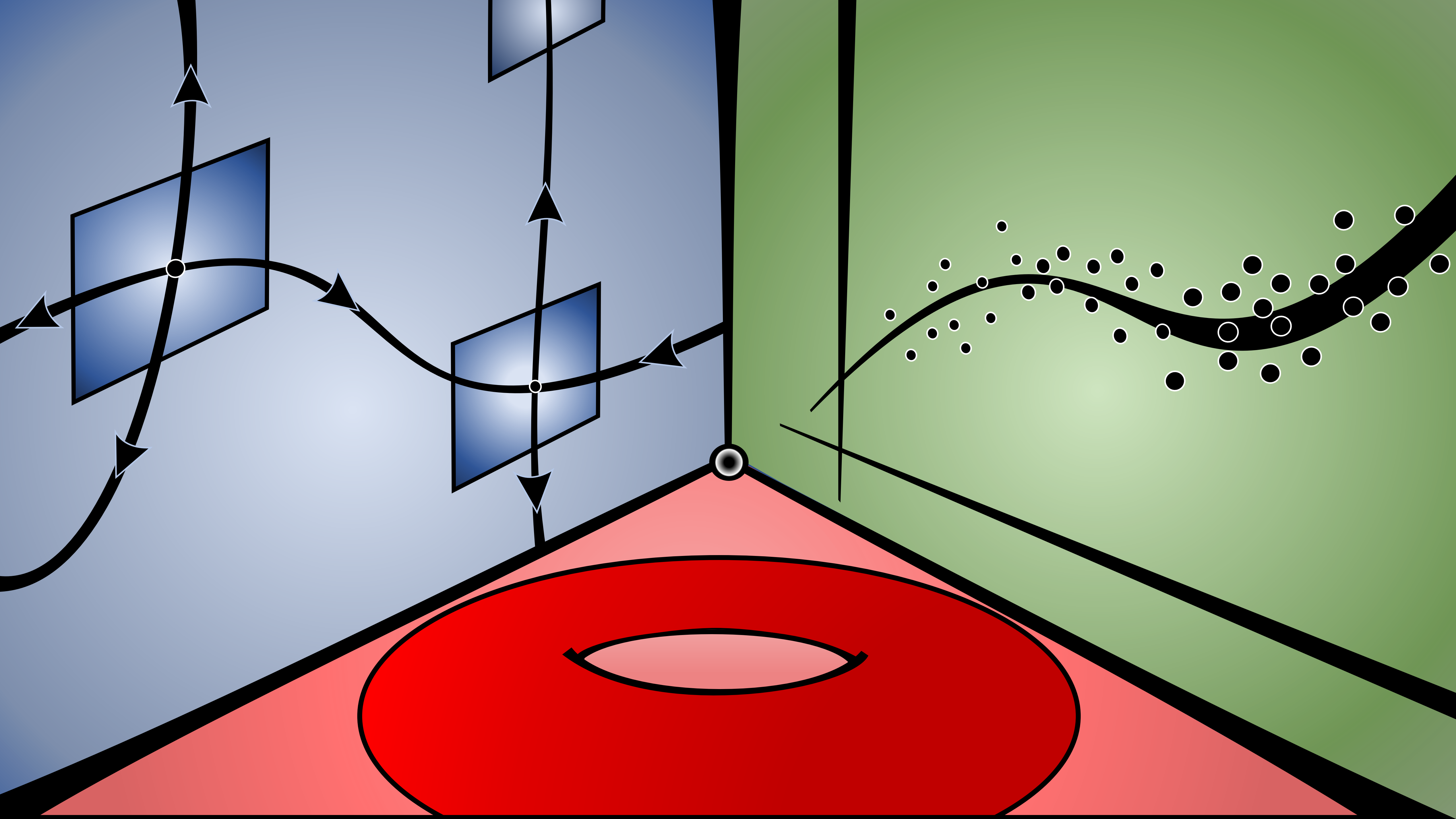
Path Signatures in Topology, Dynamics and Data Analysis

