

Giovanni Bellettini · Valentina Beorchia
Maurizio Paolini · Franco Pasquarelli

Shape Reconstruction from Apparent Contours

Theory and Algorithms

Computational Imaging and Vision

Computational Imaging and Vision

Managing Editor

MAX VIERGEVER

Utrecht University, Utrecht, The Netherlands

Series Editors

GUNILLA BORGEFORS, *Centre for Image Analysis, SLU, Uppsala, Sweden*

DANIEL CREMERS, *Technische Universität München, München, Germany*

RACHID DERICHE, *INRIA, Sophia Antipolis, France*

KATSUSHI IKEUCHI, *Tokyo University, Tokyo, Japan*

REINHARD KLETTE, *University of Auckland, Auckland, New Zealand*

ALES LEONARDIS, *ViCoS, University of Ljubljana, Ljubljana, Slovenia*

STAN Z. LI, *CASIA, Beijing & CIOTC, Wuxi, China*

DIMITRIS N. METAXAS, *Rutgers University, New Brunswick, NJ, USA*

HEINZ-OTTO PEITGEN, *CeVis, Bremen, Germany*

JOHN K. TSOTSOS, *York University, Toronto, Canada*

This comprehensive book series embraces state-of-the-art expository works and advanced research monographs on any aspect of this interdisciplinary field.

Topics covered by the series fall in the following four main categories:

- Imaging Systems and Image Processing
- Computer Vision and Image Understanding
- Visualization
- Applications of Imaging Technologies

Only monographs or multi-authored books that have a distinct subject area, that is where each chapter has been invited in order to fulfill this purpose, will be considered for the series.

Volume 44

More information about this series at
<http://www.springer.com/series/5754>

Giovanni Bellettini • Valentina Beorchia • Maurizio
Paolini • Franco Pasquarelli

Shape Reconstruction from Apparent Contours

Theory and Algorithms

Giovanni Bellettini
Department of Mathematics
University of Rome Tor Vergata
Rome
Italy

Valentina Beorchia
University of Trieste Department
of Mathematics & Geosciences
Trieste
Italy

Maurizio Paolini
Università Cattolica del Sacro Cuore
Department of Mathematics & Physics
Brescia
Italy

Franco Pasquarelli
Università Cattolica del Sacro Cuore
Department of Mathematics & Physics
Brescia
Italy

ISSN 1381-6446
Computational Imaging and Vision
ISBN 978-3-662-45190-8 ISBN 978-3-662-45191-5 (eBook)
DOI 10.1007/978-3-662-45191-5

Library of Congress Control Number: 2015933011

Springer Heidelberg New York Dordrecht London
© Springer-Verlag Berlin Heidelberg 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer-Verlag GmbH Berlin Heidelberg is part of Springer Science+Business Media
(www.springer.com)

Contents

1	A Variational Model on Labelled Graphs with Cusps and Crossings	1
1.1	The Reconstruction Problem	1
1.2	The Mumford–Shah Model	4
1.3	The Nitzberg–Mumford Model	6
1.4	Other Curvature-Depending Functionals	11
1.5	The Variational Model on Labelled Graphs	13
	References	21
2	Stable Maps and Morse Descriptions of an Apparent Contour	25
2.1	Stability of Maps	25
2.2	Stable Maps from a Two-Manifold to the Plane	31
2.3	Ambient Isotopies	36
2.4	Ambient Isotopic and Diffeomorphically Equivalent Apparent Contours	41
2.5	Morse Descriptions of an Apparent Contour	42
2.5.1	Genericity of Morse Lines in Case of No Cusps	44
2.5.2	Morse Lines in Case of Cusps: Markers	45
2.5.3	The Morse Description	48
2.5.4	Recovering the Shape from a Morse Description	50
	References	51
3	Apparent Contours of Embedded Surfaces	53
3.1	Three-Dimensional Scenes	53
3.1.1	Splitting of \mathbb{R}^3	54
3.2	Apparent Contours of Embedded Surfaces	55
3.3	The Function f_Σ	59
3.4	Labelling an Apparent Contour: The Function d_Σ	64

3.5	Ambient Isotopic and Diffeomorphically Equivalent Labelled Apparent Contours	69
3.6	Visible Contours	70
	References	72
4	Solving the Completion Problem	73
4.1	Some Concepts from Graph Theory	73
4.1.1	Contour Graphs and Visible Contour Graphs	75
4.2	Complete Contour Graphs and Labelling	77
4.3	Statement of the Completion Theorem	80
4.4	Morse Descriptions of a Visible Contour Graph	82
4.4.1	Localization	84
4.5	Proof of the Completion Theorem	84
4.5.1	Analysis at the Global Maximum and at Local Maxima	86
4.5.2	Analysis at Terminal Points	87
4.5.3	Analysis at T-Junctions	88
4.5.4	Analysis at Local Minima and at the Global Minimum	92
4.6	Examples	96
	References	100
5	Topological Reconstruction of a Three-Dimensional Scene	101
5.1	Statement of the Reconstruction Theorem	101
5.1.1	Depth-Equivalent Scenes	102
5.2	Proof of Existence	103
5.2.1	Gluing	104
5.2.2	Smooth Local Embedding of \mathcal{T} in \mathbb{R}^3	108
5.2.3	Smooth Global Embedding of M in \mathbb{R}^3	112
5.2.4	Definition of the 3D-Shape	117
5.3	Proof of Uniqueness	118
5.A	Appendix	125
	References	128
6	Completeness of Reidemeister-Type Moves on Labelled Apparent Contours	131
6.1	Moves on a Labelled Apparent Contour	133
6.1.1	List of All Simple Rules	134
6.2	Stratifications and Stratified Morse Functions	136
6.2.1	Stratifications Induced by a Stable Map	137
6.3	Informal Statement	140
6.4	Rigorous Statement	142
6.5	Proof of the Completeness Theorem	147
6.6	Completeness of Moves	153
	References	155

7	Invariants of an Apparent Contour	157
7.1	Definition of $\mathfrak{B}(\text{appcon}(\varphi))$	158
7.2	Definition of $BL(\text{appcon}(\varphi))$	160
7.3	Coincidence Between $\mathfrak{B}(\text{appcon}(\varphi))$ and $BL(\text{appcon}(\varphi))$	162
7.3.1	Proof of Coincidence Up to a Constant	163
7.3.2	Proof of Coincidence	167
7.4	Euler–Poincaré Characteristic of ∂E	171
7.5	Cell Complexes and Fundamental Groups	176
7.5.1	Cell Complexes	178
7.5.2	Fundamental Groups	179
7.6	Alexander Polynomials and Invariants of Fundamental Groups	181
7.7	Free Differential Calculus	183
7.8	Links with Two Components: Deficiency One	189
7.9	Surfaces with Genus 2: Deficiency Two	191
	References	192
8	Elimination of Cusps	195
8.1	Embedding Sign of a Cusp	196
8.2	Connectable Cusps in an Open Set	198
8.3	Statement of the Elimination Theorem	201
8.4	Proof of the Elimination Theorem	202
8.5	Application to Closed Embedded Surfaces	206
	References	207
9	The Program “Visible”	209
9.1	An Example	209
9.2	Encoding a Morse Description of the Visible Contour	212
9.2.1	Encoding the Morse Events	213
9.2.2	Implicit Orientation	213
9.2.3	The “e” Region Marking	214
9.3	Using the Program	214
9.4	Encoding a Morse Description of the Constructed Apparent Contour	215
9.5	Some Examples	216
	Reference	223
10	The Program “Appcontour”: User’s Guide	225
10.1	An Overview of the Software	226
10.2	Region Description	230
10.2.1	Extended Arcs	231
10.2.2	Describing a Region	232
10.2.3	Completeness of the Region Description	234
10.3	Encoding an Apparent Contour with Labelling	235
10.3.1	Region Description as a Stream of Characters	235
10.3.2	Morse Description	236
10.3.3	Knot Description	238

10.4	The Rules (Reidemeister-Type Moves)	240
10.4.1	Simple Rules	240
10.4.2	A Nonlocal Effect of the B Rule	244
10.4.3	Composite Rules	244
10.4.4	Inverse Rules	247
10.5	Surgeries on Apparent Contours	248
10.5.1	Vertical Surgery	248
10.5.2	Horizontal Surgery	249
10.6	Canonical Description and Comparison	249
10.6.1	On the Isomorphism Problem for Graphs	250
10.6.2	The “Regions” Graph: \mathcal{R} -Graph	250
10.6.3	The Depth-First Search of an \mathcal{R} -Graph	252
10.6.4	The Canonization Procedure	253
10.6.5	Comparison of Apparent Contours	258
10.7	Fundamental Groups and Cell Complexes	258
10.7.1	Computing the Euler–Poincaré Characteristic and the Number of Connected Components	259
10.7.2	Fundamental Groups	260
10.7.3	Invariants of Finitely Presented Groups and the Alexander Polynomial	262
10.7.4	Alexander Polynomials and Alexander Ideals in Two Indeterminates	267
10.8	The Mendes Graph	274
10.9	Invariants	275
10.9.1	Euler–Poincaré Characteristic	275
10.9.2	Bennequin Invariant	275
10.9.3	Examples of Invariants Computation	276
10.10	contour Reference Guide	280
10.10.1	Informational Commands	280
10.10.2	Operating Commands	284
10.10.3	Conversion and Standardization Commands	287
10.10.4	Cell Complex and Fundamental Group Commands	288
10.10.5	Options Specific to Fundamental Group Computations	289
10.10.6	Common Options	290
10.10.7	Direct Input of a Finitely Presented Group or an Alexander Ideal	291
10.11	showcontour Reference Guide	293
10.11.1	Producing a Proper Morse Description	293
10.11.2	From the Morse description to a polygonal drawing	294
10.11.3	Discrete Optimization of the Polygonal Drawing	294
10.11.4	Dynamic Smoothing of the Polygonal	295

10.12	Using contour in Scripts	296
10.12.1	contour_interact.sh	296
10.12.2	contour_describe.sh	297
10.12.3	contour_transform.sh	299
10.13	Example: knotted Surface of Genus 2	300
10.14	Example: Knots in a Solid Torus	301
10.15	Example: Klein Bottle and the “House with Two Rooms”	304
10.16	Example: Mixed Internal/External Knot	309
10.17	Using appcontour on Apparent Contours Without Labelling	312
10.17.1	Haefliger Sphere	312
10.17.2	Boy Surface	313
10.17.3	Milnor Curve	313
10.17.4	Millett curve	314
10.17.5	Klein bottle	315
10.A	Appendix: Practical Canonization of Laurent Polynomials	317
10.A.1	One-Dimensional Support	318
10.A.2	Two-Dimensional Support	318
	References	321
11	Variational Analysis of the Model on Labelled Graphs	323
11.1	The Action Functional	324
11.1.1	Graphs with Cusps and Curvature in L^p	324
11.1.2	The Functional	325
11.1.3	A Notion of Convergence	326
11.2	Lower Semicontinuity	327
11.3	On the Lower Semicontinuous Envelope of the Action	332
11.3.1	Limits of Labellings	335
11.3.2	Sufficient Conditions: An Example	337
11.A	Appendix A: Systems of Curves	342
11.A.1	Curves of Class pwr_c^p	342
11.A.2	Systems of Curves	344
11.A.3	Parametrizations of Complete Contour Graphs	345
11.B	Appendix B: Convergence and Compactness of Systems of Curves	346
11.B.1	Convergence	347
	References	353
	Nomenclature	355
	Index	361

Introduction

Computer vision and image processing have become active areas of basic and applied mathematical research, due to their impact in the development of new technologies and to the related interesting theoretical problems. The vastness of applications requires a multi-disciplinary study; a selection of involved mathematical areas includes, in particular, calculus of variations, optimization and partial differential equations, probability and statistics, topology and differential topology, differential and discrete geometry, affine geometry, harmonic analysis, inverse problems and numerical analysis; see for instance [2, 9, 18, 19, 26, 28, 37, 57, 62, 63, 68, 70, 75, 89]. The areas of application in ordinary life are numerous and we could mention: medical imaging (image reconstruction, interpretation and aid to diagnostics), video processing and analysis, stereo vision, 3D reconstruction and shape recognition from image sequences, and the restoration and interpretation of satellite images; we refer the reader to [69] and references therein.¹ These subjects are mostly directed by applications, but they require solid grounded theories, appropriate for instance to ensure robustness of the related algorithms.

The aim of this book is to investigate one of the central problems of computer vision,² namely the *topological and algorithmical reconstruction of a three-dimensional scene* $E \subset \mathbb{R}^3$, composed of various smooth bounded solid objects (the connected components of E) starting from information on a generic orthogonal plane projection³ of E . As explained in detail in Chap. 1, the original motivation that led us to this study came from the calculus of variations, in the effort of finding an action functional \mathcal{F} (introduced in [12]) defined on plane graphs and whose minimization should give information on the depth ordering of the various objects composing the scene. Postponing the technical discussion on the variational aspects

¹See also [17, 29, 51, 77, 78, 81].

²See for instance [30–32, 38, 39, 41–44, 55, 56, 65, 66, 76, 82, 83] and references therein.

³ In this book we shall not consider the case when two or more simultaneous projections are involved; the case of shapes evolving in time is, instead, related to ambient isotopic deformations of the objects, an issue which will be treated in various chapters.

Fig. 1 Two partially overlapping objects: neither of them is in front of the other

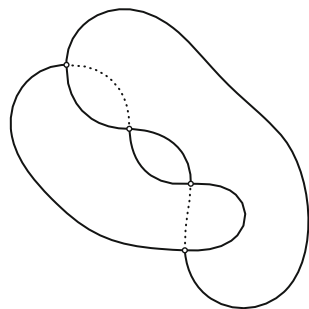
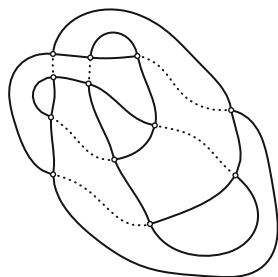


Fig. 2 A knotted torus



of the model to Chaps. 1 and 11, it is worth recalling that the functional \mathcal{F} has been introduced with the purpose of removing a difficulty in a previous model by Nitzberg and Mumford [65, 66] related to self-occlusions: in particular, in Figs. 1 and 2, we draw two interesting and typical examples of self-occlusions, which can be analysed using the functional \mathcal{F} . We also remark that, as observed in [12], admissible configurations for \mathcal{F} , which can be arguably minimizers under a certain range of parameters, may give, as a result, the illusory contours⁴ of the famous Kanizsa triangle [54], as discussed in Sect. 1.5; see, more specifically, Figs. 1.6, 1.7, and 1.8.

In order to carry out our analysis on the topological and algorithmical aspects of the reconstruction problem of a three-dimensional shape E , let us briefly explain what is the information we need on one of its stable plane projections. Denote by Σ the boundary ∂E of E ; for a given generic projection direction, let us consider the so-called *visible apparent contour* $\text{vis}(G_\Sigma) \subset \mathbb{R}^2$ of Σ , an oriented plane graph which is the natural sketch of Σ that one usually draws by hand in order to represent the scene. For instance, for the solid shape in Fig. 3, the bold curves in Fig. 5 represent the visible apparent contour. In order to have a better picture of the various graphs involved, it is often useful to imagine the shape to be semi-transparent, as in Fig. 4. Due to the genericity assumption on the projection direction, it turns out that

⁴We shall treat contours without corners; as we shall explain in Chap. 1, a slight smoothing of the original Kanizsa image does not change the qualitative properties of the example, and does not alter the presence of illusory contours.

Fig. 3 The three-dimensional scene E producing the apparent contour of Fig. 5. Image taken from [14]

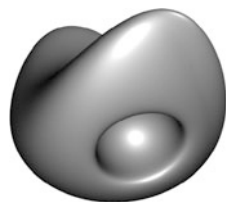


Fig. 4 The same three-dimensional scene as in Fig. 3, but made semi-transparent. Image taken from [14]

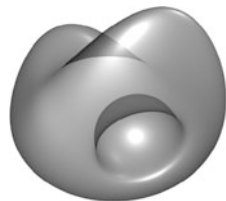
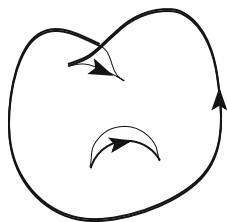


Fig. 5 The *bold graph* represents the visible part of the apparent contour of the three-dimensional scene in Fig. 3; the *whole graph* represents its apparent contour. Image taken from [14]



the singular points of $\text{vis}(G_\Sigma)$, if any, are only of two types: terminal points and T-junctions. The example of Fig. 5 shows three terminal points and one T-junction. It is not difficult to realize that $\text{vis}(G_\Sigma)$ is a subset (usually, but not always, a proper subset) of another oriented graph G_Σ , the so-called *apparent contour* of Σ^5 ; see Fig. 5 again. The graph G_Σ has two types of singular points only: cusps, arising as local completions of terminal points of $\text{vis}(G_\Sigma)$, and X-junctions (called crossings), local completions of T-junctions. Accordingly, the apparent contour of Fig. 5 has four cusps and one X-junction. It is useful to observe that G_Σ has here a geometric meaning: it is the plane projection of a finite set of smooth pairwise disjoint closed curves lying on Σ (and called *critical set*, or also *singular set*), obtained as the set of all points of Σ where the tangent plane contains the projection direction. Now, the crucial three-dimensional information carried by G_Σ is contained in a labelling [25, 53, 87, 88], which is a number $d_\Sigma(a) \in \mathbb{N}$ attached to any arc a , and representing the number of sheets of Σ in front of the part of the singular curve projecting on that arc. Accordingly, $\text{vis}(G_\Sigma)$ is the closure of the set $\{d_\Sigma = 0\}$. For instance, in Figs. 1 and 2, we have $d_\Sigma = 2$ on the dotted arcs, and $d_\Sigma = 0$ on the visible arcs. The *labelled apparent contour*, namely the pair (G_Σ, d_Σ) , is the starting point for the definition of the action functional \mathcal{F} , leading to the minimization principle described in Chaps. 1 and 11. We mention here that another

⁵Sometimes, we shall call G_Σ apparent contour of E .

symbol, denoted by f_Σ , appears in the domain of \mathcal{F} : for $x \notin G_\Sigma$ in the projection plane, the value $f_\Sigma(x)$ represents the total number of intersections between Σ and a light ray emanating from x (see Fig. 1.1 (right)). Such a locally constant function, which can be easily recovered as a doubled winding number with respect to G_Σ , is useful for various reasons, one of them being that it simplifies the presentation of the model.

Several theoretical and practical questions arise in a natural way, and show beautiful and unexpected relations between calculus of variations [62], singularity theory [4–6, 20, 86], Morse theory [10] and knot theory [73]. Moreover, we stress that the techniques we use when dealing with most of such issues fit naturally in an algorithmic setting: as we shall see, it is one of our primary goals to analyse this algorithmic part, with implementations, experiments and computed examples.

We shall be interested in investigating:

- (i) the completion problem, namely: the characterization of those plane graphs which are visible part of a labelled contour graph;
- (ii) an algorithmic construction of a completion, to be implemented as a computer program;
- (iii) the characterization of those labelled plane graphs which are apparent contours of some smooth stable three-dimensional scene;
- (iv) an algorithmic reconstruction of the topology of a three-dimensional smooth shape starting from a labelled apparent contour, and its implementation on a computer;
- (v) a list of topological invariants of three-dimensional shapes, which can be directly computed starting from the apparent contour, and that can be implemented on a computer;
- (vi) the recognition of two labelled apparent contours which are apparent contours of two ambient isotopic shapes, using a finite sequence of elementary moves, taken from a complete finite set. In other words, what are the moves on the labelled apparent contours that relate two *embedded* surfaces, deformable into each other by a smooth path of embeddings?
- (vii) a computer program aiming to implement the elementary moves on apparent contours, and, more in general, capable to manage labelled (or unlabelled) apparent contours from a structural/topological point of view;
- (viii) the problem of elimination of cusps, namely: how to use the elementary moves in order to modify a labelled apparent contour into another one without cusps, representing a three-dimensional shape, ambient isotopic to the original one;
- (ix) the generalization of some of the above problems, in particular the algorithmic parts, to more general situations, concerning for instance abstract closed (not necessarily orientable) surfaces;
- (x) a variational study of the functional \mathcal{F} , such as an investigation of the properties of sequences of labelled apparent contours having a uniform bound on the action.

Problem (i), which is global in nature, addresses necessary and sufficient conditions on an oriented plane graph K with nonexterior terminal points and T-junctions, in order to be the visible part of what we shall call a *complete labelled contour graph* (G, d) , having cusps and X-junctions. Here the function d , defined on the arcs of G , is a labelling, and must fulfill the same consistency properties⁶ shared by the function d_Σ . We can roughly rewrite (i) with the statement

$$\text{given } K \quad \exists (G, d) \text{ such that } G \supseteq K \text{ and } K = \{d = 0\}. \quad (1)$$

Following [14], in Chap. 4 we give a constructive solution to this problem (see Theorem 4.3.1, that we call the *completion theorem*), based on a suitable *Morse description* of K and G . We note here that the aim of the completion theorem is not to provide the “simplest” completion of K , whatever simplest could mean⁷; the scope of the result is to show that the conditions⁸ imposed on K are sharp, and allow us to construct at least one completion. The problem of returning, as output of the completion, a “simple” graph G is related to points (v) and (vi), and will be addressed below. An improvement of the results of [14] is given by Corollary 4.5.1, based on the introduction of the background; this represents a further degree of freedom, which allows us to fix a priori the regions of G where $f = 0$. Here the function f is defined on $\mathbb{R}^2 \setminus G$; it can be obtained as a doubled winding number with respect to G , and has to satisfy various consistency relations with the labelling d . Examples 4.6.2 and 4.6.4 clarify the interest in the use of Corollary 4.5.1, in connection with the reconstruction of the apparent contour of a standard torus; see also the example illustrated in Sect. 9.1 with the use of the `visible` program. We remark that the completion of a visible contour is inherently nonunique, even when forcing a priori the background (the region where $f = 0$); this is clarified with the example displayed in Fig. 9.10.

The Morse description, explained in Sect. 2.5, is a convenient way to encode all topological information of a graph, and fits well for practical purposes. This is clearly seen when dealing with problem (ii): the software code described in Chap. 9, in particular the `visible` program, is an actual implementation of the constructive proof given in the completion theorem. The input of the program is a Morse description of a drawing of a visible contour, see for instance Figs. 9.1, 9.9 and 9.11. The output, provided the graph is completable (namely, it satisfies the necessary and sufficient conditions of the completion theorem) is a complete labelled contour graph, still identified using a textual Morse description, and next graphically reconstructed as a drawing, using the visualization program `showcontour`. The `visible` program also recognizes those graphs K which are not completable (called “impossible graphs”), such as those described in Figs. 9.12 and 9.13.

⁶See Definition 4.2.5.

⁷For instance, a graph with a minimal number of vertices, or without cusps.

⁸See Definition 4.1.8 and Figs. 3.15 and 3.16.

The algorithmic reconstruction of the completion theorem (and of the `visible` program) strongly depends on the Morse description of the visible contour: different Morse descriptions of the same visible contour will in general lead to different structurally non-equivalent 3D shape reconstructions. Conversely, different (but structurally equivalent) visible contours described by the same Morse description will clearly lead to the same reconstruction.

The usefulness of the existence of a consistent labelling d on an oriented complete graph G is that it *characterizes* the apparent contours of smooth *stable* 3D shapes. Following closely the proof of [12] (see also [58, 87, 88]), in Theorem 5.1.1 (called the *reconstruction theorem*) we show how to reconstruct a smooth, not necessarily connected, 3D shape E starting from the labelled graph (G, d) ; namely, how to find a smooth closed surface $\Sigma := \partial E$ such that

$$G = G_{\Sigma}, \quad d = d_{\Sigma}. \quad (2)$$

The notion of stability⁹ employed in this theorem (Definition 2.1.2) goes back to the pioneering works of Whitney [86] and Thom [79], (see also Arnold [3] and Wall [84]) on singularity theory; stability turns out to be a crucial concept, and its generalizations and ramifications are of central importance in the whole book. Just to mention a few consequences of this assumption,¹⁰ it guarantees that all graphs that we consider have a finite number of nodes, that the self-intersections (X-junctions and T-junctions) are double and transverse, and that the cusps are ordinary cusps. Remarkably, stable maps from a closed two- or three-dimensional manifold to a two- or three-dimensional manifold are dense, and their singularities have been classified (see for instance [40] and references therein): these results are the cornerstone for the completeness result illustrated in Chap. 6 and, as a consequence, for a large part of the algorithms described in Chap. 10.

The cut-and-paste proof of the reconstruction theorem is topological in character and constructive. Deferring the technical details to Chap. 5, a couple of related comments are in order. The reconstructed ∂E is unique, up to transformations which do not change the order and the number of intersections of the manifold with the light rays emanating from the projection plane (and therefore do not modify the corresponding labelled apparent contour): this sort of uniqueness result is proven in Theorem 5.1.4. The proof of the reconstruction theorem furnishes an embedded smooth manifold ∂E , but not the “roundest” way to embed it in the ambient space \mathbb{R}^3 ; investigation of this latter problem is beyond the scope of the present book.

Summarizing the discussion concerning points (i)–(iii), we conclude that, starting from a visible contour graph K , we can construct a complete labelled contour graph (G, d) satisfying (1), which, in turn, provides a three-dimensional scene

⁹See, e.g., [40] and references therein.

¹⁰We assume that the boundary of the scene E is in general position with respect to the projection: using the concept of stability, this means that the restriction to ∂E of the projection is stable, see Sect. 3.2 for the details.

E satisfying (2) and fulfilling the natural sort of uniqueness for such kind of problems. The completion and the cut-and-paste procedure are automatized in a computer program. Moreover, problem (iv) is one of the issues considered in Chap. 10, which aims to be a self-contained user's guide to an original and rather complex computer program for the reconstruction of three-dimensional shapes, based on an analysis of apparent contours. The reconstruction problem is completely solved from an algorithmic point of view; the program `appcontour` reconstructs the topological structure of ∂E , in particular information such as the number of connected components of ∂E and the Euler–Poincaré characteristic of each of them can be obtained, together with information about the relative position in space allowing to distinguish, e.g., between two concentric spheres (E is a hollow sphere) and two mutually external spheres (E is a pair of solid spheres), with the commands “`contour countcc`”, “`contour extractcc`”, “`contour characteristic`”, “`contour ccordering`”, “`contour ccparent`” (see Sect. 10.10.1).

When proving results such as the reconstruction theorem, or also when analysing topological invariants of apparent contours, one realizes that a basic idea is to consider the more general concept of apparent contour of a map from a manifold into another manifold.¹¹ This is a classical topic in differential topology: see for instance [50, 60, 79, 86]. In particular, given a two-dimensional smooth closed (abstract) manifold M and a smooth stable map $\varphi : M \rightarrow \mathbb{R}^2$, the *apparent contour* $\text{appcon}(\varphi)$ of φ is the subset of \mathbb{R}^2 where the function counting the number of preimages of φ has a jump. It can be equivalently defined as the image in \mathbb{R}^2 of the *critical set* (or *singular set*) of φ in M , where the rank of the differential of φ is not maximal. The previously discussed labelled apparent contour of an embedded surface is a special case: in particular, the reconstruction theorem can be restated in terms of *factorization of maps*, as follows. Let (G, d) be a complete labelled contour graph. Then

$$G = \text{appcon}(\varphi) = G_{\Sigma} \quad \text{and} \quad d = d_{\Sigma},$$

where φ is a map from a smooth closed two-manifold M to the plane, $\Sigma := e(M)$ for a smooth embedding e of M into $\mathbb{R}^2 \times \mathbb{R}$, and φ factorizes as

$$\varphi = \pi \circ e, \tag{3}$$

where π is an orthogonal projection $\pi : \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}^2$, with Σ in general position with respect to π . Indeed, the core of the proof of the reconstruction theorem consists in producing the manifold M as a quotient, and next in embedding it in \mathbb{R}^3 ; the same theoretical procedure is next implemented in the `appcontour` program, as a starting point of the computation of the first fundamental group, as we shall see.

¹¹As we shall see, this abstract viewpoint is essential also in Chap. 6.

According to this more general viewpoint, in Chap. 2 we recall a few well-known facts from singularity theory¹² (see, e.g., [7, 8, 40] and references therein), for a stable map φ from a closed smooth manifold \mathcal{X} to a manifold \mathcal{Y} , and give some examples. It is worth recalling here that knot theory is the study of stable maps from the circle \mathbb{S}^1 to \mathbb{R}^3 . The choices

$$\mathcal{X} = M, \quad \mathcal{Y} = \mathbb{R}^2 \text{ or } \mathbb{R}^3$$

(in particular $\dim(\mathcal{X}) = 2$) can be applied to the study of apparent contours of closed not necessarily embeddable (or even not immersible) manifolds in \mathbb{R}^3 ; we quickly touch these issues (point (ix) of the above list) in Sect. 10.17, with the Boy surface (a standard immersion of the real projective plane), the Klein bottle, and examples from the literature such as the Haefliger sphere, the Millett projection of the real projective plane and the Milnor sphere. These examples lead to consider the interesting problem of apparent contours possibly without labelling, a subject that we do not want to further deepen in the present book.

Concerning point (v), in Chap. 7, we study some *invariants* of an apparent contour for a map $\varphi : M \rightarrow \mathbb{R}^2$. In the first part of the chapter (Sects. 7.1–7.3), we analyse invariants under diffeomorphisms of the target space \mathbb{R}^2 . Besides the number of cusps and of crossings of the apparent contour $\text{appcon}(\varphi)$, a third invariant has been considered in [67], and called Bennequin-type invariant, denoted by $BL(\text{appcon}(\varphi))$. This invariant, based on the Bennequin’s construction for Legendrian knots [16, 52], does not have an immediate interpretation.¹³ Following [13], in Theorem 7.3.1, we show that such an invariant can be obtained solely looking at the apparent contour, without resorting to a Legendrian lift (see, in particular, Definition 7.1.2): indeed, it turns out that the invariant can be computed only taking into account the nodes, the cusps, the extremal points with respect to some height function and the orientation of the apparent contour. Here, again, Morse descriptions of $\text{appcon}(\varphi)$ play a central role. In this way the computation can be implemented into a computer program, and this is done by the program `appcontour`, command “`contour info`”. In the last part of the chapter (Sect. 7.4), we suppose that the map φ factorizes through an embedding in \mathbb{R}^3 and an orthogonal projection as in (3). Then, we analyse some invariants of the apparent contour under diffeomorphisms of \mathbb{R}^3 . The computation [12] of the total Euler–Poincaré characteristic $\chi(\Sigma)$ of the surface ∂E of the corresponding solid shape is given in Theorem 7.4.1, in terms only of the apparent contour G_Σ : interestingly, and

¹²such as the notion of stratification [45], see Sect. 6.2.

¹³It is defined as an appropriate linking number of the Legendrian lift of $\text{appcon}(\varphi)$ in the projectivized cotangent bundle $PT^*\mathbb{R}^2$, and its computation for a given apparent contour is not trivial. More precisely, $BL(\text{appcon}(\varphi))$ is defined by taking the sum of the self-linking numbers of the liftings of the components of $\text{appcon}(\varphi)$ and the linking numbers between the liftings of two different components. The self-linking number is itself defined by also taking into account the twisting of a strip constructed by shifting points of the lifted curve by a small amount in the normal direction to the contact plane; we refer to [67] for the precise definition.

a posteriori not surprisingly, the formula for $\chi(\Sigma)$ is independent of the labelling d_Σ on G_Σ ; see the discussion in Sect. 7.4.

The chapter concludes with a number of remarks concerning the first fundamental group of E and $\mathbb{R}^3 \setminus E$. In particular, we discuss the Alexander polynomial (focusing mainly on Fox differential calculus [35, 36]) and some invariants of fundamental groups, applied to surfaces with genus two; see Sects. 7.6–7.9. These issues, as well as the actual computation of $\chi(\Sigma)$, are implemented in Chap. 10 (see Sects. 10.7 and 10.9).

Now, let us discuss the (huge) problems listed in points (vi) and (vii) and their consequences (for instance, the solution to point (viii)). To this aim, it is useful to start by recalling the well-known result [1, 21, 64, 71] of Reidemeister in knot theory, asserting that two *link diagrams*¹⁴ represent ambient isotopic links if and only if they can be related by a finite number of local Reidemeister moves or their inverses. We are interested in a similar question for two-dimensional smooth closed manifolds M embedded in \mathbb{R}^3 . Following [15],¹⁵ in Chap. 6 we prove that two generic embeddings of a closed surface M in \mathbb{R}^3 are ambient isotopic if and only if their apparent contours can be connected using only a finite set of *elementary moves* (also called *rules*, or *Reidemeister-type moves*) on labelled apparent contours and a finite number of smooth planar isotopies. We refer to Sect. 6.3 for an informal presentation of this result, which is proven in Sect. 6.5, and addressed as the *completeness theorem*. It turns out that there are six basic moves¹⁶ on an apparent contour (see Fig. 6.2 for a graphical representation) originated from a general deformation of the corresponding embedded surface; they can be used in exactly the same way as the Reidemeister moves on link diagrams. The essence of the result is that this set of moves is complete. This means that two embedded

¹⁴The diagram of a knot, or more generally of a link, is an orthogonal projection of the image of the link onto some generic plane, with the addition of the knowledge of which strand goes over at each crossing. Stability implies that transversal crossings are the only possible singularities of the diagram.

¹⁵The proof has some similarities with the one described in [24] for the embedding of surfaces in \mathbb{R}^4 .

¹⁶Namely, K (from the Russian word *kasanie* = *tangency*), L (lips), B (beak-to-beak), C (cusp-fold), S (swallow's tail) and T (triple point). This list of moves is essentially the same found in the literature for the related subject of maps from two-manifolds into \mathbb{R}^2 (see, e.g., [67]), even if the addition of the labelling entails a different classification of the list of moves. Similar classifications appear in various contexts, in particular in Thom's catastrophe theory [80] and in Cerf's theory [27], and in the paper [61] of Mond; see also the papers [59, 72, 85]. Concerning a complete set of Reidemeister moves relating two equivalent knotted surfaces in \mathbb{R}^4 , we refer to the set of moves found by Roseman [74], to the papers of Carter and Saito [22, 23] where generic embedded surfaces in \mathbb{R}^4 are considered, projected in \mathbb{R}^3 (diagram) and projected further in \mathbb{R}^2 , and to the papers [46–49] of Goryunov. We refer to [7, 8, 11, 24, 25] for further information. The results illustrated in Chap. 6 treat the case of embeddings, which are usually not considered in the literature. Considering paths of embeddings concretely means that one has to take into account the behaviour of the labelling at the “critical times” corresponding to the intersection with the strata of the so-called discriminant hypersurface.

surfaces in general position with respect to the projection, that can be deformed into each other, have apparent contours that can be connected using solely a finite sequence of such moves (and a finite number of planar isotopies). A relevant part of Chap. 10 consists in the *implementation* of the above-mentioned moves, which are essential for the results related to Chaps. 4, 7 and 9; see Sects. 10.1 and 10.2. Among the various interesting features of the program `appcontour`, the implementation of the moves allows, in several situations, to “simplify” the apparent contour, thus making possible to recognize the topology of the actual three-dimensional shape to which it corresponds (via the reconstruction theorem). It is however worth recalling that, in the simpler case of knots, there is at the moment no algorithm (and no invariant) which is capable to recognize equivalent knots. The knot group (fundamental group of the complement in \mathbb{R}^3), a powerful invariant that can be computed by `appcontour` via a presentation, is capable of distinguishing the unknot; however, it is not a manageable invariant and the problem is shifted to that of recognizing equivalent presentations of the same finitely presented group.¹⁷

A typical situation is when considering a completion of a visible contour graph provided by the completion theorem; if the completion is so complicated that the corresponding three-dimensional shape is not recognizable, we can resort to the `appcontour` program in order to try to simplify it: in several cases, this makes possible to figure out the scene (see for instance Example 4.6.4).

One interesting application of the completeness theorem is given in Chap. 8, where we give a solution to point (viii): in Theorem 8.3.2, we show that, up to \mathbb{R}^3 -ambient isotopies, any smooth closed surface embedded in \mathbb{R}^3 has an apparent contour without cusps.¹⁸ The proof of this result is based on the judicious application of various combinations of the elementary moves and their inverses. This result is, in some case, another example showing a possible way to simplify an apparent contour. Notice carefully, however, that this is not always the case: indeed, there are situations in which the elimination of all cusps is obtained at the expenses of increasing the number of crossings.

The book concludes with Chap. 11 where, following [12], we analyse some variational properties of the action functional \mathcal{F} discussed at the beginning (and described in Chap. 1). In order to minimize \mathcal{F} it is useful to deepen the study of its lower semicontinuous envelope, and this amounts in taking limits of sequences of labelled apparent contours with a uniform bound on the action. In this passage to the limit, many nice properties of labelled apparent contours (consequences of the stability assumptions) are lost. In particular, in Sect. 11.3.2 we produce some

¹⁷There are a number of software codes for the study of knots and their invariants and for the manipulation of three-manifolds, such as SnapPea, SnapPy, Orb, and Knotscape; we refer to the link <http://www.math.uiuc.edu/~nmd/computop/index.html> for further information. See also [33, 34].

¹⁸Probably the more common example is represented by the apparent contour of a torus with four cusps and two crossings, which can be modified into two concentric circles.

examples which illustrate the difficulties in identifying a notion of limit labelling defined on a limit graph.

Finally, the reference list of this book is far from being complete; we apologize for this incompleteness.

Acknowledgements The first author is partially supported by GNAMPA of INdAM. The second author is partially supported by funds of the Università degli Studi di Trieste - Finanziamento di Ateneo per progetti di ricerca scientifica - FRA 2013, and by GNSAGA of INdAM.

We are very grateful to Vicent Caselles, who prematurely died in August 2013, for his interest in our project, for the encouragement he gave us and for pointing out several related references. We are indebted to Carlo Petronio for a lot of remarks and mathematical suggestions and for pointing us various interesting articles in the literature. We are grateful to Alberto Abbondandolo, Francesca Aicardi, Walter Gerbino, Peter Giblin, Domenico Luminati, Riccardo March, Simon Masnou, Mattia Mecchia, Andrea Mennucci, Toru Ohmoto, Giovanni Paolini, Martin Rumpf and Benedikt Wirth for useful discussions.

References

1. Adams, C.A.: *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots*. American Mathematical Society, Providence (2004)
2. Alvarez, L., Guichard, F., Lions, P.-L., Morel, J.-M.: Axioms and fundamental equations of image processing. *Arch. Ration. Mech. Anal.* **123**, 199–257 (1993)
3. Arnold, V.I.: Singularities of smooth mappings. *Uspekhi Math. Nauk.* **23**, 3–44 (1968); *Russ. Math. Surv.* **34**(2), 1–42 (1979)
4. Arnold, V.I.: *The Theory of Singularities and Its Applications*. Cambridge University Press, Cambridge (1991)
5. Arnold, V.I.: *Singularities of Caustics and Wave Fronts*. *Mathematics and Its Applications*, vol. 62. Kluwer, Dordrecht (1991)
6. Arnold, V.I.: Invariants and perestroikas of plane fronts. *Proc. Steklov Inst. Math.* **209**, 11–56 (1995)
7. Arnold, V.I., Gusein-Zade, S.M., Varchenko, A.N.: *Singularities of Differentiable Maps*, vol. I. Birkhäuser, Boston (1985)
8. Arnold, V.I. (ed.), Goryunov, V.V., Lyashko, O.V., Vassiliev, V.A.: *Dynamical Systems VIII: Singularity Theory. II. Applications*. Springer, Berlin (1993)
9. Aubert, G., Kornprobst, P.: *Mathematical Problems in Image Processing*. *Applied Mathematical Sciences*, vol. 147. Springer, New York/Berlin (2002)
10. Audin, M., Damian, M.: *Théorie de Morse et Homologie de Floer*. EDP Sciences. CNRS Editions, Paris (2010)
11. Banchoff, T., Gaffney, T., McCroy, C.: *Cusps of Gauss Mappings*, vol. 55, pp. 1–88. Pitman Advanced Publishing Program, London (1982)
12. Bellettini, G., Beorchia, V., Paolini, M.: Topological and variational properties of a model for the reconstruction of three-dimensional transparent images with self-occlusions. *J. Math. Imaging Vision* **32**, 265–291 (2008)
13. Bellettini, G., Beorchia, V., Paolini, M.: An explicit formula for a Bennequin-type invariant for apparent contours. *Topology Appl.* **156**, 747–760 (2009)
14. Bellettini, G., Beorchia, V., Paolini, M.: Completion of visible contours. *SIAM J. Imaging Sci.* **2**, 777–799 (2009)
15. Bellettini, G., Beorchia, V., Paolini, M.: Completeness of Reidemeister-type moves for surfaces embedded in three-dimensional space. *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl.* **22**, 1–19 (2011)

16. Bennequin, D.: Entrelacements et équations de Pfaff. *Astérisque* **107–108**, 87–161 (1983)
17. Bishop, C.M.: *Pattern Recognition and Machine Learning*. Springer, New York (2006)
18. Blake, A., Isard, M.: *Active Contours: The Application of Techniques from Graphics, Vision, Control and Statistics to Visual Tracking of Shapes in Motion*. Springer, London (1998)
19. Blake, A., Zisserman, A.: *Visual Reconstruction*. MIT Press Series in Artificial Intelligence. MIT Press, Cambridge (1987)
20. Bruce, J.W., Giblin, P.J.: *Curves and Singularities: A Geometrical Introduction to Singularity Theory*, 2nd edn. Cambridge University Press, Cambridge (1992)
21. Burde, G., Zieschang, H.: *Knots*. W. De Gruyter, Berlin (2003)
22. Carter, J.S., Saito, M.: Syzigies among elementary string interactions in $2+1$ dimensions. *Lett. Math. Phys.* **23**(4), 287–300 (1991)
23. Carter, J.S., Saito, M.: Reidemeister moves for surfaces isotopies and their interpretation as moves to movies. *J. Knot Theory Ramif.* **2**(3), 251–284 (1993)
24. Carter, J.S., Saito, M.: *Knotted Surfaces and Their Diagrams*. Mathematical Surveys Monographs, vol. 55. American Mathematical Society, Providence (1998)
25. Carter, J.S., Kamada, S., Saito, M.: Surfaces in 4-Space. *Encyclopaedia of Mathematical Sciences*, vol. 142. Springer, Berlin (2004)
26. Caselles, V., Coll, B., Morel, J.M.: A Kanizsa programme. In: Serapioni, R., Tomarelli, F. (eds.) *Variational Methods for Discontinuous Structures*, pp. 35–55. Birkhäuser, Basel (1996)
27. Cerf, J.: La stratification naturelle des espace de fonctions différentiables réelles et le théorème de la pseudo-isotopie. *Publ. Math. Inst. Hautes Études Sci.* (39), 5–170 (1970)
28. Chan, T.F., Shen, J.: *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*. SIAM, Philadelphia (2005)
29. Chen, C.H. (ed.): *Computer Vision in Medical Imaging*. Series in Computer Vision, vol. 2. World Scientific Publishing, Singapore (2014)
30. Cipolla, R.: *Active Visual Inference of Surface Shape*. Lecture Notes in Computer Science, vol. 1016. Springer, New York (1996)
31. Cipolla, R., Giblin, P.: *Visual Motion of Curves and Surfaces*. Cambridge University Press, Cambridge (1999)
32. Cipolla, R., Fletcher, G., Giblin, P.: Surface geometry from cusps of apparent contours. *Proceedings of the 5th International Conference on Computer Vision 1995*, pp. 858–863, 20–23 June 1995
33. Delfinado, C.J.A., Edelsbrunner, H.: An incremental algorithm to Betti numbers of simplicial complexes on the 3-sphere. *Comput. Aided Geom. Des.* **12**, 771–784 (1995)
34. Dey, T.K., Guha, S.: Computing homology groups of simplicial complexes in \mathbb{R}^3 . *J. ACM* **45**, 266–287 (1998)
35. Fox, R.H.: Free differential calculus. I. *Ann. Math.* **57**, 547–560 (1953)
36. Fox, R.H.: Free differential calculus. II. The isomorphism problem of groups. *Ann. Math.* **59**, 196–210 (1954)
37. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Mach. Intell.* **6**, 721–741 (1984)
38. Giblin, P.J.: Apparent contours: an outline. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **356**, 1087–1102 (1998)
39. Giblin, P., Weiss, R.: Reconstruction of surfaces from profiles. In: *First International Conference on Computer Vision (ICCV'87)*, pp. 136–144. IEEE Computer Society Press, London (1987)
40. Golubitsky, M., Guillemin, V.: *Stable Mappings and Their Singularities*. Graduate Texts in Mathematics, vol. 14. Springer, New York/Heidelberg/Berlin (1974)
41. Golubyatnikov, V.P.: On reconstruction of transparent surfaces from their apparent contours. *J. Inverse Ill-Posed Probl.* **6**, 395–401 (1998)
42. Golubyatnikov, V.P.: *Uniqueness Questions in Reconstruction Multidimensional Objects from Tomography-Type Projection Data*. Inverse and Ill-Posed Problems Series, VSP, Utrecht (2000)
43. Golubyatnikov, V.P., Pekmen, U., Karaca, I., Ozyilmaz, E., Tantay, B.: On reconstruction of surfaces from their apparent contours and stationary phase observations. In: *Proceedings Shape*

- Modeling International '99. International Conference on Shape Modeling and Applications, 1999, pp. 116–120, 1–4 Mar 1999
44. Golubiatnikov, V.P., Karaca, I., Ozyilmaz, E., Tantay, B.: On determining the shapes of hyper-surfaces from the shapes of their apparent contours and symplectic geometry measurements. *Sib. Adv. Math.* **10**, 9–15 (2000)
 45. Goresky, M., MacPherson, R.: *Stratified Morse Theory*. Springer, Berlin (1988)
 46. Goryunov, V.V.: Singularities of projections of complete intersections. *J. Soviet Math.* **27**, 2785–2811 (1984)
 47. Goryunov, V.V.: Projections of generic surfaces with boundaries. *Adv. Soviet Math.* **1**, 157–200 (1990)
 48. Goryunov, V.V.: Monodromy of the image of a mapping. *Funct. Anal. Appl.* **25**, 174–180 (1991)
 49. Goryunov, V.V.: Local invariants of mappings of surfaces into three-space. In: Arnold, V.I., Gelfand, I.M., Smirnov, M., Retakh, V.S. (eds.) *The Arnold-Gelfand Mathematical Seminars*, pp. 223–255. Birkhäuser, Boston (1997)
 50. Haefliger, A.: Quelques remarques sur les applications différentiables d'une surface dans le plan. *Ann. Inst. Fourier Grenoble* **10**, 47–60 (1960)
 51. Herman, G.T.: *Fundamental of Computerized Tomography*. Advances in Pattern Recognition. Springer, New York (2010)
 52. Honda, K.: 3-dimensional methods in contact geometry. In: Donaldson, S.K., Eliashberg, Y., Gromov, M. (eds.) *Different Faces of Geometry*. International Mathematics Series, pp. 47–86. Kluwer, New York (2004)
 53. Huffman, D.A.: Impossible objects as nonsense sentences. In: Meltzer, B., Michie, D. (eds.) *Machine Intelligence*, vol. 6. American Elsevier Publishing Co., New York (1971)
 54. Kanizsa, G.: *Organization in Vision*. Praeger, New York (1979)
 55. Karpenko, O.A., Hughes, J.F.: *SmoothSketch: 3D free-form shapes from complex sketches*. In: *The 33rd International Conference and Exhibition on Computer Graphics and Imaging Techniques*, Boston, Massachusetts, 2006, SIGGRAPH 2006, pp. 589–598. ACM, New York (2006)
 56. Koenderink, J.J.: *Solid Shape*. MIT Press, Cambridge/London (1990)
 57. Li, S.Z.: *Markov Random Field Modeling in Computer Vision*. Springer, New York (1995)
 58. Luminati, D.: Surfaces with assigned apparent contour. *Ann. Sc. Norm. Super. Pisa Cl. Sci.* **3**, 311–341 (1994)
 59. Mancini, S., Ruas, M.A.S.: Bifurcations of generic parameter families of functions on foliated manifolds. *Math. Scand.* **72**, 5–19 (1993)
 60. Millett, K.C.: Generic smooth maps of surfaces. *Topology Appl.* **18**, 197–215 (1984)
 61. Mond, D.: On the classification of germs of maps from R^2 to R^3 . *Proc. Lond. Math. Soc.* **50**, 333–369 (1985)
 62. Morel, J.-M., Solimini, S.: *Variational Methods in Image Segmentation*. Progress in Nonlinear Differential Equations and Their Applications, vol. 14. Birkhäuser, Boston (1995)
 63. Mumford, D., Desolneux, A.: *Pattern Theory: The Stochastic Analysis of Real-World Signals*. A.K. Peters Ltd., Natick (2010)
 64. Murasugi, K.: *Knot Theory and Its Applications*. Birkhäuser, Boston (1996)
 65. Nitzberg, M., Mumford, D.: The 2.1-D sketch. In: *Proceedings of the 3rd International Conference on Computer Vision*, Osaka, 1990
 66. Nitzberg, M., Mumford, D., Shiota, T.: *Filtering, Segmentation and Depth*. Lecture Notes in Computer Science, vol. 662. Springer, Berlin (1993)
 67. Ohmoto, T., Aicardi, F.: First order local invariants of apparent contours. *Topology* **45**, 27–45 (2006)
 68. Osher, S.J., Fedkiw, R.P.: *Level Set Method and Dynamic Implicit Surfaces*. Applied Mathematical Sciences, vol. 153. Springer, New York (2003)
 69. Paragios, N., Chen, Y., Faugeras, O. (eds.): *Handbook of Mathematical Models in Computer Vision*. Springer, New York (2006)

70. Petitot, J.: *Neurogéométrie de la Vision - Modèles Mathématiques et Physiques des Architectures Fonctionnelles*. Les Editions de l'École Polytechnique, Paris (2008)
71. Reidemeister, K.: *Knotentheorie*. Springer, New York (1932/1974)
72. Rieger, J.H.: On the complexity and computation of view graph of piecewise smooth algebraic surfaces. *Philos. Trans. R. Soc. Lond. A* **354**, 1899–1940 (1996)
73. Rolfsen, D.: *Knots and Links*. AMS Chelsea Publishing, Providence (2003)
74. Roseman, D.: Reidemeister-type moves for surfaces in four dimensional space. In: Jones, V.F.R. (ed.), et al. *Knot Theory. Proceedings of the Mini-Semester, Warsaw, Poland 1995*, vol. 42, pp. 347–380. Banach Center Publications, Warszawa (1998)
75. Sapiro, G., Tannenbaum, A.: Affine invariant scale-space. *Int. J. Comput. Vision* **11**, 25–44 (1993)
76. Sato, J., Cipolla, R.: Affine reconstruction of curved surfaces from uncalibrated views of apparent contours. *IEEE Trans.* **21**, 1188–1198 (1999)
77. Scharstein, D.: *View Synthesis Using Stereo Vision*. Lecture Notes in Computer Science, vol. 1583. Springer, New York (1999)
78. Shirai, Y.: *Three-Dimensional Computer Vision*. Springer Series Symbolic Computations. Springer, New York (1987)
79. Thom, R.: Les singularités des applications différentiables. *Ann. Inst. Fourier* **6**, 43–87 (1956)
80. Thom, R.: *Stabilité Structurelle et Morphogénèse*. W.A. Benjamin, Reading (1972)
81. Toennies, K.D.: *Guide to Medical Image Analysis. Methods and Algorithms*. Springer, New York (2012)
82. Vaillant, R.: Using occluding contours for 3D object modeling. In: *The 1st European Conference on Computer Vision (ECCV'90)*, Antibes (France), pp. 454–464. Springer, New York (1990)
83. Vaillant, R., Faugeras, O.D.: Using extremal boundaries for 3D object modeling. *IEEE Trans. Pattern Anal. Mach. Intell.* **14**, 157–173 (1992)
84. Wall, C.T.C.: Lectures on C^∞ stability and classification. In: *Proceedings of Liverpool Singularities I*. Springer Lecture Notes in Mathematics, vol. 192, pp. 178–206. Springer, Berlin (1971)
85. Wassermann, G.: Stability of unfoldings in space and time. *Acta Math.* **135**, 57–128 (1975)
86. Whitney, H.: On singularities of mappings of Euclidean spaces. I. Mappings of the plane into the plane. *Ann. Math.* **62**, 374–410 (1955)
87. Williams, L.R.: *Perceptual completion of occluded surfaces*. Ph.D. Dissertation, Department of Computer Science, University of Massachusetts, Amherst (1994)
88. Williams, L.R.: Topological reconstruction of a smooth manifold-solid from its occluding contour. *Int. J. Comput. Vision* **23**, 93–108 (1997)
89. Younes, L.: *Shapes and Diffeomorphisms*. Applied Mathematical Sciences, vol. 171. Springer, Berlin (2010)

Chapter 1

A Variational Model on Labelled Graphs with Cusps and Crossings

In this chapter we review some of the variational models appearing in the mathematical literature of image segmentation. We will mainly focus attention on those models related to the problem of reconstructing a notion of order between the various objects in a three-dimensional scene. Next, we describe a variational model [14]¹ that can be considered as one of the motivations for the topological study of the apparent contours and of three-dimensional shapes made in the sequel of the book.

1.1 The Reconstruction Problem

Let $\Omega \subset \mathbb{R}^2$ be a Lipschitz bounded simply connected open set, typically $\Omega = (0, 1) \times (0, 1)$, let $I := (0, 1)$ and $\bar{I} = [0, 1]$ be the closure of I . We assume to have given an *image*, mathematically described as a function

$$g : \Omega \rightarrow \bar{I},$$

containing information on the grey-level intensity. The function g gives the brightness at each point of the plane domain Ω , and it is discontinuous along curves which correspond to sudden changes in the visible surfaces. Conventionally, we assume regions where g takes the value one (respectively zero) to correspond to the black (respectively white) colour. The final aim, which is also one of the main issues in

¹With kind permission from Springer Science+Business Media, in this chapter and in Chap. 11 we illustrate the results and report some of the figures from the quoted paper [14].

computer vision, would be to reconstruct a three-dimensional scene (a 3D-scene, or a scene for short)

$$E \subset Q := \Omega \times I \subset \mathbb{R}^3 \quad (1.1)$$

corresponding to g . By definition, a 3D scene E is the union of a finite number of connected closed pairwise disjoint solid subsets of Q with smooth boundary and not intersecting the boundary ∂Q (see Definition 3.1.1). We shall clarify in the next paragraphs what we mean by a 3D scene corresponding to g , on the basis of a variational principle.

In order to reconstruct a 3D scene E , a first problem that one has to face is the so-called *segmentation problem*, which consists in decomposing the image into regions and contours corresponding to meaningful parts of the objects of E . This decomposition is described through a pair (u, K) : the intensity function (or grey level)

$$u : \Omega \rightarrow \mathbb{R},$$

required to be as uniform as possible inside each region, with sharp transitions K , taking place across the boundaries of the regions, or on fractures interior to some region.² Such transitions are usually called discontinuity arcs. Any discontinuity arc is an intensity edge and the related problem of edge detection³ looks for the location of the sharp transitions in intensity. Edge detection requires a further linking process of the edges into global curves to achieve a segmentation. One of the advantages of the *variational approach* to image segmentation is that it unifies edge detection and linking into a single minimization process. In Sects. 1.2–1.4 we shall describe some⁴ of the approaches recently proposed in the literature, that aim to solve the segmentation problem by minimizing suitable action functionals, and to provide an optimal segmentation (u, K) of g (uniqueness of an optimal segmentation, in general, cannot be expected). These models, in particular the Nitzberg–Mumford one, can be considered as the starting point for the variational model based on the apparent contours, which is described in Sect. 1.5 and analysed in Chap. 11.

Once an optimal segmentation (u, K) has been obtained, the next step to face is the *reconstruction problem*, namely to reconstruct a three-dimensional scene E starting from (u, K) . In this process the outline of the scene (or more precisely the visible part of the apparent contour of the scene, as we shall see) should

²In the context of what we shall call “visible contours”, there are (arcs) fractures with terminal points ending inside a region; these arcs are not part of the boundary of a segmentation.

³The problem of edge detection is extensively studied in computer vision; see, for instance, [24, 45, 56, 67].

⁴It is not the aim of this chapter to give a complete overview on this argument. We refer the reader, for instance, to [3, 27, 53, 60, 66, 72, 79] for some of the topics that are not treated here, and for a more complete list of references.

correspond to the discontinuity curves K of the optimal segmentation. Observe that discontinuity curves arising from solid bodies partially or totally occluded are necessarily lost in the two-dimensional picture. Therefore the *completion problem* arises, namely the process of completing K into a new set of curves representing the projection of the whole boundary of the 3D scene. In the reconstruction of the occluded arcs, a priori information on the topology of the scene could in principle be used as a constraint. Since the final aim is to provide the scene, or at least its *topology*, information on the depth of the various solid objects must be contained in the action functional to be minimized. In a sense (that is made precise) this information appears in the Nitzberg–Mumford model. In a different way, in the variational model described in Sect. 1.5, this information is provided by the *Huffman labelling* (labelling,⁵ for short) which will enter in the domain of the action. For convenience of the reader, in Sect. 1.5 we make more explicit the relations of the variational model with the completion problem studied in Chap. 4, and with the reconstruction of the topology of the 3D scene studied in Chap. 5.

Remark 1.1.1 (On Roundedness of the 3D Shape) In this book we will not be concerned with the reconstruction of the “best shape” (i.e., the most round in some sense) that a scene may have, for instance for a given topology. Accordingly, we shall call *depth-equivalent* those scenes that differ by a homeomorphism of Q which preserves the fibres of the orthogonal projection of ∂E on Ω and the order of the points on each fibre.⁶ Then, we shall be concerned here only with the reconstruction of the depth-equivalence class $[E]$ of a scene E . We remark that the various quantities introduced in Sect. 1.5 are independent of the choice of the representative in the depth-equivalence class of E .

Before starting with a more detailed discussion, we advise the reader about some simplifications that are shared by all models that we are going to describe. These models, indeed, do not take into account what we could call “spurious” discontinuity sets, due, for instance, to

- shadows,
- patterns on the surfaces of the solid objects, caused by the presence of different materials, or by optical properties of the surfaces themselves,
- corners or edges. These however were already excluded by the smoothness requirement on E .

⁵It is worthwhile to observe that a contour may be, in general, endowed with several different labellings.

⁶Basically, two scenes E and F are depth-equivalent if they consist of the same number of connected components, and each connected component F_i of F is obtained from the corresponding connected component E_i of E through a strictly monotone map in the view direction, continuously depending on the position in Ω . In particular, if it happens that E_i is in front of the connected component E_j , then the same depth ordering is preserved for the corresponding connected components F_i and F_j of F . See Definition 5.1.2 for the details.

Concerning shadows, we shall in particular suppose that the illumination direction coincides with the projection direction of the scene on Ω .

Finally, we will not discuss the dependence of the minimizing configurations on the various possible choices of the parameters appearing in the action functionals.

1.2 The Mumford–Shah Model

The variational approach to the image segmentation problem yields a single nonlinear process, by simultaneously placing the discontinuity set K and smoothing the image only out of the boundaries. In the Mumford–Shah action functional [62], the segmentation problem is formulated as the L^2 -approximation of the given image g by means of a piecewise smooth function u , the jump of which consists of the discontinuity curves separating the regions of approximately constant, or possibly smooth, intensity. This is achieved by considering the functional

$$\text{MS}(u, K) := \int_{\Omega} (u - g)^2 dx + \alpha \int_{\Omega \setminus K} |\nabla u|^2 dx + \beta \mathcal{H}^1(K), \quad (1.2)$$

where $K \subset \overline{\Omega}$ (the closure of Ω) is a closed set, $u \in \mathcal{C}^1(\Omega \setminus K)$, \mathcal{H}^1 denotes the one-dimensional Hausdorff measure in \mathbb{R}^2 ,⁷ and α and β are nonnegative parameters. The function u is discontinuous across the set K and, outside K , it is required to be of class \mathcal{C}^1 ; in (1.2), ∇u denotes the gradient of u in $\Omega \setminus K$. The first term (sometimes called fidelity term) in the expression of the Mumford–Shah functional, namely⁸

$$\int_{\Omega} (u - g)^2 dx, \quad (1.3)$$

forces u to be close to g in the L^2 -distance. In order to make nontrivial the minimization of (1.3), some regularizing term must be added. The second term in the functional favours u to be smooth, requiring the squared L^2 gradient norm to be finite. At the same time, the last term penalizes large sets K in order to avoid a too fragmented segmentation. The variational formulation, namely the infimization of the functional $\text{MS}(u, K)$ over all closed sets $K \subset \overline{\Omega}$ and over all $u \in \mathcal{C}^1(\Omega \setminus K)$, overcomes the inconvenients of separate smoothing and detection processes, at the cost of an increased computational complexity. As already discussed in Sect. 1.1, the function u represents a denoised approximation of the image g , and K represents the set of contours of the segmentation. Because of the presence of the length term

⁷That is, $\mathcal{H}^1(K)$ is the length of K when K is sufficiently smooth, see [43].

⁸Other norms different from the L^2 -norm in (1.3) have been considered in the literature; also, suitable functions of u different from the identity can be taken into account: we refer the reader, for instance, to [44, 52] and the references therein.

$\mathcal{H}^1(K)$, this minimization process is not a straightforward matter. The proof of existence of minimizers is obtained by resorting to a weak formulation: first the existence of a weak solution is proved, then the solution of the original problem is obtained by proving regularity properties of the weak solution [34]. Weak solutions are looked for in a class of discontinuous functions introduced by De Giorgi and Ambrosio in [33], and denoted by $SBV(\Omega)$.⁹ In this way, the set K is now implicitly defined as the discontinuity set (or jump set) J_u of u , so that the weak form of the functional depends only on the function u and becomes

$$MS(u) = \int_{\Omega} (u - g)^2 dx + \alpha \int_{\Omega \setminus J_u} |\nabla u|^2 dx + \beta \mathcal{H}^1(J_u).$$

We refer the reader to the book [2] for the details and a list of related references. Mumford and Shah studied the properties of minimizers of the functional (1.2) under some simplifying assumptions. They proved that the nodes of K can only be either triple points where three curves meet with *equal angles*, or points of the boundary of Ω where one curve meets the boundary perpendicularly, or crack tips where a curve ends and meets nothing. See also [26, 32], and the books [31, 60] for more information.¹⁰

From the above results, however, it follows that relevant features for pattern recognition, such as T-junctions due to occlusions, are not allowed in a configuration minimizing the Mumford–Shah functional. Indeed, due to the presence of the length term, T-junctions are locally deformed in an unnatural triple junction at 120° angles.¹¹ One possibility that has been explored¹² is to add to the action functional a term depending on the curvature κ of the contour. Indeed, it is known

⁹This is a subspace of the space $BV(\Omega)$ of functions with bounded variation in Ω , and it is called the space of special functions of bounded variation in Ω . An example of a function in $SBV(\Omega)$ is given by the characteristic function of a finite perimeter set in Ω .

¹⁰We recall also the variational model where the total variation $\alpha \int_{\Omega} |Du|$ is considered, in place of the terms $\alpha \int_{\Omega \setminus J_u} |\nabla u|^2 dx + \beta \mathcal{H}^1(J_u)$; see [70]. The advantage of this model (originally introduced in the context of image denoising) is that the functional involved is convex; the disadvantage is its lackness of differentiability at zero and its linear growth at infinity. The discontinuities recovered by the total variation method appear less sharp with respect to the ones recovered by the Mumford–Shah functional (and seems not to be suited for the reconstruction of T-junctions). The total variation model has found (in one variant or another) many applications as a tool to compute the minimum of geometric functionals, in surface reconstruction, in the development of more sophisticated anisotropic total variation models, as a test example to develop efficient numerical schemes for nonlinear and non-differentiable functionals, or as inspiration for edge preserving regularizers. See also the book [3] for related questions and references.

¹¹Similarly, corners in case of nonsmooth shapes are smoothed out: however, in this book we will never be concerned with nonsmooth (polyhedral, for instance) 3D scenes. Apart from the discussion related to the functional in (1.5), all contours that we shall consider will be without corners.

¹²We are not concerned here with functionals depending on the Hessian of u , see [18] for more information.

that including curvature integrals in variational segmentation models allows a more accurate reconstruction of the singularities along the visible contours [30, 71]. Such a new term is also often used to solve the completion problem for occluded contours: for instance, in [47, 61] it is suggested the study of the elastica functional

$$\int (\beta + \phi_2(\kappa)) d\mathcal{H}^1, \quad \text{where } \phi_2(\xi) := \lambda \xi^2 \text{ for any } \xi \in \mathbb{R},$$

λ being a nonnegative parameter, and the integration being performed over the missing boundaries. See also [1, 5, 6]. This idea has been pursued by Nitzberg and Mumford in [64], who proposed a segmentation model which also allows regions to overlap, in order to take into account the partial occlusion of farther away objects by those nearer. See also the book [65] of Nitzberg, Mumford and Shiotani,¹³ where the reader can find a discussion on the segmentation problem with depth, on the role of T-junctions in the human visual system and on the psychology of continuation of contours, together with a list of related references.

1.3 The Nitzberg–Mumford Model

The Nitzberg–Mumford model, called the 2.1D sketch, is a variational model aiming to give a reasonable solution to the segmentation problem described in Sect. 1.1, taking into account the depth ordering between different objects, and suited for possibly restoring by completion the hidden parts of incomplete regions, with arcs of sufficiently smooth (and not necessarily rectilinear) curves.

Let $\text{Part}_{\text{overlap}}(\mathbb{R}^2)$ be the set of all finite families S of possibly overlapping closed subsets of \mathbb{R}^2 with nonempty connected interior, such that each $R \in S$ has boundary ∂R of class \mathcal{C}^2 , and

$$\bigcup_{R \in S} R = \overline{\Omega}.$$

Given $S \in \text{Part}_{\text{overlap}}(\mathbb{R}^2)$ and a partial order relation \prec on S , we enumerate the elements of S using \prec , writing S as (R_1, \dots, R_n) . The relation \prec can be interpreted as a relative depth: $R_i \prec R_j$ (or equivalently $i \prec j$) means that “ R_i is closer than R_j to the observer”, i.e., R_i occludes R_j . The sets R'_i defined as

$$\begin{aligned} R'_1 &:= R_1, \\ R'_i &:= R_i \setminus \bigcup_{j=1}^{i-1} R_j \quad \text{for } i \in \{2, \dots, n\} \end{aligned}$$

¹³In this reference the Nitzberg–Mumford model was developed further, together with a related computer algorithm which, however, does not implement a direct minimization of the functional NM in (1.4).

are therefore the “visible” parts of the objects (with respect to the choice of \prec), namely, for any $i \in \{1, \dots, n\}$ the set R'_i is interpreted as the visible part of R_i .

Now, let us denote by $\phi_{\text{NM}} : \mathbb{R} \rightarrow (0, +\infty)$ an even convex function vanishing quadratically at the origin, and linearly increasing at infinity with slope λ . The Nitzberg–Mumford functional [64, 65] has the following expression:

$$\begin{aligned} \text{NM}(S, \prec) &:= \text{NM}((R_1, \dots, R_n)) \\ &:= \sum_{i=1}^n \int_{R'_i} (g'_i - g)^2 dx + \delta \sum_{i=1}^n |R_i| + \sum_{i=1}^n \int_{\partial R_i \setminus \partial \Omega} (\beta + \phi_{\text{NM}}(\kappa)) d\mathcal{H}^1, \end{aligned} \quad (1.4)$$

where g'_i is the mean value of g on R'_i , $|R_i|$ is the area of R_i , δ is a nonnegative parameter, and κ is the curvature of the boundary of each region. Notice that:

- the first term in the expression of the functional NM depends on the order \prec (while the second and the third one do not). It encourages partitions with visible brightness g'_i close to the data.
- The second term is a penalization on the total area of the regions R_i , so that too many overlappings are penalized.¹⁴
- The last term is a penalization on the length and the curvature of all contours, which are therefore required to be short and wiggle as little as possible. The linear growth of ϕ_{NM} at infinity serves to assign a finite energy contribution to a corner of the contour, proportional to the change in tangent direction. Indeed, if the boundary of a set R in the partition is not smooth, but just piecewise smooth, the curvature depending term in (1.4) is defined as follows. Write ∂R as the image of a map γ using the arclength $s \in [0, l(\gamma)]$, where $l(\gamma) = \mathcal{H}^1(\partial R)$, and let θ denote the angle between the tangent to γ and the positive direction of the first axis. Assuming that θ is of class C^1 out of a finite set of points C , then the term $\int_{\partial R \setminus \partial \Omega} \phi_{\text{NM}}(\kappa) d\mathcal{H}^1$ is replaced by

$$\int_{[0, l(\gamma)] \setminus C} \phi_{\text{NM}}(\dot{\theta}) ds + \lambda \sum_{s \in C} |\theta^+(s) - \theta^-(s)|, \quad (1.5)$$

where $\theta^+(s)$, $\theta^-(s)$ are the two one-sided limits of θ at $s \in C$.

- The reason of the presence of the mean values g'_i is the following. Suppose to fix the pair (S, \prec) and to consider the right-hand side of (1.4) as a function of the constants g'_1, \dots, g'_n . Then, this function is minimized for the choices $g'_i = \frac{1}{|R'_i|} \int_{R'_i} g dx$ for any $i = 1, \dots, n$.

¹⁴If δ is too large, a minimizing configuration could, in principle, destroy a T-junction, transforming it into a smoothed corner.

Remark 1.3.1 A slight modification of NM consists in taking a finite ordered family (R_1, \dots, R_m) of closed subsets of \mathbb{R}^2 with nonempty connected interior and boundary of class \mathcal{C}^2 , defining $R'_1 := R_1$, $R'_i := R_i \setminus \bigcup_{j=1}^{i-1} R_j$ for $i \in \{2, \dots, m\}$, $R'_{m+1} := \Omega \setminus \bigcup_{j=1}^m R_j$ (the background) and then defining the action functional on the family \tilde{S} of these partitions, as

$$\begin{aligned} & \widetilde{\text{NM}}((R_1, \dots, R_m, R'_{m+1})) \\ &:= \sum_{i=1}^{m+1} \int_{R'_i} (g'_i - g)^2 dx + \delta \sum_{i=1}^m |R_i| + \sum_{i=1}^m \int_{\partial R_i \setminus \partial \Omega} (\beta + \phi_{\text{NM}}(\kappa)) d\mathcal{H}^1, \end{aligned} \quad (1.6)$$

where, again, g'_i is the mean value of g on the visible part R'_i of R_i . In this way, the background is not penalized. This modification corresponds to impose $R_n = \overline{\Omega}$ in the Nitzberg–Mumford model (and $n = m + 1$).

The Nitzberg–Mumford model represents a scene in terms of overlapping ordered finite families of subsets of Ω . An ordered family which minimizes NM is called an *optimal overlapping segmentation* of g . In such a segmentation possibly new contours are introduced,¹⁵ interpreted as the continuation of partly occluded contours. Interestingly, besides a segmentation of g , a configuration minimizing NM carries an order information among the various regions, so that the model identifies a sort of notion of relative depth between the objects, which tells us which region is in front and which region is in back. In addition, the model is capable to often reconstruct correctly¹⁶ the T -junctions, differently with respect to the Mumford–Shah model.

Example 1.3.2 (Partially Overlapping Disks) Take $\Omega = (0, 1) \times (0, 1)$ as usual, and

$$\begin{aligned} B^- &:= \left\{ x = (x_1, x_2) \in \Omega : \left(x_1 - \frac{3}{8} \right)^2 + x_2^2 \leq \frac{1}{16} \right\}, \\ B^+ &:= \left\{ x = (x_1, x_2) \in \Omega : \left(x_1 - \frac{5}{8} \right)^2 + x_2^2 \leq \frac{1}{16} \right\}. \end{aligned}$$

Choose a piecewise constant function g as follows: g equals to a constant in B^- and another constant in $B^+ \setminus B^-$, and $g = 0$ in $\Omega \setminus (B^- \cup B^+)$. If we consider the family $\{B^-, B^+, \Omega \setminus (B^- \cup B^+)\}$ with the order relation so that

$$R_1 = B^-, \quad R_2 = B^+, \quad R'_3 = \Omega \setminus (B^- \cup B^+),$$

¹⁵Thus giving a first guess on the completion of the occluded contours.

