DATA DEPTH:
IN BETWEEN STATISTICS AND GEOMETRY

Stanislav Nagy

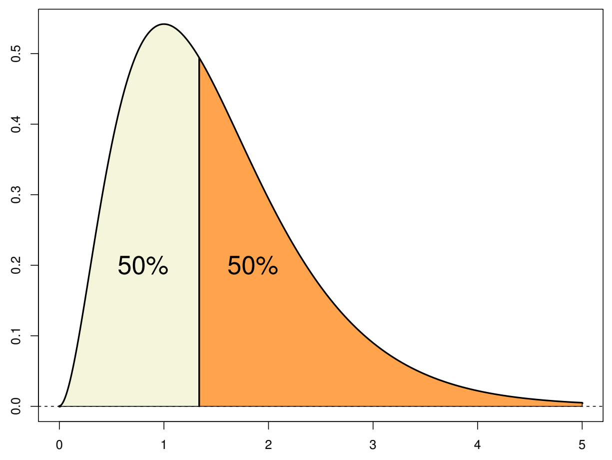
Mathematical Forum 2024

Charles University, Prague
Department of Probability and Mathematical Statistics

Motivation: The median

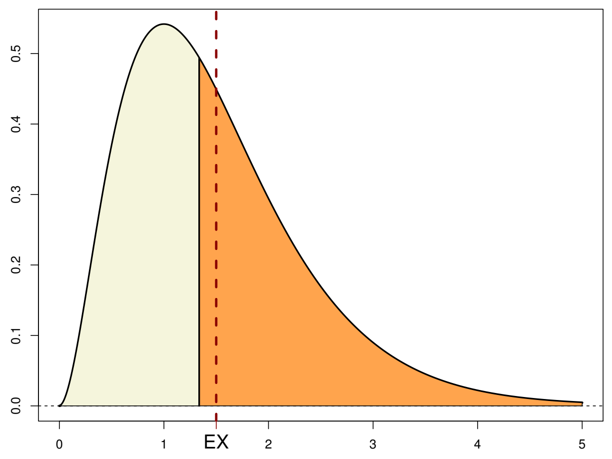
The median of $X \sim P \in \mathcal{P}(\mathbb{R})$ is any $m \in \mathbb{R}$ such that

$$\min\{\mathrm{P}(X \leq m), \mathrm{P}(X \geq m)\} \geq 1/2.$$



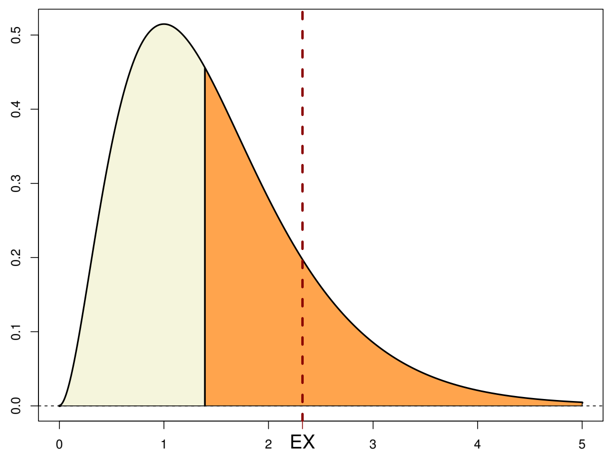
Motivation: The median

The median and **the mean** of $X \sim P \in \mathcal{P}(\mathbb{R})$



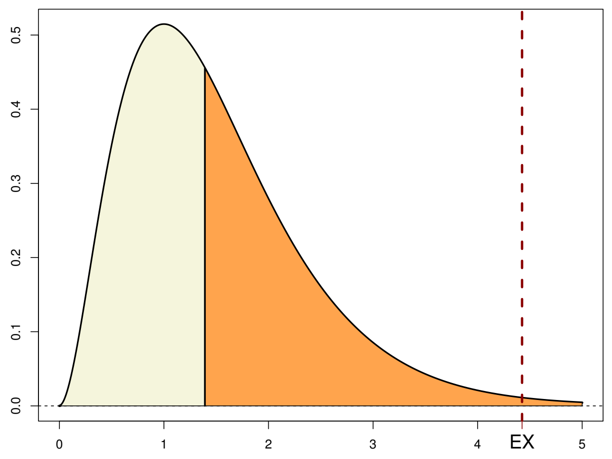
Motivation: The median

The median and the mean under contamination (5% at $x = 20$)



Motivation: The median

The median and the mean under contamination (5% at $x = 50$)



Motivation: The median

The median has a number of advantages:

- Always exists;
- Very robust (i.e., hard to disturb);
- Equivariant to monotone transformations;
- Easy to compute;
- A member of the family of quantiles $Q(\alpha) = F^{-1}(\alpha)$, $\alpha \in [0, 1]$.

Our main problem:

What is a median for multivariate data?

Motivation: The median

The median has a number of advantages:

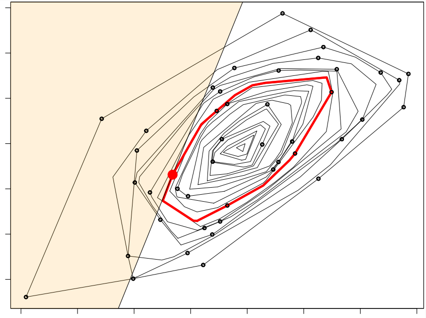
- Always exists;
- Very robust (i.e., hard to disturb);
- Equivariant to monotone transformations;
- Easy to compute;
- A member of the family of quantiles $Q(\alpha) = F^{-1}(\alpha)$, $\alpha \in [0, 1]$.

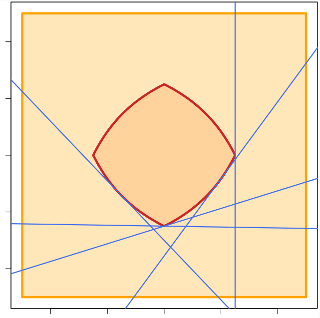
Our main problem:

What are the quantiles for multivariate data?

Statistical depth

Statistical depth function: Ordering data in multivariate spaces.



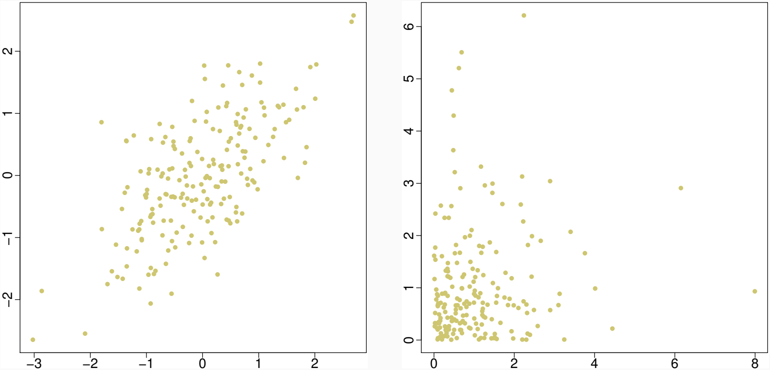


Introduced in 1975 (Tukey); studied intensively since the 1990s.

Statistical depth function

For $\mathcal{P}\left(\mathbb{R}^d\right)$ Borel probability measures on $\mathbb{R}^d$, consider the depth

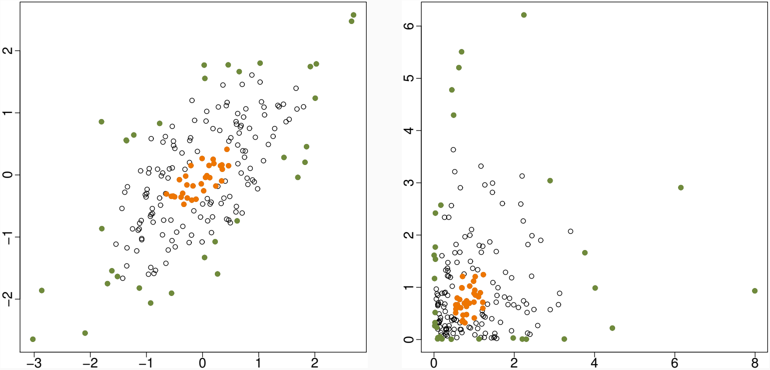
$$D\colon \mathbb{R}^d \times \mathcal{P}\left(\mathbb{R}^d\right) \to [0,1]\colon (x,P) \mapsto D(x,P).$$



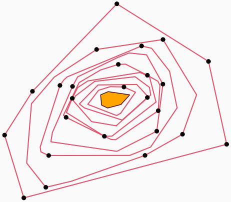
Statistical depth function

For $\mathcal{P}\left(\mathbb{R}^d\right)$ Borel probability measures on $\mathbb{R}^d$, consider the depth

$$D\colon \mathbb{R}^d \times \mathcal{P}\left(\mathbb{R}^d\right) \to [0,1]\colon (x,P) \mapsto D(x,P).$$



Principal goal: Distribution-characterizing depths



Find a depth that:

C1 characterizes probability distributions uniquely,

C2 is highly (e.g., affine) equivariant,

C3 induces robust medians, and

C4 is fast to compute.

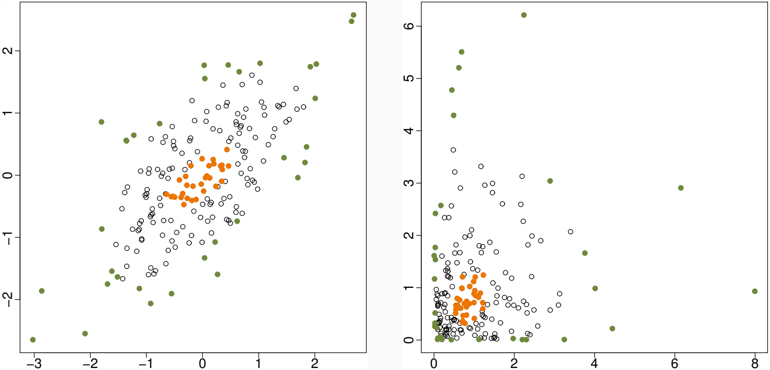
Impact: A universal framework of multivariate nonparametrics.

➥ Data exploration/statistical estimation and testing/visualisation free of parametric assumptions for complex datasets.