Vidit Nanda

Oxford

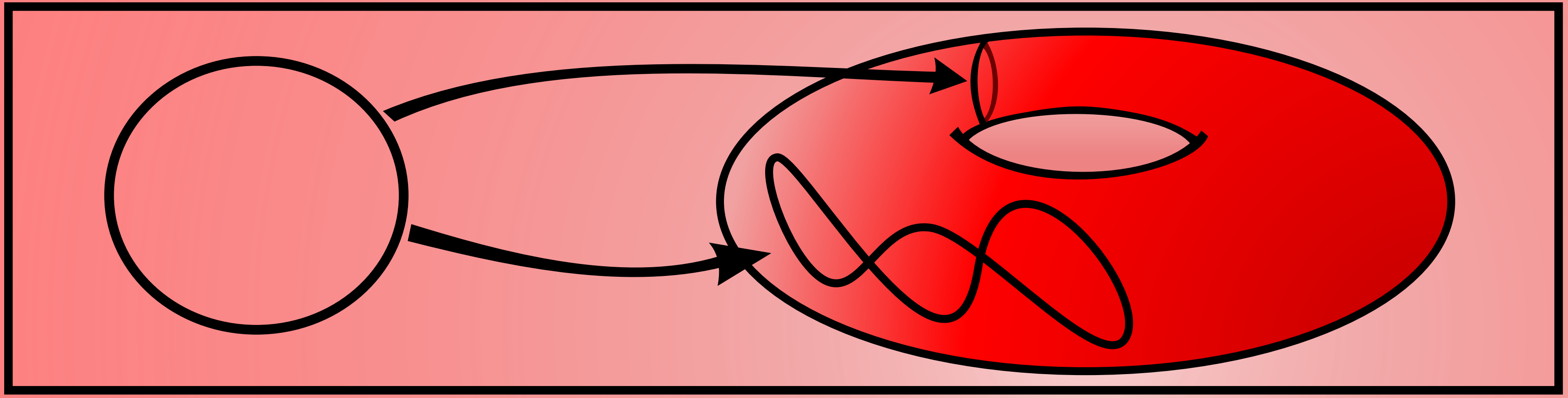
06 Aug 2020 @ DataSig

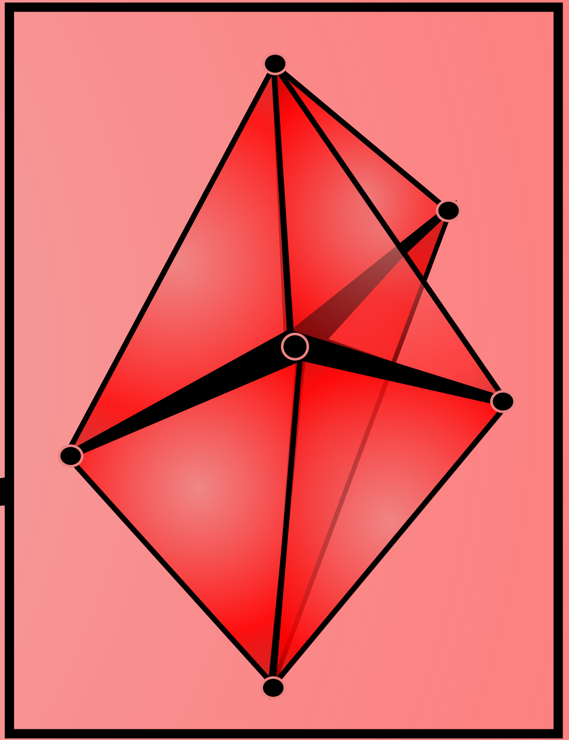
[+ Ilya Chevyrev & Harald Oberhauser]



Loop Spaces

The **loop space** ΩX of a topological space X is the set of all continuous maps $\mathbb{S}^1 \to X$ endowed with the compact-open topology





In general ΩX is not finite-dimensional (even if X is)

Ω is a homotopic adjoint to suspension: $[Y, \Omega X] \simeq [\Sigma Y, X]$

It satisfies $\pi_i(\Omega X) \simeq \pi_{i+1}(X)$

There is a nice product $\Omega X \times \Omega X \to \Omega X$ (if you're happy to choose a basepoint)

Loop spaces are ground zero for **spectra** and **stable homotopy theory**

Differential Forms

The familiar tools won't help much with computing the cohomology groups $H^\bullet(\Omega X;\mathbb{R})$ — for most X there are no known cellular models for ΩX [even $X = \mathbb{S}^n$ is hard!]

When X is a smooth manifold, one can try to exploit its **deRham complex** $C^\bullet_{dR}(X)$ to learn something about the cohomology of ΩX

$$0 \to C^0_{dR}(X) \xrightarrow{d} C^1_{dR}(X) \xrightarrow{d} C^2_{dR}(X) \xrightarrow{d} \cdots$$

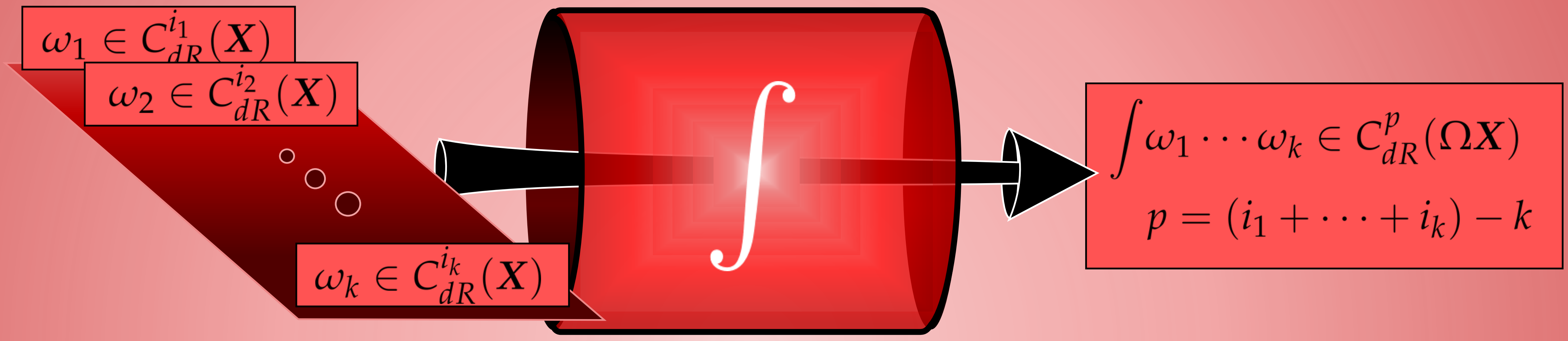
Here $C^i_{dR}(X)$ is the space of **differential** i**-forms** on X; in local coordinates $(x_1, \ldots, x_n)$ this space has a basis $\{dx_{j_1} \wedge dx_{j_2} \wedge \cdots \wedge dx_{j_i} \mid j_1 < j_2 < \cdots < j_i\}$ over the ring of smooth functions $X \to \mathbb{R}$

And d is the **exterior derivative**; for $\omega = f(x_i, \ldots, x_n)\, dx_{j_1} \wedge \cdots \wedge dx_{j_i}$,

$$d\omega = \sum_k \frac{\partial f}{\partial x_k} dx_k \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_i}$$

Idea [K.-T. Chen, 1951+]: Relate the differential graded algebra $H^\bullet_{dR}(X)$ to the differential graded algebra $H^\bullet_{dR}(\Omega X)$

