_{j,k}(t)\phi_{j',k'}(t)\,dt = \delta_{(j,k),(j',k')} \text{ for all } (j,k),(j',k') \in [2]\times[d],$$

$$\phi_{j,k}(0) = \phi_{j,k}(1), \text{ and } \int_0^1 \phi_{j,k}(t)\,dt = 0 \text{ for all } (j,k) \in [2]\times[d].$$

Parameterization of Solutions for Partial MMQV Program

Theorem (S. 2021)

If $X = U\Sigma V^T$ *with* $U = \begin{pmatrix} u_1 & u_2 & \cdot & u_d \end{pmatrix}$, *the* Φ *solves the partial MMQV program if and only if there exist unimodular constants* $\{\omega_k\}_{k=1}^d \in \mathbb{C}$ *and* $\{w_k\}_{k=1}^d \subset \{-1, 1\}$ *such that*

$$\Phi[x](t) = \sum_{k=1}^{d} \left(u_k^T x\right) \omega_k e^{2\pi i w_k k t}$$

under the identification of $\mathbb{R}^2$ *with* $\mathbb{C}$.

Example: Iris Dataset

The U matrix for the Iris dataset after standardization is

$$U = \begin{pmatrix} u_1 & u_2 & u_3 & u_3 \end{pmatrix} \approx \begin{pmatrix} 0.52 & -0.26 & 0.58 & 0.56 \\ -0.38 & -0.92 & -0.02 & -0.07 \\ 0.72 & -0.24 & -0.14 & -0.63 \\ 0.26 & -0.12 & -0.80 & 0.52 \end{pmatrix},$$

and one optimally smooth map is

$$x \mapsto \left(u_1^T x\right) e^{2\pi i t} + \left(u_2^T x\right) e^{4\pi i t} + \left(u_3^T x\right) e^{6\pi i t} + \left(u_4^T x\right) e^{8\pi i t}.$$

Proof Sketch

1. The vector-valued discrete time Fourier transform diagonalizes the differential operator
2. We get a lower bound on the MMQV by truncating the diagonal
3. Fischer-Courant ensures that the minimizers of the truncated problem match high-variability directions in the data with low-frequency Fourier components

Approximate Projective Tour Property

Using this parameterization of the system of minimizers, we have enough room to choose a particular member that is a projective tour in an asymptotic sense:

Theorem (S. 2021)

With $\omega_k = e^{2\pi i k^2/4d}$ and $w_k = 1$ for all $k \in [d]$, we have that the Φ from the last theorem satisfies the condition that $\sqrt{\frac{2}{d}}\Phi[\cdot](t)$ has singular values in the interval

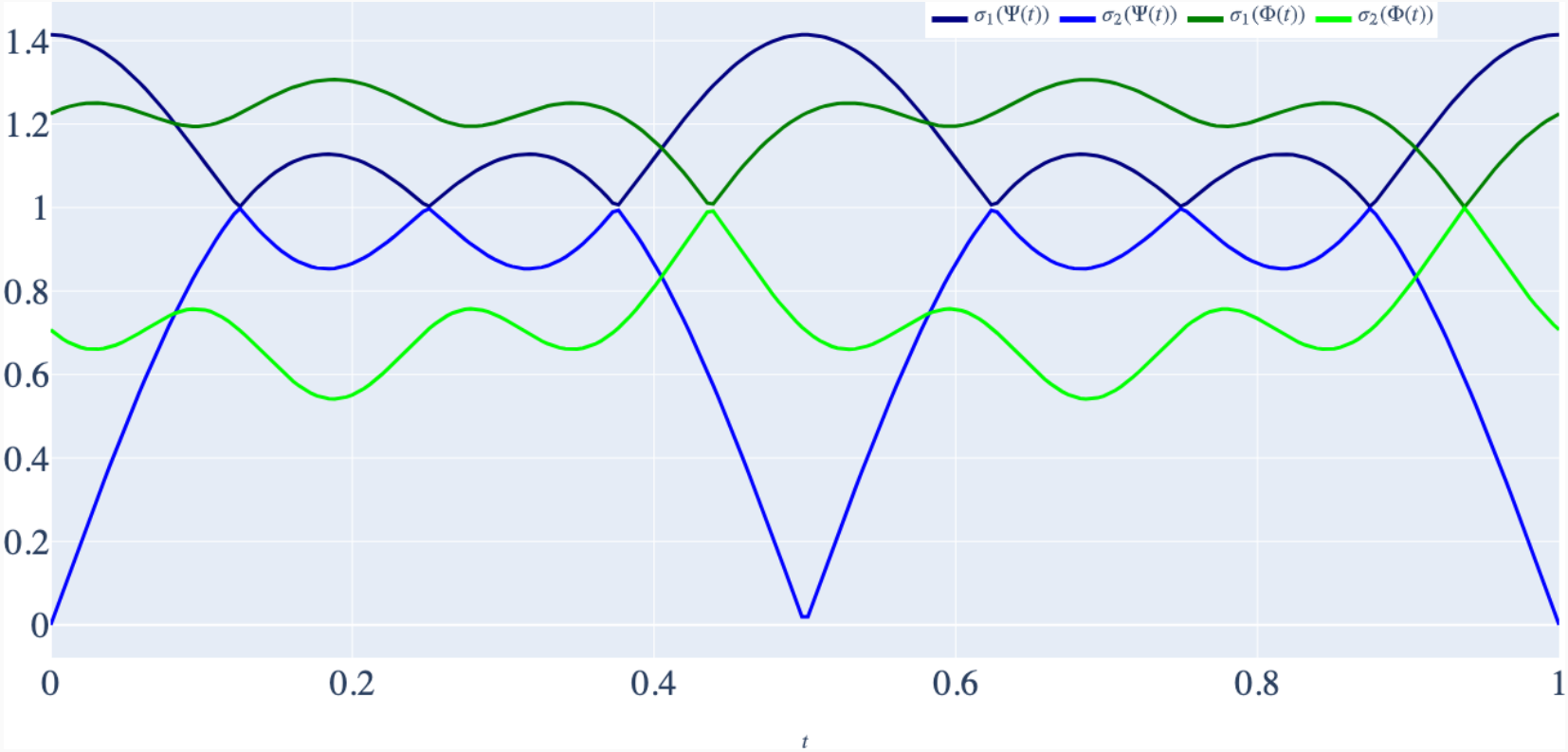
$$\left[\sqrt{1-\left(\frac{4}{\sqrt{d}}+\frac{2}{d}+\frac{2}{d^2}\right)}, \sqrt{1+\left(\frac{4}{\sqrt{d}}+\frac{2}{d}+\frac{2}{d^2}\right)}\right].$$