Halfspace depth

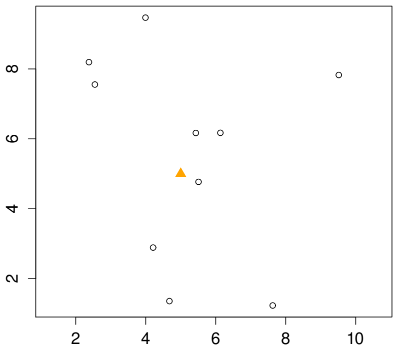
Halfspace depth (Tukey, 1975) of a point $x \in \mathbb{R}^d$ w.r.t. $P \in \mathcal{P}\left(\mathbb{R}^d\right)$

$$D(x; P) = \inf_{H \in \mathcal{H}(x)} P\left(H\right).$$



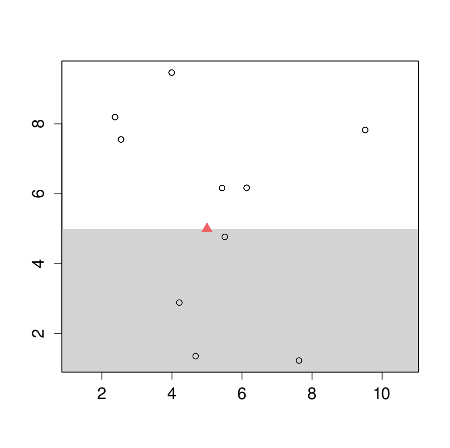
Halfspace depth

$$D\left(x; P_n\right) = \min \frac{\text{\# of observations in a halfspace that contains } x}{n}$$



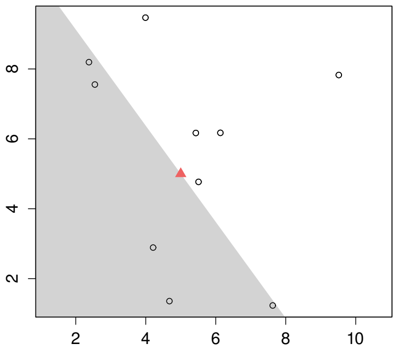
Halfspace depth

$$D(x; P_n) = \min \frac{\text{\# of observations in a halfspace that contains } x}{n}$$



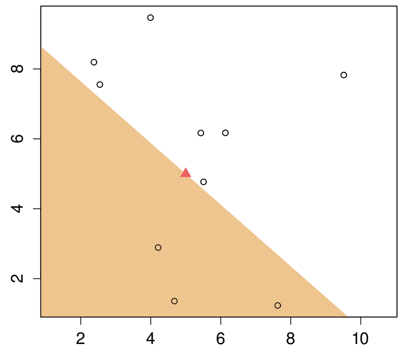
Halfspace depth

$$D(x; P_n) = \min \frac{\text{\# of observations in a halfspace that contains } x}{n}$$

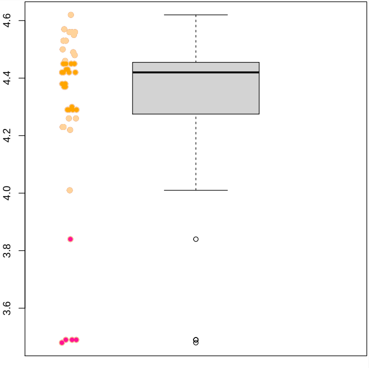


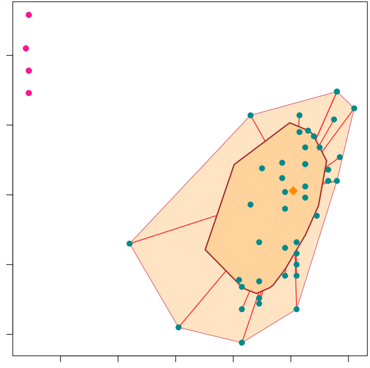
Halfspace depth

$$D(x; P_n) = \min \frac{\text{\# of observations in a halfspace that contains } x}{n}$$



Bagplot: A multivariate boxplot (Rousseeuw et al., 1999)

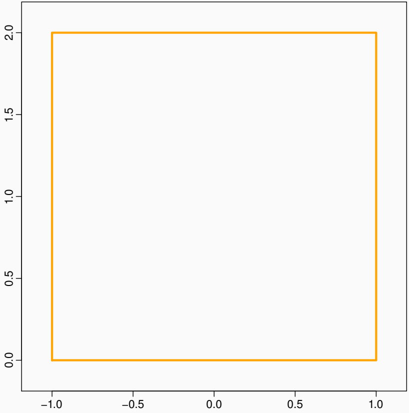




Depth: Level sets

$D(\cdot; P)$ is always quasi-concave, i.e. for each $\delta \in [0, 1]$

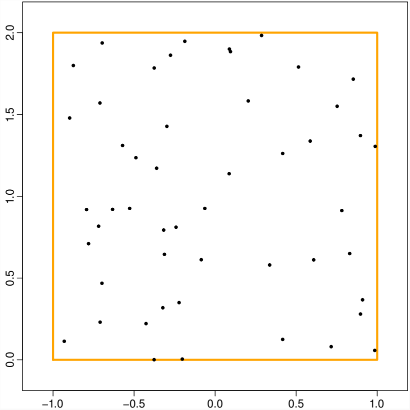
$$P_\delta = \left\{ x \in \mathbb{R}^d : D(x; P) \geq \delta \right\} \text{ is convex}$$



Depth: Level sets

$D(\cdot; P)$ is always quasi-concave, i.e. for each $\delta \in [0, 1]$

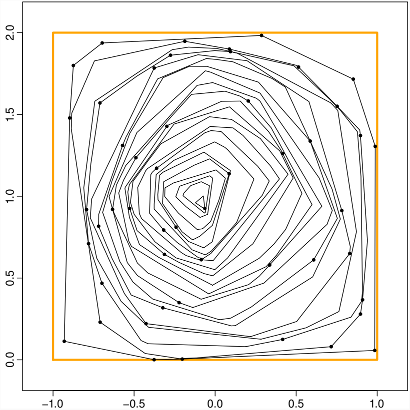
$$P_\delta = \left\{ x \in \mathbb{R}^d : D(x; P) \geq \delta \right\} \text{ is convex}$$



Depth: Level sets

$D(\cdot; P)$ is always quasi-concave, i.e. for each $\delta \in [0, 1]$

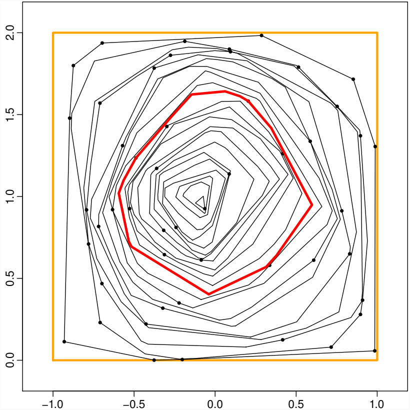
$$P_\delta = \left\{ x \in \mathbb{R}^d : D(x; P) \geq \delta \right\} \text{ is convex}$$



Depth: Level sets

$D(\cdot; P)$ is always quasi-concave, i.e. for each $\delta \in [0, 1]$

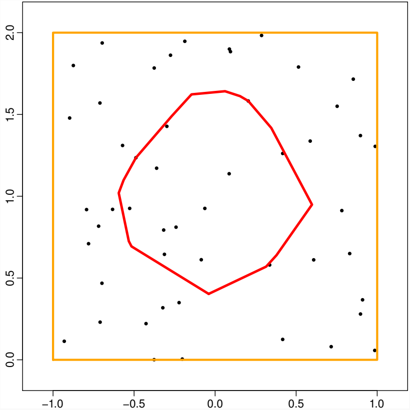
$$P_\delta = \left\{ x \in \mathbb{R}^d : D(x; P) \geq \delta \right\} \text{ is convex}$$



Depth: Level sets

$D(\cdot; P)$ is always quasi-concave, i.e. for each $\delta \in [0, 1]$

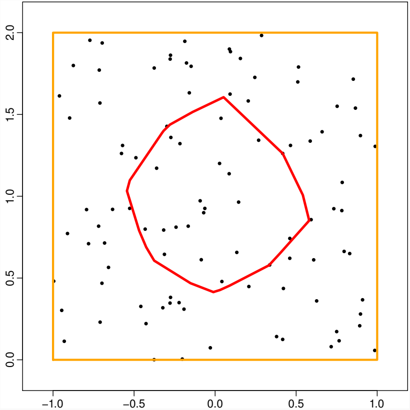
$$P_\delta = \left\{x \in \mathbb{R}^d : D(x; P) \geq \delta\right\} \text{ is convex}$$



Depth: Level sets

$D(\cdot; P)$ is always quasi-concave, i.e. for each $\delta \in [0, 1]$

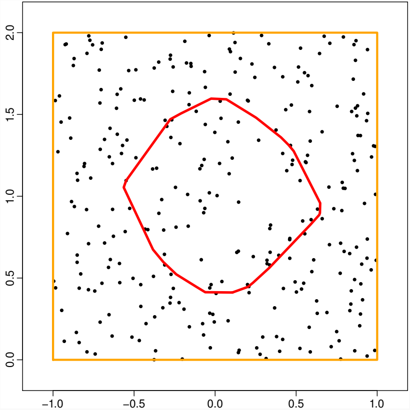
$$P_\delta = \left\{ x \in \mathbb{R}^d : D(x; P) \geq \delta \right\} \text{ is convex}$$



Depth: Level sets

$D(\cdot; P)$ is always quasi-concave, i.e. for each $\delta \in [0,1]$

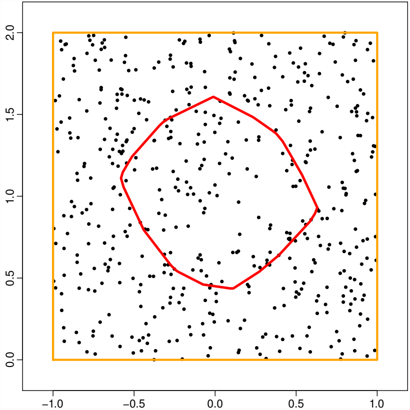
$$P_\delta = \left\{ x \in \mathbb{R}^d : D(x; P) \geq \delta \right\} \text{ is convex}$$



Depth: Level sets

$D(\cdot; P)$ is always quasi-concave, i.e. for each $\delta \in [0, 1]$

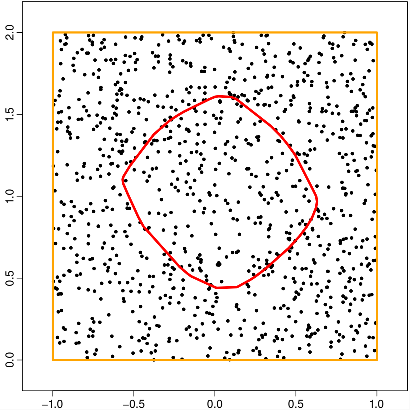
$$P_\delta = \left\{ x \in \mathbb{R}^d : D(x; P) \geq \delta \right\} \text{ is convex}$$



Depth: Level sets

$D(\cdot; P)$ is always quasi-concave, i.e. for each $\delta \in [0, 1]$

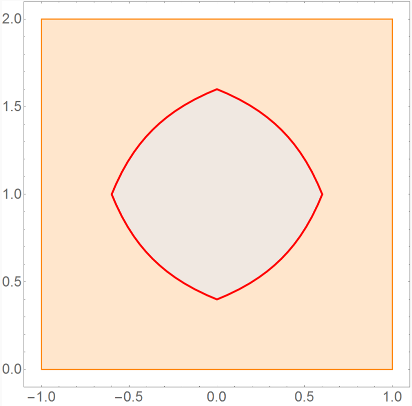
$$P_\delta = \left\{ x \in \mathbb{R}^d : D(x; P) \geq \delta \right\} \text{ is convex}$$



Depth: Level sets

We can write (Rousseeuw and Struyf, 1999; Zuo and Serfling, 2000)

$$P_\delta = \left\{x \in \mathbb{R}^d : D(x; P) \geq \delta\right\} = \bigcap \{H \in \mathcal{H} : P(H) > 1 - \delta\}.$$



Motivation: Grünbaum's inequality

Convex body $K \in \mathcal{K}^d$: compact, convex $K \subset \mathbb{R}^d$ with non-empty inter.

Identify P uniform on K with K ➥ The depth $D(x; K)$ of $K \in \mathcal{K}^d$.