Iterated Integrals



When $i_1 = \cdots = i_k = 1$, then $p = 0$ and $\int \omega_1 \cdots \omega_k$ is just a function $\Omega \boldsymbol{X} \to \mathbb{R}$; let's evaluate it at $\gamma : [0,1] \to \boldsymbol{X}$ when $\dim \boldsymbol{X} = 2$ in local coordinates (x, y)

For $\omega_1 = dx$ and $\omega_2 = dy$, this function evaluates to

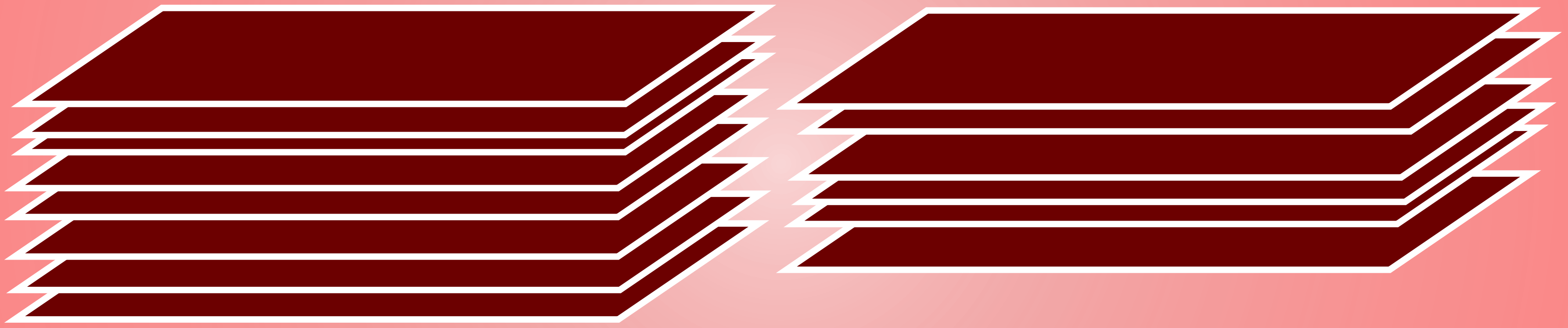
$$\left[\int \omega_1 \omega_2\right](\gamma) = \int_0^t \int_0^s \gamma_2'(t) \cdot \gamma_1'(s)\, ds\, dt$$

Thm [Chen]: The image of $\int$ is a differential graded subalgebra of $H^{\bullet}_{dR}(\Omega \boldsymbol{X})$

Shuffle Products

A permutation of $\{1, \ldots, k+\ell\}$ is called a (k, ℓ)**-shuffle** if we have both

$$\pi^{-1}(1) < \pi^{-1}(2) < \cdots < \pi^{-1}(k), \text{ and}$$
$$\pi^{-1}(k+1) < \pi^{-1}(k+2) < \cdots < \pi^{-1}(k+\ell)$$



Thm [Chen]: Given forms $\omega_1, \ldots, \omega_{k+\ell}$ in $C^\bullet_{dR}(\boldsymbol{X})$, we have

$$\left[\int \omega_1 \cdots \omega_k\right] \wedge \left[\int \omega_{k+1} \cdots \omega_{k+\ell}\right] = \sum_\pi \pm \int \omega_{\pi(1)} \cdots \omega_{\pi(k+\ell)}$$

where π ranges over all (k, ℓ)-shuffles

Picard Iteration

To solve an ODE $dx/dt = f(x,t)$, start with the constant function $x_0(t) = c$ and define

$$x_{i+1}(t) = c + \int_0^t f(x_i(t), t)dt$$

If f is Lipschitz, then the **Picard-Lindelöf** theorem applies; in this case, the limiting x_i exists and solves the ODE. Here's the simplest non-silly example [$f(x,t) = x$]

$$x_0(t) = \boxed{1}$$

$$x_1(t) = 1 + \int_0^t x_0(t)\, dt = \boxed{1 + t}$$

$$x_2(t) = 1 + \int_0^t x_1(t)\, dt = \boxed{1 + t + \frac{t^2}{2}}$$

$$x_3(t) = 1 + \int_0^t x_2(t)\, dt = \boxed{1 + t + \frac{t^2}{2} + \frac{t^3}{6}}$$

$$\vdots$$

Controlled Differential Equations

A differential equation **controlled by the path** $x : [0, 1] \to \mathbb{R}^n$ has the form

$$dy(t) = f(y(t))\, dx(t)$$

where f is a function $\mathbb{R}^n \to \mathbf{Mat}(m \times n)$. A path $y : [0, 1] \to \mathbb{R}^m$ is a solution if

$$y(t) = \int_0^t f(y(s))\, dx(s)$$

In many cases of interest, f is a nice linear map, but the controlling paths x (and hence the solutions y) are very far from smooth — so we can't take any derivatives

We say that x has **bounded variation** if the supremum over partitions $0 < t_0 < \cdots < t_k < 1$ of the quantity $\sum_i \|x(t_{i+1}) - x(t_i)\|$ is finite. Let $\mathbf{BV}(\mathbb{R}^n)$ denote the space of all bounded variation paths in n-space

Idea [T. Lyons, 1998+]: Use Picard iteration to solve CDEs where x is in $\mathbf{BV}(\mathbb{R}^n)$ and f is a linear map $\mathbb{R}^n \to \mathbf{Mat}(m \times n)$.

