トポロジーでデータ構造を解析する

佐伯 修
（九州大学 マス・フォア・インダストリ研究所）

第10回設計情報学研究会

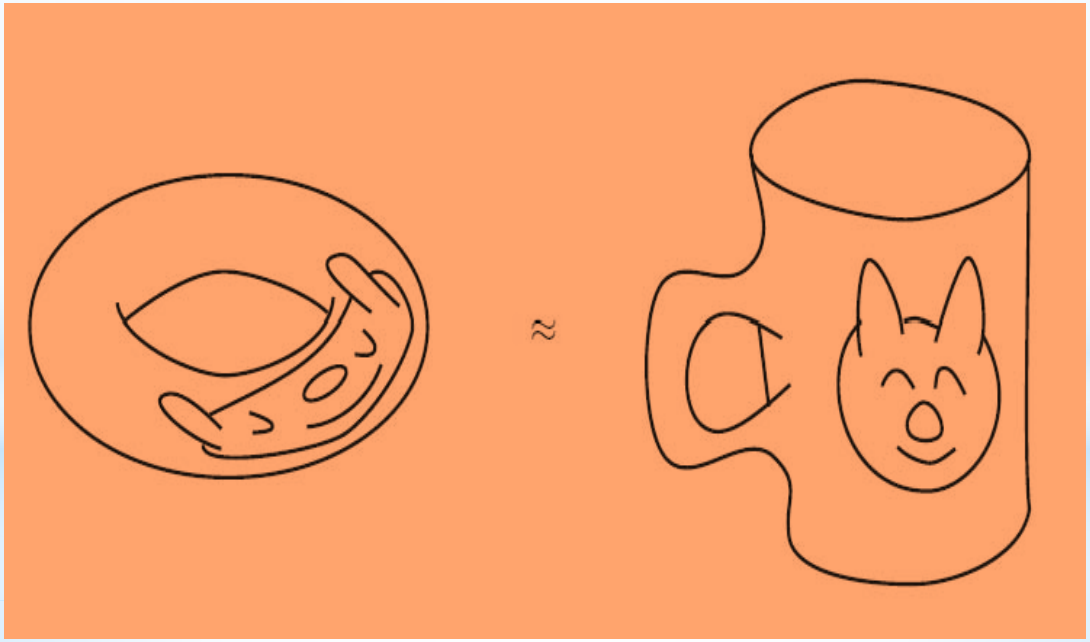
@KKRホテル熊本

2016年8月26日

トポロジーとは？

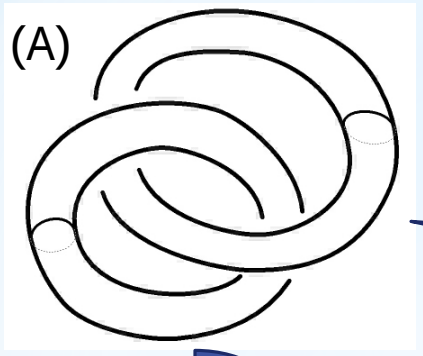
図形をゴムのようなものでできていると考え，
グニャグニャと**連続的に変形できるものは同じ**と思う，
とても**柔らかい幾何学**.

〔幾何学的対象について，連続的に変形しても変わらない性質を研究する，数学の一分野〕



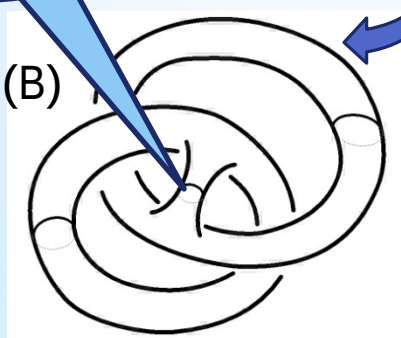
ドーナツとコーヒーカップはトポロジーでは同じ

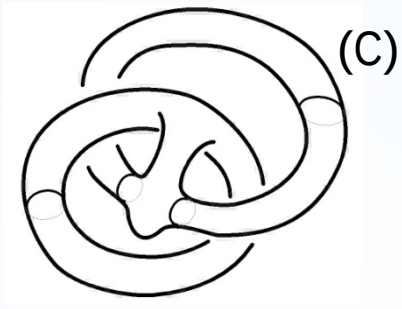
トポロジーのわざ

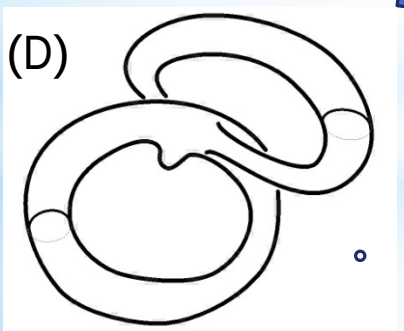


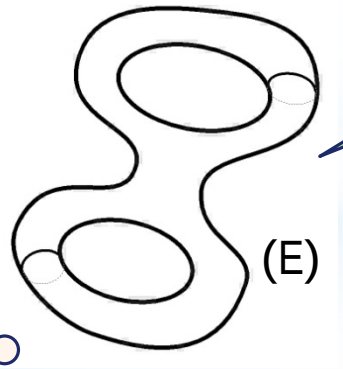
２本の輪は互いに
「絡んで」います

橋をかけると．．．









ほどけてしまいました！

トポロジーでは．．．　(A) ≠ (B) = (C) = (D) = (E)

トポロジーの発展

19世紀終わり頃に Henri Poincaré が創始した
と言われる．数学の中では比較的新しい分野.

Poincaré：“3次元多様体”のトポロジーに関する
「予想」を提出　⇒　**Poincaré予想**

20世紀:数学者は予想を解くために様々な“命がけ”の努力
⇒ トポロジーの大発展

（21世紀にPoincaré予想は Perelmanによって解決）

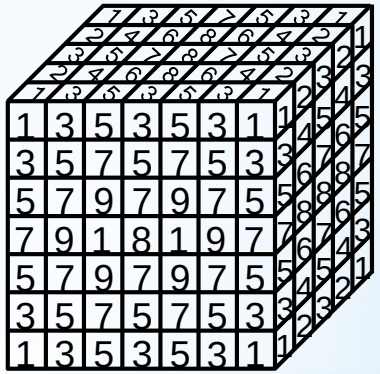
数学全般，諸科学分野への大きな影響を
与える基盤が整備された

データ解析とトポロジー

幾何構造を持った対象

↓ 実験による観測・シミュレーション

- 大量の離散データ
- 観測誤差・計算誤差



トポロジー的手法により,

- ☞ **大ざっぱ**に,
- ☞ **ロバスト**に,
- ☞ **大局的**に,

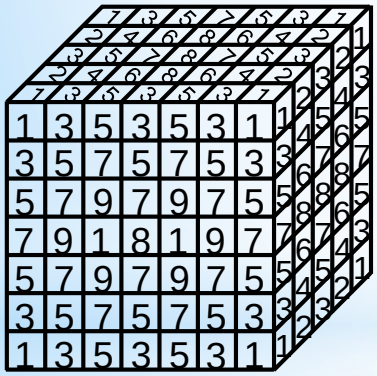
対象の特徴を抽出することが可能になる.

