First steps in topological statistics

Joint work with D. Mond, J. Smith (Warwick)

99% monomial free

Motivation: "Can watching soccer on TV cause baldness?"

little $x \in$ People

$$X(x) = \begin{cases} 2: & x \text{ has full hair} \\ 1: & x \text{ has thin hair} \\ 0: & x \text{ is bald} \end{cases}$$

$S(x) = \#$hours of soccer x watches per week

$$Y(x) = \begin{cases} 2: & S(x) \geq 2 \\ 1: & 2 > S(x) \geq 1 \\ 0: & 1 > S(x) \geq 0 \end{cases}$$

$$P(X=i \text{ and } Y=j) \neq P(X=i) \cdot P(Y=j)$$

not independent, correlated

(soccer rays?)

"Explanation": $Z(x) = \begin{cases} 1: & x \text{ male} \\ 0: & x \text{ female} \end{cases}$

$$P(X=i \;\&\; Y=j \mid Z=k) = P(X=i \mid Z=k) \cdot P(Y=j \mid Z=k)$$

(*)
$$\begin{tikzcd} & Z \arrow[dl] \arrow[dr] & \\ X & & Y \end{tikzcd}$$

$$X \perp\!\!\!\perp Y \mid Z$$

$$P(Y|X) = P(Z|X) \cdot P(Y|Z)$$

$$P(Y|X)_{ij} = P(Y_j = j \;\&\; X=i) / P(X=i)$$

matrix equation

Stochastic matrices: • $A \geq 0$, i.e., $A_{ij} \geq 0 \;\; \forall i,j$

and • $A \cdot \mathbb{1} = \mathbb{1}$, i.e., $\sum_j A_{ij} = 1 \;\; \forall i$.

Study factorizations

$$\underset{n\times m}{A} = \underset{n\times r}{B} \cdot \underset{r\times m}{C}$$

with r as small as possible — $r = \operatorname{rank} A$

"stochastic factorizations"

$$SF(A) = \{ (B,C) \mid A = BC,\ B, C \text{ stochastic},\ r = \operatorname{rank} A \}$$

$\exists C$:

~~If $A = BC$ then $L(A) \subset L(B)$~~

$L(A) =$ linear space of A

$C(A) =$ positive cone of A

$$\exists C: \quad A = BC \iff L(A) = L(B)$$

$$\exists C \geq 0: \quad A = BC \iff C(A) = C(B)$$

$$SF(A) \longleftrightarrow C(A) \subset C(\underbrace{u_1, \dots, u_r}_{\text{rows of } B}) \subset L(A) \cap \mathbb{R}^n_+$$

$$\cap \{ \textstyle\sum x_i = 1 \}$$

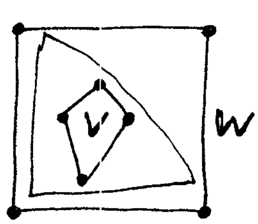
$$V \subset \Delta \subset W$$

so $\simeq \{ \Delta \mid V \subset \Delta \subset W \}$

Example: $\operatorname{rank}(A) = r = 3$

$V =$ p-gon in $\mathbb{R}^2$ $\quad p \leq m$

$W =$ q-gon in $\mathbb{R}^2$ $\quad q \leq n$

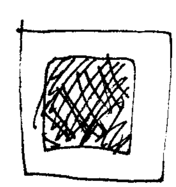


$\leftrightarrow$ factorization

$\Delta =$ triangle $\supset V$ and $\subset W$

If V is big, no factorization may be possible.

e.g.



van Straten

2

"Set of sandwiched simplices": for V, W, define

$$\Delta(V,W) := \{ \Delta \mid V \subseteq \Delta \subseteq W \}$$

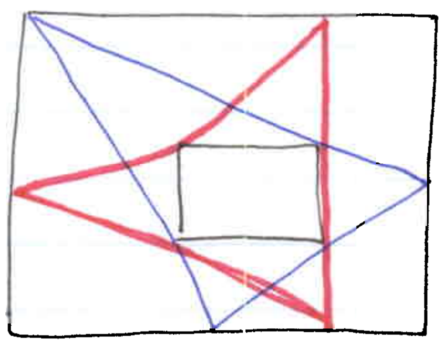
| Morse theory

Study this space — components, etc.