Solving CDEs

Replace $f : \mathbb{R}^n \to \mathbf{Mat}(m \times n)$ with a map $g : \mathbb{R}^m \to \mathbf{Mat}(n \times n)$ as follows:

$$g(u) = \Big[v \mapsto [f(v)](u)\Big] \text{ for } u \in \mathbb{R}^m \text{ and } v \in \mathbb{R}^n$$

Start Picard iteration as usual with the constant $y_0(t) = c$, and ...

$$y_1(t) = c + \int_0^t g(y_0(s))\, dx(s) = \boxed{c \cdot \left[\mathrm{Id}_n + \int_0^t dg(x(s))\right]}$$

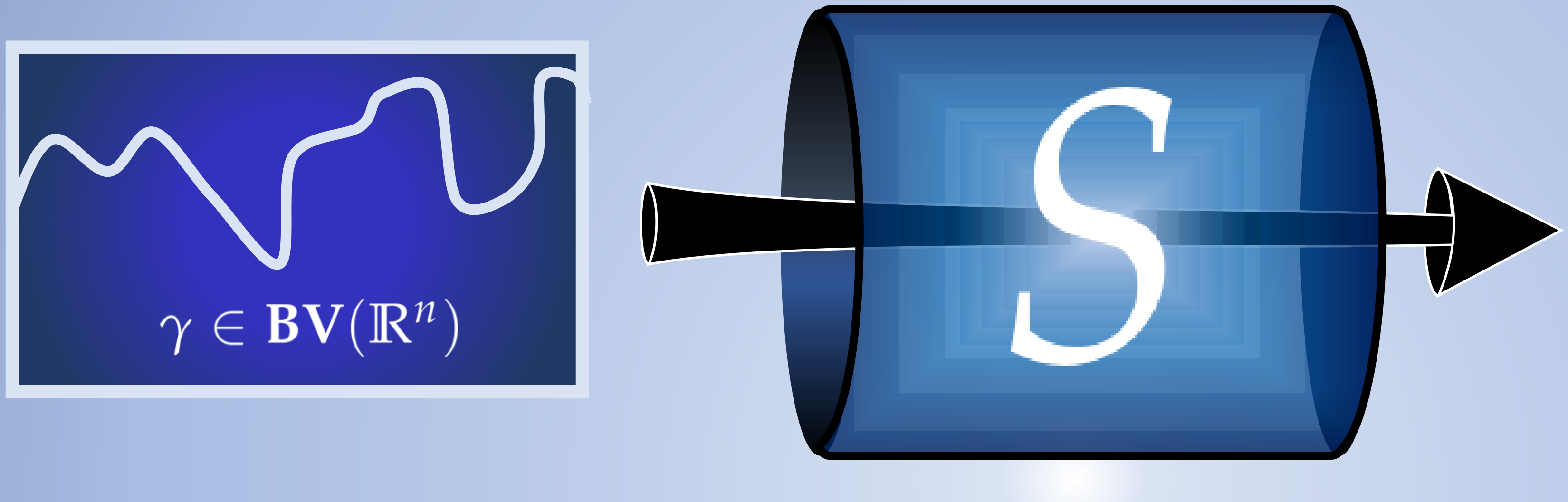
$$y_2(t) = c + \int_0^t g(y_1(s))\, dx(s) = \boxed{c \cdot \left[\mathrm{Id}_n + \int_0^t dg(x(s)) + \int_0^t \int_0^s dg(x(s)) \cdot dg(x(u))\right]}$$

$$\vdots$$

All terms have the form

$$\int_{0<t_1<\cdots<t_k<t} dg(x(t_1)) \cdots dg(x(t_k))$$

The Path Signature



The **Signature** of $\gamma \in \mathbf{BV}(\mathbb{R}^n)$ is an element $S(\gamma)$ living in the *Tensor algebra* $\mathbf{T}(\mathbb{R}^n)$

$$\mathbf{T}(\mathbb{R}^n) = \prod_{m \geq 0} (\mathbb{R}^n)^{\otimes m}$$

The m here indicates the **level** of the signature; it is customary in practice to truncate below some fixed m

The m-th level has n^m components

0	1
1	$\int_0^1 d\gamma_i(t)$
2	$\int_0^1 \int_0^t d\gamma_i(t)\, d\gamma_j(s)$
3	$\int_0^1 \int_0^t \int_0^s \cdots$

Glorious Properties

The first thing to note about $S(\gamma)$ is that its components are evaluations of Chen's iterated integral functions $[\int \omega_1 \cdots \omega_k](\gamma)$ where each ω_i is $dx_\bullet$

If $S(\gamma) = S(\gamma')$ then γ and γ' differ by a *tree-like reparametrization*

The signature satisfies $S(\gamma \star \delta) = S(\gamma) \otimes S(\delta)$ where $\star$ denotes the usual composition of paths $[0,1] \to \mathbb{R}^n$ by concatenation

Each **multi-index** $I = (i_1, \cdots, i_k)$ with $i_\bullet \in \{1, \ldots, n\}$ corresponds to a component $S_I(\gamma)$ in level k of the signature

By the **shuffle product** identity, we have

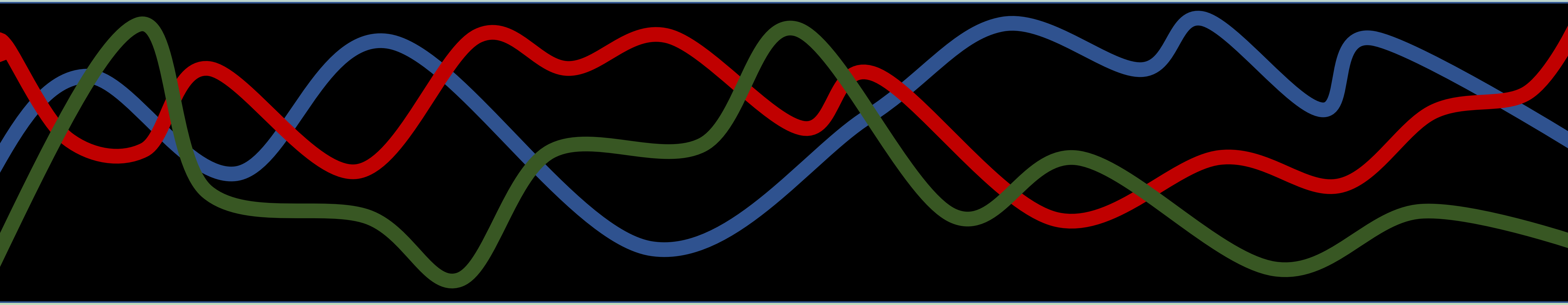
$$S_I(\gamma) \cdot S_J(\gamma) = \sum_K S_K(\gamma)$$

where K ranges over (I, J)-shuffles, so the higher-level components of S are algebraically dependent on the lower ones

