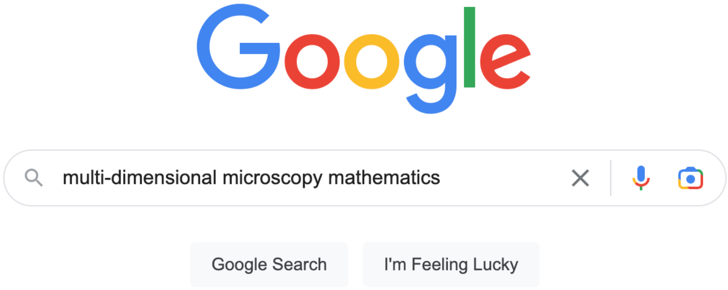
PCA is Not Dead: Vectorized persistent homology and flag medians

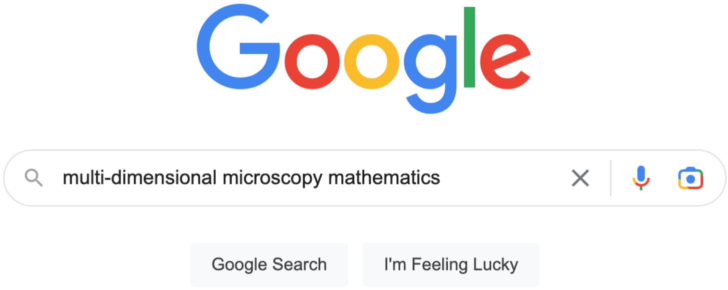
(plus other stuff)

Emily J. King

Mathematics Department, Colorado State University

Mathematical Advances for Multi-Dimensional Microscopy
IPAM, Los Angeles
October 26, 2022


Google
multi-dimensional microscopy mathematics
Google Search
I'm Feeling Lucky



Optimist's view:
This talk will be 100% controversy-free.

Tensor Decomposition?

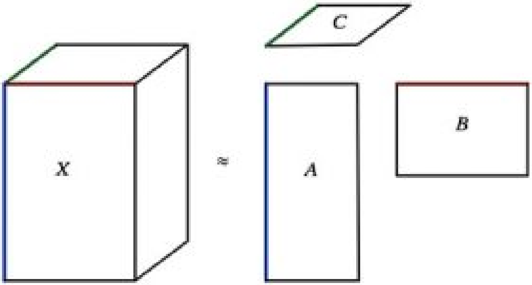


Figure from IPAM workshop page. Apparently because of Paul?

~~Tensor~~ Matrix Decomposition

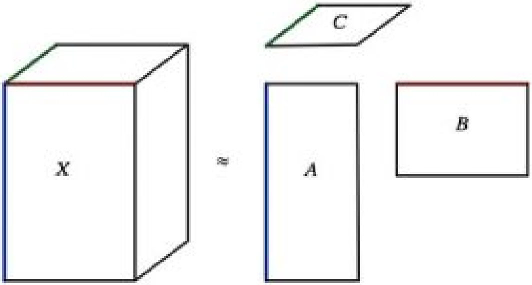
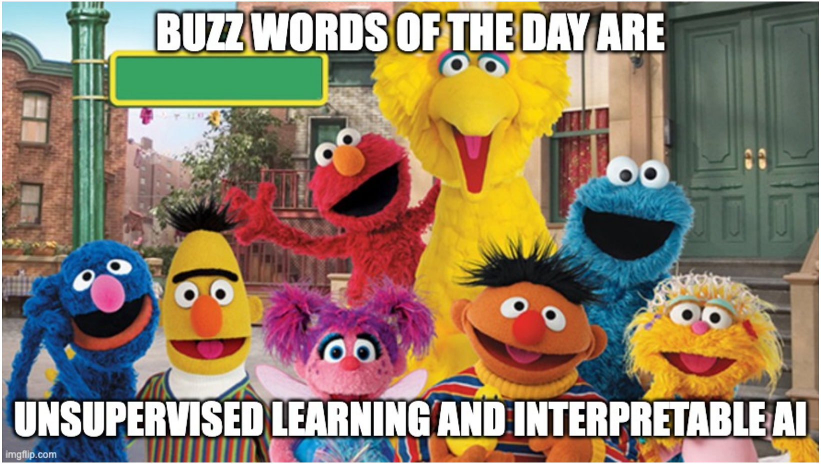


Figure from IPAM workshop page. Apparently because of Paul?



Generated using `https://imgflip.com/`

Unsupervised learning:
Can a model be learned without a large set of labeled data?

Interpretable AI:
Can a human gain intuition behind model outputs?

Desirable in fields like atmospheric/environmental science and multidimensional microscopy?,
where the amount of (labeled) training data can be small but they don't trust black boxes.

