

可微分写像の特異点論とデータ可視化

佐伯 修

九州大学マス・フォア・インダストリ研究所

1 序文

一般に、科学的なシミュレーションや実験により得られるデータは、ユークリッド空間の間の写像 $f: \mathbb{R}^n \rightarrow \mathbb{R}^p$ の離散サンプル点の集まりとして定式化できることが多い。本稿ではこうしたデータ、特に大規模なデータの特徴を、微分位相幾何学に基づいて解析し、データの可視化に役立つ技術について、特にそのために必要となる、可微分写像の特異点論における基礎的な事項を中心に解説する。

なお本文章の内容の一部は、高橋成雄氏（東京大学大学院新領域創成科学研究科）との共同研究の結果であり、一部の資料は高橋氏からご提供いただいたものである。高橋氏にはこの場を借りて感謝申し上げたい。

以下、 M を n 次元 C^∞ 級多様体、 N を p 次元 C^∞ 級多様体とし、 $n \geq p \geq 1$ と仮定する。多様体論に不慣れな読者は、 M, N はそれぞれ $\mathbb{R}^n, \mathbb{R}^p$ の開集合であると思って読み進めていただいても問題ない。そして、 $f: M \rightarrow N$ を可微分写像（より正確には C^∞ 級、すなわち無限回微分可能な写像）とする。

2 関数とレベル集合

序文のような写像のデータ解析については、まずスカラー関数の場合、すなわち $p = 1$ で $N = \mathbb{R}$ の場合が非常に良く研究されている。実際この場合は、技術的にあまり複雑なことは要求されず、種々の現実的場面で役立つことが知られている。この節ではこうしたスカラー関数 $f: M \rightarrow \mathbb{R}$ の場合について解説する。

スカラー関数の特徴解析には、次で定義されるレベル集合が重要な働きをする。

定義 2.1 値 $c \in \mathbb{R}$ に対して、

$$f^{-1}(c) = \{x \in M \mid f(x) = c\}$$

を**レベル集合**（もしくは、等値集合、等値線、等値面、*isoline*, *isosurface* 等々）という。

一般にレベル集合は $n - 1$ 次元となる ($n = \dim M$)。ただし、多様体になるとは限らない。たとえば地上の標高データであれば、 $n = 2$ であってレベル集合は等高線となる。

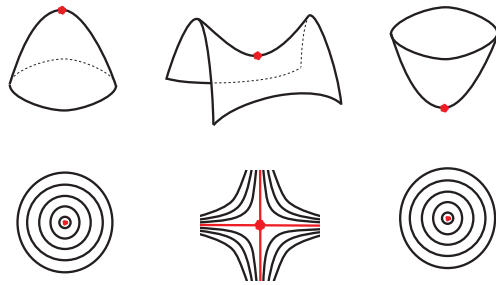


図1：特徴的な等高線

さて、たとえばそうした地上の標高データが与えられたとき、そのデータの特徴を読み取るためには何が重要であろうか？ 等高線が重要であることはもちろんであるが、中に特徴的な等高線があることに気づくであろう。それは頂上、峠、谷底に相当する（図1参照）。

これらの特徴的等高線での（標高の）値を少し変化させると、等高線が生成・消滅したり、あるいは併合・分離したりする。つまり、等高線の形が変化するところが、データの特徴を把握する際に重要である。

こうした情報を集約するために、以下の概念が非常に良く用いられる。

定義 2.2 (Reeb [11]) レベル集合の各連結成分を1点につぶしてできる図形（グラフ）を **Reeb グラフ** という（状況によっては contour tree, volume skeleton tree, topological volume skeleton, level-set graph, Stein 分解, 等々とも呼ばれる）。図2参照。数学的に厳密には、定義域 M の商空間としての位相を入れた位相空間であるが、実際には一般的なスカラー関数 f の Reeb グラフは、頂点と辺からなるグラフとなることが知られている。