Statistical Inference

The signature maps the very complicated space $\mathbf{BV}(V)$ of any vector space V into the very complicated tensor algebra $\mathbf{T}(V)$ — there is no reason to suspect a priori that this machinery can be made useful in broader contexts

But the truncated signature is a fantastic feature map for all sorts of nonlinear data that can be embedded into $\mathbf{BV}(V)$ for some artfully chosen V



Many success stories for classification tasks that involve time-varying data — handwriting and gait recognition, financial stream data analysis etc.

More Glorious Properties

Let B be the quotient of $\mathbf{BV}(V)$ by tree-like equivalence, and let $K \subset B$ be a compact subset; the signature map $S : K \to \mathbf{T}(V)$ is:

Universal: For each continuous $f : K \to \mathbb{R}$ and $\epsilon > 0$, there exists a **linear** functional $L : \mathbf{T}(V) \to \mathbb{R}$ satisfying

$$\sup_{\gamma \in K} |f(\gamma) - \langle L, S(\gamma) \rangle| < \epsilon$$

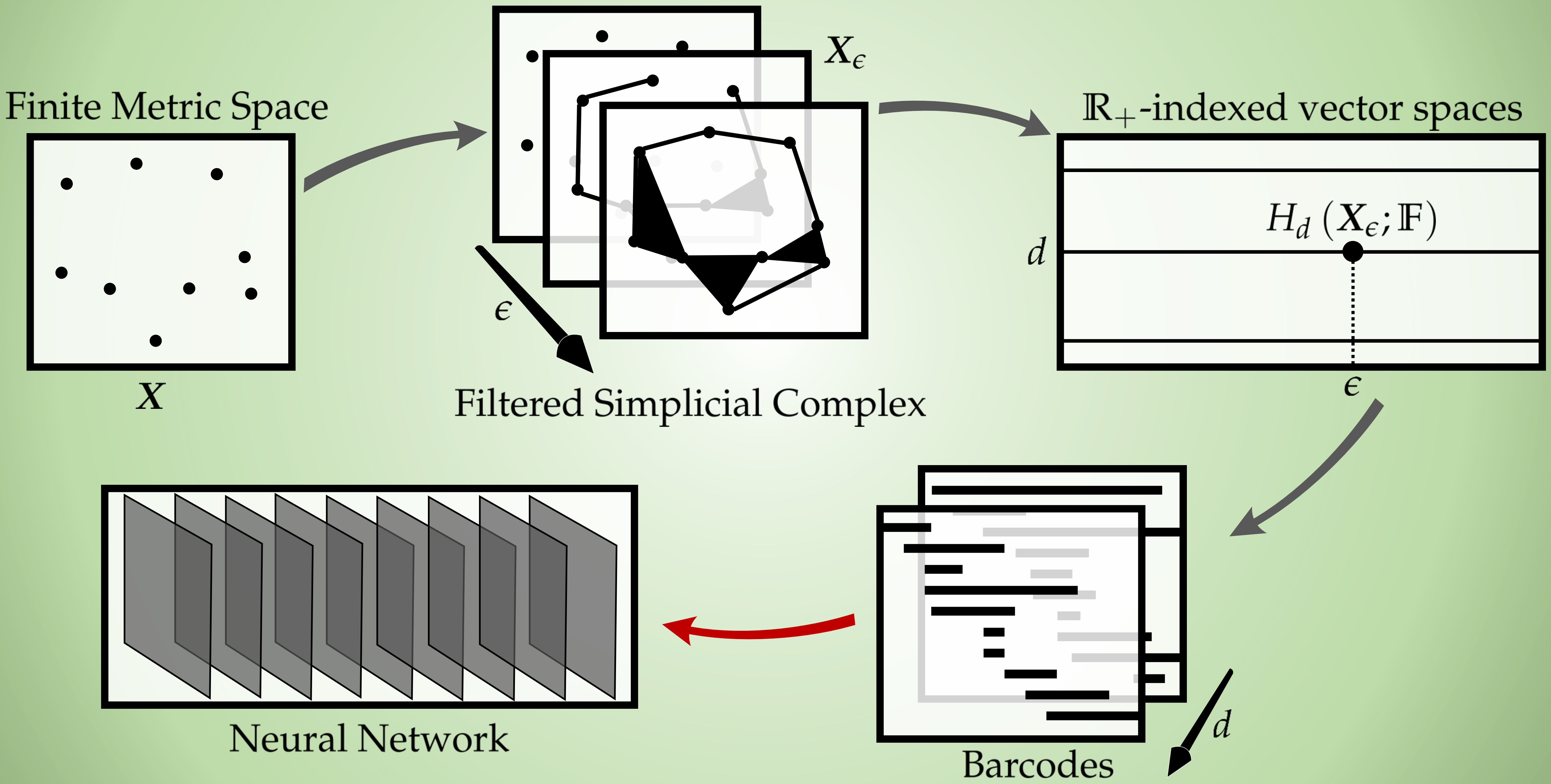
Characteristic: Borel probability measures μ on K are characterized by their signature means; namely, the map $\mathbf{Bor}(K) \to \mathbf{T}(V)$ given by

