Geometry of Algebraic Data

Sandra Di Rocco

dirocco@kth.se

Supported by

Vetenskapsrådet

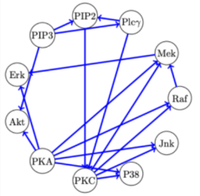
Knut and Alice Wallenberg Foundation

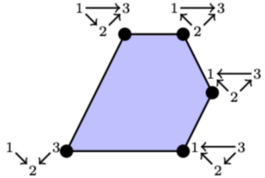
July 2026 — ICERM

Algebraic models of data

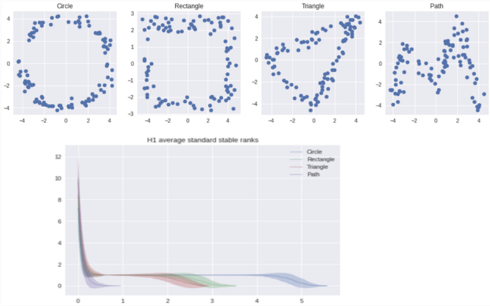
From data models to geometry

Statistical models

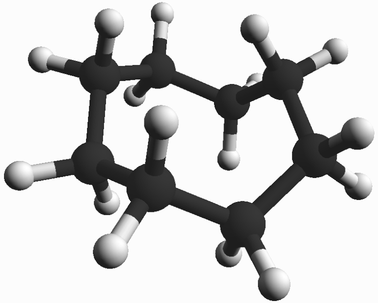




Topological analysis

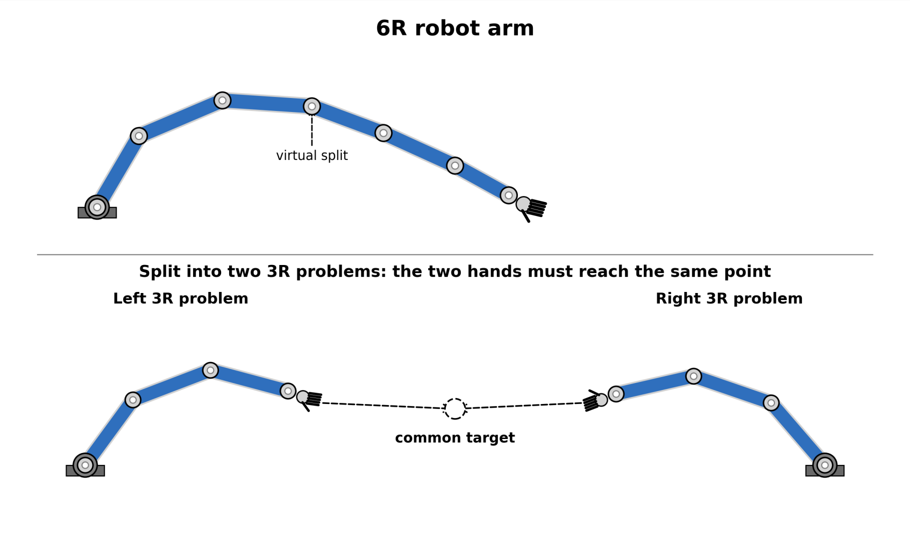


Algebraic geometry



Algebraic data: data modeled by solutions of polynomial systems.

Ex: inverse kinematics



6R IKP: general fiber of

$$\mathbb{P}^1 \times \cdots \times \mathbb{P}^1 \to Q \subset \mathbb{P}^7$$

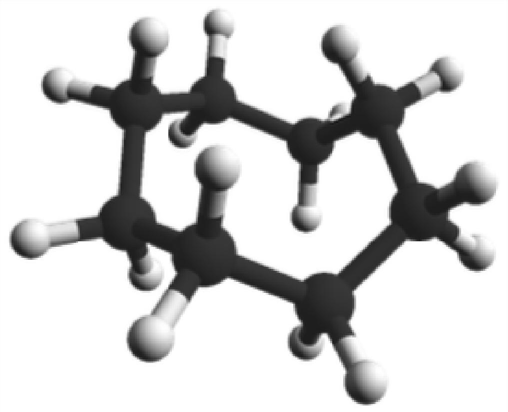
$$Q: \; x_0x_7 + x_1x_6 + x_2x_5 + x_3x_4 = 0$$

Solutions $= S_1 \cap S_2 \subset Q$:
intersection theory in the Chow ring.