¹⁶Substituting ϕ_2 in place of ϕ_{NM} does not modify this positive feature of the model.

we have: $R'_1 = R_1$, $R'_2 = R_2 \setminus R_1$, and

$$\begin{aligned} \widetilde{\text{NM}}((R_1, R_2, R'_3)) &= \int_{B^-} (g'_1 - g)^2 dx + \int_{B^+ \setminus B^-} (g'_2 - g)^2 dx + \delta |B^- \cap B^+| \\ &\quad + \int_{\partial B^-} (\beta + \phi_{\text{NM}}(\kappa)) d\mathcal{H}^1 + \int_{\partial B^+} (\beta + \phi_{\text{NM}}(\kappa)) d\mathcal{H}^1 \\ &= \delta |B^- \cap B^+| + \int_{\partial B^-} (\beta + \phi_{\text{NM}}(\kappa)) d\mathcal{H}^1 \\ &\quad + \int_{\partial B^+} (\beta + \phi_{\text{NM}}(\kappa)) d\mathcal{H}^1, \end{aligned}$$

since $g'_1 = g$ in B^- and $g'_2 = g$ in $B^+ \setminus B^-$. On the other hand, if we take the order relation so that

$$S_1 = B^+, \quad S_2 = B^-, \quad S'_3 = \Omega \setminus (B^- \cup B^+),$$

we have $S'_1 = S_1$, $S'_2 = S_2 \setminus S_1$,

$$\begin{aligned} \widetilde{\text{NM}}((S_1, S_2, S'_3)) &= \int_{B^+} (g'_1 - g)^2 dx + \int_{B^- \setminus B^+} (g'_2 - g)^2 dx + \delta |B^- \cap B^+| \\ &\quad + \int_{\partial B^-} (\beta + \phi_{\text{NM}}(\kappa)) d\mathcal{H}^1 + \int_{\partial B^+} (\beta + \phi_{\text{NM}}(\kappa)) d\mathcal{H}^1 \\ &= \int_{B^+} (g'_1 - g)^2 dx + \delta |B^- \cap B^+| + \int_{\partial B^-} (\beta + \phi_{\text{NM}}(\kappa)) d\mathcal{H}^1 \\ &\quad + \int_{\partial B^+} (\beta + \phi_{\text{NM}}(\kappa)) d\mathcal{H}^1, \end{aligned}$$

and $\int_{B^+} (g'_1 - g)^2 dx$ is positive. Therefore,

$$\widetilde{\text{NM}}((R_1, R_2, R'_3)) < \widetilde{\text{NM}}((S_1, S_2, S'_3)).$$

This inequality leads to believe that, when looking for a minimizing segmentation of g , the partition (R_1, R_2, R'_3) is more favourable with respect to (S_1, S_2, S'_3) , in which case the set B^- is in front of B^+ , and the depth order would be also reconstructed.¹⁷

¹⁷However, as in the Mumford–Shah functional, in a minimizing segmentation of g corners are smoothed out due to the presence of the term measuring the length of the contours. Indeed, it is still convenient to smooth a corner appearing in the jump set of g and then reduce the length term, at the expense of slightly increasing the other terms in the functional. This implies that corners are not sharply reconstructed, which is a phenomenon that would obviously happen also if one replaces ϕ_{NM} with ϕ_2 in (1.4).

However, another possible choice is to consider the ordered partition

$$T_1 = B^-, \quad T_2 = B^+ \setminus B^-, \quad T_3' = \Omega \setminus (B^- \cup B^+),$$

where now $T_1 \cap T_2 = \emptyset$, and T_2 is not of class \mathcal{C}^2 , and therefore the expression for the functional in (1.5) is required. We have

$$\widetilde{\text{NM}}((T_1, T_2, T_3')) = \int_{\partial B^-} (\beta + \phi_{\text{NM}}(\kappa)) d\mathcal{H}^1 + \int_{\partial(B^+ \setminus B^-)} (\beta + \phi_{\text{NM}}(\kappa)) d\mathcal{H}^1 + \lambda c,$$

where c takes into account the contribution due to the two corners of $B^+ \setminus B^-$. The comparison between $\widetilde{\text{NM}}((R_1, R_2, TR_3'))$ and $\widetilde{\text{NM}}((T_1, T_2, T_3'))$ depends now on the magnitude of the various involved parameters (in particular of δ). It is worth noticing that the partition (T_1, T_2, T_3') cannot be a minimizer (even a local minimizer), since a small smoothing of the corners makes lower the value of $\widetilde{\text{NM}}$: indeed the length term decreases most with respect to the contribution of the fidelity term.

Remark 1.3.3 (Nonlocally Constant Grey Level) The model was conceived for images having a locally constant grey level, since in (1.4) the function u is assumed to be locally constant and equals g'_i on R'_i . As observed in [65, Chapter 6], it is possible to remove such an assumption, by replacing the first sum in NM with the term¹⁸

$$\sum_{i=1}^n \int_{R'_i} (u - g)^2 dx + \alpha \sum_{i=1}^n \int_{R'_i} |\nabla u|^2 dx.$$

Various properties of a weak formulation of NM were inspected in [13], based on the previous work [15] (furtherly refined in [12]) on the elastica functional.¹⁹ A numerical minimization of the Nitzberg–Mumford functional has been explored in [39].

It is worthwhile to remark that the occlusion problem in the functional (1.4) is tackled by enforcing a priori a “global” ordering between the objects, which are therefore obtained as a stack of depth ordinate objects: in this way, *self-overlappings and interwoven shapes are excluded*.

¹⁸Again, here α is a nonnegative parameter. Note that letting $\alpha \rightarrow +\infty$ forces the function u to be piecewise constant. Notice also that, referring to the modification considered in (1.6), the term to be added reads as $\sum_{i=1}^{m+1} \int_{R'_i} (u - g)^2 dx + \alpha \sum_{i=1}^{m+1} \int_{R'_i} |\nabla u|^2 dx$.

¹⁹See [16] and the references therein for the applications of the elastica functional to computer vision.

1.4 Other Curvature-Depending Functionals

Before passing to the discussion on the apparent contours, for completeness we briefly illustrate some other variational models for image segmentation containing a curvature term. We refer also to the books [60, 68] and their list of references for a discussion on related arguments.

Functionals that include curvature terms have been considered by various authors in the recent literature on image segmentation, see, for instance, [1, 4, 7, 17–20, 22, 25, 28, 29, 35–37, 40, 55, 57, 68]. In this section we limit ourselves to describe three models related to our previous discussion.

The minimization of a functional depending both on the curvature and on the total number of singularities along the curves was proposed by Terzopoulos [73] and, independently, by Anzellotti, and reads as

$$A(u, C, P) := \int_{\Omega} (u - g)^2 dx + \alpha \int_{\Omega \setminus C \cup P} |\nabla u|^2 dx + \int_C (\beta + \phi_2(\kappa)) d\mathcal{H}^1 + \sigma \#P,$$

where C is a suitable family of plane curves²⁰ in Ω , P is the set of the endpoints of the curves in C , $\#P$ is the number of points of P and σ is a suitable positive parameter. The elastica term asks, as usual, for a greater regularity of a minimizing configuration, while the penalization of the total number of endpoints encourages the sensitivity for the recognition of the nodes. Differently with respect to the Nitzberg–Mumford model, minimizing configurations of the functional A do not necessarily destroy a corner. Therefore this model is suited for a better reconstruction of the angles in the image, but it does not recognize any order due to occlusions between the various regions.

With the purposes of incorporating the occlusion information and reconstructing the hidden contours, another variational model has been proposed in [9] (see also [8, 10]). In this model the fidelity term is the usual one as in (1.3), hence it does not depend on any notion of ordering. The order dependence due to occlusions appears in another term in the functional, that we now quickly describe. Let us denote by $BV(\mathbb{R}^2, \mathbb{Z})$ the space of functions of bounded variation taking integer values. A function $\chi \in BV(\mathbb{R}^2, \mathbb{Z})$ can be written as

$$\chi = \sum_{i \in \mathcal{I}} \alpha_i \chi_{R_i}, \quad \text{with } \alpha_i \in \mathbb{Z} \setminus \{0\} \text{ for any } i \in \mathcal{I},$$

where $\mathcal{I} \subseteq \mathbb{N}$ is at most countable, $\{R_i\}_i$ is a family of possibly overlapping finite perimeter sets with $R_i \neq R_j$ for any $i, j \in \mathcal{I}$ with $i \neq j$, and χ_{R_i} denotes the characteristic function of R_i , namely $\chi_{R_i}(x) = 1$ if $x \in R_i$ and $\chi_{R_i}(x) = 0$ if $x \notin R_i$. We call the double sequence $(\alpha_i, R_i)_{i \in \mathcal{I}}$ a representation of χ .

²⁰The functional A has been studied in [30] (see also [21] for further approximation properties), to which we refer for all details.

The functional proposed in [9] reads as follows: if $u \in SBV(\Omega)$ and $\chi \in BV(\mathbb{R}^2, \mathbb{Z})$ are such that u is of class $W^{1,2}$ out of the closure of the jump set J_χ of χ ,²¹ then

$$BM(u, \chi) := MS(u) + \inf \sum_{i \in \mathcal{I}} |\alpha_i| \overline{W}(R_i),$$

where the infimum is taken over all representations (α_i, R_i) of χ , and \overline{W} denotes the L^1 -lower semicontinuous envelope of the functional

$$A \rightarrow \int_{\Omega \cap \partial A} (\beta + \phi_{NM}(\kappa)) \, d\mathcal{H}^1. \quad (1.7)$$

In this model, J_χ represents the set of all contours, and the occluded part corresponds to $J_\chi \setminus J_u$. The function u is, as usual, the denoised grey level. However, also this model is not capable to detect occluded and occluding objects at the same time. To overcome this problem, in [11] it is proposed a different variational model allowing for interwoven and self-overlapping shapes such as those in Fig. 1 in the Introduction, which reads as follows. Let us denote by \overline{W}_p the L^1 -lower semicontinuous envelope of the functional in (1.7), where ϕ_{NM} is replaced now by

$$\phi_p(\xi) := \lambda |\xi|^p, \quad \xi \in \mathbb{R}, \quad (1.8)$$

and $p > 1$. Let $n \in \mathbb{N}$ and let $\{R_i\}_{i=1,\dots,n}$ be a family of bounded subsets of \mathbb{R}^2 with $\overline{W}_p(R_i) < +\infty$ for any $i = 1, \dots, n$. Associated with $\{R_1, \dots, R_n\}$, we introduce the function ψ defined as

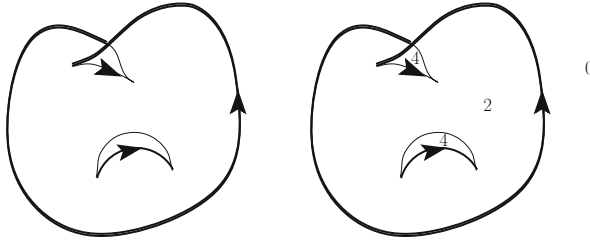


Fig. 1.1 (Left) The bold graph represents the visible part of the apparent contour of the three-dimensional scene in Fig. 3 of the Introduction; the whole graph represents its apparent contour. (Right) The values of the function f_Σ . The labelling d_Σ is not displayed here, but it can be inferred, for instance, from Fig. 3.11. Compare also with Fig. 9.11

²¹This means essentially that the jump set of u is contained in the jump set of χ .

$$\begin{cases} \psi(x) \in \{i \in \{1, \dots, n\} : x \in R_i\}, \\ \psi(x) = 0 \quad \text{if } x \notin \bigcup_{i=1}^n R_i, \end{cases}$$

so that several choices of the values of ψ are possible, in the intersection of two of the sets R_i . The function ψ takes the role of a local depth ordering associated with $\{R_1, \dots, R_n\}$, and the set $\{x \in \mathbb{R}^2 : \psi(x) = i\}$ represents the visible part of R_i , for any $i \in \{1, \dots, n\}$. Let $u \in BV_{\text{loc}}(\mathbb{R}^2)$ be piecewise constant. We denote by \mathcal{D} the set of all $(u, n, \{R_1, \dots, R_n\}, \psi)$ satisfying the inclusions

$$J_u \subseteq J_\psi \subseteq \bigcup_{i=1}^n \partial R_i.$$

We define the functional $\text{BM}_1 : \mathcal{D} \rightarrow [0, +\infty]$ as

$$\text{BM}_1(u, n, \{R_1, \dots, R_n\}, \psi) := \int_{\mathbb{R}^2} (u - g)^2 dx + \sum_{i=1}^n \overline{W}_p(R_i).$$

Observe that the dependence of BM_1 on ψ appears in the domain \mathcal{D} . A comparison between this model and the Nitzberg–Mumford model will appear in [11].

We conclude this section by recalling

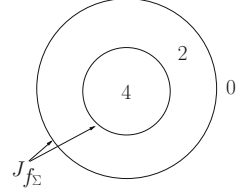
- the perspective taken in [68, 69], where the mathematics of the completion problem is inspired by the architecture of the visual cortex; see also [29], for a phenomenological variational model based on the elastica functional.
- the models for visual interpolation of missing contours in [38, 41, 42, 46, 58, 74, 75] and the references therein.

1.5 The Variational Model on Labelled Graphs

As already observed, the occlusion problem in the functional (1.4) is tackled by enforcing a priori a “global” ordering on the objects,²² which are then considered basically as flat silhouettes at constant distance from the observer. This excludes situations where two objects overlap in opposite order in different locations as in Fig. 1, or where a single body self-overlaps, like the knotted solid torus in Fig. 2. Therefore, the Nitzberg–Mumford model can represent pictures of interposed objects that are neither woven nor self-overlapping. This limitation can be considered as the main motivation for the authors of [14] to introduce the variational model on labelled graphs, that we want now to describe. As we shall see, this model is

²²See also [65, Chapter 6].

Fig. 1.2 The values of f_Σ do not identify a three-dimensional scene



strictly tied to the reconstruction problem of the 3D-scene E described in Sect. 1.1, and it requires a topological analysis on apparent contours, an argument covered in various chapters of this book. Indeed, declaring the domain $\text{Dom}(\mathcal{F})$ of the action functional \mathcal{F} requires an understanding of the structure of a labelled apparent contour. More precisely, it requires the solution of an inverse problem which is topological in character, namely the characterization of those planar graphs which are apparent contours of some three-dimensional stable scene in a given projection direction. This problem is called the *recovery problem* in computer vision, and aims to bridge the gap between three-dimensional scenes and two-dimensional contours.

Let us start with a quick description of the apparent contour²³ of the boundary Σ of a three-dimensional scene E , referring to Chap. 3 for the details. Writing a point of the set Q in (1.1) as (x_1, x_2, z) , let us consider the orthogonal projection of Σ onto the plane $\{z = 0\}$. We will interpret the direction e_3 , orthogonal to this plane, as the viewing direction: ideally, the observer is situated behind the plane, while the scene is in front of the plane, as depicted in Fig. 3.1. Associated with the choice of the projection direction, we can now consider the *critical set*, defined as the set of all points of Σ where the tangent plane contains the line spanned by e_3 . Under a suitable stability assumption,²⁴ it turns out that the critical set consists of a finite number of smooth simple pairwise disjoint closed curves. Then the *apparent contour* G_Σ of Σ , defined as the orthogonal projection on $\{z = 0\}$ of the critical set, consists of the image of a finite family of maps which are *smooth immersions* of \mathbb{S}^1 in Ω up to a finite number of points corresponding to canonical (or simple) cusps; the image of each of these maps can have a finite number of self-intersections, and also two images can intersect. In any case, the intersections (called crossings) are double and transverse.²⁵ As we have anticipated in the Introduction, in Fig. 3 we show a connected three-dimensional scene, and in Fig. 4 we show the same scene, but as if it were transparent, in order to make clear the apparent contour, which in this case has four cusps and one crossing: see Fig. 1.1 (left). On the other hand, we do not display the critical set.

Now, given a point $x \in \Omega$ out of the apparent contour, we denote by $f_\Sigma(x)$ the total number of layers of Σ which are in front of the point $(x, 0)$. Then the jump

²³The reader can look through reference [23] for an introduction to apparent contours.

²⁴See Definition 3.2.1.

²⁵In computer vision, the description of the possible singularities of the visual mapping of a smooth manifold-solid onto the image plane under parallel projection can be found, for instance, in [54].

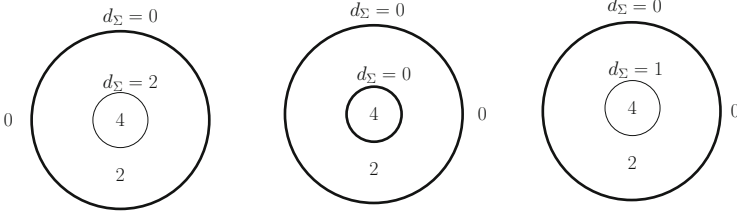


Fig. 1.3 Image taken from [14]. The values of f_Σ (the same as in Fig. 1.2) and of $d = d_\Sigma$ identify the depth-equivalence class of E . *Left*: the larger sphere is in front of the smaller one. *Centre*: the larger sphere is behind the smaller one. On the *right* the large sphere has a hole inside. Compare also with Examples 9.5.2–9.5.5

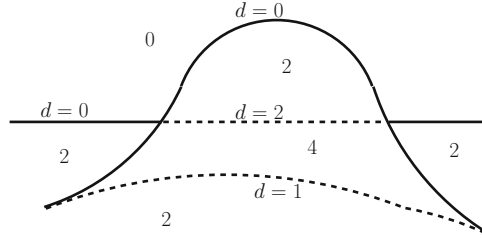


Fig. 1.4 The values of f_E and of $d = d_E$. Image taken from [14]

set J_{f_Σ} of f_Σ coincides with G_Σ , and f_Σ jumps of two units when passing from one side to the other along an arc of the apparent contour.²⁶ Moreover, J_{f_Σ} carries a natural orientation, so that the higher value of f_Σ lies locally on the left.²⁷ It is not difficult to see that the local values of f_Σ around an arc, a cusp and a crossing, respectively, are as displayed in Fig. 2.3.²⁸ In Fig. 1.1 (right) we denote the values of f_Σ for Σ as in Fig. 3.

It is worth noticing that the knowledge of f_Σ does not uniquely identify the depth-equivalence class of a three-dimensional scene, as it follows by looking at the example in Fig. 1.2. This configuration, indeed, represents three ambient isotopically inequivalent²⁹ scenes E , sharing the same values of f_Σ :

- E has two connected components: a large sphere in front of a smaller one;
- E has two connected components: a large sphere behind a smaller one;
- E is connected, and it is a sphere with a hole inside.

²⁶Notice that $f_\Sigma \in BV(\Omega, 2\mathbb{N})$, the class of all functions of bounded variation in Ω taking values in the even natural numbers (zero included).

²⁷This is the reason why, in some of the figures containing the values of f_Σ , we do not display the orientation of the apparent contour.

²⁸In Fig. 2.3 the function f_Σ is denoted by f since in the more general framework of Sect. 2.2, there is not a surface Σ embedded in \mathbb{R}^3 .

²⁹The notion of ambient isotopical equivalence of two scenes is explained in Chap. 6.

To hope uniqueness of solutions to the recovery problem up to certain transformations of \mathbb{R}^3 , it is indeed necessary to enrich G_Σ with a labelling d_Σ (sometimes called Huffman labelling [48], see also [77, 78]). Precisely, for a given point x in the relative interior of an arc, $d_\Sigma(x)$ is the number of layers of Σ in between $(x, 0)$ and the unique point on the critical set projecting on x . Namely, $d_\Sigma(x)$ gives the number of layers of Σ which are “anterior” to the point on the critical set corresponding to x .³⁰ Referring to the three possible interpretations of Fig. 1.2, we show in Fig. 1.3 the three corresponding different labellings. One can realize that f_Σ and d_Σ must satisfy a certain number of compatibility conditions, precisely the ones depicted in Fig. 3.11. It is also worthwhile to observe that the apparent contour and the two functions f_Σ and d_Σ do not depend on the choice of an element inside the depth-equivalence class of E .

In Chap. 5 we show that the labelling characterizes those graphs which are apparent contours of some three-dimensional scene. Namely, given a plane graph G the nodes of which are only cusps and crossings, enriched with a labelling d satisfying all compatibilities (called consistent labelling), there exists a smooth stable three-dimensional scene E having that labelled graph as its apparent contour, i.e., $G_\Sigma = G$ and $d_\Sigma = d$ (where, as usual, we adopt the notation $\Sigma = \partial E$). Moreover, the depth-equivalence class of E is unique. Keeping in mind this existence and uniqueness result also in connection with the reconstruction problem described in Sect. 1.1, we can finally identify the domain of the functional \mathcal{F} : $\text{Dom}(\mathcal{F})$ consists of the triplets (f, d, u) , where $f \in BV(\Omega, 2\mathbb{N})$, and

- the jump set J_f of f is an oriented graph with cusps and crossings,
- d is a compatible labelling on J_f ,
- u is in the Sobolev space H^1 out of the *visible* part $\{d = 0\}$ of J_f .

Then, given $(f, d, u) \in \text{Dom}(\mathcal{F})$, the functional \mathcal{F} reads as follows³¹:

$$\begin{aligned} \mathcal{F}((f, d, u)) := & \int_{\Omega} (u - g)^2 dx + \alpha \int_{\Omega \setminus \{d=0\}} |\nabla u|^2 dx \\ & + \int_{J_f \setminus \text{nodes}(J_f)} (\beta + \phi_p(\kappa)) d\mathcal{H}^1 + \sigma \# \text{nodes}(J_f), \end{aligned}$$

³⁰See also Fig. 3.14, which shows the labelling for Fig. 3 in the Introduction. Once we have given d_Σ , we can define the visible contour of Σ as the closure of $\{d_\Sigma = 0\}$. Its singularities are only terminal points (corresponding to cusps in the apparent contour) and T-junctions (corresponding to crossings).

³¹Chapter 11 is devoted to the mathematical study of some aspects of the functional \mathcal{F} , whose rigorous definition is given in Sect. 11.1.

where $p \in (1, 2)$,³² ϕ_p is defined in (1.8), and $\text{nodes}(J_f)$ denote the set of all cusps and crossings of J_f . Notice that:

- the fidelity term (first term in \mathcal{F}) is the same as in the Mumford–Shah functional, and therefore is independent of any occlusion ordering.
- Recall that $\{d = 0\} \subseteq J_f$ and that d depends on the values of f ; f is, in turn, determined by its jump and a suitable orientation of it (as shown in Lemma 2.2.9).

Therefore, the term $\int_{\Omega \setminus \{d=0\}} |\nabla u|^2 dx$ depends also on the orientation of J_f .

The condition $u \in H^1(\Omega \setminus \{d = 0\})$ means essentially that $J_u \subseteq \{d = 0\}$. Hence u is required to be “smooth” along the invisible part $\{d > 0\}$ of J_f . In order to decrease the action \mathcal{F} , it is therefore convenient, for a given f , to have $\{d = 0\}$ as large as possible, in order to let u jump in a larger set and consequently to reduce the value of $\mathcal{F}((f, d, u))$. The invisible part of the graph J_f is therefore “minimized”.

- The third term, which contains the curvature, is a penalization for both the visible and the invisible part of the graph. Since we assume $p > 1$, a graph with corners is excluded.
- The last term is a penalization on the total number of nodes (i.e., crossings and cusps). However, in a minimizing configuration, this does not lead to the reconstruction of corners related to the presence of polyhedral parts in the scene. The only allowed possibility happens when a corner is interpreted as a T-junction: this is exactly what happens in the Kanizsa triangle discussed below.
- We cannot in general confine ourselves to consider a piecewise constant function u out of the visible part of the apparent contour. Take, for example, the case of Fig. 1.4: assuming u to be piecewise constant out of the visible contour would imply the arcs of the apparent contour ending at a cusp not to be anymore parts of the discontinuity set of u .

Let us observe that this model has, in certain situations and with the proper choice of the parameters, a higher degree of freedom with respect to the Nitzberg–Mumford model, for what concerns the completion of the hidden contours. Consider, for instance, Fig. 1.5 (left), where we assume $g = 0$ out of the figure, $g = 1/2$ in the lower part, and $g = 1$ inside the ellipse, and two T-junctions are present. The completion of hidden contours obtained by minimizing the Nitzberg–Mumford

³²As we shall see in Chap. 5 the exponent $p = 2$ is not allowed (close to the cusps). Notice that the canonical cusp of J_f has the local form $x_2^2 = x_1^3$; hence locally around the origin on a branch of the cusp and for $x_1 > 0$, we have $\kappa(x) = \frac{3}{4} \frac{1}{x_1^{1/2} (1 + \frac{9}{4}x_1)^{3/2}}$, which belongs to L^p for $p \in [1, 2)$

but not to L^2 in a neighbourhood of the origin. It may also be useful to observe that, as in the Nitzberg–Mumford model, the values of the parameters appearing in the expression of \mathcal{F} can be related to the size of the curvature of the contour.

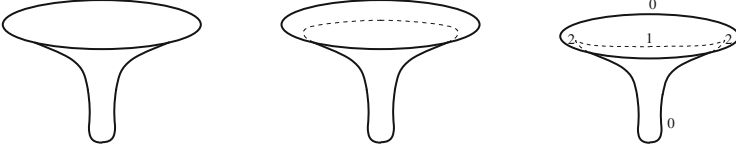


Fig. 1.5 *Centre*: the completion given by minimizing the Nitzberg–Mumford functional in (1.4), for a certain range of parameters. This configuration, as well as the one on the *right*, is admissible for the model on apparent contours. We display the values of the labelling: the integer 2 refers to the short arcs connecting a cusp to the visible part. Compare also with Example 9.5.6

model is, under certain parameters range,³³ most probably the one depicted in Fig. 1.5 (centre), which corresponds to a 3D scene consisting of *two connected components*, with the ellipsoid in front of the other connected component. On the other hand, besides this reconstruction, the model on apparent contours has another inequivalent way to resolve the 3D scene: this is shown in Fig. 1.5 (right), with $u = g$. This latter figure displays a *connected* 3D scene representing a mushroom. In this configuration (where we indicate also the values of d), there are two new nodes (the two cusps inside the ellipse) that are penalized in the functional \mathcal{F} ; however, the curvature of the invisible contour may be not so high, so that this configuration may become more favourable.

The variational model on apparent contours is more complicated than the Nitzberg–Mumford model, and allows in general for a larger number of possible topological completions of the occluded contours. However, it shares with the Nitzberg–Mumford model the limitation of computing the continuation of hidden contours using only local information on the endpoints and tangents.³⁴

It is interesting to look at the behaviour of the functional \mathcal{F} on the Kanizsa triangle [49].³⁵ Instead of the original Kanizsa triangle, we consider the picture in Fig. 1.6, where the three vertices are slightly smoothed. Such a smoothing simplifies the presentation, it does not affect our qualitative discussion, and does not affect the presence of the illusory contour. We take $g = 1$ inside the three dark regions and $g = 0$ elsewhere. We also imagine to have a very large positive parameter in front of the fidelity term, so that a minimizing configuration (f, d, u) has u very close to g .

³³Another configuration with finite action (both for NM and for \mathcal{F}) less favourable for a suitable range of parameters and also less natural, consists in splitting the set into two disjoint regions, by smoothing the two T-junctions and transforming them in two smoothed corners.

³⁴Namely, the continuation of hidden contours is computed solely from the endpoints and tangents to the visible contours at their terminal points, and not from a procedure taking also into account the global shape of the regions.

³⁵Concerning the “Gestalt school”, see, e.g., [50, 51, 59, 76].

Fig. 1.6 Smoothed Kanizsa triangle; we take $g = 1$ inside the *dark* regions and $g = 0$ outside. Image taken from [14]

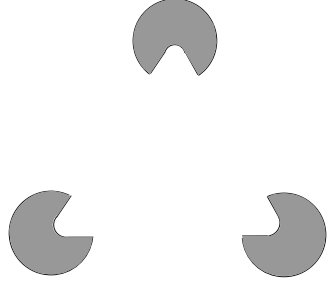
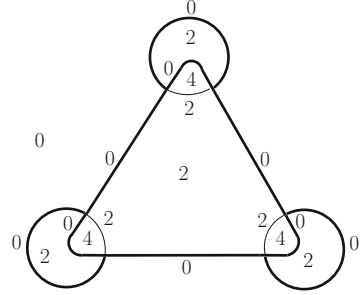


Fig. 1.7 The values of f_a and d_a . The function u_a is equal to g . Note that the visible part $\{d_a = 0\}$ of the contour consists of the whole big triangle and the union of the long arcs of the three disks. Image taken from [14]



Consider the triplet (f_a, d_a, u_a) where $u_a \equiv g$ and the values of f_a and d_a are displayed in Fig. 1.7. Note that

- the set $\{d_a = 0\}$ (visible contour) contains the three long segments composing the “illusory contour”, so that the whole big triangle with the smoothed vertices is visible and occludes part of the three circles;
- the set J_{u_a} is strictly contained in $\{d_a = 0\}$.

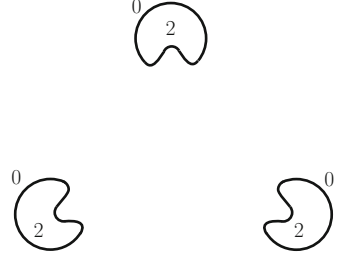
We then have

$$\mathcal{F}((f_a, d_a, u_a)) = \int_{J_{f_a} \setminus \text{nodes}(J_{f_a})} (\beta + \phi_p(\kappa)) d\mathcal{H}^1 + 6\sigma, \quad (1.9)$$

where J_{f_a} is the whole apparent contour in Fig. 1.7 and $\text{nodes}(J_{f_a})$ are the six crossings, which are penalized.

The optimality of this configuration in terms of \mathcal{F} obviously depends on the choice of the parameters. For instance, another admissible triplet (f_b, d_b, u_b) is obtained by slightly smoothing the six crossings, as displayed in Fig. 1.8, so that they do not contribute anymore to the action. The function u_b is equal to 1 inside the three smoothed regions and 0 outside, and the values of f_b and d_b are depicted in Fig. 1.8, so that the three smoothed regions are visible. Differently with respect to u_a , in this case u_b is not exactly equal to g , and this gives a nonzero contribution

Fig. 1.8 The values of f_b and d_b . The six corners of Fig. 1.6 are smoothed, and this has a cost in terms of the curvature of the contour. The function u_b is not identically equal to g . Image taken from [14]



in the expression of $\mathcal{F}((f_b, d_b, u_b))$. Notice also that $J_{u_b} = \{d_b = 0\}$. Since the corresponding apparent contour now has no nodes, we have

$$\mathcal{F}((f_b, d_b, u_b)) = \int_{\Omega} (u_b - g)^2 dx + \int_{J_{f_b}} (\beta + \phi_p(\kappa)) d\mathcal{H}^1. \quad (1.10)$$

The largest contribution on the right-hand side of (1.10) is due to the term $\int_{J_{f_b}} |\kappa|^p d\mathcal{H}^1$ around the smoothed corners, since the exponent p is larger than one.³⁶ On the other hand, the length of J_{f_b} is now smaller than the length of J_{f_a} , since the illusory arcs are not anymore in the jump set of f_b . Finally, a small nonzero contribution is present, due to the L^2 -penalization of the difference between u_b and g , which is of the order of the area of the difference between the black regions in Fig. 1.6 and their smoothed versions.

In conclusion, if β is sufficiently small with respect to the other parameters, and if the smoothing of the corners is accurate enough, comparing together formulas (1.9) and (1.10), it is reasonable to expect that

$$\mathcal{F}((f_a, d_a, u_a)) < \mathcal{F}((f_b, d_b, u_b)),$$

and that (f_a, d_a, u_a) is a triplet minimizing \mathcal{F} .

It is worth noticing that parts of the visible contour $\{d_a = 0\}$ are present also where u_a does not jump: this happens exactly on the illusory contours, which are therefore *reconstructed*.

Summarizing, our approach to the problems illustrated in Sect. 1.1 is the following.

- (1) Given g , infimize \mathcal{F} among all admissible triplets. Assuming that the problem has at least one solutions, call

$$(f, d, u)$$

a minimizer.

³⁶For a circle S_ε of radius $\varepsilon > 0$ and κ its curvature, we have $\int_{S_\varepsilon} |\kappa|^p d\mathcal{H}^1 = 2\pi\varepsilon^{1-p}$, which diverges as $\varepsilon \rightarrow 0^+$.

- (2) Find the depth-equivalence class $[E]$ of the three-dimensional scene E such that, setting $\Sigma = \partial E$,

$$f_{\Sigma} = f \quad \text{and} \quad d_{\Sigma} = d$$

via Theorems 5.1.1 and 5.1.4.

The interesting conclusion is that a three-dimensional output $[E]$ is obtained, starting from an input represented by the two-dimensional image g .

A preliminary analysis for the study of the infimization of \mathcal{F} is made in Chap. 11.

Recognizing the 3D shape, once $[E]$ is given, a problem which concretely consists in simplifying the apparent contour by means of admissible moves corresponding to isotopies of \mathbb{R}^3 , is the content of part of the book.

We conclude this chapter by observing that a direct minimization of the functional \mathcal{F} seems to be a hard problem (and essentially the same holds for the Nitzberg–Mumford functional). If one is concretely interested in finding a minimizer of \mathcal{F} , a possible strategy could be to use a descent method along the gradient, starting from some initial configuration. Such an initial configuration could be found as follows. First one approximates g with one of the available methods in segmentation and devise, in particular, a set of curves expected to be close to the visible contour of a minimizer of \mathcal{F} . Making use of the results of Chap. 4, this segmentation of g , in particular the approximation of a visible contour, can be completed into a labelled apparent contour. Such an apparent contour can finally be used as the starting configuration for the above-mentioned gradient descent method.

References

1. Ambrosio, L., Masnou, S.: A direct variational approach to a problem arising in image reconstruction. *Interfaces Free Bound.* **5**, 63–81 (2003)
2. Ambrosio, L., Fusco, N., Pallara, D.: *Functions of Bounded Variation and Free Discontinuity Problems*. Clarendon Press/Oxford University Press, New York (2000)
3. Andreu, F., Caselles, V., Mazón, J.M.: Parabolic quasilinear equations minimizing linear growth functionals. In: *Progress in Mathematics*, vol. 223. Birkhäuser, Boston (2004)
4. Anzellotti, G.: Functions of bounded variation over non-smooth manifolds and generalized curvatures. In: *Variational Methods for Discontinuous Structures (Como, 1994)*. *Progress in Nonlinear Differential Equations and Their Applications*, vol. 25, pp. 135–142. Birkhäuser, Boston (1996)
5. Ballester, C., Caselles, V., Verdera, J.: Filling-in by joint interpolation of vector fields and gray levels. *IEEE Trans. Image Proc.* **10**, 1200–1211 (2001)
6. Ballester, C., Caselles, V., Verdera, J.: Disocclusion by joint interpolation of vector fields and gray levels. *SIAM J. Multiscale Model. Simul.* **2**, 80–123 (2003)
7. Ballester, C., Bertalmio, M., Caselles, V., Shapiro, G., Verdera, J.: Filling-in by joint interpolation of vector fields and gray levels. *IEEE Trans. Image Process.* **10**, 1200–1211 (2001)
8. Bellettini, G., March, R.: Variational properties of a model for image segmentation with overlapping regions. In: *Progress in Nonlinear Differential Equations and Their Applications*, vol. 51, pp. 9–17. Birkhäuser, Boston (2002)

9. Bellettini, G., March, R.: An image segmentation variational model with free discontinuities and contour curvature. *Math. Mod. Meth. Appl. Sci.* **14**, 1–45 (2004)
10. Bellettini, G., March, R.: Asymptotic properties of the Nitzberg-Mumford variational model for segmentation with depth. In: Figueiredo, I.N., Rodrigues, J.F., Santos, L. (eds.) *Free Boundary Problems: Theory and Applications*. International Series of Numerical Mathematics, vol. 154, pp. 75–84. Birkhäuser, Basel (2007)
11. Bellettini, G., March, R.: Asymptotic properties of a variational model for image segmentation with occlusions and interwoven shapes (in preparation)
12. Bellettini, G., Mugnai, L.: Characterization and representation of the lower semicontinuous envelope of the elastica functional. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **21**(6), 839–880 (2004)
13. Bellettini, G., Paolini, M.: Variational properties of an image segmentation functional depending on contours curvature. *Adv. Math. Sci. Appl.* **5**, 681–715 (1995)
14. Bellettini, G., Beorchia, V., Paolini, M.: Topological and variational properties of a model for the reconstruction of three-dimensional transparent images with self-occlusions. *J. Math. Imaging Vision* **32**, 265–291 (2008)
15. Bellettini, G., Dal Maso, G., Paolini, M.: Semicontinuity and relaxation properties of a curvature depending functional in 2D. *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* **20**(4), 247–299 (1993)
16. Bertalmio, M., Caselles, V., Haro, G., Sapiro, G.: PDE-based image and surface inpainting. In: *Handbook of Mathematical Models on Computer Vision*, pp. 33–61. Springer, New York (2006)
17. Bertalmio, M., Caselles, V., Masnou, S., Sapiro, G.: Inpainting. In: *Encyclopedia of Computer Vision*. Springer, Berlin (2011)
18. Blake, A., Zisserman, A.: Visual reconstruction. In: *MIT Press Series in Artificial Intelligence*. MIT Press, Cambridge (1987)
19. Boscaïn, U., Duplaix, J., Gauthier, J.-P., Rossi, F.: Anthropomorphic image reconstruction via hypoelliptic diffusion. *SIAM J. Control Opt.* **50**, 1309–1336 (2012)
20. Boscaïn, U., Chertovskih, R.A., Gauthier, J.-P., Remizov, A.O.: Hypoelliptic diffusion and human vision: a semidiscrete new twist. *SIAM J. Imaging Sci.* **7**, 669–695 (2014)
21. Braides, A., March, R.: Approximation by Γ -convergence of a curvature depending functional in visual reconstruction. *Comm. Pure Appl. Math.* **59**, 71–121 (2006)
22. Bredies, K., Pock, T., Wirth, B.: A convex, lower semi-continuous approximation of Euler's elastica energy, *SIAM J. Math. Anal.*, to appear
23. Bruce, J.W., Giblin, P.J.: Curves and singularities. In: *A Geometrical Introduction to Singularity Theory*, 2nd edn. Cambridge University Press, Cambridge (1992)
24. Canny, J.: A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **8**, 679–698 (1986)
25. Cao, F., Gousseau, Y., Masnou, S., Perez, P.: Geometrically guided exemplar-based inpainting, *SIAM J. Imaging Sciences* **4**, 1143–1179 (2011)
26. Chambolle, A.: Inverse problems in image processing and image segmentation: some mathematical and numerical aspects. In: *Mathematical Problems in Image Processing*. ICTP Lecture Notes Series, vol. 2, pp. 1–94 (December 2000). ISBN 92-95003-04-7
27. Chan, T.F., Shen, J.: *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*. SIAM, Philadelphia (2005)
28. Chan, T.F., Chang, F.H., Shen, J.: Euler's elastica and curvature-based inpainting. *SIAM J. Appl. Math.* **63**, 564–592 (2002)
29. Citti, G., Sarti, A.: A cortical based model of perceptual completion in the roto-translation space. *J. Math. Imaging Vision* **24**, 307–326 (2006)
30. Coscia, A.: On curvature sensitive image segmentation. *Nonlinear Anal.* **39**, 711–730 (2000)
31. David, G.: Singular sets of minimizers for the Mumford-Shah functional. In: *Progress in Mathematics*, vol. 233. Birkhäuser, Basel (2005)
32. Dal Maso, G., Morel, J.M., Solimini, S.: A variational method in image segmentation: existence and approximation results. *Acta Math.* **168**, 89–151 (1992)

33. De Giorgi, E., Ambrosio, L.: Un nuovo funzionale del calcolo delle variazioni. *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (8) Mat. Appl.* **82**, 199–210 (1988)
34. De Giorgi, E., Carriero, M., Leaci, A.: Existence theorem for a minimum problem with free discontinuity set. *Arch. Ration. Mech. Anal.* **168**, 195–218 (1989)
35. Duits, R., Franken, E.M.: Left-invariant parabolic evolutions on $SE(2)$ and contour enhancement via invertible orientation scores, part I: linear left-invariant diffusion equations on $SE(2)$. *Quart. Appl. Math.* **68**, 293–331 (2010)
36. Duits, R., Franken, E.M.: Left-invariant parabolic evolutions on $SE(2)$ and contour enhancement via invertible orientation scores, part II: nonlinear left-invariant diffusions on invertible orientation scores. *Quart. Appl. Math.* **68**, 255–292 (2010)
37. Duits, R., van Almsick, M.: The explicit solutions of linear left-invariant second order stochastic evolution equations on the 2D Euclidean motion group. *Quart. Appl. Math.* **66**, 27–67 (2008)
38. Elder, J.H.: Bridging the dimensional gap: perceptual organization of contour into two-dimensional shape. In: Wagemans, J. (ed.) *The Oxford Handbook of Perceptual Organization*. Oxford University Press, New York (2015). Preprint available online at <http://www.gestaltrevision.be/en/ourpublications/handbook-of-perceptual-organization/accepted-chapters>
39. Esedoglou, S., March, R.: Segmentation with depth but without detecting junctions. *J. Math. Imaging Vision* **18**, 7–15 (2003)
40. Esedoglou, S., Shen, J.: Digital inpainting based on the Mumford-Shah-Euler image model. *Eur. J. Appl. Math.* **13**, 353–370 (2002)
41. Fantoni, C., Gerbino, W.: Contour interpolation by vector-field combination. *J. Vis.* **3**, 282–303 (2003)
42. Fantoni, C., Bertamini, M., Gerbino, W.: Contour curvature polarity and surface interpolation. *Vis. Res.* **45**, 1047–1062 (2005)
43. Federer, H.: *Geometric Measure Theory*. Springer, Berlin (1968)
44. Fornasier, M., March, R., Solombrino, F.: Existence of minimizers of the Mumford-Shah functional with singular operators and unbounded data. *Ann. Mat. Pura Appl.* **192**, 361–391 (2013)
45. Geman, S., Reynolds, G.: Constrained image restoration and the recovery of discontinuities. *IEEE Trans. Pattern Anal. Mach. Intell.* **14**(3), 367–383 (1992)
46. W. Gerbino: Amodally completed angles. In: Shapiro A. and Todorovi D. (eds.), *The Oxford Compendium of Visual Illusions*. New York, Oxford University Press (in press)
47. Horn, B.K.: The curve of least energy. *ACM Trans. Math. Softw.* **9**, 441–460 (1983)
48. Huffman, D.A.: Impossible objects as nonsense sentences. In: Meltzer, B., Michie, D. (eds.) *Machine Intelligence*, vol. 6. American Elsevier Publishing Co., New York (1971)
49. Kanizsa, G.: *Organization in Vision*. Praeger, New York (1979)
50. Kanizsa, G.: *Vedere e Pensare*. Collana “Biblioteca Mulino”, Il Mulino (1991)
51. Kanizsa, G.: *Grammatica del Vedere*. Saggi su Percezione e Gestalt, Collana “Biblioteca Mulino” (1979)
52. Kim, Y., Vese, L.A.: Image recovering using functions of bounded variation and Sobolev spaces of negative differentiability. *Inverse Prob. Imag.* **3**, 43–68 (2009)
53. Koenderink, J.J.: *Solid Shape*. MIT Press, Cambridge/Massachusetts/London (1990)
54. Koenderink, J.J., van Doorn, A.J.: The singularities of the visual mapping. *Biol. Cybern.* **24**, 51–59 (1976)
55. Mantegazza, C.: *Su alcune definizioni deboli di curvatura per insiemi non orientati*. Degree Thesis, Scuola Normale Superiore di Pisa (1993)
56. Marr, D., Hildreth, E.: Theory of edge detection. *Proc. Roy. Soc. Lond. B* **207**, 187–217 (1980)
57. Masnou, S.: Disocclusion: a variational approach using level lines. *IEEE Trans. Image Proc.* **11**, 68–76 (2002)
58. Masnou, S., Morel, J.-M.: On a variational theory of image a modal completion. *Rend. Sem. Mat. Univ. Padova* **116**, 211–252 (2006)
59. Metzger, W.: *Gesetze des Sehens*. Waldeman Kramer, Frankfurt (1975)
60. Morel, J.-M., Solimini, S.: Variational methods in image segmentation. In: *Progress in Nonlinear Differential Equations and their Applications*, vol.14. Birkhäuser, Boston (1995)

61. Mumford, D.: *Elastica and computer vision*. In: Bajaj, C.L. (ed.) *Algebraic Geometry and its Applications*, pp. 491–506. Springer, New York (1994)
62. Mumford, D., Shah, J.: Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.* **42**, 577–685 (1989)
63. Natterer, F., Wübbeling, F.: *Mathematical methods in image reconstruction*. In: *SIAM Monograph on Mathematical Modeling and Computation*, Philadelphia, 2001
64. Nitzberg, M., Mumford, D.: The 2.1-D sketch. In: *Proceedings of the Third International Conference on Computer Vision*, Osaka (1990)
65. Nitzberg, M., Mumford, D., Shiota, T.: *Filtering, Segmentation and Depth*. Lecture Notes in Computer Science, vol. 662. Springer, Berlin (1993)
66. Osher, S.J., Fedkiw, R.P.: *Level set method and dynamic implicit surfaces*. In: *Applied Mathematical Sciences*, vol. 153. Springer, New York (2003)
67. Paragios, N., Chen, Y., Faugeras, O. (eds.): *Handbook of Mathematical Models in Computer Vision*. Springer, New York (2006)
68. Petitot, J.: *Neurogéométrie de la Vision—Modèles Mathématiques et Physiques des Architectures Fonctionnelles*. Les Editions de l'École Polytechnique, Paris (2008)
69. Petitot, J., Tondut, Y.: Vers une neuro-géométrie. In: *Fibration Corticales, Structures the Contact et Contours Subjectif Modaus*, Mathématiques, Informatiques et Sciences Humaines, vol. 145, pp. 4–101. EHESS, Paris (1998)
70. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D* **60**, 259–268 (1992)
71. Shah, J.: *Elastica with hinges*. *J. Vis. Commun. Image Represent.* **13**, 36–43 (2002)
72. Sherzer, O., Grasmair, M., Grossauer, G., Haltmeier, M., Lenzen, F.: *Variational methods in imaging*. In: *Applied Mathematical Sciences*, vol. 167. Springer, New York (2009)
73. Terzopoulos, D.: The computation of visible-surface representations. *IEEE Trans. Pattern Anal. Mach. Intell.* **10**, 417–438 (1988)
74. Tse, P.U.: Volume completion. *Cogn. Psychol.* **39**, 37–68 (1999)
75. van Lier, R., Gerbino, W.: *Perceptual completions*. In: Wagemans, J. (ed.) *The Oxford Handbook of Perceptual Organization*. Oxford University Press, New York (2015). Preprint available online at <http://www.gestaltrevision.be/en/our-publications/handbook-of-perceptualorganization/accepted-chapters>
76. Wertheimer, M.: *Untersuchungen zur Lehre der Gestalt, II*. *Psychol. Forsch.* **4**, 301–350 (1923)
77. Williams, L.R.: *Perceptual completion of occluded surfaces*. Ph.D. dissertation, Department of Computer Science, University of Massachusetts (1994)
78. Williams, L.R.: Topological reconstruction of a smooth manifold-solid from its occluding contour. *Int. J. Comput. Vis.* **23**, 93–108 (1997)
79. Younes, L.: *Shapes and diffeomorphisms*. In: *Applied Mathematical Sciences*, vol. 171. Springer, Berlin (2010)

Chapter 2

Stable Maps and Morse Descriptions of an Apparent Contour

In this chapter we recall the notion of stable map between two manifolds.¹ It is convenient to introduce the terminology in arbitrary dimension, and in a rather abstract setting. In Chap. 3 we will be concerned with the material collected here, in the special case of a two-manifold embedded in \mathbb{R}^3 : the reader interested in the embedded case can skip this chapter, and look directly through Chap. 3. In Chap. 6 we shall make use of the stability properties of a one-parameter family of embedded surfaces. Interesting situations for us will be when the source and target manifolds have the same dimension, and in particular the two-dimensional and the three-dimensional cases.

2.1 Stability of Maps

Given a (finite dimensional) manifold \mathcal{Z} of class \mathcal{C}^∞ without boundary, we set

$$\text{Diff}(\mathcal{Z}) := \{\chi : \mathcal{Z} \rightarrow \mathcal{Z} : \chi \text{ diffeomorphism of } \mathcal{Z} \text{ of class } \mathcal{C}^\infty\}.$$

We say that $\chi \in \text{Diff}(\mathbb{R}^n)$ is supported in $B \subset \mathbb{R}^n$ if $\chi(x) = x$ for any $x \in \mathbb{R}^n \setminus B$; χ is said to have compact support if B is bounded.

We set

$$\text{Diff}_c(\mathbb{R}^n) := \{\mathcal{F} \in \text{Diff}(\mathbb{R}^n) : \mathcal{F} \text{ has compact support}\}.$$

Notice that if $\mathcal{F} \in \text{Diff}_c(\mathbb{R}^n)$, then \mathcal{F} is a positive² diffeomorphism of \mathbb{R}^n .

¹See [23, 25, 26], the books [9, 10, 18, 24], the references quoted in [27], and also [1, 17].

²The map \mathcal{F} keeps the orientation: for instance, when $n = 2$, a positively oriented Jordan curve in \mathbb{R}^2 is mapped through \mathcal{F} into a positively oriented Jordan curve.

As it is customary in differential topology, by a closed manifold we mean a compact manifold without boundary.

Let \mathcal{X} denote a closed manifold of dimension $m \geq 1$ of class \mathcal{C}^∞ (source manifold), and let \mathcal{Y} be a manifold of dimension $n \geq 1$ of class \mathcal{C}^∞ without boundary (target manifold). We denote by $\mathcal{C}^\infty(\mathcal{X}, \mathcal{Y})$ the set of all maps of class \mathcal{C}^∞ from \mathcal{X} to \mathcal{Y} , endowed with the Whitney topology [10], [15, Section 41], [2].

Definition 2.1.1 (Equivalence) Two maps $\varphi_1, \varphi_2 \in \mathcal{C}^\infty(\mathcal{X}, \mathcal{Y})$ are \mathcal{C}^∞ left–right equivalent (briefly, equivalent) if there exist $\chi \in \text{Diff}(\mathcal{X})$ and $\eta \in \text{Diff}(\mathcal{Y})$ such that

$$\varphi_2 \circ \chi = \eta \circ \varphi_1.$$

In other words, the diagram

$$\begin{array}{ccc} \mathcal{X} & \xrightarrow{\varphi_1} & \mathcal{Y} \\ \downarrow \chi & & \downarrow \eta \\ \mathcal{X} & \xrightarrow{\varphi_2} & \mathcal{Y} \end{array}$$

is commutative.

The next definition relates the notion of equivalence with the Whitney topology.

Definition 2.1.2 (Stability) A map $\varphi \in \mathcal{C}^\infty(\mathcal{X}, \mathcal{Y})$ is smoothly stable (briefly, stable) if there exists a neighbourhood $U_\varphi \subset \mathcal{C}^\infty(\mathcal{X}, \mathcal{Y})$ of φ such that any map in U_φ is equivalent to φ .

We set

$$\text{Stable}(\mathcal{X}, \mathcal{Y}) := \{\varphi \in \mathcal{C}^\infty(\mathcal{X}, \mathcal{Y}) : \varphi \text{ is stable}\}.$$

We notice that $\text{Stable}(\mathcal{X}, \mathcal{Y})$ is an open subset of $\mathcal{C}^\infty(\mathcal{X}, \mathcal{Y})$.

In view of the important role taken by the concept of stability, it is worthwhile to illustrate it with some examples.

Example 2.1.3 (Morse Functions) Suppose that $\mathcal{Y} = \mathbb{R}$. When \mathcal{X} is closed, $\varphi \in \text{Stable}(\mathcal{X}, \mathbb{R})$ if and only if φ is a Morse function (that is, it has a finite number of critical points, and each critical point is nondegenerate) with distinct critical values³ (see, e.g., [10, Chapter 3, Proposition 2.2]).

³In the terminology of Thom, Morse functions are called correct, and Morse functions with distinct critical values are called excellent [6].

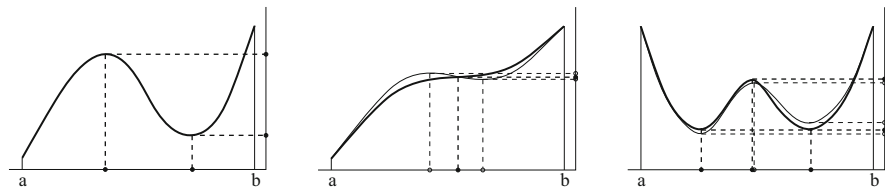


Fig. 2.1 Example 2.1.4. *Left*: a stable Morse function with two nondegenerate critical points. *Centre*: an unstable function with one degenerate critical point. A small deformation produces two critical points. *Right*: an unstable Morse function with three nondegenerate critical points

Example 2.1.4 As a special case of Example 2.1.3,⁴ suppose that $\mathcal{X} = \mathbb{R}$ and $\mathcal{Y} = \mathbb{R}$. Stability now means that the second derivative cannot vanish at stationary points (that is, critical points cannot be degenerated) and that distinct critical points cannot map onto the same value. Consider the functions in the left and right Fig. 2.1, with nondegenerate critical points in the interval (a, b) : one local maximum and one local minimum in Fig. 2.1 (left) having different critical values, and the double-well potential shaped-function with two minima at the same level and one local maximum, in Fig. 2.1 (right). Then the function in Fig. 2.1 (left) is stable; on the other hand, the one in Fig. 2.1 (right) is not stable, as well as the function in Fig. 2.1 (centre), having one horizontal inflection point. Roughly speaking, stability means that a small (in $C^\infty(\mathbb{R}, \mathbb{R})$) deformation of the function does not change the qualitative properties of its graph. It is clear that a small deformation of the function in the central picture of Fig. 2.1 alters the number of critical points, and a small deformation of the function in the right picture of Fig. 2.1 alters the number of critical values. On the other hand, for the function in Fig. 2.1 (left) it is not possible to introduce new critical points as well as to modify the local number of critical values using small deformations, since these deformations may slightly change only the horizontal or the vertical position of the critical points.

Example 2.1.5 A stable map from \mathbb{S}^1 to \mathbb{R}^2 is a smooth curve that can have a finite number of transverse self-intersections, and the number of preimages cannot be larger than two (no triple intersections). This readily generalizes to a one-dimensional closed manifold \mathcal{X} , which is the disjoint union of a finite number of copies of \mathbb{S}^1 .

Example 2.1.6 (Knots) Classical knots⁵ in \mathbb{R}^3 (respectively in \mathbb{S}^3) are stable maps from \mathbb{S}^1 to \mathbb{R}^3 (respectively to \mathbb{S}^3).

Let $\varphi \in C^\infty(\mathcal{X}, \mathcal{Y})$ and let us denote by $d\varphi$ the differential of φ .

⁴Not quite, since \mathbb{R} is not closed. However we require functions to behave “nicely” outside some interval $[a, b]$, e.g. by forcing them to have constant nonzero derivative.

⁵In this book a knot is a C^∞ embedding of \mathbb{S}^1 in \mathbb{R}^3 , hence in particular a *tame* knot in the usual terminology; see, for instance, [8, p. 5].

Definition 2.1.7 (Critical Set) The critical set of φ is defined as follows:

$$\text{crit}(\varphi) := \left\{ \xi \in \mathcal{X} : \text{rank}(d\varphi(\xi)) < \min(m, n) \right\}.$$

In the literature of singularity theory, the critical set is often also called singular set.

Notice that the critical set living on a single connected component of \mathcal{X} is not necessarily connected.

The next example, as well as Theorem 2.1.12, is of particular importance in connection with the apparent contours of embedded surfaces, which are considered in Chap. 3.

Example 2.1.8 Suppose that \mathcal{X} is two-dimensional, and let $\varphi \in \text{Stable}(\mathcal{X}, \mathbb{R}^2)$. Then, a theorem in singularity theory due to Whitney (see [25], [10, p. 191]) asserts that φ is *excellent*, namely, the following properties hold. The critical set of φ consists of folds, denoted by

$$S_1(\varphi).$$

This is the set of all points $\xi \in \mathcal{X}$ where $d\varphi(\xi)$ has rank one. $S_1(\varphi)$ is a smooth submanifold of \mathcal{X} of codimension one without boundary, hence it is the union of a finite number of smooth closed disjoint simple curves,⁶ called *fold curves*. Inside the folds, we can isolate the set of cusps,⁷ denoted by

$$S_{1_2}(\varphi) \subset S_1(\varphi).$$

This is the set of all points of $S_1(\varphi)$ where the differential of the restriction of φ to $S_1(\varphi)$ vanishes. $S_{1_2}(\varphi)$ is a smooth submanifold of codimension two, hence it is a finite set of points. The characterization of stable maps in this case is given in Theorem 2.1.12, below.

Example 2.1.9 Suppose that \mathcal{X} is two-dimensional. A classical result (see, e.g., [10, Chap. 6, Sec. 1]) asserts that a stable map $\varphi \in \mathcal{C}^\infty(\mathcal{X}, \mathbb{R}^3)$ can have (only) the following features:

- regular points: (image of) points where locally φ maps diffeomorphically onto its image;
- double curves: points of \mathbb{R}^3 having exactly two preimages;

⁶Therefore, each of these curves is the embedded image of \mathbb{S}^1 into \mathcal{X} .

⁷We warn the reader that these cusps, belonging to the source manifold \mathcal{X} , should not be confused with the cusps of an apparent contour, which lie in the target manifold \mathcal{Y} .

- pinch points (in finite number): (image of) points where the differential of φ drops rank. The standard example of a map exhibiting a pinch point is the so-called Whitney umbrella⁸;
- triple points (in finite number): points of \mathbb{R}^3 having exactly three preimages.

The next example is useful in connection with the results illustrated in Chap. 6.

Example 2.1.10 Suppose that \mathcal{X} is three-dimensional and let $\varphi \in \text{Stable}(\mathcal{X}, \mathbb{R}^3)$. Then, a theorem in singularity theory (see, e.g., [10, Chap. 7, Sec. 6]) asserts that the critical set of φ consists of folds, denoted by $S_1(\varphi)$. This is the set of all points $\xi \in \mathcal{X}$ where $d\varphi(\xi)$ has rank two. $S_1(\varphi)$ is a smooth submanifold (without boundary) of \mathcal{X} of codimension one. Inside the folds, we can isolate the pleats, denoted by $S_{1_2}(\varphi) \subset S_1(\varphi)$. This is the set of all points of $S_1(\varphi)$ where the differential of the restriction of φ to $S_1(\varphi)$ has rank one. $S_{1_2}(\varphi)$ is a smooth submanifold (without boundary) of \mathcal{X} of codimension two. Inside the pleats, we can, in turn, isolate the set of swallow's tails, denoted by

$$S_{1_3}(\varphi) \subset S_{1_2}(\varphi),$$

namely the set of all points of $S_{1_2}(\varphi)$ where the differential of the restriction of φ to $S_{1_2}(\varphi)$ vanishes. $S_{1_3}(\varphi)$ is a smooth submanifold of \mathcal{X} of codimension three, hence it is a finite set of points.

Definition 2.1.11 (Critical Value Set) Let $\varphi \in C^\infty(\mathcal{X}, \mathcal{Y})$. The critical value set of φ is defined as

$$\varphi(\text{crit}(\varphi)).$$

The critical value set, for a stable map from a closed two-dimensional manifold to the plane (also called apparent contour), will be extensively studied in the subsequent chapters. In Chap. 6 we shall consider the case of maps from a closed three-dimensional manifold into \mathbb{R}^3 (see, in particular, the stratification of \mathcal{Y} in Sect. 6.2).

The next result (see, e.g., [27] and the references therein; see also [23, pp. 61,62]) is crucial for the aims of the present book.

Theorem 2.1.12 (Equivalence to Stability in the Two-Dimensional Case) *Let \mathcal{X} be a closed two-dimensional manifold of class C^∞ , and let $\varphi \in C^\infty(\mathcal{X}, \mathbb{R}^2)$. Then φ is stable if and only if the following two conditions hold:*

- φ is excellent,
- the images of fold curves intersect only pairwise and transversally, whereas the image of any cusp does not coincide with the image of any other fold point.

⁸Sometimes called the cone on a figure-eight curve.

Moreover

- if $\xi \in \mathcal{X}$ is a fold point then, locally around ξ and $\varphi(\xi)$, there are coordinates such that φ takes the form

$$(u, v) \rightarrow (x_1, x_2) = (u^2, v);$$

- if $\xi \in \mathcal{X}$ is a cusp then, locally around ξ and $\varphi(\xi)$, there are coordinates such that φ takes the form

$$(u, v) \rightarrow (x_1, x_2) = (uv - u^3, v).$$

Example 2.1.13 (Inequivalent Maps with the Same Critical Value Set) Two maps having the same critical value set are not necessarily equivalent, as shown by the following example. Let $\mathcal{X} = \mathbb{S}^1$ and $\mathcal{Y} = \mathbb{R}$. Consider the two maps taking \mathcal{X} in \mathbb{R} , as in Fig. 2.2 (where, for convenience, the abstract manifold \mathbb{S}^1 is depicted in \mathbb{R}^2). The critical value set is in this case a finite number of (oriented) points of \mathbb{R} , and it is the same for the two maps which, however, are not equivalent. Indeed, the two correspondences between the points of the critical set and the points of the critical value set are topologically different. In the left case, the preimages of the minimal/maximal values λ_{\pm} of the singular value set are separated by critical points, while in the case on the right this is not verified.

When the dimension of \mathcal{X} is not larger than the dimension of \mathcal{Y} , we denote by

$$\text{Emb}(\mathcal{X}, \mathcal{Y})$$

the set of embeddings of \mathcal{X} into \mathcal{Y} of class C^∞ . It is possible to prove⁹ that $\text{Emb}(\mathcal{X}, \mathcal{Y})$ is open in $C^\infty(\mathcal{X}, \mathcal{Y})$.

We also recall the following result.¹⁰

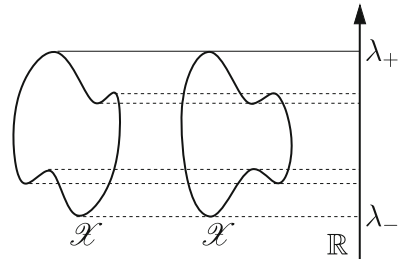


Fig. 2.2 Two maps with the same critical value set are not necessarily equivalent:
Example 2.1.13

⁹See, for instance, [10, Chapter II, Proposition 5.8].

¹⁰See [23] and [10, p. 162].

Theorem 2.1.14 (Density of Stable Maps) *Suppose that the dimensions of \mathcal{X} and \mathcal{Y} are less than or equal to three. Then $\text{Stable}(\mathcal{X}, \mathcal{Y})$ is dense in $C^\infty(\mathcal{X}, \mathcal{Y})$.*

Unless otherwise specified, all maps that we shall consider will be stable: therefore, in our discussion on apparent contours, stability will play a central role.

The case of interest in the sequel of this chapter and in Chap. 3 is that of Example 2.1.8; in the next section we shall focus on this case.

2.2 Stable Maps from a Two-Manifold to the Plane

Let us denote by M a closed manifold of class C^∞ of dimension two.¹¹ In this section we consider the case

$$\mathcal{X} = M \quad \text{and} \quad \mathcal{Y} = \mathbb{R}^2,$$

where, here and in the following, we fix the standard orientation on \mathbb{R}^2 .

Let

$$\varphi \in \text{Stable}(M, \mathbb{R}^2). \quad (2.1)$$

We specify the notation of Definition 2.1.11 in this two-dimensional case as follows.

Definition 2.2.1 (Apparent Contour) The apparent contour $\text{appcon}(\varphi)$ of φ is defined as the critical value set of φ , namely

$$\text{appcon}(\varphi) := \varphi(\text{crit}(\varphi)).$$

Notice that $\text{appcon}(\varphi)$ is nonempty; indeed, setting $\varphi = (\varphi_1, \varphi_2)$, the minimum problem $\min_{m \in M} \varphi_1(m)$ has a solution, and a minimizer is a point belonging to $\text{crit}(\varphi)$. Observe also that the critical set in a single connected component of M can give raise to an apparent contour consisting of several connected components.

Sometimes, we shall refer to $\text{appcon}(\varphi)$ as an apparent contour without labelling (or, equivalently, an unlabelled apparent contour).

Definition 2.2.2 (Component) A component of $\text{appcon}(\varphi)$ is the image through φ of a connected component of $\text{crit}(\varphi)$.

As we have seen in Theorem 2.1.12, the stability of φ ensures that the number of components of $\text{appcon}(\varphi)$ is finite, and that each component is the image of

¹¹ M is an *abstract* manifold, not necessarily oriented or connected. We shall be mostly interested (for instance, in Sect. 3.2) in the case when M can be embedded in \mathbb{R}^3 , which gives, in particular, an orientation to M .

a C^∞ map from \mathbb{S}^1 to \mathbb{R}^2 . Such a map is an immersion, up to a finite number (possibly zero) of points corresponding to *cusps*^{12,13} in the image. Each component may have a finite number of double transverse self-intersections, at points that are not cusps; furthermore, different components may have a finite number of double transverse intersections (at points that are neither cusps nor self-intersections of the same image). All such intersection points will be called *crossings* of the apparent contour of φ .

Notation: The set of all crossings of $\text{appcon}(\varphi)$ will be denoted by

$$\text{crossings}(\text{appcon}(\varphi)),$$

and the set of all cusps of $\text{appcon}(\varphi)$ will be denoted by

$$\text{cusps}(\text{appcon}(\varphi)).$$

Since $\text{appcon}(\varphi)$ is also considered as a plane graph, possibly containing closed arcs,¹⁴ it is convenient to introduce the set

$$\text{nodes}(\text{appcon}(\varphi)) := \text{cusps}(\text{appcon}(\varphi)) \cup \text{crossings}(\text{appcon}(\varphi))$$

of nodes (or vertices) of the apparent contour of φ .

Finally, we denote by

$$\text{arcs}(\text{appcon}(\varphi))$$

the set of all *relatively open* arcs of $\text{appcon}(\varphi)$; that is, a is an arc if a is a connected component of $\text{appcon}(\varphi) \setminus \text{nodes}(\text{appcon}(\varphi))$.

The arcs of φ are classified as follows:

- closed arcs (smoothly diffeomorphic to an \mathbb{S}^1);
- loops that start and end at the same node (a cusp or a crossing);
- arcs connecting two distinct nodes.

A component of $\text{appcon}(\varphi)$ corresponds to a sequence of arcs taken from $\text{arcs}(\text{appcon}(\varphi))$ and nodes taken from $\text{nodes}(\text{appcon}(\varphi))$ obtained by “glueing” arcs at cusps, and pairs of opposite arcs at crossings.

Summarizing, when $\mathcal{X} = M$ and $\mathcal{Y} = \mathbb{R}^2$, the apparent contour of the map φ has the following local structure: any point of $\text{appcon}(\varphi)$ has a neighbourhood U

¹²Locally, each cusp of the apparent contour is diffeomorphic to the simple (or ordinary, see, for instance, [3, p. 115]) cusp, which has the form $\{(x_1, x_2) : x_2^2 = x_1^3\}$ or equivalently, in a parametric form, (t^2, t^3) for a real parameter t in a neighbourhood of the origin.

¹³The component is, sometimes, called “irreducible” (with cusps and double points); see, e.g., [22].

¹⁴Compare, for instance, with Chaps. 1 (Fig. 1.3) and 4. We specify below and in Definition 4.1.1 what is a closed arc, a nonstandard feature in graph theory.

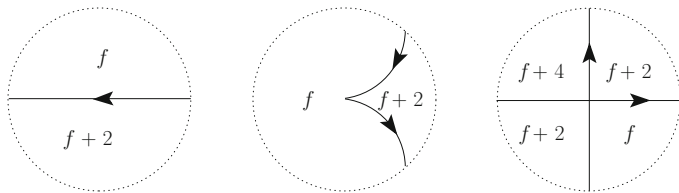


Fig. 2.3 *Left:* an arc of $\text{appcon}(\varphi)$. *Centre:* a simple cusp of $\text{appcon}(\varphi)$, i.e., the semicubic curve (t^2, t^3) . *Right:* a crossing of $\text{appcon}(\varphi)$. The labels refer to the function $f = f_\varphi$ defined in (2.2) and counting the number of preimages of φ . The apparent contour of φ is oriented according to Remark 2.2.7

in \mathbb{R}^2 such that $U \cap \text{appcon}(\varphi)$ is smoothly diffeomorphic to one of the pictures displayed in Fig. 2.3.

The complement $\mathbb{R}^2 \setminus \text{appcon}(\varphi)$ of the apparent contour is a disjoint union of a finite number of connected open sets. It is then convenient to give the following definitions.

Definition 2.2.3 (Regions) We call region of the apparent contour of φ any connected component of $\mathbb{R}^2 \setminus \text{appcon}(\varphi)$.

Definition 2.2.4 (External Region) The external region is the only unbounded region of the apparent contour of φ .

Notice that the external region is necessarily not simply connected.

We continue our description of the apparent contour of the map φ in (2.1) by introducing the function counting the number of preimages of φ . Such a function easily determines a natural orientation of $\text{appcon}(\varphi)$.

We recall that, given a set $P \subset \mathbb{R}$, we denote by $\#P$ the cardinality of P . As usual, we denote by \mathbb{N} the set of natural numbers (zero included) and by \mathbb{Z} the ring of integer numbers.

Definition 2.2.5 (The Function f_φ) We define the function $f_\varphi : \mathbb{R}^2 \rightarrow \mathbb{N}$ as follows:

$$f_\varphi(x) := \#\{m \in M : \varphi(m) = x\}, \quad x \in \mathbb{R}^2. \quad (2.2)$$

The function f_φ is finite, constant on each region of $\text{appcon}(\varphi)$, it vanishes on the external region (but *not necessarily only* in the external region), and jumps exactly on $\text{appcon}(\varphi)$. Furthermore, stability of φ implies that f_φ jumps of two units across an arc (see Theorem 2.1.12).

Remark 2.2.6 The value of f_φ on $\text{appcon}(\varphi)$ is the following:

- on an arc of $\text{appcon}(\varphi)$ it is the mean value of the two neighbouring values (an odd natural number);

- on a crossing it is the mean value of the four neighbouring values (an even natural number);
- on a cusp it is the minimum between the two neighbouring values (an even natural number).

Remark 2.2.7 (Orientation) There is a natural choice of the unit normal to the apparent contour out of the nodes; namely, the one which points toward the region where the value of f_φ is higher. Using the orientation of \mathbb{R}^2 , this determines an orientation of the tangent unit vector to the apparent contour out of the nodes, so that the higher value of f_φ is taken locally on the left; see Fig. 2.3.

We recall that the winding number $w(\gamma, x)$ of a continuous closed curve $\gamma : \mathbb{S}^1 \rightarrow \mathbb{R}^2$ with respect to a point $x \notin \gamma(\mathbb{S}^1)$ is the number of complete counterclockwise turns of the curve around x . The winding number depends on the orientation of γ , and is negative if the curve travels around x clockwise.

Definition 2.2.8 (Winding Number) The winding number

$$w(\text{appcon}(\varphi), \cdot) : \mathbb{R}^2 \setminus \text{appcon}(\varphi) \rightarrow \mathbb{Z}$$

of $\text{appcon}(\varphi)$ is defined as follows:

$$w(\text{appcon}(\varphi), x) := \sum_{\substack{\gamma : \mathbb{S}^1 \rightarrow \mathbb{R}^2 \\ \gamma(\mathbb{S}^1) \text{ component of } \text{appcon}(\varphi)}} w(\gamma, x),$$

for all $x \in \mathbb{R}^2 \setminus \text{appcon}(\varphi)$.

The function f_φ can be reconstructed from its jump set (i.e., the apparent contour) and an orientation of it. More precisely, the following observation holds.

Lemma 2.2.9 (f_φ and Winding Number) *We have*

$$f_\varphi(x) = 2 w(\text{appcon}(\varphi), x), \quad x \in \mathbb{R}^2 \setminus \text{appcon}(\varphi).$$

Proof The assertion follows by observing that the apparent contour is oriented, and that both f_φ and twice $w(\text{appcon}(\varphi), \cdot)$ satisfy the following properties:

- they are locally constant on the regions of the apparent contour,
- they vanish on the external region,
- they jump by two and with the same sign when crossing an arc of the apparent contour.

□

Remark 2.2.10 Not all oriented closed plane curves satisfying the properties described above are apparent contours of some stable map φ . For instance, it is clear that the following are necessary conditions:

- the function f_φ , as reconstructed using Lemma 2.2.9 according to the winding number, must be nonnegative in every region;
- cusps cannot be adjacent to regions with $f_\varphi = 0$. Indeed, from the local equations of φ around a cusp c given by Theorem 2.1.12, if x belongs to a sufficiently small neighbourhood of c , the points of $\varphi^{-1}(x)$ correspond to the real solutions of a cubic equation.

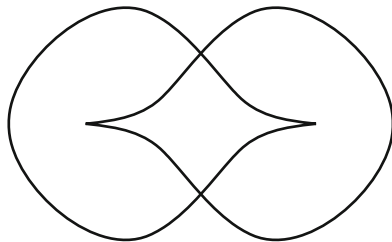
However, these two further requirements are not sufficient to identify graphs that are apparent contours of some map. For example, it can be checked that the graph shown in Fig. 3.12 (centre), the one with the “figure-eight”, satisfies the two requirements above, but cannot be obtained as an apparent contour, even in the broad context of this chapter concerning stable maps from a closed, not necessarily connected, two-manifold to \mathbb{R}^2 .

Example 2.2.11 (Haefliger Sphere) An interesting example of apparent contour, due to Haefliger¹⁵ [12], is depicted in Fig. 2.4, and corresponds to a stable map from the sphere \mathbb{S}^2 to \mathbb{R}^2 . It is worth mentioning here that this apparent contour cannot be obtained as an orthogonal projection, onto \mathbb{R}^2 , of the critical set of an immersion (self-intersections are allowed) of the sphere \mathbb{S}^2 in $\mathbb{R}^3 = \mathbb{R}^2 \times \mathbb{R}$. We shall briefly discuss this example in Sect. 10.17.1. Apparent contours of maps which factorize as an embedding (self-intersections are not allowed) and a projection will be studied in Chap. 3.

In the special case when M is *orientable*, we can distinguish positive and negative cusps¹⁶ of $\text{appcon}(\varphi)$.