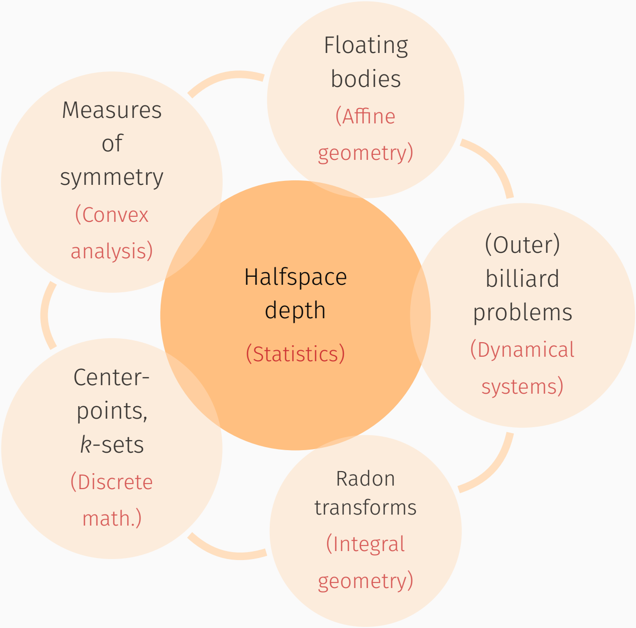
Proposition (Grünbaum, 1960)

Let $K \in \mathcal{K}^d$, and X uniform on K. Then

$$D\left(\mathsf{E}\,X; K\right) \geq \left(\frac{d}{d+1}\right)^d.$$

- $\lim_{d \to \infty} \left(\frac{d}{d+1}\right)^d = \exp(-1) \approx 0.37$.

Nonparametrics outside statistics: Halfspace depth



APPLICATIONS

DE GÉOMÉTRIE

ET

DE MÉCHANIQUE;

A LA MARINE, AUX PONTS ET CHAUSSÉES, ETC.,

POUR FAIRE SUITE

AUX DÉVELOPPEMENTS DE GÉOMÉTRIE,

PAR CHARLES DUPIN,

Membre de l'Institut de France, Académie des Sciences; ancien Secrétaire de l'Académie Ionnienne, Associé étranger de l'Institut de Naples, Associé honoraire de l'Académie royale d'Irlande, et de la Société des Ingénieurs civils de la Grande-Bretagne, Membre des Académies royales des Sciences de Stockholm, de Turin, de Montpellier, etc., de la Société des Arts de Genève, de la Société d'Encouragement pour l'industrie française, Membre du Comité consultatif des Arts et Manufactures de France, Professeur de Méchanique au Conservatoire, Officier supérieur au corps du Génie Maritime, et Membre de la Légion-d'Honneur.

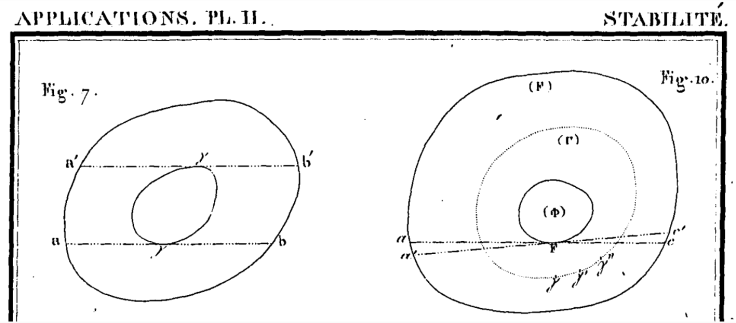
PARIS,

BACHELIER, SUCCESSEUR DE Mme. Ve. COURCIER, LIBRAIRE,

QUAI DES AUGUSTINS.

1822.

Floating body

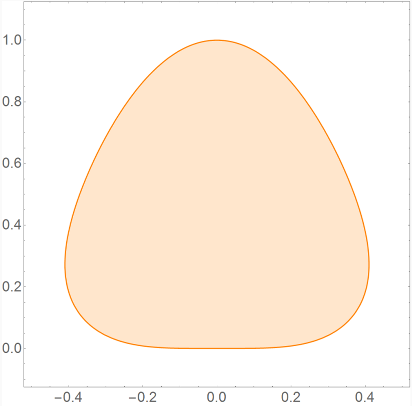


Definition (Dupin, 1822)

A convex body $K_{[\delta]}$ is the Dupin floating body of $K \in \mathcal{K}^d$ for $\delta \geq 0$ if each supporting hyperplane of $K_{[\delta]}$ cuts off a set of volume δ from K.

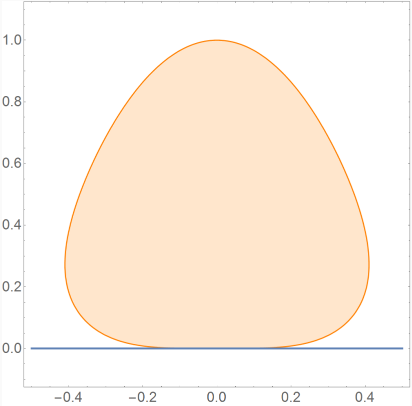
Floating body

Dupin's floating body of $K \in \mathcal{K}^2$ for $\delta = 0.3$



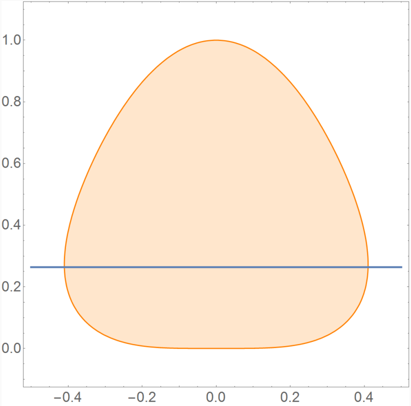
Floating body

Dupin's floating body of $K \in \mathcal{K}^2$ for $\delta = 0.3$



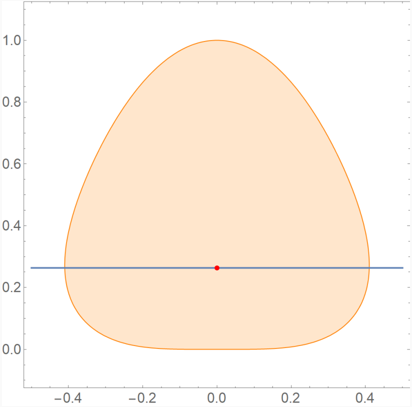
Floating body

Dupin's floating body of $K \in \mathcal{K}^2$ for $\delta = 0.3$



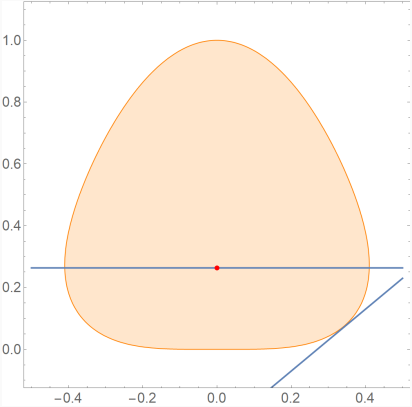
Floating body

Dupin's floating body of $K \in \mathcal{K}^2$ for $\delta = 0.3$



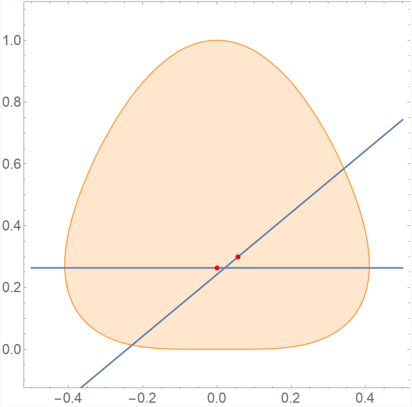
Floating body

Dupin's floating body of $K \in \mathcal{K}^2$ for $\delta = 0.3$



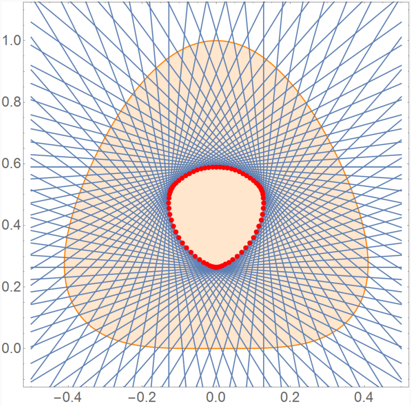
Floating body

Dupin's floating body of $K \in \mathcal{K}^2$ for $\delta = 0.3$



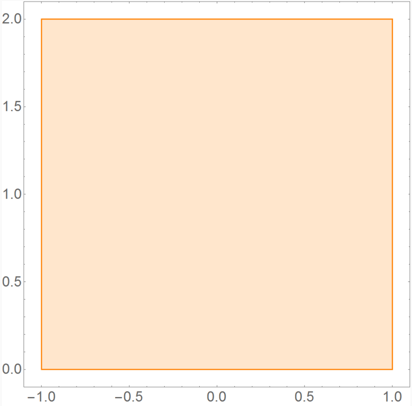
Floating body

Dupin's floating body of $K \in \mathcal{K}^2$ for $\delta = 0.3$



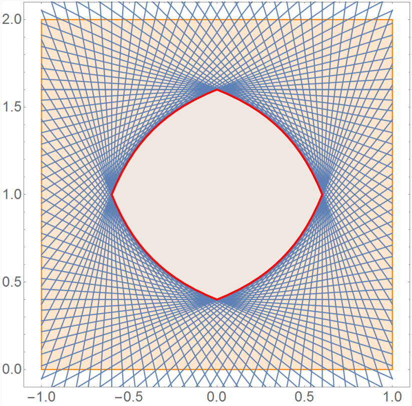
Floating body

Dupin's floating body of $K \in \mathcal{K}^2$ for $\delta = 0.3$



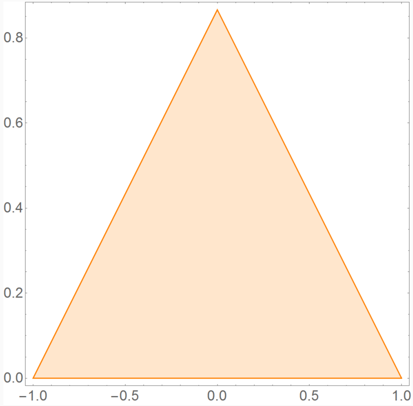
Floating body

Dupin's floating body of $K \in \mathcal{K}^2$ for $\delta = 0.3$



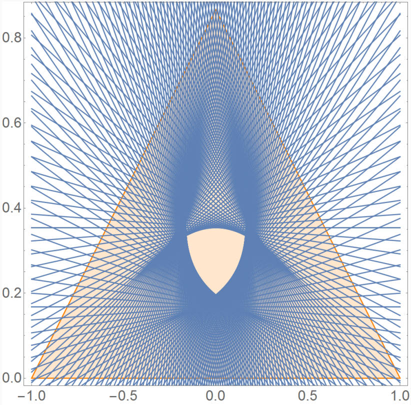
Floating body

Dupin's floating body of K does not have to exist



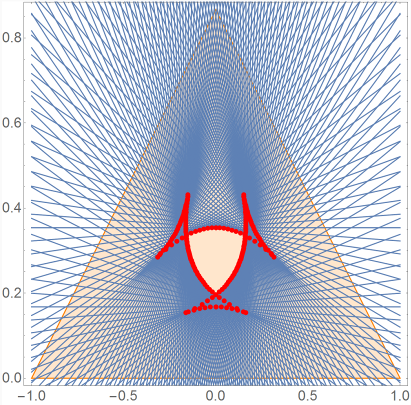
Floating body

Dupin’s floating body of K does not have to exist



Floating body

Dupin's floating body of K does not have to exist



Definition (Schütt and Werner, 1990)
Let $K \in \mathcal{K}^d$ with $\operatorname{vol}(K) = 1$ and let $\delta \in (0, 1/2)$.
The **convex floating body** of K associated with δ is given by