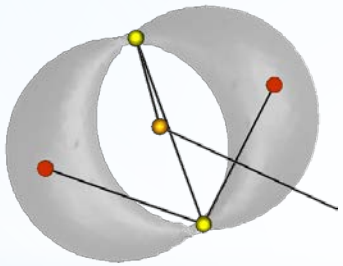
本日のテーマ：微分トポロジーを用いたデータ可視化

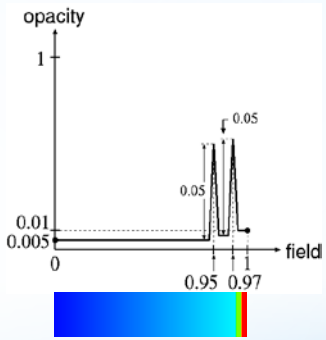
例：Direct Volume Rendering

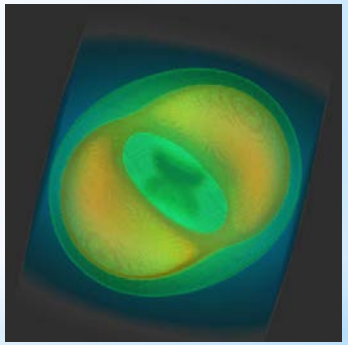
大規模データ，複雑な構造，ノイズ

3次元データ → トポロジー解析 → 伝達関数設計 → レンダリング









Reeb グラフ

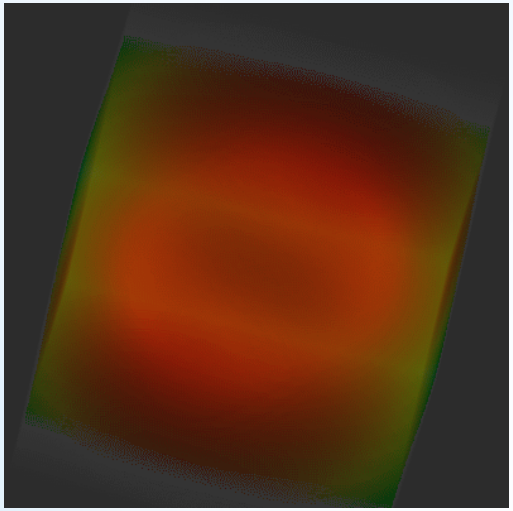
特異点論

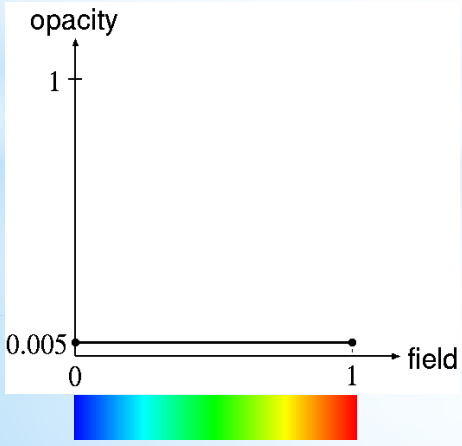
トポロジー的に意味のある曲面を抽出・強調

高橋 成雄 （会津大学 コンピュータ理工学部， IMI客員教授）

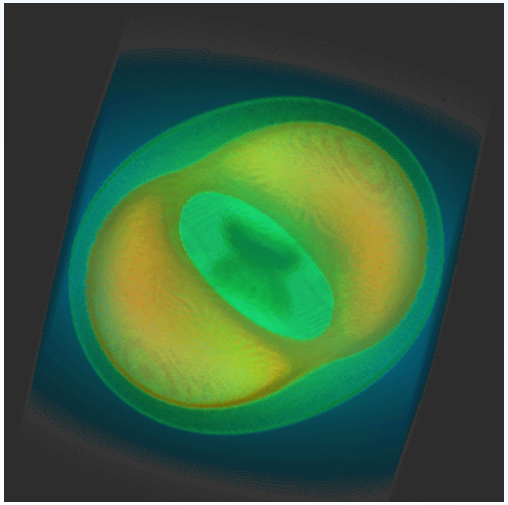
可視化の一例

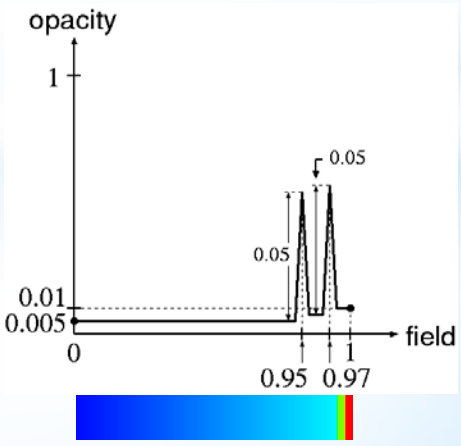
通常の可視化結果





特徴を強調した
可視化結果

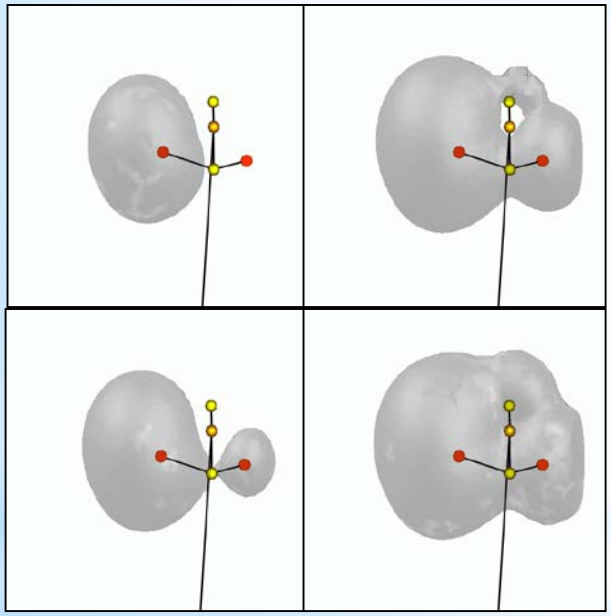




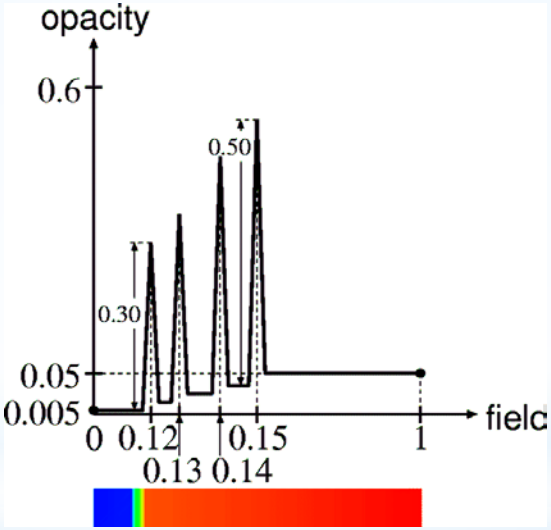
トポロジー的変化
が起きている面を
強調して可視化

例: 陽子-水素原子衝突

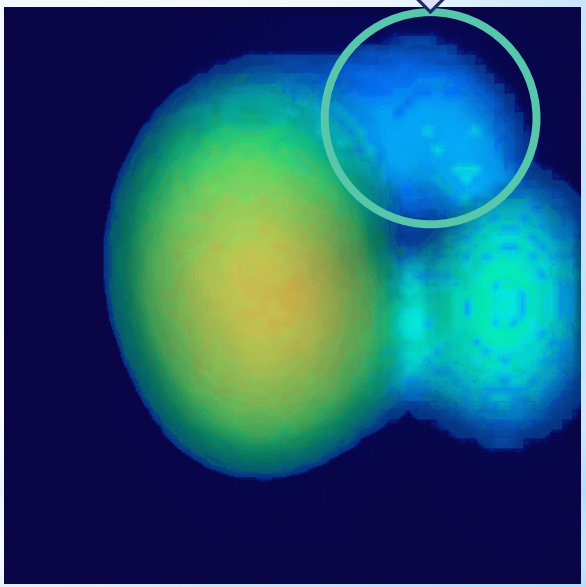
トポロジー的に強調された伝達関数により発見された現象



特徴的な「等値面」の抽出



伝達関数の設計



トポロジー可視化の結果

トポロジーでデータ構造を解析する (Part 2)

佐伯 修
（九州大学 マス・フォア・インダストリ研究所）
Joint work with
高橋成雄（会津大），櫻井大督（理研），
Hsiang-Yun Wu（慶応大），
Hamish Carr, David Duke (Univ. of Leeds),
山本卓宏（九州産業大）

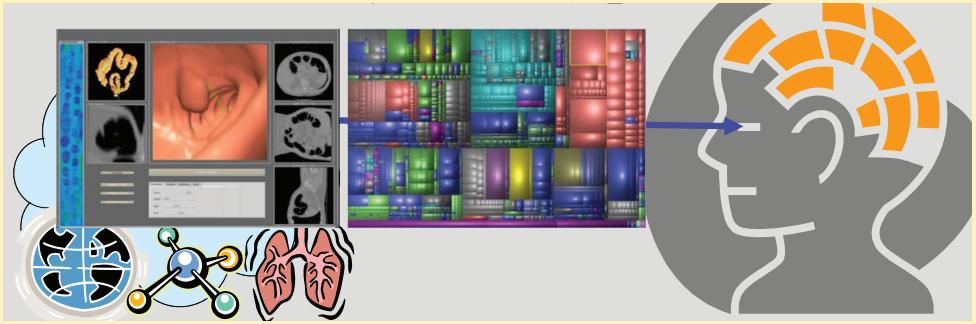
2016 年 8 月 26 日（金）　第 10 回設計情報学研究会
於 KKR ホテル熊本

§0. データ可視化

「可視化」とは？

複雑なデータの特徴を抽出して視覚的に理解できるようにする技術.
特徴抽出 & 視覚イメージで表現.

百聞は一見に如かず！



現代は**ビッグデータ**の時代. 時間とともに大量のデータが蓄積.
通常のデータベースツールでは解析が困難
⇒ **“可視化”** が重要.

トポロジー手法は近年における **one of the top ten breakthroughs** と考えられている.

§1. 単独データの可視化—関数の場合

レベル集合

N^n : n 次元の領域（もしくは「多様体」，通常は $n = 2, 3, 4$）
$f : N^n \to \mathbf{R}$　微分可能な**関数 (スカラー場)**

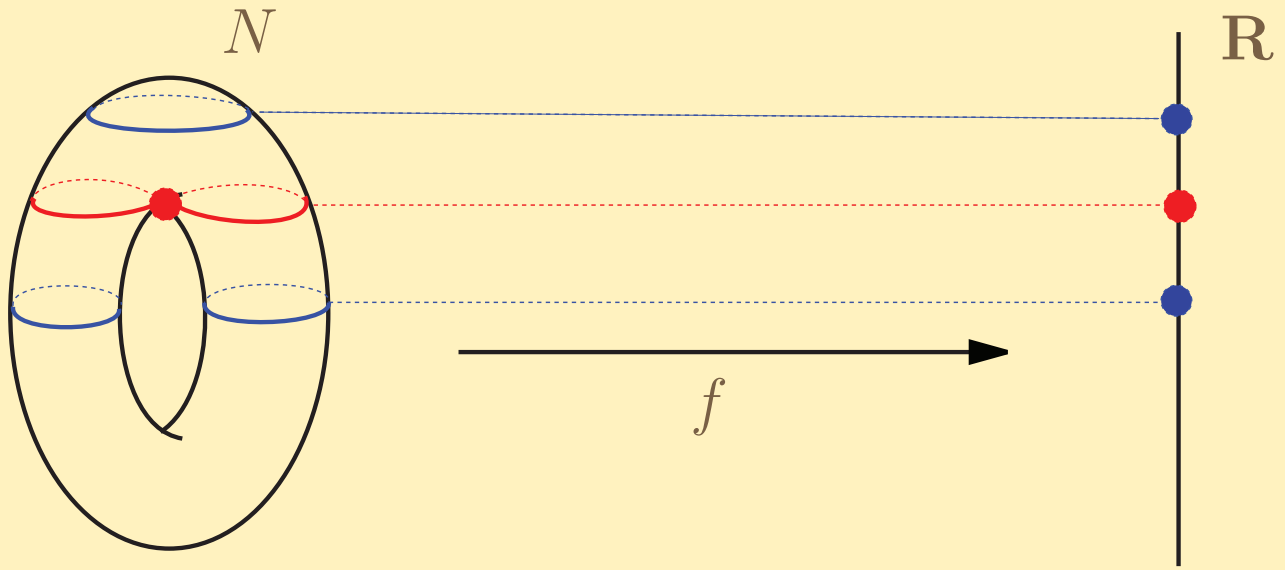
定義 1.1 $c \in \mathbf{R}$ に対して

$$f^{-1}(c) = \{p \in N^n \mid f(p) = c\}$$

を**レベル集合**という.