Definition 2.2.12 (Positive and Negative Cusps) Let M be orientable. Let $c = \varphi(m) \in \text{cusps}(\text{appcon}(\varphi))$, where $m \in M$ is the unique preimage of φ on the *critical set* of φ . Cusp c is called *positive* (respectively *negative*) if the image of a clockwise oriented small topological circle around m goes around c clockwise (respectively counterclockwise).

Fig. 2.4 Picture taken from [12]. It represents the apparent contour of a map from the sphere \mathbb{S}^2 to \mathbb{R}^2 due to Haefliger. As shown in [12], this map cannot be factorized as an immersion in \mathbb{R}^3 and a projection onto \mathbb{R}^2



¹⁵In [12, Theorem 1] the author gives necessary and sufficient conditions for the factorization of an excellent map (see Example 2.1.8) through an *immersion* and a projection.

¹⁶The notion of positive and negative cusps of Definition 2.2.12 (see, e.g., [21]) is different from the notion considered in Chap. 8 (compare Definition 8.1.2 and Remark 8.1.3).

We shall make use of the notions of positive and negative cusps in the computational part of the book; we refer, in particular, to Sect. 10.9.2. Positive and negative cusps, in the special case of labelled apparent contours, are considered in Remark 3.4.10.

2.3 Ambient Isotopies

We start by recalling from [14, p. 178] the following concepts, adapted to our case.

Definition 2.3.1 (\mathbb{R}^n -Ambient Isotopy) We say that a map

$$H : \mathbb{R}^n \times [0, 1] \rightarrow \mathbb{R}^n$$

is an \mathbb{R}^n -ambient isotopy of class \mathcal{C}^∞ (briefly, an \mathbb{R}^n -ambient isotopy), if the following properties¹⁷ hold:

- $H \in \mathcal{C}^\infty(\mathbb{R}^n \times [0, 1], \mathbb{R}^n)$,
- $H_t \in \text{Diff}(\mathbb{R}^n)$ for any $t \in [0, 1]$, where we use the notation

$$H_t(\cdot) := H(\cdot, t),$$

- $H_0 = \text{id}$ in \mathbb{R}^n .

In this book we will be interested only in \mathbb{R}^n -ambient isotopies with compact support.

Definition 2.3.2 (Compact Support) We say that the \mathbb{R}^n -ambient isotopy H has compact support if there exists a bounded subset B of \mathbb{R}^n such that

$$H_t(x) = x, \quad x \in \mathbb{R}^n \setminus B, \quad t \in [0, 1].$$

Notice that an \mathbb{R}^n -ambient isotopy can be viewed as a path

$$t \in [0, 1] \rightarrow H_t \in \text{Diff}_c(\mathbb{R}^n)$$

with origin the identity.

Throughout the book, we shall use the symbol h to denote an \mathbb{R}^2 -ambient isotopy,¹⁸ and the symbol H to denote an \mathbb{R}^3 -ambient isotopy. When, as usual, an apparent contour $\text{appcon}(\varphi)$ lies inside $\Omega = (0, 1) \times (0, 1)$, and an \mathbb{R}^2 -ambient

¹⁷The first property means that H admits an extension of class \mathcal{C}^∞ on an open set of \mathbb{R}^{n+1} containing $\mathbb{R}^n \times [0, 1]$.

¹⁸Consistently, we set $h_t(\cdot) = h(\cdot, t)$.

isotopy h with compact support is applied to $\text{appcon}(\varphi)$, we shall tacitly assume that $h(\cdot, t)$ is the identity out of a compact set contained in Ω , for any $t \in [0, 1]$.

We shall need the following technical result.

Theorem 2.3.3 (Plane or Space Isotopies with Compact Support and Diffeomorphisms) *Let $n \in \{2, 3\}$ and $\vartheta \in \text{Diff}_c(\mathbb{R}^n)$. Then there exists an \mathbb{R}^n -ambient isotopy $h : \mathbb{R}^n \times [0, 1] \rightarrow \mathbb{R}^n$ with compact support such that*

$$h_0 = \text{id}, \quad h_1 = \vartheta.$$

Proof Suppose first $n = 2$. Let us consider \mathbb{S}^2 as the compactification of \mathbb{R}^2 with the point at infinity, denoted by ∞ . Let us extend ϑ on \mathbb{S}^2 (keeping the same symbol) as $\vartheta(\infty) := \infty$. Consider the spherical inversion $\text{inv} : \mathbb{S}^2 \rightarrow \mathbb{S}^2$ defined as

$$\text{inv}(x) := \begin{cases} \frac{x}{|x|^2} & \text{if } x \in \mathbb{R}^2 \setminus \{0\}, \\ 0 & \text{if } x = \infty, \\ \infty & \text{if } x = 0, \end{cases} \quad (2.3)$$

where $|x|$ is the euclidean norm of x .

Define

$$\tilde{\vartheta}(x) := \text{inv}(\vartheta(\text{inv}(x))), \quad x \in \mathbb{S}^2.$$

Then $\tilde{\vartheta} \in \text{Diff}^+(\mathbb{S}^2)$, where $\text{Diff}^+(\mathbb{S}^2)$ denotes the set of all positive diffeomorphisms of \mathbb{S}^2 of class C^∞ . In addition, $\tilde{\vartheta}$ is the identity in a neighbourhood of the origin which, without loss of generality, we can suppose to be the unit ball $B_1 = \{x \in \mathbb{R}^2 : |x| < 1\} \subset \mathbb{S}^2$.

From [19, Theorem 1.3] and [20],¹⁹ it follows that there exists a path $\tilde{h} : \mathbb{S}^2 \times [0, 1] \rightarrow \mathbb{S}^2$ of positive C^∞ diffeomorphisms of \mathbb{S}^2 , i.e., $\tilde{h}_t \in \text{Diff}^+(\mathbb{S}^2)$ for any $t \in [0, 1]$, such that $\tilde{h}_0 = \text{id}$ in \mathbb{S}^2 and $\tilde{h}_1 = \tilde{\vartheta}$. Note that $\tilde{h}_1 = \text{id}$ in B_1 .

Now, the idea of constructing h_t directly from \tilde{h}_t by spherical inversion could not lead to the correct conclusion, since it is not guaranteed that a neighbourhood of 0 is left fixed²⁰ by \tilde{h}_t for all $t \in (0, 1)$.

We divide the proof into three steps. In the first step, we show that it is possible to modify \tilde{h} into a map \bar{h} with the same properties, and, in addition, fixing the origin for all $t \in [0, 1]$.

¹⁹We recall that, if V is a C^∞ orientable n -dimensional manifold without boundary, and if $\text{Diff}^+(V)$ denotes the group of positive diffeomorphisms of V endowed with the $C^\infty(V, V)$ topology, then the orbits coincide with the connected components; see [7, p. 1]. If $V = \mathbb{S}^2$, then the connected component of the identity coincides with the arcwise connected component of the identity (see [6, p. 1]).

²⁰Actually, even $\{0\}$ could not be left fixed by \tilde{h}_t .

Step 1. There exists a map $\bar{h} : \mathbb{S}^2 \times [0, 1] \rightarrow \mathbb{S}^2$ of class \mathcal{C}^∞ such that

- (1) $\bar{h}_t \in \text{Diff}^+(\mathbb{S}^2)$ for all $t \in [0, 1]$,
- (2) $\bar{h}_0 = \text{id}$ in \mathbb{S}^2 ,
- (3) $\bar{h}_1 = \tilde{\vartheta}$, in particular, \bar{h}_1 is the identity in B_1 ,
- (4) $\bar{h}_t(0) = 0$ for all $t \in [0, 1]$.

Let us consider the orbit of the origin, namely $\tilde{h}_t(0)$, for $t \in [0, 1]$. Up to the composition of \tilde{h}_t with a positive \mathcal{C}^∞ diffeomorphism of \mathbb{S}^2 smoothly depending on t , we can suppose that this orbit never meets ∞ , namely

$$\tilde{h}_t(0) \neq \infty, \quad \forall t \in [0, 1]. \quad (2.4)$$

In order to ensure the validity of property (3), it is sufficient to consider a rotation²¹ $R(\cdot, t)$ sending $\tilde{h}_t(0)$ into the origin, and then define

$$\bar{h}(\cdot, t) := R\left(\tilde{h}(\cdot, t), t\right), \quad t \in [0, 1].$$

Then \bar{h} satisfies the required properties.

Note that

$$\bar{h}_0 = \tilde{h}_0, \quad \bar{h}_1 = \tilde{h}_1. \quad (2.5)$$

Step 2. We can modify \bar{h} into a map (still denoted by \bar{h} and belonging to $\mathcal{C}^\infty(\mathbb{S}^2 \times [0, 1], \mathbb{S}^2)$) satisfying, besides properties (1)–(3), also a stronger version of (4), namely

(4') there exists $\rho \in (0, 1)$ such that, setting $B_\rho := \{x \in \mathbb{R}^2 : |x| < \rho\} \subset \mathbb{R}^2$,

$$\bar{h}(x, t) = x + \mathcal{O}_t(|x|^2), \quad x \in B_\rho, \quad t \in [0, 1], \quad (2.6)$$

²¹Making use of (2.4), we define $R(\cdot, t)$ as follows. Let us identify \mathbb{S}^2 with the unit sphere in \mathbb{R}^3 and endow it with parallels and meridians with the north pole identified with ∞ and $0 \in \mathbb{S}^2$ identified with the south pole. The stereographic projection from the north pole to the tangent plane at the south pole provides an identification of points of \mathbb{R}^2 with points of $\mathbb{S}^2 \setminus \{\infty\}$ (we employ a scale reduction of a factor 2 on the stereographic projection so that the equator is mapped onto the unit circle of \mathbb{R}^2). Suppose first that $\tilde{h}_t(0)$ is not the south pole. Let $P(t)$ be the intersection between the equator of \mathbb{S}^2 and the meridian passing through (the poles and) $\tilde{h}_t(0)$. Let $r(t)$ be the line (in \mathbb{R}^3) joining $p(t)$ to $q(t)$, where $p(t)$ [respectively $q(t)$] is the point on the equator having the longitude of $P(t)$ plus (respectively minus) $\pi/2$. Then $R(\cdot, t)$ is the (smallest) rotation around $r(t)$ sending $\tilde{h}_t(0)$ into the origin. This is a rotation of angle given by the latitude of $\tilde{h}_t(0)$ plus $\pi/2$, clearly this rotation takes the above-mentioned meridian into itself. If $\tilde{h}_t(0)$ is the south pole, we define $R(\cdot, t) := \text{id}_{\mathbb{S}^2}(\cdot)$. Then, $R(\cdot, t) \in \text{Diff}^+(\mathbb{S}^2)$, and R is of class \mathcal{C}^∞ and satisfies the required properties.

uniformly with respect to $t \in [0, 1]$; that is, $\mathcal{O}_t(|x|^2)$ is a smooth map of (x, t) , such that

$$\sup_{t \in [0, 1]} \frac{|\mathcal{O}_t(|x|^2)|}{|x|^2} < +\infty.$$

Given $x \in \mathbb{R}^2$, let $J_{\bar{h}_t}(0)$ be the Jacobian matrix of \bar{h}_t (with respect to x) evaluated at $x = 0$. Making use of the compactness of $[0, 1]$, let us select $\rho \in (0, 1)$ with the property that

$$(J_{\bar{h}_t}(0))^{-1}(B_\rho) \subset\subset B_1, \quad t \in [0, 1].$$

Define, for any $t \in [0, 1]$, the map $\zeta_t : \mathbb{S}^2 \rightarrow \mathbb{S}^2$ as follows:

$$\zeta_t(x) := \begin{cases} (J_{\bar{h}_t}(0))^{-1}x & \text{if } x \in B_\rho, \\ \psi(x, t) & \text{if } \rho \leq |x| < 1, \\ x & \text{if } |x| \geq 1 \text{ or } x = \infty, \end{cases}$$

where the \mathcal{C}^∞ map ψ glues the values of ζ_t in order to ensure that ζ_t belongs to $\text{Diff}^+(\mathbb{S}^2)$.

In order to get condition (4'), it is then enough to consider the composite map

$$(x, t) \in \mathbb{S}^2 \times [0, 1] \rightarrow \bar{h}(\zeta_t(x), t),$$

and to use a Taylor expansion. The uniformity of the remainder in (2.6) with respect to $t \in [0, 1]$ is a consequence of the invertibility of $J_{\bar{h}_t}(0)$ and the compactness of $[0, 1]$.

This concludes the proof of step 2. Observe that (2.5) is still valid.

Now, we have to find a neighbourhood of the origin which is fixed by \bar{h}_t .

Step 3. We can modify the map \bar{h} into a map $\underline{h} \in \mathcal{C}^\infty(\mathbb{S}^2 \times [0, 1], \mathbb{S}^2)$, satisfying, besides properties (1)–(3), also

(4'') $\underline{h}(x, t) = x$ for any $x \in B_\rho$ and any $t \in [0, 1]$.

Choose a function $\alpha : \mathbb{S}^2 \rightarrow [0, 1]$ of class \mathcal{C}^∞ such that

- $\alpha(x) = 0$ if $|x| \geq \rho$,
- $\alpha(x) = 1$ if $|x| \leq \rho/2$,
- there exists $C \in (0, +\infty)$ such that $\sup_{x \in \mathbb{S}^2} |\nabla \alpha(x)| \leq C/\rho$.

Next, define

$$\underline{h}(x, t) := \alpha(x)x + (1 - \alpha(x))\bar{h}(x, t), \quad (x, t) \in \mathbb{S}^2 \times [0, 1],$$

so that $\underline{h}(x, t) = x$ for any $x \in B_{\rho/2}$, and $\underline{h} = \bar{h}$ in $(\mathbb{S}^2 \setminus B_\rho) \times [0, 1]$.

The Jacobian matrix of \underline{h}_t at $x \in \mathbb{R}^2$ is computed as

$$[\alpha(x)\text{Id} + (1 - \alpha(x))J_{\bar{h}_t}(x)] + (x - \bar{h}_t(x)) \otimes \nabla\alpha(x),$$

where:

- the first term is an approximation of the identity as $\rho \rightarrow 0^+$ by (4)' uniformly with respect to $t \in [0, 1]$. Indeed

$$[\alpha(x)\text{Id} + (1 - \alpha(x))J_{\bar{h}_t}(x)] - \text{Id} = (1 - \alpha(x)) (J_{\bar{h}_t}(x) - \text{Id}),$$

and by (4)'

$$\|(1 - \alpha(x)) (J_{\bar{h}_t}(x) - \text{Id})\| \leq \|J_{\bar{h}_t}(x) - \text{Id}\| \leq |\mathcal{O}_t(|x|)|,$$

where $\mathcal{O}_t(|x|)$ is a smooth map of (x, t) , such that

$$\sup_{t \in [0, 1]} \frac{|\mathcal{O}_t(|x|)|}{|x|} < +\infty;$$

- the second term (supported in B_ρ) is infinitesimal as $\rho \rightarrow 0^+$: indeed, by (2.6), we have that $x - \bar{h}_t(x)$ behaves quadratically in $|x|$ (uniformly with respect to $t \in [0, 1]$), hence it is of the order of ρ^2 , while $|\nabla\alpha|$ blows up at most as C/ρ .

It follows that the determinant of the Jacobian of \underline{h} is everywhere positive. We conclude by resorting to [13, Proposition 2.30]: on the one hand, by the properties of \underline{h} , in $\mathbb{S}^2 \setminus B_\rho$ (where it coincides with \bar{h}) we have that \underline{h}_t has degree one. On the other hand, knowing that the degree is one, positivity of the determinant of the Jacobian implies that every point of \mathbb{S}^2 has exactly one preimage.

Finally, we define

$$h(x, t) := \text{inv}(\underline{h}(\text{inv}(x), t)), \quad x \in \mathbb{R}^2, t \in [0, 1].$$

Then h satisfies the required properties.

If $n = 3$, the proof follows along the lines as in the case $n = 2$, with the following differences. One has to use, in place of [19, 20], the results of [6, Section 1, pp. 1,3] (see also [5]), which gives a smooth path $\tilde{H} : \mathbb{S}^3 \times [0, 1] \rightarrow \mathbb{S}^3$ of positive \mathcal{C}^∞ diffeomorphisms of \mathbb{S}^3 such that $\tilde{H}(\cdot, 0) = \text{id}$ in \mathbb{S}^3 and $\tilde{H}(\cdot, 1) = \tilde{\mathcal{F}}(\cdot)$, where $\tilde{\mathcal{F}}$ is a positive diffeomorphism of \mathbb{S}^3 (\mathbb{R}^3 compactified with the addition of the point ∞) extending \mathcal{F} . In addition, the rotation R in the proof of the case $n = 2$ must now be replaced by the map

$$\tau : x \in \mathbb{S}^3 \rightarrow \begin{cases} x - \tilde{H}_t(0) & \text{if } x \in \mathbb{R}^3, \\ \infty & \text{if } x = \infty. \end{cases}$$

To establish the regularity of τ in a neighbourhood of ∞ , we have to use a local chart, in order to replace ∞ with 0; this can be done using the spherical inversion inv defined as in (2.3) with \mathbb{S}^3 replacing \mathbb{S}^2 . In the new chart, $\text{inv}(\tau(\text{inv}(x)))$ is of class \mathcal{C}^1 , since

$$\begin{aligned} \text{inv}(\tau(\text{inv}(x))) &= \frac{1}{|\text{inv}(x) - \tilde{H}_t(0)|^2} \left(\frac{x}{|x|^2} - \tilde{H}_t(0) \right) \\ &= \frac{1}{1 - 2x \cdot \tilde{H}_t(0) + |x|^2 |\tilde{H}_t(0)|^2} (x - |x|^2 \tilde{H}_t(0)) \\ &= x + 2(\tilde{H}_t(0) \cdot x)x + |x|^2 \tilde{H}_t(0) + \mathcal{O}(|x|^3) \quad \text{as } x \rightarrow 0, \end{aligned}$$

and the Jacobian of $\text{inv}(\tau(\text{inv}(x)))$ converges to the identity as $x \rightarrow 0$. We can now regularize by convolution the function $\text{inv}(\tau(\text{inv}(\cdot)))$, using a standard convolution kernel independent of t , and with sufficiently small support. \square

2.4 Ambient Isotopic and Diffeomorphically Equivalent Apparent Contours

In various parts of the book we will be interested in structural properties of apparent contours; it is then convenient to introduce suitable notions of isotopies and equivalence between apparent contours.

Let us denote by M_1 and M_2 two closed two-dimensional manifolds of class \mathcal{C}^∞ .

Definition 2.4.1 (Ambient Isotopic Apparent Contours) Let

$$\varphi_1 : M_1 \rightarrow \mathbb{R}^2, \quad \varphi_2 : M_2 \rightarrow \mathbb{R}^2$$

be two stable maps. We say that $\text{appcon}(\varphi_1)$ and $\text{appcon}(\varphi_2)$ are \mathcal{C}^∞ \mathbb{R}^2 ambient isotopic (briefly, ambient isotopic) if there exists an \mathbb{R}^2 -ambient isotopy $h : \mathbb{R}^2 \times [0, 1] \rightarrow \mathbb{R}^2$ with compact support such that

$$h_1(\text{appcon}(\varphi_1)) = \text{appcon}(\varphi_2),$$

and

$$f_{\varphi_2} \circ h_1 = f_{\varphi_1}.$$

Therefore, two apparent contours are ambient isotopic if there exists a path $t \in [0, 1] \rightarrow h_t(\cdot) \in \text{Diff}_c(\mathbb{R}^2)$ with origin the identity, taking at the final time one apparent contour into the other, and respecting the values of the functions $f_{\varphi_1}, f_{\varphi_2}$ (this latter condition being equivalent to consistency of the \mathbb{R}^2 -ambient isotopy with the orientation of the arcs of the apparent contour).

Definition 2.4.2 (Diffeomorphically Equivalent Apparent Contours) Let

$$\varphi_1 : M_1 \rightarrow \mathbb{R}^2, \quad \varphi_2 : M_2 \rightarrow \mathbb{R}^2$$

be two stable maps. We say that $\text{appcon}(\varphi_1)$ and $\text{appcon}(\varphi_2)$ are diffeomorphically equivalent if there exists $\vartheta \in \text{Diff}_c(\mathbb{R}^2)$ such that

$$\vartheta(\text{appcon}(\varphi_1)) = \text{appcon}(\varphi_2), \quad (2.7)$$

and

$$f_{\varphi_2} \circ \vartheta = f_{\varphi_1}. \quad (2.8)$$

Definitions 2.4.1 and 2.4.2 are equivalent, as the following result shows.

Theorem 2.4.3 (\mathbb{R}^2 -Ambient Isotopies and Diffeomorphisms) Let

$$\varphi_1 : M_1 \rightarrow \mathbb{R}^2, \quad \varphi_2 : M_2 \rightarrow \mathbb{R}^2$$

be two stable maps. Then $\text{appcon}(\varphi_1)$ and $\text{appcon}(\varphi_2)$ are ambient isotopic if and only if they are diffeomorphically equivalent.

Proof Suppose that $\text{appcon}(\varphi_1)$ and $\text{appcon}(\varphi_2)$ are ambient isotopic. Then, the map $\vartheta := h_1$ belongs to $\text{Diff}_c(\mathbb{R}^2)$, and (2.7) and (2.8) hold. Conversely, if $\text{appcon}(\varphi_1)$ and $\text{appcon}(\varphi_2)$ are diffeomorphically equivalent, then they are ambient isotopic, as a consequence of Theorem 2.3.3. \square

A first use of the diffeomorphic equivalence between apparent contours is given in the next section.

2.5 Morse Descriptions of an Apparent Contour

In this section we shall consider again the case $\mathcal{X} = M$ a closed two-dimensional manifold, and $\mathcal{Y} = \mathbb{R}^2$. In a number of problems, it is convenient to seek a way to describe an apparent contour, capable of keeping track of its topological structure, and insensitive to smooth *sufficiently small* deformations of \mathbb{R}^2 .²² We shall use this description for solving the completion problem (Chap. 4), to give a natural description of the invariant $\mathfrak{B}(\text{appcon}(\varphi))$ defined in Sect. 7.1, and, above all, for algorithmical applications (Chaps. 9 and 10). It turns out that this description can be achieved by a simple technique obtained by adapting the methods of Morse

²²This task is similar to describing a prime knot in knot theory by means of notations like the one introduced by Dowker-Thistlethwaite [16, p. 7], see also the combinatorial description of knotted surfaces in [4, pp. 21,22].

theory [2]. As seen in Sect. 2.2, an apparent contour is essentially a 1D manifold with the addition of singular points, namely crossings and cusps. This suggests the introduction of a *height function* with the usual nondegeneracy properties of a Morse function, and further requirements that take into account the singular points. Such additional requirements would be similar (and simpler, due to the lower dimension) to those that define a *stratified Morse function*,²³ the stratification of $\text{appcon}(\varphi)$ being defined as the set of crossings and cusps as the zero-dimensional stratum, and its complement as the one-dimensional stratum. By scanning the apparent contour while moving the level of the height function, we can keep track of all critical changes in the local shape of the apparent contour: local maxima/local minima are typical of standard Morse theory (for one-dimensional manifolds). To them we must add new *Morse events*: traversing of a crossing and traversing of a cusp.

However, by doing so, we shall lose information about the relative position, in the plane, of the point of the apparent contour originating the *Morse event* with respect to the rest of the apparent contour, in particular to the other *regular* points at the same level (same value of the height function). Therefore, we adapt this approach in a way that takes into account the embedding of the apparent contour in \mathbb{R}^2 . It turns out that $\text{appcon}(\varphi)$ defines a stratification of the plane having the set of crossings and cusps as the zero-dimensional stratum, its complement in $\text{appcon}(\varphi)$ as the one-dimensional stratum and $\mathbb{R}^2 \setminus \text{appcon}(\varphi)$ as the two-dimensional stratum. We could thus consider height functions defined on the whole of \mathbb{R}^2 , that are *stratified Morse functions* with respect to the given stratification (see again Definition 6.2.3). Actually, here we do not need this generality; moreover, it is convenient to depart substantially from the details of such an approach in the vicinity of a cusp, in order to simplify the resulting description.

In the end, we shall be able to describe the apparent contour in terms of a finite sequence of *events*. This sequence can be readily converted into a sequence of typographical characters to be used, for instance, as input for a computer program that implements specific computations on apparent contours as the ones described in Chaps. 9 and 10. As already said, a Morse description of an apparent contour²⁴ will also be used in Chap. 7 and, in the context of *labelled* apparent contours, in Sect. 4.4, for the completion problem of visible contours. A software program that automates the completion process will be illustrated in Chap. 9.

The Morse description will be:

- **Finite:** the description requires only a finite sequence of symbols, taken from a finite set. The action of an element of $\text{Diff}_c(\mathbb{R}^2)$ leads to a deformed apparent contour that might or might not have the same Morse description.
- **Complete:** two apparent contours with the same Morse description are ambient isotopic. This last property is crucial, and a proof will be sketched at the end of Sect. 2.5.4.

²³See [11] for the definition of stratifications and stratified maps. Compare also with Chap. 6, Definition 6.2.3.

²⁴See, e.g., [4].

Definition 2.5.1 (Morse Lines) By a *one-parameter family of Morse lines traversing* \mathbb{R}^2 we mean a C^∞ diffeomorphism²⁵ $m : \mathbb{R} \times \mathbb{R} = \mathbb{R}_s \times \mathbb{R}_\lambda \rightarrow \mathbb{R}^2 = \mathbb{R}_x^2$ which is the identity out of a bounded set of $\mathbb{R} \times \mathbb{R}$.

The first variable of m will be denoted by s and the second variable by λ . If $p \in \mathbb{R}^2$ has coordinate $x = (x_1, x_2)$, with $\lambda(p)$ we shall thus refer to the second component of the inverse m^{-1} of m , which will play the role of the aforementioned Morse height function. The first component of the inverse of m will be usually denoted by $s = s(p)$ and used as a parametrization of the lines of constant λ , called Morse lines.

For any $\lambda \in \mathbb{R}$ we set $m_\lambda(\cdot) := m(\cdot, \lambda)$, which is an oriented curve “traversing” \mathbb{R}^2 . Note that if $\lambda_1 \neq \lambda_2$ then $m_{\lambda_1}(\mathbb{R}) \cap m_{\lambda_2}(\mathbb{R}) = \emptyset$.

Since all apparent contours that we shall consider are compactly contained in the square $\Omega = (0, 1) \times (0, 1)$, we can suppose without loss of generality that m is the identity outside $\overline{\Omega}$, so that

$$m_0([0, 1]) \cap \text{appcon}(\varphi) = m_1([0, 1]) \cap \text{appcon}(\varphi) = \emptyset.$$

For any given point $p \in \mathbb{R}^2$ there is a unique value of $\lambda = \lambda(p) \in \mathbb{R}$ such that the Morse line $m_\lambda(\cdot)$ passes through p . In this way the height function $\lambda(p)$ is defined in the whole of \mathbb{R}^2 and in particular on the apparent contour.

Remark 2.5.2 (Horizontal Morse Straight Lines) Up to the action of an element in $\text{Diff}_c(\mathbb{R}^2)$, we can suppose that m is the identity everywhere, so that the Morse lines m_λ become horizontal straight lines at height λ . It is therefore equivalent to leave the apparent contour unchanged and choose m appropriately or to deform the apparent contour (by smoothly deforming the plane) and choosing m as the identity map. Note also that using $m^{-1}(p) = (s(p), \lambda(p))$ in place of $p = (x_1, x_2)$ is equivalent to a change of (curvilinear) coordinate system: in the modified coordinate system we can take m to be the identity without having to deform the apparent contour and again the Morse lines m_λ become horizontal straight lines. In Fig. 2.5 the same apparent contour is drawn in the x_1x_2 coordinate system (left picture) and in the $s\lambda$ coordinate system (right picture).

2.5.1 Genericity of Morse Lines in Case of No Cusps

Suppose for the moment that the apparent contour does not contain cusps. Given $\lambda \in [0, 1]$, we say that $m_\lambda(\cdot)$ is *generic* if $m_\lambda([0, 1])$ does not contain any crossing, it intersects $\text{appcon}(\varphi)$ in a finite set and each intersection is transverse.

²⁵Notice that m can be thought of as an element of $\text{Diff}_c(\mathbb{R}^2)$.

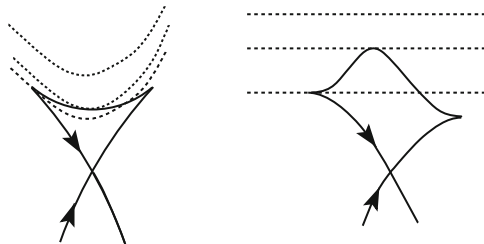


Fig. 2.5 The descriptive map m (Definition 2.5.4) can be regarded as an alternative coordinate system. In the new system (*right*) the Morse curves (*dotted lines in the left figure*) become *horizontal straight lines* (*right*). Note that the interior local maximum remains after the coordinate change

We say in turn that m is *generic* if $m_\lambda(\cdot)$ is generic for all λ except for a finite set $\{\lambda_1, \dots, \lambda_n\}$ of values called *critical levels* (or *critical values*), with $1 > \lambda_1 > \lambda_2 > \dots > \lambda_n > 0$. Moreover for any $\bar{\lambda} \in \{\lambda_1, \dots, \lambda_n\}$ we require that there exists a unique $\bar{s} \in (0, 1)$ such that $m_{\bar{\lambda}}([0, 1] \setminus \{\bar{s}\})$ intersects transversally $\text{appcon}(\varphi)$ in a finite set of points not in $\text{crossings}(\text{appcon}(\varphi))$ (i.e., $m_{\bar{\lambda}}$ would be generic but for the presence of the point $m_{\bar{\lambda}}(\bar{s})$) and, if we set $p := m_{\bar{\lambda}}(\bar{s}) \in \text{appcon}(\varphi)$, we have exactly one of the following items:

- p is a crossing with the two tangents transversal to the Morse line;
- p is an intersection point with $\text{appcon}(\varphi)$ locally on one side of the Morse line (a local maximum or a local minimum).

Since crossings are in a finite number we can readily adapt the standard arguments of Morse theory and conclude that the set of generic families of Morse lines is dense in the set of C^∞ diffeomorphisms of the plane. This means that the identity is generic up to a small deformation (equivalently, that the identity is generic if we slightly deform the apparent contour).

In view of Remark 2.5.2, we shall assume that each m_λ is a *horizontal straight line*. Therefore we call m a *generic family of horizontal Morse lines* for $\text{appcon}(\varphi)$. We choose an orientation of \mathbb{R}^2 and we take the convention that the lines foliate \mathbb{R}^2 from top to bottom as $\lambda \in \mathbb{R}$ decreases.

Recall that the family of Morse lines $m_\lambda(\cdot)$ can equivalently be seen as the family of level curves of the Morse *height* function $\lambda(p)$. In this context the *critical levels* correspond to self-intersections and (nondegenerate) critical points of the Morse function.

2.5.2 Morse Lines in Case of Cusps: Markers

Cusps of the apparent contour are treated in a special way. Indeed it is convenient to view cusps simply as marked points along an otherwise smooth arc (either closed or connecting pairs of crossings). This is accomplished for any cusp $c \in$

$\text{cusps}(\text{appcon}(\varphi))$ by devising the diffeomorphism m so that the Morse line m_λ passing through c is tangent to the cusp (namely, tangent to the two arcs ending at c). This is clearly achievable by a local modification of the diffeomorphism m in a sufficiently small ball centred at c .

Equivalently, as shown in Fig. 2.7, we can deform the apparent contour locally around the cusp and make it tangential to the Morse line through it; if m is the identity, this is equivalent to require that cusps are horizontal. Compare also Fig. 2.5, in which the use of the $s\lambda$ coordinate system allows to leave the apparent contour unchanged; see also Remark 2.5.2.

Remark 2.5.3 It is important to clarify that the above described modifications (of m or of the apparent contour) are *not* small in terms of the $W^{1,\infty}$ -norm, since we are substantially changing the derivatives of m or the slope of the apparent contour. Moreover the required tangentiality property is lost under small perturbations of m or of the apparent contour in the Whitney topology. However such a requirement is perfectly legitimate and well suited for our purposes. Indeed, if the cusp were not tangent to the Morse line, then necessarily the apparent contour would lie locally on one side, similarly to the case of local maxima/minima, and this would confuse the resulting description.

To take into account the presence of cusps we accordingly modify the definition of one-parameter family of Morse curves as follows.

Given $\lambda \in [0, 1]$, we say that $m_\lambda(\cdot)$ is *generic* if $m_\lambda([0, 1])$ does not contain any crossings *or cusps*, it intersects $\text{appcon}(\varphi)$ in a finite set and each intersection is transverse.

Definition 2.5.4 (Descriptive Map of an Apparent Contour) We say that m is a descriptive map of $\text{appcon}(\varphi)$ if:

- (1) there exists a finite set $\{\lambda_1, \dots, \lambda_n\}$ of real numbers, called critical values, with $1 > \lambda_1 > \lambda_2 > \dots > \lambda_n > 0$, such that $m_\lambda(\cdot)$ is generic for any $\lambda \in [0, 1] \setminus \{\lambda_1, \dots, \lambda_n\}$;
- (2) for any $\bar{\lambda} \in \{\lambda_1, \dots, \lambda_n\}$ there exists a unique $\bar{s} \in (0, 1)$ such that $m_{\bar{\lambda}}([0, 1] \setminus \{\bar{s}\})$ intersects $\text{appcon}(\varphi)$ in a finite set of points and each intersection is transverse and belongs to $\text{appcon}(\varphi) \setminus \text{nodes}(\text{appcon}(\varphi))$. Moreover, if we set $p := m_{\bar{\lambda}}(\bar{s}) \in \text{appcon}(\varphi)$, we have exactly one of the following items:
 - (a) p is a crossing with the two tangents transversal to the Morse line $m_{\bar{\lambda}}(\mathbb{R})$;
 - (b) p is neither a crossing nor a cusp of $\text{appcon}(\varphi)$ and is a local maximum or a local minimum, namely $\text{appcon}(\varphi)$ lies locally on one side of the Morse line $m_{\bar{\lambda}}(\cdot)$ with the values of λ being locally smaller than $\bar{\lambda}$ for a maximum and locally larger than $\bar{\lambda}$ for a minimum;
 - (c) p is a cusp and the Morse line is tangent to the two arcs ending at p ; in view of the local shape of cusps which forces the curvature to blow up both at $+\infty$ and at $-\infty$ at the cusp, the Morse line must necessarily traverse the apparent contour at p , i.e., the two arcs at p are locally in opposite sides with respect to the Morse line.

The pair $(\bar{s}, \bar{\lambda})$ in Definition 2.5.4 is called critical point; with a slight abuse of notation we shall also refer to $m_{\bar{\lambda}}(\bar{s})$ as a critical point.

Since a Morse line through a cusp crosses locally the apparent contour, from the point of view of the Morse description a cusp behaves essentially in the same way as a regular point at a regular value of λ , at which the Morse line has only transversal intersections. Moreover the local geometry of the cusp is, up to a compactly supported C^∞ diffeomorphism of \mathbb{R}^2 , completely determined and only depends on the local orientation of the two arcs ending at the cusp. Namely the cusp points to the right if the two arcs are oriented upwards, and it points to the left if the two arcs are oriented downwards; here, we suppose for definiteness that m is the identity, i.e., the Morse lines are horizontal straight lines oriented from left to right for increasing values of the parameter s .

In other words, a cusp can be reconstructed (up to the action of an element of $\text{Diff}_c(\mathbb{R}^2)$) if we substitute it with a distinguished point (that will be called a *marker* or *marked point*) on the arc, obtained by glueing the two arcs meeting at the cusp in a smooth way and with a non-horizontal tangent; this is described in Fig. 2.8. Concretely, we smooth out all cusps and only keep track of them by marking their former position along the arcs.

Definition 2.5.5 (Extended Arcs) The relatively open arcs that result after glueing at cusps will be called extended arcs.

As an example, the apparent contour of Fig. 2.6 can be conveniently described using the equivalent sketch of Fig. 2.8 in which we have five extended arcs, two of them containing two *markers* each (representing the four cusps).

We shall still regard such markers as formally representing cusps, so that they will still be listed in $\text{nodes}(\text{appcon}(\varphi))$.

Remark 2.5.6 For clarity of exposition, cusps (marked points) are considered singular points in the definition of a generic $m_\lambda(\cdot)$ (i.e., the intersection of $m_\lambda(\cdot)$ with $\text{appcon}(\varphi)$ at a marked point implies that $m_\lambda(\cdot)$ is nongeneric). However the topological structure of the apparent contour does not change when locally crossing such a critical level and the only required information for our purposes is the

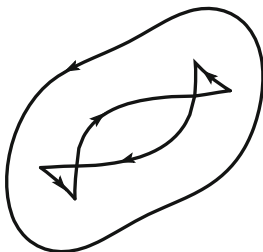


Fig. 2.6 An example of oriented apparent contour with two crossings and four cusps (M is a torus)

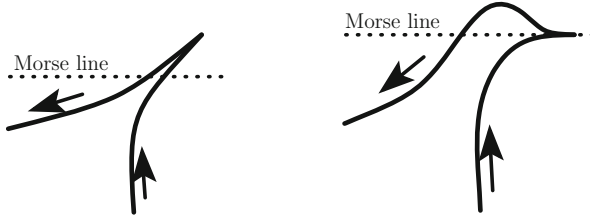


Fig. 2.7 A cusp in general position with respect to the Morse line (*left*) can be locally deformed into a cusp tangent to the Morse line (*right*)

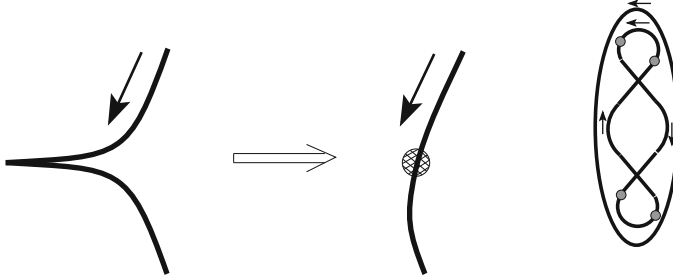


Fig. 2.8 Substitution of a marker in place of a cusp in the apparent contour of Fig. 2.6. Cusps are moved so that they have different height, which is a required condition for m to be a descriptive map using a family of horizontal Morse lines

number of cusps contained in each extended arc of $\text{appcon}(\varphi)$. Such extended arcs are curves that do not contain any crossing, which are either closed or have crossings as endpoints.

Remark 2.5.7 In Sect. 10.2 we shall introduce another description of an apparent contour, called the region description, which is also well suited for being implemented on a computer.

2.5.3 The Morse Description

A descriptive map $m : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}^2$ provides a way to encode the shape of the apparent contour, up to deformations by a C^∞ diffeomorphism of \mathbb{R}^2 (or equivalently of $\mathbb{R} \times \mathbb{R}$) with compact support. For any given $\lambda \in (0, 1)$ we have a finite set of intersections of the Morse line $m_\lambda(\cdot)$ with $\text{appcon}(\varphi)$, all of these, except at most one, are transverse, with an arc a of the apparent contour that crosses the Morse line upwards (following the canonical orientation, a crosses the Morse line with increasing values of λ) or downwards (a crosses the Morse line with decreasing values of λ). For regular values of λ (i.e., when $m_\lambda(\cdot)$ is generic), these are the only possible intersection points. Since we seek a description up to the action of $\text{Diff}_c(\mathbb{R}^2)$, it is not important to record the values of the parameter s corresponding to these crossings, but only their relative ordering together with a

sign distinguishing between upward and downward crossings. Therefore we end up with a finite sequence of symbols taken from

$$\{\uparrow, \downarrow\}.$$

If $1 > \lambda_1 > \dots > \lambda_n > 0$ are the critical values of λ , the sequence of arrows does not change for all values $\lambda \in \bigcup_{i=1}^{n-1} (\lambda_{i+1}, \lambda_i)$. Hence it suffices to choose one regular value for each interval between two consecutive critical values to represent the whole regular interval.

At a critical level $\bar{\lambda} = \lambda_i$ for some $i \in \{1, \dots, n\}$ we have a special intersection point which, according to its type and depending on the local orientation of the arcs involved, corresponds to one of the following (finite) list of symbols

$$\{\curvearrowleft, \curvearrowright, \smile, \frown, \times, \bowtie, \bowtie, \bowtie, \bowtie, \succ, \prec\}$$

which, together with the symbols \uparrow and \downarrow , allows to encode the shape information associated with $\bar{\lambda}$. Finally, the information in a regular interval can be easily inferred from the information at one of the two critical values bounding the interval, and can thus be omitted.

In the end we can encode the shape information with a list of n sequences of symbols, each taken from the lists above, exactly one of which is not in $\{\uparrow, \downarrow\}$.

Remark 2.5.8 (Morse Description and Diffeomorphic Equivalence) Two diffeomorphically equivalent apparent contours (Definition 2.4.2) do not necessarily have the same Morse description. Similarly, an apparent contour has infinitely many different Morse descriptions.

Example 2.5.9 (Morse Description of a Torus) To illustrate this we give the encoding corresponding to the apparent contour of Fig. 2.8 (right) where the four markers are just a graphical replacement for small horizontal cusps:

$$\begin{array}{c} \curvearrowleft \\ \downarrow \curvearrowleft \uparrow \\ \downarrow \prec \uparrow \uparrow \\ \downarrow \downarrow \succ \uparrow \\ \downarrow \times \uparrow \\ \downarrow \bowtie \uparrow \\ \downarrow \prec \uparrow \uparrow \\ \downarrow \downarrow \succ \uparrow \\ \downarrow \smile \uparrow \\ \smile \end{array}$$

The critical levels are the following:

Oriented global maximum: \curvearrowright . The orientation is necessarily from the right to the left.

Oriented local maximum, having one arc on the left (respectively on the right) oriented downwards (respectively upwards): $\downarrow \curvearrowright \uparrow$.

Oriented marked point $<$ having one arc on the left oriented downwards and two arcs on the right oriented upwards: $\downarrow < \uparrow \uparrow$.

Oriented marked point $>$ having two arcs on the left oriented downwards and one arc on the right oriented upwards: $\downarrow \downarrow > \uparrow$.

Oriented crossing, having one arc on the left (respectively on the right) oriented downwards (respectively upwards): $\downarrow \times \uparrow$.

Oriented crossing, having one arc on the left (respectively on the right) oriented downwards (respectively upwards): $\downarrow \times \uparrow$.

Oriented marked point $<$ having one arc on the left oriented downwards and two arcs on the right oriented upwards: $\downarrow < \uparrow \uparrow$.

Oriented marked point $>$ having two arcs on the left oriented downwards and one arc on the right oriented upwards: $\downarrow \downarrow > \uparrow$.

Oriented local minimum \curvearrowleft , having one arc on the left (respectively on the right) oriented downwards (respectively upwards): $\downarrow \curvearrowleft \uparrow$.

Oriented global minimum \curvearrowleft .

Remark 2.5.10 (Morse Description and Labelling) In Sect. 3.4 (see also Definition 4.2.5) we shall introduce a *labelling* of the apparent contour, which is, in essence, just a nonnegative integer tag attached to each arc. For such labelled apparent contours the Morse description is slightly more complex. Indeed we need to add one or more nonnegative integers to each symbol encoding the corresponding value of the labelling: one integer is enough for the symbols representing transversal crossings $\{\uparrow, \downarrow\}$ and local maxima/minima $\{\curvearrowright, \curvearrowleft, \curvearrowright, \curvearrowleft\}$, two integers are necessary to give the values of the labelling before and after a cusp for the symbols in $\{>, <\}$, and finally four integers are required for giving the labelling of the four involved arcs at a crossing for the remaining symbols.

2.5.4 Recovering the Shape from a Morse Description

Given the list of symbols obtained from an apparent contour $\text{appcon}(\varphi)$ with the procedure described above and using m as descriptive map, we can physically construct a *representative* apparent contour and prove that it is diffeomorphically equivalent to $\text{appcon}(\varphi)$ (Definition 2.4.2). From this it readily follows the **completeness** property listed at the end of the introductory text of Sect. 2.5. In the construction below all diffeomorphisms are assumed to be of class C^∞ and with compact support.

Step 1: construction of the representative when $\text{appcon}(\varphi)$ does not contain cusps.

We start by constructing a grid of relevant points as follows. Let n be the number of lines of symbols in the Morse description (number of critical levels). Let us introduce n equally spaced horizontal lines in the square $\Omega = (0, 1) \times (0, 1)$ and on each line (starting from the top) fix a number of equally spaced points in the same number as the symbols in the corresponding line of the Morse description. The constructed set of points corresponds to the intersections of the Morse line with the apparent contour at the critical levels. Now, we connect the points on two adjacent horizontal lines with oriented segments consistently with the type indicated by the corresponding symbols. This is clearly possible since the Morse description actually comes from some apparent contour. In this way we end up with an oriented *polygonal* apparent contour that can be smoothed out with standard smoothing techniques without moving the points of the grid and using only horizontal displacements. Let us call $\bar{\Psi}$ the resulting apparent contour and let \bar{m} be the corresponding descriptive map (which is actually the identity map). On one side it is clear that the Morse description of $\bar{\Psi}$ given by \bar{m} coincides with our Morse description. On the other side we can prove that $\bar{\Psi}$ is diffeomorphically equivalent to $\text{appcon}(\varphi)$. First, by preliminarily using m as a diffeomorphism of the plane, we can assume without loss of generality that m is the identity map. Then we can map the n critical levels of m onto the equally spaced lines used to construct $\bar{\Psi}$ with an increasing diffeomorphism of the x_2 -coordinate; we can thus assume that the critical levels of m are exactly the same as those of \bar{m} . Eventually, for any given height x_2 we map diffeomorphically the intersection points of m onto those of \bar{m} . This can be done in view of our construction and can also be done smoothly with respect to the x_2 -coordinate, giving the desired equivalence.

Step 2: dealing with the cusps.

Cusps can be introduced into the representative apparent contour at the end of the previous construction (obtained as if cusps were not present). We simply deform the apparent contour in a small enough neighbourhood of the cusp point by introducing a rescaled version of the standard semicubic cusp with horizontal tangent and oriented to the right or to the left according to the orientation of the arcs ending at the cusp. It is again possible to prove the diffeomorphic equivalence of $\bar{\Psi}$ with $\text{appcon}(\varphi)$ by locally modifying around cusps the diffeomorphism constructed in step 1.

References

1. Arnold, V.I., Goryunov, V.V., Lyashko, O.V., Vassiliev, V.A.: In: V.I. Arnold (ed.) *Dynamical Systems VIII. Singularity Theory. II. Applications*. Springer, Berlin (1993)
2. Audin, M., Damian, M.: *Théorie de Morse et Homologie de Floer*. EDP Sciences. CNRS Editions, Paris (2010)
3. Bruce, J.W., Giblin, P.J.: *Curves and Singularities. A Geometrical Introduction to Singularity Theory*, 2nd edn. Cambridge University Press, Cambridge (1992)
4. Carter, J.S., Kamada, S., Saito, M.: *Surfaces in 4-Space*. Encyclopaedia of Mathematical Sciences, vol. 142. Springer, Berlin (2004)

5. Cerf, J.: La Théorie de Smale sur le h -cobordisme des variétés, 1961/1962 Séminaire Henri Cartan, 1961/62, Exp. 11–13 23 pp. Secrétariat Mathématique, Paris
6. Cerf, J.: Sur le difféomorphismes de la sphère de dimension trois ($\Gamma_4 = 0$). Lecture Notes in Mathematics, vol. 53. Springer, Berlin (1968)
7. Cerf, J.: La stratification naturelle des espace de fonctions différentiables réelles et le théorème de la pseudo-isotopie. Publ. Math. Inst. Hautes Études Sci. 5–170 (1970)
8. Crowell, R.H., Fox, R.H.: Introduction to Knot Theory. Springer, New York (1977)
9. Gibson, C.G.: Singular Points of Smooth Mappings. Research Notes in Mathematics. Pitman, London (1978)
10. Golubitsky, M., Guillemin, V.: Stable Mappings and Their Singularities. Graduate Texts in Mathematics, vol. 14. Springer, New York (1974)
11. Goresky, M., MacPherson, R.: Stratified Morse Theory. Springer, Berlin (1988)
12. Haefliger, A.: Quelques remarques sur les applications différentiables d'une surface dans le plan. Ann. Inst. Fourier. Grenoble **10**, 47–60 (1960)
13. Hatcher, A.: Algebraic Topology Online Book. Cambridge University Press, Cambridge (2002)
14. Hirsch, M.W.: Differential Topology. Springer, New York (1976)
15. Kriegel, A., Michor, P.W.: The Convenient Setting of Global Analysis. Mathematical Surveys and Monographs, vol. 53. American Mathematical Society, Providence (1997)
16. Lickorish, W.B.R.: An Introduction to Knot Theory. Springer, New York (1997)
17. Lu, Y.C.: Singularity Theory and an Introduction to Catastrophe Theory. Universitext. Springer, New York (1976)
18. Martinet, J.: Singularités des Fonctions et Applications Différentiables. Pontificia Universidade de Rio de Janeiro, Rio de Janeiro (1974)
19. Munkres, J.: Differentiable isotopies on the 2-sphere. Mich. Math. J. **7**, 193–197 (1960)
20. Munkres, J.: Obstruction to the smoothing of piecewise-differentiable homeomorphisms. Ann. Math. **72**, 521–554 (1960)
21. Ohmoto, T., Aicardi, F.: First order local invariants of apparent contours. Topology **45**, 27–45 (2006)
22. Pignoni, R.: Curves and surfaces in real projective spaces: an approach to generic projections. Banach Center Publ. **88**, 335–351 (1988)
23. Thom, R.: Les singularités des applications différentiables. Ann. Inst. Fourier **6**, 43–87 (1956)
24. Thom, R.: Stabilité Structurale et Morphogénèse. W.A. Benjamin, Inc., Reading (1972)
25. Whitney, H.: On singularities of mappings of Euclidean spaces. I. Mappings of the plane into the plane. Ann. Math. **62**, 374–410 (1955)
26. Whitney, H.: Singularities of mappings of Euclidean spaces. Sympos. Internac. Topologica Algebraica Mexico **62**, 285–301 (1958)
27. Wilson, L.C.: Equivalence of stable mappings between two-dimensional manifolds. J. Differ. Geom. **11**, 1–14 (1976)

Chapter 3

Apparent Contours of Embedded Surfaces

In this chapter we adapt the notions introduced in Chap. 2 to the special case of the apparent contour of a smooth, possibly nonconnected, compact surface Σ without boundary embedded in \mathbb{R}^3 . Embeddedness allows to enrich an apparent contour with a labelling, which, in particular, permits to define the visible contour.

3.1 Three-Dimensional Scenes

Let us start by defining what we mean by a three-dimensional scene.

Definition 3.1.1 (3D Scene) By a 3D scene E (a scene, for short) we mean the closure of a bounded open subset of \mathbb{R}^3 having boundary ∂E of class C^∞ .

We recall that ∂E is of class C^∞ if the following holds: for any point $p \in \partial E$ there exist a ball B centred at p and an orthogonal coordinates system of \mathbb{R}^3 such that $B \cap \partial E$ (respectively $B \cap E$) can be written as the graph (respectively the subgraph intersected with B) of a function of class C^∞ defined in an open subset of \mathbb{R}^{n-1} .

We shall always assume that $E \subset Q = \Omega \times I$, namely that the scene is contained in the open portion Q of cylinder over the open square $\Omega = (0, 1) \times (0, 1)$.

From the definition, it follows that the set E is the union of a finite number of smooth, compact, pairwise disjoint connected components (sometimes called bodies), each connected component being the closure of a bounded open subset of \mathbb{R}^3 . In the sequel, in order to simplify notation, the boundary of E will be often denoted with the symbol Σ , which is an orientable, possibly nonconnected, closed¹

¹Recall that closed here means compact without boundary.

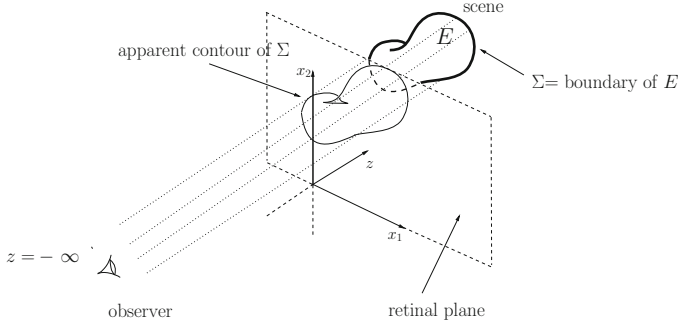


Fig. 3.1 The left-oriented orthonormal basis $\{e_1, e_2, e_3\}$. The span of e_1 and e_2 is the retinal plane, which lies in between the observer and the scene

surface of class C^∞ , embedded in \mathbb{R}^3 . It is clear that \mathbb{R}^3 is partitioned as the disjoint union of the interior of E , of Σ , and of the complement $\mathbb{R}^3 \setminus E$ of E .

In our framework, the scene is seen from the point of view of an observer lying far away from E . As explained in the Introduction and in Chap. 1, one of our main interests is to reconstruct E , or at least its topology, starting from information on a two-dimensional picture of E taken by the observer. The picture of E lies on a “retinal plane” where points of E are projected: conventionally, we imagine the retinal plane to be interposed between the observer and E itself, as in Fig. 3.1. In this way we can recover the 2D picture by looking at the intersections of the light rays connecting a point on Σ with the point of view. The actual position of the retinal plane is not really important, and positioning the plane behind the scene (with respect to the observer) would make no real difference.

We will be concerned with the case of an observer situated at infinite distance from E , and of an orthogonal projection from \mathbb{R}^3 onto the retinal plane. This suggests the following splitting.

3.1.1 Splitting of \mathbb{R}^3

Let us denote by $\{e_1, e_2, e_3\}$ a left-oriented orthonormal basis of the ambient space \mathbb{R}^3 , see Fig. 3.1, where the z -axis corresponds to e_3 . We split \mathbb{R}^3 as

$$\mathbb{R}^3 = \mathbb{R}^2 \oplus \mathbb{R},$$

so that a point $p \in \mathbb{R}^3$ has coordinates (x, z) in the basis $\{e_1, e_2, e_3\}$, with $x = (x_1, x_2) \in \mathbb{R}^2$ and $z \in \mathbb{R}$. The coordinate z has the meaning of depth and increases from the observer toward the scene. We fix the plane $\{z = 0\} = \text{span}\{e_1, e_2\}$ to be the retinal plane. Since by our convention the retinal plane lies in between the observer and E , the scene is, from the point of view of the observer, behind the

retinal plane²: points of E have positive third coordinate z , while the observer is situated at $z = -\infty$.

We denote by

$$\pi : \mathbb{R}^3 \rightarrow \mathbb{R}^2 \times \{0\}, \quad \pi(x, z) := (x, 0), \quad (3.1)$$

the orthogonal projection onto \mathbb{R}^2 ; whenever no confusion is possible, we shall often write \mathbb{R}^2 in place of $\mathbb{R}^2 \times \{0\}$. As we shall see, the apparent contour of Σ will be the planar image, via the map π , of a suitable set of curves lying on Σ . In the terminology of computer vision, the planar image of E is sometimes called the outline of the scene, while the boundary of the outline is called the contour [12, Chap. 8].

When E is considered as consisting of opaque bodies, parts of E may be occluded by other parts. As already discussed in Chap. 1, various qualitatively different cases can occur. For instance, the occlusion does not necessarily take place between two different connected components of E , since one connected component can be partially occluded by itself, as displayed in Fig. 2 of the Introduction. This shows that, in general, there is no depth ordering relation between the connected components of E . In Chap. 4 we will be concerned with the problem of reconstructing hidden portions of the apparent contour of Σ when E is opaque.³

Another case is when E is partially transparent, so that it may be possible to recognize the boundary of the outline of Σ even in presence of occlusions. Depending on the situation at hand, it might or might not be possible to infer the number of layers of Σ crossed by the light ray before hitting an occluded target point on Σ . The transparent case will be considered in Chap. 5.⁴

We shall always assume that the boundary of the scene is *in general position with respect to π* . To properly introduce this notion (Definition 3.2.1 below) we need to recall some concepts from singularity theory (see, for instance, [1, 2, 9] and the references therein).

3.2 Apparent Contours of Embedded Surfaces

We now come back to the discussion started in Sect. 2.2, focusing our attention to the case of three-dimensional scenes, therefore when our surfaces are embedded in the three-dimensional Euclidean space. Essentially, this amounts to consider particular maps φ that factorize through an embedding of M in \mathbb{R}^3 and an orthogonal

²We advise the reader that in some of the figures, it will be convenient to imagine the x_1x_2 plane as horizontal, with the z direction being vertical, and e_3 pointing downwards.

³We shall show a completion theorem starting only from the visible part of the apparent contour.

⁴We shall show a reconstruction theorem of a three-dimensional shape starting from the knowledge of the whole apparent contour, and of a consistent labelling on it.

projection $\mathbb{R}^3 \rightarrow \mathbb{R}^2$; see (3.3) below. In this situation we are allowed to identify M with its embedded image $\Sigma \subset \mathbb{R}^3$. We shall see that it is then possible to give a clear geometric meaning to the function f_φ introduced in Definition 2.2.5. Moreover, it is also possible to define a new function on the apparent contour related to the depth ordering, leading to the concept of labelling; see, e.g., [11, 17], [7, p. 19], [6, pp. 19,20], [3]. The labelling will be crucial in the process of global reconstruction of a scene starting from a suitable plane graph, as described in Chap. 5.

Let there be given a scene $E \subset \mathbb{R}^3$, and set, as usual,

$$\Sigma = \partial E. \quad (3.2)$$

Consider the splitting of \mathbb{R}^3 and the projection $\pi : \mathbb{R}^3 \rightarrow \mathbb{R}^2 \times \{0\}$ described in Sect. 3.1.1.

Definition 3.2.1 (General Position, Stable Scene) We say that Σ is in general position with respect to π , or equivalently that the scene E is stable (or also that Σ is stable), if the restriction of π to Σ ,

$$\pi|_\Sigma : \Sigma \rightarrow \mathbb{R}^2 \times \{0\},$$

is stable.

Definition 3.2.1 can be easily related to Definition 2.1.2. Indeed, take a two-dimensional smooth closed orientable manifold M with the same topology of Σ , denote by

$$\text{Emb}(M, \mathbb{R}^3) \subset \mathcal{C}^\infty(M, \mathbb{R}^3)$$

the set of all smooth embeddings⁵ of M in \mathbb{R}^3 , and take $e \in \text{Emb}(M, \mathbb{R}^3)$ so that

$$\Sigma = e(M).$$

Without loss of generality we can fix $M = \Sigma$, viewed as an abstract two-manifold, and $e = \text{id}$. Next, define $\varphi : M \rightarrow \mathbb{R}^2 = \mathbb{R}^2 \times \{0\}$ as

$$\varphi := \pi \circ e = \pi|_\Sigma. \quad (3.3)$$

Then Σ is in general position with respect to π if and only if the map φ is stable.⁶

Up to small deformations, we can assume that Σ is in general position with respect to π ,⁷ in the following sense.

⁵Following, e.g., [9], a smooth embedding of M into \mathbb{R}^3 is a smooth injective map having differential of rank 2 (maximal) at all points of M .

⁶See [13] for stability theorems of composite mappings.

⁷This could be achieved also in terms of small changes in the viewing direction, as in [16, 17].

Lemma 3.2.2 (Stable Scenes Up to Small Deformations) *Given $e \in \text{Emb}(M, \mathbb{R}^3)$ and a neighbourhood $U_e \subset \mathcal{C}^\infty(M, \mathbb{R}^3)$ of e , there exists $\hat{e} \in \text{Emb}(M, \mathbb{R}^3) \cap U_e$ such that $\hat{\Sigma} := \hat{e}(M)$ is in general position with respect to π .*

Proof The composition $\pi \circ e \in \mathcal{C}^\infty(M, \mathbb{R}^2)$ is a map between two 2-manifolds, and the source manifold M is closed. We recall now that the set of all stable maps from M to \mathbb{R}^2 is dense in $\mathcal{C}^\infty(M, \mathbb{R}^2)$ (see [9, p. 160, and Theorem 2.5 of Chapter 6, Proposition 3.3 of Chapter 2]). Therefore, given any neighbourhood $V \subset \mathcal{C}^\infty(M, \mathbb{R}^2)$ of $\pi \circ e$, there exists a stable map $\psi \in V$. We define $\hat{e} : M \rightarrow \mathbb{R}^3$ as

$$\hat{e}(m) := (\psi(m), z(m)) \in \mathbb{R}^2 \times \mathbb{R} = \mathbb{R}^3,$$

where $m \in M$ and $z(m)$ is the z coordinate of $e(m) \in \mathbb{R}^3$. In other words we use ψ to deform Σ in the x_1x_2 plane while keeping the z coordinate fixed.

In particular we conclude

$$\pi \circ \hat{e} = \psi. \quad (3.4)$$

Taking V small enough and recalling that $\text{Emb}(M, \mathbb{R}^3)$ is an open subset of $\mathcal{C}^\infty(M, \mathbb{R}^3)$, we obtain that $\hat{e} \in \text{Emb}(M, \mathbb{R}^3) \cap U_e$; the stability of ψ and (3.4) imply that $\hat{\Sigma} := \hat{e}(M)$ is in general position with respect to π . \square

Example 3.2.3 Figure 3.2 shows an example of Σ not in general position with respect to π .

Unless otherwise specified, from now on we shall always assume that Σ is in general position with respect to π .

Remark 3.2.4 (Critical Curve) Denote by $d(\pi|_\Sigma)$ the differential of $\pi|_\Sigma$. Since M and Σ can be identified via the embedding map e , the critical set $\text{crit}(\varphi)$ (where φ is as in (3.3)) can be identified with

$$\begin{aligned} & \left\{ \sigma \in \Sigma : \text{rank}(d(\pi|_\Sigma)(\sigma)) = 1 \right\} \\ &= \{ \sigma \in \Sigma : \text{the tangent plane to } \Sigma \text{ at } \sigma \text{ contains } \text{span}\{e_3\} \} \end{aligned}$$

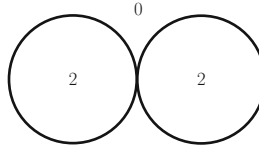


Fig. 3.2 Image taken from [3]. The scene E consists of two disjoint three-dimensional closed balls at different depth. We display the values of f_Σ (see formula (3.5)). This scene is not stable. See [9, p. 88] for more information

which, as we already know, consists of a finite set of pairwise disjoint connected components, each component being a smooth closed simple curve in Σ , and will be called the *critical curve*. At the points of the critical curve, the projection π maps the tangent plane onto a line, and e_3 generates the kernel of $d(\pi|_\Sigma)$.

Definition 3.2.5 (Apparent Contour) Let φ be as in (3.3). The apparent contour $\text{appcon}(\varphi)$ of φ , denoted by

$$G_\Sigma,$$

will be called the apparent contour of Σ .

With a small abuse of language, sometimes G_Σ will be called the apparent contour of the 3D shape E (recall (3.2)).

The reason for using the symbol G_Σ is that it reminds the notion of graph; as we shall see, graphs will be of frequent use in the next chapters. Accordingly, arcs, crossings, cusps and nodes of G_Σ are denoted by

$$\text{arcs}(G_\Sigma), \quad \text{crossings}(G_\Sigma), \quad \text{cusps}(G_\Sigma), \quad \text{nodes}(G_\Sigma),$$

respectively. The external region will be denoted by

$$\text{ext}(G_\Sigma).$$

To have a better geometric insight, let us list some of the properties of the apparent contours in this embedded case. Under our stability assumptions the apparent contour of Σ coincides with the set of all points $x \in \mathbb{R}^2$ such that there exists a point (x, z) belonging to $\pi^{-1}(x) \cap \Sigma$ where $\pi^{-1}(x)$ and Σ are nontransverse. More precisely:

- if $x \in \text{arcs}(G_\Sigma)$ then, inside the finite set of points of $\pi^{-1}(x) \cap \Sigma$ there is only one, call it (x, z) , which is a nontransverse intersection, and the intersection multiplicity at (x, z) is two.
- If $x \in \text{cusps}(G_\Sigma)$ then, inside the finite set of points of $\pi^{-1}(x) \cap \Sigma$ there is only one, call it (x, z) , which is a nontransverse intersection, and the intersection multiplicity at (x, z) is three. See also Fig. 3.3.
- If $x \in \text{crossings}(G_\Sigma)$ then, inside the finite set of points $\pi^{-1}(x) \cap \Sigma$ there are exactly two distinct points (x, z) and (x, w) which are nontransverse intersections, and the intersection multiplicity at (x, z) and at (x, w) is two. Moreover, the tangent planes to Σ at (x, z) and (x, w) are non-parallel. Figure 3.4 displays a local realization of this situation.

Fig. 3.3 A smooth fold generating a cusp in the apparent contour: here the vertical direction is the direction of the observer. Compare also with Example 3.3.3

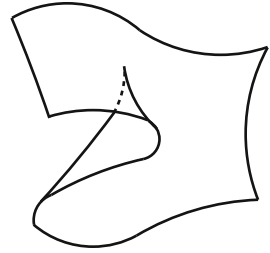
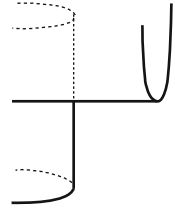


Fig. 3.4 Two smooth folds, one behind the other, generating a crossing in the apparent contour



3.3 The Function f_Σ

Assuming the factorization of G in (3.3) through the embedding e and the projection π makes it possible to give a geometric interpretation of the function $f_{\pi \circ e}$ defined in (2.2). Indeed, setting for notational simplicity

$$f_{\pi \circ e} = f_\Sigma,$$

we have

$$f_\Sigma(x) = \# \{ \pi^{-1}(x) \cap \Sigma \} = \# \{ z \in \mathbb{R} : (x, z) \in \Sigma \}, \quad x \in \mathbb{R}^2. \quad (3.5)$$

Therefore, $f_\Sigma(x)$ is the total number of intersections between Σ and the line emanating at $(x, 0)$ in the view direction. As we already know, f_Σ is constant and even on each region, and vanishes on the external region (and obviously also out of Ω). With a small abuse of notation, we denote by

$$f_\Sigma(R)$$

the constant value of f_Σ on the region $R \subset \mathbb{R}^2 \setminus G_\Sigma$. Remember that the local orientation of the apparent contour around a cusp c is such that the inside (i.e., the acute part) of the cusp lies locally on the left. Hence, in a neighbourhood of c we always have the higher number of preimages of π in the inside of the cusp (see Fig. 3.3). Accordingly, a cusp pointing to the right is always oriented upwards. The incoming arc of G_Σ at c will be sometimes denoted by c^- and the outgoing arc by c^+ . In Fig. 3.11 we show the local values of f_Σ around c .

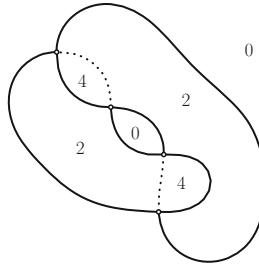


Fig. 3.5 The curve G_Σ is the apparent contour of a three-dimensional set E consisting of two interwoven connected components. The integer numbers are the values of the function f_Σ on the various regions

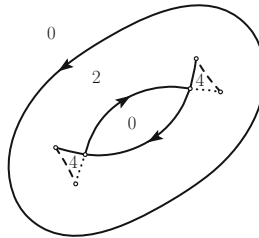


Fig. 3.6 The apparent contour of a torus, together with the values of f_Σ on the regions



Fig. 3.7 Image taken from [3]. Pair of consecutive cusps. The configuration on the right is not allowed in an apparent contour G_Σ , in view of the properties of the function f_Σ (the highest value of f_Σ is taken in the acute part of the cusp)

In Figs. 3.5 and 3.6 we show two examples of apparent contour G_Σ with the values of f_Σ on the regions; Fig. 3.5 corresponds to an occlusion between two different objects, none of them being in front of the other, and Fig. 3.6 corresponds to a torus.

In Fig. 3.7 (right) we show an example of two consecutive cusps which cannot appear in an apparent contour G_Σ .

In the last two pictures of Fig. 3.11 we show the local values of f_Σ around a crossing. If p is a crossing, the four arcs at p will be sometimes denoted by a_a^\pm and a_b^\pm (a stands for *above*, b stands for *below*) in such a way that a_a^- is incoming and is opposite to a_a^+ which is outgoing (a_a^\pm belong locally to the same component of the apparent contour). Similarly for a_b^\pm .

Definition 3.3.1 (Stratum) A stratum of region R is a pair

$$(R, r),$$

where $r \in \mathbb{N}$ is an index with $r \in \{1, \dots, f_\Sigma(R)\}$.

As a consequence, we can identify $f_\Sigma(x)$ with the number of strata of Σ in front of the point $(x, 0)$, for any x in a region.⁸

Remark 3.3.2 (The Function f_Σ on the Apparent Contour) The function f_Σ is defined at all points of \mathbb{R}^2 , not only on regions. In particular, its values on G_Σ are precisely as follows.⁹

- If a is an arc of the apparent contour of Σ , the value of f_Σ is constant on a and takes the mean value of f_Σ on the two adjacent regions. For instance, the value of f_Σ on the two vertical arcs in Fig. 3.4 is $\frac{1}{2}(2 + 4) = 3$ on the upper part, and $\frac{1}{2}(0 + 2) = 1$ on the lower one (remember that the apparent contour is oriented, and the higher value of f_Σ is taken on the region locally on the left, and equals the lower value plus two). Again with a small abuse of notation, when necessary we will denote by

$$f_\Sigma(a)$$

the value of f_Σ on the arc a .

- If x is a crossing of G_Σ , then $f_\Sigma(x)$ is the mean value of f_Σ on the four neighbouring regions. For instance, in Fig. 3.4 the value of f_Σ on the crossing is $\frac{1}{4}(0 + 2 + 2 + 4) = 2$.
- If c is a cusp of G_Σ , then $f_\Sigma(c)$ is the minimum of the values of f_Σ over the two regions adjacent to c .

Example 3.3.3 (Local Equations Around a Cusp) The simplest way to obtain an ordinary cusp point in a plane curve is to project a space twisted cubic curve, see [5, Chapter 1], [9, pp. 146,147], [8, p. 10]. Assume¹⁰ that the tangent plane to Σ at $(0, 0, 0)$ is spanned by e_1 and e_3 , and that locally around $(0, 0, 0)$ the set Σ takes the form

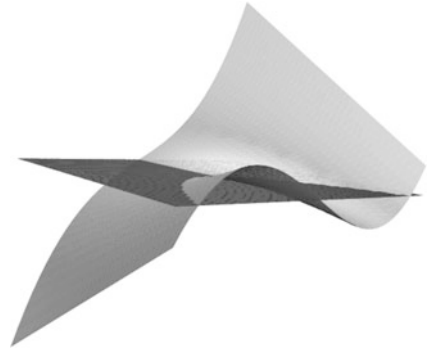
$$\{(x_1, x_2, z) : x_2 = F_c(x_1, z)\},$$

⁸The critical curve divides, with the terminology of [16], the “anterior” surfaces from the “posterior” ones.

⁹Compare with Remark 2.2.6 which deals with a more general case.

¹⁰For simplicity, in the present example, the set E is not disjoint from the retinal plane. Clearly, this assumption is irrelevant.

Fig. 3.8 Image taken from [3]. The surface represents the graph of the function F_c in (3.6). The plane is $\{(x_1, x_2, z) : x_2 = 0\}$, and is tangent to $\text{graph}(F_c)$ at the origin. The intersection of the plane and the surface consists of the parabola $\{3x_1 = z^2, x_2 = 0\}$ and the line $\{z = 0, x_2 = 0\}$



where

$$F_c(x_1, z) := \frac{1}{2} (z^3 - 3x_1 z). \quad (3.6)$$

In Fig. 3.8 we display the graph of F_c over the plane $\{x_2 = 0\}$.

Locally around $(0, 0, 0)$, Σ is parameterized as $(x_1, z) \rightarrow (x_1, F_c(x_1, z), z)$, and the tangent plane is spanned by $(1, \frac{\partial F_c}{\partial x_1}, 0)$ and $(0, \frac{\partial F_c}{\partial z}, 1)$. The condition that, on Σ , the tangent plane to Σ contains $\text{span}\{e_3\}$ (Remark 3.2.4) becomes

$$\Sigma \cap \left\{ \frac{\partial F_c}{\partial z} = 0 \right\}.$$

Therefore the *critical set* (the critical curve, in this case) takes locally the form

$$\{(x_1, x_2, z) : x_2 = F_c(x_1, z), x_1 = z^2\}, \quad (3.7)$$

and since $F_c(z^2, z) = -z^3$, it is a curve that can be parameterized locally around $(0, 0, 0)$, as

$$z \rightarrow (z^2, -z^3, z).$$

This is a smooth one-dimensional submanifold¹¹ of Σ the orthogonal projection of which, on the plane spanned by e_1 and e_3 , is the parabola $x_1 = z^2$.

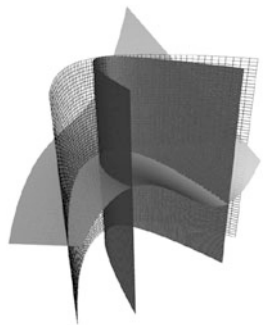
Note that the intersection multiplicity of $\pi^{-1}(0, 0) = \{x_1 = x_2 = 0\}$ with Σ is three, and the differential of the restriction of π to (3.7) vanishes at the origin. Moreover, at the origin, the order of contact [9] of $\pi^{-1}(0, 0)$ with the singular set is one.

¹¹Note that e_3 is the tangent vector to the critical curve at the origin, thus coinciding with the kernel of $d\pi|_{\Sigma}$.