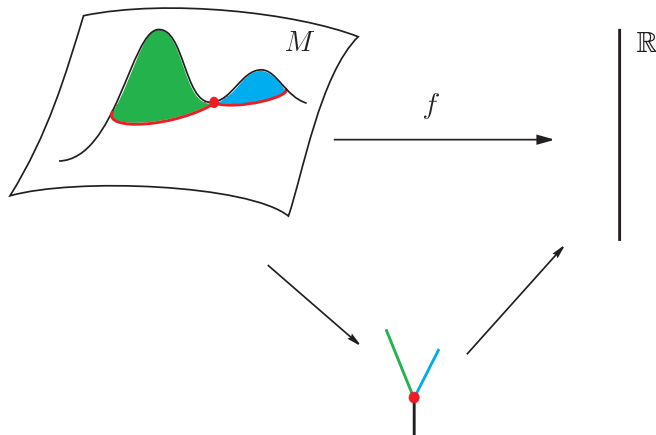


図2：Reeb グラフの例

こうしたグラフは、標高データの場合に等高線の形の変化を表現する道具[13]として非常に多くの研究がなされていると共に、様々な応用も試されている。

与えられたデータから Reeb グラフを求めるアルゴリズムは、定義域の次元 n が $n = 2, 3, 4$ のときには既に確立している（たとえば [1, 4, 10] 参照）。

さて、ここで重要な点は、Reeb グラフの頂点が、ちょうどレベル集合の形が変わるところに対応していることである。データの大局的特徴を捉えるには、こうした頂点が重要な役割を担う。これは図 1 のような点に対応している。これらは以下のように定式化できる。

定義 2.3 (1) 可微分関数 $f: M \rightarrow \mathbb{R}$ を考える。点 $x \in M$ であって、 $(x$ のまわりの局所座標について) そこでの f の 1 階の偏微分係数がすべて消える点を f の**臨界点**（または**特異点**）という。またその点での f の値を**臨界値**という。

(2) 臨界点 x で、 f の 2 階の偏微分係数からなる $n \times n$ 対称行列

$$\left(\frac{\partial^2 f}{\partial x_i \partial x_j} (x) \right)_{i,j}$$

が正則行列のとき、臨界点 x は**非退化**であるという（ここで (x_1, x_2, \dots, x_n) は、点 x のまわりの局所座標である）。

なお、どんな関数も、ほんの少し摂動すれば、臨界点はすべて非退化として良いことが知られている（たとえば [6, 7] 参照）。

微分位相幾何では以下の定理が基本的である。

定理 2.4 (Morse の補題) 非退化な臨界点のまわりで、 f は適当な局所座標により

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

と書ける（ c は定数で、臨界値に相当する）。

上の 2 次式で、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で決定される。すなわち、レベル集合の変化は臨界点の近くでのみ起こるが、その変化は上の 2 次式で完全に記述できるのである。（たとえば次元が $n = 2$ のときは、指数は 2, 1, 0 の 3 種類であって、それぞれの臨界点の近くでのレベル集合は図 1 で尽きている。）こうした意味で、レベル集合の変化を追うために Morse の補題は基本的なのである。

たとえば $n = 3$ で、 M が 3 次元ユークリッド空間の開集合のとき、レベル集合（等値面）の臨界値前後での大局的な変化は、その面が外に向かって値が大きくなるのか、小さくなるのか、つまりその面によって隠される部分の値の大小まで込めて、数学的に分類することが可能である（[14, 15] 参照）。こうした分類は、3 次元データの可視化において大変重要である。

具体例として、たとえば [3] では、陽子と水素原子の衝突における電子密度関数のシミュレーションデータにこうした手法を適用している。これは時空間データであり $n = 4$ に相当するが、実際には時刻 T を止めたときの 3 次元データの解析を行い、その時刻 T を変化させることで、Reeb グラフの変化を抽出し、特徴的な時刻を探している。これにより、衝突の際の電子分布の変化の様子とその特徴が、単なる動画による通常の可視化によるよりも、良くつかめることが解説されている。

3 可微分写像の特異点と特異ファイバー

ではいよいよ一般の可微分写像 $f: M \rightarrow N$ で、 $n = \dim M \geq p = \dim N \geq 1$ で $p = 1$ とは限らない場合を考えよう。 $N = \mathbb{R}^p$ の場合は、 p 個の成分関数を並べて、

$$f = (f_1, f_2, \dots, f_p)$$

と書けるので、 f は p 値関数（もしくは多値関数）とも呼べる。この場合は p 個の関数それぞれについて前節で述べたような解析をすることはもちろん可能である。しかし、それでは各成分関数間の関係はおろか、データの全体像も見えてこないことが多い。

そこでここでは成分関数それぞれを考えてからそれらの間の関係を探る、といった指針を捨て、1つのまとまった写像 f を考える、という立場に立ってデータ解析をしてゆこう。