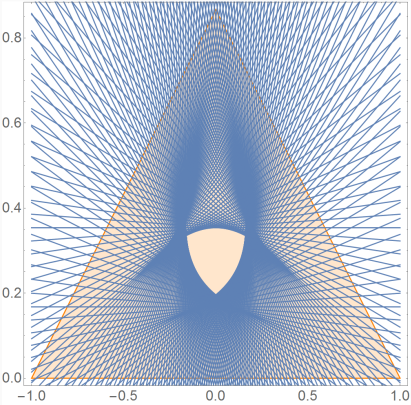
$$K_\delta = \bigcap \{H \in \mathcal{H} \colon \operatorname{vol}(K \cap H) \geq 1 - \delta\}.$$

Proposition (Schütt and Werner, 1990)

- *K_δ always exists.*
- *If $K_{[\delta]}$ exists, then $K_{[\delta]} = K_\delta$.*
- *Just as $K_{[\delta]}$, also K_δ has "nice" properties.*

Convex floating body

Convex floating body of K always exists

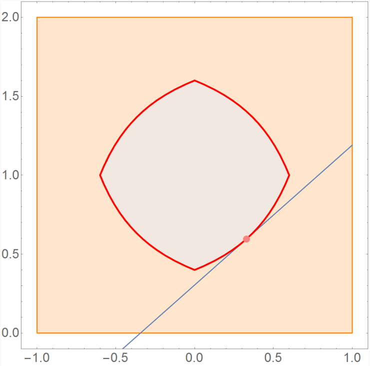


Depth: Asymptotic normality

Let $P_n \in \mathcal{P}\left(\mathbb{R}^d\right)$ be the empirical measure of n i.i.d. variables from P.

$\sqrt{n}\,(D(x; P_n) - D(x; P))$ is asymptotically normal

$\iff$ $D(x; P)$ is realised by a single halfspace $H \in \mathcal{H}(x)$ (Massé, 2004)

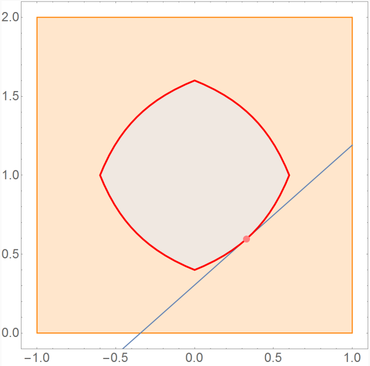


Depth: Asymptotic normality

Let $P_n \in \mathcal{P}\left(\mathbb{R}^d\right)$ be the empirical measure of n i.i.d. variables from P.

$\sqrt{n}\,(D(x;P_n) - D(x;P))$ is asymptotically normal

$$\iff \quad \text{the contour of } D(\cdot;P) \text{ is smooth at } x$$

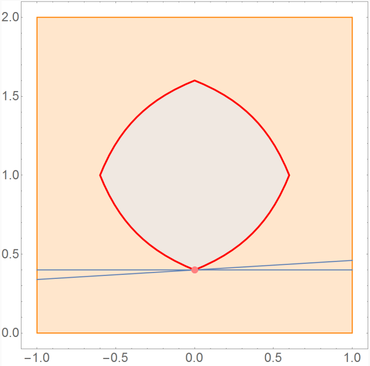


Depth: Asymptotic normality

Let $P_n \in \mathcal{P}\left(\mathbb{R}^d\right)$ be the empirical measure of n i.i.d. variables from P.

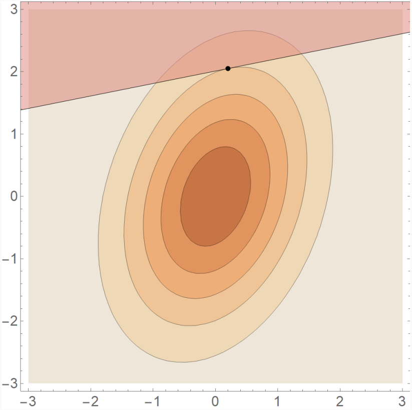
$\sqrt{n}\,(D(x; P_n) - D(x; P))$ is asymptotically normal

$\iff$ the contour of $D(\cdot; P)$ is smooth at x



Problem: Smoothness of the depth

Elliptically symmetric distributions have smooth depth contours



Smoothness of floating bodies

Problem (Massé and Theodorescu, 1994)

Is there a non-elliptical distribution with smooth depth contours?

Proposition (Meyer and Reisner, 1991)

Uniform distributions on smooth, symmetric, strictly convex bodies have smooth depth.

Open problem: An analogous result for $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ with a density?

Depth characterization conjecture

Question: (Struyf and Rousseeuw, 1999)
Does for any $P \neq Q$ in $\mathcal{P}\left(\mathbb{R}^d\right)$ exist $x \in \mathbb{R}^d$ such that $D(x;P) \neq D(x;Q)$?

Positive answers for $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ such that:

- $d = 1$ (there depth $\sim$ distribution function).
- P is purely atomic, with finitely many atoms.
  (Struyf and Rousseeuw, 1999; Koshevoy, 2002; Laketa and Nagy, 2021)
- P is atomic. (Cuesta-Albertos and Nieto-Reyes, 2008)
- P is properly integrable. (Koshevoy, 2003)
- P has a smooth density. (Hassairi and Regaieg, 2008)
- all Dupin's floating bodies of P exist.
  (Kong and Zuo, 2010; Nagy, Schütt, Werner, 2019)

Conjectured positive answer.
(Cuesta-Albertos and Nieto-Reyes, 2008; Kong and Mizera, 2012)

Depth characterization conjecture

Question: (Struyf and Rousseeuw, 1999)

Does for any $P \neq Q$ in $\mathcal{P}\left(\mathbb{R}^d\right)$ exist $x \in \mathbb{R}^d$ such that $D(x; P) \neq D(x; Q)$?

Positive answers for $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ such that:

- $d = 1$ (there depth $\sim$ distribution function).
- P is purely atomic, with finitely many atoms.
  (Struyf and Rousseeuw, 1999; Koshevoy, 2002; Laketa and Nagy, 2021)
- ~~P is atomic.~~ (Cuesta-Albertos and Nieto-Reyes, 2008)
- ~~P is properly integrable.~~ (Koshevoy, 2003)
- ~~P has a smooth density.~~ (Hassairi and Regaieg, 2008)
- all Dupin's floating bodies of P exist.
  (Kong and Zuo, 2010; Nagy, Schütt, Werner, 2019)

Conjectured positive answer.

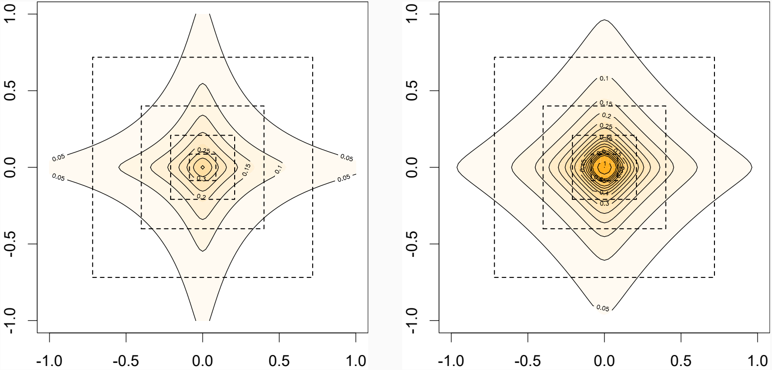
(Cuesta-Albertos and Nieto-Reyes, 2008; Kong and Mizera, 2012)

Characterization conjecture

Question: (Struyf and Rousseeuw, 1999)

Does for any $P \neq Q$ in $\mathcal{P}\left(\mathbb{R}^d\right)$ exist $x \in \mathbb{R}^d$ such that $D(x; P) \neq D(x; Q)$?

Not for $d > 1$.



Measures of symmetry of convex bodies

Definition (Grünbaum, 1963)

A mapping $\rho\colon \mathcal{K}^d \to [0,1]$ is a measure of symmetry if

- $\rho(K) = 1$ if and only if K is symmetric;
- $\rho(T(K)) = \rho(K)$ for non-singular affine transforms $T\colon \mathbb{R}^d \to \mathbb{R}^d$; and
- ρ is continuous.

The Winternitz measure of symmetry: (Winternitz, 1910s)

$$\rho(K) = 2 \max_{x \in \mathbb{R}^d} D(x; K),$$

i.e. twice the depth of the halfspace median of K.

Funk's characterization of symmetry

Theorem

$K \in \mathcal{K}^d$ is symmetric around the origin 0 if and only if

$$\mathrm{vol}\,(K \cap H) = \mathrm{vol}\,(K)\,/2$$

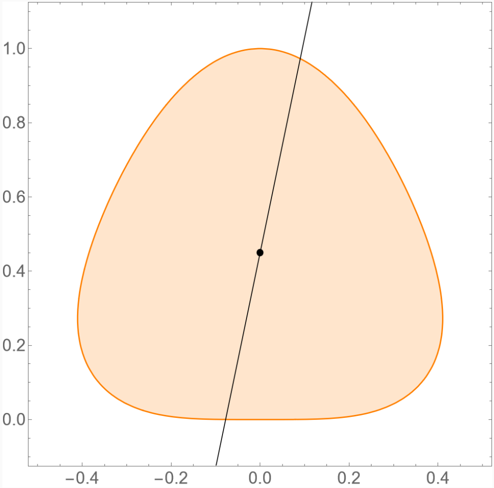
for every halfspace $H \in \mathcal{H}(0)$.

Proof:

- $\mathbb{R}^2$: easy;
- $\mathbb{R}^3$: proved in the 1910s (Funk, 1915);
- $\mathbb{R}^d$: proved **50 years later** using spherical harmonics (Schneider, 1970).