Theorem: If V is a p-gon and W is a q-gon, then the # of connected cpts of $\Delta(V,W)$ is $\leq p+q$.

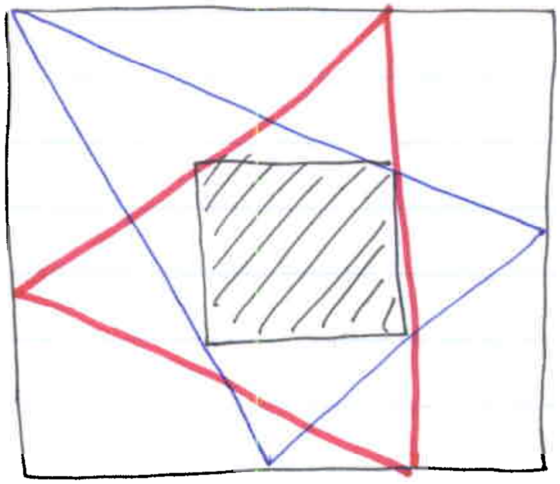
$$\text{this \# is } \begin{cases} \leq q & p=3 \\ \leq p & q=3 \\ 1 & p=q=3. \end{cases}$$

Example:



(oops) should be a triangle

8 components.

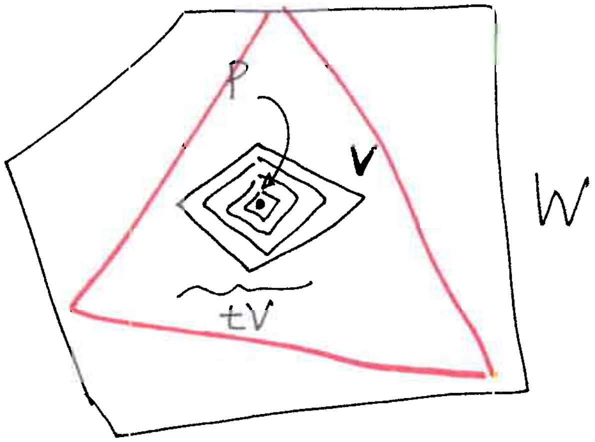


this is a bit more like it.

van Straten

3

Idea: See how $\Delta(V,W)$ changes if V grows from a point $p \in V$:



As t gets larger, at some point Δ gets forbidden.

Proposition: ① $\Delta(V,W) \simeq \Delta(V, \partial W)$
by radial projection

② $\mathring{\Delta}(P, \partial W) \simeq O(n)/S_{n+1}$ (as a topological space)

Function $f : \mathring{\Delta}(P, \partial W) \longrightarrow \mathbb{R}$

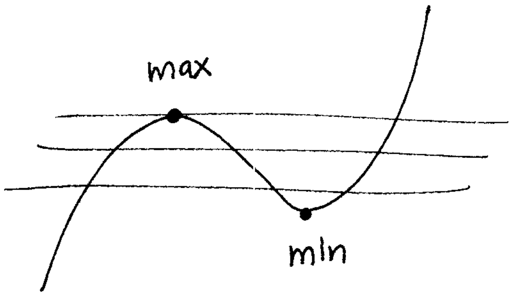
$$f(\Delta) = \min\{ t : \ tV \cap \partial\Delta \neq \emptyset \}$$

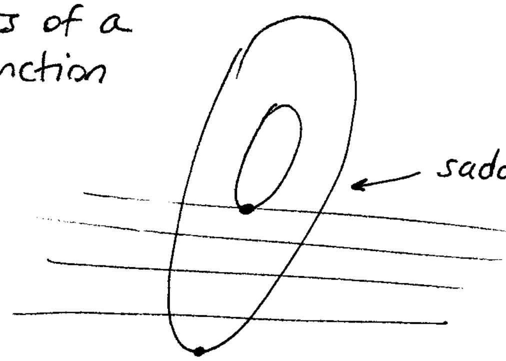
$$\boxed{\Delta(sV, W) = f^{-1}([s, \infty))}$$ $\Rightarrow$ do Morse theory.

Problems: • f not differentiable,
• space not smooth.