定義 3.1 点 $c \in N$ に対して、

$$f^{-1}(c) = \{x \in M \mid f(x) = c\}$$

を、 f のファイバーという。（ときにはレベル集合ともいう。）

たとえば、 $n = 3$ とし、 M を海水、 $f: M \rightarrow \mathbb{R}^2$ を、 $f = (\text{水温}, \text{塩分濃度})$ として得られる写像としよう。すると、状況は図3のようになる。

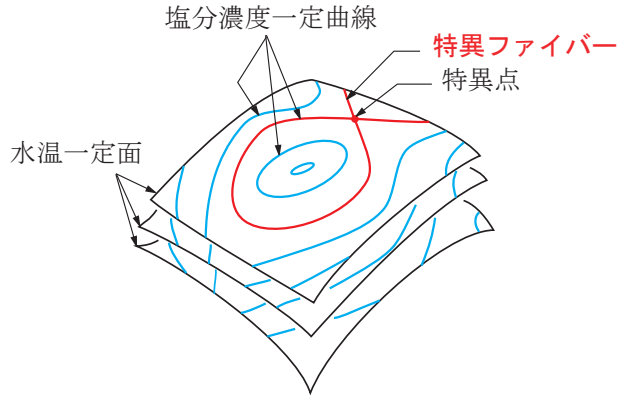


図3: 2値関数のファイバーの例

なお、特異点を含むファイバーを特異ファイバーと呼ぶが（詳細は後述）、これが多値関数データの特徴抽出において重要な役割を果たすことが、スカラー関数の場合の類似として容易に想像できよう。

定義 3.2 可微分写像 $f: M \rightarrow N$ を考える。 M の点 x に対して、 x と $f(x)$ のまわりの局所座標を選ぶ。このとき、 $df_x: \mathbb{R}^n \rightarrow \mathbb{R}^p$ を、 f のヤコビ行列（ f の成分関数の点 x での1階の偏微分係数を並べてできる、実 $p \times n$ 行列）から定まる線形写像とする。これを f の、点 x での

微分という。そして、 $\text{rank } df_x < p$ となる M の点 x を f の特異点という。(この定義は局所座標の選び方に依存しないことが容易に証明できる。) 特異点全体の集合

$$J(f) = \{x \in M \mid \text{rank } df_x < p\}$$

を f のヤコビ集合 (もしくは特異点集合) という。さらに、特異点の f による像を特異値、特異点を含むファイバーを特異ファイバーという。

一般にヤコビ集合 $J(f)$ は $p-1$ 次元となることが知られている。

さて、 $p=1$ の場合、すなわちスカラー関数の場合には Morse の補題があり、それによりレベル集合の変化が非常によく理解できた。しかしながら一般の $p \geq 2$ に対しては、そうした補題は特別な場合を除き知られていない。というよりむしろ、一般には不可能であることが数学的に示されている (詳細は [5, 8] 等を参照)。

ここではそうした Morse の補題に類似の定理が成り立つ場合について解説する。具体的には $p=2, 3$ の場合である。

まず $p=2$ の場合を見てみよう。簡単のため $n=2$ とする。すると、“一般的な写像” の特異点には、折り目とカスプの2種類しかないことが古典的に知られている ([9, 16] 等参照)。それぞれ、局所座標を用いて

$$(x, y) \mapsto (x, y^2), \quad (x, y) \mapsto (x, -xy + y^3)$$

と表される特異点である (図4参照)。

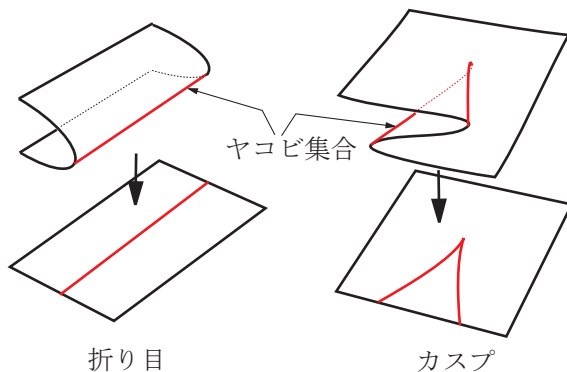


図4: 折り目とカスプ

なお、上のような分類は定義域の次元が $n \geq 3$ の場合にも同様に可能である。ただし、Morse の補題のときのように、指数によっていくつか異なるものが登場する。

たとえば可微分写像 $f: M \rightarrow N$ について、 $p = \dim N = 2$ で、 $\dim M = n \geq 3$ が奇数と仮定しよう。すると各折り目特異点に指数 λ が定義できる ($\lambda = 0, 1, \dots, (n-1)/2$)。ヤコビ集合の中で、どの点がカスプかというのは、理論的にはもちろん分かるはずであるが、データ相手にカスプを同定するためには工夫が必要である。たとえばカスプは、上述の指数が変化すると

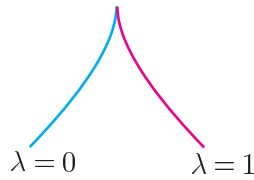


図5：カusp近くの折り目特異点の指数 ($n = 3$ の場合)

ころとして特徴付けられる (図5参照). なお, 折り目特異点の指数を計算することはそれほど (アルゴリズム的に) 難しくはない.