for all $t \in [0,1]$

Example: Iris Dataset

With U in the last example, we now map x to the function

$$\left(u_1^T x\right) e^{\pi i/8} e^{2\pi i(t+1/8)} + \left(u_2^T x\right) e^{\pi i/2} e^{4\pi i t} + \left(u_3^T x\right) e^{9\pi i/8} e^{6\pi i t} + \left(u_4^T x\right) e^{8\pi i t}.$$



Proof Sketch

1. Relax the result that $\{f_k\}_{k=1}^d \subset \mathbb{R}^2$ is tight if and only if $\sum_{k=1}^d z_k^2 = 0$ to a perturbative condition relating the sum of squares to the singular values of the F matrix.
2. Invoke uniform bounds on generalized quadratic Gaussian sums [11]:

$$S_d(\xi, \theta) = \sum_{k=1}^{d} \exp(\pi i \xi k^2 + 2\pi i k \theta) = O(1/\sqrt{\xi})$$

with $\xi = 1/d$ and $\theta = t$ to bound the complex sum of squares

Discussion

1. The proof of Theorem 1 may be reconfigured to show that 1D Andrew's plots using PCA scores optimize the MQV
2. The bounds $L^1 \leq L^2 \leq L^\infty$ have a $O(\sqrt{d})$ gap
3. Using Johnson-Lindenstrauss [1], a data set of size N may be embedded in $O(\log N)$ dimensions with small distortion, so the gap (for a general N-point dataset) becomes $O(\sqrt{\log N})$ for approximating distances visually

Curves via Frenet-Serret Equations

Moving Frames

A moving frame is a function $U : [0, 1] \to \mathcal{O}(3)$. The classic moving frame for an arc-length parameterized curve is

$$U(t) = \begin{pmatrix} \mathbf{T}(t) \\ \mathbf{N}(t) \\ \mathbf{B}(t) \end{pmatrix}$$

where $\mathbf{T}$ is the tangent to the curve, $\mathbf{N} = d\mathbf{T}/\|d\mathbf{T}\|$ is the normal to the curve, and $\mathbf{B} = \mathbf{T} \times \mathbf{N}$ is the binormal to the curve.

Classic Frenet-Serret

The classic Frenet-Serret equations are

$$\begin{pmatrix} d\mathbf{T}(t) \\ d\mathbf{N}(t) \\ d\mathbf{B}(t) \end{pmatrix} = \begin{pmatrix} 0 & \kappa(t) & 0 \\ -\kappa(t) & 0 & \tau(t) \\ 0 & -\tau(t) & 0 \end{pmatrix} \begin{pmatrix} \mathbf{T}(t) \\ \mathbf{N}(t) \\ \mathbf{B}(t) \end{pmatrix}$$

where $\kappa(t) = \|d\mathbf{T}(t)\|$ and $\tau(t) = \|d\mathbf{B}(t)\|$ are the curvature and torsion functions (respectively).