Key idea of principal component analysis (PCA):

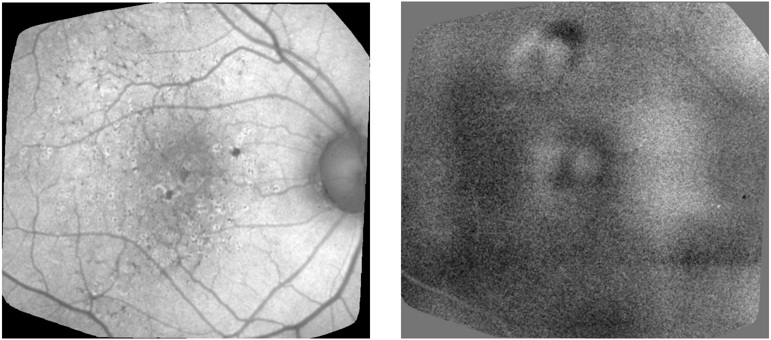
Schmidt-Eckart-Mirsky-Young Theorem:
If $\boldsymbol{USV}^*$ is the singular value decomposition (SVD) of $\boldsymbol{A}$, then

$$\boldsymbol{U}\left(\begin{array}{cccc|c} \sigma_1 & \mathbf{0} & \dots & \mathbf{0} & \\ \mathbf{0} & \sigma_2 & \dots & \mathbf{0} & \\ \vdots & \vdots & \ddots & \vdots & 0 \\ \mathbf{0} & \mathbf{0} & \dots & \sigma_k & \\ \hline & & 0 & & 0 \end{array}\right)\boldsymbol{V}^*$$

is the best (at most) rank $\boldsymbol{k}$ approximation[1] of $\boldsymbol{A}$.

[1] with respect to Frobenius and spectral norms

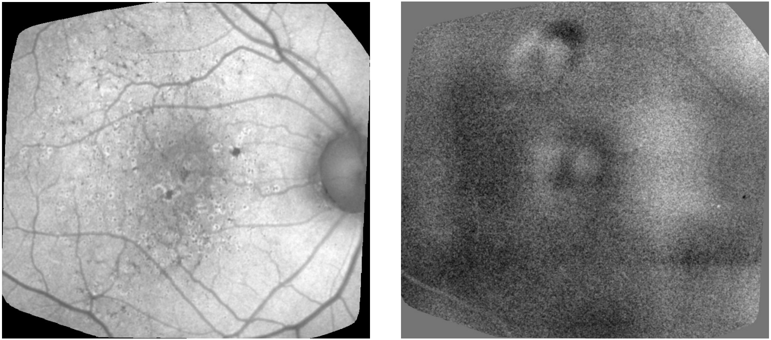
PCA: Doesn't Hurt to Try



The first and second principal components of a stack of four fundus camera images of the retina of a patient with age-related macular degeneration.

Ehler, Majumdar, K, et al. 2010; original data source: National Eye Institute

PCA: Doesn't Hurt to Try*



The first and second principal components of a stack of four fundus camera images of the retina of a patient with age-related macular degeneration.

* It obviously doesn't always work, see Ivan's talk.

Ehler, Majumdar, K, et al. 2010; original data source: National Eye Institute

Lecture Outline

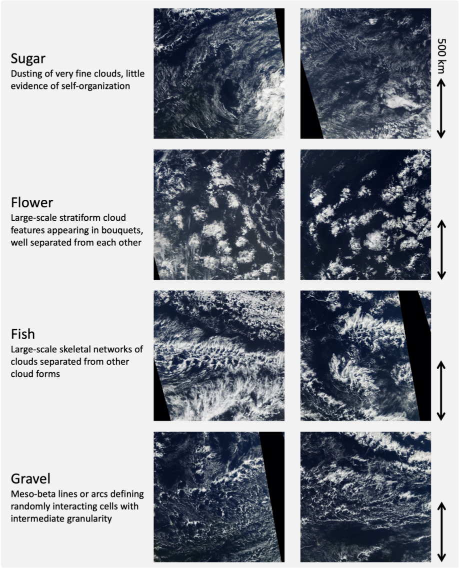
Vectorized Persistent Homology (Atmospheric Science)

Other Stuff (Y'all's Science?)

- More Persistence of Topology and Geometry
- Shearlet-Based Feature Extraction and Inpainting
- Geometry of Adversarial Attacks

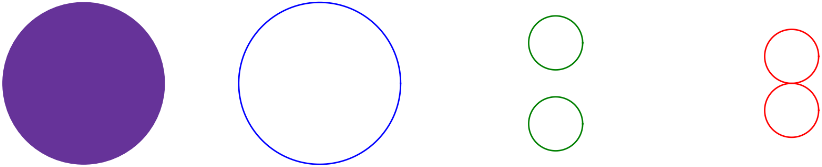
Flag Medians (Computer Vision)

Mesoscale Organization of Clouds



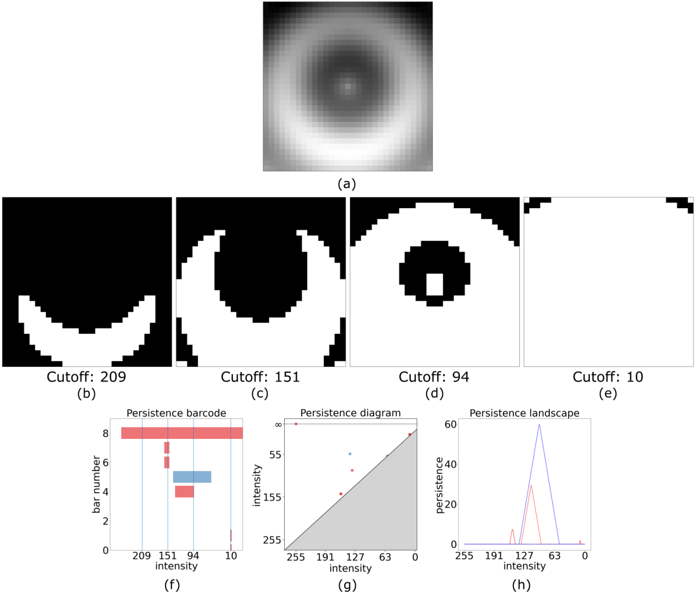
Examples of the four cloud types from the Sugar, Fish, Flowers, and Gravel dataset, Rasp et al. 2019