Fig. 3.9 Image taken from [3]. The set $\{(x_1, x_2, z) : x_1 \geq 0, x_2 = \pm x_1^{3/2}\}$ is tangent to the graph of F_c along the singular set. The intersection contains also another curve Θ , which bounds the two external layers



Fig. 3.10 Image taken from [3]. The critical curve is the intersection of the graph of F_c with the cylinder over the plane parabola $x_1 = z^2$, while Θ lies on the cylinder $x_1 = z^2/4$. The three layers diffeomorphic to the acute part of the cusp are projected to three planar regions bounded by the two parabolae



Locally around the origin of $\mathbb{R}^2 = \mathbb{R}^2_{(x_1, x_2)} = \mathbb{R}^2 \times \{0\}$, the apparent contour of Σ takes the form

$$\{(x_1, x_2) : x_2^2 = x_1^3\} = \{(x_1, x_2) : x_1 \geq 0, x_2 = \pm(x_1)^{3/2}\}$$

i.e., a simple cusp of \mathbb{R}^2 .

The intersection of Σ with the set $\{(x_1, x_2, z) : x_2 = \pm(x_1)^{3/2}\}$ (a cylinder over the cusp $\{(x_1, x_2, 0) : x_2 = \pm(x_1)^{3/2}\}$) consists of the critical curve and of a smooth curve $\Theta \subset \Sigma$ the orthogonal projection of which, on the plane $\{x_2 = 0\}$, is given by the parabola $\{(x_1, 0, z) : x_1 = z^2/4\}$ (see Figs. 3.9 and 3.10).

Indeed, points (x_1, x_2, z) in

$$\text{graph}(F_c) \cap \{(x_1, x_2, z) : x_1 \geq 0, x_2 = \pm(x_1)^{3/2}\} \quad (3.8)$$

satisfy

$$x_1^3 - \left(\frac{z^3}{2} - \frac{3}{2}zx_1 \right)^2 = 0$$

which factorizes as

$$(x_1 - z^2)^2 \left(x_1 - \frac{z^2}{4} \right) = 0.$$

Notice also that, locally, the projection map π is a diffeomorphism outside the cusp, and it is a triple covering over the acute part

$$U_{\text{inn}} := \{(x_1, x_2) : x_1 \geq 0, -x_1^{3/2} < x_2 < x_1^{3/2}\}$$

of the cusp, see Fig. 3.8. To check that the three layers forming $\pi^{-1}(U_{\text{inn}})$ are diffeomorphic one to the other, one may project onto the coordinate plane $\{x_2 = 0\}$, with the projection map $p_{x_2}(x_1, x_2, z) := (x_1, 0, z)$. Clearly p_{x_2} restricted to $\text{graph}(F_c)$ is a diffeomorphism. The three layers we are considering are bounded by (3.8) which is the union of the critical curve and of the curve Θ with

$$p_{x_2}(\Theta) = \{x_1 = z^2/4\}.$$

Hence the images under p_{x_2} of the three layers are the plane regions bounded by the image through p_{x_2} of the critical curve and by $p_{x_2}(\Theta)$, and the three regions diffeomorphic to the three layers are therefore locally given by

$$\begin{aligned} & \left\{ (x_1, z) : z \leq 0, \frac{z^2}{4} < x_1 < z^2 \right\}, \\ & \left\{ (x_1, z) : z \geq 0, \frac{z^2}{4} < x_1 < z^2 \right\}, \\ & \{(x_1, z) : x_1 > z^2\}. \end{aligned}$$

3.4 Labelling an Apparent Contour: The Function d_Σ

The factorization of φ in (3.3) allows to enrich the apparent contour with a labelling [11], which is a function defined on the arcs of G_Σ , and has the meaning of the ordering induced by the depth. The starting point is the observation that, for any $x \in \mathbb{R}^2$, there is a natural ordering between the points of the fibre $\pi^{-1}(x)$ over x , according to the value of the last coordinate in \mathbb{R}^3 , 0 being the nearest value to the point of view. This ordering induces obviously an ordering on the points of the set $\pi^{-1}(x) \cap \Sigma$ consisting of $f_\Sigma(x)$ elements, which we number from 1 to $f_\Sigma(x)$. As a consequence, if $x \in \text{arcs}(G_\Sigma)$, we can count the number of layers of Σ in front of the corresponding point of the critical curve (see Remark 3.2.4), and the resulting number will be the labelling d_Σ .

Definition 3.4.1 (Labelling) Given a point $x \in G_\Sigma$ which is not a crossing, the intersection between $\pi^{-1}(x)$ and the critical curve consists of only one element with coordinates (x, z) . Then the function

$$d_\Sigma : G_\Sigma \setminus \text{crossings}(G_\Sigma) \rightarrow \mathbb{N}$$

is defined as follows: $d_\Sigma(x)$ is the number of points of $\pi^{-1}(x) \cap \Sigma$ having coordinates (x, ζ) with $\zeta < z$.

The function d_Σ is finite, and constant on each arc; namely, if x, x' are two points on the same arc, then $d_\Sigma(x) = d_\Sigma(x')$. To address this property, we shall often say that d_Σ is locally constant on the arcs.¹² With a small abuse of notation, if a is an arc of the apparent contour G_Σ , we will denote by

$$d_\Sigma(a)$$

the value of d_Σ on a .

Definition 3.4.2 (Labelled Apparent Contour of Σ) The pair (G_Σ, d_Σ) will be called labelled apparent contour of Σ .

When no confusion is possible, sometimes also G_Σ will be called labelled apparent contour of E .

It is useful to remember the list of compatibility conditions that f_Σ and the labelling d_Σ must satisfy (consistency of a labelling).

Remark 3.4.3 (Compatibilities) The following properties hold.

- f_Σ takes nonnegative locally constant even integer values on the regions, and vanishes on the external region. It jumps by two when passing from a region to an adjacent one.
- d_Σ takes nonnegative locally constant integer values on the arcs.
- At a crossing, d_Σ jumps by two when passing from an arc to a consecutive one, without however jumping on the other pair of consecutive arcs, and respecting the compatibility conditions in the last two pictures of Fig. 3.11. The case $d_2 > d_1$ happens when there is at least one intermediate transversal layer in between the two folds.
- At a cusp, d_Σ jumps by one when passing from an arc to the adjacent one, respecting the compatibility conditions in the second and third picture of Fig. 3.11.
- If $\bar{x} \in \text{arcs}(G_\Sigma)$,

$$0 \leq d_\Sigma(\bar{x}) \leq \liminf_{x \rightarrow \bar{x}} f_\Sigma(x). \quad (3.9)$$

¹²The function d_Σ is not defined at a crossing, where it could be defined as a multifunction taking two nonnegative integer values: we shall not need such an extension.

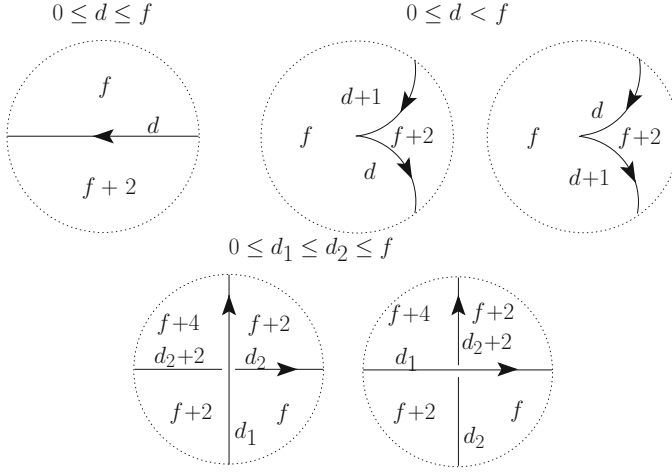


Fig. 3.11 f stands for f_Σ on the various regions; d, d_1, d_2 stand for the possible values of d_Σ on the arcs. These figures are the analog of the ones in Fig. 2.3, now enriched with the labelling. Since in the present situation we have the embedding (see (3.3)), the central picture in Fig. 2.3 concerning the cusp gives rise to two different possibilities (observe the strict inequality $d < f$) which are displayed here. In the first case d is decreasing around the cusp, while in the second case d is increasing (see Definition 8.1.1). An example is given by the apparent contour of the torus in Fig. 3.6. Similarly, the right picture in Fig. 2.3 generates two different possibilities, depending on which arc is behind the other in terms of the labelling. $d_1 < d_2$ at a crossing means that the arc labelled by d_2 is “behind” the arcs labelled by d_1 , so that there is at least one layer of Σ separating the two corresponding contours. The labelling d_Σ is therefore constant on the arcs “closest” to the eye, while it jumps across the “furthest” arcs (the higher value of d is adjacent to regions where the value of f is higher). The functions f and d must satisfy various compatibility conditions around an arc, a cusp and a crossing. See also the inequalities above the pictures. The gap shown in the pictures for arcs at a crossing is just added for visual convenience to help to distinguish the arc with larger values of d (where d jumps by two, broken arc) from the arc with smaller value of d (unbroken, closer to the eye, and called emerging; see Definition 3.4.5). Image taken from [3]

Remark 3.4.4 (The Labelling d_Σ on a Cusp) The labelling d_Σ is defined at all points of G_Σ which are not crossings. In particular, its value on a cusp c is given by the minimum of d_Σ on the two arcs of G_Σ meeting at c .

Notation The region inside (respectively outside) a cusp shall be often denoted by R_{inn} or R_{max} (respectively R_{out} or R_{min}). We shall write $f = f_{\text{inn}} = f_{\text{max}}$ in R_{max} . Around a crossing, the region where f takes the maximal (respectively minimal) value shall be denoted by R_{max} (respectively R_{min}). We shall write $f = f_{\text{max}}$ in R_{max} and $f = f_{\text{min}}$ in R_{min} .

Definition 3.4.5 (Occluding and Emerging Arcs) The two vertical arcs joining smoothly in the fourth picture of Fig. 3.11 are called occluding arcs, and the horizontal arc on the right of the occluding arcs is called emerging arc. Similarly, the two horizontal arcs joining smoothly in the fifth picture are called occluding arcs, and the vertical arc below the occluding arcs is called emerging arc.

Observe that the emerging arc is always on the right of the occluding arcs.

Lemma 3.4.6 (Parity of Cusps) *The number of cusps on a component of G_Σ is either even or zero.*

Proof At a crossing, the labelling d_Σ varies of an even ($= 0$ or 2) quantity when passing from an arc to a consecutive one. Moreover, it varies of an odd ($= 1$) quantity when passing throughout a cusp. Therefore, in order that d_Σ attains its original value at the starting point when travelling along the whole component of the apparent contour, there must be an even number of cusps.¹³ \square

A stronger version of Lemma 3.4.6 is given in Lemma 8.1.4.

Remark 3.4.7 (Number of Nodes) The number of self-intersections (if any) of a single component of an apparent contour can be either even or odd: indeed, if we insert a small loop inside one of the two parts of the figure-eight in Fig. 3.12 (left) then it is possible to check that the new graph admits a compatible labelling. On the other hand, the number of intersections of a component of an apparent contour with another (different) component is necessarily even.

Remark 3.4.8 The usefulness of the labelling stands on the fact that it encodes all three-dimensional information. Indeed, as we shall prove in Chap. 5, given a plane graph with cusps and crossings admitting a consistent labelling, it is possible to construct a stable scene¹⁴ having that graph as its apparent contour. Furthermore, the depth-equivalence class of the scene is unique. These results have been already addressed in the variational model discussed in Sect. 1.5.

Now, we describe some examples of “impossible” graphs, i.e., planar graphs for which there is no labelling satisfying all compatibility conditions.

Example 3.4.9 (Impossible Graphs) Let us consider the planar graph G in Fig. 3.12 (left). The orientation of the exterior loop is forced to be counterclockwise. On

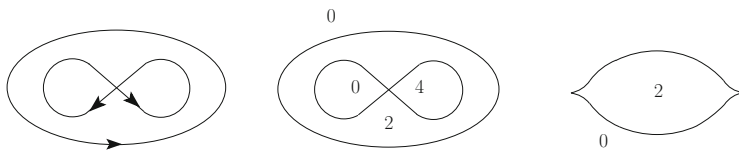
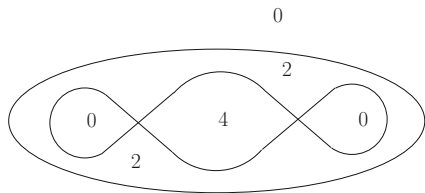


Fig. 3.12 Example 3.4.9: two examples of graphs which are not apparent contours of some Σ . Image taken from [3]

¹³For results concerning the number of cusps of apparent contours in general contexts, see [15, p. 409], [14, p. 84].

¹⁴As already anticipated in the Introduction, the existence of a labelling satisfying all compatibility conditions makes possible the construction of an abstract smooth surface M and a smooth embedding $e : M \rightarrow \mathbb{R}^3$ so that $\Sigma = e(M)$; see Theorem 5.3.1 for a precise statement.

Fig. 3.13 This graph is the apparent contour of three-dimensional scene. Image taken from [3]



the other hand, reversing the orientation of the figure-eight would not change the final conclusion, so we fix the orientation as in the figure. Accordingly, in Fig. 3.12 (centre) we display the values of twice the winding number in the various regions. To check that G admits no labelling, we argue as follows. From inequality (3.9), it follows $d = 0$ on the exterior loop and on the left part of the figure-eight. Then there is no choice of d on the right part of the figure-eight, which is consistent with the compatibility conditions specified in the last two pictures of Fig. 3.11. Hence G cannot be the apparent contour of a three-dimensional scene. The same assertion holds for the graph in Fig. 3.12 (right). Indeed, once we properly orient the graph (counterclockwise), twice the winding number on the regions takes the values indicated in the figure. Then, any choice of d (necessarily $d = 0$ on the lower arc and $d = 1$ on the upper one, or vice versa) leads to a contradiction with inequality (3.9). On the other hand, the graph in Fig. 3.13 is the apparent contour of a three-dimensional scene: define, for instance, $d = 0$ everywhere on G except for the central lower arc connecting the two crossings, where $d = 2$.

Remark 3.4.10 (Positive and Negative Cusps on Labelled Apparent Contours) The presence of a labelling on $\text{appcon}(\varphi)$ provides a standard orientation of M , obtained by identifying M with an embedded closed surface in \mathbb{R}^3 (see Theorem 5.1.1).¹⁵ In this case, recalling Definition 2.2.12, it turns out that $c \in \text{cusps}(\text{appcon}(\varphi))$ is positive (respectively negative) if the minimum of the labelling on the two arcs incident at c is even (respectively odd). This can be checked, for instance, in Fig. 8.1¹⁶; in this case, the right figure can be identified with M , and a positive small circle around the upper cusp has the opposite orientation with respect to a positive small circle around the lower cusp, due to the fact that M is oriented (by a normal vector field).

¹⁵Recall that a closed surface embedded in \mathbb{R}^3 encloses an interior, hence an outward normal is well defined; see, for instance, [10, p. 89].

¹⁶Notice carefully that the signs $+$ and $-$ in Fig. 8.1 refer to the embedding sign of a cusp (Definition 8.1.2), and not to the notion of positivity and negativity of Definition 2.2.12.

3.5 Ambient Isotopic and Diffeomorphically Equivalent Labelled Apparent Contours

In Sect. 2.4 we have introduced certain isotopies and a notion of equivalence between apparent contours. These notions, adapted to labelled apparent contours, read as follows.

Definition 3.5.1 (Ambient Isotopic Labelled Apparent Contours) We say that two labelled apparent contours $(G_{\Sigma_1}, f_{\Sigma_1}, d_{\Sigma_1})$, $(G_{\Sigma_2}, f_{\Sigma_2}, d_{\Sigma_2})$ are C^∞ ambient isotopic (briefly, ambient isotopic), if there exists an \mathbb{R}^2 -ambient isotopy $h : \mathbb{R}^2 \times [0, 1] \rightarrow \mathbb{R}^2$ with compact support such that

$$h_1(G_{\Sigma_1}) = G_{\Sigma_2},$$

such that¹⁷

$$f_{\Sigma_2} \circ h_1 = f_{\Sigma_1},$$

and such that if we extend the function d_{Σ_1} (respectively d_{Σ_2}) out of the arcs of G_{Σ_1} (respectively of G_{Σ_2}) to a given constant value, say $+\infty$ for definitiveness, then the following diagram

$$\begin{array}{ccc} \mathbb{R}^2 & \xrightarrow{h_1} & \mathbb{R}^2 \\ \downarrow d_{\Sigma_1} & & \downarrow d_{\Sigma_2} \\ \mathbb{N} \cup \{+\infty\} & \xrightarrow{\text{id}} & \mathbb{N} \cup \{+\infty\} \end{array}$$

is commutative, namely $d_{\Sigma_2} \circ h_1 = d_{\Sigma_1}$.

According to the discussion in Sect. 2.4, we can now give the following definition.

Definition 3.5.2 (Equivalence of Labelled Apparent Contours) We say that two labelled apparent contours $(G_{\Sigma_1}, f_{\Sigma_1}, d_{\Sigma_1})$, $(G_{\Sigma_2}, f_{\Sigma_2}, d_{\Sigma_2})$ are diffeomorphically equivalent if there exists $\vartheta \in \text{Diff}_c(\mathbb{R}^2)$ such that

$$\vartheta(G_{\Sigma_1}) = G_{\Sigma_2},$$

$$f_{\Sigma_2} \circ \vartheta = f_{\Sigma_1},$$

¹⁷Hence h_1 preserves the orientation of the arcs.

and the functions $d_{\Sigma_1}, d_{\Sigma_2}$, extended as in Definition 3.5.1, make commutative the following diagram

$$\begin{array}{ccc} \mathbb{R}^2 & \xrightarrow{\vartheta} & \mathbb{R}^2 \\ \downarrow d_{\Sigma_1} & & \downarrow d_{\Sigma_2} \\ \mathbb{N} \cup \{+\infty\} & \xrightarrow{\text{id}} & \mathbb{N} \cup \{+\infty\} \end{array}$$

namely, $d_{\Sigma_2} \circ \vartheta = d_{\Sigma_1}$.

Similarly to the abstract case (Theorem 2.4.3) Definitions 3.5.1 and 3.5.2 are equivalent; indeed, the following result.

Theorem 3.5.3 (\mathbb{R}^2 -Ambient Isotopies and Diffeomorphisms) *Two labelled apparent contours are ambient isotopic if and only if they are diffeomorphically equivalent.*

Proof The proof is the same as in Theorem 2.4.3. □

3.6 Visible Contours

The arcs of G_Σ where d_Σ vanishes form the *visible* part of the contour, called visible contour, since there are no layers of Σ in front of the corresponding points of the singular set. They play an important role, in particular in Chaps. 1, 4 and 9.

Definition 3.6.1 (Visible Arcs and Visible Contour) An arc where $d_\Sigma = 0$ is called a visible arc. The visible contour of G_Σ is defined as follows:

$$\text{vis}(G_\Sigma) := \text{closure of } \{x \in \text{arcs}(G_\Sigma) : d_\Sigma(x) = 0\}.$$

The visible contour is usually drawn with thick arcs in the figures.

In Fig. 3.14 we draw the apparent contour, its visible part and the labelling, for the corresponding scene in Fig. 3 of the Introduction.

Remark 3.6.2 (Structure of the Visible Contour) The structure of the visible contour shall be analysed in Chap. 4, as well as necessary and sufficient conditions for a plane graph to be the visible part of an apparent contour. It is clear that $\text{vis}(G_\Sigma)$ is oriented, its arcs are smooth, and that a node of G_Σ belonging to $\text{vis}(G_\Sigma)$ is either a T-junction or a terminal point. Moreover, at a terminal point the local shape of the incident arc is diffeomorphic to the arc of the semicubic curve (t^2, t^3) . We also anticipate here¹⁸ that:

¹⁸This list of properties originates the definition of *visible contour graph* given in Chap. 4: indeed, the visible contour is a visible contour graph in the sense of Definition 4.1.8.

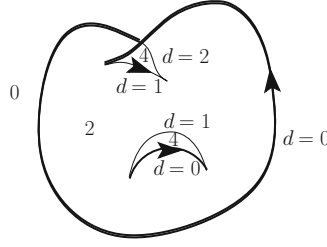


Fig. 3.14 The *bold graph* represents the visible contour of the three-dimensional scene in Fig. 3 of the Introduction; the *whole graph* represents its apparent contour (all cusps are ordinary cusps). We also depict the corresponding values of f_Σ and $d = d_\Sigma$

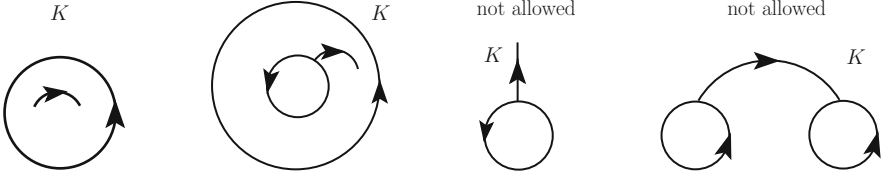


Fig. 3.15 The plane graphs in the *first two figures* satisfy all properties listed in Remark 3.6.2. The graph in the *third figure* has a terminal point which is adjacent to the external region. The arc connecting the *two circles* in the last figure has the external region also on the *left*. Image taken from [4]

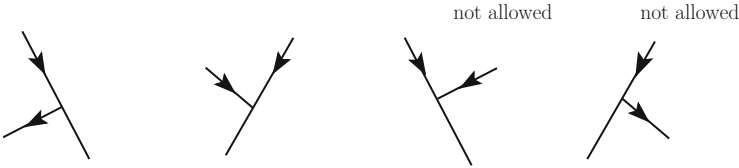


Fig. 3.16 *First two figures*: allowed orientations around a T-junction of $\text{vis}(G_\Sigma)$. *Last two figures*: forbidden orientations. Compare also with (K4) of Definition 4.1.8. Image taken from [4]

- a terminal point of $\text{vis}(G_\Sigma)$ cannot lie in the external region. For instance, the graph K in the third picture of Fig. 3.15 cannot be the visible part of an apparent contour.
- Any arc a of $\text{vis}(G_\Sigma)$ which is adjacent to the external region is oriented in such a way that the external region lies locally only on the right of a . For instance, the graph K in the last picture of Fig. 3.15 cannot be the visible part of an apparent contour.
- At a T-junction the transversal (or emerging) arc lies locally on the right of the two occluding arcs. Emerging arcs are the nearly horizontal arcs in Fig. 3.16.
- At a T-junction the two occluding arcs must join smoothly and with consistent orientation, and the emerging arc must be transverse.

- If we define the invisible contour of G_Σ as

$$\{x \in \text{arcs}(G_\Sigma) : d_\Sigma(x) > 0\},$$

then any of its arcs cannot enter the external region.

References

1. Arnold, V.I., Goryunov, V.V., Lyashko, O.V., Vassiliev, V.A.: Dynamical Systems VIII. In: Arnold, V.I. (ed.) *Singularity Theory. II. Applications*. Springer, Berlin (1993)
2. Arnold, V.I., Gusein-Zade, S.M., Varchenko, A.N.: *Singularities of Differentiable Maps, Volume I*. Birkhäuser, Boston (1985)
3. Bellettini, G., Beorchia, V., Paolini, M.: Topological and variational properties of a model for the reconstruction of three-dimensional transparent images with self-occlusions. *J. Math. Imaging Vision* **32**, 265–291 (2008)
4. Bellettini, G., Beorchia, V., Paolini, M.: Completion of visible contours. *SIAM J. Imaging Sci.* **2**, 777–799 (2009)
5. Bruce, J.W., Giblin, P.J.: *Curves and Singularities. A Geometrical Introduction to Singularity Theory*, 2nd edn. Cambridge University Press, Cambridge (1992)
6. Carter, J.S., Kamada, S., Saito, M.: Surfaces in 4-Space, *Encyclopaedia of Mathematical Sciences*, vol. 142. Springer, New York (2004)
7. Carter, J.S., Saito, M.: *Knotted Surfaces and Their Diagrams*, *Mathematical Surveys Monographs*, vol. 55. American Mathematical Society, Providence, RI (1998)
8. Francis, G.K.: *A Topological Picturebook*. Springer, New York (2007)
9. Golubitsky, M., Guillemin, V.: *Stable Mappings and Their Singularities*, *Graduate Texts in Mathematics*, vol. 14. Springer, New York/Heidelberg/Berlin (1974)
10. Guillemin, V., Pollack, A.: *Differential Topology*. Englewood Cliffs, NJ/Prentice-Hall (1974)
11. Huffman, D.A.: Impossible objects as nonsense sentences. In: Meltzer, B., Michie, D. (eds.) *Machine Intelligence*, vol. 6. American Elsevier, New York (1971)
12. Koenderink, J.J.: *Solid Shape*. MIT, Cambridge, MA/London (1990)
13. Nakai, I.: Topological stability theorem for composite mappings. *Ann. Inst. Fourier* **39**, 459–500 (1989)
14. Thom, R.: Les singularités des applications différentiables. *Ann. Inst. Fourier* **6**, 43–87 (1956)
15. Whitney, H.: On singularities of mappings of Euclidean spaces. I. Mappings of the plane into the plane. *Ann. Math.* **62**, 374–410 (1955)
16. Williams, L.R.: *Perceptual completion of occluded surfaces*. Ph.D. dissertation, Department of Computer Science, University of Massachusetts, Amherst, MA (1994)
17. Williams, L.R.: Topological reconstruction of a smooth manifold-solid from its occluding contour. *Int. J. Comput. Vision* **23**, 93–108 (1997)

Chapter 4

Solving the Completion Problem

Following [1],¹ in this chapter we show how to solve the completion problem, namely we characterize those oriented plane graphs that are visible part of an apparent contour (Theorem 4.3.1). The proof is generalized to the case where the background is not reduced to the external region. In our presentation we need some elementary concepts of the theory of oriented graphs, and the Morse description of a graph, as outlined in Sect. 2.5.3. In Chap. 9 we describe a code that automates the construction of the proof of Theorem 4.3.1.

4.1 Some Concepts from Graph Theory

In this section the properties of an apparent contour described in Chaps. 2 and 3² will be promoted to definitions. The usefulness of these definitions is validated by the completion theorem.

A nonempty plane *oriented* graph H is a graph with oriented arcs³ *embedded* in \mathbb{R}^2 . We are concerned only with compact plane graphs; for definiteness and without loss of generality, all graphs that we consider are contained in the open square $\Omega = (0, 1) \times (0, 1)$.

By

$$\text{nodes}(H)$$

¹With kind permission from SIAM Journal on Imaging Sciences: in this chapter we illustrate the results and report some of the figures from the quoted paper [1].

²In particular, the local and global structures of an apparent contour and the labelling.

³As in Chap. 2, an arc of H is relatively open.

we denote the set of nodes (or vertices) of H . We always assume that the graph is finite, namely it has a finite number of nodes. Observe that we do not require H to be connected.

We stress that we allow H to contain *closed arcs* and *loops*,⁴ which are defined as follows.

Definition 4.1.1 (Closed Arcs and Loops) Closed arcs of H are Jordan curves that have neither starting nor ending points, hence they do not contain any node. Loops, on the other hand, are arcs of H that start and end at the same node.

According to the notation introduced in Sect. 2.2, the connected components of the complement of H in \mathbb{R}^2 are called *regions* of H . The unbounded region will be called the *external region* of H , and denoted by

$$\text{ext}(H).$$

Concerning the regularity of H , we assume that H is *continuous*, which means that

$$H = \gamma(S),$$

with $\gamma : S \rightarrow \mathbb{R}^2$ a continuous map, where S is the disjoint union of a finite number (possibly zero) of copies of the unit interval $[0, 1]$ and of a finite number (possibly zero) of copies of \mathbb{S}^1 .⁵

Denoting by ∂S the boundary of S , i.e., the disjoint union of all the endpoints of the unit intervals, we have that $\text{nodes}(H) = \gamma(\partial S)$.

Definition 4.1.2 (Degree of a Node) The degree of a node p of H is the total number of incident arcs at p , and it is denoted by

$$\deg(p),$$

with the convention that loops are counted twice.

Definition 4.1.3 (Consecutive, Adjacent and Opposite Arcs) Two arcs of H are *consecutive* at a common node p if p is the terminal node (head) of an arc and the starting node (tail) of the other. Two arcs of H are *adjacent* at the node p if they are incident in p and locally they are adjacent to the same region. If p is a node of H of degree 4, we say that two incident arcs are *opposite* if they are not adjacent.

⁴The presence of closed arcs makes our notion of graph different from the classical one.

⁵The image of $[0, 1]$ is an oriented arc (possibly oriented loop) and the image of \mathbb{S}^1 is an oriented closed arc.

4.1.1 Contour Graphs and Visible Contour Graphs

For the discussion on completion of graphs, it is convenient to introduce the following concept.

Definition 4.1.4 (Contour Graph) We say that H is a contour graph if $H = \gamma(S)$ is a nonempty compact finite oriented continuous plane graph, and the following conditions hold:

- (i) if $t_1, t_2 \in S$ are such that $\gamma(t_1) = \gamma(t_2)$, then either $t_1 = t_2$ or $t_1, t_2 \in \partial S$.
- (ii) If $p \in \text{nodes}(H)$, then $\deg(p) \in \{1, 2, 3, 4\}$, and
 - (ii-1) if $\deg(p) = 1$, we refer to p as a *terminal point*;
 - (ii-2) if $\deg(p) = 2$, we refer to p as a *cusp*. In this case the two arcs incident in p must be consecutive;
 - (ii-3) if $\deg(p) = 3$, we refer to p as a *T-junction*. In this case we distinguish one of the incident arcs and we call it *transversal*; the remaining two arcs are called *occluding*, and must be consecutive;
 - (ii-4) if $\deg(p) = 4$, we refer to p as a *crossing*; each pair of opposite arcs incident at p must be consecutive.

Remark 4.1.5 Condition (i) means that distinct arcs can only meet at a node. The terminology that we use in (ii-2) (respectively in (ii-3)) is motivated by the fact that p will play the role of a semicubic cusp point (respectively a T-junction) in a visible contour.⁶

Remark 4.1.6 (Components) Note that

- we allow for loops, and for multiple arcs that connect the same two nodes.
- In view of the consecutiveness assumptions, we can always *continue* arcs (i.e., follow the corresponding consecutive arc) that end on a cusp or on a crossing, or if the arc is an *occluding* arc of a T-junction. In case of a crossing, *continuation* arcs must be opposite. The *components* of H that we obtain by continuing arcs as far as possible are either closed oriented arcs or arcs that start and end at a terminal point or at a T-junction as transversal arcs.

Definition 4.1.7 (Interior Terminal Points) Let H be a contour graph. We say that a terminal point of H is interior if it is not adjacent to $\text{ext}(H)$.

Remembering the structural properties of a visible contour listed in Remark 3.6.2, we now give the following definition.^{7,8}

⁶Compare Sect. 2.2.

⁷The smoothness requirements (K1), (K5) and (K6) in Definition 4.1.8, as well as conditions (G1) and (G4) in Definition 4.2.2 below, are not essential for the proof of Theorem 4.3.1. However, they are important for the reconstruction results of a 3D scene described in Chap. 5.

⁸Definition 4.1.8 is better understood looking also at Fig. 3.16 of Chap. 3.

Definition 4.1.8 (Visible Contour Graph) Let K be a contour graph. We say that K is a visible contour graph if the following properties hold:

- (K1) the arcs of K are of class C^∞ ;
- (K2) any node of K is either a T-junction or a terminal point;
- (K3) any arc a of K adjacent to $\text{ext}(K)$ is oriented in such a way that $\text{ext}(K)$ lies locally only on the right of a ;
- (K4) at a T-junction of K the *transversal* arc lies locally on the right of the two occluding arcs. Such an arc will also be called the *emerging* arc of p ;
- (K5) at a T-junction of K the two occluding arcs join in a C^∞ way and with consistent orientation, and the emerging arc is transverse;
- (K6) the local shape of the incident arc at a terminal point of K is smoothly diffeomorphic to one of the two arcs of the semicubic curve $(t^2, -t^3)$ with t in a neighbourhood of the origin. The orientation of the arc must be consistent to that of the semicubic curve as given by increasing values of t .

With a small abuse of language, the union of the two occluding arcs at a T-junction will be sometimes referred to as the occluding arc.

Remark 4.1.9 (Terminal Points Are Interior) Requirement (K3) implies that all terminal points of K are interior.

An example of contour graph where property (K4) is not satisfied is given in Fig. 9.12 of Chap. 9; see also Fig. 9.13.

Remark 4.1.10 (Background) In Definition 4.1.8 we could substitute the set $\text{ext}(K)$ with a larger set $\text{background}(K)$ (*background*), union of some of the regions of K including the *external* region, and check whether requirement (K3) still holds with $\text{background}(K)$ in place of $\text{ext}(K)$. If this is the case, then $\text{background}(K)$ will be called an *admissible background*. This remark is motivated by the fact that the completion result stated in Theorem 4.3.1 can be obtained in such a way that $\text{background}(K)$ coincides with the set $\{f = 0\}$ of the reconstructed apparent contour. Note that the notion of interior terminal point p in Definition 4.1.7 will also change accordingly, by requiring p not to be adjacent to the larger set $\text{background}(K)$. See also Remark 4.3.4.

Example 4.1.11 Let K be the visible part of an apparent contour of a smooth closed surface $\Sigma = \partial E$ embedded in \mathbb{R}^3 as in Sect. 3.6, namely

$$K := \text{vis}(G_\Sigma).$$

Then K is a visible contour graph. As we shall see, the content of the completion theorem and the results of Chap. 5 allow to prove that also the converse statement holds.⁹

⁹It is not difficult to see that the conditions defining a visible contour graph K are necessary for K to be the visible part of an apparent contour; compare, for instance, Sect. 3.6.

Notation According also to the notation used in Chap. 3, a visible contour graph shall usually be denoted with the symbol K . In addition, it will be usually depicted with thick arcs in the figures.

Remark 4.1.12 (Number of T-Junctions and Terminal Points) Let K be a visible contour graph. It follows from basic graph theory that if q_1, \dots, q_h are the T-junctions of K and p_1, \dots, p_k are the terminal points (where h, k are two nonnegative integer numbers), then $h + k$ is even; this also follows from the fact that $h + k$ is twice the number of nonclosed components of K , where the components are introduced in Remark 4.1.6.

4.2 Complete Contour Graphs and Labelling

Let us start with the following definition.

Definition 4.2.1 (Winding Number) Let H be a contour graph. The winding number of H ,

$$w(H, \cdot) : \mathbb{R}^2 \setminus H \rightarrow \mathbb{Z},$$

is defined as follows:

$$w(H, x) := \sum_{\substack{\gamma: \mathbb{S}^1 \rightarrow \mathbb{R}^2 \\ \gamma(\mathbb{S}^1) \text{ component of } H}} w(\gamma, x), \quad x \in \mathbb{R}^2 \setminus H.$$

Part of the following definition is suggested by the relations between the function f_φ and the winding number described in Lemma 2.2.9.

Definition 4.2.2 (Complete Contour Graph) Let G be a contour graph. We say that G is a complete contour graph if the following properties hold:

- (G1) the arcs of G are of class \mathcal{C}^∞ up to their relative closure;
- (G2) if $p \in \text{nodes}(G)$, then p is either a crossing or a cusp, and in this case we write $p \in \text{crossings}(G)$ and $p \in \text{cusps}(G)$ respectively;
- (G3) the function $f : \mathbb{R}^2 \setminus G \rightarrow 2\mathbb{Z}$ defined as

$$f(x) := 2w(G, x), \quad x \in \mathbb{R}^2 \setminus G, \quad (4.1)$$

is nonnegative;

- (G4) at a crossing of G the opposite arcs join in a \mathcal{C}^∞ way and the two pairs of opposite arcs cross transversally;
- (G5) at a cusp of G the local shape of the two incident arcs is smoothly diffeomorphic to the semicubic curve $(t^2, -t^3)$ with t in a neighbourhood

of the origin. The orientation of the arcs must be consistent to that of the semicubic curve as given by increasing values of t .

Crossings of G will be also called transversal crossings.

Remark 4.2.3 The function f defined in (G3) has the following properties:

- it is locally constant, i.e., it is constant on each region of the complement of G ;
- $f = 0$ in $\text{ext}(G)$;
- if $x \in G \setminus \text{nodes}(G)$ and if $f_+(x)$, $f_-(x)$ are the values of f on the two sides of x , then

$$|f_+(x) - f_-(x)| = 2;$$

- along an arc of G , the region where the value of f is higher lies locally on the left.

Remark 4.2.4 Requirement (G5) in particular excludes the (impossible) configuration of Fig. 3.7, right. See also the discussion about the orientation of the apparent contour near a cusp at the beginning of Sect. 3.3.

We have already discussed in Chap. 3 what is a labelling for an apparent contour, see in particular Sect. 3.4. We now define a similar concept for a complete contour graph. In Theorem 4.3.1, one of the main difficulties is due to the fact that one has to construct a completion of a visible graph, *together with a labelling*.

Definition 4.2.5 (Labelling on a Complete Contour Graph) Let G be a complete contour graph and let $f : \mathbb{R}^2 \setminus G \rightarrow 2\mathbb{N}$ be twice the winding number of G . A labelling of G is a function

$$d : G \setminus \text{nodes}(G) \rightarrow \mathbb{N}$$

satisfying the following properties:

- (L1) d is locally constant, i.e., d is constant on each arc of G .
- (L2) Let a be an arc of G and $d(a)$ be the value of d on a . If R_1 , R_2 are the two regions adjacent to a and $f(R_1)$ and $f(R_2)$ are the values of f in R_1 and R_2 respectively, then

$$d(a) \leq \min\{f(R_1), f(R_2)\}.$$

- (L3) The compatibility conditions between f and d in Fig. 3.11 must be satisfied at the crossings and at the cusps of G .

The labelling function d satisfying all properties listed in Definition 4.2.5 will be sometimes called *consistent labelling*.

We shall occasionally need to extend the function d at the cusps of G , as the minimum value of d in the two adjacent arcs; in this way, the function d will be defined on $G \setminus \text{crossings}(G)$ (with values in \mathbb{N}).

Definition 4.2.6 (Complete Labelled Contour Graph) A complete contour graph endowed with a labelling is called a complete labelled contour graph.

We can easily provide a number of necessary conditions for a complete contour graph to admit a labelling; for example, we cannot have cusps adjacent to the external region, see the second point in Remark 2.2.10. Nevertheless, we do not have an elementary characterization (which does not amount to test all possible feasible values of d) of those complete contour graphs that admit at least one labelling. An example of complete contour graph admitting more than one consistent labelling is the one discussed in Chap. 1, in connection with Fig. 1.3.

Remark 4.2.7 (Parity of Cusps) It readily follows from Definition 4.2.5 that the number of cusps of each component of a complete labelled contour graph is even.¹⁰

Notation with a slightly redundant notation, we sometimes denote by (f, d) a labelling and by

$$(G, f, d)$$

a complete labelled contour graph, even if f is directly deduced from G via formula (4.1).

Example 4.2.8 (Apparent Contour as a Complete Labelled Contour Graph) Let $\Sigma = \partial E$ be a smooth closed surface embedded in \mathbb{R}^3 , in general position with respect to π . Following the notation of Sects. 3.2–3.4, let G_Σ be the apparent contour of Σ , f_Σ be as in (3.5), and d_Σ be the labelling of Definition 3.4.1. Then it follows from the discussion in Chap. 3 that

$$(G_\Sigma, f_\Sigma, d_\Sigma)$$

is a complete labelled contour graph, that we shall also call labelled apparent contour.

Definition 4.2.9 (Visible and Invisible Arcs) Let (G, f, d) be a complete labelled contour graph. An arc of G where $d = 0$ is called a visible arc of G , and the closure of the set $\{d = 0\}$ is called the *visible part* of G and it is denoted by

$$\text{vis}(G).$$

An arc of G having $d > 0$ is called an invisible arc, and the complement $G \setminus \text{vis}(G)$ of the visible part of G is called the invisible part of G .

Remark 4.2.10 The visible part of G is a contour graph having as nodes only T-junctions and terminal points: a T-junction corresponds to a transversal crossing of G , while a terminal point corresponds to a cusp of G . In addition, the visible part of

¹⁰It is sufficient to repeat the proof of Lemma 3.4.6 in Chap. 3; compare also with Lemma 8.1.4.

G satisfies conditions (K1)–(K6) listed in Definition 4.1.8, hence the visible part of G is a visible contour graph.

We are now in a position to give a precise meaning to the notion of completion of a graph.

Definition 4.2.11 (Completable Contour Graphs) Let K be a contour graph. We say that K is completable if there exists a complete labelled contour graph G such that K is the visible part of G .

Definition 4.2.12 (Completion) When K and G are as in Definition 4.2.11, the graph G is called a completion of K .

Notation: according to our previous notation, and for clarity of exposition, when necessary in the figures the visible part of G will be depicted with thick arcs, and its invisible part with thin arcs.

4.3 Statement of the Completion Theorem

The completion theorem characterizes those graphs which are the visible part of an apparent contour. As we shall see, the proof is constructive and relies on a Morse description of K described in Sect. 4.4.

Theorem 4.3.1 (Completion) *Let K be a contour graph. Then K is completable if and only if K is a visible contour graph.*

In other words, let K be a plane graph oriented in such a way that the unbounded connected component of $\mathbb{R}^2 \setminus K$ lies locally only on the right of K . Suppose that the nodes of K are only T-junctions and terminal points, that all terminal points of K are interior (Remark 4.1.9) and that at a T-junction condition (K4) is met. Then there exist a complete contour graph G and a labelling of G such that K is the visible part of G . Since from the results of Chap. 5 it follows that G is originated from a smooth closed surface (not necessarily connected) $\Sigma = \partial E$ embedded in \mathbb{R}^3 , we conclude that K is completable if and only if there exists a 3D shape E such that

$$K = \text{vis}(G_\Sigma).$$

Before starting the proof of the theorem, it is worthwhile to make two other observations.

Remark 4.3.2 (Nonuniqueness) The completion of K constructed in the proof of the theorem is nonunique; compare, for instance, with Example 4.6.4.

Remark 4.3.3 (On the Optimality of a Completion) We have already observed that the proof of Theorem 4.3.1 is constructive; however, in many cases the resulting completion could be very far from “optimal”. The concept of optimality here is

difficult to make rigorous: we can only say that, in certain situations, it may happen that the completion given by the algorithmic proof of Theorem 4.3.1 does not provide (even topologically) the 3D shape E usually inferred by the human brain. In other words, the reconstruction of the stable 3D scene E having G as apparent contour and K as visible apparent contour, is not unique, and the choice of the “best” reconstruction may depend on a subjective interpretation of the scene.

Remark 4.3.4 In the proof of Theorem 4.3.1 we construct a complete labelled contour graph (G, f, d) so that $K = \text{vis}(G)$. For simplicity of exposition of the iterative algorithm in the proof, we construct the labelling in such a way that

$$\{f = 0\} = \text{ext}(K);$$

see requirement (P2_i) in the proof. This constraint can be removed, by assigning a priori a part of the plane, $\text{background}(K)$ (*background*), containing $\text{ext}(K)$, where $\{f = 0\}$; compare with Remark 4.1.10.

In the proof of Theorem 4.3.1 we describe the graphs using a family m of horizontal Morse lines $m_\lambda(\mathbb{R})$ travelling from top to bottom. The proof, based on an iterative argument, consists in showing that a completion with a labelling of K in correspondence of the i -th level of the Morse function can be continued up to the $(i + 1)$ -th level. One has to consider all types of Morse events and show that *locally* the completion of the graph can be continued beyond such events. A number of new *invisible* arcs, specified by the order of their intersection with the Morse line and by a value of the labelling d , are added during this process. Instead of giving explicit formulas of these intersections, we shall simply graphically illustrate the local shape of such invisible arcs in the neighbourhood of the Morse line. Roughly speaking, these invisible arcs are continued downwards as far as possible, till one meets as a Morse event either a local minimum or the global minimum of K . The operation needed at this final stage is a nonlocal one, and indeed this is the most delicate part of the proof, since one needs to join together different invisible arcs in order to “close the regions”. This will be the *global* part of the proof.

Remark 4.3.5 The invisible arcs introduced in our construction can intersect K only at a T-junction or at a terminal point, giving raise to a crossing or a cusp respectively, or they can cross a visible arc with a value of d that jumps of two units with strictly positive values, thus creating a new crossing. These invisible arcs will be glued together only in a basin around a local minimum of K , while around a T-junction or a terminal point they will be simply extended downwards.

When λ crosses a critical value, only a few of the segments in which $m_\lambda(\mathbb{R})$ is divided by the transversal intersections with K are involved in the Morse event; all dangling invisible arcs contained in the remaining segments (those not involved in the Morse event) are simply extended smoothly downwards without creating any local minimum/maximum, and without intersecting each other (hence in particular keeping their relative ordering from left to right).

Remark 4.3.6 (Keeping the External Region or the Background) Since we cannot insert any invisible arc in the region $\text{ext}(K)$, it will follow at the end of the proof that

$$\text{ext}(K) = \text{ext}(G).$$

Furthermore we have the even stronger condition $\text{ext}(K) = \{f = 0\}$. If we substitute the larger set $\text{background}(K)$ in place of $\text{ext}(K)$ in $(P2_i)$ of the proof (provided condition (K3) of Definition 4.1.8 is met with $\text{background}(K)$ in place of $\text{ext}(K)$), then this condition becomes

$$\text{background}(K) = \{f = 0\}.$$

4.4 Morse Descriptions of a Visible Contour Graph

The structural properties of a graph is encoded in a Morse description, which is based on the existence of a Morse height function. We have already discussed this argument in Sect. 2.5 of Chap. 2. Remember that in Sect. 2.5 we have defined what is a one-parameter family $m : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}^2$ of Morse lines traversing \mathbb{R}^2 , and the meaning of $s(p)$, $\lambda(p)$ and m_λ .¹¹ Following what we have done in Sect. 2.5.2 for apparent contours, we introduce the notion of descriptive map of a visible contour graph.

More precisely, let K be a visible contour graph. Given $\lambda \in \mathbb{R}$, we say that m_λ is *generic* if $m_\lambda([0, 1]) \cap \text{nodes}(K) = \emptyset$, $m_\lambda([0, 1]) \cap K$ is finite, and each intersection is transverse.

We now give the following definition.¹²

Definition 4.4.1 (Descriptive Map of a Visible Contour Graph) Let K be a visible contour graph. We say that m is a descriptive map of K if:

- (1) there exists a finite set $\{\lambda_1, \dots, \lambda_n\}$ of real numbers, called critical values, with $1 > \lambda_1 > \lambda_2 > \dots > \lambda_n > 0$, such that that $m_\lambda(\cdot)$ is generic for any $\lambda \in [0, 1] \setminus \{\lambda_1, \dots, \lambda_n\}$;

¹¹Recall also that $m_0([0, 1]) \cap K = m_1([0, 1]) \cap K = \emptyset$.

¹²See Figs. 4.1 and 4.2a for (m1), Figs. 4.8 and 4.10 for (m2), Figs. 4.5a and c for (m5), Figs. 4.6a, 4.7a for (m6), Fig. 4.4a and b for (m3), Fig. 4.4c and d for (m4).

- (2) for any $\bar{\lambda} \in \{\lambda_1, \dots, \lambda_n\}$ there exists a unique $\bar{s} \in (0, 1)$ such that $m_{\bar{\lambda}}([0, 1] \setminus \{\bar{s}\})$ intersects K in a finite set of points and each intersection is transverse and is not a node of K . Moreover, if we set $p := m_{\bar{\lambda}}(\bar{s}) \in K$ and $\bar{\ell} := m_{\bar{\lambda}}(\mathbb{R})$, we have exactly one of the following Morse events:
- (m1) $p \notin \text{nodes}(K)$ and K lies locally below $\bar{\ell}$ (p is a strict *local maximum* of K with respect to $\bar{\ell}$);
 - (m2) $p \notin \text{nodes}(K)$ and K lies locally above $\bar{\ell}$ (p is a strict *local minimum* of K with respect to $\bar{\ell}$);
 - (m3) p is a terminal point, with the arc lying locally above $\bar{\ell}$ (ending terminal point); moreover, $\bar{\ell}$ is tangent to the visible arc at p ;
 - (m4) p is a terminal point, with the arc lying locally below $\bar{\ell}$ (starting terminal point); moreover, $\bar{\ell}$ is tangent to the visible arc at p ;
 - (m5) p is a T-junction, and locally two arcs of K , one of which being the transversal arc, lie above $\bar{\ell}$ and the third one lies below $\bar{\ell}$ (ending T-junction);
 - (m6) p is a T-junction, and locally two arcs of K , one of which being the transversal arc, lie below $\bar{\ell}$ and the third one lies above $\bar{\ell}$ (starting T-junction).

According to the orientation of the arcs of K , each of the items (m1)–(m6) split in a certain number of different cases (compare, for instance, with Example 4.4.3), as we shall see in the proof of Theorem 4.3.1.

Remark 4.4.2 (Tangency of the Morse Line at a Terminal Point) In Sect. 2.5.2 we have discussed the reasons for a Morse line to be tangent at a cusp of an apparent contour. Essentially for the same reason and since a terminal point p gives raise to a cusp in the completion procedure described in Theorem 4.3.1,¹³ in the present context of visible contour graphs we require a further condition, which is always possible with an appropriate choice of the map m : the arc at p is forced to stay locally at one side of the Morse line. This can be achieved as a consequence of the fact that the curvature of a visible arc blows up with a sign near a terminal point (condition (K6) in Definition 4.1.8).

Example 4.4.3 For the visible part of the graph in Fig. 3.14, the Morse description reads as follows: a global maximum, a local maximum oriented from right to left, a T-junction, a terminal point, another local maximum oriented from left to right, a terminal point, another terminal point, and a global minimum.

¹³We need to avoid that the Morse line passing through p has locally both arcs of the completed contour on the same side, an inconvenient fact since it makes p to behave like a local maximum/minimum which is undesirable, since p (being a node of K) already forces the Morse line through p to be critical.

4.4.1 Localization

For the proof of the completion theorem it is convenient to give a localized version of Definition 4.2.12.

Let K be a contour graph, let \mathfrak{m} be a generic family of horizontal Morse lines for K , and for $\lambda \in [0, 1]$ let U_λ^+ be the open half-plane above $\mathfrak{m}_\lambda(\mathbb{R})$. Let λ_i and λ_{i+1} be two consecutive critical values of \mathfrak{m} , and let

$$\mu \in (\lambda_{i+1}, \lambda_i).$$

Set

$$K_\mu := K \cap U_\mu^+,$$

and let $\text{nodes}(K_\mu)$ be the set of nodes of K_μ . Finally, let G_μ be a contour graph.

Definition 4.4.4 (Localized Completion in a Half-Plane) We say that (G_μ, f_μ, d_μ) is a completion of K_μ in U_μ^+ if

- there exists a function $f_\mu : U_\mu^+ \setminus G_\mu \rightarrow 2\mathbb{N}$ such that condition (G3) of Definition 4.2.2 holds, with U_μ^+ in place of \mathbb{R}^2 and G_μ in place of G ;
- if $p \in \text{nodes}(G_\mu)$, then p is either a crossing, or a cusp, or a terminal point. Moreover all terminal points of G_μ lie on $\mathfrak{m}_\mu([0, 1])$;
- there exists $\epsilon > 0$ small enough such that the arcs of G_μ intersecting the strip $U_{\mu-\epsilon}^+ \setminus \overline{U}_\mu^+$ are transverse to $\mathfrak{m}_\lambda([0, 1])$ for all $\lambda \in [\mu - \epsilon, \mu]$;
- there exists a function $d_\mu : G_\mu \setminus \text{nodes}(G_\mu) \rightarrow \mathbb{N}$ satisfying (L1)–(L3) of Definition 4.2.5 with G_μ in place of G and f_μ in place of f ;
- $K_\mu \cap U_\mu^+$ coincides with $\{d_\mu = 0\}$.

The pair (f_μ, d_μ) is called a labelling of G_μ in U_μ^+ .

Definition 4.4.5 (Dangling Arcs) The arcs of $G_\mu \setminus K_\mu$ traversing the strip $U_{\mu-\epsilon}^+ \setminus \overline{U}_\mu^+$ will be called the dangling arcs of G_μ .

4.5 Proof of the Completion Theorem

We are now in a position to start the proof of Theorem 4.3.1. We have already seen that if K is completable then K is a visible contour graph. Therefore, what remains to prove is the converse statement. Take a visible contour graph K . Let \mathfrak{m} be a generic family of horizontal Morse lines for K and let $\lambda_1, \dots, \lambda_n$ be the critical values of \mathfrak{m} with $1 > \lambda_1 > \lambda_2 > \dots > \lambda_n > 0$.

Given $\lambda \in [0, 1]$, denote as before by U_λ^+ the open half-plane above the Morse line $m_\lambda(\mathbb{R})$. Take noncritical values μ_1, \dots, μ_{n+1} with

$$1 > \mu_1 > \lambda_1 > \mu_2 > \dots > \lambda_n > \mu_{n+1} > 0.$$

Given $i \in \{1, \dots, n\}$, let us suppose that the following properties are verified:

- (P1_i) a localized completion (G_i, f_i, d_i) of K is constructed in $U_{\mu_i}^+$;
- (P2_i) the labelling (f_i, d_i) satisfies the further condition

$$f_i = 0 \text{ only in } U_{\mu_i}^+ \cap \text{ext}(K). \quad (4.2)$$

Note that (P1₁) and (P2₁) are satisfied.

Our aim is to show that we can construct a completion $(G_{i+1}, f_{i+1}, d_{i+1})$ of K in $U_{\mu_{i+1}}^+$, in such a way that properties (P1_{i+1}) and (P2_{i+1}) hold, thus going beyond the critical value λ_i . In order to do this, we need to take into account all possible Morse events at level λ_i , and provide a certain number of rules for constructing the completion and the labelling around each event in such a way to keep valid properties (P1_{i+1}) and (P2_{i+1}).

Once we show this step, the proof of the theorem concludes after $n + 1$ steps, by setting

$$(G, f, d) := (G_{n+1}, f_{n+1}, d_{n+1}).$$

Indeed, since λ_n necessarily corresponds to the global minimum, we have $U_{\mu_{n+1}}^+ \supseteq K$, and our construction will give a completion

$$G_{\mu_{n+1}} \subseteq \overline{\Omega \setminus \text{ext}(K)} \subseteq U_{\mu_{n+1}}^+.$$

Notation In the figures, the dotted lines denote the Morse lines $m_{\mu_i}([0, 1])$ and $m_{\mu_{i+1}}([0, 1])$. The bold arcs are the arcs of K , while the thin arcs are added ones in the completion procedure. The inductively given values of f at step i in $U_{\mu_i}^+$ are denoted by the index i , for instance f_i^1, f_i^2 and so on. We adopt a similar notation for the inductively given values of d on the arcs, using d_i^1, d_i^2 , etc. For simplicity, in the figures the new values f_{i+1}, f_{i+1}^j , and d_{i+1}, d_{i+1}^j on the various regions and arcs will be simply denoted by f, f^j , and d, d^j respectively.

Since the Morse lines travel from top to bottom as λ decreases, it is natural to begin the list of Morse events at level λ_i with the local maxima, the first one being obviously the global maximum.

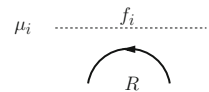
4.5.1 Analysis at the Global Maximum and at Local Maxima

There are two possible orientations of the involved arc of K , depicted in Figs. 4.1 and 4.2a, where R denotes the region bounded by the arc.

- Case 1: Fig. 4.1. From (K3) of Definition 4.1.8 it follows that R cannot be contained in the exterior region $\text{ext}(K)$. Therefore, we define $f := f_i + 2$ in R .
- Case 2: Fig. 4.2a, which describes a local maximum that necessarily is not the global maximum. In view of the orientation of the involved arc of K and the inductive hypothesis (P2_i) it follows that $f_i \geq 2$. We need to distinguish two further subcases, depending on whether region R is contained in the exterior region of K or not.
 - Case 2a: $R \subseteq \text{ext}(K)$. In order to fulfill requirement (P2_{i+1}), we need to define $f := 0$ in R . This is compatible with the inductively given value f_i only if $f_i = 2$, a situation in which we do not add any local completion and we simply proceed to the next Morse event. On the other hand, if $f_i > 2$ we add the proper number $f_i/2 - 1$ of arcs (depicted with thin arcs) with the correct orientation as in Fig. 4.2b: the figure displays the construction in the particular case $f_i = 6$. In this way we reduce f of two units across each new arc so that we can match the value $f = 0$ in R . Then we assign any value of $d > 0$ to each new arc (so that they become invisible arcs), with the only requirement that (L2) of Definition 4.2.5 is satisfied: the simplest choice is of course $d = 1$. Note that we have created $f_i - 2$ new terminal points just below the Morse line $m_{\mu_{i+1}}(\mathbb{R})$. Such terminal points must be continued up to the subsequent Morse event.
 - Case 2b: $R \cap \text{ext}(K) = \emptyset$. If $f_i > 2$ we define $f := f_i - 2$ in R , so that (P2_{i+1}) is satisfied, and there is no need to add any local completion. On the other hand, if $f_i = 2$ we add an arc with the reversed orientation as in Fig. 4.2c, so that we can increase the value of f to $f = 4$ above the visible arc, and consequently define $f := 2$ in R . The value of $d > 0$ on the new (invisible) arc is adjusted to fulfill condition (L2): again the simplest choice is to assign $d = 1$. In this case we have created 2 new terminal points just below the Morse line $m_{\mu_{i+1}}(\mathbb{R})$, to be continued up to the subsequent Morse event.

To conclude the analysis at local maxima we need to specify how to deal with dangling invisible arcs that might be present on the Morse line $m_\lambda(\mathbb{R})$, $\mu_i > \lambda > \mu_{i+1}$, in the open segment S_λ between the consecutive transversal intersections of $m_\lambda(\mathbb{R})$ with K , where the new local maximum Morse event occurs. As shown in Fig. 4.3 this segment gets split into two parts S_λ^- and S_λ^+ , $\lambda < \lambda_i$ separated by the new region R below the local maximum. For each dangling arc in S_λ we are free to choose one of the two new segments S_λ^- and S_λ^+ , provided we preserve the relative

Fig. 4.1 Local or global maximum. Image taken from [1]



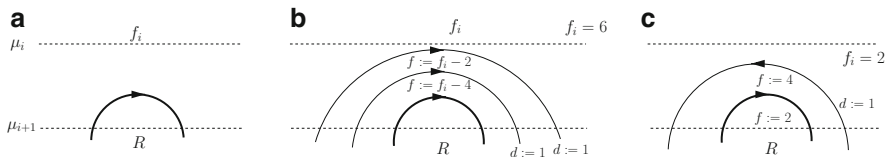


Fig. 4.2 (a): local but not global maximum. (b): the case $R \subseteq \text{ext}(K)$, and the local addition of new invisible arcs when $f_i = 6$. (c): the case $R \cap \text{ext}(K) = \emptyset$ and $f_i = 2$. Image taken from [1]

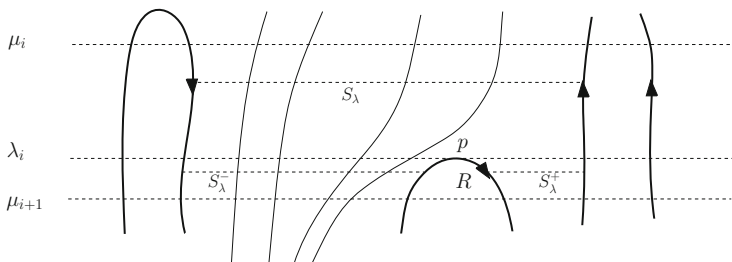


Fig. 4.3 At a local maximum p we conventionally choose to continue all dangling invisible arcs on the left. Image taken from [1]

ordering. We shall conventionally make here the choice of continuing all dangling arcs by traversing the left segment S_λ^- , as shown in Fig. 4.3.

Being concluded the casistics of the local maxima, we continue the list of Morse events at level λ_i with the analysis at terminal points and T-junctions in the order.

4.5.2 Analysis at Terminal Points

Let us observe preliminarily that from the assumptions on K and requirement (P2_i) we have $f_i > 0$ in a neighbourhood of a terminal point p .

We distinguish two possibilities: ending terminal points (Fig. 4.4 a, b, e and f) and starting terminal points (Fig. 4.4 c, d, g and h).

- Ending terminal point. Let f_i^l (respectively f_i^r) be the inductively given value of f on the adjacent left (respectively right) region to the arc of K , where left and right are taken with respect to the orientation of the arc. There are two cases to be considered, depicted in Figs. 4.4a and b.
- Case 1: Fig. 4.4a. We locally complete the arc as in Fig. 4.4e, with a new arc starting at p and below the Morse line, forming a simple cusp with the correct orientation. We also set $d := 1$ on this new arc and set the values of f in the inside of the cusp and in the complement as in the figure. Notice that the labelling is locally consistent, in the sense that it locally meets the requirements of Definition 4.2.5.

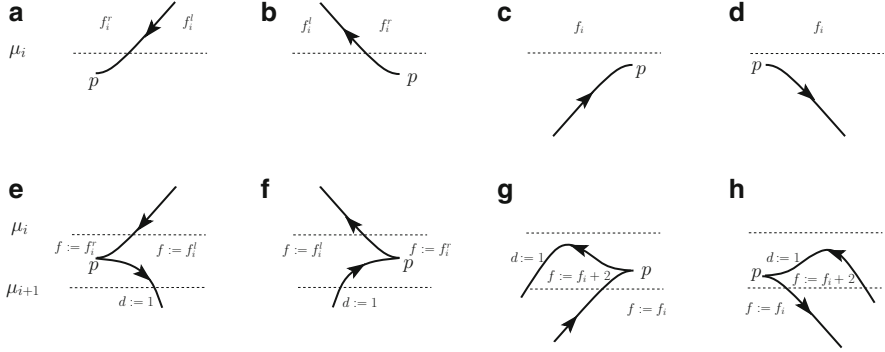


Fig. 4.4 (a), (e): ending terminal point p : case 1. (b), (f): ending terminal point p : case 2. (c), (g): starting terminal point p : case 1. (d), (h): starting terminal point p : case 2. Image taken from [1]

- Case 2: Fig. 4.4b: it is similar to case 1, and the local construction is displayed in Fig. 4.4f.
- Starting terminal point. There are two cases to be considered, depicted in Figs. 4.4c and d.
- Case 1: Fig. 4.4c. By the inductive assumption (P2_i) it follows that $f_i \geq 2$. We proceed as in Fig. 4.4g: namely, we add an invisible arc at p with $d := 1$, and forming a simple cusp at p , with $f = f_i + 2$ inside the cusp. Observe that the labelling is locally consistent.
- Case 2: it is similar to case 1, and it is shown in Figs. 4.4d and h.

For a starting terminal point the segment of the Morse line where the Morse event occurs gets split into two. If there are dangling arcs, we can proceed similarly to the case of the local maxima, see Fig. 4.3.

4.5.3 Analysis at T-Junctions

We distinguish two possibilities: ending T-junctions (Figs. 4.5a and c) and starting T-junctions (Figs. 4.6a and 4.7a).

- Ending T-junction. We are in the situation of Fig. 4.5a or c. Suppose that there are $k_i \geq 0$ dangling invisible arcs between the two points z and w , with the corresponding labelling values $\{f_i^j\}_{j=1,\dots,k_i+1}$ (for the regions) and $\{d_i^j\}_{j=1,\dots,k_i}$ (for the arcs), with $d_i^j \geq 1$ for any $j \in \{1, \dots, k_i\}$. If $k_i = 0$ (i.e., no dangling arcs are present), we have $f_i^1 := f_i$ and $d_i^1 := d_i$. We now distinguish two cases, depending on the local orientation of K .

Case 1: Orientation as in Fig. 4.5a. Notice that the orientation and the inductive assumption (P2_i) imply that $R \cap \text{ext}(K) = \emptyset$.

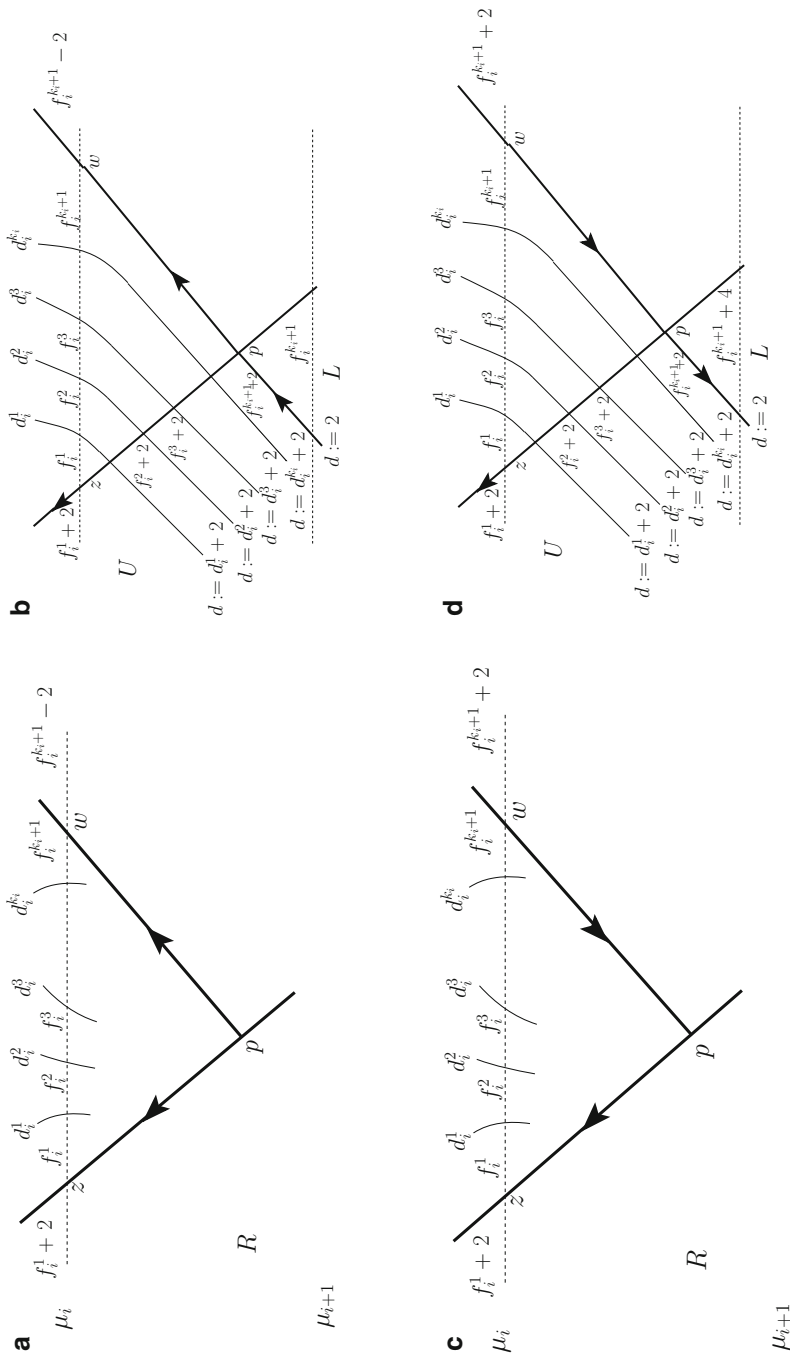


Fig. 4.5 (a): ending T-junction: case in which there are k_i dangling invisible arcs between the two points z and w . (b): all dangling invisible arcs can be locally continued on the *left* with a consistent labelling. (c): ending T-junction. (d): case 2. Again, all invisible arcs can be locally continued on the *left* with a consistent labelling. Image taken from [1]

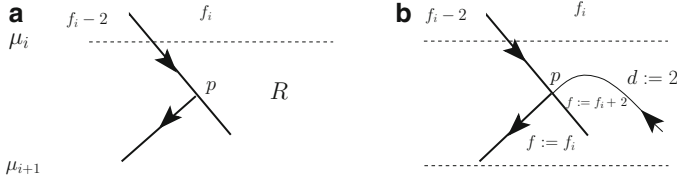


Fig. 4.6 (a): starting T-junction. (b): the local continuation of the transversal arc. Image taken from [1]

Firstly we locally continue the transversal arc joining p and w , downwards on the left with the correct orientation as in Fig. 4.5b, setting $d := 2$ on the local continuation of the arc and $f := f_i^{k_i+1}$ in the lower region L (if no dangling arcs are present $f := f_i^1$ in L).

Now, let us denote by U the region lying above the just constructed new invisible arc. We claim that it is possible to locally continue all dangling arcs inside U as in Fig. 4.5b, downwards on the left, and to assign a consistent local labelling. Note that each arc creates a crossing between z and p in the local completion. We define the local labelling in Fig. 4.5b as follows. Let $l \in \{1, \dots, k_i\}$; on the local completion of the l -th dangling arc we set $d := d_i^l + 2$. Next, the (consistent) values of f are indicated in Fig. 4.5b. Observe that this construction works *independently of the orientation of the dangling arcs*: this is the reason why we do not indicate the orientation of the dangling arcs in Fig. 4.5b. One then directly checks that all local consistency properties required for the functions f and d are satisfied.

Case 2: Orientation as in Fig. 4.5c. Again, due to the local orientation of K , we have $R \cap \text{ext}(K) = \emptyset$. We firstly locally continue the transversal arc joining p and w downwards on the left as in Fig. 4.5d, setting $d := 2$ on the local continuation of the arc and $f := f_i^1 + 4$ in region L . The local completion of the dangling arcs and the assignment of the labelling is then exactly the same as in Fig. 4.5b, see Fig. 4.5d.

– Starting T-junction. We are in the situation of Fig. 4.6a or Fig. 4.7a. Note that in both figures $R \cap \text{ext}(K) = \emptyset$. Again, we need to distinguish two cases, depending on the local orientation of K .

Case 1: Orientation as in Fig. 4.6a. By the inductive hypothesis (P2_i) it follows that $f_i \geq 2$. We locally continue the transversal arc at p upwards on the right with the correct orientation as in Fig. 4.6b, setting $d := 2$ on the local continuation of the arc and defining the values of f as in the figure.

Case 2: Orientation as in Fig. 4.7a. As a consequence of the constraint (K2) we have $U \cap \text{ext}(K) = \emptyset$ and $R \cap \text{ext}(K) = \emptyset$, so that by the inductive hypothesis (P2_i) it follows $f_i \geq 4$. Firstly, we locally continue the transversal arc at p on the right with the correct orientation as in Fig. 4.7b, and we set $d := 2$ on the local continuation. We now need to distinguish two subcases.

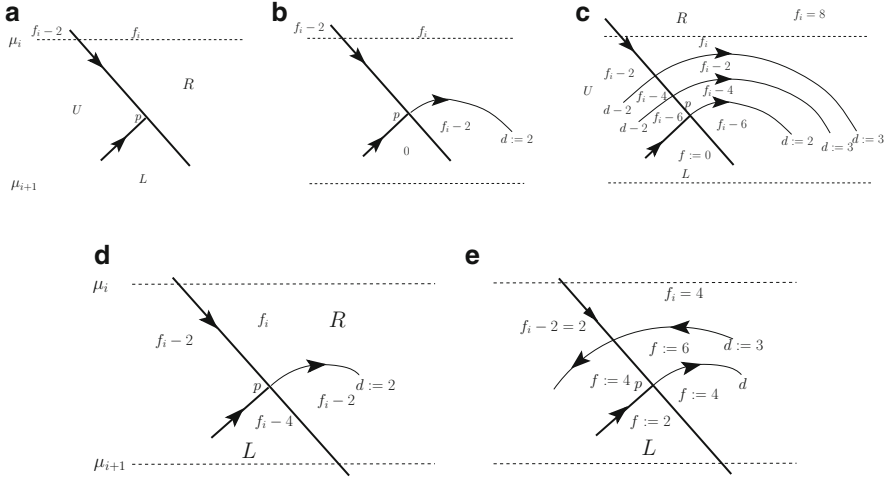


Fig. 4.7 (a): starting T-junction. (b): local continuation, case in which $L \subseteq \text{ext}(K)$ and $f_i = 4$. (c): Case in which $L \subseteq \text{ext}(K)$ and $f_i > 4$; we take as an example $f_i = 8$. (d): case in which $L \cap \text{ext}(K) = \emptyset$ and $f_i > 4$. (e): case in which $L \cap \text{ext}(K) = \emptyset$ and $f_i = 4$. Image taken from [1]

Case 2a: $L \subseteq \text{ext}(K)$. We necessarily have to define $f := 0$ in L . If $f_i = 4$, then it is enough to construct the labelling as in Fig. 4.7b. If $f_i > 4$ we need to introduce the proper number $f_i/2 - 2$ of oriented arcs as in Fig. 4.7c, in order to match the value $f := 0$ in the lower region L . On the invisible arcs (but to the lowest passing through p) we define $d := 3$. One checks that the local labelling is consistent at the new crossings on the arc separating region U from region R .

Case 2b: $L \cap \text{ext}(K) = \emptyset$. If $f_i > 4$, then we locally complete the graph as in Fig. 4.7d. If $f_i = 4$ we need to introduce an oriented arc (with reversed orientation) as in Fig. 4.7e in order to obtain a value of $f > 0$ in the lower region L . One checks that the resulting local labelling is consistent. This concludes the analysis at T-junctions.