Symmetric Frenet-Serret

We consider an arc-length parameterized curve with a moving frame that satisfies

$$\begin{pmatrix} d\mathbf{T} \\ d\mathbf{N}_1 \\ d\mathbf{N}_2 \end{pmatrix} = \begin{pmatrix} 0 & \kappa_1(t) & \kappa_2(t) \\ -\kappa_1(t) & 0 & 0 \\ -\kappa_2(t) & 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{T} \\ \mathbf{N}_1 \\ \mathbf{N}_2 \end{pmatrix}$$

This is different from the usual Frenet-Serret system because we are not writing the frame in terms of the normal and binormal.

Consequences

Since $d\mathbf{T} = \kappa_1 \mathbf{N}_1 + \kappa_2 \mathbf{N}_2$, the curve obtained by integrating $\mathbf{T}$ has the curvature function

$$\kappa(t) = \|d\mathbf{T}(t)\| = \sqrt{\kappa_1(t)^2 + \kappa_2(t)^2}.$$

Moreover, the binormal satisfies

$$\mathbf{B}(t) = \frac{\kappa_2(t)}{\kappa(t)}\mathbf{N}_1(t) - \frac{\kappa_1(t)}{\kappa(t)}\mathbf{N}_2(t) = \sin(\theta(t))\mathbf{N}_1(t) - \cos(\theta(t))\mathbf{N}_2(t),$$

and taking a derivative yields

$$\mathbf{B}' = \theta' \cos(\theta)\mathbf{N}_1 + \theta' \sin(\theta)\mathbf{N}_2.$$

Hence, the torsion of the curve is $\tau(t) = |\theta'|$.

Consequences

Viewing $\kappa_1(t) + i\kappa_2(t) = \kappa(t)e^{i\theta(t)}$, we have

$$\begin{aligned}
\frac{d\kappa_1}{dt}(t)^2 + \frac{d\kappa_2}{dt}(t)^2 &= \left| \frac{d}{dt}\left[\kappa(t)e^{i\theta(t)}\right] \right|^2 \\
&= \left| \frac{d\kappa}{dt}(t)e^{i\theta(t)} + i\frac{d\theta}{dt}(t)\kappa(t)e^{i\theta(t)} \right|^2 \\
&= \left| \frac{d\kappa}{dt}(t) + i\frac{d\theta}{dt}(t)\kappa(t) \right|^2 \\
&= \left(\frac{d\kappa}{dt}(t) \right)^2 + \left(\kappa(t)\tau(t)\right)^2 .
\end{aligned}$$

Therefore the MMQV program provides control over the derivative of the curvature function and the inner product of the squared curvature function and the squared torsion in L^2.

Consequences

1. Isometry is now to a space of "symmetric" curvature functions, so the isometry preserves distances through relative curvatures
2. Isotropic isometry ensures that directions of evolution of a given curve are unbiased
3. Minimizing the mean quadratic variation minimizes the square L^2 norm of the derivative of curvature and the L^2 inner product of the squared torsion function and the squared curvature function

Computation: Rodrigues Formula

To obtain the curve from MMQV embeddings into symmetric curvature functions, we compute an integral of a matrix exponential

$$e_1^T \int_0^t \exp \begin{pmatrix} 0 & \lambda_1(s) & \lambda_2(s) \\ -\lambda_1(s) & 0 & 0 \\ -\lambda_2(s) & 0 & 0 \end{pmatrix} ds$$

which becomes

$$e_1^T \int_0^t \left(I_3 + \sin(\lambda(s))B(s) + (1 - \cos(\lambda(s)))B(s)^2 \right) ds$$

for

$$B(t) = \frac{1}{\lambda(s)} \begin{pmatrix} 0 & \lambda_1(s) & \lambda_2(s) \\ -\lambda_1(s) & 0 & 0 \\ -\lambda_2(s) & 0 & 0 \end{pmatrix}, \lambda(s) = \sqrt{\lambda_1(s)^2 + \lambda_2(s)^2}.$$

Hyper Frames

Frames

A frame for a Hilbert space $\mathbb{H}$ with inner product $\langle \cdot, \cdot \rangle$ is a collection $\{f_\alpha\}_{\alpha \in \Lambda} \subset \mathbb{H}$ such that there exists constants $0 < A \leq B$ with

$$A\|x\|^2 \leq \sum_{\alpha \in \Lambda} |\langle x, f_\alpha \rangle|^2 \leq B\|x\|^2 \text{ for all } x \in \mathbb{H}$$

If A and B are optimal, we call them the frame constants, and if $A = B$ the frame is called tight.