Betti numbers: learn homology in one slide

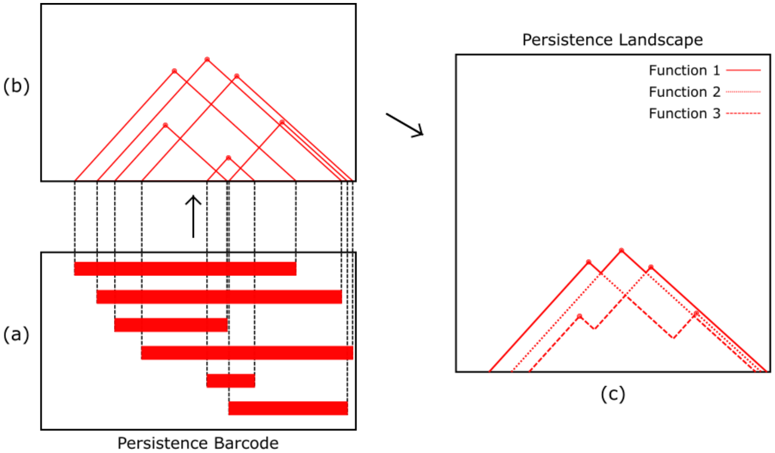


b_0	1	1	2	1
b_1	0	1	2	2

SUPERLEVELSET PERSISTENT HOMOLOGY

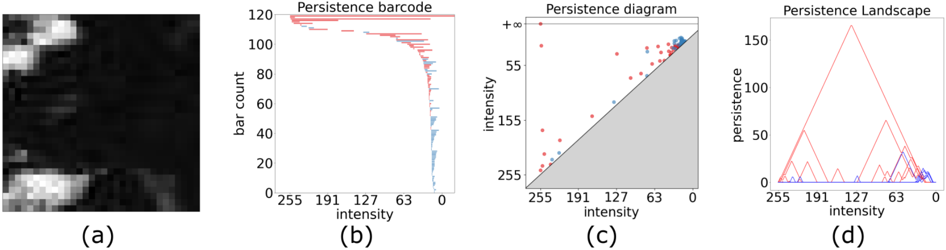


ver Hoef, Adams, K, Ebert-Uphoff 2022



The process of computing a vectorized persistence landscape from a barcode. In (c), only the first three landscape functions are shown. In this case, the first function would be sampled, followed by the second then the third to create a vector.

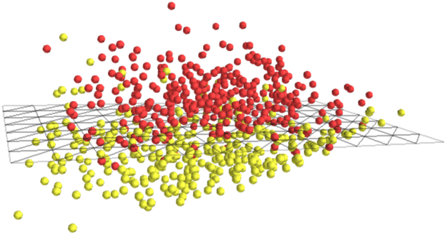
ver Hoef, Adams, K, Ebert-Uphoff 2022



(a) A **32** $\times$ **32** sample image from the grayscale MODIS imagery used in the sugar, flowers, fish, and gravel dataset, along with (b) its persistence barcode, (c) persistence diagram, and (d) persistence landscape with the first 5 piecewise-linear functions for each of the $\boldsymbol{H_0}$ and $\boldsymbol{H_1}$ classes.

ver Hoef, Adams, K, Ebert-Uphoff 2022

Key idea: PCA may be applied as an unsupervised learning method after using an interpretable method like superlevelset persistent homology.



After mapping to $\mathbb{R}^{2000}$ by vectorizing superlevel set persistence landscapes to capture geometric traits, PCA performed down to $\mathbb{R}^3$ – capturing about **90%** of the variation – for separation.

ver Hoef, Adams, K, Ebert-Uphoff 2022

Lecture Outline

Vectorized Persistent Homology
(Atmospheric Science)

Other Stuff
(Y'all's Science?)

- More Persistence of Topology and Geometry
- Shearlet-Based Feature Extraction and Inpainting
- Geometry of Adversarial Attacks

Flag Medians
(Computer Vision)

Lecture Outline

Vectorized Persistent Homology
(Atmospheric Science)

Other Stuff
(Y'all's Science?)

- More Persistence of Topology and Geometry
- Shearlet-Based Feature Extraction and Inpainting
- Geometry of Adversarial Attacks

Flag Medians
(Computer Vision)

Key Idea:

Bryan and I had a few offline talks about the importance of mesoscale features,

on a larger scale than noise

on a smaller scale than curvature / average characteristics

Persistent Homology of Point Clouds

Key Idea: Track the birth and death of various dimensional holes to distinguish structure from chaos.