J. M. Selig, Geometric Fundamentals of Robotics, 2nd ed., Springer, (2005),

Di Rocco–Eklund–Sommese–Wampler, Algebraic $\mathbb{C}^*$-actions and the inverse kinematics of a general 6R manipulator, Appl. Math. and Comp. 216 (2010)

Ex: configurations of cyclooctane



Eight carbon atoms in a ring, two hydrogen atoms on each. Closed 8R kinematic chain:

13 equations in 15 variables

Computational evidence indicates that the *shape* has two (irreducible) components: Circular and Klein.

W. M. Brown, S. Martin, S. N. Pollock, E. A. Coutsias, J.-P. Watson, Algorithmic dimensionality reduction for molecular structure analysis, J. Chem. Phys. 129, 064118 (2008)

S. Martin, A. Thompson, E. A. Coutsias, J.-P. Watson, Topology of cyclo-octane energy landscape, J. Chem. Phys. 132

Distance geometry of algebraic varieties

From real geometry to algebraic counts

$$f_1, \dots, f_s \in \mathbb{R}[x_1, \dots, x_n]$$

Real

$$X_{\mathbb{R}} \subset \mathbb{R}^n$$

data, distances, sampling

From real geometry to algebraic counts

$$f_1, \dots, f_s \in \mathbb{R}[x_1, \dots, x_n]$$

Real	**Complex**
$X_{\mathbb{R}} \subset \mathbb{R}^n$	$X_{\mathbb{C}} \subset \mathbb{C}^n$
data, distances, sampling	all algebraic solutions

From real geometry to algebraic counts

$$f_1, \dots, f_s \in \mathbb{R}[x_1, \dots, x_n]$$

Real

$X_{\mathbb{R}} \subset \mathbb{R}^n$

data, distances, sampling

Complex

$X_{\mathbb{C}} \subset \mathbb{C}^n$

all algebraic solutions

Projective

$X = \overline{X_{\mathbb{C}}} \subset \mathbb{P}^n_{\mathbb{C}}$

degrees, intersections

From real geometry to algebraic counts

$$f_1, \dots, f_s \in \mathbb{R}[x_1, \dots, x_n]$$

Real

$X_{\mathbb{R}} \subset \mathbb{R}^n$

data, distances, sampling

Complex

$X_{\mathbb{C}} \subset \mathbb{C}^n$

all algebraic solutions

Projective

$X = \overline{X_{\mathbb{C}}} \subset \mathbb{P}^n_{\mathbb{C}}$

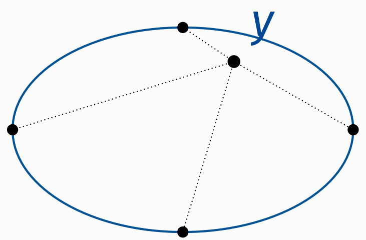
degrees, intersections

distances $\leftsquigarrow$ the quadric $q(u) = \sum_0^n a_i u_i^2$

Distance to a point: EDD

For $y \in \mathbb{R}^n$, critical points of $x \mapsto \|x - y\|^2$ on X:

$$x - y \in N_x X.$$



the ellipse: EDD = 4

Euclidean distance degree

EDD(X) = # complex critical points

Draisma–Horobeţ–Ottaviani–Sturmfels–Thomas, FoCM 2016

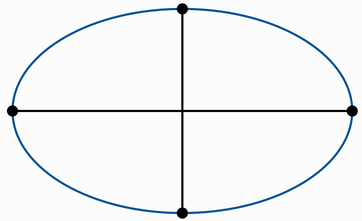
Self-distance: bottlenecks

A **bottleneck**: a pair $x, y \in X$ whose chord is normal at both ends,

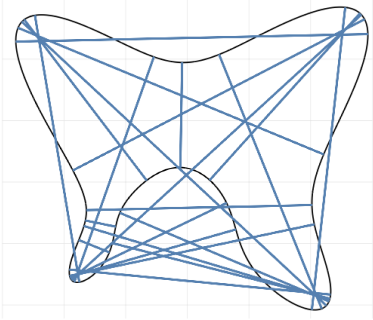
$$x - y \perp T_x X, \qquad x - y \perp T_y X.$$

Critical points of the *squared distance between pairs* — how X approaches itself.