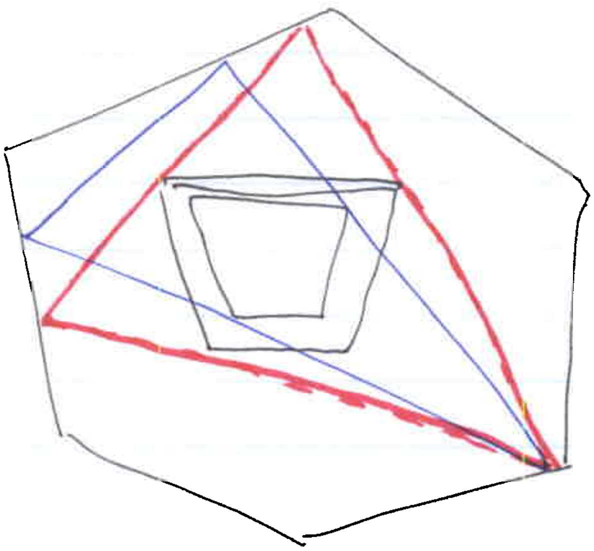
$f = \min\{f_1, \dots, f_n\}$ where $f_i = \min\{t \mid tV \cap \partial\Delta^i \neq \emptyset\}$

Morse theory: study topology of space by looking at level sets of a function





<u>Lipschitz category</u>: <u>critical points</u>

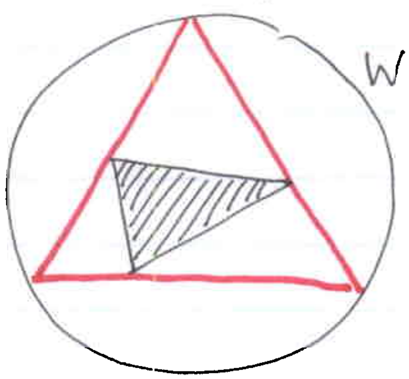


"endangered" triangle

$\Delta \to \Delta$ path in space of triangles

For critical points — need <u>tritangent</u> Δ.

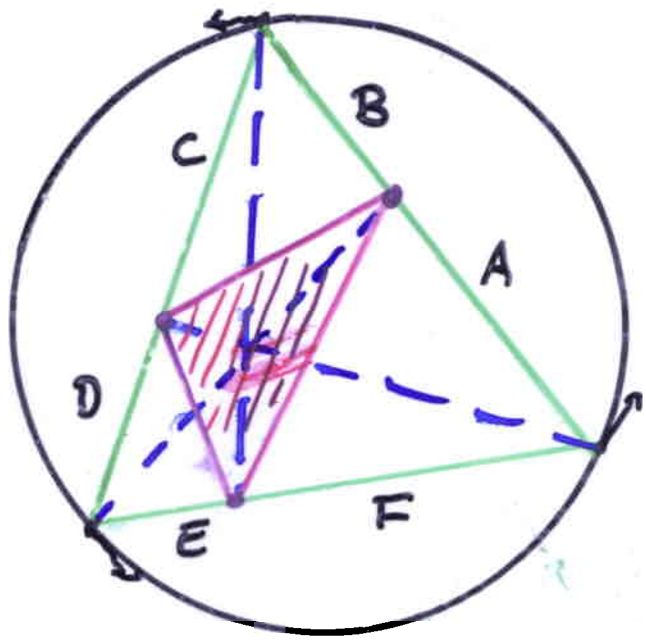
W = ~~area~~ disk, ∂W = circle



See transparencies: Cevian Triangle, etc.

<u>Theorem</u>: $H^q(\Delta(V,W);\mathbb{Z}) = 0$ for $q > n^2 - n - 1$.