Edelsbrunner-Harer [2] は, 可微分写像を区分的線形写像で近似し, そのヤコビ集合を求めるアルゴリズムを提唱した. 理論的にはそれで求められるはずであるが, 実データに対して適用すると, 曲線となるべき部分がギザギザになってしまうなど, 一般にきれいに出力されない. 図6は, この抽出アルゴリズムを組み合わせて計算された特異ファイバーの例を示しており, 2つのスカラー関数の逆像の共通部分であるファイバーが, その関数値の変化に応じて位相的な変化を起こしている様子が見て取れる. (それぞれ, 定義域 \mathbb{R}^3 と値域 \mathbb{R}^2 がペアになっており, \mathbb{R}^2 の黒い点に対応する特異ファイバーが \mathbb{R}^3 で赤い曲線として描かれている. 左のペアは1つの連結成分が生成・消滅する場合を, 右のペアは2つの連結成分が併合・分岐する場合を表している.) しかしながら, この特異ファイバーの抽出手法は, 解析関数の事例では適切な微分位相幾何特徴をもたらすが, 実際のボリュームデータなどに適用すると, ノイズや離散データの扱いのために, 定義域の多くが特異ファイバーとして覆い尽くされてしまうなどの問題が存在する.

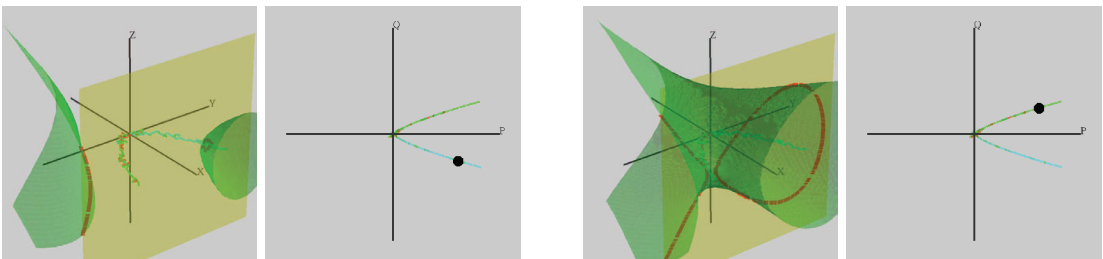


図6：3次元から2次元への写像の特異ファイバーの例

さらに, 特異点や特異ファイバーがどこにあるかは分かっても, 各特異点や特異ファイバーの型は不明である. 微分位相幾何学における可微分写像の特異点論を用いれば, そうした特異点や特異ファイバーの型の同定がある程度可能になり, 大規模データの解析, 及び可視化に大きく貢献できる可能性がある.

4 可視化のために

多値関数データの可視化のためには,

- (1) ヤコビ集合の特定
- (2) 各特異点の型の特定
- (3) ヤコビ集合像の特定
- (4) ヤコビ集合像が仕切る値域多様体の各領域上のファイバーの特定

が必要である．特に上記(4)のためには，特異ファイバーと，その近くでのファイバーの変化を特定することが不可欠となる． $\dim M = n = 3, N = \mathbb{R}^2$ のときの写像 $f: M \rightarrow \mathbb{R}^2$ について，ファイバー変化の一例を図7に示す（赤線がヤコビ集合像を表し，それで仕切られた各領域上のファイバーが黒で，ヤコビ集合像の交点以外の点上のファイバーが青で，交点上のファイバーが緑で，それぞれ描かれている）．