$$\mu \mapsto \mathbb{E}_{\gamma \sim \mu}[S(\gamma)]$$

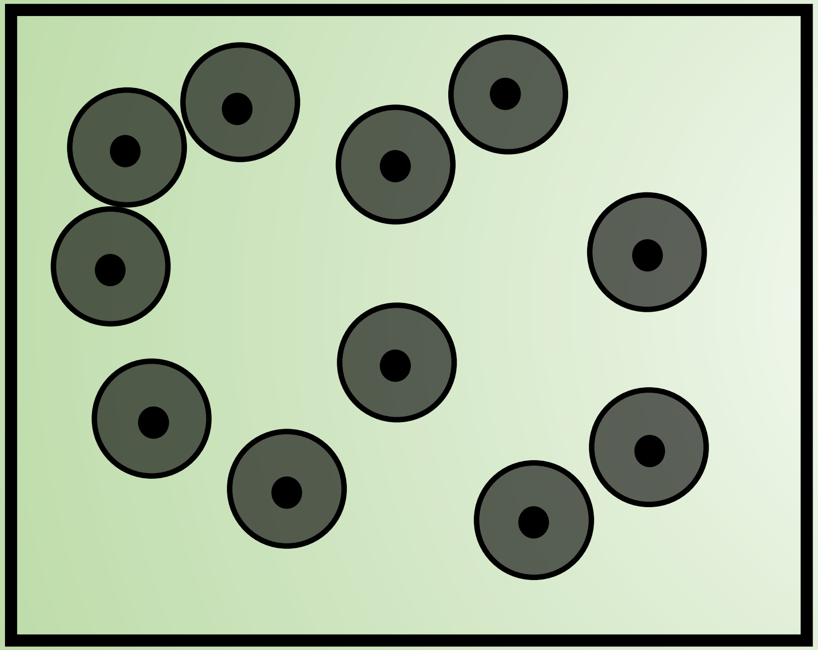
is injective!

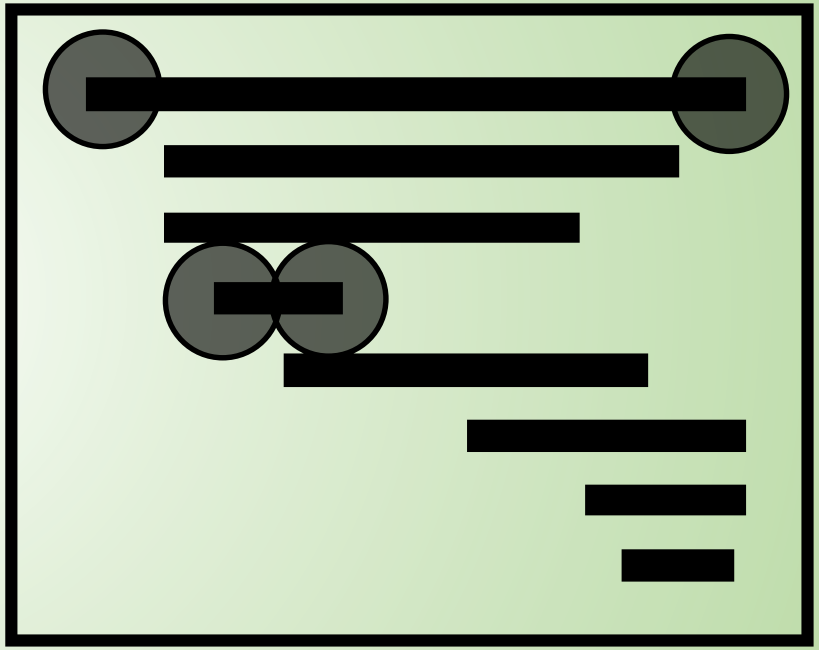
Kernelizable: If V is a Hilbert space then the map $\kappa : K \times K \to \mathbb{R}$ given by setting $\kappa(\gamma, \gamma') = \langle S(\gamma), S(\gamma') \rangle$ forms a bounded, continuous, universal and characteristic kernel

Topological Data Analysis



Stability Theorem





Feature Maps for Barcodes

The space **Bar** of persistent homology barcodes is hideously non-linear, so even defining averages in a consistent way is impossible. So there have been some efforts to find good feature maps for machine learning from barcodes (also called *persistence diagrams*)