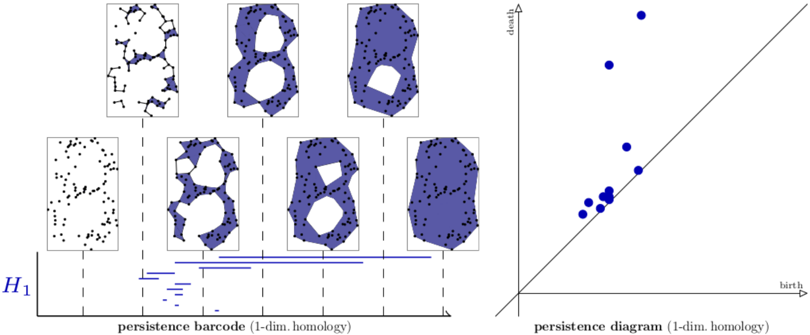
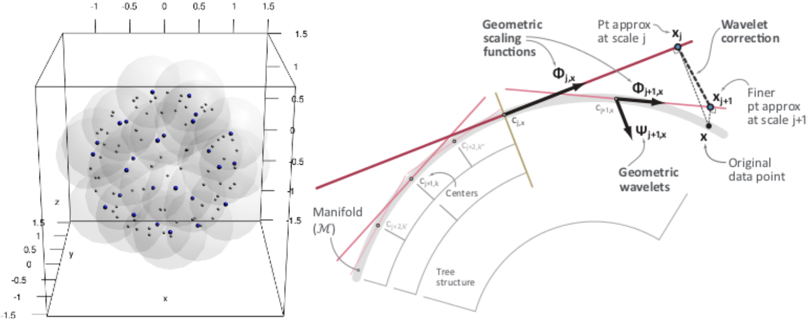


Image source: Jan Senge (Uni Bremen)

Geometric Multiresolution Analysis

Geometric multiresolution analysis is local PCA plus multiscale structure.



Source: Strawn 2015; E. Monson via Allard, Chen, Maggioni 2011

Note! "Geometric multiresolution analysis" also used to describe shearlet/curvelet analysis, especially ≈10 years ago.

Lecture Outline

Vectorized Persistent Homology
(Atmospheric Science)

Other Stuff
(Y'all's Science?)

- More Persistence of Topology and Geometry
- Shearlet-Based Feature Extraction and Inpainting
- Geometry of Adversarial Attacks

Flag Medians
(Computer Vision)

Key idea of applied harmonic analysis:

For data f,

$$\left|\langle f, \varphi_j \rangle\right|,$$

where $\{\varphi_j\}_j$ is often the orbit of a unitary representation, is the "amount" of a trait parameterized by j in f.

E.g.,
Fourier: $\left|\langle f, e^{2\pi i \gamma \cdot} \rangle\right|$ is the "amount" of frequency γ in f.

Key idea of applied harmonic analysis:

For data f,

$$\left|\langle f, \varphi_j \rangle\right|,$$

where $\{\varphi_j\}_j$ is often the orbit of a unitary representation, is the "amount" of a trait parameterized by j in f.

E.g.,

Fourier: $\left|\langle f, e^{2\pi i \gamma \cdot} \rangle\right|$ is the "amount" of frequency γ in f.

Shearlet: $\left|\langle f, 2^{3j/4} \psi(S_\ell A_{2^j} \cdot - k) \rangle\right|$ is the "amount" of features of scale j and orientation ℓ at position k in f.

Key idea of neural networks:

Pairing non-linearity with smartly chosen linear maps is very powerful.

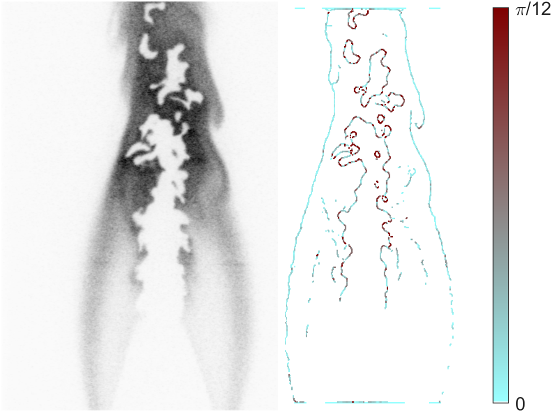
Flame Image

Planar Laser-Induced Fluorescence (PLIF) image of excited ***OH***.
Data source: Johannes Kiefer

Understanding the location and curvature of flame fronts is critical to better understand turbulence and verify computer simulations.

Flame Front Detection with Shearlets

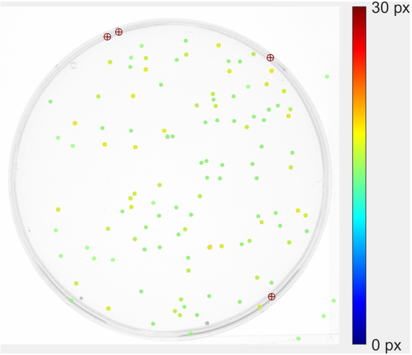
$$\mathrm{E}_{\psi}(f,x,s) = \max\left\{0,\ \frac{\left|\sum_{a\in A}\langle f,\psi^{(o)}{}_{a,s,x}\rangle\right| - \sum_{a\in A}\left|\langle f,\psi^{(e)}{}_{a,s,x}\rangle\right| - \beta\,|A|\,K_{\psi}}{|A|\max_{a\in A}\left|\langle f,\psi^{(o)}{}_{a,s,x}\rangle\right| + \epsilon}\right\},$$



L: PLIF visualization of OH. R: Detected flame front with curvature using shearlets.