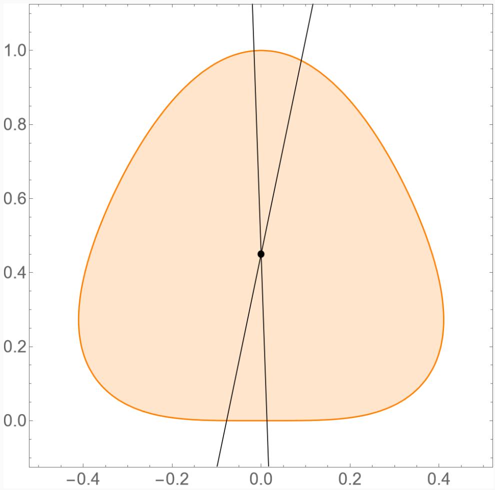
Funk theorem: Proof for $d = 2$

$$D(x; K) = 1/2 \implies K \in \mathcal{K}^d \text{ is symmetric around } x$$



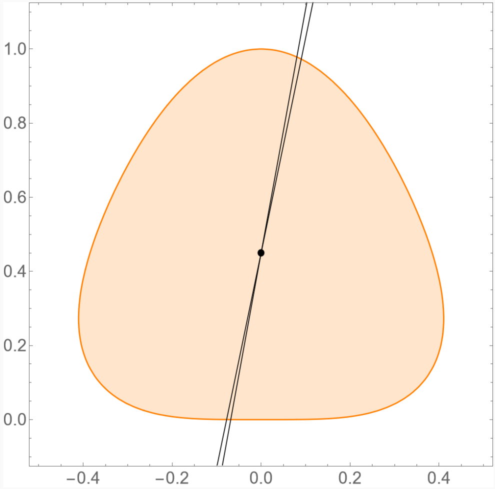
FUNK THEOREM: PROOF FOR $d = 2$

$$D(x; K) = 1/2 \implies K \in \mathcal{K}^d \text{ is symmetric around } x$$



Funk theorem: Proof for $d = 2$

$$D(x; K) = 1/2 \implies K \in \mathcal{K}^d \text{ is symmetric around } x$$



Symmetry of measures in statistics

A measure $P \in \mathcal{P}(\mathbb{R}^d)$ with $X \sim P$ is called (Zuo and Serfling, 2000)

- halfspace symmetric around $x \in \mathbb{R}^d$ if $D(x; P) \geq 1/2$,
- angularly symmetric around $x \in \mathbb{R}^d$ if

$$\frac{X - x}{\|X - x\|} \overset{d}{=} -\frac{X - x}{\|X - x\|}.$$

Theorem (Funk, 1915 and Schneider, 1970)

A uniform distribution on $K \in \mathcal{K}^d$ is halfspace symmetric if and only if it is angularly symmetric.

Zuo and Serfling (2000)

Theorem 2.6. *Suppose a random vector X is halfspace symmetric about a unique point $\theta \in \mathbb{R}^d$, and either*

(1) *X is continuous, or*
(2) *X is discrete and $P(X = \theta) = 0$.*

Then X is angularly symmetric about θ.

Proof:

To prove that (iv) $\Rightarrow$ (i), take $d = 2$ for the sake of simplicity. First we show that if $P(X \in H) = P(X \in -H)$ for any closed halfspace H with the origin on the boundary, then

$$P(X \in H_1 \cap H_2) = P(X \in -H_1 \cap -H_2) \tag{A.1}$$

Dutta, Ghosh, Chaudhuri (2011)

Theorem 2. *Suppose that* $\mathbf{X}$ *is a* d*-dimensional random vector with a probability distribution which has its half-space median at* $\boldsymbol{\mu} \in \mathbb{R}^d$*. Then, the half-space depth of* $\boldsymbol{\mu}$ *will be* 0.5 *if and only if* $(\mathbf{X} - \boldsymbol{\mu})/\|\mathbf{X} - \boldsymbol{\mu}\|_2$ *and* $(\boldsymbol{\mu} - \mathbf{X})/\|\mathbf{X} - \boldsymbol{\mu}\|_2$ *are identically distributed.*

Proof:

First, we shall prove it for the bivariate case, that is, $d = 2$. Without loss of generality, we assume that $\boldsymbol{\mu} = \mathbf{0}$. Let Z be the angle between the positive side of the x_1-axis and the random vector $\mathbf{X}$ (measured counterclockwise from the x_1-axis). Now, consider a straight line which

Rousseeuw and Struyf (2004)

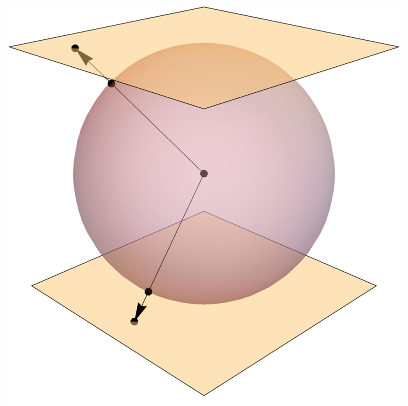
Corollary 1. *When P has a density, then P is angularly symmetric about some $\boldsymbol{\theta}_0$ if and only if*

$$\max_{\boldsymbol{\theta}} \operatorname{ldepth}(\boldsymbol{\theta}) = \tfrac{1}{2}.$$

Proof: For any dimension *d*.

Idea of the proof (Rousseeuw and Struyf, 2004)

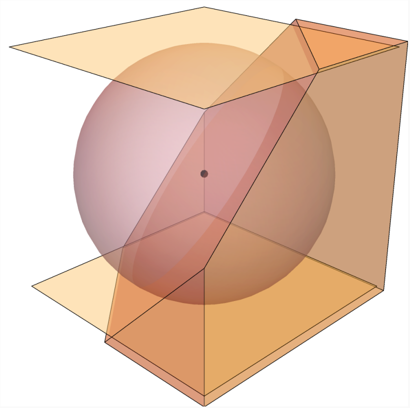
(i). The map $x \mapsto (x_1/|x_d|, x_2/|x_d|, \ldots, x_d/|x_d|)$ takes $\mathcal{H}(0)$ to halfspaces inside hyperplanes $H^{\pm} = \{x \in \mathbb{R}^d : x_d = \pm 1\}$.



(ii). Apply the Cramér-Wold theorem (Cramér and Wold, 1936) in $\mathbb{R}^{d-1}$.

Idea of the proof (Rousseeuw and Struyf, 2004)

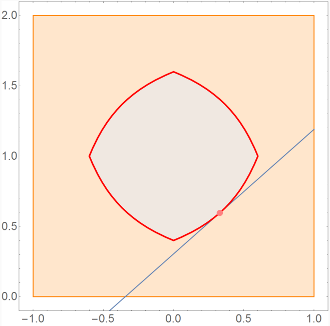
(i). The map $x \mapsto (x_1/|x_d|, x_2/|x_d|, \ldots, x_d/|x_d|)$ takes $\mathcal{H}(0)$ to halfspaces inside hyperplanes $H^{\pm} = \{x \in \mathbb{R}^d : x_d = \pm 1\}$.



(ii). Apply the Cramér-Wold theorem (Cramér and Wold, 1936) in $\mathbb{R}^{d-1}$.

Minimizing halfspace and barycentric cut

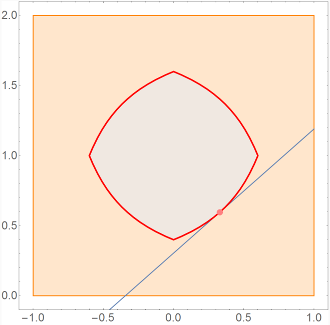
- $H \in \mathcal{H}(x)$ is a minimizing halfspace of P at x if $P(H) = D(x; P)$.
- A hyperplane ∂H is a barycentric cut of P at x if the centroid of the cut (conditional distribution) of P by ∂H is x.



➟ For $K \in \mathcal{K}^d$, the boundary of any minimizing halfspace is a barycentric cut (Blaschke, 1917).

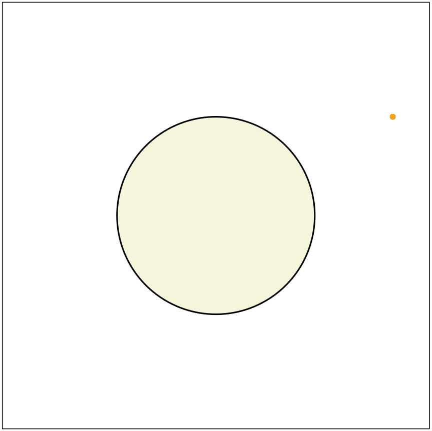
Reconstructing K from its floating body

- For $K \in \mathcal{K}^d$, the boundary of any minimizing halfspace is a barycentric cut (Blaschke, 1917).
- Starting from a single point $y \in \partial K$, reconstruct the boundary of K.
- Outer billiards with K_δ as a table (Tabachnikov, 1995).

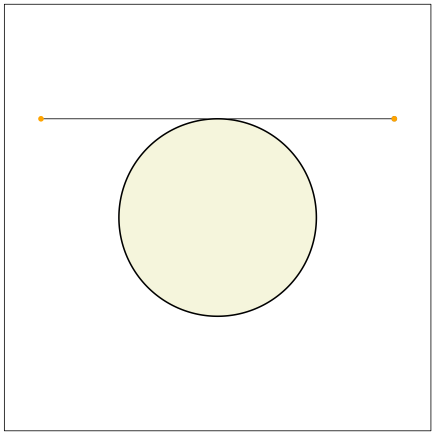


Outer billiard: Circle I

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.

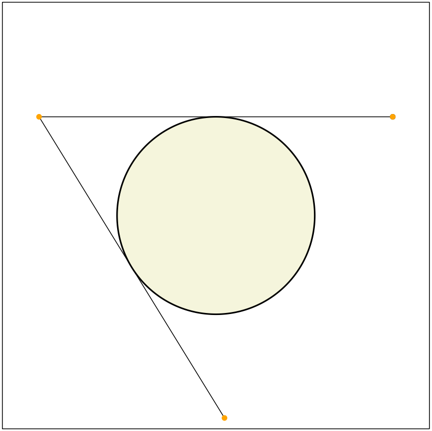


Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



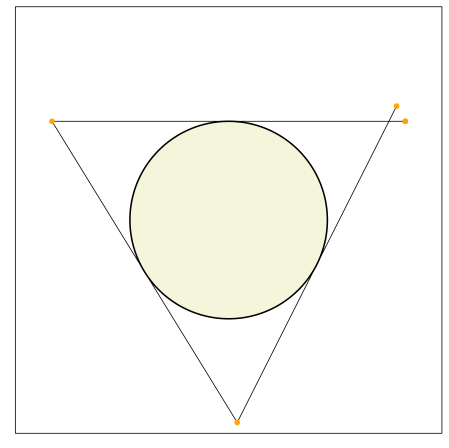
Outer billiard: Circle I

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



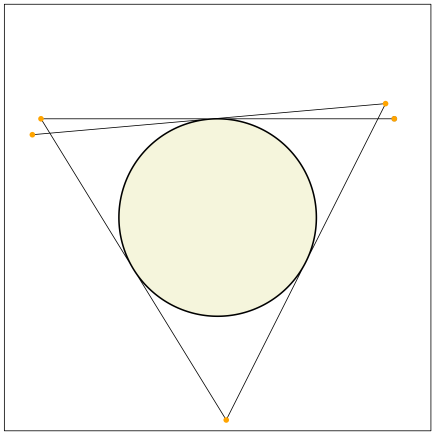
Outer billiard: Circle I

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



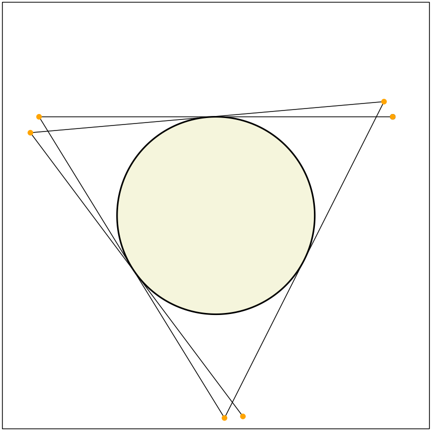
Outer billiard: Circle I

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



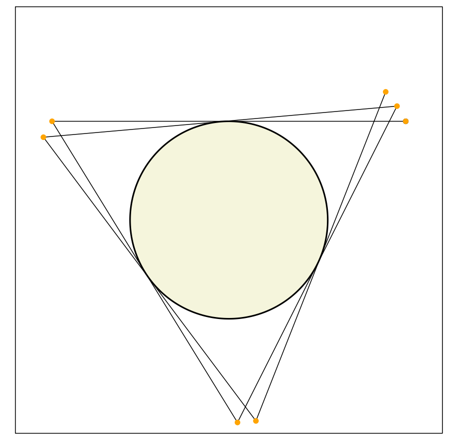
Outer billiard: Circle I

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.