Cevian Triangle

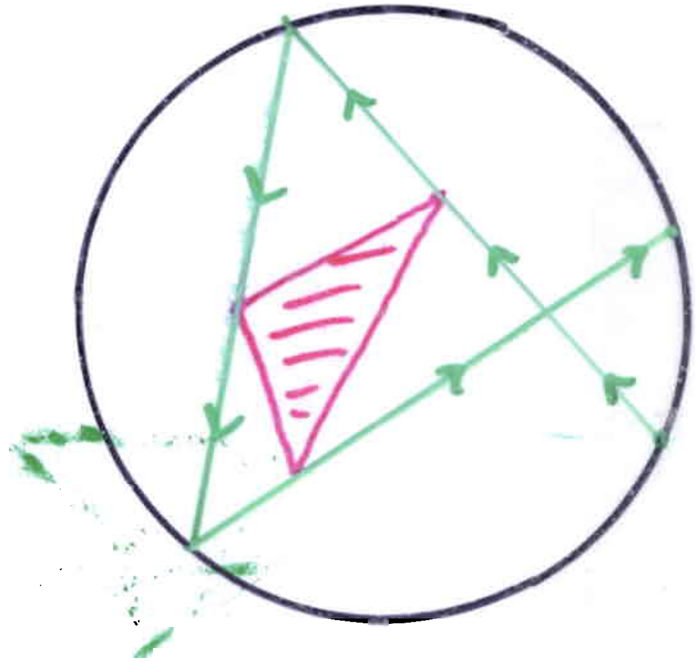


$$\frac{A}{B} \cdot \frac{C}{D} \cdot \frac{E}{F} = 1$$

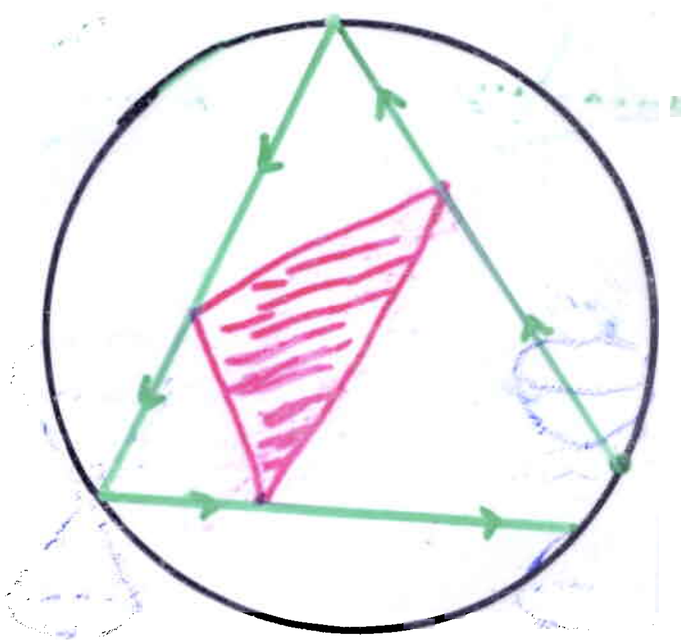
$$\Longleftrightarrow$$

$\triangle$ critical point of f.

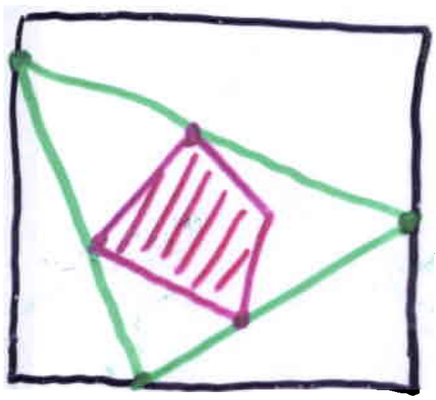
Slighty smaller.



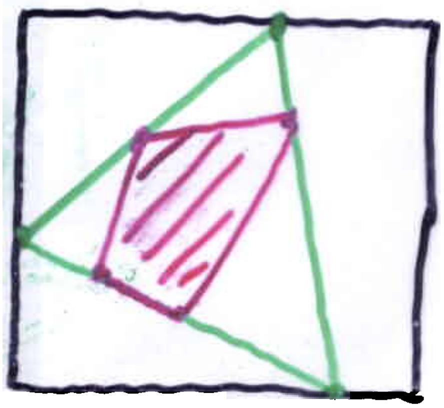
Slightly bigger.



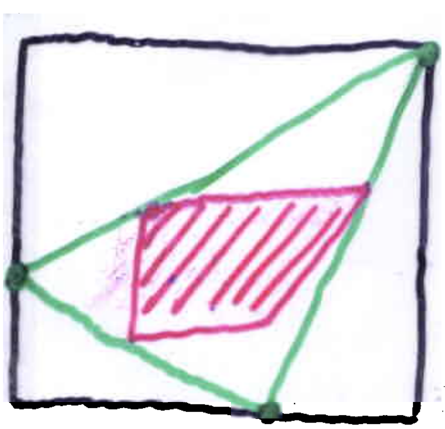
Critical Points.



Saddle



local Maximum



local Maximum.

if a "Cevian Condition" is satisfied.

Cevian Saddle.