Now, we analyse the merging part of the proof, and consider those Morse events for which the invisible dangling arcs obtained in the previous steps must be suitably joined each other. This is the case, for instance, when the Morse line intersects the global minimum p of K , and there are invisible dangling arcs inside the region U which is locally above p and bounded by K . In this situation the invisible arcs cannot go out of U since the region below p (denoted by R in the corresponding figures) is contained in $\text{ext}(K)$ (compare with Remark 4.3.6). This is the most delicate part of the proof. As we shall see, besides new crossings, in some cases it will be necessary also to introduce new cusps.

4.5.4 Analysis at Local Minima and at the Global Minimum

There are two cases to be considered: Figs. 4.8a and 4.10a (this latter case happens, for instance, for the global minimum, and is the most involved).

Case 1: Fig. 4.8a. In view of the orientation of the involved arc, we necessarily have $R \cap \text{ext}(K) = \emptyset$. If there are no dangling invisible arcs between the two points z and w , we simply define $f := f_i + 2$ in R (and of course no local continuation is required).

If there are $k_i \geq 1$ dangling invisible arcs between z and w , denote by $\{f_i^j\}_{j=1, \dots, k_i+1}$ the corresponding labelling values for the regions, and by $\{d_i^j\}_{j=1, \dots, k_i}$ the labelling for the arcs, with $d_i^j \geq 1$ for any $j \in \{1, \dots, k_i\}$, see Fig. 4.8b. Note that $f_i^1 \geq 2$, since no invisible arc can be adjacent to $\text{ext}(K)$. We then continue downwards the arcs inside region R , increasing of two units the corresponding value of d , and this is done independently of the orientation of each arc. Observe that in this process we create k_i new crossings, as indicated in Fig. 4.8c. The values of the function f are correspondingly increased of two units passing from region U to region R . One checks that the local labelling around each new crossing is consistent, *independently of the orientation of the dangling arcs*.

Case 2: Fig. 4.10a. If there are no dangling invisible arcs between z and w , we simply keep the value of f in R as the one given by the inductive step.

If there are $k_i \geq 1$ dangling invisible arcs between z and w , as, for instance, in Fig. 4.9a, we need to distinguish two further subcases.

Case 2a: $R \cap \text{ext}(K) = \emptyset$. In this case $f_i^1 \geq 4$. If there is some dangling arc γ having labelling $d = d(\gamma) = 1$ then, by inserting one cusp (similarly to what done, for instance, in Fig. 4.9c) to γ , we increase the value of $d(\gamma)$ of one unit. After this operation, we can therefore suppose that the value of d on each dangling invisible arc in region U is larger than or equal to 2, namely

$$d_i^j \geq 2 \text{ inside } U \text{ for any } j \in \{1, \dots, k_i\}. \quad (4.3)$$

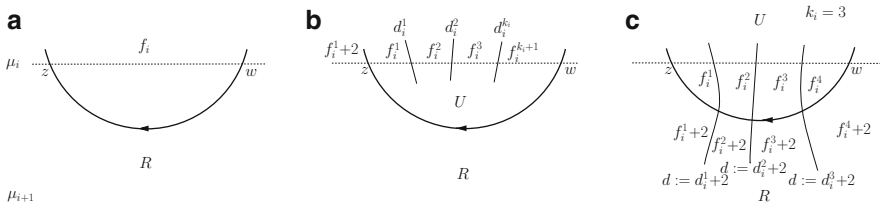


Fig. 4.8 (a): local but not global minimum: $R \cap \text{ext}(K) = \emptyset$. (b): independently of the orientation of the dangling invisible arcs, we continue these arcs inside region R . In (c) we show the local completion when $k_i = 3$. Image taken from [1]

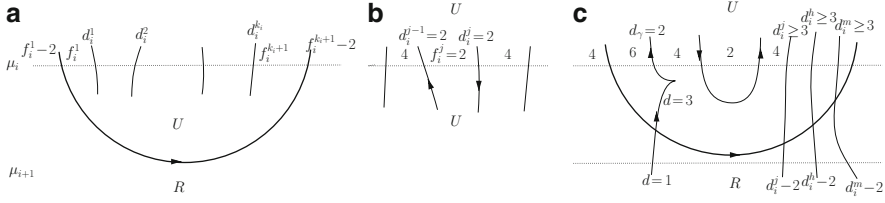


Fig. 4.9 (a): the case in which $R \cap \text{ext}(K) = \emptyset$. (b): if there is a region inside U where $f_i^j = 2$, then necessarily we have the depicted values of f and d , and the depicted orientations. The two arcs with $d_i = 2$ are then smoothly joined remaining inside U , as in (c). (c): example of the procedure adopted in case 2a of the proof. Image taken from [1]

Moreover, observe that the dangling arcs having $d_i^j = 2$ cannot be continued downwards since, due to the local orientations, the new value of d should decrease of two units (remember Fig. 3.11), thus vanishing on the local continuation of the arc.¹⁴ This is not allowed, as the continued arcs cannot be visible.

If there is $j \in \{1, \dots, k_i + 1\}$ such that $f_i^j = 2$, using also (4.3), we are necessarily in the local situation displayed in Fig. 4.9b. Then the two arcs locally bounding the region where $f_i^j = 2$ can be smoothly joined one to each other, remaining inside U , similarly to what displayed in Fig. 4.9c. From now on these joined arcs will be no longer considered as dangling arcs, and therefore we shall not consider them anymore.

Let γ be a dangling invisible arc with $d =: d(\gamma) = 2$. In view of the previous construction, necessarily the local value of f on the regions on the two sides of γ is larger than or equal to 4. Then by adding one cusp (as shown in Fig. 4.9c) we can increase the value of $d(\gamma)$ to $d(\gamma) = 3$ on γ . We can therefore assume that

$$\text{all dangling arcs have } d_i^j \geq 3. \quad (4.4)$$

We are now in the position to continue all dangling arcs under consideration from region U to region R , similarly to what depicted¹⁵ in Fig. 4.9c, where now the values of f_i^j and d_i^j are decreased of two units when passing from U to R . The resulting values of f are strictly positive, thus meeting the inductive assumptions. Also, the resulting values of d are strictly positive, thanks to (4.4).

Case 2b: $R \subseteq \text{ext}(K)$, see Fig. 4.10b. Necessarily $f_i = 0$ in R , therefore no dangling invisible arc is allowed to leave U . Observe that

¹⁴For example, if the dangling arc is oriented downwards (respectively upwards), the local labelling at the corresponding crossing is the same as in the penultimate (respectively last) picture of Fig. 3.11.

¹⁵If the dangling arc is oriented downwards, then the new formed cusp points on the left.

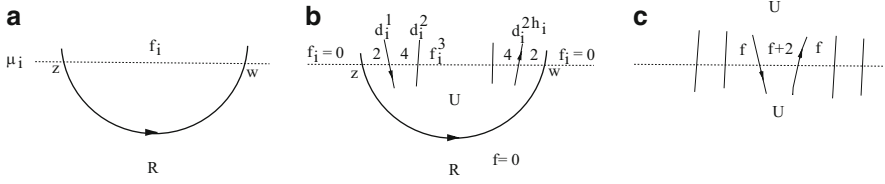


Fig. 4.10 (a): local or global minimum. R can be contained in $\text{ext}(K)$, as in Figure (b) (and as in the case of the global minimum). In that case the total number of dangling invisible arcs is even ($= 2h_i$, $h_i \geq 1$), moreover, the values of f and the orientations are necessarily as in the figure. (c): necessarily there are two contiguous arcs, among the dangling arcs of Figure (b), oriented in this way. Image taken from [1]

- $f_i = 2$ in the region between the arc of K and the first dangling invisible arc starting from the left, since $f_i = 0$ only in the exterior region by the inductive assumption ($P2_i$). Similarly, $f_i = 2$ in the region between the arc of K and the first dangling invisible arc starting from the right. See Fig. 4.10b.
- The total number of dangling arcs is even. We denote this number by $2h_i$.
- Since dangling invisible arcs cannot be adjacent to $\text{ext}(K)$, it follows that $f_i = 4$ in the region between the first and the second dangling invisible arc starting from the left (if $h_i > 1$). For the same reason, $f_i = 4$ in the region between the first and the second dangling invisible arc starting from the right, see Fig. 4.10b. In particular, the orientation of the first dangling invisible arc starting from the left is downwards, whereas the orientation of the first dangling invisible arc starting from the right is upwards, as shown in Fig. 4.10b.
- There exists *at least one pair of contiguous dangling invisible arcs with the local orientations as in Fig. 4.10c*. They are found by looking for the maximal value of f_i^j , $j = 1, \dots, 2h_i$.

Fix a pair of contiguous dangling invisible arcs as in Fig. 4.10c, and let $j, j+1$ be the integers indexing the arcs of the pair (see Fig. 4.11a). If $d_i^j = d_i^{j+1}$, we smoothly join the two arcs remaining inside region U as in Fig. 4.11b.

Suppose now that $d_i^j \neq d_i^{j+1}$ (Fig. 4.11c), and set $d_{\max} := \max\{d_i^j, d_i^{j+1}\}$. Adding the appropriate number of cusps on the arc which has the lower value of d , we can make the two invisible arcs to have the same value of $d = d_{\max}$: see Fig. 4.11c, where we have considered the case $d_i^{j+1} = d_i^j + 2$, and therefore we have added two cusps with the correct orientation on the j -th invisible arc. Then we can argue as in Fig. 4.11b, by smoothly joining the two arcs with the same value of $d = d_{\max}$ remaining inside U (Fig. 4.11c). Now, we consider all the remaining arcs: again we observe that there exists at least a pair of contiguous arcs with the orientations and the local values of f as in Fig. 4.10c. We then argue as before adding the proper number of cusps, and we smoothly join together the two contiguous arcs remaining inside U , without intersecting the other dangling arcs. After a finite number of these operations, all dangling arcs are smoothly joined pairwise. One checks that the constructed labelling is consistent, and this concludes the proof of the theorem. \square

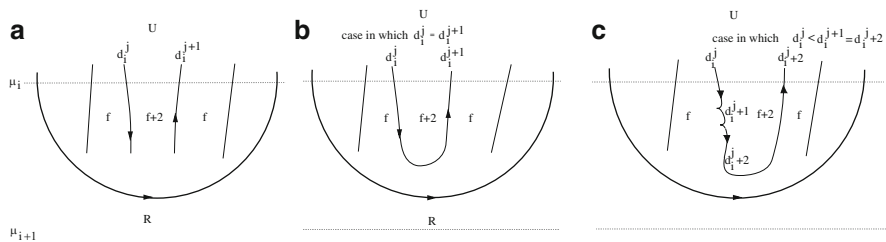


Fig. 4.11 (b): if two contiguous arcs of (a) have the same value of d we smoothly join them remaining inside L . Here again $R \subseteq \text{ext}(K)$. (c): case in which $d_i^j \neq d_i^{j+1}$, for instance $d_i^j < d_i^{j+1}$; we show the example when $d_i^j + 2 = d_i^{j+1}$, so that we add two cusps to the left arc in order to increase the value of d of two units. Image taken from [1]

We shall see in Example 4.6.4 that the constructive proof of Theorem 4.3.1 provides, for K as in Fig. 4.15a, the completion G of Fig. 4.15b. This is the apparent contour of a solid shape topologically equivalent to a sphere (and not to a torus), and this is a consequence of requirement (4.2). This requirement can be relaxed: more precisely, the region $\{f = 0\}$ can be larger than the external region $\text{ext}(K)$, provided we strengthen condition (K3) in Definition 4.1.8. Under this assumption, it turns out that the resulting completion is given by the apparent contour of a torus. We observe that deciding whether f vanishes also in some nonexterior region is a piece of information that must be known a priori.

Corollary 4.5.1 (Completion with Assigned $\{f = 0\}$) *Let K be a contour graph satisfying conditions (K1), (K2) and (K4)–(K6) of Definition 4.1.8. Suppose in addition that $\text{background}(K)$ is a union of connected components of $\mathbb{R}^2 \setminus K$ and satisfies the following conditions:*

- (K3') $\text{background}(K)$ contains $\text{ext}(K)$,
- (K3'') any arc a of K which is adjacent to $\text{background}(K)$ is oriented in such a way that $\text{background}(K)$ lies locally only on the right of a .

Then there exists a complete labelled contour graph (G, f, d) such that K is the visible part of G , and furthermore

$$\{f = 0\} = \text{background}(K).$$

Proof It is the same¹⁶ as the proof of Theorem 4.3.1, by replacing $\text{ext}(K)$ with $\text{background}(K)$. \square

Remark 4.5.2 The local shape of K (respectively G), especially near a terminal point (respectively a cusp), is inessential for the validity of Theorem 4.3.1. Indeed it is always possible to recover the required regularity of G (properties (G1), (G4)

¹⁶Notice that in the proof of Theorem 4.3.1 the connectedness of $\text{ext}(K)$ is not used.

and (G5) of Definition 4.2.2) *a posteriori* by means of local deformations. This is also possible with no modification of the visible part K of G provided K already has the required regularity properties (K1), (K5) and (K6) listed in Definition 4.1.8.

The completion of a visible contour graph is far from being unique. Therefore, any invariant (in whatever sense) of a completion cannot be an invariant of the original visible contour graph. We refer to Sect. 9.2.3 for a further discussion.

Remark 4.5.3 (Dependence on the Morse Description) The algorithmic proof of Theorem 4.3.1 depends on the Morse description of the visible contour: different Morse descriptions of the same visible contour will in general lead to inequivalent 3D shape reconstructions.

4.6 Examples

The aim of Theorem 4.3.1, as opposed, for instance, to that of [2], is not to provide a *good* solution, i.e., a completion that corresponds to the 3D object that our brain infers from the contour sketch, but only to rigorously show that the set of possible 3D (smooth) objects that give rise to the given visible contour is nonempty. In some cases, however, the result of Theorem 4.3.1 coincides with the natural interpretation of the original visible contour. In other cases, instead, the completion given by Theorem 4.3.1 may not give, as a result, what one commonly expects. One can then try to apply Corollary 4.5.1. Let us illustrate these assertions with some examples.

Example 4.6.1 (Bean) Let us consider the visible contour of Fig. 4.12a, which satisfies all assumptions (K1)–(K6) of Definition 4.1.8. Applying the construction described in Theorem 4.3.1, and employing the Morse description (see Definition 4.4.1), we proceed as follows:

- a (global) maximum, Morse event (m1). No addition of invisible arcs is required (Fig. 4.1);
- a (local) maximum oriented from right to left, Morse event (m1). Again, no addition of invisible arcs is required (Fig. 4.1);

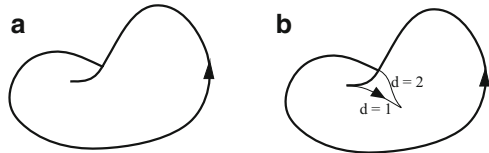


Fig. 4.12 Example 4.6.1.
Image taken from [1]

- a T-junction, Morse event (m5). We are in the situation of Fig. 4.5c¹⁷; since we have no dangling arcs we simply locally complete the transversal arc by creating a new dangling arc with $d = 2$, oriented upwards;
- a terminal point, Morse event (m3). We are in the situation of Fig. 4.4a and we will proceed as in Fig. 4.4e by adding a new dangling arc with $d = 1$ oriented downwards;
- a (global) minimum, Morse event (m2). We are in the situation of Fig. 4.10b with two ($h_i = 1$) dangling arcs (with opposite orientation) with $d = 1$ and $d = 2$; we will connect them after the addition of a cusp.

The result is depicted in Fig. 4.12b, which represents a *bean*. This is the *natural* interpretation of the original visible contour.

The next example is peculiar: the visible contour K does not need any completion since it already satisfies all properties required to be a complete contour graph. The choice $G = K$ gives therefore the simplest possible completion of K and the resulting 3D object E is a solid torus. However the procedure illustrated in the proof of Theorem 4.3.1 gives a different (but perfectly legitimate) result.

Example 4.6.2 (Torus, I) Let us consider the visible contour K of Fig. 4.13a consisting of a pair of concentric circles with opposite orientation, the larger one necessarily oriented counterclockwise. This contour satisfies all requirements (K1)–(K6) of Definition 4.1.8. Notice that it does not contain any node, so that the trivial choice $G = K$ is already a solution to the completion problem. Such a G is the apparent contour of a torus viewed from above. The constructive procedure in the proof of Theorem 4.3.1 gives an unexpected result: indeed, due to requirement (P2_i) we cannot assign the (otherwise legitimate) value $f = 0$ inside the inner circle. The obtained completion is the apparent contour G of Fig. 4.13b with a value of $d = 1$ on the intermediate (invisible) circle. Compare case 2b in Sect. 4.5.1. The 3D reconstruction of such a G has the shape of a sphere with a large excavation viewed from above; it could resemble to an *amphora* viewed from above. These two very different 3D reconstructions (which also have different topological types) are completely indistinguishable by means of their *visible* apparent contours. On the other hand, imposing $f = 0$ also in the inside of the inner circle gives, using

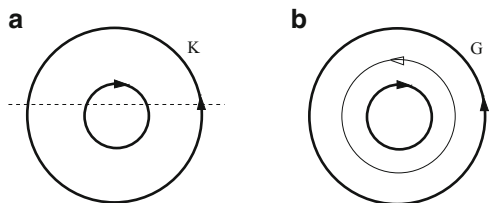


Fig. 4.13 (a): Example 4.6.2. In (b): we depict the resulting completion G , given K , provided by Theorem 4.3.1. Image taken from [1]

¹⁷With a left–right reflection. Notice that reflecting the values of f still results in an orientation of the two arcs from right to left.

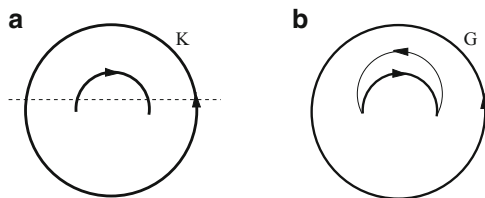


Fig. 4.14 (a): Example 4.6.3. Note that this example is identical to Example 4.6.2 (Fig. 4.13) if we restrict to the first two Morse events (above the *dotted line*). The completion G in (b): is “optimal”. Image taken from [1]

Corollary 4.5.1, the apparent contour of the torus ($G = K$), which is the expected result.

Example 4.6.3 (Sphere with a Cave) Consider the visible contour K of Fig. 4.14a. It consists of an external circle oriented counterclockwise containing an arc with two terminal points oriented from left to right. Note carefully that the visible contours of the two examples in Figs. 4.13a and 4.14a coincide whenever restricted to the half-plane above the dotted lines. Therefore their Morse description is identical until the Morse line reaches the dotted lines, and consequently the initial construction of the completion in Theorem 4.3.1 (based on the Morse description) will necessarily be the same in both cases up to that line.

The resulting G that we obtain from the construction of the proof of Theorem 4.3.1 is shown in Fig. 4.14b. The reconstructed 3D object E could resemble a sphere with a cave, and in this case the reconstruction seems to be the expected one.

In both Examples 4.6.2 and 4.6.3 our procedure requires the addition of an invisible arc between the two local maxima Morse events; this addition is actually redundant in the first example, but it is essential in the second one since otherwise we would get a negative value of f locally in the region immediately below the internal visible arc.

Example 4.6.4 (Torus, II) A more involved example is shown in Fig. 4.15a. In order to distinguish the values of f from the values of d , we depict the values of f inside a small circle. The sequence of Morse events includes two T-junctions and two terminal points. Hence, following the constructions described in the proof of Theorem 4.3.1, we create a total of 6 dangling arcs as shown in the lower part of Fig. 4.15b. In particular note that the first T-junction forces us to add an invisible arc with $d = 3$ in order to have $f > 0$ inside the *hole* of the original contour, in respect to the inductive requirement (P2_i). The result is rather complex and its analysis can be carried out with the aid of the `appcontour` software described in Chap. 10.

In Chap. 9 we describe an actual implementation of the constructive proof of Theorem 4.3.1 in a software code named `visible`. Its output is a Morse description that can be directly used as input to the `appcontour` program.

Figure 4.16a shows the resulting apparent contour as reconstructed by the software; the reader should check that this is equivalent to Fig. 4.15b up to ambient

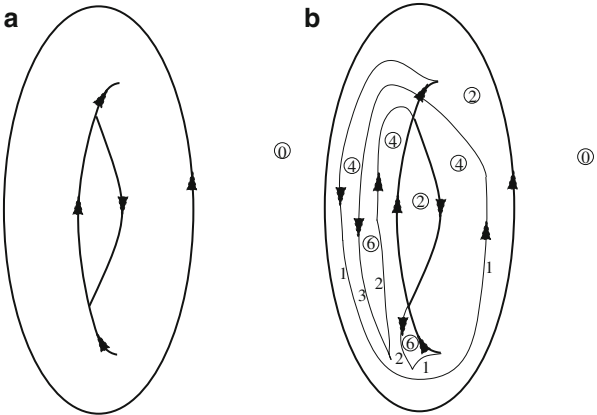


Fig. 4.15 Example 4.6.4. The values of f are depicted inside a small circle. The completion in (b) of (a) is the apparent contour of a surface ambient isotopic to a sphere. Image taken from [1]

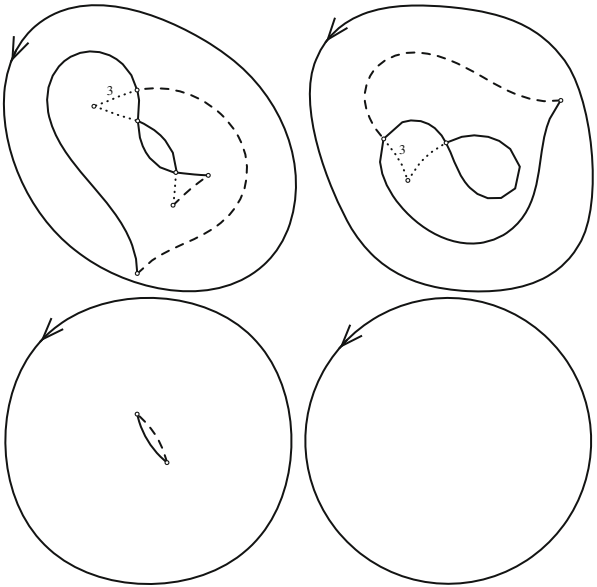


Fig. 4.16 Example 4.6.4. Use of the software `appcontour` to understand the constructed apparent contour. The four pictures show the initial contour (ambient isotopically equivalent to Fig. 4.15b) and the result of successive application of *rules* that correspond to 3D isotopic deformations of the originating 3D shape; $d = 1$ on the *dashed arcs* and $d = 3$ on the *dotted ones*. Image taken from [1]

isotopies. A number of properties of the apparent contour are also computed, in particular the number of connected components ($= 1$) and the Euler characteristic ($= 2$) of the 3D surface.

Using `appcontour` it is also possible to apply *rules* (similar to the *Reidemeister moves* for knots) to the apparent contour, that correspond to ambient isotopic deformations of the 3D surface; one of the applicable rules (denoted *S* by the software) allows to simplify the *swallow's tail* and the result is shown in Fig. 4.16b. Now rule *C* allows to untangle one of the two cusps, and the result is shown in Fig. 4.16c. Finally, rule *L* produces the apparent contour of a sphere (Fig. 4.16d).

We remark that forcing $f = 0$ in the innermost region of Fig. 4.15a and making use of Corollary 4.5.1, produces, as a result of the algorithmic proof of Theorem 4.3.1, the usual torus: see the discussion in Example 9.1.

References

1. Bellettini, G., Beorchia, V., Paolini, M.: Completion of visible contours. *SIAM J. Imag. Sci.* **2**, 777–799 (2009)
2. Karpenko, O.A., Hughes, J.F.: SmoothSketch: 3D free-form shapes from complex sketches, The 33rd. International Conference and Exhibition on Computer Graphics and Imaging Techniques, Boston, Massachusetts, 2006, SIGGRAPH 2006, pp. 589–598. ACM, New York (2006)

Chapter 5

Topological Reconstruction of a Three-Dimensional Scene

Following closely [1],^{1,2} in this chapter we characterize those planar graphs contained in Ω that are apparent contours of a stable smooth 3D scene $E \subset Q = \Omega \times (-1, 1)$. As we shall see, the conditions imposed on a graph for being a complete labelled contour graph are sufficient for our purposes.

The notation concerning graphs is introduced in Chap. 4; in particular, complete labelled contour graphs are introduced in Definition 4.2.6. The meaning of a 3D scene E is explained in Definition 3.1.1, and the concept of stable scene is explained in Definition 3.2.1.

5.1 Statement of the Reconstruction Theorem

The reconstruction result (Theorems 5.1.1 and 5.1.4) *characterizes* those graphs which are apparent contours of a smooth three-dimensional scene. As we shall see, the proof of this assertion is rather long, and splits into an existence and a uniqueness part.

The existence result,³ the proof of which is given in Sect. 5.2, can be stated as follows.

¹With kind permission from Springer Science+Business Media, in this chapter and in Chap. 11 we illustrate some results and report some of the figures from the quoted paper [1].

²See also the panelling construction in [11] and [2–6, 12].

³See again [11, 12], and also [8].

Theorem 5.1.1 (Existence) *Let (G, f, d) be a complete labelled contour graph. Then there exists a stable 3D scene $E \subset Q$ such that*

$$f = f_\Sigma \quad \text{and} \quad d = d_\Sigma,$$

where $\Sigma := \partial E$.

It is worthwhile to observe that part of the proof of Theorem 5.1.1, specifically the construction of the topological manifold \mathcal{T} described in Sect. 5.2.1, is actually implemented in the `appcontour` program. This piece of information is used in the computation of the CW complex, and in the computation of the first fundamental group of E and $\mathbb{R}^3 \setminus E$, as described in Sect. 10.7.

In order to state the uniqueness part of the reconstruction result, we need to introduce a suitable notion of equivalence.

5.1.1 Depth-Equivalent Scenes

Let $E \subset Q$ be a three-dimensional scene. For any $x \in \Omega$, we define $E_x := \{z \in (-1, 1) : (x, z) \in E\} = \pi^{-1}(x) \cap E$ as the one-dimensional slice of E passing through the point $(x, 0)$. We call E_x the *fibre* of E over $(x, 0)$ in the direction of the eye.

Since we are not interested in reconstructing the precise depth of the points on the surfaces, we have to consider equivalent those three-dimensional scenes which differ to each other by a homeomorphism preserving the order of the points on the fibres.

Therefore, we need the following definition, already mentioned at the beginning of Chap. 1.

Definition 5.1.2 (Depth-Equivalence Class of a 3D Scene) Let $E_1, E_2 \subset Q$ be two 3D scenes. We say that E_1 and E_2 are depth-equivalent, and we write $[E_1] = [E_2]$, if there exists a homeomorphism $\Theta : Q \rightarrow Q$ of the form

$$\Theta(x, z) = (x, \Theta_x(z)), \quad (x, z) \in Q,$$

with Θ_x strictly increasing for any $x \in \Omega$, and

$$1_{E_1} = 1_{E_2} \circ \Theta.$$

Namely, if E_1 and E_2 are depth-equivalent, then, for any $x \in \Omega$, there is a continuous strictly monotone map Θ_x taking the fibre $E_{1,x}$ of E_1 over $(x, 0)$ onto the fibre $E_{2,x}$ of E_2 , and varying continuously with respect to x .

Remark 5.1.3 Let E_1 and E_2 be depth-equivalent, let f_1 and f_2 be the corresponding functions as defined in (3.5); that is, for any $x \in \Omega$,

$$f_1(x) = \#\{\pi^{-1}(x) \cap \Sigma_1\}, \quad f_2(x) = \#\{\pi^{-1}(x) \cap \Sigma_2\},$$

where

$$\Sigma_1 = \partial E_1, \quad \Sigma_2 = \partial E_2.$$

Let d_1 and d_2 be the corresponding labellings as in Definition 3.4.1. Then

$$f_1 = f_2 \quad \text{and} \quad d_1 = d_2.$$

We are now in a position to state the uniqueness part of the reconstruction problem, which is proven in Sect. 5.3.

Theorem 5.1.4 (Uniqueness) *Let $E_1, E_2 \subset Q$ be two stable 3D scenes, such that*

$$f_{\Sigma_1} = f_{\Sigma_2} \quad \text{and} \quad d_{\Sigma_1} = d_{\Sigma_2}.$$

Then

$$[E_1] = [E_2].$$

5.2 Proof of Existence

The aim of this section is to prove Theorem 5.1.1.

Denote by

$$R_1, \dots, R_n$$

the regions⁴ of G (see Sect. 4.1).

The proof splits into various steps. In the first step we define a two-dimensional abstract topological manifold \mathcal{T} by properly glueing the boundaries of the regions. The manifold \mathcal{T} is obtained as a quotient of a set D , a sort of disjoint union⁵ of the closure of the regions, each region appearing the proper number of copies. This construction will be implemented in the program `appcontour` in Chap. 10, for instance in Sect. 10.7.1.

⁴Remember that, by definition, the regions are open.

⁵We recall that, if X_1, \dots, X_m are sets, the disjoint union $\coprod_{i=1}^m X_i$ is defined as $\cup_{i=1}^m \{(x, i) : x \in X_i\} = \cup_{i=1}^m (X_i \times \{i\})$, and the disjoint union topology on $\coprod_{i=1}^m X_i$ is defined as follows: $A \subset \coprod_{i=1}^m X_i$ is open if $A \cap (X_i \times \{i\})$ is open for any $i = 1, \dots, m$.

According to our usual notation (Sect. 3.3), for any $i = 1, \dots, n$ we denote by $f(R_i)$ the value of the restriction of f to region R_i .⁶ We also let \mathbb{N}_+ be the set of all positive natural numbers.

5.2.1 Glueing

Define

$$D := \{(x, i, r) \in \Omega \times \mathbb{N} \times \mathbb{N}_+ : i \in \{1, \dots, n\}, x \in \overline{R}_i, r \in \{1, \dots, f(R_i)\}\}.$$

The set D can be interpreted as the disjoint union $\coprod_{i=1}^n D_i$, where

$$D_i := \{(x, r) \in \Omega \times \mathbb{N}_+ : x \in \overline{R}_i, r \in \{1, \dots, f(R_i)\}\}.$$

Notice that the set D_i is, in turn, the disjoint union of $f(R_i)$ copies of \overline{R}_i , the strata⁷ of R_i . We endow D with the disjoint union topology.

Remark 5.2.1 By definition, the set D_i is empty if $f(R_i) = 0$: this happens in particular when R_i is the external region.

Now, we want to suitably “glue” together the boundaries of the various strata of the regions, and this will be done by introducing an equivalence relation on D . The idea is to paste ordered pairs of arcs (preserving the same orientation) of the strata using the information given by the values of the labelling d .

Let $(x_1, i_1, r_1), (x_2, i_2, r_2) \in D$. We say that (x_1, i_1, r_1) and (x_2, i_2, r_2) are equivalent, and we write $(x_1, i_1, r_1) \sim (x_2, i_2, r_2)$, if

$$x_1 = x_2 =: x,$$

and either $i_1 = i_2$ and $r_1 = r_2$, or one of the following four cases hold (the second case splits into cases 2.1–2.4, the third one into cases 3.1–3.5 and the fourth one into cases 4.1–4.3):

1. $x \notin G$.

In this case x belongs to only one region, so that $i_1 = i_2$. Then $r_1 = r_2$.

Therefore, glueing does not occur for points in the regions;

2. $x \in a \in \text{arcs}(G)$.

Let for notational simplicity d ($= d(x) = d(a)$) be the value of the labelling d at x , and denote by R_{\max} and R_{\min} the two regions at opposite sides of the arc

⁶Remember from Definitions 4.2.1 and 4.2.2 that $f(x)$ is independent of the choice of $x \in R_i$.

⁷Recall that a stratum of R_i is a pair (R_i, r) with $r \in \{1, \dots, f(R_i)\}$; see Definition 3.3.1.

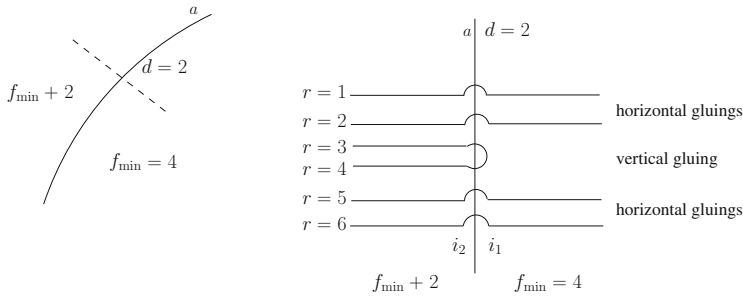


Fig. 5.1 Four horizontal glueings and one vertical glueing along the arc a . Image taken from [1]

a , with corresponding values $f_{\max} := f(R_{\max}) = f(R_{\min}) + 2 =: f_{\min} + 2$. Then

- 2.1. $i_1 \neq i_2$, $r_1 = r_2 \leq d$. We glue along a the first d strata of the regions at opposite sides of a . We call such an identification *horizontal glueing along a* , see Fig. 5.1;
- 2.2. $i_1 = i_2$, $R_{i_2} = R_{\max}$, $d + 1 \leq r_1 < r_2 \leq d + 2$. We glue the $(d + 1)$ -th and the $(d + 2)$ -th strata of R_{\max} along a . We call such an identification *vertical glueing along a* , see Fig. 5.1;
- 2.3. $i_1 \neq i_2$, $R_{i_1} = R_{\min}$, $d + 3 \leq r_2 = r_1 + 2$, provided $f_{\min} > d$. We glue along a (horizontal glueing) the last $f_{\min} - d$ strata of the two regions at opposite sides of a ;
- 2.4. $i_1 \neq i_2$, $R_{i_1} = R_{\max}$, $d + 3 \leq r_1 = r_2 + 2$, provided $f_{\min} > d$. Same identification as in 2.3, with exchanged i_1 and i_2 .

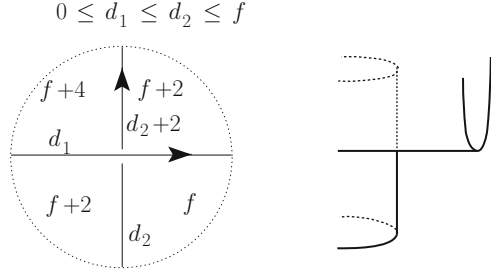
While in item 2 we have been concerned with the identification of arcs, in item 3 we identify quadruples of points, and in item 4 we identify either pairs of points or quadruples of points. We still use the words horizontal and vertical glueings *at the point x* , considered as limit cases of the horizontal and vertical glueings introduced in item 2.

3. $x \in \text{crossings}(G)$.

Concerning the notation around x , we refer to Fig. 5.2. which coincides with the last picture in Fig. 3.11 (the case of the penultimate picture being similar⁸). Note that $d_1 \leq d_2$. There are two strata folding at depths $d_1 + 1$ and $d_2 + 3$ (items 3.2 and 3.4 below concern those strata in the order). The remaining strata do not fold around x , and are divided into three groups and horizontally glued at x (items 3.1, 3.3 and 3.5 below).

⁸For the glueing concerning the penultimate picture of Fig. 3.11, it is sufficient to repeat items 3.1–3.5, with $i_+ -$ replaced by $i_- +$.

Fig. 5.2 *Left:* compatibility conditions between f and d around a crossing. *Right:* Local realization in space ($f = 0, d_1 = d_2 = 0$)



Denote by R_{i++} (respectively R_{i--}) the region where f takes its maximum (respectively minimum) value among the four regions around x , and by R_{i+-} (respectively R_{i-+}) the region in the first (respectively third) quadrant. Then

- 3.1. $r_1 = r_2 \leq d_1$, provided $d_1 > 0$. The first d_1 strata in front of the two folded strata are glued horizontally four by four at x ;
- 3.2. $(i_1, r_1), (i_2, r_2) \in \{(i_{+-}, d_1 + 1), (i_{+-}, d_1 + 2), (i_{++}, d_1 + 1), (i_{++}, d_1 + 2)\}$.

Remembering the glueing described for the arcs, and looking at the first folding, at x we glue all together

- vertically the two strata of region R_{i+-} ,
- vertically the two strata of region R_{i++} ,
- horizontally the first two strata of regions R_{i+-} and R_{i++} ,
- horizontally the second two strata of regions R_{i+-} and R_{i++} .

- 3.3. $(i_1, r_1), (i_2, r_2) \in \{(i_{--}, r), (i_{-+}, r), (i_{+-}, r + 2), (i_{++}, r + 2)\}$ for $r \in \{d_1 + 1, \dots, d_2\}$, provided $d_2 > d_1$. The $d_2 - d_1$ strata intermediate between the two folded strata are glued horizontally four by four at x at the proper depth.

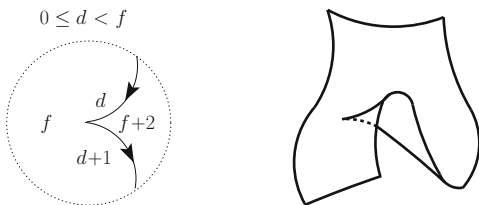
- 3.4. $(i_1, r_1), (i_2, r_2) \in \{(i_{-+}, d_2 + 1), (i_{-+}, d_2 + 2), (i_{++}, d_2 + 3), (i_{++}, d_2 + 4)\}$.

Remembering the glueing described for the arcs, and looking at the last folding, at x we glue all together

- vertically the two strata of region R_{i-+} ,
- vertically the two strata of region R_{i++} ,
- horizontally the first stratum of region R_{i-+} with the third stratum of region R_{i++} ,
- horizontally the second stratum of region R_{i-+} with the fourth stratum of region R_{i++} .

- 3.5. $(i_1, r_1), (i_2, r_2) \in \{(i_{--}, r), (i_{-+}, r + 2), (i_{+-}, r + 2), (i_{++}, r + 4)\}$ for $r \in \{d_2 + 1, \dots, f_{\min}\}$, provided $f_{\min} > d_2$. The remaining $f_{\min} - d_2$ strata, behind the two folded strata, are glued horizontally four by four at x .

Fig. 5.3 *Left*: compatibility conditions between f and d around a cusp. *Right*: local realization in space ($d = 0$)



4. $x \in \text{cusps}(G)$.

Let us denote by $R_{i_{\text{inn}}}$ the region containing the acute part of the cusp, and by $R_{i_{\text{out}}}$ the other region. We also let f_{out} be the value of f in $R_{i_{\text{out}}}$; recall that $f = f_{\text{out}} + 2$ in $R_{i_{\text{inn}}}$. Then

4.1. $r_1 = r_2 \leq d$, provided $d > 0$. The first d strata in front of the cusp are glued horizontally two by two at x (identification of two points).

4.2. $(i_1, r_1), (i_2, r_2) \in \{(i_{\text{out}}, d+1), (i_{\text{inn}}, d+1), (i_{\text{inn}}, d+2), (i_{\text{inn}}, d+3)\}$.

Remembering the glueing described for the arcs, and referring⁹ to Fig. 5.3 (which coincides with the third picture in Fig. 3.11), at x we glue all together (identification of four points)

- vertically the first and the second stratum of $R_{i_{\text{inn}}}$, as a limit case along the arc labelled by d ,
- horizontally the first stratum of $R_{i_{\text{inn}}}$ with the stratum of $R_{i_{\text{out}}}$, as a limit case along the arc labelled by d ,
- vertically the second and the third stratum of $R_{i_{\text{inn}}}$, as a limit case along the arc labelled by $d+1$,
- horizontally the third stratum of $R_{i_{\text{inn}}}$ with the stratum of $R_{i_{\text{out}}}$, as a limit case along the arc labelled by $d+1$;

4.3. $(i_1, r_1), (i_2, r_2) \in \{(i_{\text{out}}, r), (i_{\text{inn}}, r+2)\}$ for $r \in \{d+2, \dots, f_{\text{out}}\}$, provided $f_{\text{out}} > d+1$ (recall that $f_{\text{out}} \geq d+1$). The last $f_{\text{out}} - (d+1)$ strata are glued horizontally two by two at x (identification of two points).

Remark 5.2.2 We can rephrase the previous definitions as follows.

- Condition 3.2 is equivalent to

$$(x, i_{+-}, d_1 + 1) \sim (x, i_{+-}, d_1 + 2) \sim (x, i_{++}, d_1 + 1) \sim (x, i_{++}, d_1 + 2).$$

- Condition 3.4 is equivalent to

$$(x, i_{-+}, d_2 + 1) \sim (x, i_{-+}, d_2 + 2) \sim (x, i_{++}, d_2 + 3) \sim (x, i_{++}, d_2 + 4).$$

⁹The case of the second picture of Fig. 3.11 can be treated in a similar manner.

- Condition 4.1 is equivalent to $(x, i_{\text{inn}}, r) \sim (x, i_{\text{out}}, r)$ for any $r \in \{1, \dots, d\}$.
- Condition 4.2 is equivalent to

$$(x, i_{\text{out}}, d+1) \sim (x, i_{\text{inn}}, d+1) \sim (x, i_{\text{inn}}, d+2) \sim (x, i_{\text{inn}}, d+3).$$

- Condition 4.3 is equivalent to

$$(x, i_{\text{out}}, r) \sim (x, i_{\text{inn}}, r+2), \quad r \in \{d+2, \dots, f_{\text{out}}\}.$$

It is possible to check that \sim is an equivalence relation on D , and that all elements of D belong to an equivalence class. If $(x, i, r) \in D$, we denote by

$$[(x, i, r)] = \{(y, k, \rho) \in D : (y, k, \rho) \sim (x, i, r)\}$$

the equivalence class of (x, i, r) . Observe that the cardinality of an equivalence class is

$$\#[(x, i, r)] = \begin{cases} 1 & \text{if } x \notin G, \\ 2 & \text{if } x \in G \setminus \text{nodes}(G), \\ 4 & \text{if } x \in \text{crossings}(G), \\ 2 \text{ or } 4 & \text{if } x \in \text{cusps}(G). \end{cases}$$

We define \mathcal{T} as the quotient of D with respect to the equivalence relation \sim , i.e.,

$$\mathcal{T} := D/\sim = \{[(x, i, r)] : (x, i, r) \in D\}, \quad (5.1)$$

which we endow with the quotient topology.¹⁰ The set \mathcal{T} has a natural structure of topological manifold, as described in a precise way in the appendix of this chapter.

5.2.2 Smooth Local Embedding of \mathcal{T} in \mathbb{R}^3

We now show that \mathcal{T} can be endowed with a structure of smooth manifold, and that it can be smoothly *locally* embedded in \mathbb{R}^3 in such a way that the projection map keeps the correct apparent contour. The explicit expressions of the embeddings will be useful also in connection with the uniqueness result (Theorem 5.1.4).

1. $x \notin G$. In this case the local charts which make \mathcal{T} a manifold of class \mathcal{C}^∞ around x are obtained using (x_1, x_2) as parameters. Assume, without loss of generality,

¹⁰That is, if $q : D \rightarrow \mathcal{T}$ is the quotient map, then $U \subseteq \mathcal{T}$ is open if and only if $q^{-1}(U)$ is open in D .

that $x = (0, 0)$ and that $(0, 0)$ is contained in region R_i , so that $f(0, 0) = f(R_i)$. Fix $r \in \{1, \dots, f(0, 0)\}$. Select a neighbourhood $U \subset \subset R_i$ of $(0, 0)$. We define the map $\psi : U \rightarrow \mathcal{T}$ as follows:

$$\psi(x_1, x_2) := [(x_1, x_2), i, r], \quad (x_1, x_2) \in U.$$

Then ψ is a homeomorphism between U and $\psi(U)$, and this latter is a neighbourhood of $[((0, 0), i, r)]$ in \mathcal{T} . The transition functions between two intersecting neighbourhoods $U \subset \mathbb{R}^2_{(x_1, x_2)}$ are the identity; in particular, they are of class \mathcal{C}^∞ . Hence, \mathcal{T} becomes a two-dimensional manifold of class \mathcal{C}^∞ around x .

Now, let

$$e_R : \psi(U) \rightarrow \mathbb{R}^3$$

be defined as follows:

$$e_R([(x_1, x_2), i, r)] := (x_1, x_2, z_r), \quad (x_1, x_2) \in \psi(U),$$

where¹¹

$$z_r := -1 + 2 \frac{r}{f(R_i) + 1}.$$

Then e_R is a smooth local embedding of $\psi(U)$ in \mathbb{R}^3 .

2. $x \in a \in \text{arcs}(G)$. As we shall see, and in contrast with the previous case, here the local charts which make \mathcal{T} a manifold of class \mathcal{C}^∞ around x are obtained using (x_1, z) as parameters. Assume that $x = (0, 0)$; let $U \subset \Omega$ be a sufficiently small neighbourhood of x , in such a way that $a \cap U$ can be written as $x_2 = g_a(x_1)$, for x_1 in an open interval $I \subset \mathbb{R}$, for a function $g_a \in \mathcal{C}^\infty(I)$ satisfying $g_a(0) = g'_a(0) = 0$. Suppose in addition that $f = f_{\min} = 0$ in $U \cap \{x_2 < g_a(x_1)\}$ (hence $f = 2$ in $U \cap \{x_2 > g_a(x_1)\}$), the general case being an easy generalization.¹² Denote by R_i the region containing $U \cap \{x_2 > g_a(x_1)\}$. Note that $d = 0$ in $a \cap U$. Set

$$V_r := \left\{ [(x_1, x_2), i, r] : (x_1, x_2) \in U, x_2 \geq g_a(x_1) \right\}, \quad r \in \{1, 2\},$$

¹¹Observe that $z_r \in (-1, 1)$, so that all points that we consider belong to \mathcal{Q} . Moreover $z_{r_1} < z_{r_2}$ if $r_1, r_2 \in \{1, \dots, f(0, 0)\}$ and $r_1 < r_2$.

¹²If $f_{\min} > 0$, it is enough to consider the strata that are transverse in correspondence of U , and that are either in front of a parametrized stratum, or behind it, in dependence of the index r and of the value of d .

and

$$V_a := V_1 \cup V_2,$$

which is a neighbourhood of $[(0, 0), i, 1]$ in \mathcal{T} .

Define, for (x_1, z) in a neighbourhood¹³ W of $(0, 0)$,

$$G_a(x_1, z) := z^2 + g_a(x_1). \quad (5.2)$$

Clearly, if W is small enough, we have $G_a \in \mathcal{C}^\infty(W)$. Let $\psi_a : W \rightarrow \mathcal{T}$ be the parametrization map defined as follows:

$$\psi_a(x_1, z) := \begin{cases} ((x_1, G_a(x_1, z)), i, 1) & \text{if } z < 0, \\ ((x_1, G_a(x_1, z)), i, 2) & \text{if } z \geq 0. \end{cases}$$

Provided W is small enough, we can ensure that $\psi_a(W) \subset V_a$. One checks that ψ is a homeomorphism between W and $\psi_a(W)$, so that the chart $\psi_a^{-1} : \psi_a(W) \rightarrow W$ is well defined and continuous.

Moreover, it turns out that the transition functions between two open sets W of $\mathbb{R}_{(x_1, z)}^2$ (the domains of two parametrizations as above) having nonempty intersection are of class \mathcal{C}^∞ . In addition, the transition functions between two open sets, one of the form U as in case 1, and the other one of the form W , are of class \mathcal{C}^∞ .¹⁴

We are now in a position to construct a smooth local embedding in \mathbb{R}^3 . Let

$$e_a : V_a \rightarrow \mathbb{R}^3$$

be defined as follows:

$$\begin{aligned} e_a([(x_1, z), i, 1)]) &:= (x_1, G_a(x_1, z), z), \\ e_a([(x_1, z), i, 2)]) &:= (x_1, G_a(x_1, z), z), \end{aligned} \quad (x_1, z) \in \psi(W). \quad (5.3)$$

Note that e_a is of class \mathcal{C}^∞ ,

$$e_a(V_a) = \{(x_1, x_2, z) : x_2 = G_a(x_1, z)\} = \text{graph}(G_a)$$

locally around the origin, and that the tangent plane to $e_a(V_a)$ at $(0, 0, 0) \in \mathbb{R}_{(x_1, x_2)}^2 \times \mathbb{R}_z$ is $\{x_2 = 0\}$.

¹³We suppose, as usual, that $|z| < 1$.

¹⁴The reason being that the function $\rho \in (0, +\infty) \rightarrow \sqrt{\rho}$ is of class \mathcal{C}^∞ .

Let us show that the projection map π leaves invariant the apparent contour. The singular set is given by

$$e_a(V_a) \cap \left\{ (x_1, x_2, z) : \frac{\partial}{\partial z} G_a(x_1, z) = 0 \right\} = e_a(V_a) \cap \{z = 0\} = \{(x_1, g_a(x_1), 0)\},$$

whose orthogonal projection on the plane $\{z = -1\}$ coincides with $\{x_2 = g_a(x_1)\}$.

3. $x \in \text{crossings}(G)$. In this case the embedding is constructed by repeating twice the previous argument, making a suitable translation of the strata corresponding to the highest value of d , and then inserting, if necessary, the proper number of transversal layers at the correct depth.

Concerning the next case, it is helpful to recall the discussion made in Example 3.3.3.

4. $x \in \text{cusps}(G)$. Assume that $x = (0, 0)$, that the two arcs meeting at x belong, locally, to $\{(x_1, x_2) \in \Omega : x_1 \in [0, \delta)\}$ for a suitable $\delta > 0$ small enough, and that have horizontal tangent line at 0. Up to a diffeomorphism of \mathbb{R}_x^2 , we know that the two arcs have equation

$$x_2 = \pm x_1^{3/2}, \quad x_1 \in [0, \delta),$$

for $\delta > 0$ sufficiently small.

We can suppose that the three involved strata are numbered with $r = 1, 2, 3$. Consequently, we can assume¹⁵ that $f_{\min} = 1$ in $R_{\text{out}} = R_{r_2}$ (hence $f = 3$ in $R_{\text{inn}} = R_{r_1}$) and $d = 1$ in the lower arc of the cusp. Hence $d = 0$ in the upper arc of the cusp; see Fig. 5.3.

We want to use (x_1, z) as local parameters. Let

$$W := (-\varepsilon, \varepsilon) \times (-\delta, \delta) \subset \mathbb{R}_{(x_1, z)}^2$$

be a sufficiently small rectangular neighbourhood of $(0, 0)$. Let V_c be a neighbourhood of $[(0, 0), i_{\text{inn}}, 1)]$ in \mathcal{T} , which is of the form

$$\begin{aligned} V_c := & \left\{ ((x_1, x_2), i_{\text{inn}}, r) : (x_1, x_2) \in U, x_1 \geq 0, -x_1^{3/2} < x_2 < x_1^{3/2}, r \in \{1, 2, 3\} \right\} \\ & \cup \{((x_1, x_2), i_{\text{out}}, 1) : (x_1, x_2) \in U, x_1 < 0\} \\ & \cup \left\{ ((x_1, x_2), i_{\text{out}}, 1) : (x_1, x_2) \in U, x_1 \geq 0, x_2 > x_1^{3/2} \text{ or } x_2 < -x_1^{3/2} \right\}, \end{aligned}$$

¹⁵For simplicity, here f takes odd positive integer values: in order our discussion to be included in the standard framework where f takes values in $2\mathbb{N}$, it is enough to add a transversal layer at the proper depth.

for a suitable sufficiently small neighbourhood U of the origin. Define

$$F_c(x_1, z) := \frac{1}{2} (z^3 - 3zx_1), \quad (x_1, z) \in W.$$

Let $\psi_c : W \rightarrow \mathcal{T}$ be the parametrization map defined as follows:

$$\psi_c(x_1, z) := \begin{cases} (x_1, F_c(x_1, z), i_{\text{inn}}, 2) & \text{if } x_1 > z^2, \\ (x_1, F_c(x_1, z), i_{\text{inn}}, 3) & \text{if } \frac{z^2}{4} < x_1 < z^2 \text{ and } z < 0, \\ (x_1, F_c(x_1, z), i_{\text{inn}}, 1) & \text{if } \frac{z^2}{4} < x_1 < z^2 \text{ and } z > 0, \\ (x_1, F_c(x_1, z), i_{\text{out}}, 1) & \text{elsewhere.} \end{cases}$$

Provided W is small enough, we can ensure that $\psi_c(W) \subset V_c$. Then ψ_c is a homeomorphism between W and $\psi_c(W)$, so that the chart $\psi_c^{-1} : \psi_c(W) \rightarrow W$ is well defined and continuous.

One checks that the transition functions between two open sets W of $\mathbb{R}_{(x_1, z)}^2$ (the domains of two parametrizations as above) having nonempty intersection are of class C^∞ . In addition, the transition functions between two open sets, one of the form U as in case 1, and the other of the form W , are of class C^∞ .

We are now in a position to construct the smooth local embedding of \mathcal{T} in \mathbb{R}^3 . Let

$$e_c : V_c \rightarrow \mathbb{R}^3$$

be defined as

$$\begin{aligned} e_c([(x_1, z), i_{\text{inn}}, 1)]) &:= (x_1, F_c(x_1, z), z), \\ e_c([(x_1, z), i_{\text{inn}}, 2)]) &:= (x_1, F_c(x_1, z), z), \\ e_c([(x_1, z), i_{\text{inn}}, 3)]) &:= (x_1, F_c(x_1, z), z), \\ e_c([(x_1, z), i_{\text{out}}, 1)]) &:= (x_1, F_c(x_1, z), z), \end{aligned} \quad (x_1, z) \in \psi_c(W).$$

The fact that the projection π leaves invariant the apparent contour follows from the discussion in Example 3.3.3.

We denote by M the topological manifold \mathcal{T} endowed with the smooth differential structure described above.

5.2.3 Smooth Global Embedding of M in \mathbb{R}^3

In this section, we show that the local parametrizations introduced in Sect. 5.2.2 can be glued in a smooth way, in the intersection of two coordinate neighbourhoods.

This will allow us to use a partition of unity, which, in the end, will lead to the C^∞ global embedding of M in \mathbb{R}^3 .

We start by introducing a class of multifunctions in the variables (x_1, x_2) . Let $B \subset \mathbb{R}_{(x_1, x_2)}^2$ be an open ball centred at the origin, and denote by $\mathcal{P}(\mathbb{R})$ the set of all subsets of \mathbb{R} . Let $g \in C^\infty(\mathbb{R})$ be a function, satisfying $g(0) = 0$, and let $\text{graph}(g) := \{(x_1, g(x_1)) : x_1 \in \mathbb{R}\}$.

Definition 5.2.3 (Smooth Two-Valued Function) Let $\zeta : B \cap \{x_2 \geq g(x_1)\} \rightarrow \mathcal{P}(\mathbb{R})$. We say that ζ is a smooth two-multifunction with ordered branches glued along $B \cap \text{graph}(g)$, if

$$\zeta(x) = \{\zeta^-(x), \zeta^+(x)\}, \quad x \in B \cap \{x_2 \geq g(x_1)\},$$

where $\zeta^\pm : B \cap \{x_2 \geq g(x_1)\} \rightarrow \mathbb{R}$ are two functions of class C^∞ such that

$$\zeta^- < \zeta^+ \text{ in } B \cap \{x_2 > g(x_1)\},$$

$$\zeta^- = \zeta^+ \text{ on } B \cap \text{graph}(g).$$

We use the notation

$$\zeta = [\zeta^-, \zeta^+]$$

to indicate the smooth two-valued function ζ , and we set

$$\text{graph}(\zeta) := \text{graph}(\zeta^-) \cup \text{graph}(\zeta^+) \subset \mathbb{R}_{(x_1, x_2)}^2 \times \mathbb{R}_z.$$

Definition 5.2.4 (Smoothly Glued Branches) We say that the two branches of ζ are smoothly glued along $B \cap \text{graph}(g)$, if $\text{graph}(\zeta)$ can be seen as the graph of a function ϕ of class C^∞ with respect to the variables (x_1, z) , such that

$$\frac{\partial^2 \phi}{\partial^2 z}(x_1, g(x_1)) > 0.$$

An example of a smooth two-valued function with smoothly glued branches is given by the right-hand side of formula (5.3) (recall (5.2)).

Lemma 5.2.5 (Square-Root of a Smooth Function) Let $I \subset \mathbb{R}$ be an open interval containing the origin, and let $\theta : I \rightarrow [0, +\infty)$ be a function with the following properties:

- $\theta \in C^\infty(I)$,
- $\theta(z) = 0$ if and only if $z = 0$,
- $\theta''(0) > 0$.

Then θ can be written as follows:

$$\theta = \delta^2 \quad \text{in } I,$$

where $\delta \in C^\infty(I)$ has the following properties:

- $\delta(0) = 0$,
- $\delta(z) < 0$ for any $z \in I$ with $z < 0$, and $\delta(z) > 0$ for any $z \in I$ with $z > 0$,
- $\delta'(0) > 0$.

Precisely

$$\delta(z) := \begin{cases} \frac{z}{|z|} \sqrt{\theta(z)} & \text{if } z \in I \setminus \{0\}, \\ 0 & \text{if } z = 0. \end{cases}$$

Proof Define

$$h(z) := \begin{cases} \frac{\theta(z)}{z^2} & \text{if } z \in I \setminus \{0\}, \\ \frac{\theta''(0)}{2} & \text{if } z = 0. \end{cases}$$

One checks¹⁶ that $h \in C^\infty(I, (0, +\infty))$, therefore $\sqrt{h} \in C^\infty(I)$. Since $\delta(z) = z\sqrt{h(z)}$, it follows that $\delta \in C^\infty(I)$. Observing¹⁷ that $\delta'(0) > 0$, it follows that δ is locally invertible around zero. \square

Lemma 5.2.6 (Horizontal Interpolation) *Let $I \subset \mathbb{R}$ be an open interval, and let $a, b \in I$ with $a < b$. Suppose that $\delta_1, \delta_2 \in C^\infty(I)$ are two functions having the following properties:*

- $\delta_1(z) < 0$ for any $z \in I$ with $z < a$, and $\delta_1(z) > 0$ for any $z \in I$ with $z > a$,
- $\delta'_1(a) > 0$,
- $\delta_2(z) < 0$ for any $z \in I$ with $z < b$, and $\delta_2(z) > 0$ for any $z \in I$ with $z > b$,
- $\delta'_2(b) > 0$.

Then, given $\omega \in C^\infty(\mathbb{R}, [0, 1])$ with $\omega'(0) = 0$, and setting

$$c := (1 - \omega(0))a + \omega(0)b,$$

¹⁶For example, let us check that $h \in C^1(I)$. For $z \neq 0$ we have $h'(z) = \frac{z\theta'(z) - 2\theta(z)}{z^3}$, so that, applying twice de l'Hôpital's theorem, we have $\lim_{z \rightarrow 0} h'(z) = \lim_{z \rightarrow 0} \frac{\theta''(z)}{6} = \frac{\theta''(0)}{6}$, and therefore h is differentiable at the origin. In a similar manner, one proves that all derivatives of h are continuous in I .

¹⁷For $z > 0$ we have $\theta(z) = \frac{z^2}{2}\theta''(\tau)$ and $\theta'(z) = z\theta''(\nu)$, for two suitable points $\tau, \nu \in (0, z)$. Hence $\delta'(z) = \frac{\theta'(z)}{2\sqrt{\theta(z)}} = \frac{\theta''(\nu)}{\sqrt{2\theta''(\tau)}}$, and therefore $\delta'(0) = \lim_{z \rightarrow 0^+} \delta'(z) = \sqrt{\frac{\theta''(0)}{2}} > 0$.

there exist a neighbourhood J of c and a function $\delta_3 \in C^\infty(J)$, such that, for any $\delta \in \mathbb{R}$ with $|\delta|$ sufficiently small, if we set

$$z_3 := (1 - \omega(\delta))z_1 + \omega(\delta)z_2, \quad \text{with } z_1, z_2 \in I \text{ so that } \delta_1(z_1) = \delta \text{ and } \delta_2(z_2) = \delta,$$

we have $z_3 \in J$ and

$$\delta = \delta_3(z_3).$$

Proof The function δ_1 is locally invertible around a with a smooth inverse, and, in the same manner, δ_2 is locally invertible around b with a smooth inverse. Hence, for $\delta \in \mathbb{R}$ with $|\delta|$ sufficiently small, we can consider the C^∞ -function

$$h(\delta) := (1 - \omega(\delta))\delta_1^{-1}(\delta) + \omega(\delta)\delta_2^{-1}(\delta).$$

Using the assumption $\omega'(0) = 0$, we have

$$h'(0) = (1 - \omega(0))\frac{1}{\delta_1'(a)} + \omega(0)\frac{1}{\delta_2'(b)} > 0.$$

Inverting h in a suitable neighbourhood of zero, we can now define, locally around $c = h(0)$, the function δ_3 as

$$\delta_3 := h^{-1}.$$

Then the assertion of the lemma follows from the inverse function theorem. \square

Proposition 5.2.7 (Interpolation of Two Multifunctions) *Let R be a region, and let B be a sufficiently small open ball contained in Ω and intersecting G and R . Let $[\zeta_1^-, \zeta_1^+] : B \cap \overline{R} \rightarrow \mathcal{P}(\mathbb{R})$ (respectively $[\zeta_2^-, \zeta_2^+] : B \cap \overline{R} \rightarrow \mathcal{P}(\mathbb{R})$) be a smooth two-valued function smoothly glued along $B \cap G$ of the form (5.3). Let $w_i, i = 1, 2$ be two nonnegative functions of class $C^\infty(B)$ with*

$$w_1 + w_2 = 1 \quad \text{in } B \cap \overline{R}.$$

Define

$$\eta^\pm := w_1\zeta_1^\pm + w_2\zeta_2^\pm \quad \text{in } B \cap \overline{R}, \tag{5.4}$$

and

$$\eta := [\eta^-, \eta^+] : B \cap \overline{R} \rightarrow \mathcal{P}(\mathbb{R}).$$

Then η is a smooth two-valued function, smoothly glued along $B \cap G$.

Proof We need to prove the smoothness assertion in a neighbourhood of a point x belonging to the relative interior of an arc $a \subseteq G$. Assume without loss of generality that $x = 0$, and that locally a can be written as $a = \{x_2 = g_a(x_1)\}$, for a function g_a of class \mathcal{C}^∞ with $g_a(0) = g'_a(0) = 0$, and that \overline{R} lies locally in $\{x_2 \geq g_a(x_1)\}$. Hence, locally, ζ_1^\pm and ζ_2^\pm are defined for $x_2 \geq g(x_1)$.

By assumption we have that, for $i \in \{1, 2\}$,

$$\{(x_1, x_2, \zeta_i^-(x_1, x_2)) : (x_1, x_2) \in \overline{B} \cap \overline{R}\} \cup \{(x_1, x_2, \zeta_i^+(x_1, x_2)) : (x_1, x_2) \in \overline{B} \cap \overline{R}\}$$

can be viewed as the graph $\{x_2 = \phi_i(x_1, z)\}$ of a function ϕ_i , of class \mathcal{C}^∞ in its domain. In addition

- the function $z \rightarrow \phi_i(x_1, z)$ is strictly convex in a neighbourhood of $z = 0$,
- locally $\phi_i(x_1, z) \geq g_a(x_1)$, with the equality only¹⁸ if

$$z = z_i(x_1) := \zeta_i^\pm(x_1, g_a(x_1)),$$

namely $\phi_i(x_1, z_i(x_1)) = g_a(x_1)$.

Define, for a fixed choice of x_1 ,

$$\theta_i(z) := \phi_i(x_1, z + z_i(x_1)) - g_a(x_1), \quad i \in \{1, 2\},$$

for z in a suitable open interval $(-\sigma, \sigma)$ of $0 \in \mathbb{R}$. The function θ_i is of class $\mathcal{C}^\infty((-\sigma, \sigma))$ and satisfies the assumptions of Lemma 5.2.5, since $\theta_i(z) = 0$ only if $z = 0$ and $\theta_i''(0) > 0$.

Therefore there exists $\sigma > 0$ such that θ_i can be locally represented as

$$\theta_i = (\delta_i)^2 \quad \text{in } (-\sigma, \sigma),$$

for an invertible function $\delta_i \in \mathcal{C}^\infty((-\sigma, \sigma))$, which we shall use a new variable in place of the square-root of x_2 .

Now, we write the function $w_1(x_1, x_2)$ (and, as a consequence, also the function $w_2(x_1, x_2) = 1 - w_1(x_1, x_2)$) in the hypothesis of the proposition, as $\tilde{w}_i(\delta)$ in the new coordinates (x_1, δ) , related to (x_1, x_2) via the \mathcal{C}^∞ -map

$$w_1(x_1, x_2) = w_1(x_1, \delta^2 + g_a(x_1)) =: \tilde{w}_1(\delta).$$

Observe that

- $\tilde{w}'_1(0) = 0$,
- from Lemma 5.2.5, $\delta'_i(0) > 0$.

¹⁸Recall that $\zeta_i^+ = \zeta_i^-$ at $(x_1, g_a(x_1))$.

Therefore, we are in a position to apply Lemma 5.2.6 with the choice $\omega = \tilde{w}_1$, and we get the function δ_3 . Let us define $\theta_3(x_1, \cdot) := \delta_3^2(\cdot)$, and

$$\phi_3(x_1, z) := \theta_3(x_1, z) + g_a(x_1).$$

We have $\phi_3(x_1, z) \geq g_a(x_1)$; moreover, ϕ_3 assumes its minimal value at $z = \eta^\pm(x_1, g_a(x_1))$. Eventually, it is possible to check that the two branches of the inverse of ϕ_3 are exactly the functions η^\pm defined in (5.4). \square

5.2.3.1 Partition of Unity

Let us cover G and all regions with a *finite* family of overlapping open balls, with the property that any node of G belongs to only one ball. We have to show that in the intersection of two balls of the covering the definitions given in Sect. 5.2.2 can be glued in a smooth way. Since a node belongs to only one ball, we need to prove our claim only locally around the relative interior of the arcs of G , therefore we shall be concerned only with glueing two multifunctions of the form (5.3). Then the assertion follows from Proposition 5.2.7.

Taking a partition of unity¹⁹ relatively to the above-mentioned covering, we finally obtain the global C^∞ -embedding of M in \mathbb{R}^3 .

5.2.4 Definition of the 3D-Shape

Since M is two-dimensional, closed and embeddable in \mathbb{R}^3 , the image Σ of the embedding (an orientable surface) divides the space into various connected components (see [7, p. 89]), only one of which is unbounded. We define the 3D shape E as the set of points of \mathbb{R}^3 that are connected with infinity with a generic curve²⁰ intersecting Σ an odd number of times. The image of the embedding²¹ is therefore $\partial E = \Sigma$.

¹⁹We recall that if A_1, \dots, A_n is a finite covering of Ω , a partition of unity subordinated to the covering is given by a family of C^∞ functions $\lambda_1, \dots, \lambda_n : \Omega \rightarrow [0, 1]$ such that $\sum_{i=1}^n \lambda_i(x) = 1$ for any $x \in \Omega$.

²⁰Generic here means the following: the curve has only a finite number of intersection with the image of the embedding, and each intersection is transverse.

²¹Since Σ is orientable, also M turns out to be orientable; in this book, we shall always choose the orientation on M consistently with the induced orientation on Σ .

Remark 5.2.8 (Trivial Covering) The restriction of $\pi_{|\partial E}$ to the preimage $\pi_{|\partial E}^{-1}(R_i)$ is a covering that, in principle, could be nontrivial when R_i is not simply connected: this is not the case here, since the manifold that we want to construct is embedded, and therefore we are allowed to properly order the preimages of the points of R_i .²²

A careful analysis of the local reconstructions described above shows that the solid shape E is stable: in the proof of this assertion one needs to use various facts, for instance, the nondegeneracy assumption $\theta''(0) > 0$ of Lemma 5.2.5.

The proof of Theorem 5.1.1 is concluded. \square

In Theorem 7.4.1 we prove a formula (see [1]) which allows to compute the Euler–Poincaré characteristic of Σ from the apparent contour.

5.3 Proof of Uniqueness

The aim of this section is to prove Theorem 5.1.4. Set $f := f_{\Sigma_1} = f_{\Sigma_2}$ and $d := d_{\Sigma_1} = d_{\Sigma_2}$. Denote by G the apparent contour of Σ_1 (and of Σ_2), namely G is the jump set J_f of f .

We denote, as usual, by R_1, \dots, R_n the regions of G . If $x \in G$, we denote by $R_{\max}(x)$ (respectively $R_{\min}(x)$) the region adjacent to x where $f = f_{\max}$ (respectively $f = f_{\min}$).

Let us start by recalling that the function f is defined at every point of G in a natural way,²³ namely:

$$x \in G \setminus \text{nodes}(G) \Rightarrow f(x) = \frac{f^-(x) + f^+(x)}{2},$$

$$\begin{aligned} x \in \text{crossings}(G) \Rightarrow f(x) &= \frac{1}{4} (f_{\min}(x) + 2(f_{\min}(x) + 2) + f_{\min}(x) + 4) \\ &= f_{\min}(x) + 2, \end{aligned}$$

$$x \in \text{cusps}(G) \Rightarrow f(x) = f_{\text{out}}(x) = f_{\min}(x).$$

²²An example of a nontrivial covering can be constructed by taking the Klein bottle as M , constructed as the square $[0, 1] \times [0, 1]$ with identification of the two horizontal sides, the two vertical sides are also identified but with reversed orientation: $(0, m_2)$ is identified with $(1, 1 - m_2)$. The map φ can then be constructed as $(m_1, m_2) \in M \rightarrow \rho(\cos \theta, \sin \theta)$ with $\theta = 2\pi m_1$ and $\rho = 3 + \cos(2\pi m_2)$. The apparent contour consists of two concentric circles of radii 2 and 4.

²³Compare with Remark 3.3.2: function f , on G_Σ , counts the actual number of intersections of the light ray with the surfaces.

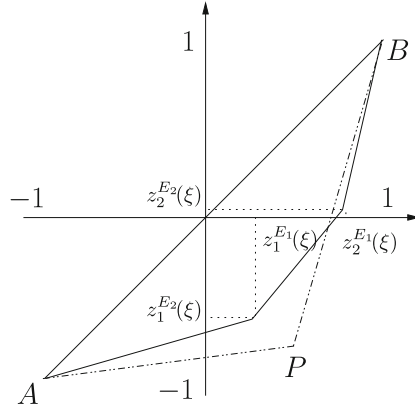


Fig. 5.4 E_1 and E_2 are topologically a three-dimensional ball and $[E_1] = [E_2]$, with the apparent contour (coinciding here for simplicity with the visible contour) which is (topologically) a circle. The bisectrix corresponds to the limit of the piecewise linear interpolation from the side of $R_{\min}(x)$. On the other hand, as ξ converges to $x \in G$ from the side of $R_{\max}(x)$, and as $z_1^{E_1}(\xi), z_2^{E_1}(\xi)$ collapse, and similarly $z_1^{E_2}(\xi), z_2^{E_2}(\xi)$ collapse, the limit piecewise linear interpolation is the *dotted graph*, passing through a suitable point P . The discontinuity between such a piecewise linear function and the bisectrix is apparent. Image taken from [1]

Given $h \in \{1, 2\}$ and $x \in \Omega$, we let $z_1^{E_h}(x), \dots, z_{f(x)}^{E_h}(x)$ be the depth (z -coordinate) of each intersection of ∂E_h with the light ray $\pi^{-1}(x)$ hitting x , ordered as

$$-1 < z_1^{E_h}(x) < \dots < z_{f(x)}^{E_h}(x) < 1.$$

For any $x \in \Omega$ we let

$$\ell_x : z \in [-1, 1] \rightarrow \ell_x(z) \in [-1, 1]$$

be the continuous piecewise linear interpolating function satisfying

$$\begin{cases} \ell_x(z_r^{E_1}(x)) = z_r^{E_2}(x) & \text{if } r \in \{1, \dots, f(x)\}, \\ \ell_x(\pm 1) = \pm 1, \end{cases} \quad (5.5)$$

see Fig. 5.4. Then

- given any $i \in \{1, \dots, n\}$ and $z \in [-1, 1]$, the function

$$x \in R_i \rightarrow \ell_x(z) \quad \text{is continuous,}$$

because, for any $r \in \{1, \dots, f(x)\}$, the points $z_r^{E_1}(x)$ and $z_r^{E_2}(x)$ depend in a continuous way on x , when x varies continuously in region R_i .

- Given $z \in [-1, 1]$,

$$x \in G \implies \lim_{\substack{\xi \in R_{\max}(x) \\ \xi \rightarrow x}} \ell_\xi(z) = \ell_x(z),$$

and in the limit

$$R_{\max}(x) \ni \xi \rightarrow x \in G,$$

two distinct elements of the set $\{z_1^{E_1}(\xi), \dots, z_{f_{\max}(\xi)}^{E_1}(\xi)\}$ collapse to the same point, and the same happens for two elements of the set $\{z_1^{E_2}(\xi), \dots, z_{f_{\max}(\xi)}^{E_2}(\xi)\}$.

- After collapsing to the same point, these elements suddenly vanish when ξ leaves the closure of $R_{\max}(x)$ and enters one of the adjacent regions: as a consequence,

$$x \rightarrow \ell_x(z) \text{ is not continuous at } x \in G,$$

see again Fig. 5.4. Such discontinuities prevent us to use the function ℓ_x for constructing the homeomorphism Θ_x which appears in the statement of the theorem.

Indeed, in the sequel we shall properly modify ℓ_x (recall Definition 5.1.2).

- For any $x \in \Omega$ the function $z \in [-1, 1] \rightarrow \ell_x(z)$ is strictly increasing.

In order to get the continuity with respect to x also at points of G , we shall modify the definition of ℓ_x in a neighbourhood of G . Set, for notational simplicity,

$$z_r^{E_1} = z_r^1, \quad z_r^{E_2} = z_r^2, \quad r \in \{1, \dots, f(x)\}.$$

We consider separately the cases of points on an arc, of crossings and of cusps.