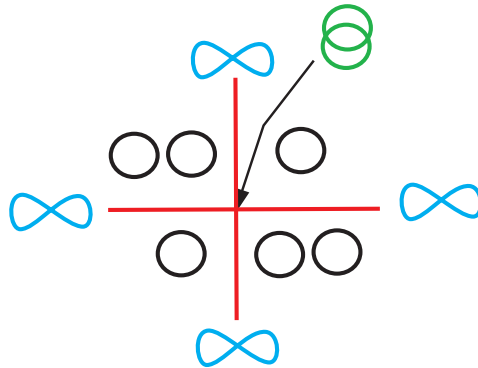


図7：3次元から2次元への写像のファイバー変化の一例

図7を見てもわかるように，ファイバーには，その複雑度に応じて階層構造がある．正確には，ファイバーの「型」 \mathcal{F} に対して，

$$\mathcal{F}(f) = \{y \in N \mid \text{ファイバー } f^{-1}(y) \text{ が } \mathcal{F} \text{ 型}\}$$

という値域多様体 N 内の部分集合を考え，

$$\kappa = \dim N - \dim \mathcal{F}(f)$$

を，ファイバーの型 \mathcal{F} の余次元という． $n = 3, p = 2$ のときに現れるファイバーの余次元の例を図8に示す．

では次に $n = 4, p = 3$ の場合を見てみよう（詳細は[12]を参照）．これはたとえば，時空間上で与えられたデータの3つ組の解析に相当する．まず，“一般的な写像” $f: M \rightarrow N$ ($\dim M = 4, \dim N = 3$) のヤコビ集合像は， N 内の特異点を持った曲面となって現れる．その各点の近くでの様子は図9のように分類される．

たとえば図9の(5)のような点の近くでのファイバー変化は，何種類かあるが，そのうちの一つは図10のようになる．

そして， $n = 4, p = 3$ の場合の特異ファイバーのリストは図11のようになる．




-  最も複雑. $\kappa = 2$, 離散的に現れる.
-  次に複雑. $\kappa = 1$, 曲線に沿って現れる.
-  最も単純. $\kappa = 0$, 面に沿って現れる.

図 8 : 3次元から2次元への写像に現れるファイバーの例とその余次元

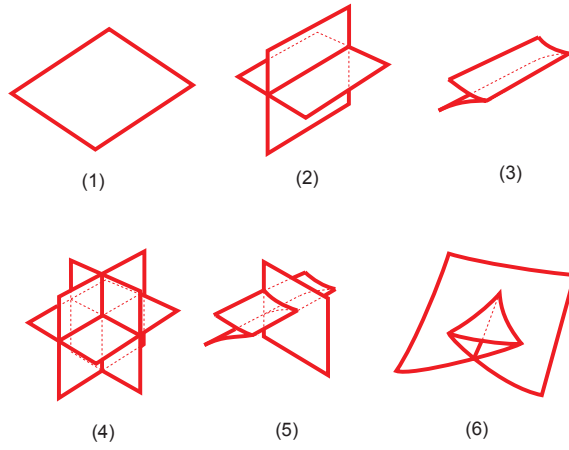


図 9 : 4次元から3次元への写像のヤコビ集合像の局所形

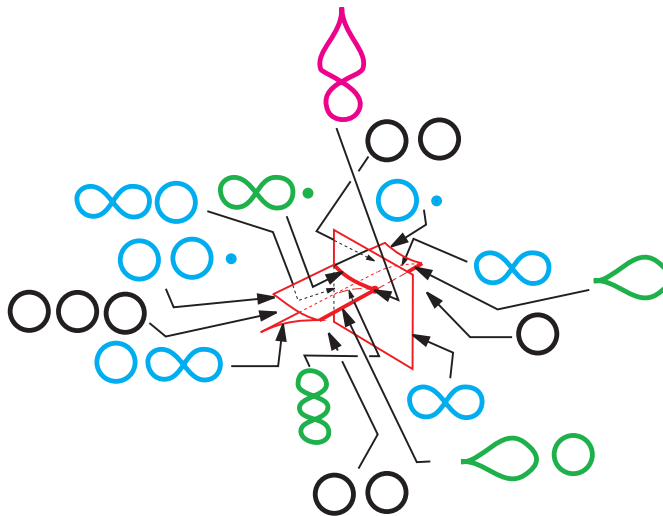


図 10 : 4次元から3次元への写像のファイバー変化の一例




























| | | | | | | | |
|--------------|---|---|---|---|---|--|--|
| $\kappa = 1$ |  |  | | | | | |
| $\kappa = 2$ |  |  |  |  |  |  | |
| $\kappa = 3$ |  |  |  |  |  |  | |
| |  |  |  |  |  |  | |
| |  |  |  |  |  |  | |
| |  | | | | | | |