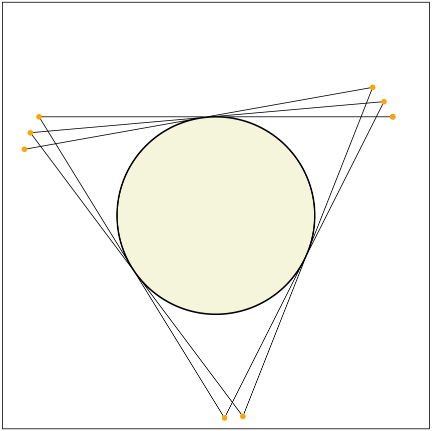


Outer billiard: Circle I

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.

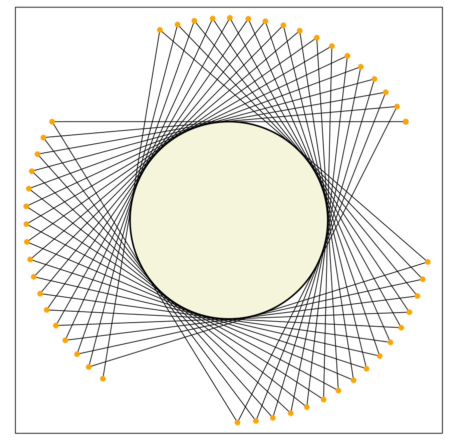


Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



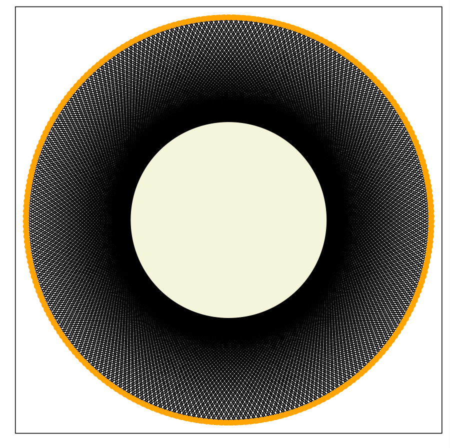
Outer billiard: Circle I

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



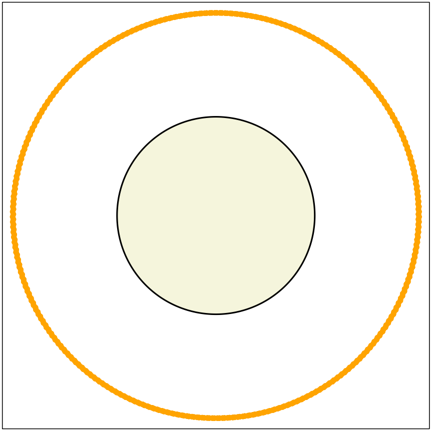
Outer billiard: Circle I

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



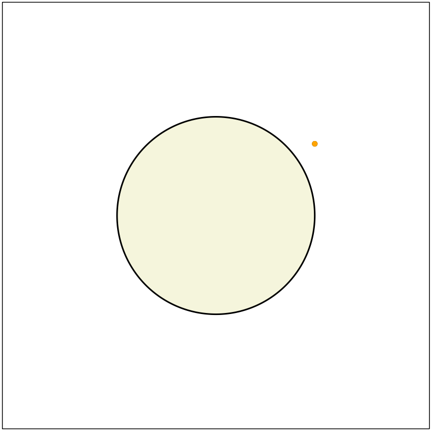
Outer billiard: Circle I

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



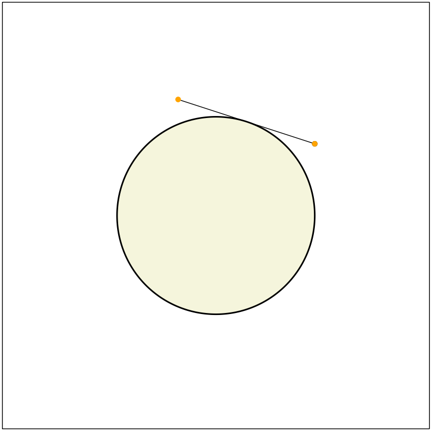
Outer billiard: Circle II

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



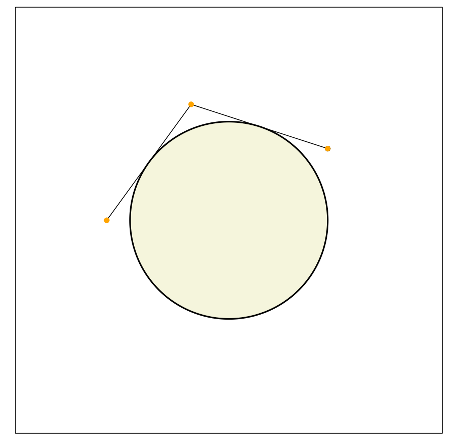
Outer billiard: Circle II

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.

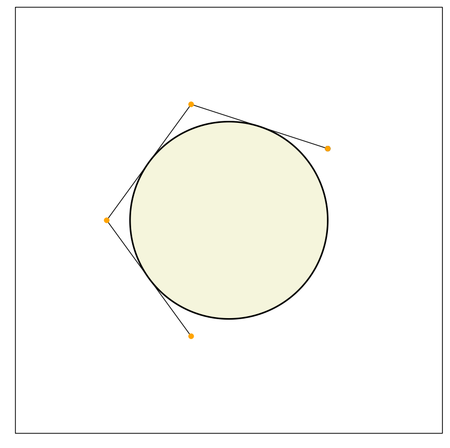


Outer billiard: Circle II

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.