$$b(X) = \min_{\text{bottlenecks}} \frac{1}{2}\|x - y\|$$

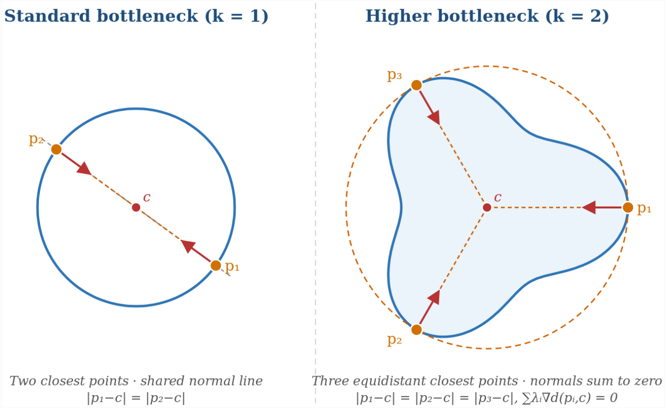


the two bottlenecks of an ellipse



Quartic curve $x^4 + y^4 + 1 - 4y - x^2y^2 - 4x^2 - x - 2y^2 = 0$ in $\mathbb{R}^2$ and its 22 bottleneck lines.

Weak feature size



a geometric k-bottleneck: z in the convex hull of its k nearest points

$$\mathrm{wfs}(X) = \inf\{d_X(z) : z \text{ a } k\text{-bottleneck}\}$$

Grove–Shiohama 1977; Chazal–Lieutier 2004

Theorem DR–Edwards–Eklund–Gäfvert–Hauenstein, SIAGA 2023

Generic complete intersections of degree ≥ 3 have *finitely many* $k+1$-bottlenecks.

Theorem Lerario–Horobeţ 2026

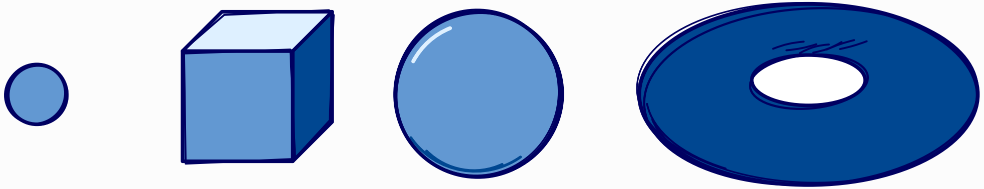
For general complete intersections,

$$\mathrm{wfs}(X) = b(X) :$$

the weak feature size is computed by ordinary bottlenecks.

Sampling and topology

The shape of data

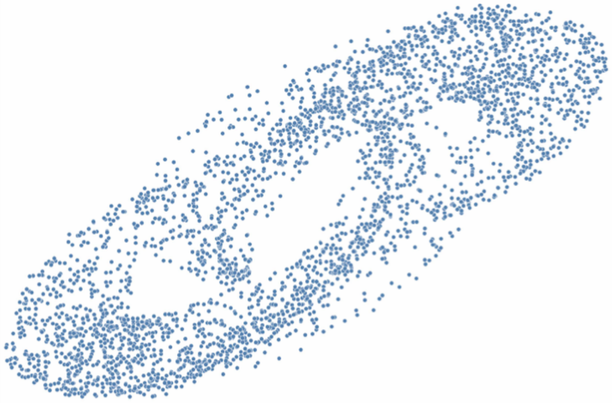


Topological information

The shape is encoded as a string of **Betti numbers** $(\beta_0, \beta_1, \ldots), \beta_i = \dim H^i(X, \mathbb{Z}_2)$

Space	β_0	β_1	β_2
Point	1	0	0
Cube	1	0	1
Sphere	1	0	1
Torus	1	2	1

Density



Definition

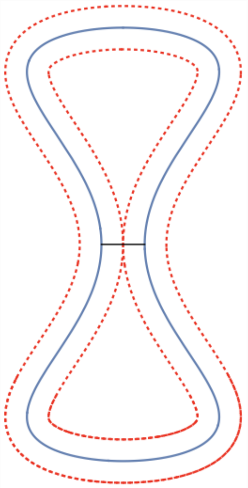
A finite set $P \subset \mathbb{R}^n$ is an ε**-sample** of a compact set $X \subset \mathbb{R}^n$ if every point of X lies within distance ϵ of P, i.e.

$$X \subseteq \bigcup_{p \in P} B(p, \varepsilon).$$

Goal

Find the optimal ε that recovers the topology of X.

How well can a variety be approximated?



medial axis: centers with ≥ 2 nearest points

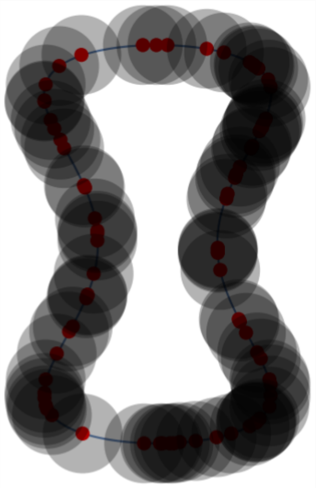
The **reach** is the distance to the medial axis (Federer) 1959:

$$\tau_X = \min\{\text{curvature radius}, \ b(X)\}.$$

Two obstructions to sparse approximation:

- local: high curvature,
- global: narrow bottlenecks.