Statistical Topological Data Analysis using Persistence Landscapes

Pete
Depa
Cleve
Cleve

Sliced Wasserstein Kernel for Persistence Diagrams

Persistence Images: A Stable Vector Representation of Persistent Homology

Henry Ad
Francis Mo

Kernel Method for Persistence Diagrams via Kernel Embedding and Weight Factor

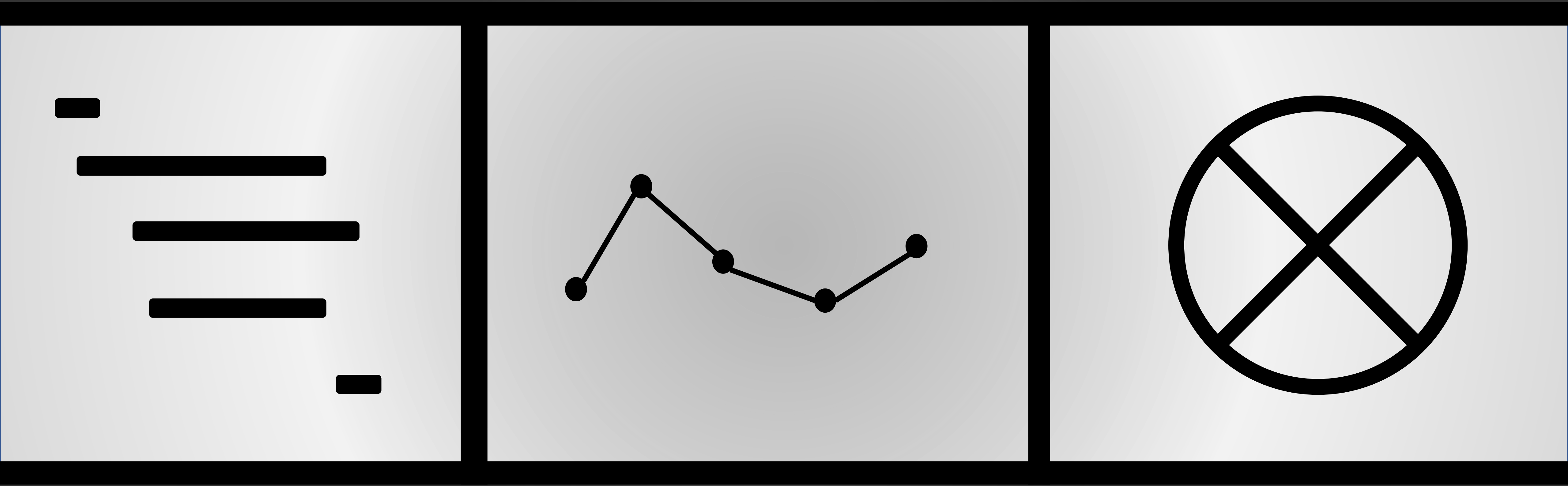
Genki Kusan
Graduate Schoo

Approximating Continuous Functions on Persistence Diagrams Using Template Functions

Jose A. Perea JOPEREA@MSU.EDU
Elizabeth Munch MUNCHELI@EGR.MSU.EDU
Department of Computational Mathematics, Science, and Engineering; and Department of Mathematics

Firas A. Khasawneh KHASAWN3@EGR.MSU.EDU
Department of Mechanical Engineering

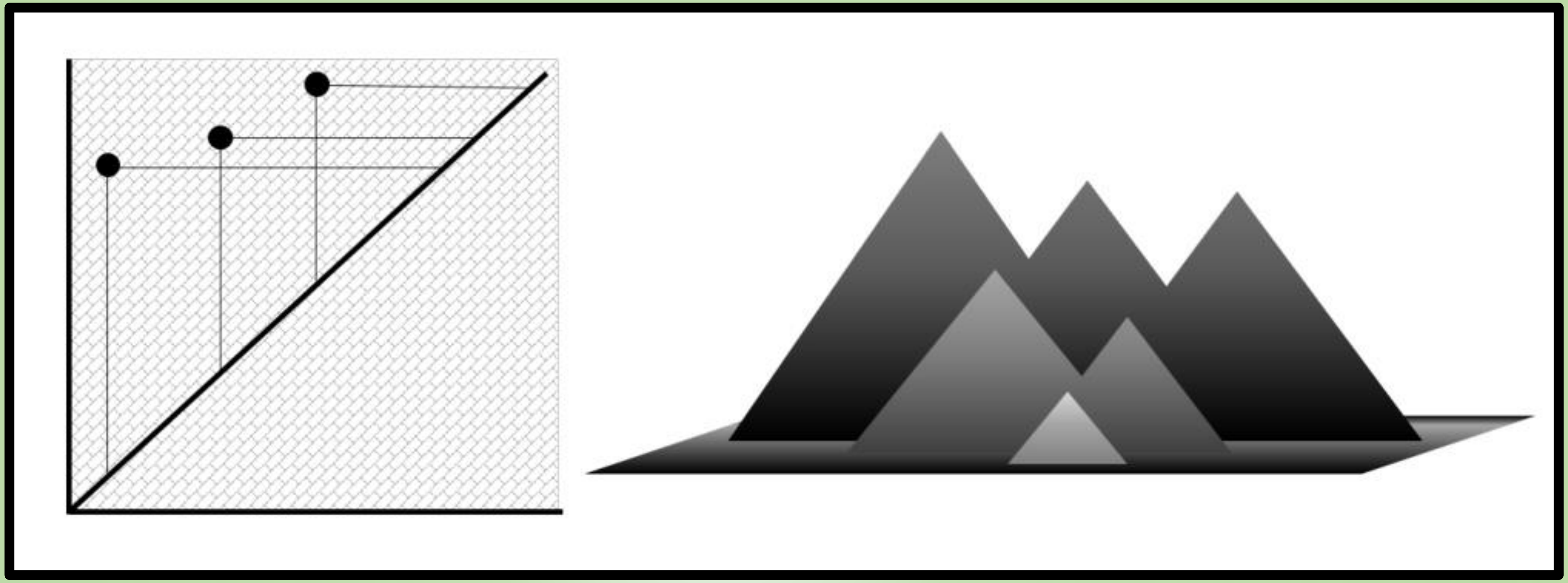
$$\mathbf{Bar} \xrightarrow{\iota_\bullet} \mathbf{BV}(V) \xrightarrow{S} \mathbf{T}(V)$$



The maps $\iota_\bullet$ are *persistence path embeddings*

And S is the usual signature map obtained by iterated integration

The Landscape Embedding



Barcodes can be transformed into **landscape** functions (Bubenik, 2012) and the assignment is stable

Paths obtained as antiderivatives of landscapes are stable

Not difficult to extract landscape from barcode, but signatures of integrated landscapes are hard

The Betti Embedding

β

For each dim d and scale t, let $\beta_d(t)$ be the number of bars in the d-th barcode that contain t

The assignment $t \mapsto \beta_d(t)$ gives curves in $\mathbb{R}^m$, where m is the largest dimension of simplices encountered

Not stable, but quite computable and contains a lot of topologically invariant information