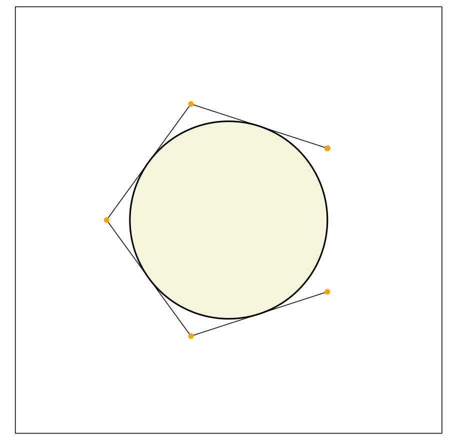


Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



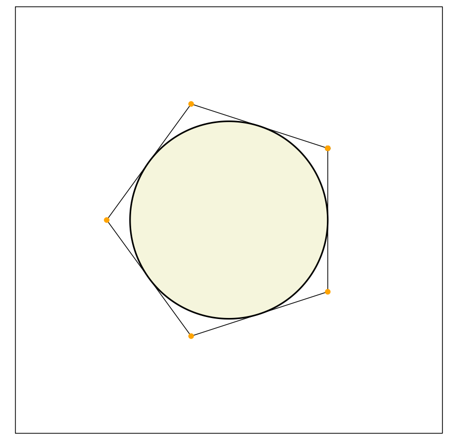
Outer billiard: Circle II

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



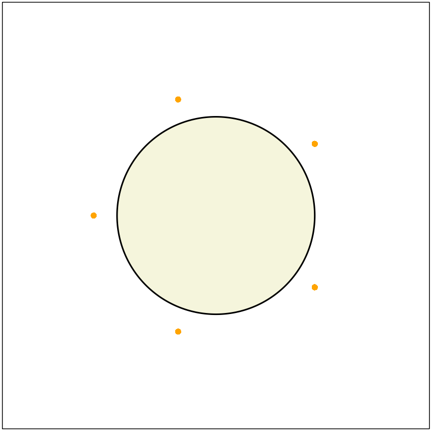
Outer billiard: Circle II

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



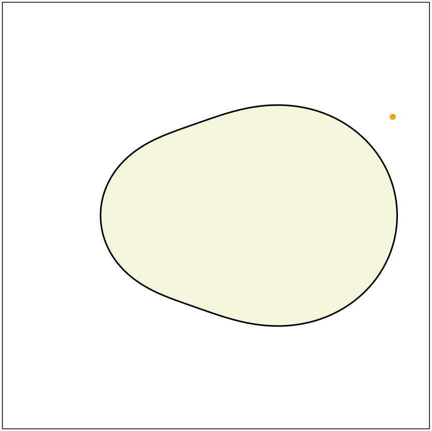
Outer billiard: Circle II

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



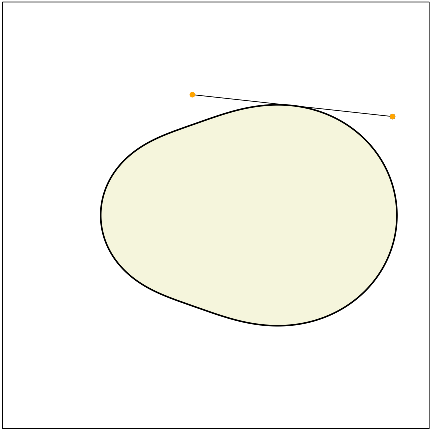
Outer billiard: Egg I

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



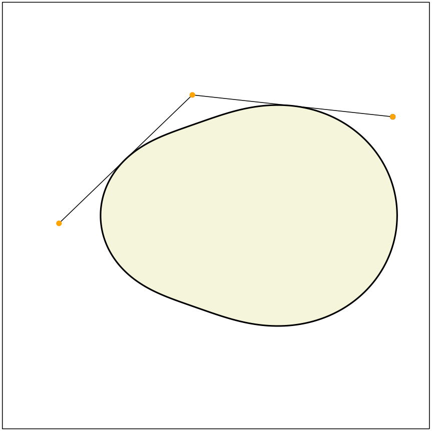
Outer billiard: Egg I

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



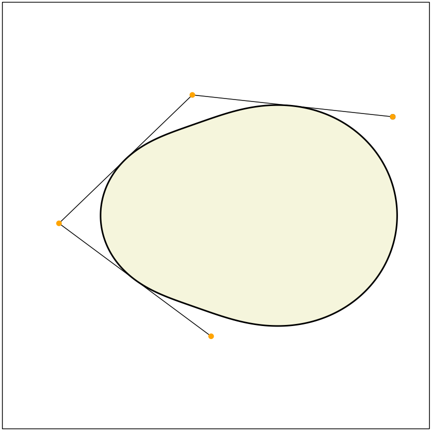
Outer billiard: Egg I

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



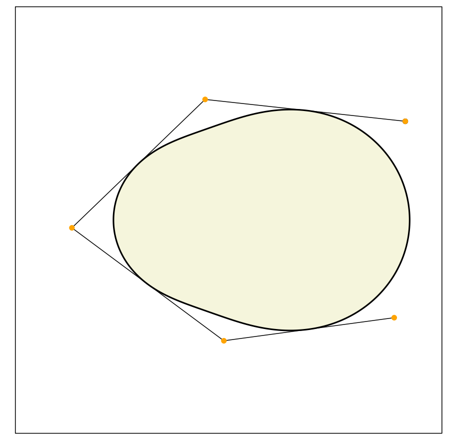
Outer billiard: Egg I

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



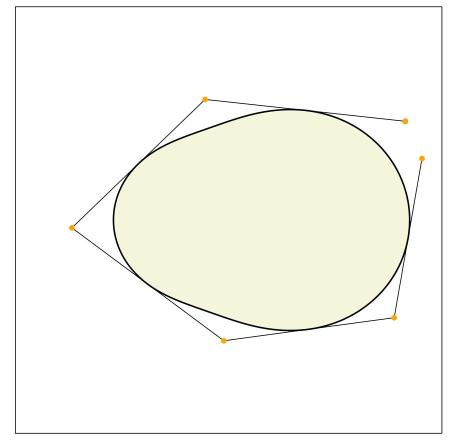
Outer billiard: Egg I

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



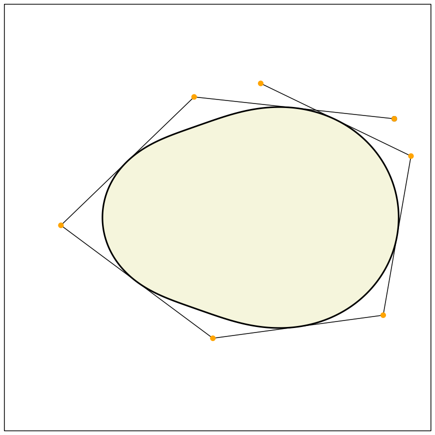
Outer billiard: Egg I

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



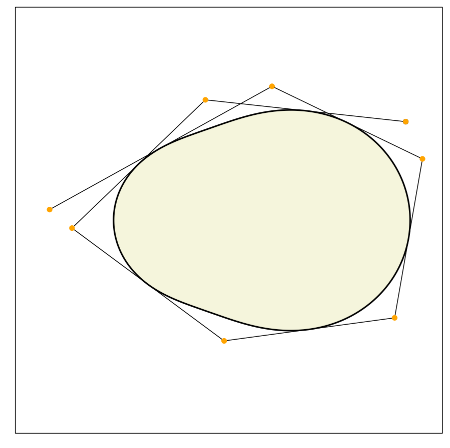
Outer billiard: Egg I

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



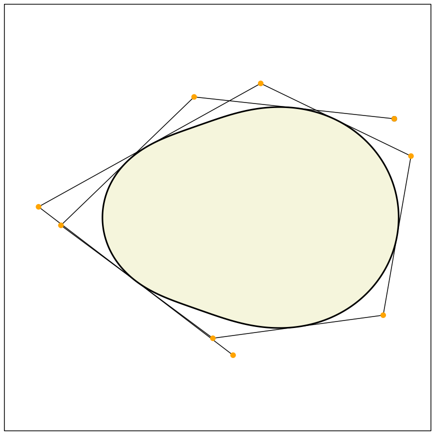
Outer billiard: Egg I

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.

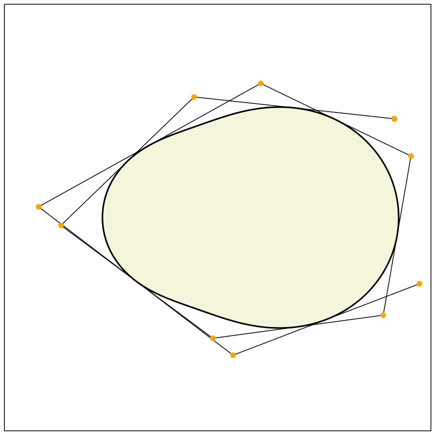


Outer billiard: Egg I

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.

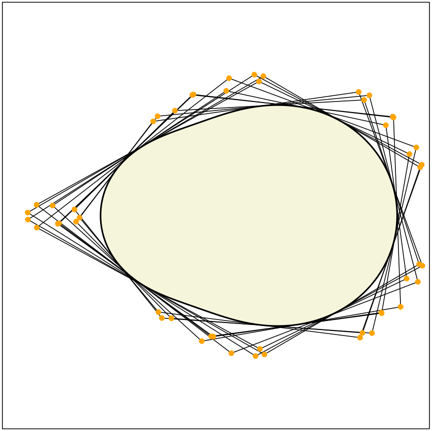


Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



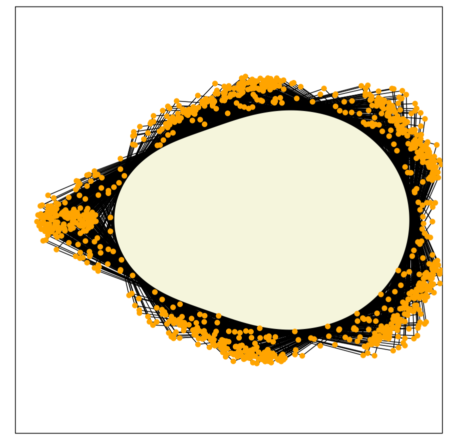
Outer billiard: Egg I

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



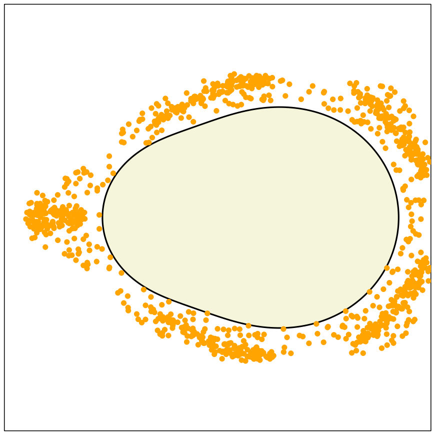
Outer billiard: Egg I

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



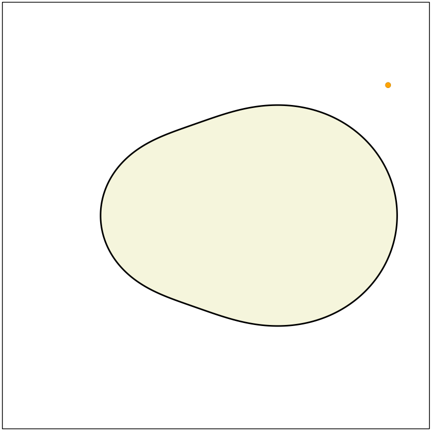
Outer billiard: Egg I

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



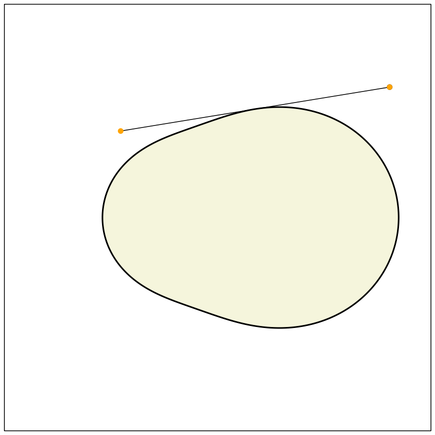
Outer billiard: Egg II

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



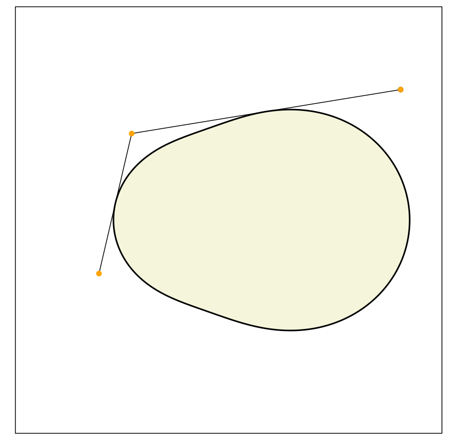
Outer billiard: Egg II

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



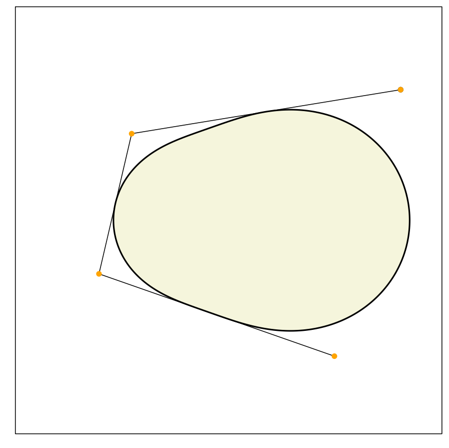
Outer billiard: Egg II

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.

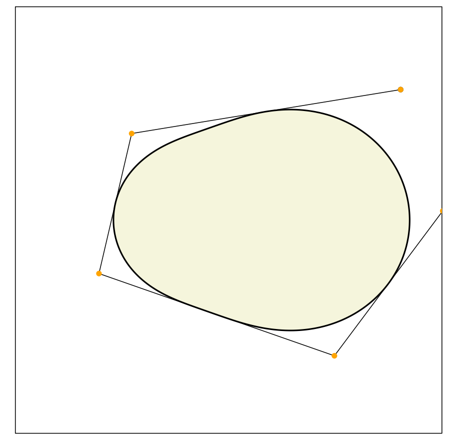


Outer billiard: Egg II

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.