1. $x \in a \in \text{arcs}(G)$. For $h \in \{1, 2\}$ and $\xi \in R_{\max}(x)$, we denote by

$$z_{--}^h(\xi), z_{-}^h(\xi), z_{+}^h(\xi), z_{++}^h(\xi) \in \{-1, z_1^h(\xi), \dots, z_{f_{\max}(\xi)}^h(\xi), 1\},$$

the depths of the four consecutive intersections between the fibre $\pi^{-1}(\xi)$ and ∂E_h , with $z_{--}^h(\xi) < z_{-}^h(\xi) < z_{+}^h(\xi) < z_{++}^h(\xi)$, such that $z_{-}^h(\xi)$ and $z_{+}^h(\xi)$ collapse as $\xi \rightarrow x$, i.e.,

$$\lim_{\substack{\xi \in R_{\max}(x) \\ \xi \rightarrow x}} z_{-}^h(\xi) =: z_{-}^h(x) = z_{+}^h(x) := \lim_{\substack{\xi \in R_{\max}(x) \\ \xi \rightarrow x}} z_{+}^h(\xi),$$

with the convention that

$$f(\xi) = 2 \implies z_{--}^h(\xi) = -1 \quad \text{and} \quad z_{++}^h(\xi) = 1,$$

a case which happens when $d(x) = 0$.

It is worthwhile to observe that

$$z_{--}^h(y) \text{ and } z_{++}^h(y) \text{ are defined also for } y \in R_{\min}(x). \quad (5.6)$$

For $h \in \{1, 2\}$ define

$$\tilde{\lambda}_h(x) := \frac{z_{++}^h(x) - z_{--}^h(x)}{z_{++}^h(x) - z_{--}^h(x)} \in (0, 1).$$

Now, we work on a one-sided strip of a , from the side of $R_{\min}(x)$. We take $\bar{\delta} > 0$ sufficiently small, and, given $y \in \bar{R}_{\min}(x)$ with $0 \leq \text{dist}(y, G) \leq \bar{\delta}$, we denote by $x(y)$ a point on G nearest to y , i.e.,

$$\text{dist}(y, G) = |y - x(y)|.$$

We take $\bar{\delta} > 0$ sufficiently small, so that $x(y)$ is uniquely defined.²⁴ We also restrict ourselves to those $y \in R_{\min}(x)$ for which

$$x(y) \in G \setminus \text{nodes}(G) \quad (5.7)$$

and x stays at some small positive distance from the nodes. We let $S_-^a \subset R_{\min}(x)$ be the one-sided strip of a consisting of the points y with the properties listed above. We introduce a “fictitious” point $\hat{z}_{--}^{E_h}(y) \in (-1, 1)$ as a suitable convex combination of $z_{--}^h(y)$ and $z_{++}^h(y)$ (recall (5.6)). More precisely, we extend $\tilde{\lambda}_h$ on the side of $R_{\min}(x)$ as follows: we set

$$\mu_h(y) := \frac{\text{dist}(y, G)}{2\bar{\delta}} + \frac{\bar{\delta} - \text{dist}(y, G)}{\bar{\delta}} \tilde{\lambda}_h(x(y)), \quad h \in \{1, 2\},$$

and we let

$$\lambda_h(\eta) := \begin{cases} \tilde{\lambda}_h(x) & \text{if } \eta = x \in a, \\ \mu_h(y) & \text{if } \eta = y \in S_-^a \text{ is such that } x(y) \text{ satisfies (5.7).} \end{cases}$$

Observe that λ_h is continuous at the points $x \in a$ under consideration. In addition,

$$\text{dist}(y, G) = \bar{\delta} \Rightarrow \lambda_h(y) = 1/2.$$

²⁴This can be done because the arcs of G are of class C^∞ , and in proximity of a cusp we are considering the region where $f = f_{\min}$.

Now define, for all $y \in S_-^a$ such that $x(y) = x$,

$$\hat{z}_-^h(y) := \lambda_h(y)z_{--}^h(y) + (1 - \lambda_h(y))z_{++}^h(y) \in (z_{--}^h(y), z_{++}^h(y)), \quad h \in \{1, 2\}.$$

Then

$$\begin{aligned} \hat{z}_-^h(x) &= \lambda_h(x)z_{--}^h(x) + (1 - \lambda_h(x))z_{++}^h(x) \\ &= \frac{(z_{++}^h(x) - z_{--}^h(x))z_{--}^h(x) + (z_{++}^h(x) - z_{--}^h(x) - z_{++}^h(x) + z_{--}^h(x))z_{++}^h(x)}{z_{++}^h(x) - z_{--}^h(x)} \\ &= z_{--}^h(x), \end{aligned}$$

and

$$\text{dist}(y, G) = \bar{\delta} \implies \hat{z}_-^h(y) = \frac{1}{2} (z_{--}^h(y) + z_{++}^h(y)), \quad h \in \{1, 2\}. \quad (5.8)$$

Hence, for $y \in R_{\min}(x(y))$ such that $\text{dist}(y, G) = \bar{\delta}$, the fictitious point is exactly the mean value of two consecutive intersections for both surfaces ∂E_1 and ∂E_2 . Referring to Fig. 5.4, in this way, we linearly connect the polygonal APB to the segment AB , and this is done by moving point B onto the midpoint of AB , as the distance of y from G passes from 0 to $\bar{\delta}$.

For any $y \in S_-^a$ we let

$$\ell_y^a : [-1, 1] \rightarrow [-1, 1]$$

be the continuous piecewise linear interpolating function satisfying the analog of (5.5), with the addition of the fictitious point,²⁵ and such that $\ell_y^a(\pm 1) = \pm 1$.

Observe that (5.8) entails that the resulting interpolation is globally affine, and hence ℓ_y^a matches with ℓ_y , which is kept for those $y \in R_{\min}(x)$ such that $\text{dist}(y, G) > \bar{\delta}$. This observation will be used later on when defining the final homeomorphism (see (5.9), below).

2. $p \in \text{crossings}(G)$. If $d_2 > d_1$ (i.e., there is a transverse layer separating the two foldings) it is sufficient to repeat the arguments of case 1, separately for each of the two foldings. Therefore, we can assume $d_1 = d_2$. We denote by a_1 the arc where the labelling does not jump when passing through p (emerging arc), and a_2 the other arc (broken arc). Along a_1 let z_1^h be the collapsing value of the intersection of the light ray with the layer from the side where f is larger, and define z_2^h in a similar manner with respect to a_2 , for $h \in \{1, 2\}$. We let $z_-^h(\xi)$ (possibly equal to -1) and $z_+^h(\xi)$ (possibly equal to 1) be the intersections of

²⁵Namely, in the list of the z -coordinates of all intersections of ∂E_h with $\pi^{-1}(y)$, we now insert also the depth $\hat{z}_-^h(y)$ of the fictitious point.

the light ray $\pi^{-1}(\xi)$ with the layers, just before and after the two foldings. As in case 1, we introduce two fictitious points $\hat{z}_1^h(y)$ (respectively $\hat{z}_2^h(y)$) in a curved rectangle S^{cr} of width $\bar{\delta} > 0$ sufficiently small, having p as a vertex, and located on the side where f is smaller, around a_1 (respectively a_2). In order to ensure the monotonicity of the interpolant, we use the factor $2/3$ (respectively $1/3$) as the limit weight in the convex combination of z_{--}^h and z_{++}^h . We can then define the piecewise linear interpolant

$$\ell_y^{\text{cr}}$$

by adding the fictitious points in the regions where they are defined, and keep ℓ_y elsewhere. In order to connect this construction with the previous one, we also need two small rectangular regions separating the rectangle with each one-sided strip constructed in case 1, where the limit weight passes linearly from $1/3$ to $1/2$.

3. $p \in \text{cusps}(G)$. For $h \in \{1, 2\}$ and $\xi \in \bar{R}_{\text{inn}}(p) = \bar{R}_{\text{max}}(p)$, we denote by

$$z_{-}^h(\xi), z_0^h(\xi), z_{+}^h(\xi),$$

the depth of the three intersection points collapsing to $z_0^h(p)$ at p , with $z_{-}^h(\xi) < z_0^h(\xi) < z_{+}^h(\xi)$; whereas, if $y \in \bar{R}_{\text{out}}(p) = \bar{R}_{\text{min}}(p)$, we denote by $z_{\pm}^h(y)$ the depth of the unique intersection point.

We denote by a_{-} (respectively a_{+}) the arc of G where d has the lower (respectively larger) value. We already know²⁶ that, on the arcs a_{\pm} , two elements of the set $\{z_{-}^h, z_0^h, z_{+}^h\}$ collapse: precisely, we have, for $h \in \{1, 2\}$,

$$\begin{aligned} z_{-}^h(\xi) &= z_0^h(\xi) < z_{+}^h(\xi), & \xi \in a_{-}, \\ z_{-}^h(\xi) &< z_0^h(\xi) &= z_{+}^h(\xi), & \xi \in a_{+}, \\ z_{-}^h(p) &= z_0^h(p) &= z_{+}^h(p). \end{aligned}$$

If $\bar{\delta} > 0$ is small enough, given $y \in R_{\text{out}}(p) = R_{\text{min}}(p)$ with $0 \leq \text{dist}(y, G) \leq \bar{\delta}$, we denote by $x(y)$ the uniquely defined point on G nearest to y . We suppose for definitiveness

$$x(y) \in a_{+}.$$

²⁶Recall that, by definition, the arcs are relatively open.

We introduce the depth $\hat{z}_0^h(y)$ of a fictitious point, for y in a one-sided strip $S_-^c \subset \overline{R}_{\text{out}}(p)$ of width $\bar{\delta}$, as follows: for $y \in S_-^c$ such that $x(y) = x$,

$$\hat{z}_0^h(y) := \frac{\text{dist}(y, G)}{\bar{\delta}} z_{\pm}^h(y) + \frac{\bar{\delta} - \text{dist}(y, G)}{\bar{\delta}} \left(z_{\pm}^h(y) - z_{\pm}^h(x(y)) + z_0^h(x(y)) \right), \quad h \in \{1, 2\},$$

where we observe that $z_{\pm}^h(y)$ joins continuously with $z_{\pm}^h(y)$ across a_+ (and continuously with $z_+^h(y)$ across a_-).

Note that

$$\hat{z}_0^h(y) > z_{\pm}^h(y) \quad \text{provided } \text{dist}(y, G) < \bar{\delta} \quad (\text{and } x(y) \neq p), \quad h \in \{1, 2\},$$

and

$$\text{dist}(y, G) = \bar{\delta} \implies \hat{z}_0^h(y) = z_{\pm}^h(y), \quad h \in \{1, 2\}.$$

One checks that \hat{z}_0^h is well defined and continuous in S_-^c . For any $y \in S_-^c$ we let

$$\ell_y^c$$

be the continuous piecewise linear interpolating function satisfying the analog of (5.5), with the addition of the fictitious point at depth \hat{z}_0^h , and such that $\ell_y^c(\pm 1) = \pm 1$.

Similarly to the case of crossings, we need two small interfacial regions where we join ℓ_y^a with ℓ_y^c , keeping the required continuity and monotonicity.

Now, we collect the local interpolants defined in cases 1–3, and define the final homeomorphism Θ . We set

$$\Theta_y := \begin{cases} \ell_y^c & \text{if } y \in S_-^c, \\ \ell_y^{\text{cr}} & \text{if } y \in S_-^{\text{cr}}, \\ \ell_y^a & \text{if } y \in S_-^a, \\ \ell_y & \text{elsewhere in } \Omega. \end{cases} \quad (5.9)$$

This concludes the proof, since the map Θ is the homeomorphism required in Definition 5.1.2, which makes E_1 and E_2 depth-equivalent. \square

The map $\Theta : Q \rightarrow Q$ constructed in the proof of Theorem 5.1.4 is a homeomorphism and for fixed x , $\Theta(x, \cdot)$ is piecewise linear (hence Lipschitz continuous). One can show that Θ is Lipschitz continuous. Providing a map Θ of class C^∞ seems to be more complicated, and we do not insist on this.

We have seen that existence of a consistent labelling on a complete contour graph is equivalent to the embeddability of a closed surface M in \mathbb{R}^3 corresponding to that graph. The pair (f, d) allows us to construct the abstract surface \mathcal{T} with a cut and paste technique, which turns out to be diffeomorphic to M : the function f allows to

fix the number of copies of each region, and d determines the way to paste them. The statements of Theorems 5.1.1 and 5.1.4 can be rewritten in terms of factorization of maps as follows.

Theorem 5.3.1 (Reconstruction in Terms of Maps) *Let (G, f, d) be a complete labelled contour graph. Then there exist*

- a closed two-manifold M of class C^∞ ,
- a map $\varphi \in C^\infty(M, \mathbb{R}^2)$,
- an embedding $e : M \rightarrow \mathbb{R}^2 \times \mathbb{R}$ of class C^∞ ,
- an orthogonal projection $\pi : \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}^2$ on the first factor,

such that

- (i) $\Sigma := e(M)$ is in general position with respect to π ,
- (ii) φ factorizes as

$$\varphi = \pi \circ e,$$

- (iii) $G = \text{appcon}(\varphi)$, $f = f_\varphi$ and $d = d_\Sigma$.

Moreover, if for $i = 1, 2$, M_i is a smooth closed two-manifold of class C^∞ , $\varphi_i \in C^\infty(M_i, \mathbb{R}^2)$, and $e_i : M_i \rightarrow \mathbb{R}^2 \times \mathbb{R}$ is an embedding of class C^∞ , such that

$$G = \text{appcon}(\varphi_1) = \text{appcon}(\varphi_2), \quad f = f_{\varphi_1} = f_{\varphi_2}, \quad d = d_{\Sigma_1} = d_{\Sigma_2},$$

then there exists a homeomorphism $\Psi : M_1 \rightarrow M_2$ such that the following diagram commutes:

$$\begin{array}{ccc} M_1 & \xrightarrow{\Psi} & M_2 \\ \downarrow \varphi_1 & & \downarrow \varphi_2 \\ \mathbb{R}^2 & \xrightarrow{\text{id}} & \mathbb{R}^2 \end{array}$$

The map Ψ is constructed using Theorem 5.1.4, by means of the map Θ . We actually expect to be able to choose $\Psi : M_1 \rightarrow M_2$ to be a C^∞ diffeomorphism but that, at least, would require Θ to be in $C^\infty(Q, Q)$.

5.A Appendix

In this appendix we show that the structure of topological manifold on \mathcal{T} (defined in (5.1)) can be constructed independently of the embedding shown in Sect. 5.2.2. This observation could be useful in situations which are more general than those considered in the present chapter. Our aim here is to construct an atlas of continuous maps, in such a way that \mathcal{T} is locally a copy of \mathbb{R}^2 around each of its points.

1. $x \notin G$. Assume that $x = (0, 0) \in R_i$ is a point in region R_i , and let $r \in \{1, \dots, f(R_i)\}$. For (s, τ) in a neighbourhood A of the origin of \mathbb{R}^2 , define the map

$$\varphi : A \rightarrow \mathcal{T}, \quad \varphi(s, \tau) := ((s, \tau), i, r). \quad (5.10)$$

Then φ is a continuous local parametrization around x of the r -th stratum of region R_i .

2. $x \in a \in \text{arcs}(G)$. Assume that $x = (0, 0)$, with $d := d(x) = d(a)$, and that a , locally around x , is described by $\{x_2 = 0\}$ (the general case being an easy extension). Suppose also that $f = f_{\min}$ locally in $\{x_2 < 0\}$. Write $R_{\max} = R_{i_+}$ for some index $i_+ \in \{1, \dots, n\}$. For (s, τ) in a neighbourhood A of the origin of \mathbb{R}^2 , define the map

$$\varphi_a : A \rightarrow \mathcal{T}, \quad \varphi_a(s, \tau) := \begin{cases} ((s, \tau), i_+, d + 1) & \text{if } \tau \geq 0, \\ ((s, -\tau), i_+, d + 2) & \text{if } \tau \leq 0. \end{cases} \quad (5.11)$$

Formula (5.11) gives a local parametrization of a neighbourhood of the point in the preimage of x on the singular set; concerning the other preimages, it is enough to argue as in formula (5.10).

Then, recalling item 2.2 in Sect. 5.2.1 and the definition (5.1) of \mathcal{T} and its topology, it follows that φ_a is a continuous local parametrization.

3. $x \in \text{crossings}(G)$. Assume that $x = (0, 0)$ is a crossing between two arcs a and γ as in the right picture of Fig. 5.2 (the remaining cases being similar, as well the corresponding general cases of two curved arcs). Let a (respectively γ) be described, locally around x , by $\{x_2 = 0\}$ (respectively $\{x_1 = 0\}$), and let $d_1 = d(a)$ and $d_2 = d(\gamma)$. We have $f = f_{\min}$ in region $R_{i_{--}}$ (locally contained in $\{x_1 > 0, x_2 < 0\}$), where $i_{--} \in \{1, \dots, n\}$, hence $f = f_{\max}$ in region $R_{i_{++}}$ for some index $i_{++} \in \{1, \dots, n\}$. As before, $R_{i_{+-}}$ (respectively $R_{i_{-+}}$) is the region locally contained in $\{x_1 > 0, x_2 > 0\}$ (respectively in $\{x_1 < 0, x_2 < 0\}$) as displayed in the right picture of Fig. 5.2. For (s, τ) in a neighbourhood A of the origin of \mathbb{R}^2 , define the two maps $\varphi_{\text{cr}}^{(1)} : A \rightarrow \mathcal{T}$ and $\varphi_{\text{cr}}^{(2)} : A \rightarrow \mathcal{T}$ as follows:

$$\varphi_{\text{cr}}^{(1)}(s, \tau) := \begin{cases} ((s, \tau), i_{+-}, d_1 + 1) & \text{if } s \geq 0, \tau \geq 0, \\ ((s, -\tau), i_{+-}, d_1 + 2) & \text{if } s \geq 0, \tau \leq 0, \\ ((s, \tau), i_{++}, d_1 + 1) & \text{if } s \leq 0, \tau \geq 0, \\ ((s, -\tau), i_{++}, d_1 + 2) & \text{if } s \leq 0, \tau \leq 0, \end{cases} \quad (5.12)$$

$$\varphi_{\text{cr}}^{(2)}(s, \tau) := \begin{cases} ((-s, \tau), i_{++}, d_2 + 3) & \text{if } s \geq 0, \tau \geq 0, \\ ((-s, \tau), i_{-+}, d_2 + 1) & \text{if } s \geq 0, \tau \leq 0, \\ ((s, \tau), i_{++}, d_2 + 4) & \text{if } s \leq 0, \tau \geq 0, \\ ((s, \tau), i_{-+}, d_2 + 2) & \text{if } s \leq 0, \tau \leq 0. \end{cases} \quad (5.13)$$

Formulas (5.12) and (5.13) give a local parametrization of a neighbourhood of the two points in the preimage of x which lie on the singular set; concerning the other preimages, it is enough to argue as in formula (5.10).

Then, recalling item 3 in Sect. 5.2.1 and definition (5.1) of \mathcal{T} and its topology, it follows that $\varphi_{\text{cr}}^{(1)}, \varphi_{\text{cr}}^{(2)}$ are continuous local parametrizations.

4. $x \in \text{cusps}(G)$. Assume that the cusp is located at $x = (0, 0)$, and that the two arcs adjacent at x have, locally around x , the equation $x_2^2 = x_1^3$ (the general case being an easy extension). Moreover, assume that the labelling is d on the upper arc and $d + 1$ on the lower arc, as in the right picture of Fig. 5.3 (the case of the left picture being similar). For (s, τ) in a neighbourhood A of the origin of \mathbb{R}^2 , define the map

$$\varphi_c : (s, \tau) \in A \rightarrow ((s, x_2(s, \tau), i(s, \tau), r(s, \tau)) \in \mathcal{T}$$

as follows:

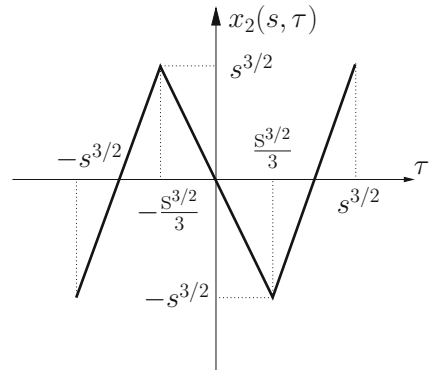
- if $(s, \tau) \in A$ and $\tau^2 > s^3$ then $\varphi_c(s, \tau) := ((s, \tau), i_{\text{out}}, d + 1)$;
 if $(s, \tau) \in A$ and $\tau^2 \leq s^3$ (in particular $s \geq 0$) then

$$\varphi_c(s, \tau) := \begin{cases} ((s, 3\tau + 2s^{3/2}), i_{\text{inn}}, d + 3) & \text{if } -s^{3/2} \leq \tau \leq -\frac{s^{3/2}}{3}, \\ ((s, -3\tau), i_{\text{inn}}, d + 2) & \text{if } -\frac{s^{3/2}}{3} \leq \tau \leq \frac{s^{3/2}}{3}, \\ ((s, 3\tau - 2s^{3/2}), i_{\text{inn}}, d + 1) & \text{if } \frac{s^{3/2}}{3} \leq \tau \leq s^{3/2}, \end{cases} \quad (5.14)$$

see Fig. 5.5. Formula (5.14) gives a local parametrization of a neighbourhood of a point in the preimage of x which lies on the singular set; concerning the other preimages, it is enough to argue as in formula (5.10).

Recalling item 4 in Sect. 5.2.1 and definition (5.1) of \mathcal{T} and its topology, it follows that φ is a continuous local parametrization.

Fig. 5.5 Plot of the graph of the function $\tau \rightarrow x_2(s, \tau)$ in (5.14), for a fixed value of $s > 0$ small enough. Image taken from [1]



It can be checked that the transition functions expressing the coordinate changes are continuous. We have therefore constructed a \mathcal{C}^0 topological manifold \mathcal{T} of dimension two; note that \mathcal{T} is a Hausdorff compact topological space.²⁷

References

1. Bellettini, G., Beorchia, V., Paolini, M.: Topological and variational properties of a model for the reconstruction of three-dimensional transparent images with self-occlusions. *J. Math. Imaging Vision* **32**, 265–291 (2008)
2. Cipolla, R., Giblin, P.: *Visual Motion of Curves and Surfaces*. Cambridge University Press, Cambridge (1999)
3. Golubyatnikov, V.P.: On reconstruction of transparent surfaces from their apparent contours. *J. Inverse Ill-Posed Probl.* **6**, 395–401 (1998)
4. Golubyatnikov, V.P.: Uniqueness questions in reconstruction multidimensional objects from tomography-type projection data. In: *Inverse and Ill-Posed Problems Series*. VSP, Utrecht (2000)
5. Golubyatnikov, V.P., Karaca, I., Ozyilmaz, E., Tantay, B.: On determining the shapes of hyper-surfaces from the shapes of their apparent contours and symplectic geometry measurements. *Sib. Adv. Math.* **10**, 9–15 (2000)
6. Golubyatnikov, V.P., Pekmen, U., Karaca, I., Ozyilmaz, E., Tantay, B.: On reconstruction of surfaces from their apparent contours and the stationary phase observations. In: *Proceedings of International Conference on Shape Modeling and Applications (Shape Modeling International '99)*, pp. 116–120 (1999)
7. Guillemin, V., Pollack, A.: *Differential Topology*. Englewood Cliffs, Prentice-Hall (1974)
8. Karpenko, O.A., Hughes, J.F.: SmoothSketch: 3D free-form shapes from complex sketches. In: *The 33rd International Conference and Exhibition on Computer Graphics and Imaging Techniques*, Boston, Massachusetts, SIGGRAPH 2006, pp. 589–598. ACM, New York (2006)
9. Munkres, J.R.: *Topology*, vol. xvi, 2nd. edn. Upper Saddle River, Prentice Hall (2000)

²⁷If cusps are not present, the quotient space \mathcal{T} is a compact Hausdorff space, which is locally homeomorphic to an open 2-ball (see, e.g., [9, 12 Thms. 74.1, 77.5]). In case of a connected quotient surface \mathcal{T} , we also recall (see, e.g., [10, Thm. 1]) that the genus of a single compact surface embedded in \mathbb{R}^3 can be computed from the apparent contour.

10. Pignoni, R.: On surfaces and their contour. *Manuscripta Math.* **72**, 223–249 (1991)
11. Williams, L.R.: Perceptual completion of occluded surfaces. Ph.D. dissertation, Department of Computer Science, University of Massachusetts, Amherst (1994)
12. Williams, L.R.: Topological reconstruction of a smooth manifold-solid from its occluding contour. *Int. J. Comput. Vision* **23**, 93–108 (1997)

Chapter 6

Completeness of Reidemeister-Type Moves on Labelled Apparent Contours

In this chapter we illustrate the results and report the figures from the paper [3]. More specifically, we shall prove that there exists a finite set of simple, or elementary, moves (also called rules) on labelled apparent contours, such that the following property holds¹: the images Σ_1 and Σ_2 of two stable *embeddings* of a closed smooth (not necessarily connected) surface M in \mathbb{R}^3 are isotopic if and only if their apparent contours can be connected using finitely many isotopies of \mathbb{R}^2 , and a finite sequence of elementary moves or of their inverses (sometimes called “reverses”). The completeness of this result² is of crucial conceptual importance in this book, because it helps in recognizing the topology of a 3D-shape, once its apparent contour is given. We stress that the result refers to embedded (and not immersed) surfaces. We also notice that, forgetting for a moment about the labelling, the list of elementary moves coincides with the classification considered in Chap. 7, on the codimension one stratum of the discriminant hypersurface in $C^\infty(M, \mathbb{R}^2)$. Roughly, the isotopy (consisting of *embeddings*) connecting Σ_1 and Σ_2 intersects the discriminant hypersurface only on the codimension one strata, and such intersections are transverse. We refer to Sect. 7.3 for more details.

The proof of the completeness result relies, basically, on the *classification of singularities* of a stable³ map from a closed *three-manifold* \mathcal{S} into a *three manifold* \mathcal{T} , which was briefly discussed in Example 2.1.10, and on the *density* of $\text{Stable}(\mathcal{S}, \mathcal{T})$ in $C^\infty(\mathcal{S}, \mathcal{T})$.

¹See Theorem 6.0.3 and Corollary 6.6.5 for a precise statement.

²Namely, the fact that there are *no other* moves, besides those in the list of Sect. 6.1, necessary to connect two apparent contours of isotopic surfaces.

³In [6, Definition 2] a different notion of equivalence between maps is introduced. Such a definition can be more suitable when the target space is the cartesian product of a two-dimensional manifold with \mathbb{R} .

The program `appcontour` described in Chaps. 9 and 10, and the results of Chap. 8 are heavily based on the use of the elementary moves devised in the completeness theorem.

Recall from Sect. 2.3 the notion of \mathbb{R}^3 -ambient isotopy with compact support.

Definition 6.0.1 (\mathbb{R}^3 -Ambient Isotopic Embeddings) Let e_1, e_2 be two C^∞ embeddings in \mathbb{R}^3 of a closed C^∞ surface M . We say that e_1 and e_2 are $C^\infty\mathbb{R}^3$ -ambient isotopic (briefly, ambient isotopic), if there exists an \mathbb{R}^3 -ambient isotopy $H : \mathbb{R}^3 \times [0, 1] \rightarrow \mathbb{R}^3$ with compact support, such that

$$H_1 \circ e_1 = e_2 \quad \text{in } M.$$

Let us denote by M_1 and M_2 two closed two-dimensional manifolds of class C^∞ .

Definition 6.0.2 (Ambient Isotopic Surfaces) Let

$$e_1 : M_1 \rightarrow \mathbb{R}^3, \quad e_2 : M_2 \rightarrow \mathbb{R}^3$$

be two stable C^∞ embeddings. We say that $\Sigma_1 := e_1(M)$ and $\Sigma_2 := e_2(M)$ are $C^\infty\mathbb{R}^3$ -ambient isotopic (briefly, ambient isotopic), if there exists an \mathbb{R}^3 -ambient isotopy $H : \mathbb{R}^3 \times [0, 1] \rightarrow \mathbb{R}^3$ with compact support such that

$$H_1(\Sigma_1) = \Sigma_2.$$

Note that if Σ_1 and Σ_2 are ambient isotopic, then M_1 and M_2 are smoothly diffeomorphic.

The next result ensures that, in order to check that $\Sigma_1 := e_1(M)$ and $\Sigma_2 := e_2(M)$ are ambient isotopic, it is sufficient to find $\mathcal{F} \in \text{Diff}_c(\mathbb{R}^3)$ such that $\mathcal{F}(\Sigma_1) = \Sigma_2$. In this way, the dependence on “time” $t \in [0, 1]$ is suppressed. We refer to [5, p. 10] (see also [7]) for similar properties, in the case of topological embeddings of knots.

Theorem 6.0.3 (\mathbb{R}^3 -Ambient Isotopies and Diffeomorphisms) *Let e_1, e_2 be two stable C^∞ embeddings in \mathbb{R}^3 of a closed C^∞ surface M . The two following assertions are equivalent:*

- e_1 and e_2 are ambient isotopic;
- there exists $\mathcal{F} \in \text{Diff}_c(\mathbb{R}^3)$ such that $\mathcal{F} \circ e_1 = e_2$.

Proof It follows from Theorem 2.3.3. □

In a similar manner, it is possible to prove the following result.

Theorem 6.0.4 *Let $e_1 : M_1 \rightarrow \mathbb{R}^3, e_2 : M_2 \rightarrow \mathbb{R}^3$ be two C^∞ embeddings. The two following assertions are equivalent:*

- $\Sigma_1 := e_1(M_1)$ and $\Sigma_2 := e_2(M_2)$ are ambient isotopic;
- there exists $\mathcal{F} \in \text{Diff}_c(\mathbb{R}^3)$ such that $\mathcal{F}(\Sigma_1) = \Sigma_2$.

6.1 Moves on a Labelled Apparent Contour

In this section we list the *moves*, or *rules* (namely, local topological modifications) on a labelled apparent contour. As we shall see, there are six basic moves, that correspond to a general deformation of the corresponding embedded surface; they can be used in exactly the same way as the Reidemeister moves on link diagrams. We show that this set of moves is complete (see Theorem 6.4.6 and Corollary 6.6.2). This essentially means that two embedded surfaces in general position with respect to a fixed projection, that can be deformed into each other by an \mathbb{R}^3 -ambient isotopy, have apparent contours that can be connected using only a finite sequence of such moves and a finite set of \mathbb{R}^2 -ambient isotopies.

Definition 6.1.1 (Reidemeister-Type Moves (Or Rules)) The moves on a labelled apparent contour are denoted by

$$K, L, B, C, S, T.$$

They are defined in Fig. 6.2, by identifying a box in \mathbb{R}^2 diffeomorphic to the box on the left side of the picture and replacing it with a box diffeomorphic to the box on the right.

We require that the moves leave unchanged a (small) neighbourhood of the boundary of the box. The letters are motivated by the following terminology:

- K kasanie (= tangency),
- L lips,
- B beak-to-beak,
- C cusp-fold,
- S swallow's tail,
- T triple point.

Remark 6.1.2 (Inverse Moves) Except for the move T, the same definition as Definition 6.1.1 can be given by switching the role of the two boxes: this is equivalent to reverse the orientation of the t -axis, and to consider the inverse moves as temporal inverse moves (see also Sect. 10.4.4). The corresponding moves are denoted by

$$K^{-1}, L^{-1}, B^{-1}, C^{-1}, S^{-1}.$$

The (direct) moves K, L, B, C, S, T are chosen in such a way that they simplify the local topology of the apparent contour (i.e., they decrease the number of crossings/cusps). There is no distinction between direct and inverse moves of type T, as we shall explain in the sequel.

Since the apparent contours that we consider are oriented, different orientations determine different moves (see also Sect. 10.4.1); for simplicity of notation in Fig. 6.2, the orientations are often not depicted. Moreover, we do not display the values of f before and after the moves, since they can be inferred from the orientation of the apparent contour (see Lemma 2.2.9).

6.1.1 List of All Simple Rules

Now, we list the simple (or elementary) moves, taking into account the labelling (see also Sect. 10.4.1).

- *The four moves of type K.* The moves of type K are subdivided into four different cases as follows. First of all, possibly applying a rotation of 180° , we can suppose that the arc having the two extremal points on the left is in front of the other arc. Accordingly, the four moves are classified on the basis of the four possible orientations of the two arcs, and they are denoted by

$$K0, K1, K1b, K2,$$

as shown in Fig. 6.1 (see again Sect. 10.4.1). These four moves are parametrized by two nonnegative integer numbers d and k , as explained in Sect. 6.1.3 below.⁴

- *The moves L and B.* Possibly applying a rotation of 180° , we can assume that the upper arc has the highest value of d . Then there is one move L and one move B, as depicted in Fig. 6.2.⁵
- *The eight moves of type C.* We divide the moves of type C into two groups of four different types. First we consider the case when the cusp is in front of the (vertical) arc. The first of the pictures for C in Fig. 6.2 is, in turn, divided into four cases, depending⁶ on whether the value of d decreases (decreasing cusp) or increases (increasing cusp) when travelling along the cusp, and on the orientation of the vertical arc. In the second picture the cusp is behind the vertical arc: similarly as before, we have four cases. The set of values taken by d along the

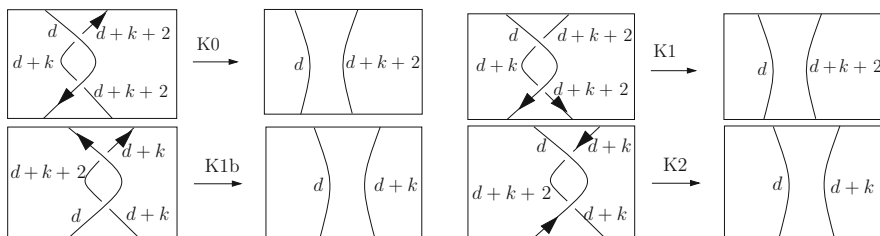


Fig. 6.1 Image taken from [3]. List of K-moves. Here $d \in \mathbb{N}$ and $k \in \mathbb{N}$; compare also with Fig. 3.11

⁴A realization in space of these moves involves two folds of the surface which can be “far one from the other”.

⁵The move L can be realized in space by considering the surface in Fig. 1.4, by gradually reducing the “hill”. The inverse of a move B can be realized by straightening the central part of a depression in a long “wave” with two parallel arcs corresponding to the crease and the valley of the wave.

⁶Compare also with Definition 8.1.1.

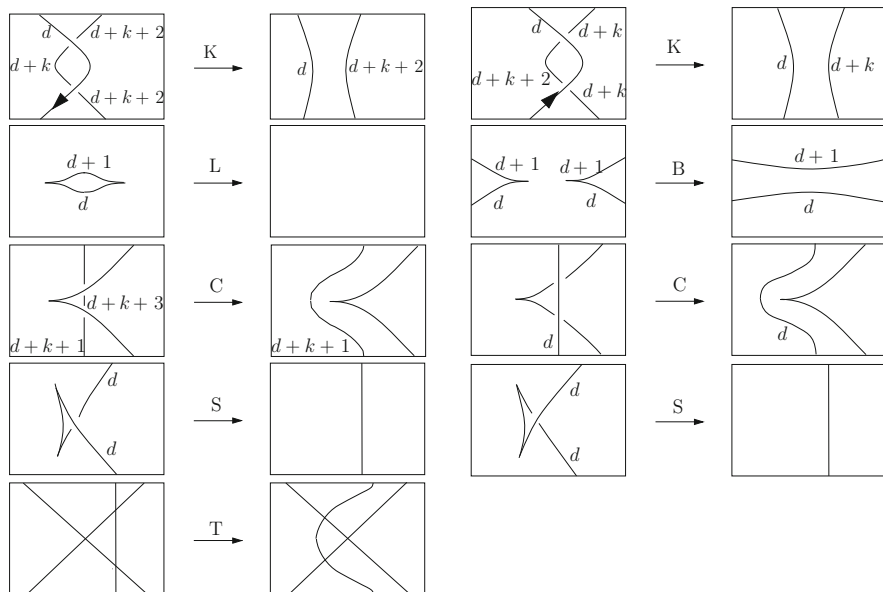


Fig. 6.2 The basic Reidemeister-type moves on a labelled apparent contour. Image taken from [3]

cusped arc are $\{d+k, d+k+1, d+k+2, d+k+3\}$. Again the meaning of the two parameters d and k is explained in Sect. 6.1.3.

- *The two moves of type S.* The moves of type S are divided into two groups: in the first picture of Fig. 6.2 the value of d jumps up by two at the crossing and the two cusps are decreasing, whereas in the second picture the two cusps are increasing.
- *The 16 moves of type T.* The three arcs carry the ordering given by their relative depth (increasing values of d); we can always rotate the picture so that the *nearest* arc (lowest d) is the vertical one. Then, we have two different possibilities for the position of the intermediate and furthest arcs. Each of the three arcs can be oriented in two ways: the internal triangular region can lie on the left or on the right. In the end we have 16 different possibilities which however also account for the corresponding *time reversed* moves, in the sense that the inverse of a T move is still a T move (differently to what happens for all other moves). If d_1 , k_1 and k_2 denote respectively the number of layers in front of the first fold (nearest arc), the number of layers interposed between the first and second fold and the number of layers interposed between the second and the third fold, then $d = d_1$ on the first arc, $d_1 + k_1$ and $d_1 + k_1 + 2$ are the two values of d on the second arc, and the values of d on the third arc are contained in the set $\{d_1 + k_1 + k_2 + i : i = 0, 2, 4\}$, the precise values depending on the orientation of the first and second arcs.

Composition of simple moves will be largely used in Chaps. 8 and 10.

Remark 6.1.3 (Number of Layers Involved) In order to make a complete classification of the moves, the number of layers, at different depths, of the corresponding embedded surface⁷ must be taken into account: this introduces further degrees of freedom in the list of different moves, as follows. Moves L, B and S have one nonnegative integer parameter d , counting the number of layers in front of the fold. Moves of type K and C have two nonnegative integer parameters d and k , counting the number of layers in front of the first fold, and the number of layers in between the two folds. Moves of type T have three nonnegative integer parameters, given respectively by the number of layers in front of the first fold, by the number of layers between the first and the second fold, and by the number of layers between the second and the third fold.

Remark 6.1.4 Moves B, L and S elide pairs of cusps: we refer to Sect. 8.1 for more information.

We can conclude this section with the following useful notion, which will be generalized further in Definition 8.3.1.

Definition 6.1.5 (Reidemeister-Equivalence of Labelled Contour Graphs) We say that two labelled apparent contour graphs are Reidemeister-equivalent if they can be connected by using a finite sequence of direct or inverse Reidemeister-type moves, and a finite number of \mathbb{R}^2 -ambient isotopies with compact support.

More generally, if $\mathcal{O} \subset \mathbb{R}^2$ is a bounded open set, we say that two labelled apparent contour graphs are Reidemeister-equivalent in \mathcal{O} if they can be connected by using a finite sequence of direct or inverse Reidemeister-type moves in \mathcal{O} , and a finite number of \mathbb{R}^2 -ambient isotopies compactly supported in \mathcal{O} .

6.2 Stratifications and Stratified Morse Functions

We shall recall here briefly a few facts about singularity theory (see, e.g., [1, 2, 8, 20, 22–24] and the references therein, and Sect. 2.1). Let \mathcal{S} denote a closed smooth manifold of dimension three, and let \mathcal{T} be a smooth manifold of dimension three without boundary. Remembering the concept of stability given in Definition 2.1.2, we recall that if F is stable, then any map equivalent to F is stable. Moreover, for the present choice of the dimension of \mathcal{S} and \mathcal{T} , we have⁸ that $\text{Stable}(\mathcal{S}, \mathcal{T})$ is dense in $\mathcal{C}^\infty(\mathcal{S}, \mathcal{T})$.

⁷See Chap. 5.

⁸See Theorem 2.1.14.

6.2.1 Stratifications Induced by a Stable Map

In Example 2.1.10 we have seen that a map $F \in \text{Stable}(\mathcal{S}, \mathcal{T})$ has a critical set in the manifold \mathcal{S} consisting of:

- *Folds*. Denoted by $S_1(F)$, it is a smooth submanifold of \mathcal{S} of codimension 1;
- *Pleats*. Denoted by $S_{1_2}(F) \subset S_1(F)$, it is a smooth submanifold of \mathcal{S} of codimension 2 in \mathcal{S} ;
- *Swallow's tails in \mathcal{S}* . Denoted by $S_{1_3}(F) \subset S_{1_2}(F)$, it is a finite set of points of \mathcal{S} .

These submanifolds allow to stratify \mathcal{S} as follows. We define

$$\begin{aligned} X_0(F) & \text{ the set of regular points of } F, \\ X_1(F) & := S_1(F) \setminus S_{1_2}(F), \\ X_2(F) & := S_{1_2}(F) \setminus S_{1_3}(F), \\ X_3(F) & := S_{1_3}(F), \end{aligned} \tag{6.1}$$

the index i in $X_i(F)$ denoting the codimension in \mathcal{S} .

Observe that \mathcal{S} is the union of the mutually disjoint smooth submanifolds $X_j(F)$, and

$$\overline{X_j(F)} = \bigcup_{j \leq h \leq 3} X_h(F), \quad j \in \{0, 1, 2, 3\}.$$

Definition 6.2.1 (Stratification of \mathcal{S}) We call $\{X_0(F), X_1(F), X_2(F), X_3(F)\}$ the stratification of \mathcal{S} associated with the stable map F , and $X_i(F)$ is called the stratum of codimension i , for any $i = 0, 1, 2, 3$.

When no confusion is possible, for notational simplicity, we drop the dependence on F of the sets on the left-hand side of (6.1), thus setting

$$X_i := X_i(F).$$

We are now in a position to introduce the stratification of \mathcal{T} . By [8, Chap. 7, Theorem 6.3], if $F \in \text{Stable}(\mathcal{S}, \mathcal{T})$, the images of the strata X_i through F must intersect transversally. On $F(X_0)$ there are no conditions. The set $F(X_1)$, if it self-intersects, it self-intersects transversally, the resulting intersection is a set of **double curves** of codimension 2 in \mathcal{T} (points having two singular preimages, i.e., two preimages in $S_1(F)$) and a set of **triple points** (codimension 3, points with three singular preimages) in \mathcal{T} . Moreover, $F(X_1)$ intersects $F(X_2)$ transversally, giving a finite set of **cusp-fold** points in \mathcal{T} . The remaining cases have dimension that is too low to give rise to any intersection set. Therefore, we define the following subsets

of the target manifold \mathcal{T} (the index i in $Y_i^j(F)$ denoting the codimension in \mathcal{T} and the superscript j denoting the number of singular preimages):

- $Y_0(F)$ is the set of all $\eta \in \mathcal{T}$ such that no element in $F^{-1}(\eta)$ belongs to $X_1 \cup X_2 \cup X_3$; hence $Y_0(F) \cap F(\mathcal{S}) \subseteq F(X_0)$;
- $Y_1^1(F)$ is the set of all $\eta \in \mathcal{T}$ such that $F^{-1}(\eta)$ has one element in X_1 and the other elements in X_0 . We call $Y_1^1(F)$ the set of **fold surfaces**. It carries a natural orientation, since it separates points where the number of preimages of F jumps of two units;
- $Y_2^1(F)$ is the set of all $\eta \in \mathcal{T}$ such that $F^{-1}(\eta)$ has one element in X_2 , and the other elements in X_0 . We call $Y_2^1(F)$ the set of **cusp curves**;
- $Y_2^2(F)$ is the set of all $\eta \in \mathcal{T}$ such that $F^{-1}(\eta)$ has two elements in X_1 and the other elements in X_0 . We call $Y_2^2(F)$ the set of **double curves**;
- $Y_3^1(F)$ is the set of all $\eta \in \mathcal{T}$ such that $F^{-1}(\eta)$ has one element in X_3 , and the other elements in X_0 . We call $Y_3^1(F)$ the set of **swallow's tails**;
- $Y_3^2(F)$ is the set of all $\eta \in \mathcal{T}$ such that $F^{-1}(\eta)$ has one element in X_2 , one element in X_1 and the other elements in X_0 . We call $Y_3^2(F)$ the set of **cusp-fold points**;
- $Y_3^3(F)$ is the set of all $\eta \in \mathcal{T}$ such that $F^{-1}(\eta)$ has three elements in X_1 and the other elements in X_0 . We call $Y_3^3(F)$ the set of **triple points**.

Figure 6.3 shows an example of the strata sets $Y_2^2(F)$, $Y_2^1(F)$ (twice), $Y_3^1(F)$, $Y_3^2(F)$, $Y_3^3(F)$ in the order.

Now, we can define a natural stratification of the target manifold \mathcal{T} in the smooth submanifolds $Y_0(F)$, $Y_1(F)$, $Y_2(F)$, $Y_3(F)$, where

$$\begin{aligned} Y_1(F) &:= Y_1^1(F), \\ Y_2(F) &:= Y_2^1(F) \cup Y_2^2(F), \\ Y_3(F) &:= Y_3^1(F) \cup Y_3^2(F) \cup Y_3^3(F). \end{aligned} \tag{6.2}$$

When no confusion is possible, for simplicity of notation, we drop the dependence on F in (6.2), thus setting $Y_j^i := Y_j^i(F)$ and $Y_j := Y_j(F)$.

Definition 6.2.2 (Stratification of \mathcal{T}) The set $\{Y_0(F), Y_1(F), Y_2(F), Y_3(F)\}$, denoted also by $\{Y_j\}_F$, is called the stratification of \mathcal{T} induced by the stable map F , and $Y_j(F)$ is called the stratum of codimension j , for any $j = 0, 1, 2, 3$.

We conclude this section by recalling the definition of a stratified Morse function.⁹

⁹See, e.g., [9] and the references therein, or also [15, p. 597, 600, 601]. Usually, a stratified Morse function takes real values; for technical reasons, we consider here the slightly different case of a function taking values in \mathbb{S}^1 , but the definition is essentially the same.

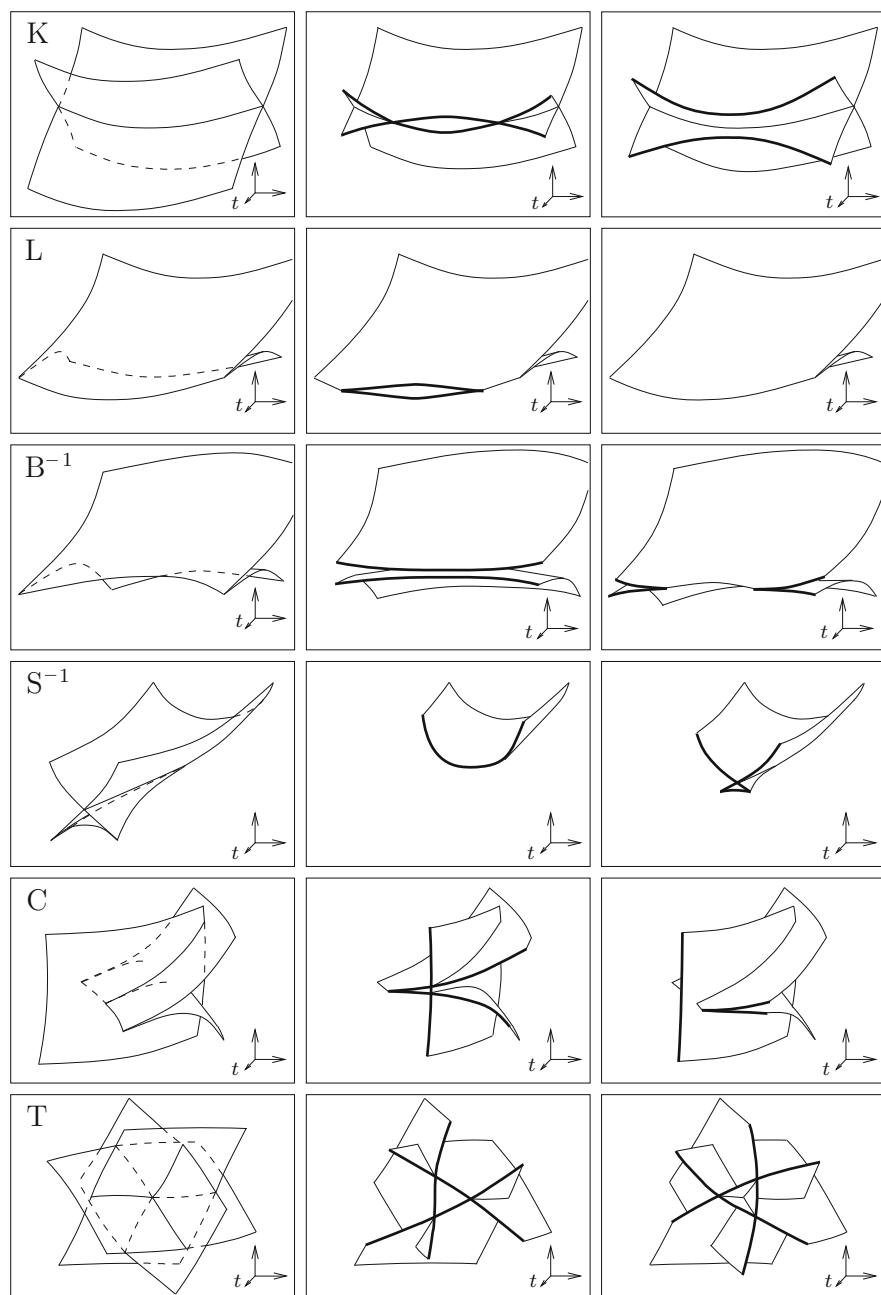


Fig. 6.3 Image taken from [3]. Each row represents the fold surfaces Y_1 near different types of critical points for p : K, L, inverse of B, inverse of S, C, T. The second and third pictures on each row show the slice at $t = \text{const}$ [apparent contour, remember equality (6.4)] before and after the critical time. Compare with Corollary 6.6.2

Definition 6.2.3 (Stratified Morse Function) Let $F \in \text{Stable}(\mathcal{S}, \mathcal{T})$ and let

$$u : \mathcal{T} \rightarrow \mathbb{S}^1$$

be a function of class \mathcal{C}^∞ . We say that u is a stratified Morse function on \mathcal{T} endowed with the stratification $\{Y_j\}_F$ induced by F , if the following three conditions hold:

- for any $j \in \{0, 1, 2\}$ the restriction $u|_{Y_j}$ of u to stratum Y_j is a Morse function,¹⁰ and the set $\text{crit}(u|_{Y_j})$ of its critical points is finite;
- the critical values $\bigcup_{j \in \{0, 1, 2\}} u|_{Y_j}(\text{crit}(u|_{Y_j})) \cup u(Y_3)$ are distinct¹¹;
- if $j \in \{1, 2, 3\}$ and $\bar{\eta} \in Y_j$, then the kernel $\ker(du_{\bar{\eta}})$ of the differential of u at $\bar{\eta}$ does not contain any limit of a sequence of tangent spaces to Y_h at $\eta^{(k)} \in Y_h$, where $0 \leq h < j$ and $\lim_{k \rightarrow +\infty} \eta^{(k)} = \bar{\eta}$.

As we shall see, the stratified Morse function of interest here is the second projection $p : \mathbb{R}^2 \times \mathbb{S}^1 \rightarrow \mathbb{S}^1$. In this case, $\ker(dp_{\bar{\eta}})$ is the plane $\{t = 0\}$ (t the variable in \mathbb{S}^1); this is the frontal plane in Fig. 6.3. The last condition in Definition 6.2.3 therefore means that, at a point of a stratum, the limit tangent spaces coming from the strata with higher dimension are not parallel to the plane $\{t = 0\}$.

6.3 Informal Statement

A rough version of the result we are interested in (Theorem 6.4.6 and its corollaries) can be stated as follows. Let be given two stable scenes E_1 and E_2 ; think, for instance, of $\Sigma_1 := \partial E_1$ as the surface in Fig. 3 of the Introduction, and $\Sigma_2 := \partial E_2$ as the standard round sphere. Consider the corresponding labelled apparent contours, namely the one in Fig. 5.1 of the introduction (for Σ_1), and the round circle (for Σ_2). The statement consists of two parts. Suppose first that we are able to show that the two apparent contours are Reidemeister-equivalent. Then

$$\Sigma_2 = \mathcal{F}(\Sigma_1), \tag{6.3}$$

for some $\mathcal{F} \in \text{Diff}_c(\mathbb{R}^3)$, and therefore Σ_1 and Σ_2 are \mathbb{R}^3 -ambient isotopic (by Theorem 6.0.3). The existence of \mathcal{F} follows by observing that each move can be realized as the composition of a diffeomorphism of \mathbb{R}^3 , which is the identity out of a neighbourhood of the region corresponding to the move, and a fixed projection.

Conversely, suppose that for some $\mathcal{F} \in \text{Diff}_c(\mathbb{R}^3)$ equality (6.3) holds. Then we show that the two apparent contours are Reidemeister-equivalent. The rest of this

¹⁰Hence, the critical points of $u|_{Y_j}$ are nondegenerate.

¹¹By definition, points of Y_3 are considered as critical points of u .

chapter is essentially devoted to the proof of this latter implication, which runs as follows. Let be given an orthogonal projection $\pi : \mathbb{R}^3 = \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}^2$ and a smooth closed surface M . We apply Theorem 6.0.3, where we identify Σ_1 with the surface M , so that $e_1 := \text{id}_{\Sigma_1}$. Calling $e_2 := \mathcal{F} \circ e_1$, from Theorem 6.0.3 we can find an \mathbb{R}^3 -ambient isotopy H as in Definition 6.0.1. In particular, Σ_1 and Σ_2 are isotopic (Definition 6.4.3), via an isotopy

$$\gamma \in \mathcal{C}^\infty(M \times [0, 1], \mathbb{R}^3),$$

a path between an initial embedding ($t = 0$) and a final embedding ($t = 1$) of M in \mathbb{R}^3 , both in general position with respect to π . Since we need to deal with closed manifolds \mathcal{S} , \mathcal{T} , we extend γ in a smooth periodic way to a map defined on $\mathcal{S} := M \times \mathbb{S}^1$. Let us interpret as time the last coordinate $t \in \mathbb{S}^1$, and denote by (m, t) the coordinates of the points of \mathcal{S} . Let us now consider the *level-preserving map*

$$F_\gamma : \mathcal{S} \rightarrow \mathcal{T} := \mathbb{R}^2 \times \mathbb{S}^1,$$

obtained as the composition of the track $(m, t) \in M \times \mathbb{S}^1 \rightarrow (\gamma(m, t), t) \in \mathbb{R}^3 \times \mathbb{S}^1$ of the isotopy γ with the projection $(x, z, t) \in \mathbb{R}^3 \times \mathbb{S}^1 \rightarrow (x, t) \in \mathbb{R}^2 \times \mathbb{S}^1$. Namely,

$$F_\gamma(m, t) = (\varphi_t(m), t),$$

where

$$\varphi_t : M \rightarrow \mathbb{R}^2, \quad \varphi_t(m) := (\gamma_1(m, t), \gamma_2(m, t)).$$

In other words,

$$F_\gamma = (\pi \circ \gamma, \text{id}_{\mathbb{S}^1}).$$

Provided F_γ is *stable*, its critical set $\text{crit}(F_\gamma)$ (Definition 2.1.7) gives a stratification of \mathcal{S} into smooth submanifolds as explained in Sect. 6.2, and, in a natural way, also a stratification $\{Y_0, Y_1, Y_2, Y_3\} = \{Y_j\}_{F_\gamma}$ of \mathcal{T} . The critical value set of F_γ can be written as

$$F_\gamma(\text{crit}(F_\gamma)) = Y_1 \cup Y_2 \cup Y_3,$$

where we recall that Y_1 is the stratum of fold surfaces, Y_2 the stratum of cusp curves and double curves, and Y_3 the discrete stratum of cusp-fold points, swallow's tails and triple points. The family of apparent contours relating the two embeddings then satisfies the useful equality (see Remark 6.4.5):

$$F_\gamma(\text{crit}(F_\gamma)) = \bigcup_{t \in \mathbb{S}^1} (\text{appcon}(\varphi_t) \times \{t\}). \quad (6.4)$$

Let us now consider the second projection $p : \mathcal{T} \rightarrow \mathbb{S}^1$, $p(y, t) = t$. Provided p is a *stratified Morse function* (when \mathcal{T} is endowed with the stratification $\{Y_j\}_{F_\gamma}$), it turns out that Y_3 and the critical points of the restriction $p|_{Y_2}$ determine the complete list of moves on the apparent contours (see Corollary 6.6.2 and Fig. 6.3). Ensuring that p is a stratified Morse function means in particular, in our context, that the tangent spaces to the various strata are *transverse* to the planes $\{t = \text{const}\}$.

One technical point in the proof consists in showing that F_γ can be slightly deformed into a stable map, keeping its level-preserving structure, and then perturbed once more in order to make p stratified: this is the content of Theorem 6.4.6 (based on Proposition 6.5.1, Lemma 6.5.3 and Corollary 6.5.2).

6.4 Rigorous Statement

In this section M denotes a C^∞ two-dimensional closed manifold (hence, not necessarily connected). We recall from [11, p. 177] the following concepts (see also [12, p. 33]).

Definition 6.4.1 (Isotopy) We say that a map $\gamma \in C^\infty(M \times [0, 1], \mathbb{R}^3)$ is an isotopy from M to \mathbb{R}^3 , and we write

$$\gamma \in \text{Isot}(M \times [0, 1], \mathbb{R}^3),$$

if, for any $t \in [0, 1]$, the map $\gamma(\cdot, t) : M \rightarrow \mathbb{R}^3$ is an embedding.

Given $\gamma \in \text{Isot}(M \times [0, 1], \mathbb{R}^3)$ and $t \in [0, 1]$, we often shall write

$$\gamma_t(\cdot) = \gamma(\cdot, t).$$

Definition 6.4.2 (Isotopic Embeddings) Let e_1 and e_2 be two smooth embeddings of M in \mathbb{R}^3 . We say that e_1 and e_2 are isotopic, if

$$\exists \gamma \in \text{Isot}(M \times [0, 1], \mathbb{R}^3) : \quad \gamma_0 = e_1 \quad \text{and} \quad \gamma_1 = e_2. \quad (6.5)$$

Next, we introduce what we mean by an isotopy between two surfaces Σ_1 and Σ_2 .

Definition 6.4.3 (Isotopic Surfaces) Let Σ_1 and Σ_2 be the images of two smooth embeddings of M in \mathbb{R}^3 . We say that Σ_1 and Σ_2 are isotopic if

$$\exists \gamma \in \text{Isot}(M \times [0, 1], \mathbb{R}^3) : \quad \gamma_0(M) = \Sigma_1 \quad \text{and} \quad \gamma_1(M) = \Sigma_2. \quad (6.6)$$

With a small abuse of language, if γ , Σ_1 , Σ_2 are as in (6.6), we say that γ is an isotopy between Σ_1 and Σ_2 , or also that Σ_1 and Σ_2 are isotopic via the isotopy γ .

Notice that if e_1 and e_2 are two ambient isotopic smooth embeddings of M in \mathbb{R}^3 via an \mathbb{R}^3 -ambient isotopy H , then e_1 and e_2 are isotopic, as it follows by defining the isotopy γ as $\gamma_t := H_t \circ e_1$ for any $t \in [0, 1]$.¹²

As already remarked in Sect. 6.3, we shall consider a slight modification of the concept of isotopy, since, in order to apply the results of Sect. 6.2 on singularity theory, we need to consider maps defined on a closed manifold; therefore, we perform the following operations. We first reparametrize the map $\gamma(m, \cdot)$ by composing it with a strictly increasing $C^\infty([0, 1], [0, 1])$ function having vanishing derivatives of all orders at 0 and 1. We still denote by γ this composition and by t the new variable, so that

$$\frac{\partial^k \gamma}{\partial t^k}(m, t)|_{t=0} = \frac{\partial^k \gamma}{\partial t^k}(m, t)|_{t=1} = 0, \quad k \in \mathbb{N}, k \geq 1.$$

We next extend γ on the whole of $M \times \mathbb{R}$ by reflecting it about 0 and 1, resulting in a C^∞ periodic function of period 2 in the variable t . If we identify $\mathbb{R}/[0, 2]$ with \mathbb{S}^1 , we obtain a smooth function, still denoted by γ , defined on the closed smooth manifold $M \times \mathbb{S}^1$ with values in \mathbb{R}^3 . In this way 0 and 1 are two distinct points in the oriented circle \mathbb{S}^1 .

From now on, we set

$$\mathcal{S} := M \times \mathbb{S}^1 \tag{6.7}$$

(the source manifold) and

$$\mathcal{T} := \mathbb{R}^2 \times \mathbb{S}^1 \tag{6.8}$$

(the target manifold). Variables in \mathcal{S} are denoted by

$$(m, t) \quad \text{with } m = (m_1, m_2) \in M \text{ (locally),} \quad \text{and } t \in \mathbb{S}^1.$$

Moreover, we denote by (x, z) a point of $\mathbb{R}^3 = \mathbb{R}^2 \times \mathbb{R}$ where $x = (x_1, x_2) \in \mathbb{R}^2$ and $z \in \mathbb{R}$.

Variables in \mathcal{T} are denoted by

$$(x, t) \quad \text{with } x = (x_1, x_2) \in \mathbb{R}^2, \quad \text{and } t \in \mathbb{S}^1.$$

¹²The converse statement also holds true, as a consequence of the Isotopy Extension Theorem (see, for instance, [11, Theorem 1.3, p. 180], see also [19, pp. 157–201]). Namely, suppose that γ , e_1 and e_2 are as in (6.5). Then, γ induces an isotopy from $e_1(M)$ to \mathbb{R}^3 , which extends to an \mathbb{R}^3 -ambient isotopy with compact support.

Moreover, we denote by

$$p : \mathcal{T} \rightarrow \mathbb{S}^1$$

the projection (that can be considered as a “height” function, or also a “time” function) on the second factor,

$$p(x, t) := t, \quad x \in \mathbb{R}^2, \quad t \in \mathbb{S}^1, \quad (6.9)$$

and we let

$$\pi : \mathbb{R}^3 = \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}^2$$

be the projection on the first factor, defined by $\pi(x, z) := x$.

Definition 6.4.4 (The Level-Preserving Map F_α) Let $\alpha \in \mathcal{C}^\infty(\mathcal{S}, \mathbb{R}^3)$. We associate¹³ with α the map $F_\alpha \in \mathcal{C}^\infty(\mathcal{S}, \mathcal{T})$, defined as

$$F_\alpha(m, t) := (\pi(\alpha(m, t)), t), \quad (m, t) \in \mathcal{S}. \quad (6.10)$$

Remark 6.4.5 Writing γ in components as $\gamma = (\gamma_1, \gamma_2, \gamma_3)$, it is immediately seen that the differential of F_γ has rank 2 at $(\bar{m}, \bar{t}) \in \mathcal{S}$ if and only if the differential of the map

$$\varphi_{\bar{t}} := \pi(\gamma(\cdot, \bar{t})) : m \in M \rightarrow \varphi_{\bar{t}}(m) := (\gamma_1(m, \bar{t}), \gamma_2(m, \bar{t})) \in \mathbb{R}^2$$

has rank one at $\bar{m} \in M$. In particular, formula (6.4) relating the critical value set of F_γ with the apparent contour of φ_t ¹⁴ holds true.

Let Σ_1 and Σ_2 be two isotopic surfaces embedded in \mathbb{R}^3 , and let γ be the isotopy. In order to prove the completeness of the set of moves (Sect. 6.6) we need that $F_\gamma \in \text{Stable}(\mathcal{S}, \mathcal{T})$ and, *at the same time*, that $p : \mathcal{T} \rightarrow \mathbb{S}^1$ is a stratified Morse function on \mathcal{T} endowed with the stratification $\{Y_j\}_{F_\gamma}$. Following a terminology similar to that of [18, Sect. 5], [17, p. 350], in this case we say that F_γ is prepared for moves.

Recalling the definition of general position of an embedded surface with respect to a projection (Definition 3.2.1), we can now state the main result of this chapter.

¹³Notice that the map $\alpha \in \mathcal{C}^\infty(\mathcal{S}, \mathbb{R}^3) \rightarrow F_\alpha \in \mathcal{C}^\infty(\mathcal{S}, \mathcal{T})$ is continuous.

¹⁴Note also that, defining f_{φ_t} as in (2.2), we have $f_{\varphi_t}(x) = \#\{m \in M : F_\gamma(m, t) = (x, t)\}$ for any $(x, t) \in \mathcal{T}$.

Theorem 6.4.6 *Let M be a C^∞ closed two-dimensional manifold. Let \mathcal{S} and \mathcal{T} be defined as in (6.7) and (6.8), respectively. Let $\pi : \mathbb{R}^3 = \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}^2$ be the orthogonal projection, let $e_1, e_2 \in \text{Emb}(M, \mathbb{R}^3)$, and suppose that the two surfaces*

$$\Sigma_1 := e_1(M), \quad \Sigma_2 := e_2(M)$$

are in general position with respect to π . Let

$$\gamma \in \text{Isot}(\mathcal{S}, \mathbb{R}^3)$$

be an isotopy between Σ_1 and Σ_2 . Then, for any neighbourhood $U_{e_j} \subset \text{Emb}(M, \mathbb{R}^3) \subset C^\infty(M, \mathbb{R}^3)$ of e_j , $j = 1, 2$, and for any neighbourhood $U_\gamma \subset C^\infty(\mathcal{S}, \mathbb{R}^3)$ of γ , there exists a map¹⁵

$$\tilde{\gamma} \in U_\gamma \cap \text{Isot}(\mathcal{S}, \mathbb{R}^3)$$

so that $\tilde{\gamma}_0 \in U_{e_1}$, $\tilde{\gamma}_1 \in U_{e_2}$, $\tilde{\Sigma}_1 := \tilde{\gamma}_0(M)$ and $\tilde{\Sigma}_2 := \tilde{\gamma}_1(M)$ are in general position with respect to π , and satisfying the following properties:

$$F_{\tilde{\gamma}} \in \text{Stable}(\mathcal{S}, \mathcal{T}), \tag{6.11}$$

and $p : \mathcal{T} \rightarrow \mathbb{S}^1$ is a stratified Morse function on \mathcal{T} endowed with the stratification $\{Y_j\}_{F_{\tilde{\gamma}}}$ induced by the map $F_{\tilde{\gamma}}$.

The proof of Theorem 6.4.6 is postponed in Sect. 6.5; as we shall see, the critical points of the restriction of p to the codimension two and three strata of a suitable stratification of \mathcal{T} will determine the list of moves.

Remark 6.4.7 (Critical Points of the Restrictions of p) Let $\alpha \in C^\infty(\mathcal{S}, \mathbb{R}^3)$ be such that

$$F_\alpha \in \text{Stable}(\mathcal{S}, \mathcal{T}),$$

and let $\{Y_j\}_{F_\alpha}$ be the stratification of \mathcal{T} induced by F_α . Recalling definition (6.9) of p and the fact that Y_0 is an open set, it is immediate to check that

$$p|_{Y_0} \text{ has no critical points.}$$

Moreover, observing that a point $(x, t) \in Y_1$ is critical for $p|_{Y_1}$ if and only if the tangent plane to Y_1 at (x, t) is parallel to the plane $\{t = 0\}$, we have that

$$p|_{Y_1} \text{ has no critical points.}$$

¹⁵For $t \in [0, 1]$, we use the notation $\tilde{\gamma}_t(\cdot) = \tilde{\gamma}(\cdot, t)$.

Indeed, let $(\bar{m}, \bar{t}) \in X_1$ be such that $F_\alpha(\bar{m}, \bar{t}) \in Y_1$, and let us consider the differential $d(F_{\alpha|X_1})_{(\bar{m}, \bar{t})}$ of the restriction of F_α to X_1 , at (\bar{m}, \bar{t}) . Then $d(F_{\alpha|X_1})_{(\bar{m}, \bar{t})}(X_1)$ coincides with the tangent space to Y_1 at $F_\alpha(\bar{m}, \bar{t})$, and $d(F_{\alpha|X_1})_{(\bar{m}, \bar{t})}(X_1) \subseteq (dF_\alpha)_{(\bar{m}, \bar{t})}(X_1)$. But the definition of X_1 implies that $d(F_{\alpha|X_1})_{(\bar{m}, \bar{t})}(X_1)$ is two-dimensional. On the other hand, the rank of the differential of F_α at $(\bar{m}, \bar{t}) \in \mathcal{S}$ is two and therefore also $(dF_\alpha)_{(\bar{m}, \bar{t})}(X_1)$ is two-dimensional, and we deduce $d(F_{\alpha|X_1})_{(\bar{m}, \bar{t})}(X_1) = (dF_\alpha)_{(\bar{m}, \bar{t})}(X_1)$. Thus, the tangent space to Y_1 at $F_\alpha(\bar{m}, \bar{t})$ coincides with $(dF_\alpha)_{(\bar{m}, \bar{t})}(X_1)$. The last column of the matrix of $(dF_\alpha)_{(\bar{m}, \bar{t})}$ is $(0, 0, 1)$: this implies that the tangent space to Y_1 at $F_\alpha(\bar{m}, \bar{t})$ contains a vector of the form $(c_1, c_2, 1)$ for some $c_1, c_2 \in \mathbb{R}$. In particular the tangent plane to Y_1 at $F_\alpha(\bar{m}, \bar{t})$ is transverse to $\{t = \bar{t}\}$, and this holds uniformly with respect to the points in Y_1 .

Remark 6.4.8 Let $(\bar{x}, \bar{t}) \in Y_2 \cup Y_3$ be a point which is limit of a sequence of points $(x^{(k)}, t^{(k)}) \in Y_1$ as $k \rightarrow +\infty$. Assume that the limit T of the sequence of tangent planes to Y_1 at $(x^{(k)}, t^{(k)}) \in Y_1$ exists. Then, by Remark 6.4.7 and by continuity, still $(c_1, c_2, 1)$ is one of the two vectors spanning T . Therefore T is transverse to $\{t = \bar{t}\}$ at (\bar{x}, \bar{t}) .

It may happen that a curve in Y_2 having an endpoint in Y_3 has there a tangent line contained in a plane parallel to $\{t = 0\}$, as we now show; this kind of examples motivate Lemma 6.5.3.

Example 6.4.9 The function p could not be a stratified Morse function on \mathcal{S} , endowed with the stratification $\{Y_j\}_{F_y}$, since the third condition of Definition 6.2.3 may fail. Indeed, we can construct a map $\alpha \in \mathcal{C}^\infty(\mathcal{S}, \mathbb{R}^3)$ with $F_\alpha \in \text{Stable}(\mathcal{S}, \mathcal{T})$, having a triple point at $(\bar{x}, \bar{t}) = (0, 0) \in Y_3^3$, with one of the double curves in Y_2^2 parallel to $\{t = 0\}$. For instance, it is enough to consider a map $F_\alpha \in \text{Stable}(\mathcal{S}, \mathcal{T})$ having, locally around $(0, 0)$, the fold surfaces of the form $\{x_1 = \pm t\}$ and $\{x_2 = 0\}$. These folds are obviously mutually transverse, $\{x_1 = t = 0\}$ is locally one of the double curves and it is parallel to the plane $\{t = 0\}$.

Remark 6.4.10 (Swallow's Tail Singularity) Up to a change of variables in \mathcal{S} and \mathcal{T} , a swallow's tail singularity at the origin has the local description¹⁶

$$\eta_1 = \xi_1 \xi_2 + \xi_1^2 \xi_3 + \xi_1^4, \quad \eta_2 = \xi_2, \quad \eta_3 = \xi_3.$$

There are two cusp curves and one double curve originating at the singularity with a common tangent vector $(0, 0, -1)$; moreover, at the singularity, all fold surfaces are locally tangent to the plane $\{\eta_1 = 0\}$. We can provide two simple realizations in our context of the canonical representation above. The choice

$$m = (\xi_1, \xi_2), \quad x = (\eta_1, \eta_2), \quad t = \xi_3 = \eta_3,$$

¹⁶See, for instance, [8, p. 176, 177], and more generally [13, 14].

corresponds to the move S, whereas the choice

$$m = (\xi_1, \xi_3), \quad x = (\eta_1, \eta_3), \quad t = \xi_2 = \eta_2,$$

(whence $\pi(\alpha(m, t)) = (tm_1 + m_2m_1^2 + m_1^4, m_2)$) corresponds to an evolution that is degenerate at $t = 0$: indeed, the corresponding apparent contour has a cusp with one of the two departing arcs that is (locally) completely contained in another arc of the contour.

6.5 Proof of the Completeness Theorem

Following closely [3], we split the proof of Theorem 6.4.6 into various steps. The first step concerns the map $\alpha \rightarrow F_\alpha$: essentially, it says that given a neighbourhood N_α of α , we can find a neighbourhood V_{F_α} of F_α , such that any $G \in V_{F_\alpha}$ is equivalent to a map having the third component equal to t .