一般に，レベル集合は $n - 1$ 次元.
$n = 2$ なら等値線，$n = 3$ なら等値面，等.

レベル集合の例



レベル集合を追うことで，データの全体としての特徴がつかめる．
レベル集合のトポロジー変化が起こるところが大事．
$\Longleftrightarrow$ **臨界点**を含むレベル集合

臨界点：関数 f の１階の偏微分係数がすべてゼロとなる点

Morseの補題

$f : N^n \to \mathbf{R}$　微分可能な関数（スカラー場）

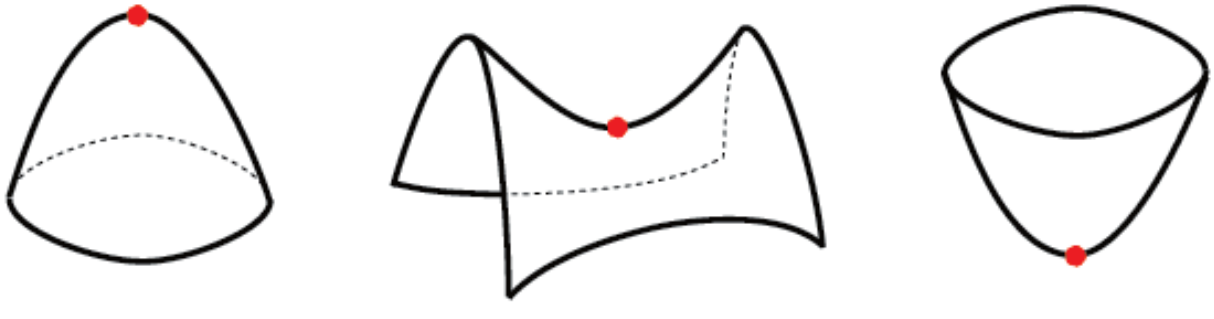
定理 1.2 (Morseの補題) f が十分に**一般的**であれば，各**臨界点**のまわりで局所座標をうまくとれば，f は以下のように記述される．

$$f = \pm x_1^2 \pm x_2^2 \pm \cdots \pm x_n^2 + c$$

（c は定数.）

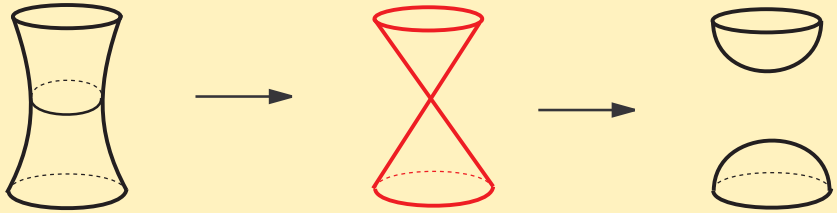
マイナス「$-$」の符号の個数を，臨界点の**指数**という．
臨界点のトポロジーは指数によって完全に記述される．
レベル集合の変化を調べるにはMorseの補題が本質的！

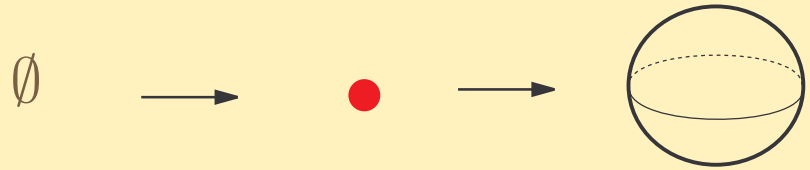
$n = 2$ のとき：一般的な関数に対し臨界点（等高線のトポロジーが変化する点）での等高線の変化は次の 3 種類のみ.



3 次元の場合

$f : N^3 \to \mathbf{R} \quad (\dim N^3 = 3)$
レベル集合は曲面（「特異点」を持つかも知れない）.

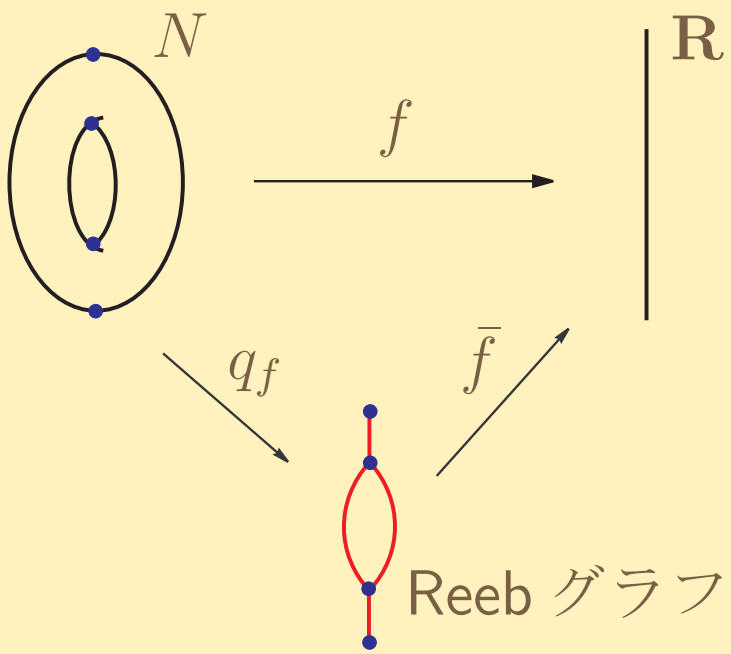




臨界点を含む等値面のまわりでの，等値面のトポロジー変化は，上の 2 種類と，左右を逆にした 2 種類の，合計 4 種類のみ.

Reeb グラフ

$f : N^n \to \mathbf{R}$　関数
各レベル集合の各**連結成分**を 1 点につぶして得られるグラフ
$\Longrightarrow$ **Reeb グラフ**という [Reeb, 1946]

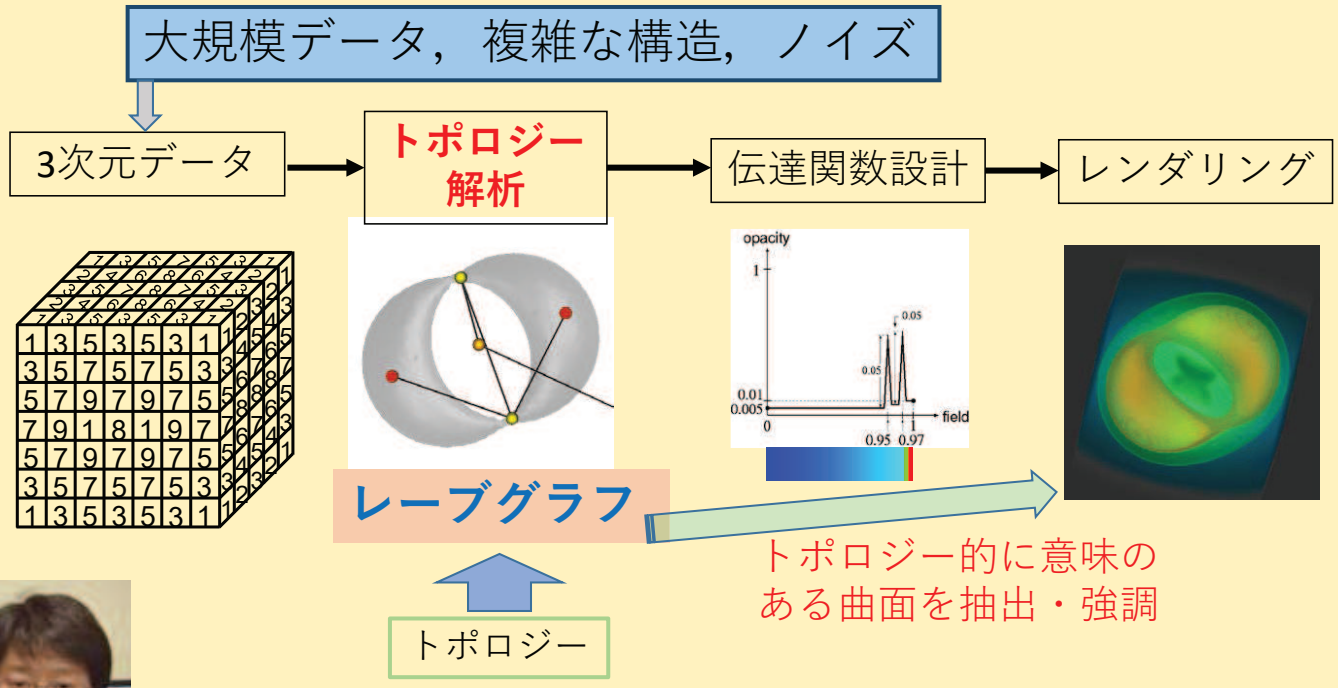


Reeb グラフの頂点 $\Longleftrightarrow$ 関数の臨界点
関数の可視化のためには Reeb グラフはなくてはならない.