K., Reisenhofer, et al. 2015; Kiefer, Reisenhofer, K. 2015; Reisenhofer, K 2019

Cell Colony Detection



Estimated colony widths. **Source of original petri dish image**: Quentin Geissmann 2013

In comparison with three other automated cell colony counting methods, SymFD is best or second best.

(Handles webcam images uniformly better.)

Reisenhofer, K 2019; cf. probably waaaaaaaay faster method from Eric's talk

Inpainting Examples

An issue in data analysis is that of incomplete data. Some examples of desired inpainting / image completion / data recovery are

- the repair of scratched photos and audio recordings,
- the completion of microscopy data,
- the filling in of webcam videos of your grad student, and
- the removal of overlaid text in images, or . . .

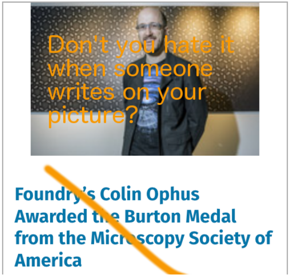


Image source: Hadn't yet brought up Colin.

Cf. Nigel's talk

INPAINTING EXAMPLES, CONT.

- …the erasure of people from history who got in your way.

Original photo Gazeta 1937. Doctored photo unknown source. Downloaded from `https://www.indexoncensorship.org/2017/08/commissar-vanishes/`.

Key Idea:

Assume that
$\mathbf{\Phi_1}, \mathbf{\Phi_2}$ "nice" dictionaries incoherent to each other but coherent to desired structures in the data

P_K throws away data from known degraded region

Solve

$$(x_1^\star, x_2^\star) = \mathbf{argmin}_{y_1, y_2} \|\mathbf{\Phi}_1^* y_1\|_1 + \|\mathbf{\Phi}_2^* y_2\|_1$$
$$\text{s.t.} \quad P_K x = P_K(y_1 + y_2).$$

Elad, Starck, Querre, & Donoho 2005; Stürck 2015 (M.S. thesis from my group); K. & Murphy 2017;
for the general idea of sparsity regularization being powerful, cf. Bryan's talk

SHEARLET-BASED MCA

(a) Masked image

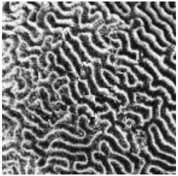
(b) MCA, SSIM = 0.7978

(c) Iterative thresholding, SSIM = 0.8813

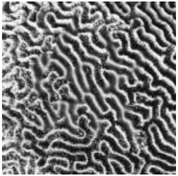
(d) Shearlet-based MCA, SSIM = 0.8831

Stürck 2015 (M.S. thesis from my group),
cf. inpainting of similar degradation in Nigel's talk and Juan-Carlos's excitement

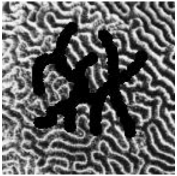
What is Successful Inpainting?, I



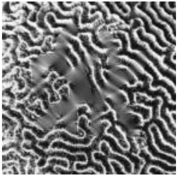
Patch-based, t=3.74s



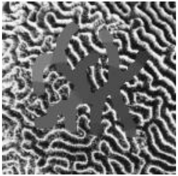
NLM, t=210.75s



Masked image

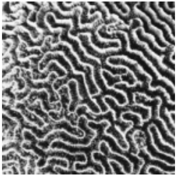


Shearlet (s12d1), t=26.29s

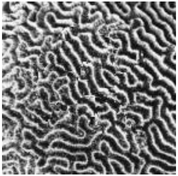


TV, t=2.04s

K 2013 (not put in a paper)

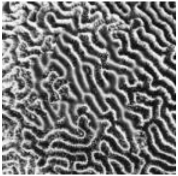


Original image



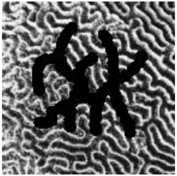
Patch-based, t=3.74s

PSNR=17.5652dB, SSIM=0.8555

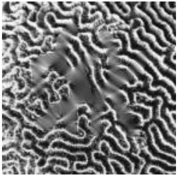


NLM, t=210.75s

PSNR=18.7336dB, SSIM=0.8720

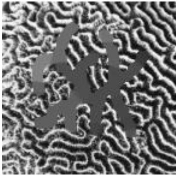


Masked image



Shearlet (s12d1), t=26.29s

PSNR=19.4429dB, SSIM=0.8542

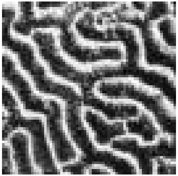


TV, t=2.04s

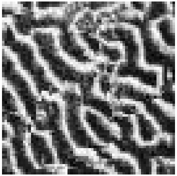
PSNR=18.6828dB, SSIM=0.8241

K 2013 (not put in a paper)

WHAT IS SUCCESSFUL INPAINTING?, I

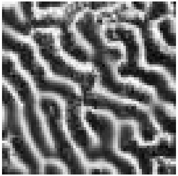


Original image



Patch-based, t=3.74s

PSNR=17.5652dB, SSIM=0.8555



NLM, t=210.75s

PSNR=18.7336dB, SSIM=0.8720

Masked image

Shearlet (s12d1), t=26.29s

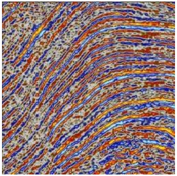
PSNR=19.4429dB, SSIM=0.8542

TV, t=2.04s

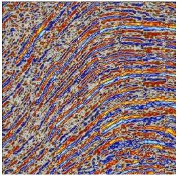
PSNR=18.6828dB, SSIM=0.8241

K 2013 (not put in a paper)