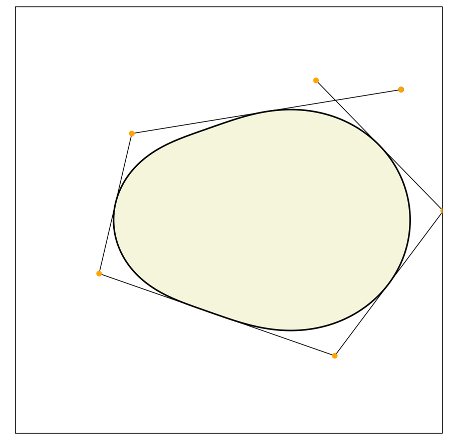


Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



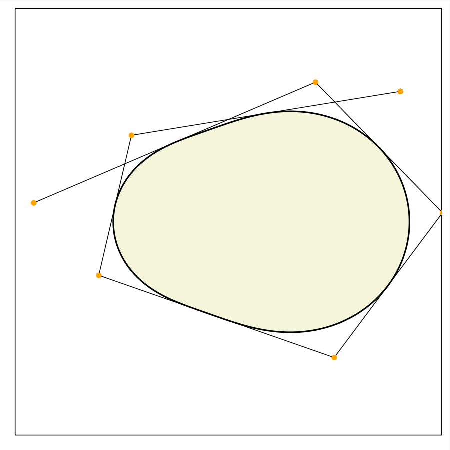
Outer billiard: Egg II

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



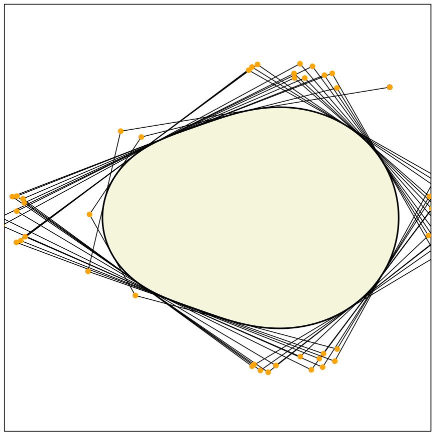
Outer billiard: Egg II

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



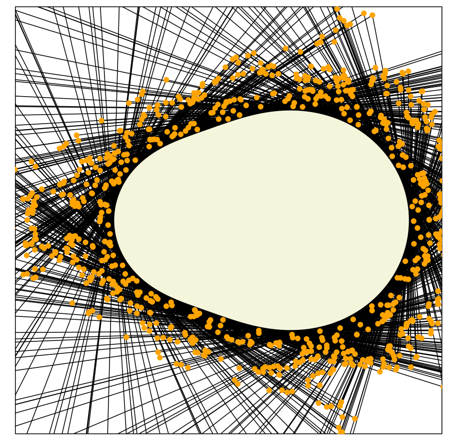
Outer billiard: Egg II

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



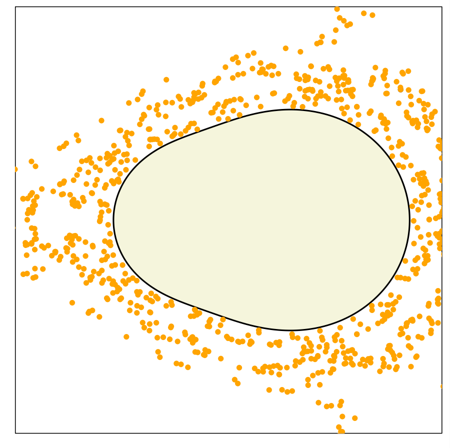
Outer billiard: Egg II

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



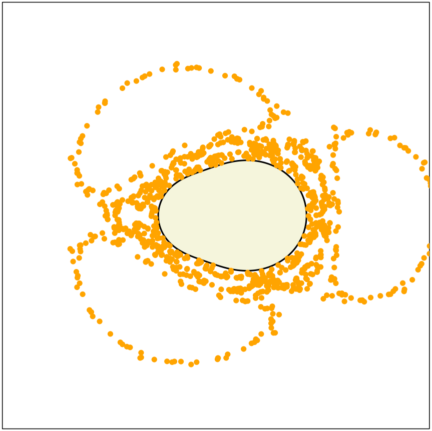
Outer billiard: Egg II

Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



Outer billiard: Egg II

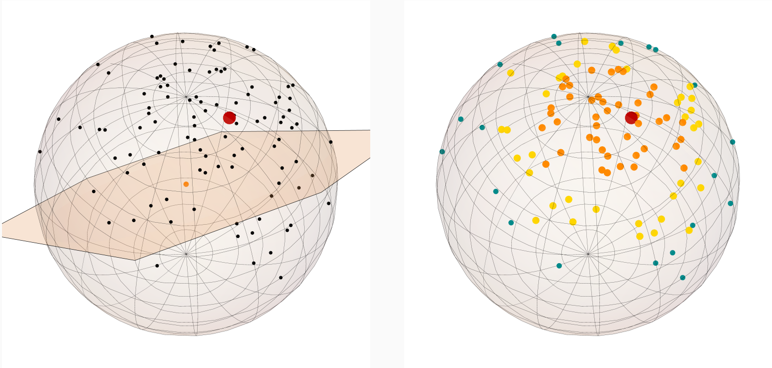
Reconstructing $K \in \mathcal{K}^2$ from its floating body using barycentric cuts.



Depth for directional data

Angular halfspace depth (Small, 1987) of $x \in \mathbb{S}^{d-1}$ w.r.t. $P \in \mathcal{P}\left(\mathbb{S}^{d-1}\right)$

$$aHD(x; P) = \inf \left\{ P(H) : H \in \mathcal{H}(0), x \in H \right\}.$$



Angular halfspace depth: almost forgotten

Pandolfo, Paindaveine, Porzio (2018):

> *"The main drawback of the angular halfspace depth is the computational effort it requires, especially for higher dimensions."*

In *R*, function ***sdepth*** in package ***depth*** (Genest et al., 2019)

- works only for $d = 2, 3$,
- computing *aHD* for **a single point** $x \in \mathbb{S}^2$:
  - for sample size $n = 200$ takes 7 seconds,
  - for sample size $n = 500$ takes 2 minutes,
  - for sample size $n = 1000$ takes forever.

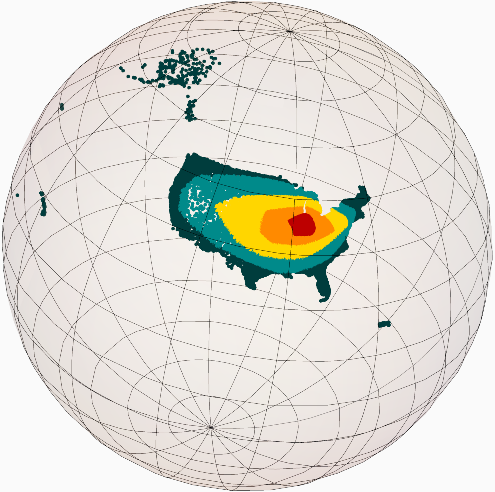
A GEOMETRIC ALGORITHM C2(M): $d = 3$ (IN SECONDS)

$d = 3$	single Point		$m = 1000$ points		all data points	
n	C2(m)	*sdepth*	C2(m)	*sdepth*	C2(m)	*sdepth*
40	0.00030	0.00030	0.00308	0.250	0.00024	0.0103
80	0.00075	0.00182	0.00697	1.88	0.00089	0.143
160	0.00292	0.0136	0.0165	13.6	0.00318	2.35
320	0.0110	0.111	0.0421	111	0.0120	35.6
640	0.0436	1.08	0.109	995	0.0477	637
1280	0.180	8.90	0.315	8690	**0.197**	**11100**
2560	0.786	72.1	1.11		0.797	
5120	3.18	583	3.83		3.47	
10240	12.7		14.1		14.0	
20480	61.3		55.7		67.1	
40960	239		244		278	
81920	1010		1020		1170	

Algorithm *sdepth* was implemented in *C++* to get a fair comparison.

Example: Depth of US counties

Exact *aHD* of all $n = 33\,144$ US counties, computation time <2 mins.



aHD-median: Vermillion County, Indiana ($aHD = 14\,571/n$).

Conclusion

Quantiles and multivariate data:

- Different approaches; inherently geometric.
- Halfspace depth and the floating body are the same concept.
- Halfspace depth does not characterize distributions.
- **Huge overlaps of statistics with geometry**.

What we do not know:

- When are floating bodies smooth?
- When does halfspace depth characterize distributions?
- Is the triangle characterized by its halfspace depth?
- How to reconstruct P from its depth? (Homothety conjecture)
- Structural properties of the depth level sets?
- How to compute the median efficiently?

Selected literature

[1] David L. Donoho and Miriam Gasko. Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.*, 20(4):1803–1827, 1992.

[2] Charles Dupin. *Applications de Géométrie et de Méchanique*. Bachelier, Paris, 1822.

[3] Branko Grünbaum. Partitions of mass-distributions and of convex bodies by hyperplanes. *Pacific J. Math.*, 10:1257–1261, 1960.

[4] Peter J. Rousseeuw and Ida Ruts. The depth function of a population distribution. *Metrika*, 49(3):213–244, 1999.

[5] Carsten Schütt and Elisabeth M. Werner. The convex floating body. *Math. Scand.*, 66(2):275–290, 1990.

[6] Christopher G. Small. Measures of centrality for multivariate and directional distributions. *Canad. J. Statist.*, 15(1):31–39, 1987.

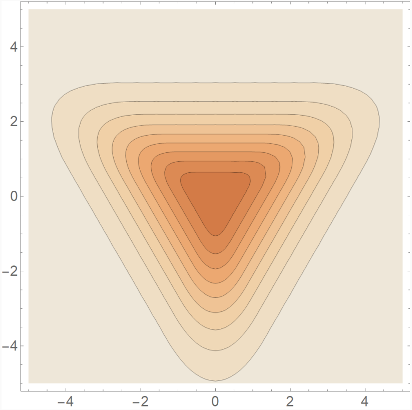
[7] Anja Struyf and Peter J. Rousseeuw. Halfspace depth and regression depth characterize the empirical distribution. *J. Multivariate Anal.*, 69(1):135–153, 1999.

[8] John W. Tukey. Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians, Vol. 2*, pages 523–531, 1975.

[9] Yijun Zuo and Robert Serfling. General notions of statistical depth function. *Ann. Statist.*, 28(2):461–482, 2000.

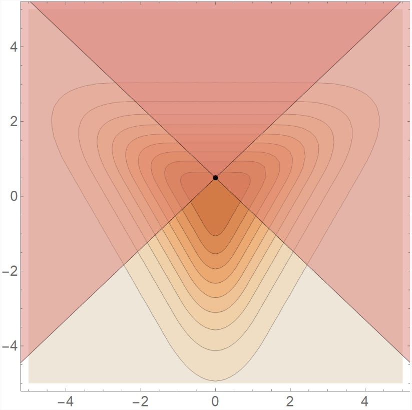
Problem: Smoothness of the depth

Smooth quasi-concave density is not sufficient for smooth depth



Problem: Smoothness of the depth

Smooth quasi-concave density is not sufficient for smooth depth



Depth characterization: Proof I

A measure $P \in \mathcal{P}\left(\mathbb{R}^d\right)$ is called α-symmetric (Eaton, 1981) if

$$\psi(t) = \int_{\mathbb{R}^d} \exp\left(\mathrm{i}\langle t, x\rangle\right) \,\mathrm{d}\, P(x) = \xi\left(\|t\|_\alpha\right) \qquad \text{for all } t \in \mathbb{R}^d$$

for some $\xi\colon \mathbb{R} \to \mathbb{R}$. For $X = (X_1, \ldots, X_d) \sim P$, these measures satisfy

$$\langle X, u\rangle \stackrel{d}{=} \|u\|_\alpha X_1 \qquad \text{for all } u \in \mathbb{S}^{d-1}.$$

For the depth of α-symmetric P

$$D\left(x; P\right) = \inf_{u \in \mathbb{S}^{d-1}} \mathrm{P}\left(\langle X, u\rangle \leq \langle x, u\rangle\right) = \inf_{u \in \mathbb{S}^{d-1}} \mathrm{P}\left(\|u\|_\alpha X_1 \leq \langle x, u\rangle\right)$$
$$= \mathrm{P}\left(X_1 \leq \inf_{u \in \mathbb{S}^{d-1}} \langle x, u\rangle \,/\, \|u\|_\alpha\right) = F_1\left(-\|x\|_\beta\right)$$

for β the conjugate index to α, and F_1 the c.d.f. of X_1.

Depth characterization: Proof II

Fix $\gamma \in (0,1)$ and take $\psi_\alpha(t) = \exp\left(-\|t\|_\alpha^\gamma\right)$ for $\gamma \leq \alpha \leq 1$. Then

- Measure P_α with characteristic function ψ_α exists (Lévy, 1937);
- The conjugate index to $\alpha \leq 1$ is $\beta = \infty$; and
- For the characteristic function of X_1 with $X \sim P_\alpha$ we have

$$\mathsf{E} \exp\left(\mathrm{i}\, t X_1\right) = \exp\left(-|t|^\gamma\right) \quad \text{for all } t \in \mathbb{R},$$

  i.e. F_1 does not depend on α.

All $P_\alpha \in \mathcal{P}\left(\mathbb{R}^d\right)$ have the same depth

$$D\left(x; P_\alpha\right) = F_1\left(-\|x\|_\infty\right) \quad \text{for all } x \in \mathbb{R}^d.$$

Depth characterization: Proof III

For $\gamma = 1/2$, the density of P_α with $\alpha = 1$ (left) and $\alpha = 1/2$ (right).