3次元データ解析例

Reeb グラフの応用例 [Takahashi–Takeshima–Fujishiro, 2004] : **Topological Volume Skeletonization and its Application to Transfer Function Design**

例：Direct Volume Rendering



高橋 成雄 氏（会津大学 コンピュータ理工学部 コンピュータ理工学科 情報システム学部門， IMI客員教授）

§2. 複数データの同時可視化—写像の場合

単独のスカラー場ではなく，**複数のデータを同時に解析**する場合はどうしたら良いか？

（おそらく数理的困難のため）そうした**多値関数データ**の解析はつい最近始まったばかり．

N^n: n 次元の領域
$f : N^n \to \mathbf{R}^m \ (m \geq 1)$　微分可能な**写像** (or **多値関数場**)

$$f(x) = (f_1(x), f_2(x), \ldots, f_m(x))$$

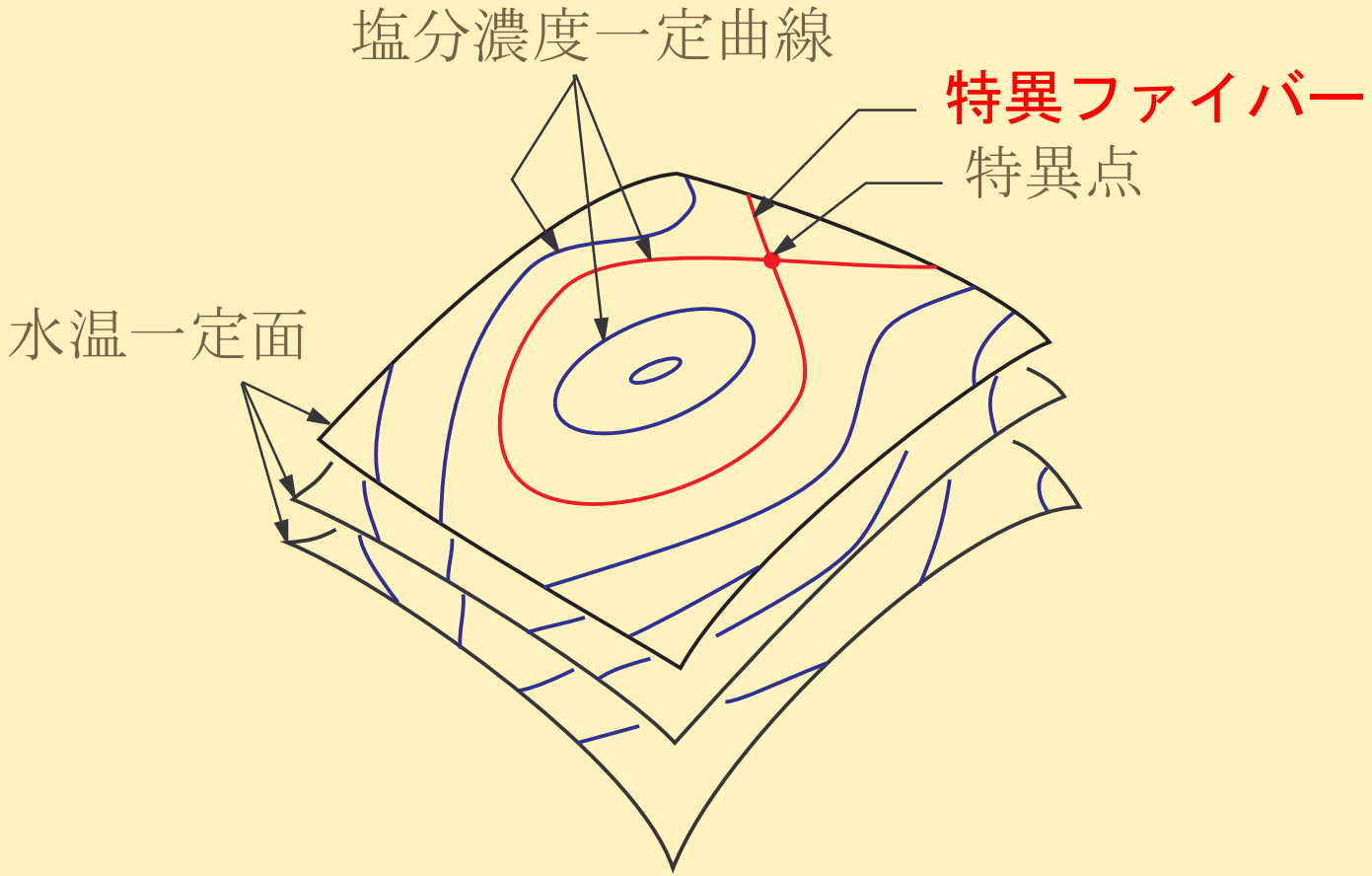
定義 2.1
$\boldsymbol{c} \in \mathbf{R}^m$ に対して，$f^{-1}(\boldsymbol{c})$ は**ファイバー**と呼ばれる．

一般にファイバーは $n - m$ 次元．以下 $n \geq m$ を仮定する．

多値関数データの可視化は，**ファイバーのトポロジー変化**を追うことで可能となる（はず）．

ファイバーの例

$n = 3$, N^3 : 海水, $f : N^3 \to \mathbf{R}^2$
$f =$ (水温, 塩分濃度)



特異点を含むファイバーを**特異ファイバー**と呼ぶ.

数学的定式化

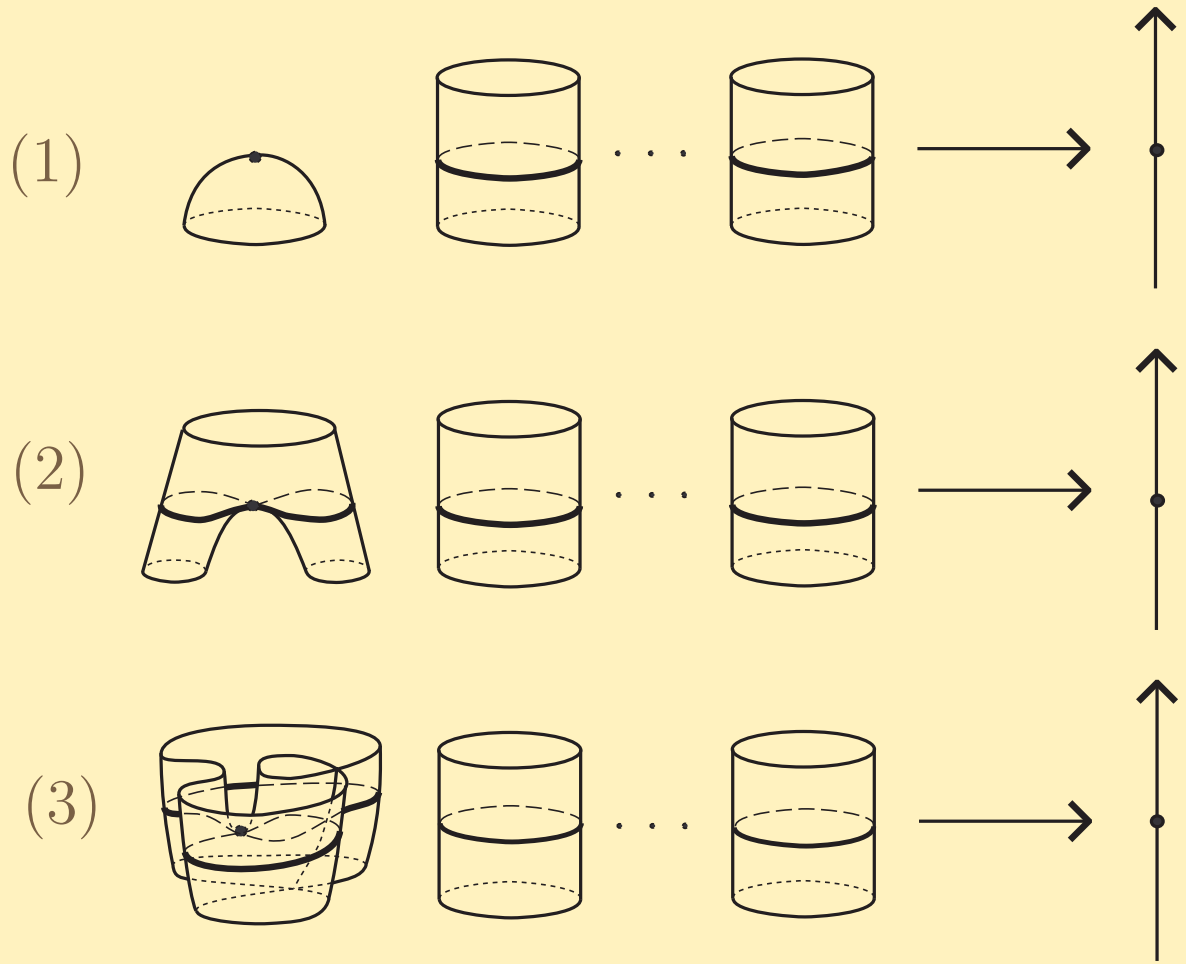
数学的には，**ファイバー**を，逆像に沿った**写像芽**として定式化.