What is Successful Inpainting?, II

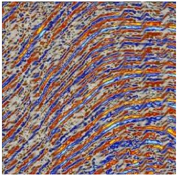


Original image



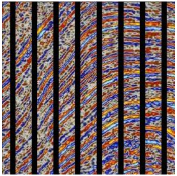
Patch-based, t=67.07s

PSNR=17.2594dB, SSIM=0.7658

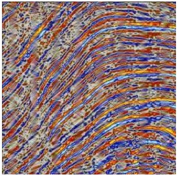


Non-local means, t=887.89s

PSNR=18.6481dB, SSIM =0.7957

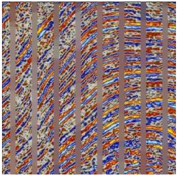


Masked image



Shearlet (s12d2), t=400.48s

PSNR=19.8313dB, SSIM=0.8209

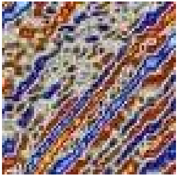


TV, t=24.65s

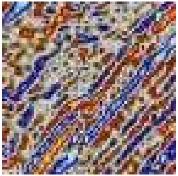
PSNR=19.1411dB, SSIM=0.7755

K 2013 (not put in a paper)

WHAT IS SUCCESSFUL INPAINTING?, II

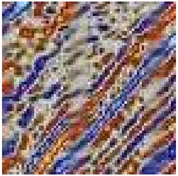


Original image



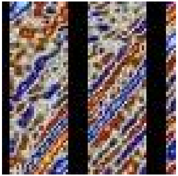
Patch-based, t=67.07s

PSNR=17.2594dB, SSIM=0.7658

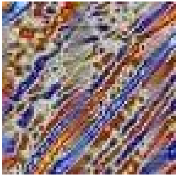


Non-local means, t=887.89s

PSNR=18.6481dB, SSIM =0.7957

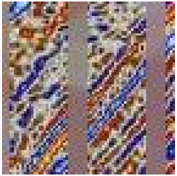


Masked image



Shearlet (s12d2), t=400.48s

PSNR=19.8313dB, SSIM=0.8209



TV, t=24.65s

PSNR=19.1411dB, SSIM=0.7755

K 2013 (not put in a paper)

Lecture Outline

Vectorized Persistent Homology
(Atmospheric Science)

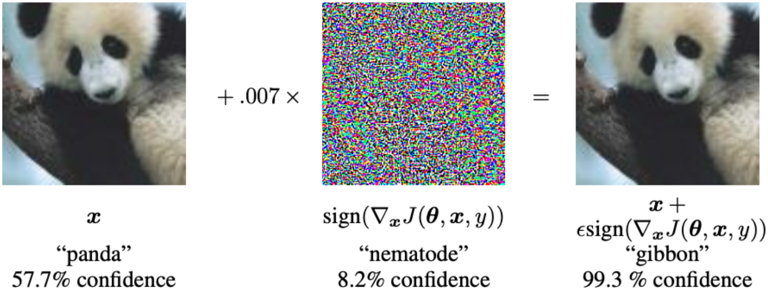
Other Stuff
(Y'all's Science?)

- More Persistence of Topology and Geometry
- Shearlet-Based Feature Extraction and Inpainting
- Geometry of Adversarial Attacks

Flag Medians
(Computer Vision)

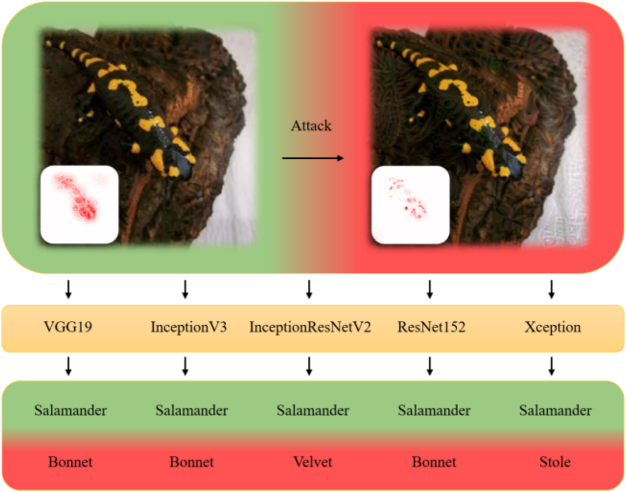
Images input to neural networks performing classification tasks may be slightly perturbed to cause the network to strongly misclassify them.

Cf. some of the discussion after Huolin's talk



A now "classic" example of an adversarial attack on GoogLeNet.

Goodfellow, Shlens, Szegedy 2015



DAmageNet is a blackbox adversarial attack algorithm that generates perturbations of images, misclassified by many standard deep neural networks.

Chen, Huang, He, Sun 2019; Chen, He, Sun, Yang, Huang 2020

I am interested in leveraging manifold and polyhedral geometry and graph theory to better understand the success of adversarial attacks.