Proposition 6.5.1 (On Level-Preserving Paths) *Let $\alpha \in C^\infty(\mathcal{S}, \mathbb{R}^3)$. For any neighbourhood N_α of α in $C^\infty(\mathcal{S}, \mathbb{R}^3)$, there exists a neighbourhood V_{F_α} of F_α in $C^\infty(\mathcal{S}, \mathcal{T})$ such that the following property holds:*

$$\forall G \in V_{F_\alpha} \quad \exists \bar{\alpha} \in N_\alpha : G \circ \chi = F_{\bar{\alpha}} \text{ for some } \chi \in \text{Diff}(\mathcal{S}). \quad (6.12)$$

In particular,

$$F_{\bar{\alpha}} \text{ is equivalent to } G.$$

Proof Let $N_\alpha \subset C^\infty(\mathcal{S}, \mathbb{R}^3)$ be a neighbourhood of α . If we choose a sufficiently small neighbourhood V_{F_α} of F_α , we can ensure that any $G \in V_{F_\alpha}$, that we write componentwise as

$$G = (G_1, G_2, G_3) : \mathcal{S} = M \times \mathbb{S}^1 \rightarrow \mathbb{R}^2 \times \mathbb{S}^1 = \mathcal{T},$$

has the following property: for any $m \in M$, the function

$$t \in \mathbb{S}^1 \rightarrow G_3^m(t) := G_3(m, t) \in \mathbb{S}^1$$

is close to the identity in $C^\infty(\mathbb{S}^1, \mathbb{S}^1)$, and therefore it is invertible. Let us denote by

$$g^m : \mathbb{S}^1 \rightarrow \mathbb{S}^1$$

the inverse of G_3^m , so that

$$g^m(G_3^m) = \text{id}_{\mathbb{S}^1} \quad \text{and} \quad G_3^m(g^m) = \text{id}_{\mathbb{S}^1}.$$

Notice that the map $(m, t) \in \mathcal{S} \rightarrow g^m(t) \in \mathbb{S}^1$ is smooth. Denoting by

$$\alpha_z$$

the third component of α , we define $\bar{\alpha} : \mathcal{S} \rightarrow \mathbb{R}^3$ as follows:

$$\bar{\alpha}(m, t) := \left(G_1(m, g^m(t)), G_2(m, g^m(t)), \alpha_z(m, t) \right), \quad (m, t) \in \mathcal{S}.$$

Since $\bar{\alpha}$ depends continuously on G , possibly reducing V_{F_α} , we can ensure that

$$\bar{\alpha} \in N_\alpha.$$

By definition,

$$F_{\bar{\alpha}}(m, t) = \left(G_1(m, g^m(t)), G_2(m, g^m(t)), t \right), \quad (m, t) \in \mathcal{S}.$$

Now, if we consider the map $\chi : (m, t) \in \mathcal{S} \rightarrow \chi(m, t) := (m, g^m(t)) \in \mathcal{S}$, we have

$$\chi \in \text{Diff}(\mathcal{S}).$$

Since $G \circ \chi = F_{\bar{\alpha}} = \text{id}_{\mathcal{S}} \circ F_{\bar{\alpha}}$, the thesis of the proposition follows. \square

Corollary 6.5.2 *Let $\alpha \in \mathcal{C}^\infty(\mathcal{S}, \mathbb{R}^3)$. For any neighbourhood N_α of α in $\mathcal{C}^\infty(\mathcal{S}, \mathbb{R}^3)$, there exists $\bar{\alpha} \in N_\alpha$ such that*

$$F_{\bar{\alpha}} \in \text{Stable}(\mathcal{S}, \mathcal{T}). \quad (6.13)$$

Proof Let V_{F_α} and $\bar{\alpha}$ be as in Proposition 6.5.1. Since $\text{Stable}(\mathcal{S}, \mathcal{T})$ is dense in $\mathcal{C}^\infty(\mathcal{S}, \mathcal{T})$, it follows that $V_{F_\alpha} \cap \text{Stable}(\mathcal{S}, \mathcal{T})$ is nonempty. Therefore, choosing G in (6.12) with the further property that $G \in \text{Stable}(\mathcal{S}, \mathcal{T})$, we deduce that $F_{\bar{\alpha}}$ is equivalent to a map in $\text{Stable}(\mathcal{S}, \mathcal{T})$. As a consequence, (6.13) holds true. \square

In the next lemma¹⁷ we perform a further perturbation of a stable map F_β , in order to get a new map which is prepared for moves.

Lemma 6.5.3 (Existence of Stable $F_{\hat{\beta}}$ and Stratified p) *Let $\beta \in \mathcal{C}^\infty(\mathcal{S}, \mathbb{R}^3)$ be such that $F_\beta \in \text{Stable}(\mathcal{S}, \mathcal{T})$. For any neighbourhood W_β of β in $\mathcal{C}^\infty(\mathcal{S}, \mathbb{R}^3)$,*

$$\exists \hat{\beta} \in W_\beta \quad \text{such that} \quad F_{\hat{\beta}} \in \text{Stable}(\mathcal{S}, \mathcal{T}), \quad (6.14)$$

¹⁷See [18, Proposition 5.4] for related problems.

and

$$p : \mathcal{T} \rightarrow \mathbb{S}^1 \text{ is a stratified Morse function on } \mathcal{T} \text{ endowed with } \{Y_j\}_{F_\beta}. \quad (6.15)$$

Proof Since $F_\beta \in \text{Stable}(\mathcal{S}, \mathcal{T})$, from Remark 6.4.7 it follows that $p|_{Y_0(F_\beta)}$ and $p|_{Y_1(F_\beta)}$ have no critical points. Example 6.4.9 shows, however, that p is not necessarily a stratified Morse function on \mathcal{T} endowed with $\{Y_j\}_{F_\beta}$. Therefore, we need to slightly perturb F_β (see (6.19) and (6.23), below), into a new map $F_{\hat{\beta}}$ in order to obtain (6.15).

Let W_β be a neighbourhood of β in $\mathcal{C}^\infty(\mathcal{S}, \mathbb{R}^3)$. Recalling Definition 6.2.3, in order to prove the thesis, we have to show that (6.14) holds, and the function $p : \mathcal{T} \rightarrow \mathbb{R}$ satisfies the following three properties:

- (1) if $(\bar{x}, \bar{t}) \in Y_3(F_{\hat{\beta}})$, then all curves in $Y_2(F_{\hat{\beta}})$ having (\bar{x}, \bar{t}) as an endpoint cannot have a limit tangent line at (\bar{x}, \bar{t}) contained in the plane $\{t = \bar{t}\}$;
- (2) the critical points of $p|_{Y_2(F_{\hat{\beta}})}$ are nondegenerate;
- (3) the critical values of $p|_{Y_2(F_{\hat{\beta}})}$ are distinct and distinct from $p(Y_3(F_{\hat{\beta}}))$, in turn consisting of distinct points of \mathbb{S}^1 .

Let us consider the stratification $\{Y_j\}_{F_\beta}$ induced by F_β , and let $(\bar{x}, \bar{t}) \in Y_3(F_\beta)$. Note that, if $(\bar{x}, \bar{t}) \in Y_3^3(F_\beta)$, then only one of the three double curves meeting at (\bar{x}, \bar{t}) may have a limit tangent line contained in $\{t = \bar{t}\}$; indeed, if two of them share this property, then there is a fold surface in $Y_1(F_\beta)$ having the tangent plane at (\bar{x}, \bar{t}) parallel to $\{t = \bar{t}\}$, which is in contradiction with Remark 6.4.7. Recall also that, if $(\bar{x}, \bar{t}) \in Y_3^1(F_\beta)$ is a swallow's tail, then the two cusp curves and the double curve meeting at (\bar{x}, \bar{t}) have the same tangent vector there.

Firstly, we want to show that we can achieve condition (1). Assume that there is a curve contained in $Y_2(F_\beta) \cup \{(\bar{x}, \bar{t})\}$ having (\bar{x}, \bar{t}) as an endpoint, and with a limit tangent line at (\bar{x}, \bar{t}) contained in $\{t = \bar{t}\}$. Let $\lambda = (\lambda_1, \lambda_2, \lambda_3) \in \mathcal{C}^\infty([0, 1], \mathcal{T})$ be a regular parametrization of such a curve, having (\bar{x}, \bar{t}) as initial point, so that

$$\lambda(0) = (\bar{x}, \bar{t}) \quad \text{and} \quad \lambda'_3(0) = 0.$$

Let us select a function $a \in \mathcal{C}^\infty(\mathcal{T})$ with compact support, and satisfying

$$a(\bar{x}, \bar{t}) = 0, \quad (6.16)$$

$$\frac{d}{d\sigma} a(\lambda(\sigma))|_{\sigma=0} \neq 0. \quad (6.17)$$

Let

- $\mathcal{O} \subset \mathcal{T}$ be a neighbourhood of (\bar{x}, \bar{t}) small enough so that

$$\overline{\mathcal{O}} \cap (Y_3(F_\beta) \setminus \{(\bar{x}, \bar{t})\}) = \emptyset;$$

- $\varrho \in \mathcal{C}^\infty(\mathcal{T})$ be a nonnegative function, having support contained in \mathcal{O} , and which is constantly equal to one in a small neighbourhood of (\bar{x}, \bar{t}) ;
- $\varepsilon \in \mathbb{R}$.

Define the function $j : \mathcal{T} \rightarrow \mathbb{R}$ as

$$j(x, t) := t + \varepsilon \varrho(x, t) a(x, t), \quad (x, t) \in \mathcal{T}.$$

Then, $j \in \mathcal{C}^\infty(\mathcal{T})$ and, provided $|\varepsilon|$ is sufficiently small, we have that for any $x \in \mathbb{R}^2$ the function $t \in \mathbb{S}^1 \rightarrow j(x, t) \in \mathbb{S}^1$ is invertible. Hence, if we consider the map $\Psi : \mathcal{T} \rightarrow \mathcal{T}$ defined by

$$\Psi := (\text{id}_{\mathbb{R}^2}, j),$$

we have

$$\Psi \in \text{Diff}(\mathcal{T}). \quad (6.18)$$

Define $G : \mathcal{S} \rightarrow \mathcal{T}$ as

$$G := \Psi \circ F_\beta. \quad (6.19)$$

Thus, G is equivalent to F_β , and therefore, since by assumption $F_\beta \in \text{Stable}(\mathcal{S}, \mathcal{T})$, we also have

$$G \in \text{Stable}(\mathcal{S}, \mathcal{T}).$$

Now, let us consider the stratification

$$\{Y_0(G), Y_1(G), Y_2(G), Y_3(G)\} \quad (6.20)$$

induced on \mathcal{T} by G : by (6.18) and (6.19) it follows that such a stratification is the image through Ψ of the stratification induced by F_β . Equality (6.16) implies that $(\bar{x}, \bar{t}) \in Y_3(G)$; moreover $\Psi(\lambda([0, 1])) \subset Y_2(G) \cup \{(\bar{y}, \bar{t})\}$ is regularly parametrized in a neighbourhood of (\bar{x}, \bar{t}) by $\sigma \in [0, 1] \rightarrow \Psi(\lambda(\sigma)) = (\lambda_1(\sigma), \lambda_2(\sigma), j(\lambda(\sigma)))$. Since

$$\frac{d}{d\sigma} j(\lambda(\sigma))|_{\sigma=0} = \lambda'_3(0) + \varepsilon \frac{d}{d\sigma} a(\lambda(\sigma))|_{\sigma=0} = \varepsilon \frac{d}{d\sigma} a(\lambda(\sigma))|_{\sigma=0}, \quad (6.21)$$

equality (6.17) guarantees that the right-hand side of (6.21) is nonzero. It follows that assertion (1) is satisfied for the stratification (6.20); namely, if $(\bar{x}, \bar{t}) \in Y_3(G)$, then all curves in $Y_2(G)$ having (\bar{x}, \bar{t}) as an endpoint cannot have a limit tangent line at (\bar{x}, \bar{t}) contained in the plane $\{t = \bar{t}\}$.

Applying Proposition 6.5.1 with the choice

$$\alpha = \beta, \quad N_\alpha = W_\beta,$$

we obtain a corresponding neighbourhood V_{F_β} of F_β . If $|\varepsilon|$ is sufficiently small, we have

$$G \in V_{F_\beta}.$$

From (6.19) it follows

$$G(m, t) = (\pi \circ \beta(m, t), i^m(t)), \quad (m, t) \in \mathcal{S},$$

where

$$i^m(t) := t + \varepsilon \chi(F_\beta(m, t)) a(F_\beta(m, t)), \quad (m, t) \in \mathcal{S}.$$

Denote by $\chi : \mathcal{S} \rightarrow \mathcal{S}$ the map defined by $\chi(m, t) := (m, (i^m)^{-1}(t))$. Then

$$\chi \in \text{Diff}(\mathcal{S}), \quad (6.22)$$

and

$$G \circ \chi \in V_{F_\beta}.$$

From (6.12), it follows that

$$\exists \hat{\beta} \in W_\beta \text{ such that } G \circ \chi = F_{\hat{\beta}}. \quad (6.23)$$

Since $G \in \text{Stable}(\mathcal{S}, \mathcal{T})$, we deduce that

$$F_{\hat{\beta}} \in \text{Stable}(\mathcal{S}, \mathcal{T}).$$

Moreover, (6.22) and (6.23) imply that the stratification of \mathcal{S} associated with G is the image through χ of the stratification associated with F_β , and that the stratification of \mathcal{T} induced by G coincides with the stratification induced by $F_{\hat{\beta}}$. We conclude therefore that the stratification induced by $F_{\hat{\beta}}$ satisfies condition (1).

Now, if we replace the function a in the previous argument with the function $a + b$, where $b : \mathcal{T} \rightarrow \mathbb{R}$ is a smooth function satisfying

$$b(\bar{x}, \bar{t}) \neq 0, \quad db_{(\bar{x}, \bar{t})} = 0,$$

we obtain that $p(Y_3)$ consists of distinct points of \mathbb{S}^1 .

Considering the stratification of \mathcal{T} induced by $F_{\hat{\beta}}$, from condition (1) we deduce that $p|_{Y_2(F_{\hat{\beta}})}$ has no critical points on the boundary points of $Y_2(F_{\hat{\beta}})$. Therefore, being all critical points of $p|_{Y_2(F_{\hat{\beta}})}$ interior to any arc of $Y_2(F_{\hat{\beta}})$, we can argue using one-dimensional Morse theory, and obtain assertion (2). Also assertion (3) follows in a standard way. \square

We are now in the position to prove Theorem 6.4.6.

For $j = 0, 1$, let $U_{e_j} \subset \mathcal{C}^\infty(M, \mathbb{R}^3)$ be a neighbourhood of e_j . Let $U_\gamma \subset \mathcal{C}^\infty(\mathcal{S}, \mathbb{R}^3)$ be a neighbourhood of γ . Let us apply Corollary 6.5.2 to $\alpha = \gamma$ and with the choice of a neighbourhood N_α of α satisfying

$$N_\alpha \subseteq U_\gamma.$$

We obtain a corresponding map $\bar{\gamma} \in N_\alpha$ such that

$$F_{\bar{\gamma}} \in \text{Stable}(\mathcal{S}, \mathcal{T}). \quad (6.24)$$

Possibly reducing N_α , we can assume that

$$\bar{\gamma}_0 \in U_{e_0} \quad \text{and} \quad \bar{\gamma}_1 \in U_{e_1}.$$

Since (6.24) holds, we can now apply Lemma 6.5.3 to $\beta = \bar{\gamma}$, taking as W_β a neighbourhood of $\bar{\gamma}$ satisfying $W_\beta \subseteq N_\alpha$. We obtain a corresponding map $\hat{\beta} \in W_\beta$. If we set

$$\tilde{\gamma} := \hat{\beta},$$

it follows that

$$\tilde{\gamma} \in W_\beta \subseteq N_\alpha \subseteq U_\gamma,$$

and $\tilde{\gamma}$ satisfies (6.11) and the (last) assertion of the statement of the theorem, concerning the stratifiability of p . Moreover $\tilde{\gamma}_0 \in U_{e_0}$ and $\tilde{\gamma}_1 \in U_{e_1}$.

Since $\gamma_t \in \text{Emb}(M, \mathbb{R}^3)$ and $\text{Emb}(M, \mathbb{R}^3)$ is open in $\mathcal{C}^\infty(M, \mathbb{R}^3)$, it follows, possibly reducing W_β , that $\tilde{\gamma}_0(M)$ and $\tilde{\gamma}_1(M)$ are isotopic (Definition 6.4.2). Since by hypothesis Σ_1 and Σ_2 are in general position with respect to π , we have that $\pi|_{\Sigma_1} \in \text{Stable}(\Sigma_1, \mathbb{R}^2)$ and $\pi|_{\Sigma_2} \in \text{Stable}(\Sigma_2, \mathbb{R}^2)$. Whence, possibly reducing W_β , and recalling that $\text{Stable}(\Sigma_j, \mathbb{R}^2)$ is open in $\mathcal{C}^\infty(\Sigma_j, \mathbb{R}^2)$ for $j = 1, 2$, it follows that

$$\pi|_{\tilde{\gamma}_0(M)} \in \text{Stable}(\Sigma_1, \mathbb{R}^2) \quad \text{and} \quad \pi|_{\tilde{\gamma}_1(M)} \in \text{Stable}(\Sigma_2, \mathbb{R}^2).$$

The proof of the theorem is concluded. \square

6.6 Completeness of Moves

Let π , e_1 , e_2 , $\tilde{\gamma}$, \mathcal{T} , $\{Y_j\}_{F_{\tilde{\gamma}}}$ and p be as in Theorem 6.4.6. By stability and compactness, the set $\text{crit}(p|_{Y_2})$ of critical points of $p|_{Y_2}$ is finite. Since also Y_3 consists of isolated points, it follows that $p|_{Y_2}(\text{crit}(p|_{Y_2})) \cup p(Y_3)$ is a *finite* set of points of \mathbb{S}^1 .

Definition 6.6.1 (Critical Times) We call

$$p|_{Y_2}(\text{crit}(p|_{Y_2})) \cup p(Y_3)$$

the set of critical times.

If t_0 is not a critical time, the apparent contour (given by a time-slice, i.e., the transversal intersection of $Y_1 \cup Y_2 \cup Y_3$ with $\{t = t_0\}$) varies smoothly, and its topology does not change. Hence, if $t_1 < t_2$ are such that the time interval $[t_1, t_2]$ does not contain any critical time, we can find a smooth path $[t_1, t_2] \rightarrow \text{Diff}_c(\mathbb{R}^2)$ of diffeomorphisms of \mathbb{R}^2 compactly supported in a fixed bounded set, connecting the apparent contours at times t_1 and t_2 . Moreover, in view of the *classification of singularities of stable mappings between three-manifolds* (Sect. 6.2.1), we obtain the following result (compare with Fig. 6.3; in that figure, notice carefully the direction of t , with respect to which local maxima and minima are considered).

Corollary 6.6.2 (Completeness) *A point $(x, t) \in \text{crit}(p|_{Y_2}) \cup Y_3$ lies in one of the following classes, each determining a move in the list of Sect. 6.1:*

- $(x, t) \in Y_2^2$ is a local maximum (respectively local minimum) of a double curve: moves of type K (respectively of type K^{-1}).
- $(x, t) \in Y_2^1$ is a local maximum (respectively local minimum) of a cusp curve that bounds folds going downwards (respectively upwards): move L (respectively move L^{-1}).
- $(x, t) \in Y_2^1$ is a local maximum (respectively local minimum) of a cusp curve that bounds folds going upwards (respectively downwards): move B (respectively move B^{-1}).
- $(x, t) \in Y_3^1$: moves of type S .
- $(x, t) \in Y_3^2$: moves of type C .
- $(x, t) \in Y_3^3$: moves of type T .

The precise one-to-one correspondence between the list of moves of Sect. 6.1 and the list in Corollary 6.6.2 is provided by the following observation.

Remark 6.6.3 A more refined classification of each move can be obtained by looking at the orientation of the various folds involved, and at the relative depth of the preimages in the critical set $\bar{X}_1 = S_1(F_{\tilde{\gamma}})$ (see Sect. 6.2.1) with respect to the regular preimages. For example, a point in $Y_3^3(F_{\tilde{\gamma}})$ has three distinct preimages in $X_1(F_{\tilde{\gamma}})$, which can be ordered according to the z -coordinate (dropped by the

projection π). Each of the three involved folds (which are transversal to “time” direction t) carries a natural orientation and hence contributes with a sign. The relative depth of the three singular preimages with respect to the remaining regular preimages provides for the three nonnegative integer parameters, as explained in Sect. 6.1.3. Note that a cusp curve forces the orientation of the two adjacent folds.

Remark 6.6.4 Corollary 6.6.2 is consistent with [4, Theorem 3.5.5], where isotopies of surfaces Γ immersed in \mathbb{R}^3 , realized as projections of embedded surfaces in \mathbb{R}^4 , possibly with pinch points, are considered. The authors of [4] seek all codimension one singularities of maps $\mathbb{R}^3 \supset \Gamma \rightarrow \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$. The majority of these singularities are related to self-intersections and pinch points of Γ , which are excluded in our setting. Moreover, they also consider singularities that arise from the presence of a height function $\mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$, which we do not need here. The remaining codimension one singularities correspond to those listed in Corollary 6.6.2.¹⁸

Corollary 6.6.5 *Let $\pi : \mathbb{R}^3 = \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}^2 \times \{0\}$ be the projection on the first factor; let $e_1, e_2 \in \text{Emb}(M, \mathbb{R}^3)$, and suppose that the two surfaces*

$$\Sigma_1 := e_1(M), \quad \Sigma_2 := e_2(M)$$

are in general position with respect to π . Let $\varphi_1 := \pi \circ e_1$ and $\varphi_2 := \pi \circ e_2$. Then Σ_1 and Σ_2 are isotopic if and only if their apparent contours

$$\text{appcon}(\varphi_1), \text{appcon}(\varphi_2) \subset \mathbb{R}^2$$

are Reidemeister-equivalent.

Remark 6.6.6 (Reidemeister Moves on Knots) A proof similar to the one described in Sects. 6.5 and 6.6 can be used to show the classical result of Reidemeister, on the completeness of the three Reidemeister moves for knots and links. The proof is obtained by replacing M with the disjoint union \mathcal{U} of a finite number of copies of \mathbb{S}^1 , and by recalling the classification of singularities of a map from $\mathcal{U} \times \mathbb{S}^1$ to $\mathbb{R}^2 \times \mathbb{S}^1$. More specifically, the second and third Reidemeister moves correspond, respectively, to the double curves and to the triple points, while the first Reidemeister move corresponds to the Whitney umbrella (see Example 2.1.9).

¹⁸To be more specific, [4, Fig. 9(row 1, column 1)] corresponds to move L; [4, Fig. 9(2,1)] corresponds to move B; [4, Fig. 9(3,1)] corresponds to move S; [4, Fig. 10(1,1)] corresponds to move K; [4, Fig. 10(1,2)] corresponds to move C; [4, Fig. 10(2,1)] corresponds to move T. Theorem [4, 3.2.3] is a simpler version of [4, Theorem 3.5.5], in which the height function is not considered, and hence is more close to Corollary 6.6.2. It follows by combining the classifications given by [10, 16, 21].

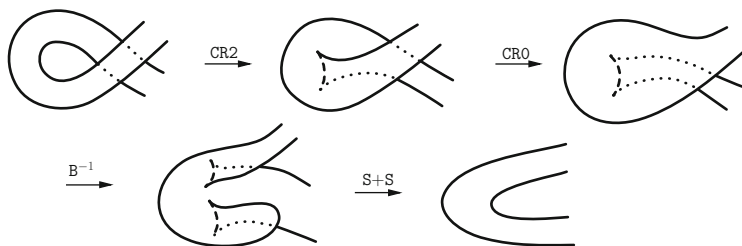


Fig. 6.4 The first Reidemeister move is obtained with a suitable sequence of compositions of the moves for apparent contours, by considering the apparent contour of a thin tubular neighbourhood of the knot (or link); $d = 1$ on the *dashed* curves, and $d = 2$ on the *dotted* ones

Considering a sufficiently thin tubular neighbourhood of the knot (or link), the second and third Reidemeister moves can be obtained, respectively, by means of an iterated application of the K and T move. The first Reidemeister move, instead, is obtained with the sequence of composite rules displayed in Fig. 6.4; composite rules are defined in Sect. 10.4.3, and are extensively used by the program `appcontour`.

References

1. Arnold, V.I., Gusein-Zade, S.M., Varchenko, A.N.: *Singularities of Differentiable Maps*, vol. I. Birkhäuser, Boston (1985)
2. Arnold, V.I., Goryunov, V.V., Lyashko, O.V., Vassiliev, V.A. (eds.): *Dynamical Systems VIII. Singularity Theory. II. Applications*. Springer, Berlin (1993)
3. Bellettini, G., Beorchia, V., Paolini, M.: Completeness of Reidemeister-type moves for surfaces embedded in three-dimensional space. *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl.* **22**, 1–19 (2011)
4. Carter, J.S., Rieger, J.H., Saito, M.: A combinatorial description of knotted surfaces and their isotopies. *Adv. Math.* **127**, 1–51 (1997)
5. Crowell, R.H., Fox, R.H.: *Introduction to Knot Theory*. Springer, New York (1977)
6. Favaro, L.A., Mendes, C.M.: Global stability for diagrams of differentiable applications. *Ann. Inst. Fourier (Grenoble)* **36**, 133–153 (1986)
7. Fisher, G.M.: On the group of all homeomorphisms of a manifold. *Trans. Am. Math. Soc.* **97**, 193–212 (1960)
8. Golubitsky, M., Guillemin, V.: *Stable Mappings and Their Singularities*. Graduate Texts in Mathematics, vol. 14. Springer, New York/Heidelberg/Berlin (1974)
9. Goresky, M., MacPherson, R.: *Stratified Morse Theory*. Springer, Berlin (1988)
10. Goryunov, V.V.: Monodromy of the image of a mapping. *Funct. Anal. Appl.* **25**, 174–180 (1991)
11. Hirsch, M.W.: *Differential Topology*. Springer, Academic Press, New York (1976)
12. Kosinski, A.A.: *Differential Manifolds*. Academic, New York (1993)
13. Morin, B.: Formes canoniques des singularités d’une application différentiable. *C. R. Acad. Sci. Paris* **260**, 5662–5665 (1965)
14. Morin, B.: Formes canoniques des singularités d’une application différentiable. *C. R. Acad. Sci. Paris* **260**, 6503–6506 (1965)

15. Pignoni, R.: Density and stability of Morse functions on a stratified space. *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* **6**, 593–608 (1979)
16. Rieger, J.H.: On the complexity and computation of view graph of piecewise smooth algebraic surfaces. *Phil. Trans. R. Soc. Lond. A* **354**, 1899–1940 (1996)
17. Roseman, D.: Reidemeister-type moves for surfaces in four dimensional space. In: Jones, V.F.R., et al. (eds.) *Knot Theory. Proceedings of the Mini-Semester, Warsaw, Poland 1995*, vol. 42, pp. 347–380. Banach Center Publications, Warsaw (1998)
18. Roseman, D.: Elementary moves for higher dimensional knots. *Fundam. Math.* **184**, 291–310 (2004)
19. Thom, R.: La classifications des immersions. *Seminaire Bourbaki*, 157-01–157-11 (1957)
20. Thom, R.: Ensembles and morphismes stratifiés. *Bull. Am. Math. Soc.* **70**, 240–284 (1969)
21. West, J.M.: The differential geometry of the crosscap. Ph.D. thesis, University of Liverpool, England (1995)
22. Whitney, H.: Elementary structures of real algebraic varieties. *Ann. Math.* **66**, 545–556 (1957)
23. Whitney, H.: Local properties of analytic varieties. In: Cairns, S.S. (ed.) *Differential and Combinatorial Topology*, pp. 205–244. Princeton University Press, Princeton (1965)
24. Whitney, H.: Tangents to an analytic variety. *Ann. Math.* **81**, 496–549 (1965)

Chapter 7

Invariants of an Apparent Contour

The aim of this chapter is to illustrate some interesting invariants of apparent contours and labelled apparent contours. These invariants can be numbers, groups, polynomials; invariance here means that they are insensitive to certain transformations, that will be specified case by case.

As an example, let us consider an apparent contour $\text{appcon}(\varphi) \subset \mathbb{R}^2$ of a stable map $\varphi : M \rightarrow \mathbb{R}^2$ (Definition 2.2.1). Then, it is clear that the number of components of $\text{appcon}(\varphi)$ is invariant under \mathbb{R}^2 -ambient isotopies. Moreover, the number of crossings of $\text{appcon}(\varphi)$ is

- invariant under diffeomorphic equivalence (Definition 2.4.2),
- invariant under the action of $\text{Diff}(M)$.

On the other hand, supposing that φ factorizes as an embedding in \mathbb{R}^3 and a projection on \mathbb{R}^2 (formula (3.3)), then the number of crossings of $\text{appcon}(\varphi)$ is not invariant under the action of a diffeomorphism of \mathbb{R}^3 .

The number of crossings is one of the three linearly independent first order Vassiliev-type invariants [23, 45], [29, Chapter 12], [14] in the sense¹ considered in [33], the other two being the number of cusps and a Bennequin-type invariant [10], which will be described later on in this chapter. When $\text{appcon}(\varphi)$ is a labelled apparent contour (Example 4.2.8), it is worthwhile to recall that the number of crossings is not invariant under the Reidemeister-type moves described in Chap. 6.

When M is orientable, we can distinguish positive and negative cusps of $\text{appcon}(\varphi)$ (Definition 2.2.12), the number of which is easily seen to be, separately, invariant under the action of $\text{Diff}_c(\mathbb{R}^2)$.

¹By invariant, the authors of [33] mean a locally constant function on the set of stable mappings; see Sect. 7.3. See also [23, 32] and the references therein. The invariants considered in [33] turn out to be also invariants under diffeomorphic equivalence (Definition 2.4.2).

7.1 Definition of $\mathfrak{B}(\text{appcon}(\varphi))$

Let M be a two-dimensional closed (not necessarily connected) manifold of class \mathcal{C}^∞ , and let $\varphi : M \rightarrow \mathbb{R}^2$ be a stable map. In this section, following [9],² we associate with $\text{appcon}(\varphi)$ a number, denoted by $\mathfrak{B}(\text{appcon}(\varphi))$, computable only in terms of the apparent contour.³ It turns out that $\mathfrak{B}(\text{appcon}(\varphi))$ is invariant under the action of $\text{Diff}_c(\mathbb{R}^2)$; hence, if $\text{appcon}(\varphi_1)$ and $\text{appcon}(\varphi_2)$ are diffeomorphically equivalent (Definition 2.4.2), then

$$\mathfrak{B}(\text{appcon}(\varphi_1)) = \mathfrak{B}(\text{appcon}(\varphi_2)).$$

In Sect. 7.3 we shall prove that $\mathfrak{B}(\text{appcon}(\varphi))$ turns out to be the Bennequin-type invariant considered in [33], the definition of which is based on Legendrian lifts.

Recalling the Morse description in Sect. 2.5, we denote by $\text{maxmin}_{\text{or}}(\text{appcon}(\varphi))$ (oriented maximum/minimum points) the set of all pairs (p, or) , where $p \in \text{appcon}(\varphi)$ is a local maximum or minimum point, and “or” is the local orientation of $\text{appcon}(\varphi)$ at p . We shorthand $\text{maxmin}_{\text{or}}(\text{appcon}(\varphi))$ with the following *symbols*⁴:

$$\text{maxmin}_{\text{or}}(\text{appcon}(\varphi)) = \left\{ \curvearrowleft, \curvearrowright, \smile, \frown \right\}.$$

Also, we indicate by $\text{crossings}_{\text{or}}(\text{appcon}(\varphi))$ (respectively $\text{cusps}_{\text{or}}(\text{appcon}(\varphi))$) the set of all pairs consisting of a crossing (respectively a (horizontal) cusp, or more conveniently here a marked point, see Sect. 2.5.2) of $\text{appcon}(\varphi)$ and the local orientation of $\text{appcon}(\varphi)$. We shorthand $\text{crossings}_{\text{or}}(\text{appcon}(\varphi))$ and $\text{cusps}_{\text{or}}(\text{appcon}(\varphi))$ as follows:

$$\begin{aligned} \text{crossings}_{\text{or}}(\text{appcon}(\varphi)) &= \left\{ \nearrow, \nwarrow, \searrow, \swarrow \right\}, \\ \text{cusps}_{\text{or}}(\text{appcon}(\varphi)) &= \{ \prec, \succ \}, \end{aligned}$$

where we notice that the orientation around a cusp is not indicated, since it is determined by the cusp direction: namely, \prec is oriented downwards, and \succ is oriented upwards.

²With kind permission from Elsevier, in this section and in Sects. 7.2 and 7.3 we illustrate the results and report some of the figures from the quoted paper [9].

³ $\mathfrak{B}(\text{appcon}(\varphi))$ is automatically computed in the `appcontour` program (Chap. 10).

⁴Compare also with Sect. 2.5.3.

We set

$$\begin{aligned} & \text{crit}_{\text{or}}(\text{appcon}(\varphi)) \\ & := \text{maxmin}_{\text{or}}(\text{appcon}(\varphi)) \cup \text{crossings}_{\text{or}}(\text{appcon}(\varphi)) \cup \text{cusps}_{\text{or}}(\text{appcon}(\varphi)). \end{aligned}$$

Recall, as done in Sect. 2.5.3, that if λ is a critical level, we classify it according to the (transversal) intersections of a Morse line $m_\lambda(\cdot)$ with $\text{appcon}(\varphi)$ lying on the left and on the right of the corresponding critical point.

We indicate by a vertical oriented arrow on the left (respectively on the right) of a singular point $p \in \text{nodes}(\text{appcon}(\varphi))$ the orientation of a left (respectively right) arc to p .

Following the notation in Definition 2.2.5, we recall that

$$f_\varphi(x) := \#\{m \in M : \varphi(m) = x\}, \quad x \in \mathbb{R}^2.$$

Using Remark 2.2.6, which specifies the values of f_φ on $\text{appcon}(\varphi)$, we can proceed with the following concept [9].

Definition 7.1.1 (The Function \mathfrak{b}) Let $\varphi : M \rightarrow \mathbb{R}^2$ be a stable map. Let $s \in \text{maxmin}_{\text{or}}(\text{appcon}(\varphi)) \cup \text{cusps}_{\text{or}}(\text{appcon}(\varphi))$, and $p_s \in \text{appcon}(\varphi)$ be the corresponding point. We define

$$\mathfrak{b}(s) := \begin{cases} f_\varphi(p_s) & \text{if } s \in \{\smile, \frown\}, \\ -f_\varphi(p_s) & \text{if } s \in \{\smile, \frown\}, \\ -f_\varphi(p_s) - \frac{1}{2} & \text{if } s \in \{\succ, \prec\}. \end{cases} \quad (7.1)$$

The contribution of $s \in \text{crossings}_{\text{or}}(\text{appcon}(\varphi))$ is defined as follows:

$$\mathfrak{b}(\times) := +1, \quad \mathfrak{b}(\times) := -1, \quad \mathfrak{b}(\times) := +1, \quad \mathfrak{b}(\times) := -1. \quad (7.2)$$

Definition 7.1.2 (The Number $\mathfrak{B}(\text{appcon}(\varphi))$) Let $\varphi : M \rightarrow \mathbb{R}^2$ be a stable map. We define

$$\mathfrak{B}(\text{appcon}(\varphi)) := \sum_{s \in \text{crit}_{\text{or}}(\text{appcon}(\varphi))} \mathfrak{b}(s). \quad (7.3)$$

Some comments are in order.

- The number $\mathfrak{B}(\text{appcon}(\varphi))$ only takes into account the nodes, the cusps, the extremal points (with respect to the height function) and the orientation of the apparent contour, and its definition relies on a Morse description of $\text{appcon}(\varphi)$.

- The independence of $\mathfrak{B}(\text{appcon}(\varphi))$ of the Morse description⁵ of $\text{appcon}(\varphi)$ is a consequence of Theorem 7.3.1 (see Corollary 7.3.3). In a similar way, one checks that $\mathfrak{B}(\text{appcon}(\varphi))$ is invariant under the action of any element of $\text{Diff}_c(\mathbb{R}^2)$.
- The definition of $\mathfrak{B}(\text{appcon}(\varphi))$ does not require the construction of the Legendrian lift (see Sect. 7.2).
- In the case of Legendrian knots, the so-called Thurston–Bennequin invariant (see, e.g., [44, p. 360] and the references therein), reminds definition (7.3).

The computation of $\mathfrak{B}(\text{appcon}(\varphi))$ is implemented in the program `appcon-tour`; the results obtained on some examples are presented in Sect. 10.9.2.

7.2 Definition of $BL(\text{appcon}(\varphi))$

Let us briefly recall from [33] the construction of $BL(\text{appcon}(\varphi))$. In the following \mathcal{O} is an open disk containing $\text{appcon}(\varphi)$. The construction consists of various steps.

- Step 1.* Each component of $\text{appcon}(\varphi)$ is lifted by using the direction of the (projectivized) cotangent line as an additional dimension (*Legendrian lift*, see, for instance, [34]). We can take the additional coordinate as an angle $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ (direction of the normal to $\text{appcon}(\varphi)$) with extremal values identified to each other. The Legendrian lift of $\text{appcon}(\varphi)$ obtained in this way is a set of (pairwise disjoint) smooth closed oriented curves embedded in $\mathcal{O} \times [-\frac{\pi}{2}, \frac{\pi}{2}]$ with top and bottom faces identified, which is a set \hat{T} with the topology of a solid torus.
- Step 2.* One takes a thin strip along each component γ of the Legendrian lift, by moving it of a small amount $\pm\epsilon$ in the normal direction to the contact plane⁶ at the given point. We denote by $\tilde{\gamma}$ the boundary of this strip: this may consist of one or two components, according to whether the strip is a Moebius band or a cylinder.

⁵An informal way to realize that $\mathfrak{B}(\text{appcon}(\varphi))$ is independent of the Morse description is the following. Suppose we are given two Morse descriptions of $\text{appcon}(\varphi)$. Possibly composing with two elements of $\text{Diff}_c(\mathbb{R}^2)$, we can suppose that the Morse lines of both the two Morse descriptions are horizontal and straight. The original apparent contour $\text{appcon}(\varphi)$ is changed, under the action of these two diffeomorphisms, into two new apparent contours, say G and G' . Let us construct by hand an \mathbb{R}^2 -ambient isotopy taking G into G' . One then classifies the events that appear in the path of diffeomorphisms taking G into G' , which are the following: local maxima or minima can be created or destroyed, and one checks that in both cases, (7.3) is unchanged. In addition, the number of crossings does not change, since a diffeomorphism of \mathbb{R}^2 can only locally “rotate” and translate a crossing. When performing a local rotation, local maxima and minima are introduced, and one checks directly the invariance using definition (7.3) and Definition 7.1.1. Also the number of cusps does not change; moreover, cusps have been previously transformed into transversal marked points: applying a local rotation to an arc containing a marked point, the invariance follows from the definition, since the weight is independent of the orientation of the arc.

⁶That is, the tangent line to $\text{appcon}(\varphi)$ times \mathbb{R} .

Step 3. The solid torus \hat{T} can be identified with a solid torus

$$T \subset \mathbb{R}^3,$$

obtained by rotating \mathcal{O} around some line $r \subset \mathbb{R}^3$ by the angle 2θ , $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, with r chosen so that

$$r \cap \overline{\mathcal{O}} = \emptyset \quad \text{and} \quad r \subset \{z = 0\}, \quad (7.4)$$

$\{z = 0\}$ being the plane containing \mathcal{O} . We obtain a set of oriented closed nonintersecting curves $\Gamma \subset T$, together with the corresponding strip boundaries $\tilde{\Gamma} \subset T$, where each Γ (respectively $\tilde{\Gamma}$) is the image in T of $\gamma \subset \hat{T}$ (respectively $\tilde{\gamma} \subset \hat{T}$), obtained using such an identification.

Step 4. For each Γ one computes a self-linking number, which is obtained by first projecting Γ and $\tilde{\Gamma}$ onto some generic plane, and then counting crossings of the projected Γ and of the projected $\tilde{\Gamma}$ with an appropriate sign,⁷ and with the following appropriate weight: 1 for self-crossings of the projected Γ and $\frac{1}{2}$ for self-crossings of the projected $\tilde{\Gamma}$. When counting self-crossings of the projected $\tilde{\Gamma}$ one omits crossing points that are inherited by self-crossings of the projected Γ ; in other words, one omits self-crossings of the projected $\tilde{\Gamma}$ that are at $O(\epsilon)$ distance from self-crossings of the projected Γ . The resulting quantity can be shown to be invariant under Reidemeister moves on the projections of Γ and $\tilde{\Gamma}$. This entails that the final result is in fact independent of the choice of the generic projection plane.

Step 5. One also computes the linking numbers of any pair of distinct components Γ_1 and Γ_2 of the Legendrian lift. We use the same sign convention as before at each mutual crossing of the projection onto some plane; in this case, the displaced curves $\tilde{\Gamma}_i$ are not involved. The resulting quantity is also invariant under Reidemeister moves on the corresponding projections.

Step 6. Finally,

$$BL(\text{apcon}(\varphi))$$

is defined as the sum of the self-linking and linking numbers of the Legendrian lift.

⁷According to the vector product of the tangent vector to the path above and the tangent vector to the path below.

It turns out that $BL(\text{appcon}(\varphi))$ is an invariant⁸ of the apparent contour $\text{appcon}(\varphi)$ under the action of $\text{Diff}_c(\mathbb{R}^2)$.

7.3 Coincidence Between $\mathfrak{B}(\text{appcon}(\varphi))$ and $BL(\text{appcon}(\varphi))$

Following closely [9], in this section we prove that $\mathfrak{B}(\text{appcon}(\varphi))$ coincides with $BL(\text{appcon}(\varphi))$. In order to do this, we need to recall some results mentioned in [33] (see also [5, 41] and [36]).

Let us denote by

$$\text{Unstable}(M, \mathbb{R}^2) \subset \mathcal{C}^\infty(M, \mathbb{R}^2)$$

the collection of all unstable maps in $\mathcal{C}^\infty(M, \mathbb{R}^2)$ (sometimes called discriminant hypersurface⁹). The codimension one strata of $\text{Unstable}(M, \mathbb{R}^2)$ are classified¹⁰ as follows¹¹: L (lips), B (beak-to-beak), S (swallow's tail), K_0 , K_1 , K_2 (kasanie, or tangency), T_0 , T_1 (triple points), C_0 , C_1 (cusp-crossing). We refer to [33, Figure 2], where it is shown that each stratum can be realized locally as the projection on \mathbb{R}^2 of a surface immersed in \mathbb{R}^3 (see also Fig. 6.3, which however refers to the less general case of labelled apparent contours).

Each stratum is locally cooriented, positively in the direction where the number of cusps and crossings is increasing; as for T_0 and T_1 , for which the number of cusps and crossings does not change, the coorientation is toward the region where the number of preimages of the newly formed triangle is higher [33, p. 30].

Given the ten local strata of the discriminant hypersurface, the corresponding small letters (l , b , s , k_0 , k_1 , k_2 , t_0 , t_1 , c_0 , c_1) denote functions whose jump is equal to one at every crossing of the stratum in the positive direction defined by

⁸As we have already said in the Introduction, $BL(\text{appcon}(\varphi))$ is called a Bennequin-type invariant; see, e.g., [36]. In [33] it is proved that all local first order Vassiliev-type invariants of $\text{appcon}(\varphi)$ are a combination of the number of cusps of $\text{appcon}(\varphi)$, the number of crossings of $\text{appcon}(\varphi)$, and of $BL(\text{appcon}(\varphi))$.

⁹Or also the bifurcation set of $\mathcal{C}^\infty(M, \mathbb{R}^2)$ (see [42, 43] and [30]). Recall that, if a map φ belongs to $\text{Unstable}(M, \mathbb{R}^2)$, then every neighbourhood of φ contains maps not equivalent to φ . We refer to the survey article [13] for more information.

¹⁰Not surprisingly, such a classification is similar to the one in Chap. 6; this becomes reasonable, if one interprets φ as the first two coordinates of a *local* embedding of M in \mathbb{R}^3 . Notice carefully that M is, in this chapter, an abstract two-manifold, therefore there is no labelling on $\text{appcon}(\varphi)$. In contrast, in Chap. 6 only labelled apparent contours are taken into account, and for this reason the number of possible cases is much larger.

¹¹See, e.g., [2–4, 6, 7] and [22].

the coorientation: $\Delta l = \Delta b = \Delta s = \Delta k_0 = \Delta k_1 = \Delta k_2 = \Delta t_0 = \Delta t_1 = \Delta c_0 = \Delta c_1 = 1$. The invariant $BL(\text{appcon}(\varphi))$, which we are interested in, has jump given by

$$\Delta BL(\text{appcon}(\varphi)) = \Delta l + \Delta b + 2\Delta k_0 - 2\Delta k_1 + 2\Delta k_2.$$

Theorem 7.3.1 (Coincidence) *Let M be a closed two-dimensional manifold (not necessarily orientable) of class C^∞ . Let $\varphi : M \rightarrow \mathbb{R}^2$ be a stable map. We have*

$$\mathfrak{B}(\text{appcon}(\varphi)) = BL(\text{appcon}(\varphi)). \quad (7.5)$$

The next section is devoted to the proof of Theorem 7.3.1. We shall split the proof, which is the same as in [9], into two parts. First we check that both $\mathfrak{B}(\text{appcon}(\varphi))$ and $BL(\text{appcon}(\varphi))$ jump of the same amount when crossing each codimension one stratum of the discriminant hypersurface. As a consequence, we have that $\mathfrak{B}(\text{appcon}(\varphi)) = BL(\text{appcon}(\varphi))$ up to a constant. In the second part of the proof, we show that actually this constant is zero.

7.3.1 Proof of Coincidence Up to a Constant

We start the first part of the proof, by localizing the degeneration in a sufficiently small box. Let α be the value of f_φ on the right of the box; in terms of α , we can write the contribution to $\mathfrak{B}(\text{appcon}(\varphi))$ inside the box.

We use a family of Morse horizontal straight lines (see Sect. 2.5) travelling downwards, as the parameter λ decreases. The box A is a local description before the degeneration, and the box B after the degeneration (see Figs. 7.1 and 7.2).

1. Stratum L. The contribution given by box A is clearly zero; in box B, the function f_φ assumes value $\alpha + 2$ inside the contour, $\alpha + 1$ at the local maximum and minimum points, α at the two cusps. Hence, recalling the definition of \mathfrak{b} given in (7.1), we have

$$\text{in box B : } \begin{cases} \mathfrak{b}(\curvearrowright) = 1 + \alpha, \\ \mathfrak{b}(\succ) = -\alpha - \frac{1}{2}, \\ \mathfrak{b}(\prec) = -\alpha - \frac{1}{2}, \\ \mathfrak{b}(\curvearrowleft) = 1 + \alpha. \end{cases}$$

Therefore $\mathfrak{B}(\text{box B}) = \mathfrak{b}(\curvearrowright) + \mathfrak{b}(\succ) + \mathfrak{b}(\prec) + \mathfrak{b}(\curvearrowleft) = 1$, $\mathfrak{B}(\text{box A}) = 0$,

$$\Delta \mathfrak{B} := \mathfrak{B}(\text{box B}) - \mathfrak{B}(\text{box A}) = 1.$$

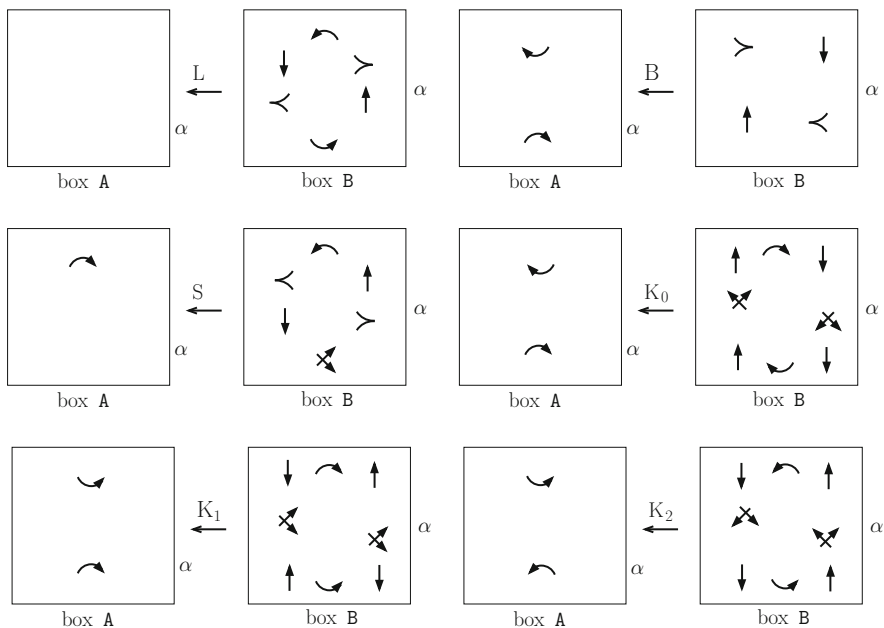


Fig. 7.1 Strata L, B, S, K_0 , K_1 , K_2 . Image taken from [9]

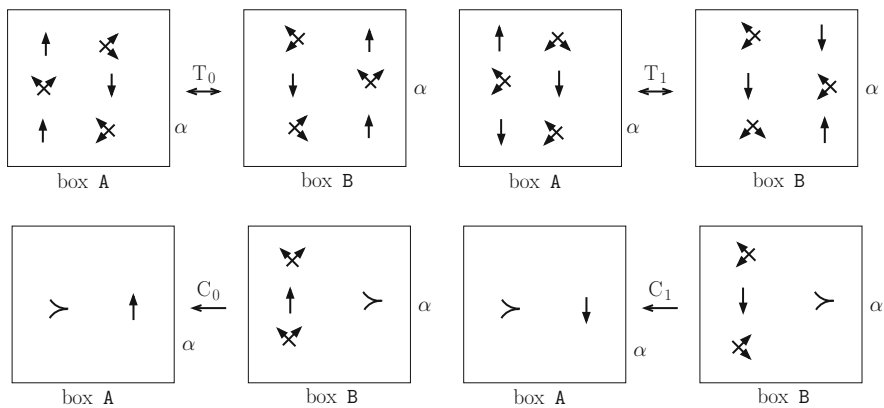


Fig. 7.2 Strata T_0 , T_1 , C_0 , C_1 . Image taken from [9]

2. Stratum B. Here the value of f_ψ at the extremal points is $\alpha - 1$, the value at the cusps is $\alpha - 2$. Therefore, we have

$$\text{in box A : } \begin{cases} b(\curvearrowright) = 1 - \alpha, \\ b(\curvearrowleft) = 1 - \alpha. \end{cases}$$

$$\text{in box B : } \begin{cases} \mathfrak{b}(\succ) = -(\alpha - 2) - \frac{1}{2} = \frac{3}{2} - \alpha, \\ \mathfrak{b}(\prec) = -(\alpha - 2) - \frac{1}{2} = \frac{3}{2} - \alpha. \end{cases}$$

Therefore $\mathfrak{B}(\text{box A}) = 2 - 2\alpha$, $\mathfrak{B}(\text{box B}) = 3 - 2\alpha$ and $\Delta\mathfrak{B} = 1$.

3. Stratum S. Before the degeneration the values of f_φ outside the contour are α and $\alpha - 2$, hence we have a value of $\alpha - 1$ at the extremal point in box A, $\alpha + 1$ at the extremal point in box B and α at the two cusps. Therefore

$$\text{in box A : } \quad \mathfrak{b}(\curvearrowright) = 1 - \alpha.$$

$$\text{in box B : } \begin{cases} \mathfrak{b}(\curvearrowleft) = 1 + \alpha, \\ \mathfrak{b}(\prec) = -\alpha - \frac{1}{2}, \\ \mathfrak{b}(\succ) = -\alpha - \frac{1}{2}, \\ \mathfrak{b}(\bowtie) = 1. \end{cases}$$

It follows that $\mathfrak{B}(\text{box A}) = 1 - \alpha$, $\mathfrak{B}(\text{box B}) = 1 - \alpha$ and $\Delta\mathfrak{B} = 0$.

4. Stratum K_0 . After the degeneration we have $f_\varphi = \alpha - 4$ in the internal region; the value of f_φ is $\alpha - 1$ at the two extremal points in box A and $\alpha - 3$ at the two extremal points in box B. Therefore

$$\text{in box A : } \begin{cases} \mathfrak{b}(\curvearrowright) = 1 - \alpha, \\ \mathfrak{b}(\curvearrowleft) = 1 - \alpha. \end{cases}$$

$$\text{in box B : } \begin{cases} \mathfrak{b}(\curvearrowright) = 3 - \alpha, \\ \mathfrak{b}(\bowtie) = -1, \\ \mathfrak{b}(\times) = -1, \\ \mathfrak{b}(\curvearrowleft) = 3 - \alpha. \end{cases}$$

It follows that $\mathfrak{B}(\text{box A}) = 2 - 2\alpha$, $\mathfrak{B}(\text{box B}) = -2\alpha + 4$ and $\Delta\mathfrak{B} = 2$.

5. Stratum K_1 . In box A the values of f_φ are $\alpha + 2$, α and $\alpha - 2$ respectively in the top, middle and bottom region, hence we have values of $\alpha + 1$ and $\alpha - 1$ at the extremal points. The same values are achieved in box B. Therefore

$$\text{in box A : } \begin{cases} \mathfrak{b}(\curvearrowright) = 1 + \alpha, \\ \mathfrak{b}(\curvearrowleft) = 1 - \alpha. \end{cases}$$

$$\text{in box B : } \begin{cases} \mathfrak{b}(\curvearrowright) = -\alpha - 1, \\ \mathfrak{b}(\curvearrowleft) = +1, \\ \mathfrak{b}(\curvearrowright) = +1, \\ \mathfrak{b}(\curvearrowleft) = \alpha - 1. \end{cases}$$

It follows that $\mathfrak{B}(\text{box A}) = 2$, $\mathfrak{B}(\text{box B}) = 0$ and $\Delta\mathfrak{B} = -2$.

6. Stratum K_2 . Here the values of f_φ at the extremal points are $\alpha + 1$ in box A and $\alpha + 3$ in box B. Therefore

$$\text{in box A : } \begin{cases} \mathfrak{b}(\curvearrowright) = 1 + \alpha, \\ \mathfrak{b}(\curvearrowleft) = 1 + \alpha. \end{cases}$$

$$\text{in box B : } \begin{cases} \mathfrak{b}(\curvearrowright) = \alpha + 3, \\ \mathfrak{b}(\curvearrowleft) = -1, \\ \mathfrak{b}(\curvearrowright) = -1, \\ \mathfrak{b}(\curvearrowleft) = \alpha + 3. \end{cases}$$

It follows that $\mathfrak{B}(\text{box A}) = 2 + 2\alpha$, $\mathfrak{B}(\text{box B}) = 2\alpha + 4$ and $\Delta\mathfrak{B} = 2$.

7. Stratum T_0 . We have

$$\text{in box A : } \begin{cases} \mathfrak{b}(\curvearrowright) = +1, \\ \mathfrak{b}(\curvearrowleft) = -1, \\ \mathfrak{b}(\curvearrowright) = +1. \end{cases} \quad \text{in box B : } \begin{cases} \mathfrak{b}(\curvearrowright) = +1, \\ \mathfrak{b}(\curvearrowleft) = -1, \\ \mathfrak{b}(\curvearrowright) = +1. \end{cases}$$

Therefore $\mathfrak{B}(\text{box A}) = 1$, $\mathfrak{B}(\text{box B}) = 1$ and $\Delta\mathfrak{B} = 0$.

8. Stratum T_1 . We have

$$\text{in box A : } \begin{cases} \mathfrak{b}(\curvearrowright) = -1, \\ \mathfrak{b}(\curvearrowleft) = +1, \\ \mathfrak{b}(\curvearrowright) = +1. \end{cases} \quad \text{in box B : } \begin{cases} \mathfrak{b}(\curvearrowright) = +1, \\ \mathfrak{b}(\curvearrowleft) = +1, \\ \mathfrak{b}(\curvearrowright) = -1. \end{cases}$$

Therefore $\mathfrak{B}(\text{box A}) = 1$, $\mathfrak{B}(\text{box B}) = 1$ and $\Delta\mathfrak{B} = 0$.

9. Stratum C_0 . Going from right to left in box A the function f_φ assumes the values α , $\alpha + 2$, $\alpha + 4$ in the three regions. The value at the cusp is $\alpha + 2$ in box A and α in box B. Therefore

$$\text{in box A : } \mathfrak{b}(\succ) = -\alpha - \frac{5}{2}.$$

$$\text{in box B : } \begin{cases} \mathfrak{b}(\bowtie) = -1, \\ \mathfrak{b}(\succ) = -\alpha - \frac{1}{2}, \\ \mathfrak{b}(\searrow) = -1. \end{cases}$$

It follows that $\mathfrak{B}(\text{box A}) = -\alpha - \frac{5}{2}$, $\mathfrak{B}(\text{box B}) = -\alpha - \frac{5}{2}$ and $\Delta\mathfrak{B} = 0$.
10. Stratum C_1 . We have

$$\text{in box A : } \mathfrak{b}(\succ) = -\alpha + \frac{3}{2}.$$

$$\text{in box B : } \begin{cases} \mathfrak{b}(\bowtie) = +1, \\ \mathfrak{b}(\succ) = -\alpha - \frac{1}{2}, \\ \mathfrak{b}(\searrow) = +1. \end{cases}$$

Therefore $\mathfrak{B}(\text{box A}) = -\alpha + \frac{3}{2}$, $\mathfrak{B}(\text{box B}) = -\alpha + \frac{3}{2}$ and $\Delta\mathfrak{B} = 0$.

Now, it is proven in [33, Lemma 5.4 and Section 2] that the variation of $BL(\text{appcon}(\varphi))$, before and after a codimension one degeneration, is the same as the one computed above. Therefore, there exists a constant C such that

$$\mathfrak{B}(\text{appcon}(\varphi)) - BL(\text{appcon}(\varphi)) = C. \quad (7.6)$$

7.3.2 Proof of Coincidence

We pass to the second part of the proof of Theorem 7.3.1. We shall provide a constructive argument, based on an alternative derivation of $\mathfrak{B}(\text{appcon}(\varphi))$; this argument will show that the additive constant C , found in Eq. (7.6), actually vanishes.

The key tool is the invariance under Reidemeister moves of the self-linking and linking numbers of the Legendrian lift (see Sect. 7.1). This allows us to perform a deformation (see (7.7), below) of the Legendrian lift, and to choose a projection plane to compute the self-linking and linking numbers directly on the original apparent contour $\text{appcon}(\varphi)$. This deformation, however, cannot be done directly, since the Legendrian lift in the solid torus T involves a rotation around some line r , with r satisfying (7.4).

For convenience, let us translate \mathcal{O} so that the straight line r passes through the origin, with equation $\{x_1 = z = 0\}$. If $(P, \theta) \in \gamma \subset \hat{T}$, $P = (P_1, P_2)$ is a point on a component γ of the Legendrian lift, the corresponding point under the identification of \hat{T} with T is given by

$$(P_1 \cos(2\theta), P_2, P_1 \sin(2\theta)).$$

Given ζ a small positive number, we map (P, θ) in

$$(P_1 \cos(2\zeta\theta), P_2, P_1 \sin(2\zeta\theta)) \quad \text{if } \theta \in (-\frac{\pi}{2}, \frac{\pi}{2}), \quad (7.7)$$

$$\{(P_1 \cos(\omega\pi), P_2, P_1 \sin(\omega\pi)) : \zeta \leq |\omega| \leq 1\} \quad \text{if } |\theta| = \frac{\pi}{2}. \quad (7.8)$$

Note that when $|\theta| = \frac{\pi}{2}$ this gives a set of values connecting nearby points along a large circle. Notice that, as ζ grows to 1, this identification is continuously deformed to the original one of step 3, Sect. 7.2.

For $\zeta > 0$ very small, the set of angles from $-\zeta\pi$ to $\zeta\pi$ live in a set which approaches a shallow cylinder; a further small deformation can be applied to get the exact cylinder

$$C = \mathcal{O} \times (-\zeta, \zeta);$$

see Fig. 7.3. We denote by Γ_ζ and $\tilde{\Gamma}_\zeta$ the components of the Legendrian lift thus deformed.

We cannot yet project onto \mathcal{O} , because the projection is nongeneric, since we have distinct points of Γ_ζ and $\tilde{\Gamma}_\zeta$ (those on the top/bottom face of C) that project onto the same point in $\text{appcon}(\varphi)$ with the same (horizontal) tangent. We will then slightly perturb Γ_ζ and $\tilde{\Gamma}_\zeta$, e.g., by *cutting* the apparent contour at points with horizontal tangent, which we can suppose to be local maxima or minima and then moving a little bit apart the two sides. Figure 7.3 represents the construction for an apparent contour which is just a circle: the Legendrian lift Γ_ζ (thicker curves) has two components, since the normal line to $\text{appcon}(\varphi)$ attains each direction twice (at

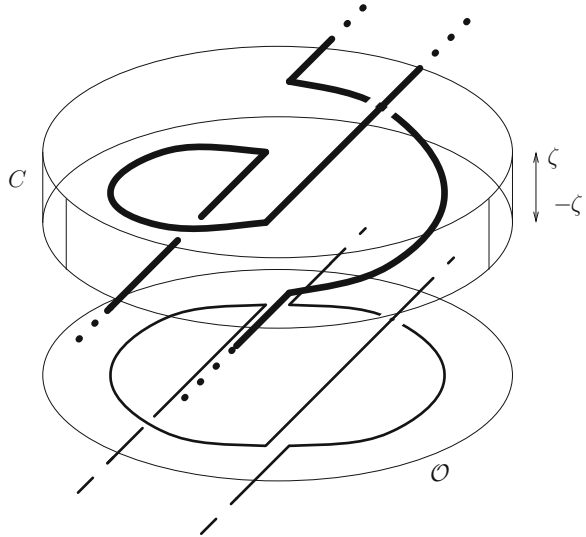


Fig. 7.3 Legendrian lift Γ_ζ (thick lines) of a closed circle and projection back on \mathcal{O} (solid lines). Image taken from [9]

opposite points on the circle). For the sake of clarity the long arcs of Γ_ζ connecting the top to the bottom endpoints of the two components are shown as straight thick lines at 45 degrees; one should imagine the same arcs at an angle close to zero (so that their projections on the plane $\{z = 0\}$ are almost horizontal with positive slope). The result of the projection back onto \mathcal{O} is also shown in Fig. 7.3, with the signature of the crossing points depending on which arc crosses over the other.

Without loss of generality (and possibly applying a deformation of $\text{appcon}(\varphi)$) we can assume that

- the two tangent lines at a crossing are not horizontal, and hence both components of $\text{appcon}(\varphi)$ cross the horizontal line transversally;
- each horizontal line crosses the apparent contour in at most one critical point (a cusp or a crossing or a maximum/minimum);
- all cusps have horizontal tangent with the two branches lying at opposite sides of the tangent. This is a source of nongenericity, since horizontal tangent implies that cusp points end up on the top and bottom faces of the cylinder C . However by slightly curving the cusps and introducing a new minimum/maximum point we can resolve this issue.

We are now in a position to compute all contributions to $BL(\text{appcon}(\varphi))$ described in steps 4 and 5 of Sect. 7.2, by considering all symbols $s \in \text{crit}_{\text{or}}(\text{appcon}(\varphi))$.

- $s \in \text{crossings}_{\text{or}}(\text{appcon}(\varphi))$. We need to recover the signature of the crossing points and add it to the linking and self-linking numbers between the components Γ_ζ . In the Legendrian lift the *decreasing* branch crosses over the *increasing* branch, so that we end up with $\mathfrak{b}(s)$ as defined in (7.2).
- $s \in \text{cusps}_{\text{or}}(\text{appcon}(\varphi))$. For definiteness, let us assume that the cusp is pointing to the right. The horizontal tangent at the cusp is a source of nongenericity; therefore, we need to deform it by, e.g., curving it slightly upwards, thus introducing a new minimum point corresponding to a symbol of type \curvearrowright that must then be taken into account. Moreover this deformation also introduces a new crossing between the horizontal tangent to the minimum point and one of the branches of the cusp; the corresponding contribution shall be taken into account later on. The cusp itself is lifted in the Legendrian lift with a vertical tangent: it is therefore necessary a further deformation in order to achieve genericity. This produces a self-crossing of $\tilde{\Gamma}_\zeta$ with positive signature and hence contributes with $\frac{1}{2}$ in $BL(\text{appcon}(\varphi))$.
- $s \in \text{maxmin}_{\text{or}}(\text{appcon}(\varphi))$. Take for definiteness a local minimum p_s . Such a point corresponds to a long arc that connects the top face (left half of the *cut* minimum point) to the bottom face (right half of the *cut* minimum point) of the cylinder C . Up to a deformation, we can assume that these arcs project on curves that travel almost horizontally towards the right, with a nearby crossing over $\text{appcon}(\varphi)$ on the right of p_s and possibly other crossings over $\text{appcon}(\varphi)$ that lie on the right of p_s , make a large curve and reach the minimum point from the left, possibly crossing below arcs of $\text{appcon}(\varphi)$ lying on the left of p_s . To account

for these crossings we define an integer-valued function which is *nonlocal*, in the sense that it depends on the orientation of the left and right arcs to the symbol.

Definition 7.3.2 (The Function w) Let $p \in \mathbb{R}^2$ and ℓ be the horizontal straight line passing through p , and denote by $\ell_l \subset \ell$ (respectively $\ell_r \subset \ell$) the open half-line starting at p and lying on the left (respectively on the right) of p . We define

$$w(p) := w_{\text{left}}(p) + w_{\text{right}}(p),$$

$$w_{\text{left}}(p) := \#_{\text{left}} \downarrow - \#_{\text{left}} \uparrow, \quad w_{\text{right}}(p) := \#_{\text{right}} \uparrow - \#_{\text{right}} \downarrow,$$

where $\#_{\text{left}} \downarrow$ (respectively $\#_{\text{left}} \uparrow$) denotes the number of points of $\ell_l \cap \text{appcon}(\varphi)$ where $\text{appcon}(\varphi)$ traverses ℓ_l downwards (respectively upwards), and $\#_{\text{right}} \uparrow$ (respectively $\#_{\text{right}} \downarrow$) denotes the number of points of $\ell_r \cap \text{appcon}(\varphi)$ where $\text{appcon}(\varphi)$ traverses ℓ_r upwards (respectively downwards).

Note that¹²

$$p \notin \text{appcon}(\varphi) \Rightarrow w(p) = f_{\varphi}(p). \quad (7.9)$$

When $p = p_s$ is the point corresponding to $s \in \max_{\text{min}_{\text{or}}}(\text{appcon}(\varphi))$, the resulting $w(p_s)$ accounts for all crossings far from p_s . It has the correct sign, with respect to the contribution in the linking and self-linking numbers, if $s \in \{\smile, \frown\}$, whereas it has the opposite sign if $s \in \{\curvearrowright, \curvearrowleft\}$. Indeed, in the latter case, the line ℓ is traversed from right to left. The contribution coming from cusps (see the case $s \in \text{cusps}_{\text{or}}(\text{appcon}(\varphi))$ above) always corresponds to the symbol \curvearrowright , so that in this case w has the opposite sign. We still need to take into account the crossing near p_s ; in all cases, when $s \in \max_{\text{min}_{\text{or}}}(\text{appcon}(\varphi))$, a direct check shows that this crossing is always positive, whereas in case of cusps we have two *new* crossings with opposite signature, so that we have no contribution to the self-linking number.

We end up with a contribution:

$$\begin{aligned} 1 + w(p_s) & \text{ if } s \in \{\smile, \frown\}, \\ 1 - w(p_s) & \text{ if } s \in \{\curvearrowright, \curvearrowleft\}, \\ \frac{1}{2} - w(p_s) & \text{ if } s \in \{<, >\}, \end{aligned} \quad (7.10)$$

where in the cusp case we also take into account the self-linking due to intersections of $\tilde{\Gamma}_{\zeta}$.

¹²Indeed, it suffices to check that both functions are zero far from $\text{appcon}(\varphi)$ and that both have the same jumps when crossing $\text{appcon}(\varphi)$.

Claim Let $s \in \maxmin_{\text{or}}(\text{appcon}(\varphi)) \cup \text{cusps}_{\text{or}}(\text{appcon}(\varphi))$ and $p_s \in \text{appcon}(\varphi)$ be the corresponding point. Then

$$\begin{aligned} s \in \{\searrow, \swarrow\} &\Rightarrow w(p_s) = f_\varphi(p_s) - 1, \\ s \in \{\nearrow, \nwarrow\} &\Rightarrow w(p_s) = f_\varphi(p_s) + 1, \\ s \in \{>, <\} &\Rightarrow w(p_s) = f_\varphi(p_s) + 1. \end{aligned}$$

Indeed, recalling the definition of w , (7.9) and Remark 2.2.6, the case where p_s is a local maximum (respectively minimum) follows by computing f_φ at a point slightly above (respectively below) p_s . If p_s is a cusp, we compute w at a point z slightly above p_s , obtaining $w(p) = w(p_s) - 1$ for both type of cusps, and the result follows from $f_\varphi(p) = f_\varphi(p_s)$.

Finally, we can now prove (7.5): it was already observed that crossings contribute in the same way. For the other type of symbols the result follows directly by recalling Definition 7.1.2 and the contributions to $BL(\text{appcon}(\varphi))$ given by (7.10), in light of the claim. \square

Since the definition of $BL(\text{appcon}(\varphi))$ does not involve any Morse description of the apparent contour, from Theorem 7.3.1 we deduce the following result.

Corollary 7.3.3 (Independence) $\mathfrak{B}(\text{appcon}(\varphi))$ is independent of the Morse description of $\text{appcon}(\varphi)$.

7.4 Euler–Poincaré Characteristic of ∂E

In the previous sections of this chapter we have studied certain \mathbb{R}^2 -ambient isotopic invariants of a (not necessarily labelled) apparent contour. The main results of this section are, instead, confined to labelled apparent contours; however, we will be concerned with a stronger invariance, namely, invariants of the corresponding 3D shapes¹³ E under \mathbb{R}^3 -ambient isotopies.

One of the main examples of such invariants is represented by the right-hand side of formula (7.11), below. More specifically, we recall from Theorem 5.1.1 that a complete labelled contour graph (G, d) can be interpreted as the apparent contour of a surface $\Sigma = \partial E$, where Σ is the image of an embedding in \mathbb{R}^3 of a smooth two-dimensional closed manifold M . Hence (G, d) inherits all topological information of Σ , and therefore one expects that it should be possible to compute from it the Euler–Poincaré characteristic $\chi(\Sigma)$. As we shall see, this is indeed the case: $\chi(\Sigma)$ can be computed solely in terms of the apparent contour¹⁴ [8]. Even more, it turns out that the actual computation of $\chi(\Sigma)$ from (G, d) *does not* involve the labelling d . A posteriori, this is not surprising, since $\chi(\Sigma)$ depends only on the manifold M .

¹³See also [17, Definition (1.19), p. 107] for related subjects.

¹⁴Therefore, the resulting construction is not only invariant under diffeomorphic equivalence of apparent contours (Definition 2.4.2), but also under \mathbb{R}^3 -ambient isotopies.

This suggests that the computed value is the Euler–Poincaré characteristic of the source manifold also in the more general setting of apparent contours of abstract manifolds. Interestingly, this turns out to be the case,¹⁵ provided we conventionally define (as it is customary) the Euler–Poincaré characteristic of a nonconnected manifold as the sum of the characteristics of each connected component (total characteristic). We refer to Sect. 10.7.1 for an implementation of the formula in the next result, and for further information.

Theorem 7.4.1 (Total Characteristic from the Apparent Contour) *Let (G, f, d) and $\Sigma = \partial E$ be as in Theorem 5.1.1. Then the Euler–Poincaré characteristic $\chi(\Sigma)$ of Σ is given by*

$$\chi(\Sigma) = F_{\text{regions}}(G) - S_{\text{arcs}}(G) + V_{\text{crossings}}(G) + V_{\text{cusps}}(G), \quad (7.11)$$

where

$$\begin{aligned} F_{\text{regions}}(G) &:= \sum_{R \text{ region of } G} \left(2 - \#\{\text{connected components of } \partial R\} \right) f(R), \\ S_{\text{arcs}}(G) &:= \sum_{\substack{a \text{ arc of } G, \\ a \cap \text{nodes}(G) \neq \emptyset}} \frac{f^+(a) + f^-(a)}{2}, \\ V_{\text{crossings}}(G) &:= \sum_{p \in \text{crossings}(G)} (f_{\min}(p) + 2), \\ V_{\text{cusps}}(G) &:= \sum_{c \in \text{cusps}(G)} f_{\min}(c). \end{aligned} \quad (7.12)$$

In particular, $\chi(\Sigma)$ is independent of d .

Proof Let us consider a curvilinear partition \mathcal{P} of Σ , that is a finite family of closed connected subsets $T \subseteq \Sigma$ having the following properties:

- each T , called face, is the homeomorphic image of a closed planar polygon,
- $\bigcup_{T \in \mathcal{P}} T = \Sigma$,
- the intersection of two distinct faces is either empty, or a vertex, or an edge.¹⁶

¹⁵See, e.g., [26], [35, Thm. 1] and [27].

¹⁶By a vertex (respectively an edge) of \mathcal{P} we mean the homeomorphic image of a vertex (respectively an edge) of a closed planar polygon.

It is well known¹⁷ that

$$\chi(\Sigma) = F - S + V, \quad (7.13)$$

where F is the number of faces, S the number of edges and V the number of vertices of the curvilinear partition.

Remembering the definition of the projection π in (3.1), we have that the graph G induces a partition of Σ , obtained from the closed graph¹⁸

$$\pi^{-1}(G) \cap \Sigma,$$

the set of vertices of which is, by definition, $\pi^{-1}(\text{nodes}(G)) \cap \Sigma$. Then the preimage through π of each connected component of the complement of G with connected boundary, each arc of $\pi^{-1}(G) \cap \Sigma$ containing at least one vertex and each vertex are considered as elements of a partition of Σ .

We observe that this partition may differ from a curvilinear partition of Σ , since there may be:

- loops (i.e., boundary of faces) with no vertices,
- faces which are not simply connected (hence with nonconnected boundary).