Topology from samples



Theorem

If $\varepsilon < \tau_X$, the homology of X is recovered from an ε-sample.

Niyogi–Smale–Weinberger 2005; Chazal–Lieutier 2005; Cohen–Steiner et al. 2007; Cucker–Krick–Shub 2018; Bürgisser–Cucker–Lairez 2019

Challenge

Computations for τ_X collapses under high curvature, while $\mathrm{wfs}(X)$ does not.

A sampling algorithm

$X \subset \mathbb{R}^n$ smooth compact, $d = \dim X$. For $t \subset \{1, \dots, n\}$: π_t = projection onto the coordinates in t, $G_t(\delta)$ = randomly shifted δ-grid.

1. Slice

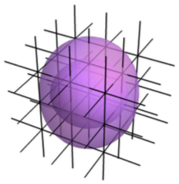
$$E_\delta = \bigcup_{|t|=d} \bigcup_{g \in G_t(\delta)} X \cap \pi_t^{-1}(g)$$

fibers = affine spaces of codimension d: finitely many points each, by homotopy continuation

2. Complete (if $d > 1$)

$$E'_\delta = \bigcup_{k<d} \bigcup_{|t|=k,\ g \in G_t(\delta)} I\left(X \cap \pi_t^{-1}(g),\ q\right)$$

critical points of the distance from generic q on each slice (EDD): they hit *every connected component*



Output: $S_\delta = E_\delta \cup E'_\delta$

HomotopyContinuation.jl

Guarantees

Theorem Di Rocco–Eklund–Gäfvert , Math. of Comp. 2022

If $0 < \delta\sqrt{n} < \min\{\varepsilon, 2\, b(X)\}$, the output is an ε-sample.

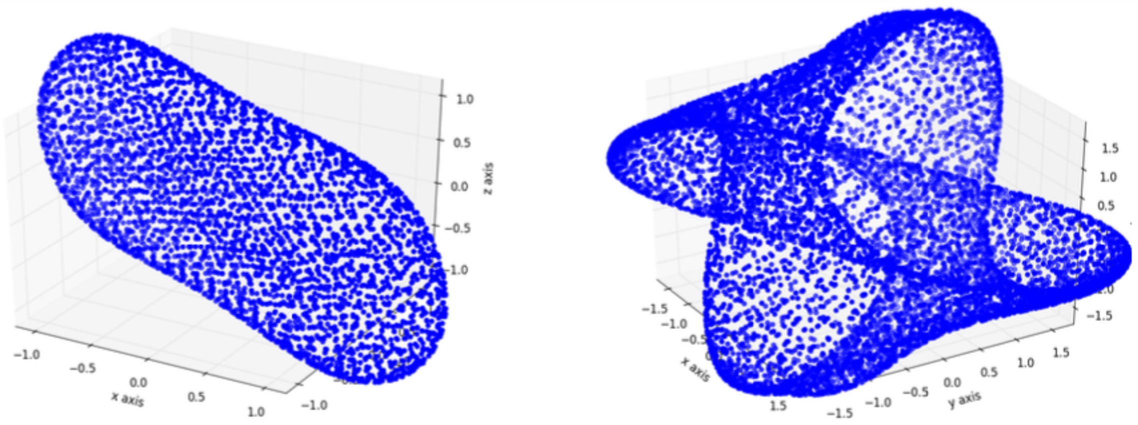
Theorem Di Rocco–Eklund–Gäfvert Math. of Comp. 2022

If moreover $\varepsilon < \mathrm{wfs}(X)$ then

$$H_i(S_\varepsilon) \cong H_i(X), \qquad i = 0, 1.$$

where S_ε is a modified Vietoris-Rips complex.

Cyclooctane, revisited



Spherical component
$\beta = (1, 0, 1)$

Klein component
$\beta = (1, 2, 0)$

Algorithm with $\varepsilon < b(C)$ gives the sample

CHomP, Kaczynski-Mischaikow-Mrozek, Computational Homology, 2004.