定義 2.2 $f : N^n \to \mathbf{R}_1^m$, $g : L^n \to \mathbf{R}_2^m$ 微分可能写像
$\boldsymbol{c} \in \mathbf{R}_1^m$ と $\boldsymbol{d} \in \mathbf{R}_2^m$ に対して，$\boldsymbol{c}$ 上のファイバーと $\boldsymbol{d}$ 上のファイバーが**同値**（あるいは，2 点が同じ**ファイバー型**を持つ）
$\Longleftrightarrow$ $\boldsymbol{c}$ の近傍 $U \subset \mathbf{R}_1^m$，$\boldsymbol{d}$ の近傍 $V \subset \mathbf{R}_2^m$ があって，次の可換図式が成り立つ：

$$\begin{array}{ccc} (f^{-1}(U), f^{-1}(\boldsymbol{c})) & \xrightarrow{\cong} & (g^{-1}(V), g^{-1}(\boldsymbol{d})) \\ \Big\downarrow{\scriptstyle f} & & \Big\downarrow{\scriptstyle g} \\ (U, \boldsymbol{c}) & \xrightarrow{\cong} & (V, \boldsymbol{d}) \end{array}$$

キーポイント. ファイバーが同値なら，近くのファイバー達も全部同値！「ファイバー型」がわかれば，近くでのファイバーのトポロジー変化が完全にわかる.

ファイバーの分類例



曲面上の Morse 関数の特異ファイバー

3 次元 2 値関数の場合

N^3: 3 次元領域（滑らかな**境界**を持つと仮定）
$f : N^3 \to \mathbf{R}^2$　**一般的な写像**（より正確には C^∞ **安定写像**）
f は局所的に以下のように記述される（Morse の補題の一般化）.

(a)　領域 N^3 の内部 $\mathrm{Int}\, N^3$ の点 p のまわり：

$$(x,y,z) \mapsto \begin{cases} (x,y), & p\text{: 正則点,} \\ (x,y^2+z^2), & p\text{: 定値折り目,} \\ (x,y^2-z^2), & p\text{: 不定値折り目,} \\ (x,y^3+xy-z^2), & p\text{: カスプ.} \end{cases}$$

境界に沿った特異点

(b) N^3 の境界 ∂N^3 の点 p のまわり（以下では，$\mathrm{Int}\, N^3 \leftrightarrow \{z > 0\}$ で，$\partial N^3 \leftrightarrow \{z = 0\}$ となるような局所座標を考える）：

$$(x, y, z) \mapsto \begin{cases} (x, y), & p\text{: } f|_{\partial N^3} \text{ の}\textbf{正則点}, \\ (x, y^2 + z), & p\text{: }\textbf{境界定値折り目}, \\ (x, y^2 - z), & p\text{: }\textbf{境界不定値折り目}, \\ (x, y^3 + xy + z), & p\text{: }\textbf{境界カスプ}. \end{cases}$$

または

$$(x, y, z) \mapsto (x, y^2 + xz \pm z^2), \qquad p\text{: } B_2 \textbf{ 点}.$$

特異ファイバーの分類

定理 2.3 $f : N^3 \to \mathbf{R}^2$ を C^∞ 安定写像とする.
f の各**特異ファイバー**は（正則ファイバー成分である円周や線分部分を除けば）以下のいずれかと同値.

(1) $\widetilde{\mathrm{bI}}^{\mu}$ $(2 \leq \mu \leq 10)$ （*List 1*, 余次元 $\kappa = 1$）,

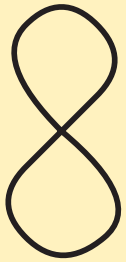
(2) 非連結なファイバー $\widetilde{\mathrm{bII}}^{\mu,\nu} = \widetilde{\mathrm{bI}}^{\mu} \sqcup \widetilde{\mathrm{bI}}^{\nu}$ $(2 \leq \mu \leq \nu \leq 10)$（余次元 $\kappa = 2$）,

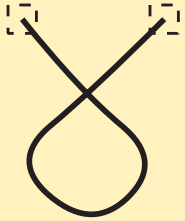
(3) *List 2 & 3*（余次元 $\kappa = 2$）.