Fusion Frames

A fusion frame [5] of $\mathbb{R}^d$ of rank k is a collection of rank k orthogonal projections $\{P_j\}_{j=1}^M$ such that there exist constants $0 < A \leq B$ with

$$A\|x\|^2 \leq \sum_{\alpha \in \Lambda} \|P_j x\|^2 \leq B\|x\|^2 \text{ for all } x \in \mathbb{R}^d.$$

Super Frames

A super frame [4] of $\mathbb{R}^d$ to $\mathbb{R}^2$ is a pair of frames $F = \{f_j\}_{j=1}^M$ and $G = \{g_j\}_{j=1}^M$ of $\mathbb{R}^d$ such that there exist constants $0 < A \leq B$ with

$$A(\|x\|^2 + \|y\|^2) \leq \sum_{j=1}^{M} (\langle f_j, x\rangle + \langle g_j, y\rangle)^2 \leq B(\|x\|^2 + \|y\|^2)$$

for all $x, y \in \mathbb{R}^d$.

If $A = B = 1$ we have a Parseval super frame, which is equivalent to

$$x \mapsto \begin{pmatrix} \langle f_1, x\rangle & \cdots & \langle f_M, x\rangle \\ \langle g_1, x\rangle & \cdots & \langle g_M, x\rangle \end{pmatrix}$$

is an isotropic isometry from $\mathbb{R}^d$ to the set of $\mathbb{R}^2$-valued functions over $[M]$.

Hyper Frames

A hyper frame of $\mathbb{R}^d$ to $\mathbb{R}^2$ is a Parseval super frame $\{F, G\}$ such that the d by d matrices

$$P_j = \frac{N}{d} \begin{pmatrix} f_j & g_j \end{pmatrix} \begin{pmatrix} f_j & g_j \end{pmatrix}^T$$

form a fusion frame of $\mathbb{R}^d$ of rank 2.

These are the finite-dimensional analogues of systems that satisfy the isotropic isometry and projective tour conditions!

Complex Characterization of Planar Hyper Frames

Proposition

Suppose $\{F, G\}$ *is a super frame of* $\mathbb{R}^d$ *to* $\mathbb{R}^2$ *and set* $Z = F + iG$. *Then* $\{F, G\}$ *is a hyper frame of* $\mathbb{R}^d$ *to* $\mathbb{R}^2$ *if and only if*

$$ZZ^T = 0, ZZ^* = 2I_d, diag(Z^T Z) = 0, diag(Z^* Z) = \frac{2d}{N}\mathbf{1}_M.$$

Open problems: constructions and hyper-frame potential minimization

Conclusion and Future Directions

Sup Norm Embeddings

Linial, London, and Rabinovich [8] provide isometric embeddings of finite metric spaces into ℓ^∞. Is there an aesthetically appealing way to isometrically embed a finite collection of data into $(L^\infty)^3$?

Hausdorff Metric Embeddings

Since we want static curves in $\mathbb{R}^3$, is it possible to compute smooth embeddings $x \mapsto \gamma_x$ such that

$$\|x - y\| = \min_t \max_s \|\gamma_x(t) - \gamma_y(t)\|?$$

Closed Curve Problem

Ideally, we are targeting closed curves, but inducing closed curves from embeddings into curvature/torsion is complicated by a lack of useful resolution for the closed curve problem [3]

Questions?

References I

[1] Nir Ailon and Bernard Chazelle. The fast johnson–lindenstrauss transform and approximate nearest neighbors. *SIAM Journal on computing*, 39(1):302–322, 2009.

[2] David F Andrews. Plots of high-dimensional data. *Biometrics*, pages 125–136, 1972.

[3] Josu Arroyo, Oscar J Garay, and José J Mencía. When is a periodic function the curvature of a closed plane curve? *The American Mathematical Monthly*, 115(5):405–414, 2008.

[4] Radu V Balan. Multiplexing of signals using superframes. In *Wavelet Applications in Signal and Image Processing VIII*, volume 4119, pages 118–129. International Society for Optics and Photonics, 2000.

[5] Peter G Casazza and Gitta Kutyniok. Frames of subspaces. *Contemporary Mathematics*, 345:87–114, 2004.

References II

[6] Alfred Inselberg. The plane with parallel coordinates. *The visual computer*, 1(2):69–91, 1985.

[7] Joseph B Kruskal. *Multidimensional scaling*. Number 11. Sage, 1978.

[8] Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15 (2):215–245, 1995.

[9] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008.

[10] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger. Umap: Uniform manifold approximation and projection. *Journal of Open Source Software*, 3(29):861, 2018.

References III

[11] RB Paris. An asymptotic expansion for the generalised quadratic gauss sum revisited. *Journal of Classical Analysis*, 5(1):15–24, 2014.

[12] PA Tukey and JW Tukey. Graphical display of data sets in 3 or more dimensions. *Interpreting multivariate data*, 189:275, 1981.