σ(Jamil, Liu, Cole), Blanchard, K, Kirby, Peterson 2022 (on arXiv pending AFRL approval), 2022+

Lecture Outline

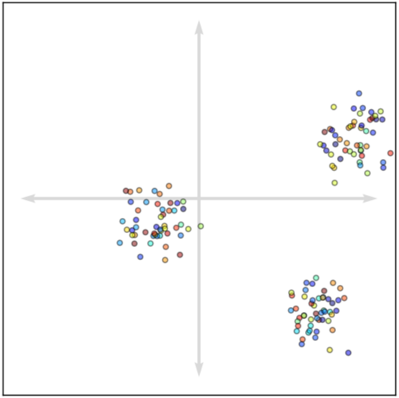
Vectorized Persistent Homology
(Atmospheric Science)

Other Stuff
(Y'all's Science?)

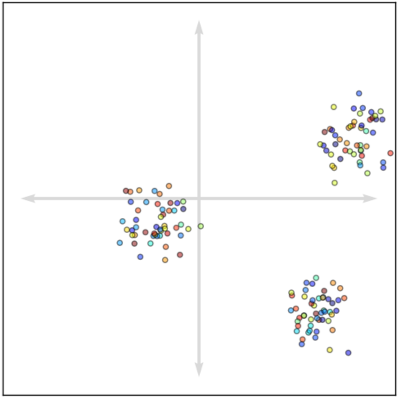
- More Persistence of Topology and Geometry
- Shearlet-Based Feature Extraction and Inpainting
- Geometry of Adversarial Attacks

Flag Medians
(Computer Vision)

Given a data set(s) $\boldsymbol{X} \subset \mathbb{R}^d$, what is (are) the point(s) in $\mathbb{R}^d$ that best represents $\boldsymbol{X}$?



Given a data set(s) $\boldsymbol{X} \subset \mathbb{R}^d$, what is (are) the point(s) in $\mathbb{R}^d$ that best represents $\boldsymbol{X}$?



How sensitive should the prototype be to outliers?

Prototypes of data classes $X = \{x_i\}_{i=1}^s \subset \mathbb{R}^d$ often solve

$$\min_{y \in A} \sum_{i=1}^{s} \|x_i - y\|_2^p .$$

Prototypes of data classes $X = \{x_i\}_{i=1}^s \subset \mathbb{R}^d$ often solve

$$\min_{y \in A} \sum_{i=1}^{s} \|x_i - y\|_2^p.$$

$A = \mathbb{R}^d, p = 2$: centroid

Prototypes of data classes $X = \{x_i\}_{i=1}^s \subset \mathbb{R}^d$ often solve

$$\min_{y \in A} \sum_{i=1}^{s} \|x_i - y\|_2^p .$$

$A = \mathbb{R}^d, p = 2$: centroid
(just the component-wise average)

Prototypes of data classes $X = \{x_i\}_{i=1}^s \subset \mathbb{R}^d$ often solve

$$\min_{y \in A} \sum_{i=1}^{s} \|x_i - y\|_2^p .$$

$A = \mathbb{R}^d, p = 2$: centroid
(just the component-wise average)

$A = \mathbb{R}^d, p = 1$: geometric median

Prototypes of data classes $X = \{x_i\}_{i=1}^s \subset \mathbb{R}^d$ often solve

$$\min_{y \in A} \sum_{i=1}^{s} \|x_i - y\|_2^p.$$

$A = \mathbb{R}^d, p = 2$: centroid
(just the component-wise average)

$A = \mathbb{R}^d, p = 1$: geometric median

$A = X, p = 1$: medoid

Prototypes of data classes $X = \{x_i\}_{i=1}^s \subset \mathbb{R}^d$ often solve

$$\min_{y \in A} \sum_{i=1}^{s} \|x_i - y\|_2^p .$$

$A = \mathbb{R}^d, p = 2$: centroid
(just the component-wise average)

$A = \mathbb{R}^d, p = 1$: geometric median

$A = X, p = 1$: medoid

$p = 2$: closed form solution as a least squares problem

$p = 1$: less sensitive to outliers

Weiszfeld Algorithm I

Key ideas:

Weiszfeld Algorithm I

Key ideas:

Finding centroids is easy.

Weiszfeld Algorithm I

Key ideas:

Finding centroids is easy.

A geometric median $\boldsymbol{y}$ of $\{\boldsymbol{x}_i\}_{i=1}^{s}$ satisfies

$$\boldsymbol{y} = \left(\sum_{i=1}^{s} \frac{\boldsymbol{x}_i}{\|\boldsymbol{x}_i - \boldsymbol{y}\|_2}\right) \left(\sum_{k=1}^{s} \frac{1}{\|\boldsymbol{x}_k - \boldsymbol{y}\|_2}\right)^{-1}.$$

Weiszfeld Algorithm II

Given $\{x_i\}_{i=1}^{s} \subset \mathbb{R}^d$,

compute until converged:

$$y = \frac{1}{s}\sum_{i=1}^{s} \omega_i x_i, \quad \omega_i = \frac{p}{\|x_i - y\|_2}\left(\sum_{k=1}^{s} \frac{1}{\|x_k - y\|_2}\right)^{-1}.$$

Weiszfeld 1937; briefly mentioned as a technique in Sandra's talk

Weiszfeld Algorithm II

Given $\{x_i\}_{i=1}^{s} \subset \mathbb{R}^d$,

compute until converged:

$$y = \frac{1}{s}\sum_{i=1}^{s} \omega_i x_i, \quad \omega_i = \frac{p}{\|x_i - y\|_2}\left(\sum_{k=1}^{s}\frac{1}{\|x_k - y\|_2}\right)^{-1}.$$

Such algorithms are called iteratively reweighted least squares (IRLS).