特異ファイバー, List 1

$\kappa = 1$

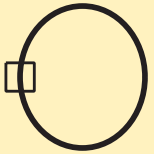
$\widetilde{\mathrm{bI}}^2$

$\widetilde{\mathrm{bI}}^3$ 

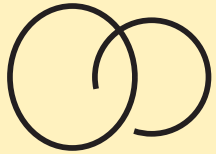
$\widetilde{\mathrm{bI}}^4$ 

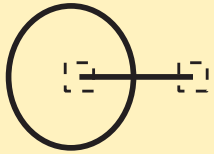
$\widetilde{\mathrm{bI}}^5$

$\widetilde{\mathrm{bI}}^6$

$\widetilde{\mathrm{bI}}^7$ 

$\widetilde{\mathrm{bI}}^8$

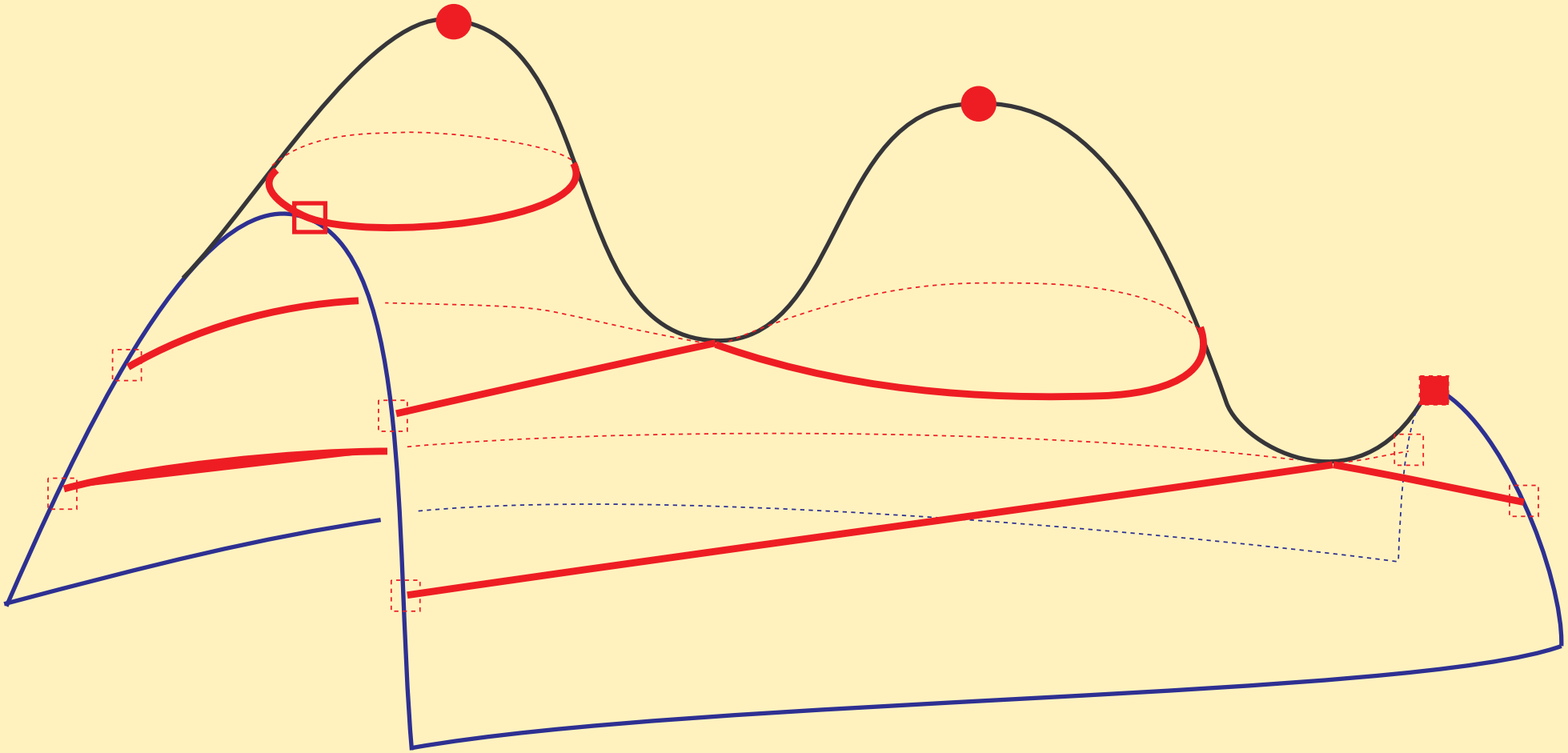
$\widetilde{\mathrm{bI}}^9$ 

$\widetilde{\mathrm{bI}}^{10}$ 

■ □ N の境界との接点

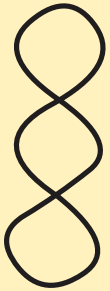
□ N の境界との横断的交点

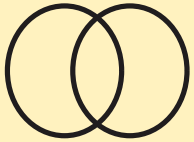
余次元１特異ファイバー

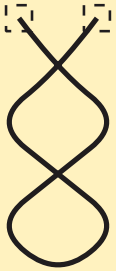


特異ファイバー, List 2

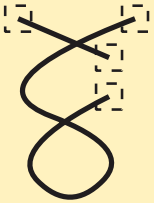
$\kappa = 2$

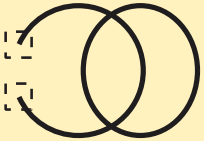
$\widetilde{\mathrm{bII}}^{11}$ 

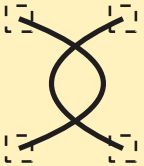
$\widetilde{\mathrm{bII}}^{12}$ 

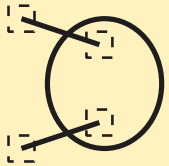
$\widetilde{\mathrm{bII}}^{13}$ 

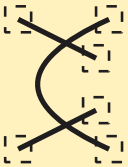
$\widetilde{\mathrm{bII}}^{14}$

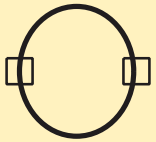
$\widetilde{\mathrm{bII}}^{15}$ 

$\widetilde{\mathrm{bII}}^{16}$ 

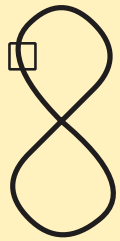
$\widetilde{\mathrm{bII}}^{17}$ 

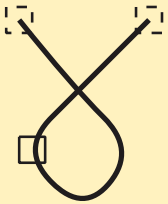
$\widetilde{\mathrm{bII}}^{18}$ 

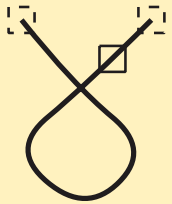
$\widetilde{\mathrm{bII}}^{19}$ 

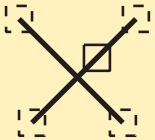
$\widetilde{\mathrm{bII}}^{20}$ 

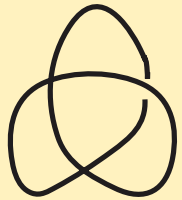
$\widetilde{\mathrm{bII}}^{21}$

$\widetilde{\mathrm{bII}}^{22}$ 

$\widetilde{\mathrm{bII}}^{23}$ 

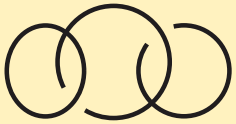
$\widetilde{\mathrm{bII}}^{24}$ 

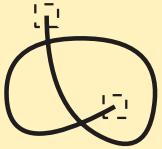
$\widetilde{\mathrm{bII}}^{25}$ 

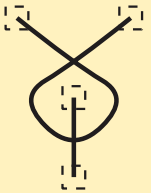
$\widetilde{\mathrm{bII}}^{26}$ 

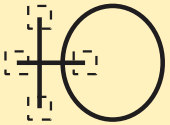
特異ファイバー, List 3

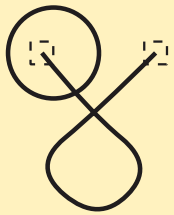
$\widetilde{\mathrm{bII}}^{27}$

$\widetilde{\mathrm{bII}}^{28}$ 

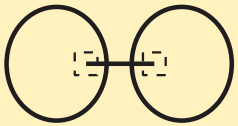
$\widetilde{\mathrm{bII}}^{29}$ 

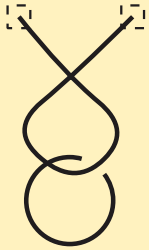
$\widetilde{\mathrm{bII}}^{30}$ 

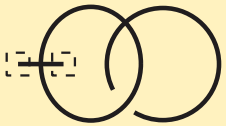
$\widetilde{\mathrm{bII}}^{31}$ 

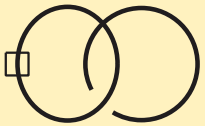
$\widetilde{\mathrm{bII}}^{32}$ 

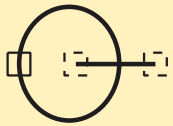
$\widetilde{\mathrm{bII}}^{33}$

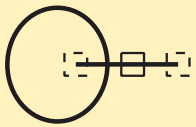
$\widetilde{\mathrm{bII}}^{34}$ 

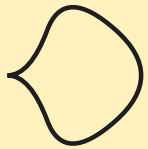
$\widetilde{\mathrm{bII}}^{35}$ 

$\widetilde{\mathrm{bII}}^{36}$ 

$\widetilde{\mathrm{bII}}^{37}$ 

$\widetilde{\mathrm{bII}}^{38}$ 

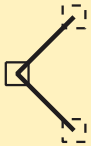
$\widetilde{\mathrm{bII}}^{39}$ 

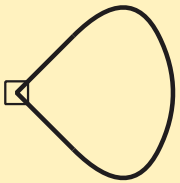
$\widetilde{\mathrm{bII}}^{a}$ 

$\widetilde{\mathrm{bII}}^{b}$

$\widetilde{\mathrm{bII}}^{c}$

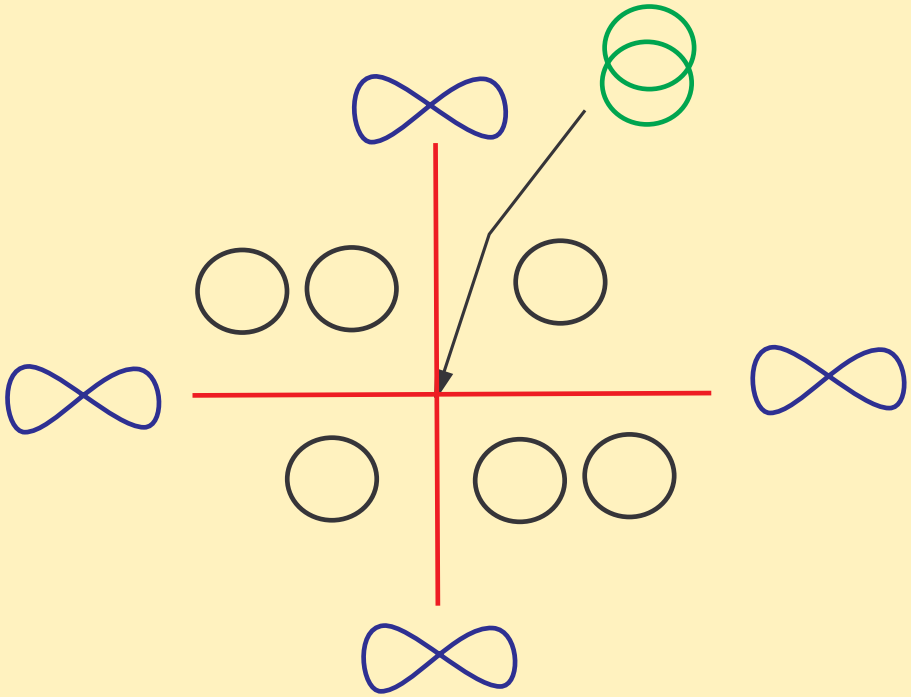
$\widetilde{\mathrm{bII}}^{d}$

$\widetilde{\mathrm{bII}}^{e}$ 

$\widetilde{\mathrm{bII}}^{f}$ 

ファイバー変化の例

以下，f の**特異点**（1 階の偏微分係数からなるヤコビ行列の rank < 2 となる点）からなる集合を $S(f)$ $(\subset N^3)$ と書く．



$f(S(f))$ と $\mathbf{R}^2 \setminus f(S(f))$ の各領域上のファイバー達

各**特異ファイバー**に対して上のような図が描ける．

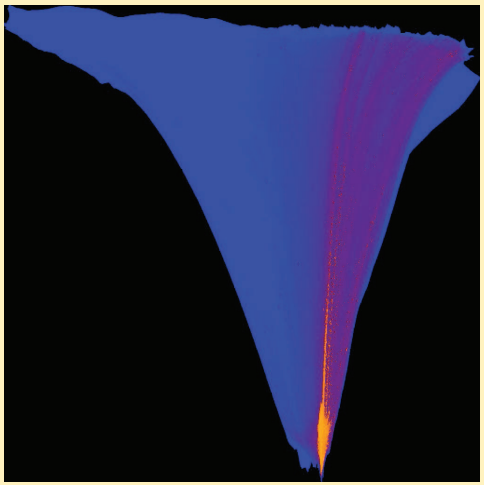
$\Longrightarrow$ **ファイバーの位相的変化**を追うことが可能に！

§3 データ可視化

既存の手法

以下，写像（2 値関数）$f : N^3 \to \mathbf{R}^2$ のサンプルデータが，3 次元領域 N^3 内の**離散集合**で与えられているものとする．

[Bachthaler & Weiskopf, 2008] **Continuous Scatterplots**:
$f(S(f))$ がぼんやりと観察できる．



The “Hurricane Isabel” data set was produced by the Weather Research and Forecast (WRF) model, courtesy of NCAR and the U.S. National Science Foundation (NSF).

上の図は，国立研究開発法人 理化学研究所 計算科学研究機構可視化技術研究チーム 提供．

既存の手法（続）

[Lehmann & Theisel, 2010]
Discontinuities in Continuous Scatterplots:

これらの手法は，数学的理論を本質的には使っていなかった.

可視化のために

多値関数データの可視化のためには以下が必要：

1. 領域内の特異点集合 $S(f)$ の同定，各特異点の**特異点型**の同定.
2. 特異点集合像 $f(S(f))$ の同定，各ファイバーの**ファイバー型**の同定.

[Edelsbrunner & Harer, 2002]
Jacobi Sets of Multiple Morse Functions:
特異点集合（Jacobi 集合とも呼ばれる）$S(f)$ を得るためのアルゴリズムを提唱.
しかし特異点型の同定はできなかった.

我々の新手法

多値関数データの解析のため，**Joint Contour Net** (= JCN) の概念を用いる [Carr & Duke, 2013].
Range を，似たような振る舞いをする領域に効率的に分割する手法.
– New Technology in Compter Science –

$f = (f_1, f_2) : N^3 \to \mathbf{R}^2$ を考える.
サンプルデータが，3 次元領域 N^3 内の離散集合で与えられているものとする.

JCN の「肝」は，f の値を**量子化**すること.