Counting contacts: projective geometry

Polar classes

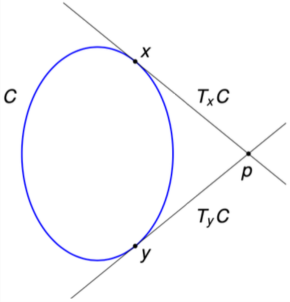
Tangency conditions against linear subspaces define classes

$$c_i(J_1(\mathbb{O}_X(1)) = [p_i(X)] \in A^i(X), \qquad i = 0, \dots, \dim X,$$

intrinsic projective invariants of $X \subset \mathbb{P}^n$.

Segre, Noether, Severi, Todd 1870s – 1930s;
Piene, Kleiman, Fulton 1970s –1980s

locus where $T_x X$ fails to meet a general $L \cong \mathbb{P}^{[n-m]+(i-2)}$ transversely



Di Rocco–Eklund–Peterson, *Numerical polar calculus and cohomology of line bundles*, (2018)

Theorem (DHOST 2016)

For a general metric, $\mathrm{EDD}(X) = \sum_i \deg p_i(X)$.

Bottlenecks are also polar

Theorem Di Rocco–Eklund–Weinstein, SIAGA 2020

For smooth m-dimensional $X \subset \mathbb{P}^n$ in general position,

$$\boxed{\mathrm{BND}(X) = \sum_{i=0}^{m-1} \epsilon_i^2 + d^2 - \deg B_m(h, p_1, \dots, p_m)}$$

$\epsilon_i = \sum_{j=0}^{m-i} \deg(p_j)$, $\qquad d = \deg X$.

curves in $\mathbb{P}^2$: $\mathrm{BND} = d^4 - 4d^2 + 3d$

curves in $\mathbb{P}^3$: $\mathrm{BND} = 10d^2 + 12dg + 4g^2 - 24d - 18g + 14$

surfaces in $\mathbb{P}^5$: $\mathrm{BND} = \epsilon_0^2 + \epsilon_1^2 + d^2 - \deg(3h^2 + 6hp_1 + 12p_1^2 + p_2)$

An algorithm (Macaulay2) computes the polynomial for every dimension.

Two open questions, work in progress, upcoming ICERM program

Generic vs. special

A *general* quartic surface in $\mathbb{C}^3$: BND = 2220.

One *specific* quartic: BND = 1390.

Take-away

Counts are stable and polar in the generic regime.
Special varieties and special *metrics* lie on degeneracy loci where counts drop.

[ED discriminant]: Draisma–Horobeţ–Ottaviani–Sturmfels–Thomas 2016; Horobeţ 2017; Joiţa–Siersma–Tibăr 2024; Rydell–Horobeţ 2025. [Voronoi cells]: Alexandr–Heaton 2020; Cifuentes–Ranestad–Sturmfels–Weinstein 2022; Brandt–Weinstein 2024.

A general frame

Interpolating object:

Work in progress with Lorenzo Barbato, Antonio Lerario (SISSA), Giacomo Maletto (KTH)

For a pair $X, Y \subseteq \mathbb{R}^n$: describe and count the critical points of $\operatorname{dist}(Y, \cdot)|_X$.

$$\begin{array}{c} \mathrm{EDD}(X) \\ k = 0, Y = \{y\} \end{array} \longleftarrow \begin{array}{c} D_k(X, Y) \\ \text{general pair} \end{array} \longrightarrow \begin{array}{c} \mathrm{BND}(Y) \\ k = 1, X = \mathbb{R}^n \end{array}$$

Challenges

1. finiteness.

   General Hypersurfaces: "Morse theory of Euclidean distance functions and applications to real algebraic geometry," Guidolin–Lerario–Ren–Scolamiero, 2025.

2. count: multiple polar geometry, multiple jets.

Metric Algebraic Geometry: Going Global

ICERM Semester Program

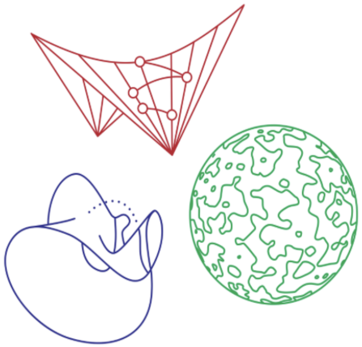
January 20 – April 23, 2027

Workshops:

- Algebraic Distance Optimization Feb 15–19
- Integral Geometry of Algebraic Varieties Mar 15–19
- Sampling and Data in Algebraic Geometry Apr 5–9

Organizers: P. Breiding, S. Di Rocco, J. Kileel, K. Kohn, A. Lerario, C. Meroni, J. Rodriguez, A. Seigal

Metric Algebraic Geometry: Going Global



Jan 20 - Apr 23, 2027
Semester Program

Thank you!