Weiszfeld 1937; briefly mentioned as a technique in Sandra's talk

Geometry on the Grassmannian

Let $\mathbf{Gr}(r, d)$ denote the Grassmannian of r-dimensional subspaces of $\mathbb{R}^d$.

For $W_j \in \mathbf{Gr}(r, d)$, let P_j denote the orthogonal projection onto W_j.

Between W_j and W_k, the

- chordal distance is $\frac{1}{\sqrt{2}} \left\| P_j - P_k \right\|_{Fro}$ and the
- geodesic distance is $\|\Theta\|_2$,
  where Θ is a vector of the square-root-arccos of the singular values of $P_j P_k$.

Subspace Prototypes

Given $W_1, W_2, \ldots, W_s$ where $W_i \in \mathbf{Gr}(k_i, n)$ and d chordal or geodesic distance, what is

$$\min_{Y \in \mathbf{Gr}(r,n)} \sum_{i=1}^{s} d(W_i, Y)^p?$$

d chordal distance, $p = 2$: flag mean
(r-dimensional PCA)

d geodesic distance, $p = 2$: Karcher mean
(only for $k_1 = \ldots = k_s$)

d geodesic distance, $p = 1$: ℓ_2-median
(only for $k_1 = \ldots = k_s$)

Draper, Kirby, et al. 2014; Karcher 1977; Fletcher, Venkatasubramanian, Joshi 2009; Mankovich, K, Peterson, Kirby 2022

Subspace Prototypes

Given $W_1, W_2, \ldots, W_s$ where $W_i \in \mathbf{Gr}(k_i, n)$ and d chordal or geodesic distance, what is

$$\min_{Y \in \mathbf{Gr}(r,n)} \sum_{i=1}^{s} d(W_i, Y)^p?$$

d chordal distance, $p = 2$: flag mean
(r-dimensional PCA)

d geodesic distance, $p = 2$: Karcher mean
(only for $k_1 = \ldots = k_s$)

d geodesic distance, $p = 1$: ℓ_2-median
(only for $k_1 = \ldots = k_s$)

d chordal distance, $p = 1$: flag median

Draper, Kirby, et al. 2014; Karcher 1977; Fletcher, Venkatasubramanian, Joshi 2009; Mankovich, K, Peterson, Kirby 2022

FLAGIRLS

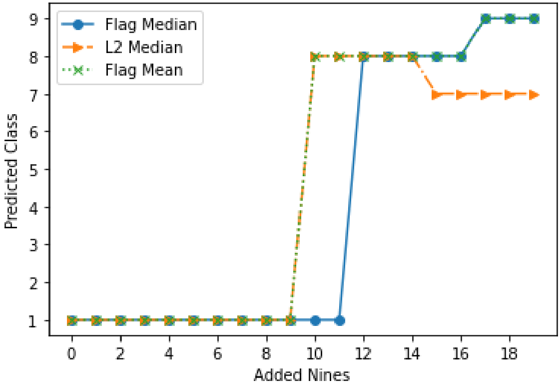
Key Idea:

As with Weiszfeld, we leverage the fact that the flag mean solution has a closed form (here: via PCA).

We solve a sequence of flag mean problems with weighted subspace orthonormal bases (ONB).

The weights were determined using Lagrangians.

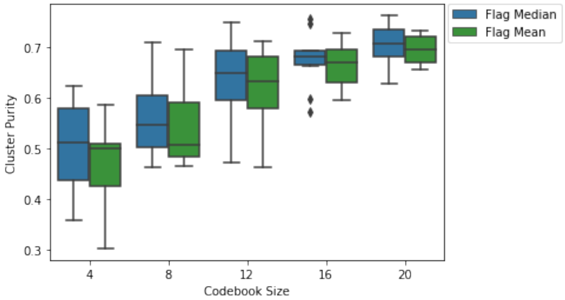
Mankovich, K, Peterson, Kirby 2022



The neural network predicted class of the prototype for the dataset with $\mathbf{20}$ examples of $\mathbf{1}$'s and $\boldsymbol{i}$ examples of $\mathbf{9}$'s from MNIST with $\boldsymbol{i} = \mathbf{0, 1, 2, \ldots, 19}$.

Mankovich, K, Peterson, Kirby 2022

An example of frames from the YouTube dataset. Liu, Luo, Shah 2009



An LBG implementation on the YouTube dataset. The results of 2 different implementations of LBG for codebook sizes **4, 8, 12, 16** and **20**. The flag median outperforms flag mean for all codebook sizes.

Mankovich, K, Peterson, Kirby 2022

CONCLUSION

Original meme downloaded from `https://memegenerator.net/instance/53842731/mark-twain-reports-of-my-death-are-greatly-exaggerated`

Thanks for your time
and trypophobia I now have!

For papers and software:

https://www.math.colostate.edu/~king

CodEx Seminar:
A pan-university, remote seminar on the theory and applications of harmonic analysis, combinatorics, and algebra

https://www.math.colostate.edu/~king/codex