点 $\boldsymbol{c} \in \mathbf{R}^2$ の代わりに，小さなピクセル $P \subset \mathbf{R}^2$ を考える.
逆像 $f^{-1}(\boldsymbol{c})$ の代わりに，**太ったファイバー** $f^{-1}(P)$ を考える.

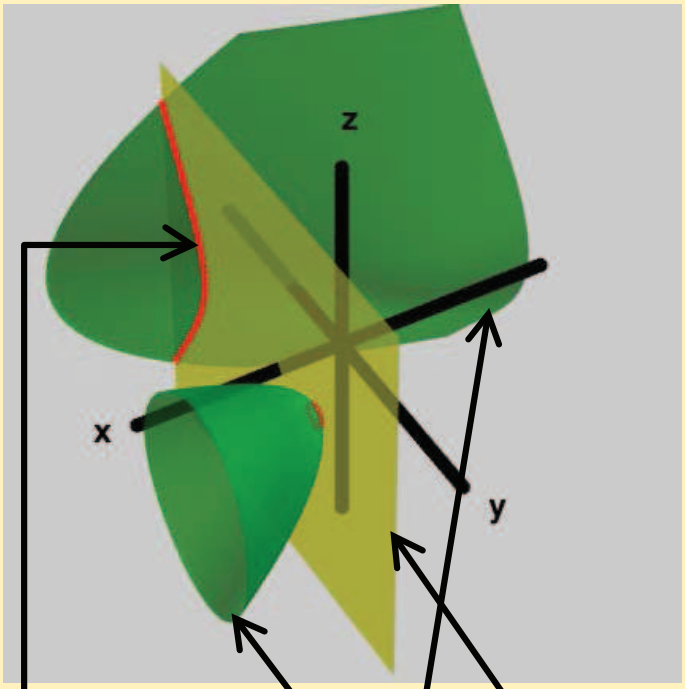
こうして特異ファイバーがロバストに同定可能となる.

サンプル：多項式写像 (1)

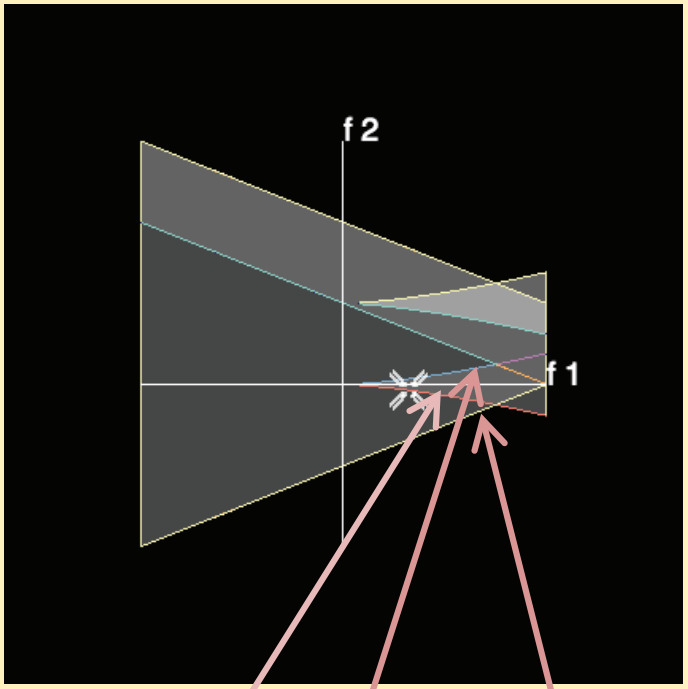
$$f(x,y,z) = (x, y^3 - xy + z^2)$$

ファイバーの生成消滅が観察できる.

3次元領域



Range



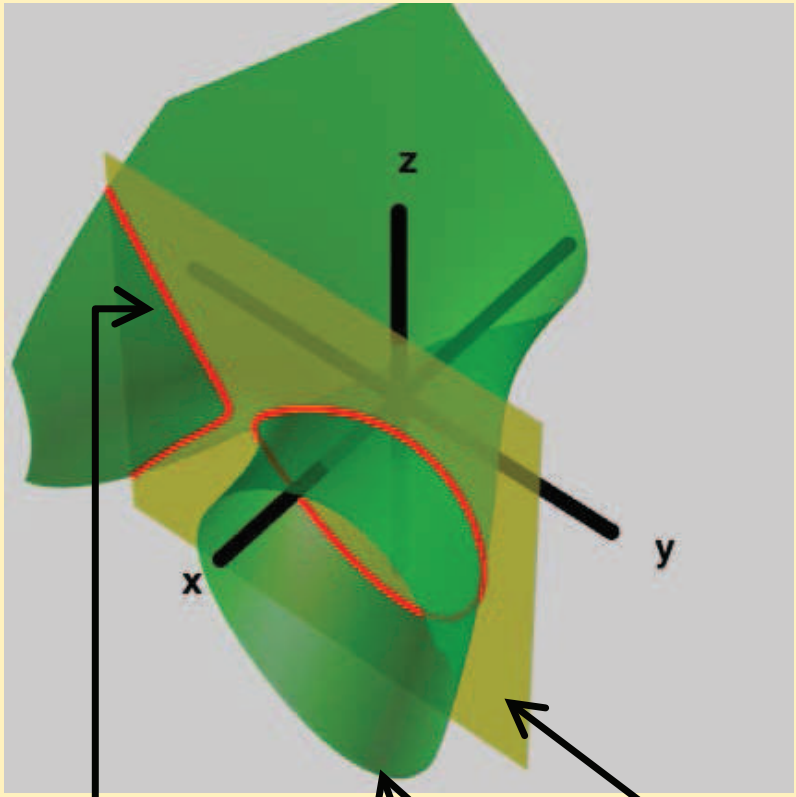
$f^{-1}(3\varepsilon^2, -2\varepsilon^3)$

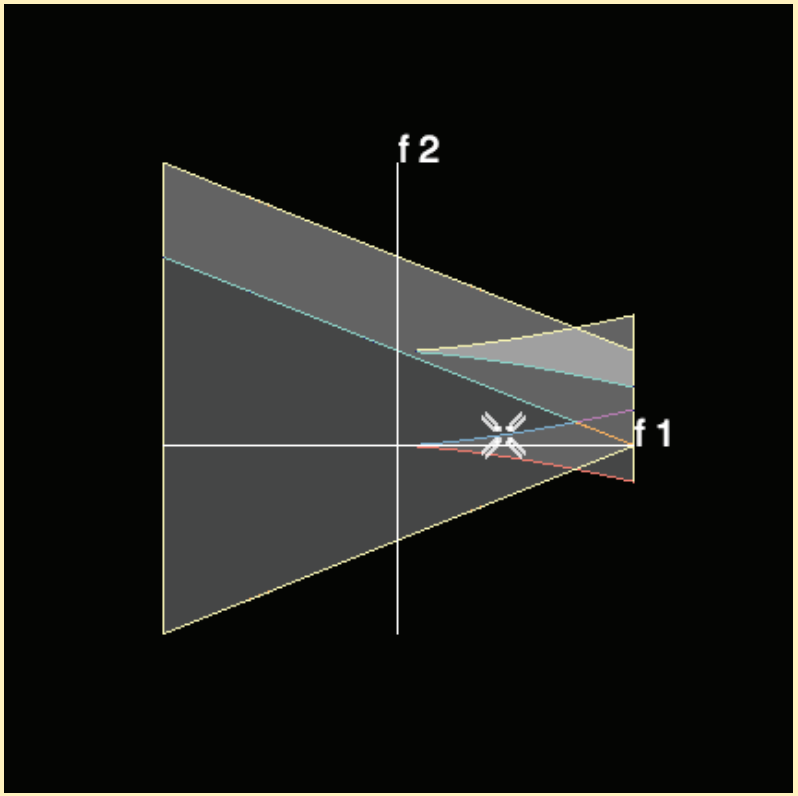
$x = 3\varepsilon^2$

$y^3 - xy + z^2 = -2\varepsilon^3$

Sample: サンプル：多項式写像 (2)

同じ写像 $f(x, y, z) = (x, y^3 - xy + z^2)$
ファイバーの併合・分離が観察できる.