図 11 : 4次元から3次元への写像に現れる特異ファイバーのリスト

こうしたリストを用いれば、データ解析において、特徴的なファイバーとその型を抽出することが可能となり、データの可視化に貢献できることが大いに期待される。

応用の可能性の例としては、たとえばCTスキャンデータがある。CTスキャンは、ある方向に関する断層画像を積み重ねることで、3次元データを計測する。その際、複数の方向からの断層画像群を計測することで、取得したい3次元データの復元精度を上げることが考えられる。1方向だけ考えればReebグラフを構築することに相当するので、こうして2つ（以上）の方向を考え、Reebグラフの合成表現、あるいは特異ファイバーのつながりを表すグラフを構築すれば、1方向だけ考えていたのでは見えなかった特徴も抽出できるであろう。

さて、こうして見てきた微分位相幾何特徴抽出手法は、通常ではハンドリングできないほどの膨大な取得データに内在する重要な特徴を、少ないデータ量で効率的に画像として表現できる特長を持っている。もちろん、視覚情報の（情報）帯域が数値データなどの文字情報に比べると格段に広いのもご利益の一つである。こうした手法が今後、データ解析において大活躍することになるであろうと期待される所以である。

参考文献

- [1] H. Carr, J. Snoeyink, and U. Axen, Computing contour trees in all dimensions, Computational Geometry: Theory and Applications **24** (2003), 75–94.

- [2] H. Edelsbrunner and J. Harer, Jacobi sets of multiple Morse functions, Foundations of computational mathematics: Minneapolis, 2002, pp. 37–57, London Math. Soc. Lecture Note Ser., Vol. 312, Cambridge Univ. Press, Cambridge, 2004.
- [3] I. Fujishiro, R. Otsuka, S. Takahashi, and Y. Takeshima, T-Map: A topological approach to visual exploration of time-varying volume data, in “High-Performance Computing” (Eds. J. Labarta, K. Joe, and T. Sato), pp. 176–190, Lecture Notes in Computer Science, Vol. 4759, Springer, Berlin, Heidelberg, 2008.
- [4] X. Ge, I. Safa, M. Belkin, and Y. Wang, Data skeletonization via Reeb graphs, Twenty-Fifth Annual Conference on Neural Information Processing Systems 2011, pp. 837–845.
- [5] J.N. Mather, Stability of C^∞ mappings. VI: The nice dimensions, Proc. Liverpool Singularities–Symposium, I (1969/70), pp. 207–253, Lecture Notes in Math., Vol. 192, Springer, Berlin, 1971.
- [6] 松本幸夫, Morse 理論の基礎, 岩波書店, 2005.
- [7] J. Milnor, Morse theory. Based on lecture notes by M. Spivak and R. Wells, Ann. of Math. Studies, No. 51, Princeton Univ. Press, Princeton, N.J., 1963.
- [8] 西村尚史, 福田拓生, 特異点と分岐, 特異点の数理 2, 共立出版, 2002.
- [9] 野口広, 福田拓生, 復刊 初等カタストロフイー, 共立出版, 2002.
- [10] V. Pascucci, G. Scorzelli, P.-T. Bremer, and A. Mascarenhas, Robust on-line computation of Reeb graphs: Simplicity and speed, ACM Trans. Graph. **26**, no. 3, (2007), Article 58, 58.1–58.9.
- [11] G. Reeb, Sur les points singuliers d’une forme de Pfaff complètement intégrable ou d’une fonction numérique, C. R. Acad. Sci. Paris **222** (1946), 847–849.
- [12] O. Saeki, Topology of singular fibers of differentiable maps, Lecture Notes in Math., Vol. 1854, Springer–Verlag, Berlin, 2004.
- [13] S. Takahashi, T. Ikeda, Y. Shinagawa, T.L. Kunii, and M. Ueda, Algorithms for extracting correct critical points and constructing topological graphs from discrete geographical elevation data, Computer Graphics Forum **14** (1995), 181–192.
- [14] S. Takahashi, Y. Takeshima, and I. Fujishiro, Topological volume skeletonization and its application to transfer function design, Graphical Models **66** (2004), 24–49.
- [15] Y. Takeshima, S. Takahashi, I. Fujishiro, and G.M. Nielson, Introducing topological attributes for objective-based visualization of simulated datasets, in “Proc. Volume Graphics 2005”, pp. 137–145, 2005.
- [16] H. Whitney, On singularities of mappings of euclidean spaces. I. Mappings of the plane into the plane, Ann. of Math. (2) **62** (1955), 374–410.