The Euler Embedding

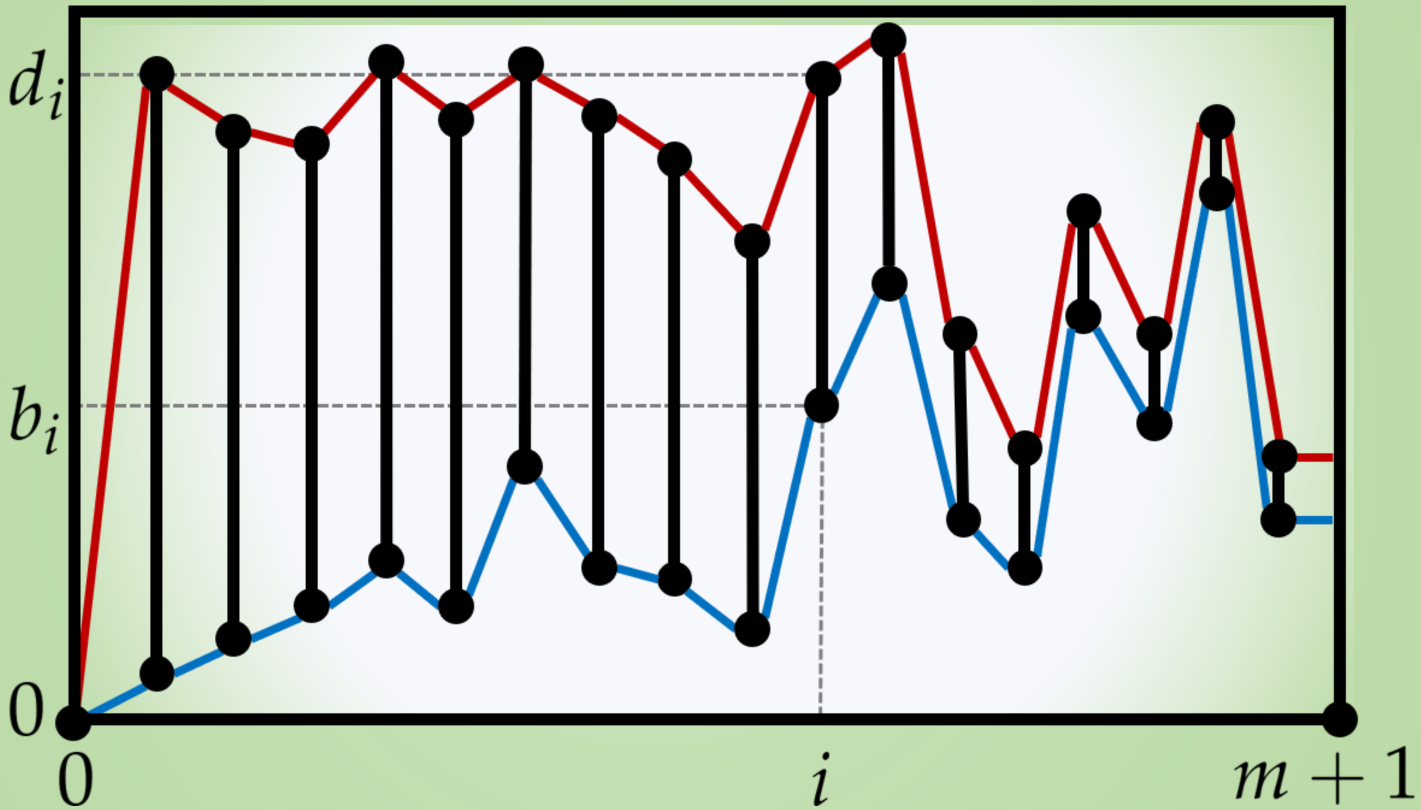
χ

The easiest persistence path to compute: $\chi(t) = \sum_d (-1)^d \beta_d(t)$

Or better yet, just compute alternating counts of simplices!

Also hopelessly unstable

E The Envelope Embedding



Arrange bars in descending order of length, then connect-the-dots

Longer, stable bars appear before shorter, unstable ones

Method	Textures	Orbits	Shapes
k_{SW}	96.8 ± 1.0	94.6 ± 1.3	95.8 ± 1.6
Φ_{PI}	93.7 ± 1.0	$\mathbf{99.86 \pm 0.21}$	90.3 ± 2.3
k_E	90.4 ± 1.5	96.6 ± 0.9	92.7 ± 1.5
k_χ	94.9 ± 0.6	NA	92.4 ± 3.0
k_β	$\mathbf{97.8 \pm 0.2}$	NA	93.0 ± 3.0
Φ_E	88.1 ± 0.8	98.1 ± 1.0	95.0 ± 0.9
Φ_χ	92.9 ± 0.7	98.8 ± 0.6	98.0 ± 1.1
Φ_β	96.6 ± 0.6	97.7 ± 0.8	$\mathbf{98.1 \pm 0.7}$

Persistence Paths and Signature Features in Topological Data Analysis

Ilya Chevyrev, Vidit Nanda, and Harald Oberhauser

Abstract—We introduce a new feature map for barcodes as they arise in persistent homology computation. The main idea is to first realize each barcode as a path in a convenient vector space, and to then compute its path signature which takes values in the tensor algebra of that vector space. The composition of these two operations — barcode to path, path to tensor series — results in a feature map that has several desirable properties for statistical learning, such as universality and characteristicness, and achieves state-of-the-art results on common classification benchmarks.

Index Terms—Topological data analysis, barcodes, signature features, kernel learning

IEEE TPAMI , DOI: 10.1109/TPAMI.2018.2885516

T Lyons, M Caruana, T Levy
Differential equations driven by rough paths

P Friz and M Hairer
A course on rough paths

I Chevyrev and A Kormilitzin
The signature method in machine learning

K-T Chen
Iterated path integrals

C Giusti and D Lee
Iterated integrals and population times series analysis