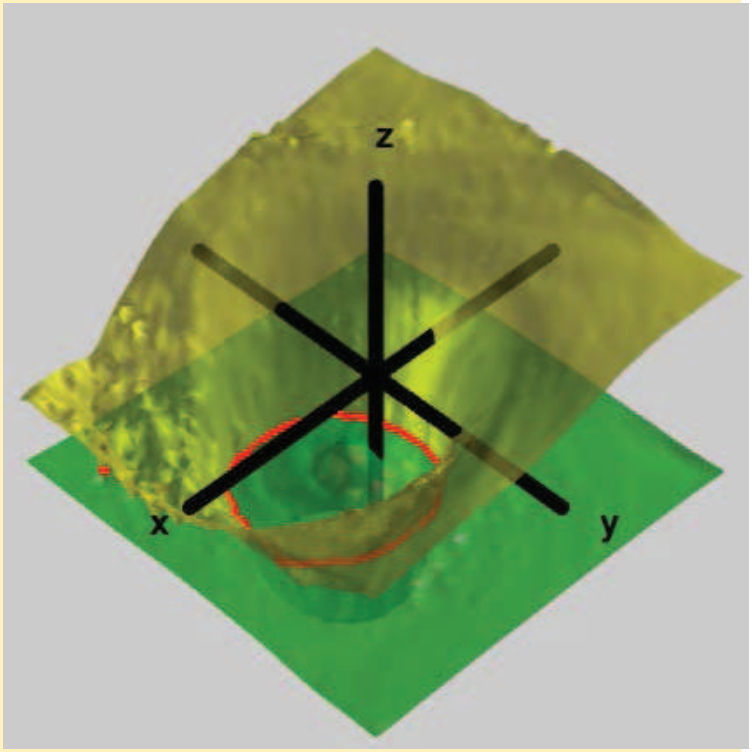
$f^{-1}(3\varepsilon^2, 2\varepsilon^3)$

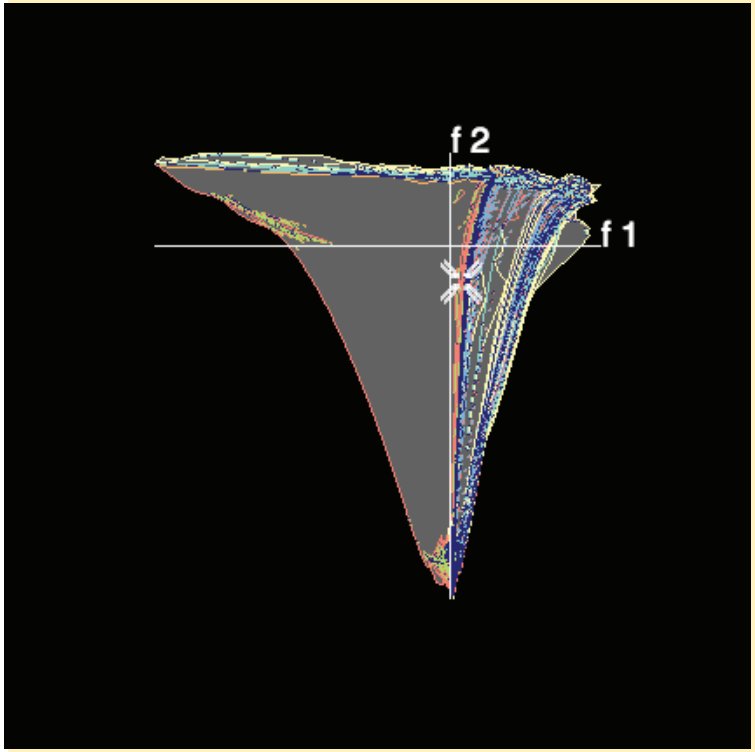
$x = 3\varepsilon^2$

$y^3 - xy + z^2 = 2\varepsilon^3$

Hurricane Isabel data (1)

Volume data for the Hurricane Isabel
f = (Pressure, Temperature)

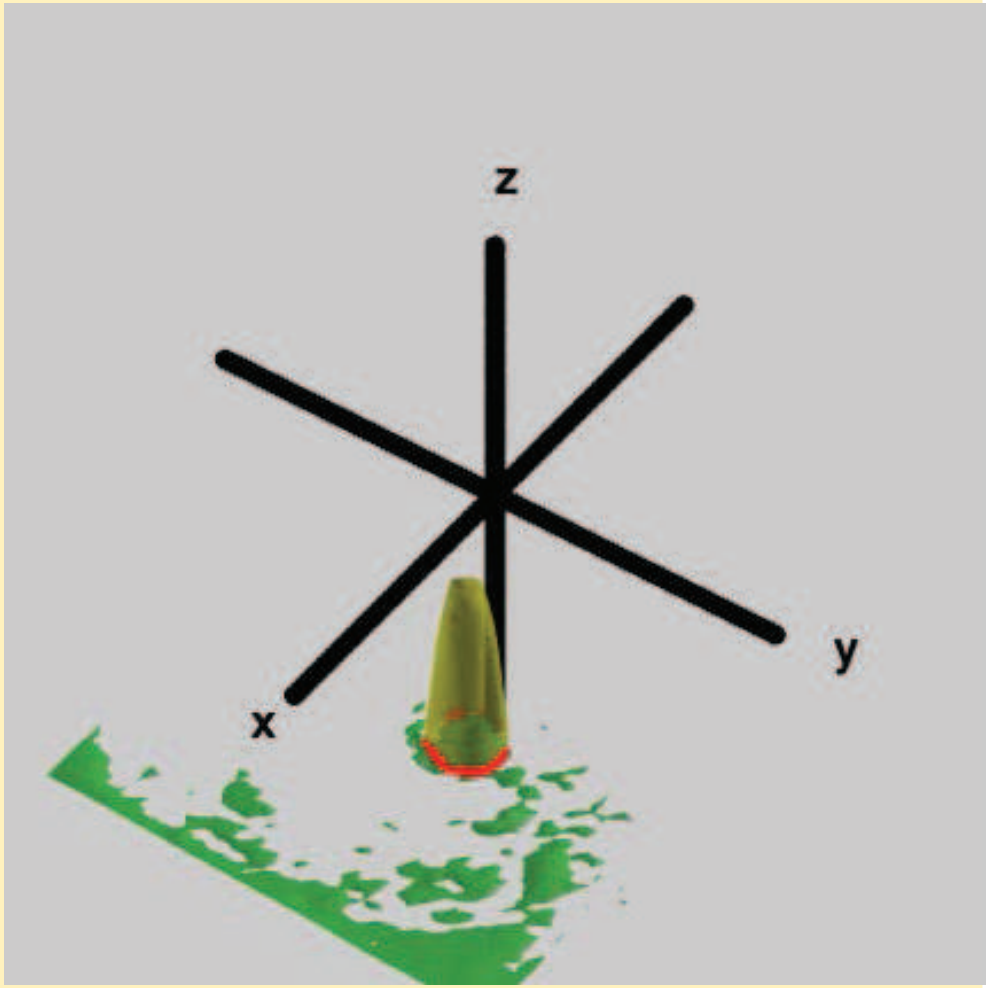


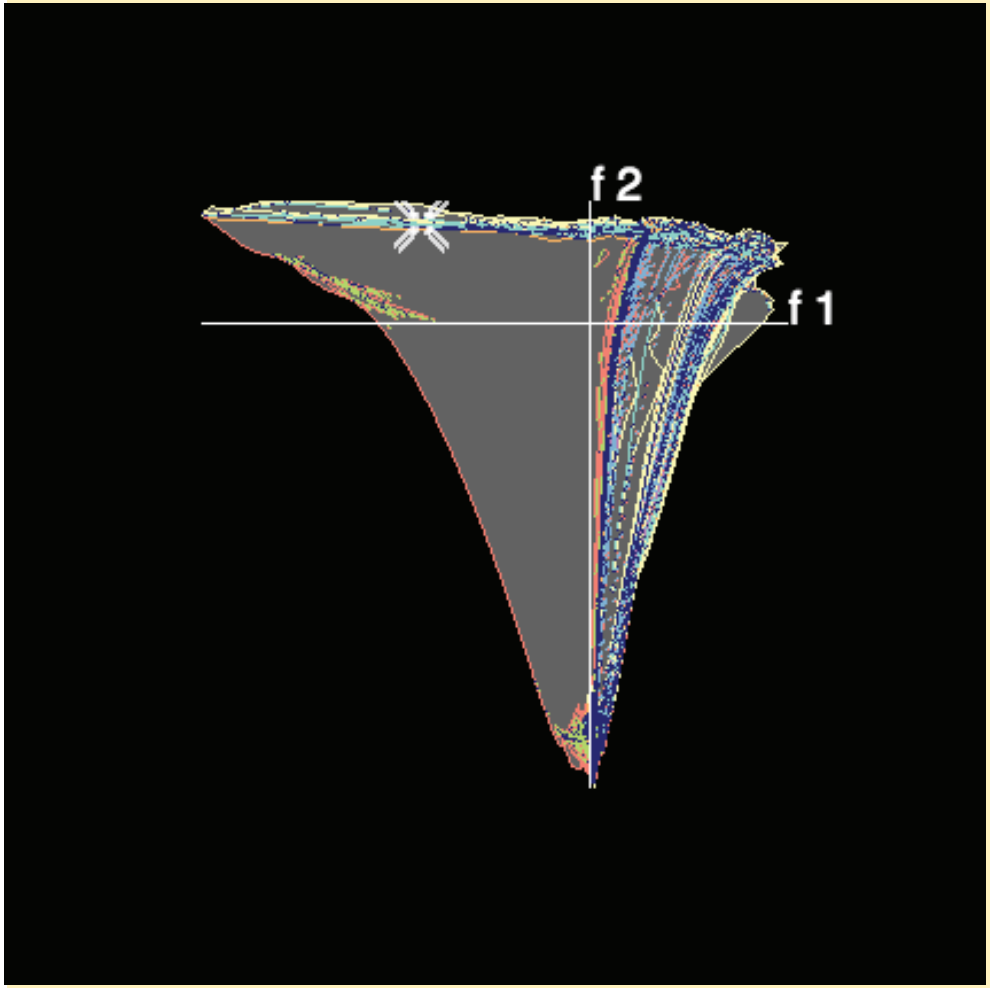


"Hurricane Isabel" data: produced by the Weather Research and Forecast (WRF) model, courtesy of NCAR and the U.S. National Science Foundation (NSF).

右の「×」印が左のファイバーに対応.

Hurricane Isabel data (2)





まとめ

微分可能写像の**特異点理論**を使うことで,

- 一般的に表れる**特異点型**や**ファイバー型**を**リストアップ**することができる.
- これにより，**特異点**と**ファイバー**，そしてそれらの型を**同定**できる.

これにより科学的データの可視化に大きく貢献できる.

実は，こうした可視化手法は数学における特異点理論の研究も大いに助ける！

Data Visualization

Singularity Theory