However, these loops do not contribute in the computation of $\chi(\Sigma)$ in (7.13). Indeed, adding v vertices to a loop without vertices splits it into v edges, and therefore the S and V contributions relatively to this loop satisfy $-S + V = 0$.

Now, let us compute the contribution to $\chi(\Sigma)$ of the above-mentioned non-simply connected faces. Let $R \subset \Sigma$ be a (connected) face with nonconnected boundary, and let

$$c_R := \# \{\text{connected components of } \partial R\}.$$

Denote by $\partial_{\text{ext}} R \subset \partial R$ the external curve, bounding the union of all other connected components $\partial_{\text{inn}}^{\text{cc}} R$ of ∂R . To embed the given configuration in a curvilinear partition, we add at least two arcs joining $\partial_{\text{inn}}^{\text{cc}} R$ with $\partial_{\text{ext}} R$. If $\partial_{\text{ext}} R$ and all $\partial_{\text{inn}}^{\text{cc}} R$ contain some vertex, we may choose the joining arcs without adding any new vertex. In this way R splits into c_R connected faces, and we have added $2(c_R - 1)$ new arcs. Hence the contribution to $\chi(\Sigma)$ is

$$c_R - 2(c_R - 1) = 2 - c_R.$$

¹⁷Recall that the Euler–Poincaré characteristic of a CW complex is $\sum_d (-1)^d \# \{d\text{-dimensional cell}\}$; see, for instance, [40, p. 429], [11].

¹⁸This graph contains the singular curve (see Remark 3.2.4). The corresponding partition has been considered, in more generality, for instance in [46].

If a connected component C of ∂R contains no vertices, then C is a closed loop, and we have to introduce some new vertices besides the new arcs; however, from the preceding arguments, any number of new vertices on a closed loop gives no contribution to the total sum.

From Theorems 5.1.1 and 5.1.4 it follows that Σ is homeomorphic to a quotient \mathcal{T} (see formula (5.1)), obtained by pasting in a suitable way the various copies of the boundaries of the regions. The statement of the proposition follows by recalling that, in the cut-and-paste construction of \mathcal{T} ,

- a region R appears $f(R)$ times, and

$$\begin{aligned} F_{\text{regions}}(G) &= \sum_{\substack{R \text{ region of } G \text{ with} \\ \partial R \text{ connected}}} f(R) \\ &+ \sum_{\substack{R \text{ region of } G \text{ with} \\ \partial R \text{ not connected}}} \left(2 - \#(\text{connected components of } \partial R)\right) f(R), \end{aligned}$$

- an arc a appears $\frac{f^+(a) + f^-(a)}{2}$ times,
- a crossing p appears $(f_{\min}(p) + 2)$ times,
- a cusp c appears $f_{\min}(c)$ times.

□

It is immediate to check that the right-hand side of (7.11) is invariant under \mathbb{R}^2 -ambient isotopies.

Remark 7.4.2 (Additivity on Connected Components) In Theorem 7.4.1 the manifold Σ is *not* supposed to be connected. Assume, for example, that Σ consists of two connected components, namely $\Sigma = \Sigma_1 \cup \Sigma_2$, so that $\chi(\Sigma) = \chi(\Sigma_1) + \chi(\Sigma_2)$. From (7.11) we deduce the following formula, which is not directly evident. Let G_{Σ_1} (respectively G_{Σ_2}) be the apparent contour of Σ_1 (respectively of Σ_2). Then

$$\begin{aligned} &F_{\text{regions}}(G_{\Sigma_1 \cup \Sigma_2}) - S_{\text{arcs}}(G_{\Sigma_1 \cup \Sigma_2}) + V_{\text{crossings}}(G_{\Sigma_1 \cup \Sigma_2}) + V_{\text{cusps}}(G_{\Sigma_1 \cup \Sigma_2}) \\ &= F_{\text{regions}}(G_{\Sigma_1}) - S_{\text{arcs}}(G_{\Sigma_1}) + V_{\text{crossings}}(G_{\Sigma_1}) + V_{\text{cusps}}(G_{\Sigma_1}) \\ &\quad + F_{\text{regions}}(G_{\Sigma_2}) - S_{\text{arcs}}(G_{\Sigma_2}) + V_{\text{crossings}}(G_{\Sigma_2}) + V_{\text{cusps}}(G_{\Sigma_2}). \end{aligned}$$

Remark 7.4.3 (Characteristic of the Interior and the Exterior) In the special case where ∂E is connected, formula (7.12) allows us to deduce the Euler–Poincaré

characteristic of the solid set E and of its complement $\mathbb{R}^3 \setminus E$ from the apparent contour G , since¹⁹

$$\chi(E) = \frac{1}{2}\chi(\partial E), \quad \chi(\mathbb{R}^3 \setminus E) = \frac{1}{2}\chi(\partial E) + 1.$$

It is customary to compactify \mathbb{R}^3 into \mathbb{S}^3 : in this case, one obtains the more symmetric formula

$$\chi(E) = \chi(\mathbb{S}^3 \setminus E) = \frac{1}{2}\chi(\partial E).$$

Obviously, these numbers do not identify the \mathbb{R}^3 -ambient isotopy equivalence class of the embedding of the set E , as the following example shows.

Example 7.4.4 Let

- E_0 be a knotted solid torus,
- E_1 be the standard solid torus,
- E_2 be a sphere with two small two-dimensional disks removed around the north and south poles, and with a knotted gallery connecting the two removed disks (the so-called knotted anti-torus, see, for instance, [1, Fig. 7], and Fig. 10.20).

Then

$$\chi(\partial E_0) = \chi(\partial E_1) = \chi(\partial E_2),$$

so that

$$\chi(E_0) = \chi(E_1) = \chi(E_2),$$

and

$$\chi(\mathbb{S}^3 \setminus E_0) = \chi(\mathbb{S}^3 \setminus E_1) = \chi(\mathbb{S}^3 \setminus E_2).$$

On the other hand, E_0 is not \mathbb{R}^3 -ambient isotopic to E_1 (and to E_2), and E_1 is not \mathbb{R}^3 -ambient isotopic to E_2 .

In the next section we shall be concerned with invariants attached to a labelled apparent contour, under ambient isotopies of \mathbb{R}^2 and also of \mathbb{R}^3 , in the sense specified at the beginning of Sect. 7.4.

¹⁹Suppose that \mathcal{M} is a compact oriented connected three-manifold with boundary, and consider the double $D(\mathcal{M})$ of \mathcal{M} , obtained by identifying $\partial\mathcal{M}$ with the boundary of a copy $-\mathcal{M}$ of \mathcal{M} , with opposite orientation. Then [37, p. 261] $\chi(D(\mathcal{M})) = 0$. On the other hand, it is possible to show that $\chi(D(\mathcal{M})) = 2\chi(\mathcal{M}) - \chi(\partial\mathcal{M})$.

7.5 Cell Complexes and Fundamental Groups

In knot theory one of the most important invariants²⁰ is the topology of the complement of a (tame) knot in \mathbb{R}^3 , and in particular its fundamental group π_1 . This suggests to focus our attention on the topological properties of the three sets E , $\Sigma = \partial E$ and $\mathbb{R}^3 \setminus E$, the solid set (the inside), its boundary and its complement in \mathbb{R}^3 (the outside). Of these, the less interesting is Σ , since its topology is completely determined by its Euler–Poincaré characteristic $\chi(\Sigma)$. Indeed, as it is well known, all closed two-dimensional manifolds are completely classified [21, 39], and in particular an oriented manifold can only be a sphere with zero or more “handles” (a torus, for example, has the topology of a sphere with one handle). Also note that the fundamental group of the outside $\mathbb{R}^3 \setminus E$ does not change if we compactify \mathbb{R}^3 and consider $\mathbb{S}^3 \setminus E$,²¹ so that in all considerations concerning the fundamental group we shall often interchangeably use the most convenient one.

If $\Sigma = \partial E$ consists of a single connected component, we can distinguish the following cases based on the genus g of the surface,²² i.e., Σ is a topological sphere with g handles.

- $g = 0$: The surface $\Sigma = \partial E$ is a topological sphere (smoothly embedded in \mathbb{R}^3). The generalized version of the Schoenflies theorem [12] (see also, e.g., [31]) implies that E is topologically a solid sphere (a 3-ball). This is also true for the outside, provided we compactify $\mathbb{R}^3 \subset \mathbb{S}^3$ by adding the point at infinity. As a consequence both E and $\mathbb{S}^3 \setminus E$ are contractible sets and hence their first fundamental groups are trivial. Removal of the point at infinity from $\mathbb{S}^3 \setminus E$ obtaining $\mathbb{R}^3 \setminus E$ does not change its fundamental group, although $\mathbb{R}^3 \setminus E$ is no longer contractible.
- $g = 1$: The surface $\Sigma = \partial E$ is now a topological torus. Working in \mathbb{S}^3 it can be shown (we refer, e.g., to [38, p. 107]) that Σ bounds a solid torus on either side (or both). This means that either E is a solid torus, hence a (deformed) tubular neighbourhood of a knot in \mathbb{R}^3 , or its complement is a solid torus in \mathbb{S}^3 , or both. The latter case corresponds to the so-called *unknot*, a knot that can be deformed into a circle. An example of the first case is illustrated in Fig. 10.18 (the trefoil knot), an example of the second case is illustrated in Fig. 10.20. The fundamental groups of E and $\mathbb{R}^3 \setminus E$ allow us to identify the three cases; the unknotting theorem [38, p. 103] ensures that having the infinite cyclic group \mathbb{Z} as fundamental group of both the inside and the outside indicates that Σ is ambient isotopic to the standard torus. Observe that in any case either E or $\mathbb{S}^3 \setminus E$ can be retracted onto a one-dimensional set (an \mathbb{S}^1). Both the inside and the outside have a fundamental group

²⁰Under \mathbb{R}^3 -ambient isotopies with compact support.

²¹Beware however that $\mathbb{R}^3 \setminus E$ and $\mathbb{S}^3 \setminus E$ are not, in general, homotopically equivalent; for example, the complement of a solid sphere in \mathbb{S}^3 is contractible, whereas the complement in \mathbb{R}^3 is not.

²²The Euler–Poincaré characteristic and the genus are related by $\chi(\Sigma) = 2 - 2g$.

having infinite cyclic commutator quotient (the abelianized is the infinite cyclic group \mathbb{Z}).

- $g > 1$: In this case we cannot say much about E or $\mathbb{S}^3 \setminus E$. Sometimes E can be retracted onto a one-dimensional set, a bouquet of more than one possibly entangled loops (see Fig. 10.6) sometimes this is true for the complement in \mathbb{S}^3 (see Fig. 10.37), sometimes this is not true for either (see Fig. 10.38).

The picture above is however not complete. A connected component C of (say) E (all considerations are symmetrically valid for $\mathbb{S}^3 \setminus E$) might have a nonconnected boundary consisting of some of the connected components of Σ , for example if Σ is made of three concentric spheres, the “inside” E will consist of two connected (solid) components: an internal solid ball and a surrounding hollow sphere, the latter has a boundary consisting of two of the three spheres composing Σ . This example is not that interesting since the fundamental group of C is insensitive to the presence of a connected component of the boundary with the topology of the sphere (because the complement of that sphere is contractible due to the three-dimensional version of the Schoenflies theorem [12]).

On the contrary, if there are more than one component of ∂C with positive genus, then the fundamental group is greatly influenced by the relative position of the boundary components. As an example consider a scene E consisting of r solid tori obtained as tubular neighbourhood of a (tame) link with r components in \mathbb{R}^3 . The fundamental group of the outside $\mathbb{S}^3 \setminus E$ is the *link group* and carries important information on the topology of the link, often allowing to recognize links that cannot be split by ambient isotopy. As an illustration, see Fig. 10.22 for three examples of such situation. Of course we could have even more involved situations, such as the one illustrated in Fig. 7.4. It has a fundamental group that can be presented as

$$\langle x_1, x_2, x_3; x_1 x_2 x_3 x_1^{-1} x_2^{-1} x_3^{-1} \rangle$$

having \mathbb{Z}^3 , the free abelian group of rank 3, as its abelianization.

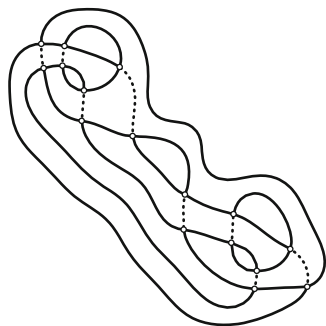
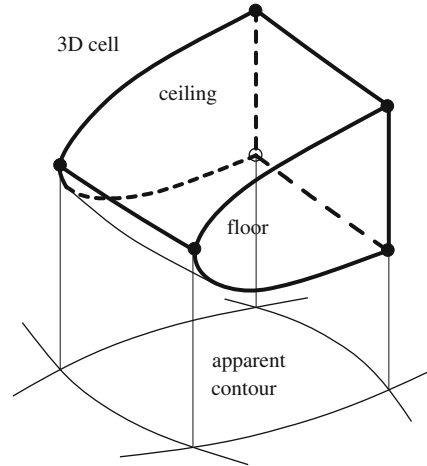


Fig. 7.4 A torus and a double torus in a mutually linked position

Fig. 7.5 A typical 3D cell is bounded below and above by two strata corresponding to a region of the apparent contour G_Σ . We call these two faces *floor* and *ceiling*. Arcs of the apparent contour may give rise to two-dimensional cells; for instance, the vertical front face arises from the corresponding arc. Nodes of the apparent contour may give rise to one-dimensional cells: for instance, the right *bold vertical* segment arises from the corresponding node



7.5.1 Cell Complexes

We observe that each of the three sets E , Σ , $\mathbb{R}^3 \setminus E$ can be decomposed in cells to form a CW complex [24]. The surface Σ is a two-dimensional complex, whereas the two solid sets E and $\mathbb{R}^3 \setminus E$ are three-dimensional complexes.

For definiteness, let us focus our attention on the solid set E . The construction of the complex, which is then implemented in the program `appcontour` (Sect. 10.7, Chap. 10), proceeds as follows. Each element of the apparent contour (regions, arcs, crossings, cusps) can be lifted in \mathbb{R}^3 to produce one or more cells of dimension three (regions), two (regions and arcs), one (arcs and crossings), zero (crossings and cusps). For instance, a three-dimensional cell is a solid portion of \mathbb{R}^3 bounded from below and above by two strata (floor and ceiling) of the surface that project onto a region R of the apparent contour (see Fig. 7.5), and bounded on the sides by “walls” that project onto the arcs that bound R .

We proceed in a similar manner for the “outside” $\mathbb{R}^3 \setminus E$. In order to avoid problems with infinite cells, we fix a sufficiently large ball B containing the solid set E , and $\mathbb{R}^3 \setminus E$ will be replaced by $B \setminus E$; this has no effect on the homotopy type. The apparent contour of $B \setminus E$ can be easily obtained by placing the original apparent contour inside a large \mathbb{S}^1 and increasing by one the original labelling.

In order to meet the definition of a CW complex it is sometimes necessary to introduce *cutting* elements to the apparent contour; in particular, this is necessary for regions that are not simply connected: one cutting arc for each hole of the region, and for closed arcs: one cutting node, that is actually an endpoint of a cutting arc.

Remark 7.5.1 The resulting cell complex is not necessarily connected, even for a connected E . For instance, the outside of a hollow sphere has two connected components, one homotopic to a point (the internal void) and the other homotopic to an \mathbb{S}^2 (the outside of the “filled” sphere).

A key property of the 3D cells is that they are never glued together at floors or ceilings.²³ In other words, a ceiling of a 3D cell does not bound any other 3D cell. This allows us to perform a deformation retract that carries the original complex to a new one in which both the 3D cell and its ceiling have been removed. By repeating this procedure we can substitute the original complex with a homotopically equivalent two-dimensional one.

This fact is crucial, since it is much easier to encode the adjacency information for a 2D cell complex: after arbitrarily choosing an orientation for faces and arcs, the boundary of a face is simply an ordered list of 1D cells (arcs) each with a sign indicating whether the arc is positively or negatively oriented with respect to the face; the boundary of an arc is just an ordered pair of 0D cells (nodes).

All properties of E (respectively $\mathbb{R}^3 \setminus E$) up to homotopy are now completely encoded in the combinatorial structure of the corresponding cell complex, in particular so is its fundamental group π_1 .

7.5.2 Fundamental Groups

The cell complex constructed in Sect. 7.5.1 is of little direct use, and in particular it is **not** invariant under \mathbb{R}^3 -ambient isotopies, since the construction is based on a choice of a projection direction, and it changes when we change the projection direction. However, its topological structure is indeed invariant under \mathbb{R}^3 -ambient isotopies, and in particular so are quantities like the number of connected components or (more interestingly) the first fundamental group π_1 of each connected component.

Two \mathbb{R}^3 -ambient isotopic scenes have homotopically equivalent “outsides” leading to isomorphic fundamental groups, in other words each of the fundamental groups (up to isomorphism) provides an invariant (a quite powerful one, actually) of the 3D scene E up to ambient isotopy. Consequently non-isomorphic fundamental groups correspond to necessarily non-equivalent scenes. On the contrary isomorphic fundamental groups do not entail²⁴ that the corresponding scenes are ambient isotopic. This is the case even in the special case of knots and links (modelled by using knotted tubes embedded in \mathbb{R}^3), for example the two chiral versions of the trefoil knot (Fig. 10.18) are not ambient isotopic, yet they lead to isomorphic fundamental groups.²⁵

²³The reason being that Σ separates E (the interior) from $\mathbb{R}^3 \setminus E$ (the exterior): the interior is below the ceiling, and consequently it cannot be above it at the same time.

²⁴There are exceptions, most notably the unknotting theorem [38, p. 103] asserts that “trivial” fundamental groups for Σ with the topology of a torus imply that the scene is ambient isotopic to the standard solid torus.

²⁵Proving that the trefoil knot cannot be deformed into its specular version, although apparently obvious, requires quite sophisticated techniques, beyond the scope of this book.

For definiteness, let us focus attention on the solid set $\mathbb{R}^3 \setminus E$ (outside).²⁶ For simplicity we shall assume that it is connected, otherwise (the hollow sphere is an example) all subsequent considerations are to be carried out separately for each connected component.

From the constructed cell complex we can obtain a *presentation* of its fundamental group by using standard procedures (see, e.g., [24]). For a two-dimensional cell complex (our case) this amounts in finding a spanning tree for the one-skeleton, listing the remaining arcs as generators of the group and constructing a relator for each of the 2D faces. A presentation of a group G is written as $\langle x_1, \dots, x_n; r_1, \dots, r_m \rangle$ where x_1, \dots, x_n are specific elements (generators) of G that generate all other elements and r_1, \dots, r_m are words in the generators and their inverses that correspond to the unit element of G .²⁷

It should be noted, however, that dealing with a finite presentation of a group is far from conclusive: we inevitably incur in algorithmic problems, most notably the *isomorphism problem*²⁸ of deciding whether two different finite presentations describe isomorphic groups. More precisely, the isomorphism problem²⁹ has been proved to be undecidable: there does not exist a computer algorithm that correctly solves every instance of the isomorphism problem regardless of how much time is allowed for the algorithm to run [28].

The so-called Tietze transformations can be used to manipulate a finite presentation with the intent of transforming one presentation into the other. If this can be done, then we have isomorphic groups. Conversely proving that two presentations describe different groups amounts in finding some group invariants that differ, e.g., the dimension of the abelianization, which however gives very little information.

²⁶This is the most interesting choice in the case E is a knotted solid torus (tubular neighbourhood of a knot), or a union of knotted solid tori, tubular neighbourhood of a link. Indeed in this case the fundamental group of $\mathbb{R}^3 \setminus E$ is simply called the *knot group* (respectively *link group*), whereas the fundamental group of E simply reduces to \mathbb{Z} , the infinite cyclic group, for each connected component. Of course in our broader context we can imagine situations, such as the sphere with a knotted tunnel of Fig. 10.20, where the interesting solid set is E itself.

²⁷Strictly speaking, x_1, \dots, x_n are free generators of the free group F of rank n ; r_1, \dots, r_m are elements of F and G is the quotient $G = F/H$ where H is the smallest normal subgroup of F containing r_1, \dots, r_m .

²⁸Although as a matter of fact the isomorphism problem is decidable if restricted to special classes of finitely presented groups. Among these, interestingly, we also find the fundamental groups of three-manifolds. A quite interesting post on this subject by Henry Wilton can be found at the web address <http://ldtopology.wordpress.com/2010/01/26/3-manifold-groups-are-known-right/> (May 21, 2014).

²⁹The related *word problem* (respectively *conjugacy problem*) of deciding whether two words define the same element (respectively conjugate elements) in a finitely presented group is also undecidable in general.

7.6 Alexander Polynomials and Invariants of Fundamental Groups

Following [28], the isomorphism problem for finitely presented groups can be alleviated by the computation of invariants: objects that are invariant under Tietze transformations, so that they are independent of the actual presentation of the group. Two presentations having different invariants necessarily describe groups that are not isomorphic.

The simplest of such invariants is the *abelianized* of the group, obtained as the quotient G/G' of the group by its commutator, where G stands for the *finitely presented* fundamental group and G' is its commutator, defined as the smallest normal subgroup containing all elements of the form $aba^{-1}b^{-1}$. The resulting group, which turns out to be isomorphic to the first homology group of the set, does not provide any particularly new information, since it is a finitely generated free abelian group having rank r , i.e., isomorphic to \mathbb{Z}^r , that can be recovered using the Euler–Poincaré characteristic of each of the connected components of the surface $\Sigma = \partial E$ that are adjacent to the connected component C of $\mathbb{R}^3 \setminus E$ under consideration.³⁰

Direct computation of the rank r from a presentation of the fundamental group of C requires the construction of a so-called *preabelian* presentation, with a procedure that resembles gaussian elimination of matrices [28, Section 3.3].

For a general presentation we can construct an integer-valued matrix, the so-called *exponent sum* matrix, by considering the sum of the exponents in the relator r_j of the generator x_i , obtaining the element in row i , column j . If the number m of relators (number of columns of the exponent sum matrix) is less than the number n of generators (rows of the exponent sum matrix), it is customary to add $n - m$ zero-valued columns on the right, so that we can assume $m \geq n$. For a preabelian presentation this matrix is diagonal with non-negative diagonal elements d_1, \dots, d_n satisfying the property that d_i divides d_{i+1} , $i \in \{1, \dots, n-1\}$. This is the so-called *Smith normal form* of the matrix.³¹ In particular we have that $d_i = 1$ for $i \in \{1, \dots, i_0\}$, $d_i > 1$ for $i = i_0 + 1, \dots, i_1$, $d_i = 0$ for $i \in \{i_1 + 1, \dots, n\}$, with appropriate choice of the values $0 \leq i_0 \leq i_1$. The diagonal entries different from 0 and 1 are the so-called invariant factors, and their presence gives rise to the so-called torsion part of the abelianized group; hence, if $i_0 = i_1$, there is no torsion part. The number $r = n - i_1$ of zeroes in the diagonal is the rank of the abelianized G/G' .

The structure of G/G' can be directly read from the diagonal entries of the exponent sum matrix of (any) preabelian presentation as follows.

³⁰The rank r is actually equal to the Betti number b_1 of the component C that we are considering. The other nonzero Betti numbers are b_0 , which is equal to 1, since we are restricting to a single connected component of $\mathbb{R}^3 \setminus E$, and b_2 : the number of “voids” (cavities) in C , equal to the number of connected components of the complement of C (which is also the number of connected components of Σ adjacent to C) decreased by one. The Euler–Poincaré characteristic of C is given by $b_0 - b_1 + b_2$.

³¹It can be defined for any matrix with entries in a principal ideal domain.

- The entries equal to 1 do not count; they correspond to generators contained in the commutator G' , and vanish in the quotient G/G' .
- Each invariant factor $d_i > 1$ corresponds to a factor of G/G' isomorphic to the cyclic group \mathbb{Z}_{d_i} of order d_i .
- Each zero entry in the diagonal corresponds to a factor isomorphic to the free cyclic group \mathbb{Z} .

Constructing a preabelian presentation is also a crucial step towards the computation of a more sophisticated invariant, the *Alexander polynomial*. This is a widely known knot invariant (see, for instance, [15, 16]), often computed starting from a knot diagram. In our context a knot is substituted by the surface Σ bounding a knotted solid torus E ; since we are not aware of any easy way to recover the knot diagram from the apparent contour of a (possibly widely) deformed knotted solid torus, we cannot take advantage of techniques based on the knot diagram. However, it turns out that the Alexander polynomial can be viewed actually as an invariant of the fundamental group (this is not true for other important knot invariants, such as the Jones polynomial). Following [28, Section 3.4], in the special case where the abelianized G/G' of the fundamental group G is isomorphic to the infinite cyclic group \mathbb{Z} we have $i_0 = i_1 = n - 1$ and the Alexander polynomial is computed as an invariant of the group G'/G'' , i.e., of the abelianized of the commutator subgroup of G . Unfortunately, even for a finitely presented G , its commutator G' is not in general finitely presented, as it might require countably many generators; however, a finite presentation can be recovered in the more general context of modules over the ring L of Laurent polynomials with integer coefficients in one indeterminate, i.e., polynomials where the indeterminate is allowed to have a negative (integral) exponent.

If G is the fundamental group of a knot it turns out that there is a preabelian presentation of G with one less relator than generators (deficiency equal to one); in this case, the exponent sum matrix of the corresponding presentation of G' in the context of L -modules turns out to be a square $(n-1) \times (n-1)$ matrix having Laurent polynomials with integer coefficients as elements. The Alexander polynomial Δ is its determinant.³² A straightforward computation shows that evaluation at 1 of the elements of the $(n-1) \times (n-1)$ matrix produces the $(n-1) \times (n-1)$ principal minor of the exponent sum matrix of the preabelian presentation. Having G/G' isomorphic to \mathbb{Z} we conclude that it is the identity matrix, so that $\Delta(1) = 1$.

If t denotes the indeterminate of the Laurent polynomials in L , it turns out that the Alexander polynomial is defined up to multiplication/division by t and up to the transformation $t \rightarrow 1/t$.³³ Other than that, different Alexander polynomials

³²In [20] the emphasis is given to the ideal of L generated by the Alexander polynomial, indicated by ε_1 , see Sect. 7.7.

³³The mapping $t \rightarrow 1/t$ corresponds to the automorphism of the ring \mathbb{Z} mapping the (multiplicative) generator t onto its inverse. This is the unique nontrivial automorphism of \mathbb{Z} . Interestingly, it turns out that the Alexander polynomial of a knot is invariant under such a transformation, up to multiplication by a power of t ; this is not the case for a generic finitely presented group with

denote different (non-isomorphic) finitely presented groups, which in turn implies non-equivalency (up to \mathbb{R}^3 -ambient isotopies) of the original apparent contours.

7.7 Free Differential Calculus

For a finitely presented group that does not have the infinite cyclic group \mathbb{Z} as its abelianization, the construction of the Alexander polynomial described in Sect. 7.6 does not apply. This is the case, e.g., for a link group, which in our context corresponds to E consisting of a number of solid tori, possibly mutually entangled. It also happens when $\Sigma = \partial E$ is a connected surface with genus $g > 1$.

The approach based on the *free differential calculus* of Fox [18, 19] provides a way to construct invariants of a finitely presented group G that can be successfully applied in the more general case $G/G' \cong \mathbb{Z}^r$. For $r = 1$ we obtain the same Alexander polynomial described in Sect. 7.6.

For a fixed n we shall denote by X be the free group with n generators x_1, \dots, x_n . It is composed of words in $x_1, \dots, x_n, x_1^{-1}, \dots, x_n^{-1}$ that can be *multiplied* by juxtaposition. The special empty word, denoted by 1, is the unit element of X . Words can be *reduced* by removing adjacent occurrences of a generator and its inverse; a *reduced* word is a word that cannot be reduced.

Let $G = \langle x_1, \dots, x_n; r_1, \dots, r_m \rangle = X/H$ be the finitely presented group obtained by choosing H as the smallest normal subgroup of X containing the m relators $r_1, \dots, r_m \in X$. The projection map will be denoted by

$$\phi : X \rightarrow G = X/H.$$

With $\mathbb{Z}X$ we denote the group ring of X , i.e., the ring containing the finite formal linear combinations of elements of X with integral coefficients.

Following [18], the *free partial derivatives*

$$\frac{\partial}{\partial x_j} : X \rightarrow \mathbb{Z}X, \quad j \in \{1, \dots, n\},$$

infinite cyclic commutator quotient. The symmetry of the coefficients of the Alexander polynomial is a nontrivial fact, consequence of the Poincaré Duality isomorphism. It is known that any Laurent polynomial having symmetric coefficients and that evaluates to ± 1 for $t = 1$ is the Alexander polynomial of some knot [25].

with respect to the generators x_1, \dots, x_n , are defined by the identities

$$\begin{aligned}\frac{\partial 1}{\partial x_j} &= 0, \\ \frac{\partial x_i}{\partial x_j} &= \delta_{ij}, \quad i, j \in \{1, \dots, n\}, \\ \frac{\partial(ab)}{\partial x_j} &= \frac{\partial a}{\partial x_j} + a \frac{\partial b}{\partial x_j}, \quad a, b \in X.\end{aligned}$$

From the identities above, by differentiating $x_i x_i^{-1} = 1$ it follows that

$$\frac{\partial x_i^{-1}}{\partial x_j} = \begin{cases} -x_i^{-1} & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

The free partial derivative is then extended by linearity on the group ring $\mathbb{Z}X$ as

$$\frac{\partial}{\partial x_j} \sum_{a \in X} \alpha_a a = \sum_{a \in X} \alpha_a \frac{\partial a}{\partial x_j},$$

with $\alpha_a \in \mathbb{Z}$ and $\alpha_a = 0$ for all but a finite number of elements $a \in X$.

It turns out that the free partial derivative of a reduced word $a \in X$ with respect to the generator x_j is a signed sum of elements $\hat{a}_k \in X$, one for each occurrence of x_j or x_j^{-1} in a , negative in the latter case, with \hat{a}_k consisting of the initial segment of a up to the relevant occurrence of x_j excluded if the exponent of x_j is 1, included if the exponent of x_j is -1 .

We clarify this with an example. Take $a = x_1 x_2 x_1^{-1} x_2^{-1}$; then

$$\frac{\partial a}{\partial x_1} = 1 - x_1 x_2 x_1^{-1}, \quad \frac{\partial a}{\partial x_2} = x_1 - x_1 x_2 x_1^{-1} x_2^{-1}.$$

Following [19], the Jacobian J of the presentation $\langle x_1, \dots, x_n; r_1, \dots, r_m \rangle$ of G is the $(m \times n)$ -matrix containing all the free partial derivatives of the m relators with respect to the n generators,

$$J = \begin{bmatrix} \frac{\partial r_1}{\partial x_1} & \dots & \frac{\partial r_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial r_m}{\partial x_1} & \dots & \frac{\partial r_m}{\partial x_n} \end{bmatrix}.$$

With the notation J^ϕ we indicate that the matrix entries are to be interpreted as elements of $\mathbb{Z}G$.

Remark 7.7.1 (Dependence on the Presentation) The Jacobian matrix depends on the presentation of G and a change of presentation induces a modification of the matrix. In particular, the Tietze elementary transformations induce the following actions on the Jacobian matrix:

- exchange of two rows or two columns;
- add to a row a left-multiple³⁴ of another row;
- add to a column a right-multiple of another column;
- add a new zero-valued row;
- add a new row and a new column with all zero-valued entries except the entry corresponding to the intersection of the new row and column, where a 1 (unit element of G) is placed.

Since $\mathbb{Z}G$ is not abelian, working directly with the Jacobian matrix is impractical. In order to make the computation manageable it is convenient to project the matrix onto a simpler one by using a projection of G onto an abelian group. Two of such maps are particularly useful: the *trivial map*

$$o : G \rightarrow \{1\},$$

projection of G on the trivial quotient group G/G , and the abelianizing map

$$\psi : G \rightarrow G/G'$$

with G' being, as usual, the commutator subgroup. From now on, we shall assume that the quotient G/G' is isomorphic to \mathbb{Z}^r for some $r \in \mathbb{N}$.³⁵

Under the trivial map o , the Jacobian matrix is mapped onto an integral-valued matrix that, after transposition, coincides with the exponent sum matrix introduced in Sect. 7.6. It allows to identify the structure of the commutator quotient G/G' .

Under the abelianizing map ψ , the entries of the Jacobian matrix are mapped onto the commutative ring $\mathbb{Z}(G/G')$, isomorphic to the ring L of the Laurent polynomials in r indeterminates and with integral coefficients.

³⁴Recall that $\mathbb{Z}X$ and $\mathbb{Z}G$ are noncommutative rings.

³⁵This is always the case for G the first fundamental group of sets of the form $\Sigma = \partial E$, E or $\mathbb{R}^3 \setminus E$. This follows using the Mayer–Vietoris exact sequence [24] on the two solid sets E , $\mathbb{R}^3 \setminus E$, and their common boundary Σ . Indeed, let $\rho > 0$ and consider the open sets $E_\rho^+ := \{x \in \mathbb{R}^3 : \text{dist}(x, E) < \rho\}$ and $E_\rho^- := \{x \in \mathbb{R}^3 : \text{dist}(x, \mathbb{R}^3 \setminus E) > \rho\}$, so that $\mathbb{R}^3 = E_\rho^+ \cup (\mathbb{R}^3 \setminus E_\rho^-)$. We have the short exact sequence $0 = H_2(\mathbb{R}^3) \rightarrow H_1(E_\rho^+ \cap (\mathbb{R}^3 \setminus E_\rho^-)) \rightarrow H_1(E_\rho^+) \oplus H_1(\mathbb{R}^3 \setminus E_\rho^-) \rightarrow H_1(\mathbb{R}^3) = 0$. Hence $H_1(E_\rho^+ \cap (\mathbb{R}^3 \setminus E_\rho^-))$ and $H_1(E_\rho^+) \oplus H_1(\mathbb{R}^3 \setminus E_\rho^-)$ are isomorphic. Since $H_1(\partial E)$ is (for $\rho > 0$ sufficiently small) isomorphic to $H_1(E_\rho^+ \cap (\mathbb{R}^3 \setminus E_\rho^-))$ which is a direct product of copies of \mathbb{Z} , it follows that also $H_1(E_\rho^+)$ and $H_1(\mathbb{R}^3 \setminus E_\rho^-)$ are direct products of copies of \mathbb{Z} , and the assertion follows.

Definition 7.7.2 (Elementary Ideals) For $d \in \{0, \dots, n-1\}$ the d -th *elementary ideal* ε_d is the ideal of L generated by all minor determinants of order $n-d$ of the projected Jacobian matrix $J^{\psi\phi}$. They form an ascending chain $\varepsilon_0 \subseteq \varepsilon_1 \subseteq \dots \subseteq \varepsilon_{n-1}$. It is customary to define ε_d for $d < 0$ as $\varepsilon_d := \{0\}$ and for $d \geq n$ as $\varepsilon_d := L$.

The elementary ideals are invariant under the Jacobian matrix modifications induced by the Tietze transformations, so that ultimately they are independent of the presentation and are invariants of G . *They however depend on how we choose the isomorphism between G/G' and \mathbb{Z}^r .*

- For simplicity, from now on we shall restrict our analysis to the case of rank $r = 2$; however, most of the results extend easily to the general case.

Remark 7.7.3 (Laurent Polynomials) The ring L of Laurent polynomials in two indeterminates u, v and integral coefficients is a commutative ring whose invertible elements (units) are all monomials of the form $\pm u^\alpha v^\beta$. It is not a *principal domain* ring; it follows that the elementary ideals ε_d are not, in general, principal ideals. The first nontrivial ideal ε_1 in the case of a knot group is however a principal ideal generated by the Alexander polynomial.

The dependence of the elementary ideals on the actual choice of the isomorphism $G/G' \cong \mathbb{Z}^2$ entails that they are not by themselves invariants of G . In order to compute an invariant we need to identify all possible such isomorphisms. In other words we need to find all automorphisms of \mathbb{Z}^2 and consider as equivalent ideals that correspond under such an automorphism. In the special case $r = 1$ (fundamental group of a knot) there is only one nontrivial automorphism of \mathbb{Z} , the one mapping the generator t onto its inverse $1/t$ as already observed at the end of Sect. 7.6. For $r \geq 2$ the situation is not so simple; in what follows we shall briefly investigate the subject for the special case $r = 2$.

Remark 7.7.4 (Automorphisms of \mathbb{Z}^2) With abuse of notation we shall still denote by ψ the composite map ψ followed by the isomorphism between G/G' and \mathbb{Z}^2 ;³⁶ in this way, the elementary ideals become ideals of the ring L . However they will depend on the choice of the isomorphism between G/G' and \mathbb{Z}^2 : different isomorphisms are related to each other by an automorphism of the group \mathbb{Z}^2 onto itself, or, equivalently, by a change of base of \mathbb{Z}^2 . A generic change of base is of the form $u_B = u^\alpha v^\beta$, $v_B = u^\gamma v^\delta$ where

$$B = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix}, \quad B^{-1} = \begin{bmatrix} \hat{\alpha} & \hat{\gamma} \\ \hat{\beta} & \hat{\delta} \end{bmatrix}. \quad (7.14)$$

Here B is a matrix in $\text{GL}(2, \mathbb{Z})$ with integral entries and determinant ± 1 , B^{-1} is its inverse and u, v are a base of \mathbb{Z}^2 . We use here multiplicative notation, for consistency

³⁶Strictly speaking, the isomorphism between the corresponding group rings induced by the isomorphism between G/G' and \mathbb{Z}^2 .

with the usual notation for Laurent polynomials in two indeterminates u and v . The original base u, v can be recovered from u_B, v_B using the inverse matrix B^{-1} as $u = u_B^{\hat{\alpha}} v_B^{\hat{\beta}}, v = u_B^{\hat{\gamma}} v_B^{\hat{\delta}}$.

Definition 7.7.5 (Change of Base) Given $B, B^{-1} \in \text{GL}(2, \mathbb{Z})$ as in (7.14) the “change of variables” induced by the change of base defined by B is defined for $p \in L$ as

$$p_B(u, v) = p(u^\alpha v^\beta, u^\gamma v^\delta).$$

Clearly we can recover the original polynomial with $p = (p_B)_{B^{-1}}$. If $\Lambda \subseteq L$, then we denote by Λ_B the corresponding set $\Lambda_B = \{p_B : p \in \Lambda\}$.

Proposition 7.7.6 *If $B \in \text{GL}(2, \mathbb{Z})$ denotes a change of variables and $p, q \in L$, then $(p + q)_B = p_B + q_B$ and $(pq)_B = p_B q_B$. If I is an ideal of L , then I_B is itself an ideal of L .*

Proof The first two assertions are a direct check. That I_B is closed under addition and multiplication follows from the first two assertions. If $p \in I$ and $q \in L$, set $\hat{q} = q_{B^{-1}}$. Then $qp_B = (q_{B^{-1}}p)_B \in I_B$ since I is an ideal. \square

Definition 7.7.7 (Equivalent Ideals) We say that two ideals I, \hat{I} of L are equivalent if there exists a change of base B such that $\hat{I} = I_B$.

In the special case where $I = (p)$ is a principal ideal, generated by $p \in L$, an equivalent ideal \hat{I} is itself principal and it is generated by a polynomial q that is related to p by a change of base and a unit of L , which is to say that

$$q = \pm m p_B \tag{7.15}$$

where B denotes a change of base and m is a monomial $u^s v^t$, $s, t \in \mathbb{Z}$. The combination “change of base” and “multiplication by m ” can be interpreted as an affine invertible transformation in the group \mathbb{Z}^2 . This justifies the equivalence relation defined as follows.

Definition 7.7.8 (Base-Equivalent Polynomials) Two polynomials p, q are *base-equivalent* if there exists a change of base B , a monomial m and $\varepsilon = \pm 1$ such that p and q are related by (7.15)

Remark 7.7.9 In order to compute the projected Jacobian matrix $J^{\psi\phi}$ we first need to identify a base for G/G' and compute the image of the generators x_1, \dots, x_n in terms of that base with respect to the projection ψ . This can be done by preliminarily computing a preabelian presentation of G . Indeed for a preabelian presentation, the first $n - 2$ generators are all mapped into the identity element of G/G' , whereas the last 2, once projected, are generators of G/G' (hence also providing an isomorphism

of G/G' with \mathbb{Z}^2),³⁷ they will be the two indeterminates of the Laurent polynomial in L . This remark generalizes in a straightforward way to the generic case of rank r .

Finally, we need to introduce the *fundamental ideal*.

Definition 7.7.10 (Fundamental Ideal) The fundamental ideal $\mathcal{E} \subseteq L$ is the ideal generated by $u - 1$ and $v - 1$.

Lemma 7.7.11 *The fundamental ideal \mathcal{E} is characterized by*

$$\mathcal{E} = \{p \in L : p(1, 1) = 0\}.$$

Proof The inclusion $\mathcal{E} \subseteq \{p \in L : p(1, 1) = 0\}$ is clear. On the other hand, take $p \in L$ such that $p(1, 1) = 0$. After possibly multiplying p by a suitable unit monomial $u^\alpha v^\beta$, we can assume that all exponents are nonnegative. We now think of p as a polynomial in the indeterminate v having coefficients that are polynomials in u (with integral coefficients). Euclidean division of p by $v - 1$ does not require any division (direct check), so that it can be carried out in the ring of polynomial in v with coefficients that are polynomial in u . We obtain

$$p(u, v) = q(u, v)(v - 1) + r(u),$$

where $q \in L$ and the rest r has forcibly degree 0 with respect to v since the divisor $v - 1$ has degree one. Requirement $p(1, 1) = 0$ now implies that $r(1) = 0$, so that r is divisible by $u - 1$ with a quotient $s(u)$ having integral coefficients since again the euclidean division by $u - 1$ does not require any division. We end up with

$$p(u, v) = q(u, v)(v - 1) + s(u)(u - 1),$$

and the proof is complete. \square

Since evaluation in $(1, 1)$ of a polynomial is invariant under a change of base of \mathbb{Z}^2 we immediately have:

Corollary 7.7.12 (Invariance of \mathcal{E}) *\mathcal{E} is invariant under a change of base of \mathbb{Z}^2 .*

³⁷Note that when the presentation is directly obtained from a knot/link diagram using the Wirtinger technique [38], then the relation of the generators with their projection is easily obtained: all generators associated with the same link component project to the same element, whereas generators associated with different link component (one per component) project onto a base of the quotient group.

7.8 Links with Two Components: Deficiency One

In the case of a link of two components, we have a finitely presented group with a presentation of deficiency 1 and commutator quotient of rank 2, for example as the fundamental group of the outside of two entangled solid tori.³⁸

The Jacobian matrix J is, in this case, rectangular with one more column than rows. The presentation $\langle x_1, \dots, x_n; r_1, \dots, r_{n-1} \rangle$ can be supposed to be preabelian, meaning that the projected Jacobian $J^{\circ\phi}$ has elements

$$\left(\frac{\partial r_i}{\partial x_j} \right)^{\circ} = \left(\frac{\partial r_i}{\partial x_j} \right)^{\circ\phi} = \begin{cases} 1 & \text{if } i \leq n-2 \text{ and } i = j, \\ 0 & \text{otherwise.} \end{cases} \quad (7.16)$$

The isomorphism $\psi : G/G' \rightarrow \mathbb{Z}^2$ can then be chosen so that

$$\psi(x_{n-1}) = u, \quad \psi(x_n) = v, \quad \psi(x_j) = 1, \quad j \in \{1, \dots, n-2\}, \quad (7.17)$$

with u, v generators of \mathbb{Z}^2 .

Theorem 7.8.1 *Let G be a finitely presented group with $G/G' \cong \mathbb{Z}^2$ with a preabelian presentation $G = \langle x_1, \dots, x_n; r_1, \dots, r_{n-1} \rangle$ of deficiency 1. Define ψ as in (7.17). Then there exist $q_1, \dots, q_{n-1} \in L$ such that*

$$\begin{aligned} \left(\frac{\partial r_i}{\partial x_{n-1}} \right)^{\psi\phi} &= q_i(u, v)(v-1), \\ \left(\frac{\partial r_i}{\partial x_n} \right)^{\psi\phi} &= -q_i(u, v)(u-1), \quad i \in \{1, \dots, n-1\}. \end{aligned} \quad (7.18)$$

In particular, the elementary ideal ε_1 is given by

$$\varepsilon_1 = (\Delta)\mathcal{E},$$

the product of the fundamental ideal and the principal ideal generated by a certain polynomial $\Delta \in L$ (Alexander polynomial) [20, p. 131].

Proof For any $i \in \{1, \dots, n-1\}$, using [18, equation (2.3)] we obtain the identity

$$r_i = 1 + \sum_{j=1}^{n-2} \frac{\partial r_i}{\partial x_j} (x_j - 1) + \frac{\partial r_i}{\partial x_{n-1}} (x_{n-1} - 1) + \frac{\partial r_i}{\partial x_n} (x_n - 1) \quad (7.19)$$

³⁸More generally this is the case for the inside and the outside of Σ made of two connected components of genus 1, i.e., two toric surfaces.

where r_i is viewed as an element of $\mathbb{Z}X$ (having $(r_i)^o = 1$). Direct computation using (7.16) leads to $(r_i)^\psi = 1$, moreover $(x_j - 1)^\psi = 0$ for all $j \in \{1, \dots, n-2\}$. Projecting (7.19) through ψ then leads to

$$\alpha(u, v)(u - 1) + \beta(u, v)(v - 1) = 0 \quad (7.20)$$

where $\alpha(u, v) = \left(\frac{\partial r_i}{\partial x_{n-1}}\right)^\psi \in L$ and $\beta(u, v) = \left(\frac{\partial r_i}{\partial x_n}\right)^\psi \in L$. Evaluating (7.20) in $v = 1$ (respectively in $u = 1$) shows that $\alpha(u, v)$ is divisible by $v - 1$ (respectively $\beta(u, v)$ is divisible by $u - 1$). Euclidean division by $u - 1$ and $v - 1$ is safe in L since it does not require divisions and we can conclude that

$$\alpha(u, v) = q_i(u, v)(v - 1), \quad \beta(u, v) = -q_i(u, v)(u - 1)$$

which is (7.18). The elementary ideal ε_1 is generated by the determinants of the n minors of order $n - 1$ of the $(n - 1) \times n$ projected Jacobian matrix. All minors containing both the last two columns have vanishing determinant, since the last two columns are multiples of the same column vector, so that we only need to compute two determinants, obtained by removing one of the last two columns. They are of the form $\Delta(u, v)(v - 1)$ and $-\Delta(u, v)(u - 1)$ respectively, where Δ is the determinant of the $(n - 1) \times (n - 1)$ matrix obtained by replacing the last two columns of the projected Jacobian with the column formed by the elements q_i , $i \in \{1, \dots, n - 1\}$, which concludes the proof. \square

Theorem 7.8.2 *The evaluation $|\Delta(1, 1)| \in \mathbb{N}$, where $\Delta \in L$ is defined in Theorem 7.8.1, is invariant under changes of base of \mathbb{Z}^2 .*

Proof This follows from Corollary 7.7.12 and the invariance of the evaluation in $(1, 1)$ of a Laurent polynomial under a change of base. \square

Following [20, p. 132] if G is the fundamental group of a two-components link and the generators x_{n-1} and x_n are meridians of each of the two links, then the evaluations $\Delta(u, 1)$ and $\Delta(1, v)$ verify

$$\Delta(u, 1) = (1 + u + u^2 + \dots + u^{q-1})\Delta_1(u)$$

$$\Delta(1, v) = (1 + v + v^2 + \dots + v^{q-1})\Delta_2(v)$$

where $q \in \mathbb{N}$ is the (absolute value of the) linking number between the two links and Δ_1, Δ_2 are the Alexander polynomials of the first and second components respectively (the two knots obtained by removing the other component).

For a generic preabelian presentation we cannot expect the last two generators to correspond to meridians of the link, however in view of the invariance result given by Theorem 7.8.2 we can nevertheless recover the linking number of the two-components link from G as follows.

Corollary 7.8.3 *The absolute value q of the linking number between the two components of a two-components link can be recovered from the elementary ideal $\varepsilon_1 = (\Delta)\mathcal{E}$ as*

$$q = |\Delta(1, 1)|$$

Proof The result is true for the particular presentation where the two generators u, v of \mathbb{Z}^2 correspond to meridians of the two components; then, we conclude in view of the invariance Theorem 7.8.2. \square

Remark 7.8.4 The elementary ideal ε_1 is **not** an invariant of G by itself, rather we must take into account the possible changes of base of \mathbb{Z}^2 , and consider the whole equivalence class $[\varepsilon_1]$ of ε_1 given by Definition 7.7.7. A practical consequence is that in order to conclude that two finitely presented groups are not isomorphic we need a way to ensure that the computed elementary ideals ε_1 are not base-equivalent. Computation of a *canonical representative* in the equivalence class $[\varepsilon_1]$ clearly solves the problem. This is easily done in the case of Alexander polynomials in one indeterminate, however for polynomials in two or more indeterminates practical computation of a canonical representative is not straightforward since there are an infinite number of different changes of base, so we cannot directly compute all representatives and select an optimal one using a lexicographic comparison technique. This will be discussed in Sect. 10.7.4.

Remark 7.8.5 The free group of rank 2 is a special case, in this case the elementary ideal is trivially $\varepsilon_1 = L$. The reverse implication is not to our knowledge true, so that computing the first elementary ideal is not sufficient in proving that the fundamental group is the free group. However, there is an *unknotting Theorem* also in the case of links [29, Theorem 4.2],³⁹ so that knowing that the fundamental group is free is sufficient to conclude that the link is split (made of “far away” copies of a perfect circle).

7.9 Surfaces with Genus 2: Deficiency Two

Computation of the fundamental group of each side of a surface Σ having genus 2 will also lead to a finitely presented group G with $G/G' \cong \mathbb{Z}^2$, now with a presentation with deficiency 2.

Let $G = \langle x_1, \dots, x_n; r_1, \dots, r_{n-2} \rangle$ be a preabelian presentation. Deficiency 2 implies that the elementary ideal ε_1 is trivial and gives no information, so that we shall compute the subsequent elementary ideal ε_2 . We proceed similarly to Sect. 7.8, but now we have an $(n - 2) \times n$ projected Jacobian matrix, having the

³⁹The unknotting Theorem is valid more generally for links with m components, in which case the fundamental group is the free group of rank m .

last two columns with elements of the form $q_i(u, v)(v - 1)$, $i \in \{1, \dots, n - 2\}$ and $-q_i(u, v)(u - 1)$, $i \in \{1, \dots, n - 2\}$ respectively.

There are $\binom{n}{2}$ minors of order $n - 2$, however most of them (those containing both the last two columns) have zero determinant. One of the remaining minors is given by the first $n - 2$ columns, let's call $w \in L$ its determinant. Recalling the form of the trivial projection $(J)^{o\phi}$ for a preabelian presentation we immediately deduce that $w(1, 1) = 1$, so that $w \notin \mathcal{E}$.

For any choice of one of the first $n - 2$ columns we have two minors obtained by removing that column and one of the last two. From the structure of the last two columns however, the determinant can be computed by substituting the column vector formed by the q_i , $i \in \{1, \dots, n - 2\}$ in place of the chosen column, call w_i its determinant, and then multiplying w_i times $v - 1$ or $u - 1$ respectively.

Summarizing, the elementary ideal ε_2 is generated by the principal ideal (w) and the ideal \mathcal{W} obtained by multiplying the fundamental ideal times the ideal generated by w_1, \dots, w_{n-2}

$$\mathcal{W} = (w_1, \dots, w_{n-2})\mathcal{E}, \quad \varepsilon_2 = (w, \mathcal{W}).$$

Remark 7.9.1 Again, the elementary ideal ε_2 is **not** an invariant of G by itself, rather we must take into account the possible changes of base of \mathbb{Z}^2 , and consider the whole equivalence class $[\varepsilon_2]$ of ε_2 given by Definition 7.7.7.

Remark 7.9.2 Deciding whether or not two non-principal ideals are the same cannot be done simply by comparing the two sets of generating polynomials, we also need a way to decide whether a given polynomial belongs to the ideal generated by a given set of polynomials. An alternative and more appealing approach consists in finding a canonical set of generating polynomials. This problem will be discussed in Sect. 10.7.4.

Remark 7.9.3 The free group of rank 2 is a special case, in this case the elementary ideal is trivially $\varepsilon_2 = L$. The reverse implication is not to our knowledge true, so that computing the second elementary ideal is not sufficient in proving that the fundamental group is the free group. Also, we are not aware of an analogue of the *unknotting Theorem* for the case of the inside/outside of a surface of genus 2, so that knowing that the fundamental group of e.g., the inside is free does not imply that the solid set E (inside of Σ) can be retracted onto a one-dimensional object.

References

1. Arnaud, H.: On the recognition of tori embedded in R^3 . IMAGEN-A, Publishing House of the University of Seville **1**, 65–72 (2010)
2. Arnold, V.I.: Indices of singular points of 1-forms on a manifold with boundary, the convolution of invariants of groups generated by reflections, and the singular projections of smooth surfaces. Uspekhi Math. Nauk. **34**(2), 3–38 (1979), Russian Math. Surveys **34**(2), 1–42 (1979)

3. Arnold, V.I.: The Theory of Singularities and its Applications. Cambridge University Press, Cambridge (1991)
4. Arnold, V.I.: Singularities of Caustics and Wave Fronts. In: Mathematics and its Application, vol. 62. Kluwer, Dordrecht (1991)
5. Arnold, V.I.: Topological Invariants of Plane Curves and Caustics. In: University Lecture Series, vol. 5. American Mathematical Society, Providence, RI (1994)
6. Arnold, V.I.: Invariants and perestroikas of plane fronts. In: Proceedings of the Steklov Institute of Mathematics, vol. 209 (1995)
7. Arnold, V.I., Goryunov, V.V., Lyashko, O.V., Vassiliev, V.A.: In: Arnold V.I. (ed.) Dynamical Systems VIII. Singularity Theory. II. Applications. Springer, Berlin (1993)
8. Bellettini, G., Beorchia, V., Paolini, M.: Topological and variational properties of a model for the reconstruction of three-dimensional transparent images with self-occlusions. *J. Math. Imaging Vision* **32**, 265–291 (2008)
9. Bellettini, G., Beorchia, V., Paolini, M.: An explicit formula for a Bennequin-type invariant for apparent contours. *Topology Appl.* **156**, 747–760 (2009)
10. Bennequin, D.: Entrelacements et équations de Pfaff. *Astérisque* **107–108**, 87–161 (1983)
11. Brasselet, J.-P.: Poincaré-Hopf theorems on singular varieties. In: Brasselet, J.-P., Damon, J., Tráng, L.D. Oka, M. (eds.) Singularities in Geometry and Topology. Proceedings of the Trieste Singularity Summer School and Workshop ICTP, pp. 57–80. World Scientific Publishing (2007)
12. Brown, M.: A proof of the generalized Schoenflies theorem. *Bull. Am. Math. Soc.* **66**, 74–76 (1960)
13. Callahan, J.: Singularities and plane maps. *Am. Math. Mon.* **81**, 211–240 (1974)
14. Chmutov, S., Duzhin, S., Mostovoy, J.: Vassiliev Knot Invariants. Cambridge University Press, Cambridge (2012)
15. Crowell, R.H., Fox, R.H.: Introduction to Knot Theory. Springer, New York (1977)
16. de Rham, G.: Introduction aux polynômes d'un noeud. *Ens. Math.* **XIII**, 187–194 (1967)
17. Dimca, A.: Singularities and Topology of hypersurfaces. In: Universitext. Springer, New York (2006)
18. Fox, R.H.: Free differential calculus. I. *Ann. Math.* **57**, 547–560 (1953)
19. Fox, R.H.: Free differential calculus. II: the isomorphism problem of groups. *Ann. Math.* **59**, 196–210 (1954)
20. Fox, R.H.: A quick trip through knot theory. In: M.K. Fort jun. (ed.), Topology of 3-manifolds and Related Topics. Proceedings of the University of Georgia Institute, pp. 120–167. Prentice-Hall, New Jersey (1961)
21. Francis, G.K., Weeks, J.R.: Conway's ZIP proof, *Am. Math. Mon.* **106**, 393–399 (1999)
22. Gibson, C.G., Hobbs, C.A.: Singularities of general two-dimensional planar motion. *J. Math. New Zealand* **25**, 141–163 (1996)
23. Hacon, D., Mendes de Jesus, C., Romero Fuster, M.C.: Global topological invariants of stable maps from a surface to the plane. In: Proceedings of the 6th Workshop on Real and Complex Singularities, 2001. Lecture Notes in Pure Applied Mathematics, vol. 232, pp. 227–235 (2003)
24. Hatcher, A.: Algebraic Topology Online Book. Cambridge University Press, Cambridge (2002)
25. Kawauchi, A.: A Survey of Knot Theory. Birkhäuser, Basel (1996)
26. Levine, H.I.: Stable maps: an introduction with low dimensional examples. *Bol. Soc. Bras. Mat.* **7**, 145–184 (1972)
27. Levine, H.I.: Computing the Euler characteristic of a manifold with boundary. *Proc. Am. Math. Soc.* **123**, 2563–2567 (1995)
28. Magnus, W., Karrass, A., Solitar, D.: Combinatorial Group Theory: Presentations of Groups in Terms of Generators and Relations. Dover Publications, New York (1976)
29. Manturov, V.: Knot Theory. Chapman & Hall/CRC, Boca Raton (2004)
30. Mather, J.N.: Infinite dimensional group actions, Cartan Fest. *Analyse et Topologie*, Astérisque Paris, Soc. Math. de France. **32, 33**, 165–172 (1976)
31. Moise, E.E.: Geometric Topology in Dimensions 2 and 3. In: Graduate Texts in Mathematics, vol. 47. Springer, New York (1977)

32. Ohmoto, T.: Vassiliev type invariants for generic mappings, revisited. In: *Real and Complex Singularities. Contemporary Mathematics*, vol. 569, pp. 143–159. American Mathematical Society, Providence, RI (2012)
33. Ohmoto, T., Aicardi, F.: First order local invariants of apparent contours. *Topology* **45**, 27–45 (2006)
34. Petitot, J.: *Neurogéométrie de la Vision - Modèles Mathématiques et Physiques des Architectures Fonctionnelles*, Les Editions de l'École Polytechnique, Paris (2008)
35. Pignoni, R.: On surfaces and their contour. *Manuscripta Math.* **72**, 223–249 (1991)
36. Polyak, M.: On the Bennequin invariant of Legendrian curves and its quantization. *C. R. Acad. Sci., Paris Sér. I* **322**, 77–82 (1996)
37. Richeson, D.: *Euler's Gem: The Polyhedron Formula and the Birth of Topology*. Princeton University Press, New Jersey (2008)
38. Richeson, D.: *Knots and Links*. AMS Chelsea Publishing, Canada (2003)
39. Seifert, H., Threlfall, W.: *Lehrbuch der Topologie*, Teubner, Leipzig, 1934. Translated into English as *A Textbook of Topology*. Academic Press, New York (1980)
40. Spivak, M.: *A Comprehensive Introduction to Differential Geometry*, 3rd edn. vol. 1. Publish or Perish, Inc., Houston, Texas (1999)
41. Tabachnikov, S.L.: Computation of the Bennequin invariant of a Legendrian curve from the geometry of its front. *Funct. Anal. Appl.* **22**, 89–90 (1988)
42. Thom, R.: The bifurcation subset of a space of maps. In: *Manifolds-Amsterdam. Lecture Notes in Mathematics*, vol. 197, pp. 202–208. Springer, Berlin (1971)
43. Thom, R.: *Stabilité Structurelle et Morphogénèse*. W.A. Benjamin, Inc., Reading, Massachusetts (1972)
44. Torisu, I.: On the additivity of the Thurston-Bennequin invariant of Legendrian knots. *Pac. J. Math.* **210**, 359–365 (2003)
45. Vassiliev, V.A.: Cohomology of knot spaces. *Adv. Sov. Math.* **21**, 23–69 (1990)
46. Wilson, L.C.: Equivalence of stable mappings between two-dimensional manifolds. *J. Differ. Geom.* **11**, 1–14 (1976)

Chapter 8

Elimination of Cusps

In this chapter we show that the apparent contour of a stable embedded closed smooth (not necessarily connected) surface can be modified, using some of the moves illustrated in Chap. 6, to obtain an apparent contour *without cusps*.¹ More precisely, recalling the notion of Reidemeister-equivalence (Definition 6.1.5) and its generalization on open sets (Definition 8.3.1) our aim is to prove (Theorem 8.3.2) that *any complete labelled contour graph is Reidemeister-equivalent to a complete labelled contour graph without cusps* (actually, in Theorem 8.3.2 we prove a stronger statement, since the thesis is localized in suitable open subsets of \mathbb{R}^2). Together with the results of Chap. 6, it follows that, up to \mathbb{R}^3 -ambient isotopies, any smooth closed surface embedded in \mathbb{R}^3 has an apparent contour without cusps. The prototypical example is maybe given by the torus, where its apparent contour with four cusps (Fig. 2.6) is Reidemeister-equivalent to two concentric circles (properly oriented).

The apparent contour is modified only in a local way; the new apparent contour differs from the old one by the presence of suitable “doubled long” arcs,² constructed so as to connect a cusp with another suitably chosen cusp. As already remarked in the Introduction, it is worth noticing that it may happen that the number of crossings of the new apparent contour is considerably higher than the number of crossings of the original apparent contour. Therefore, the simplification of the apparent contour due to “annihilation” of cusps could be obtained at the expense of a complication in terms of arcs and crossings.

¹A map having an apparent contour without cusps is usually called a fold map: see [3, 5, 7] for related references, and also [9, p. 403], [1, 2, 6].

²On the surface embedded in \mathbb{R}^3 having the graph as its apparent contour (compare with Theorem 5.1.1), these long arcs correspond to a pair of folds on the surface, forming a sort of wrinkle.

To state the elimination result, two notions are required: the embedding sign of a cusp, and the connectibility of two cusps in an open set. These two concepts are introduced and discussed in the next two sections, respectively.

8.1 Embedding Sign of a Cusp

Let (G, f, d) be a complete labelled contour graph (Definition 4.2.6). Recalling the notation of Sect. 3.3, if $c \in G$ is a cusp, the *incoming arc* of G at c will be denoted by c^- and the *outgoing arc* by c^+ . Moreover, we let $d(c^-)$ (respectively $d(c^+)$) be the value of d on c^- (respectively on c^+). From the requirements that we have on a labelling (see Fig. 3.11), it follows that

$$|d(c^-) - d(c^+)| = 1.$$

Remember that the function d is defined on $G \setminus \text{crossings}(G)$, hence it has a well-defined value at a cusp c , and it is also convenient to recall that

$$d(c) = \min(d(c^-), d(c^+))$$

(see Remark 3.4.4).

Definition 8.1.1 (Increasing and Decreasing Cusps) We say that the cusp c of G is *increasing* (respectively *decreasing*) if d is increasing (respectively decreasing) at c , i.e., if $d(c) = d(c^-)$ (respectively $d(c) = d(c^+)$).

For instance, the third picture of Fig. 3.11 shows an increasing cusp, and the second picture a decreasing cusp.

Now, we assign a sign to a cusp c , related to the value of d at the incoming arc at c , see Fig. 8.1.

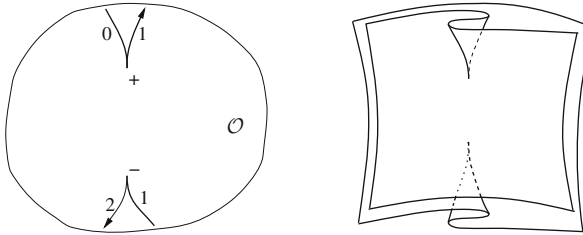


Fig. 8.1 Example of cusps that cannot be eliminated in \mathcal{O} because they are not connectable. The signs \pm refer to the embedding sign of the cusps

Definition 8.1.2 (Embedding Sign of a Cusp) Let c be a cusp of G . We define the embedding sign of c as

$$\text{embsign}(c) := (-1)^{d(c^-)}.$$

Remark 8.1.3 The embedding sign of a cusp c , which is the parity of $d(c^-)$, differs from the classical notion³ [4, 8] of parity of a cusp of an apparent contour of a map from an oriented two-manifold into the plane. In our context (since an embedded surface is necessarily orientable), the classical parity of a cusp would be simply given by the parity of $d(c)$.⁴

It has been already observed in Lemma 3.4.6 that the number of cusps of a component of the apparent contour is even. The following stronger statement holds.

Lemma 8.1.4 (Alternate Embedding Sign) Let C be a component of G . Then

$$\sum_{c \in \text{cusps}(C)} \text{embsign}(c) = 0. \quad (8.1)$$

Moreover, two consecutive cusps of C have alternate embedding sign.

Proof Let $c_1, c_2 \in \text{cusps}(C)$ be two consecutive cusps⁵ of C . By the properties of d on the arcs⁶ of G , we have that $\text{embsign}(c_1) = -\text{embsign}(c_2)$, and (8.1) follows. \square

The strategy of the proof of the elimination of cusps is based on the application on G of the Reidemeister-type moves described in Chap. 6. It is then worth noticing an invariance property of the embedding sign.

Remark 8.1.5 (Invariance of Embedding Sign Under Reidemeister-Type Moves) Move B, as well as moves L and S, elides two cusps of opposite embedding sign (belonging to two possibly different connected components of G), whereas all other Reidemeister-type moves do not involve cusps. Therefore, the sum of the embedding sign of all cusps is invariant under the application of one of the (direct or inverse)

³See Definition 2.2.12.

⁴For example, in [4, remark following Theorem 2] if we orient the embedded surface according to the inner normal, and the critical curve with the orientation induced by the arcs of the apparent contour, then the definition of sign of a cusp coincides with $(-1)^{d(c)}$.

⁵The argument is of course valid also in the presence of crossings in between the two cusps.

⁶Recall that, when traversing a crossing, either d remains constant, or it jumps by two units, so that its parity remains constant.

Reidemeister-type moves on apparent contours. This is consistent with (8.1), in the following sense:

- if, after the application of one of the above-mentioned moves, two different connected components C_1 and C_2 join into one single connected component C , then

$$\sum_{c \in \text{cusps}(C_1) \cup \text{cusps}(C_2)} \text{embsign}(c) = \sum_{c \in \text{cusps}(C)} \text{embsign}(c); \quad (8.2)$$

- if, after the application of one of the above-mentioned moves, one connected component C splits into two different connected components C_1 and C_2 , then (8.2) holds.

Observe that move B either splits a component of the apparent contour into two new components or joins two distinct components into a single new component; for this reason, we cannot speak of invariance under Reidemeister-type moves for a single component. Similarly, move L completely eliminates a component with exactly two cusps with opposite embedding sign (the reverse happens for the inverse L^{-1} of move L).

8.2 Connectable Cusps in an Open Set

In this section we introduce the notion of connectable cusps relatively to an open set.

We start by formalizing the concept of adjacency between strata already used in Sect. 5.2.1 (see, for instance, Fig. 5.1).

Definition 8.2.1 (Horizontal and Vertical Adjacency) The strata (R_1, r_1) and (R_2, r_2) are *horizontally* adjacent if

- $R_1 \neq R_2$,
- there exists an arc a of G with $a \subseteq \partial R_1 \cap \partial R_2$,
- either $r_1 = r_2 \leq d(a)$, or $r_2 = r_1 - 2 \geq d(a) + 1$ if $f(R_2) < f(R_1)$, or $r_1 = r_2 - 2 \geq d(a) + 1$ if $f(R_1) < f(R_2)$.

In this situation the arc a is said to give horizontal adjacency between the strata.

The strata (R_1, r_1) and (R_2, r_2) are *vertically* adjacent if

- $R_1 = R_2$,
- $|r_2 - r_1| = 1$,
- there exists an arc $a \subseteq \partial R_1$ of G having R_1 on the left, and such that $d(a) + 1 = \min(r_1, r_2)$.

In this situation the arc a is said to give vertical adjacency between the strata.

Now, let \mathcal{O} be an open connected subset of \mathbb{R}^2 with boundary $\partial\mathcal{O}$ of class \mathcal{C}^1 ; $\mathcal{O} = \mathbb{R}^2$ is allowed, in particular $\partial\mathcal{O}$ can be empty.

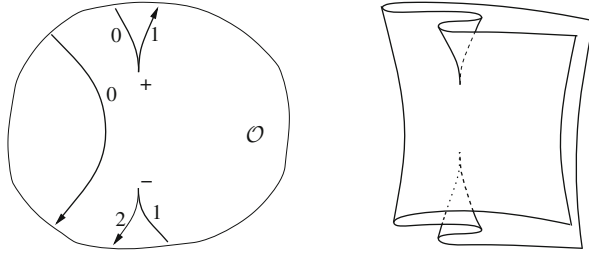


Fig. 8.2 Example of cusps that are connectable, hence they can be eliminated in \mathcal{O} . Notice the *bold left arc in the left picture*, which is not present in the left picture of Fig. 8.1, and makes now connectable the two cusps in \mathcal{O} . The signs \pm refer to the embedding signs of the cusps. Observe that the move B cannot be applied (see Fig. 6.2), because the two cusps lie in different strata

Definition 8.2.2 (Generic Position) We say that \mathcal{O} is in generic position with respect to G if

$$\text{nodes}(G) \cap \partial\mathcal{O} = \emptyset,$$

and all intersections between $\partial\mathcal{O}$ and G are transverse, and in finite number.

From now on in this chapter, \mathcal{O} is in generic position with respect to G .

Definition 8.2.3 (\mathcal{O} -horizontal and \mathcal{O} -Vertical Adjacencies) Let us consider two strata corresponding to regions that intersects \mathcal{O} . The two strata are said \mathcal{O} -horizontally adjacent (respectively \mathcal{O} -vertically adjacent) if the arc that gives horizontal adjacency (respectively vertical adjacency) intersects \mathcal{O} .

Example 8.2.4 Take, in Fig. 8.2, $R_1 = R_2$ as the complement in \mathcal{O} of the inside of the two cusps on the left of the arc a having⁷ $d(a) = 0$, and $r_1 = 1$ and $r_2 = 2$. Then (R_1, r_1) and (R_2, r_2) are \mathcal{O} -vertically adjacent. On the other hand, take $R_1 = R_2$ to be the complement in \mathcal{O} of the inside of the two cusps in Fig. 8.1 and $r_1 = 1$ and $r_2 = 2$. Then (R_1, r_1) and (R_2, r_2) are not \mathcal{O} -vertically adjacent, because of the lackness of an arc a giving vertical adjacency in \mathcal{O} . An example of horizontal adjacency is given in Fig. 8.3.

We are now in a position to introduce the notion of connectable strata.

Definition 8.2.5 (Connectable Strata in \mathcal{O}) Two strata are connectable in \mathcal{O} if there exists a finite sequence of \mathcal{O} -horizontally or \mathcal{O} -vertically adjacent strata connecting them.

⁷When G is the apparent contour of an embedded surface Σ , the arc a results from a folding of Σ .

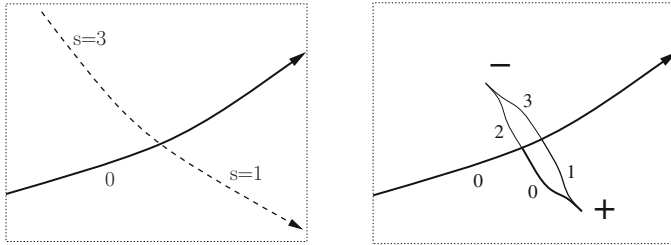


Fig. 8.3 The *dashed curve* represents the image of γ , and is oriented from *left to right*; the *bold arc* belongs to G . On the *right* we show the combined elementary moves for a horizontal adjacency: the (labelled) contour on the right is obtained by performing the inverse of an L move and the inverse of a C move (see Fig. 6.2), thus creating two new cusps on opposite sides of the bold arc. The signs \pm refer to the embedding sign of the cusps. Note that the value of s at cusp c equals $d(c) + 1$. All other horizontal adjacencies are obtained from this one by left–right reflection, front–back inversion and addition of a positive constant to the labelling d . As explained in the proof of Theorem 8.3.2, the remaining (not displayed in the right figure) long *dashed arcs* are doubled, and cusps with the proper value of d are inserted so as to annihilate the cusps in the right figure by using the move B, see Fig. 8.7. The arcs and crossings remaining after this operation belong to \tilde{G} but not to G . The prototypical case of a fold “behind” the curve γ corresponds to substituting: in the left picture, 1 in place of 0 and $s = 1$ in place of $s = 3$; in the right picture, the label 0 in place of 2, and 1 in place of 3, while the value $d = 0$ is replaced by the sequence 1, 3, 1

Definition 8.2.6 (Cusp Pointing Into a Stratum) Let c be a cusp of G and (R, r) be a stratum. We say that c points into (R, r) if

- $c \in \partial R$ and R is locally on the right of c^- and c^+ , namely $R = R_{\min}$,
- $r = d(c) + 1$.

Note that given a cusp c there is exactly one stratum such that c points into that stratum. Referring to Example 8.2.4 (Fig. 8.2) we have that the upper cusp points into $(R_1, r_1) = (R_1, 1)$, and the same happens in Fig. 8.1.

We now introduce the notion of connectable cusps, which takes a crucial role in the proof of the elimination theorem.

Definition 8.2.7 (Connectable Cusps in \mathcal{O}) Let $c_1, c_2 \in \mathcal{O}$ be two cusps of G . Let (R_1, r_1) and (R_2, r_2) be two strata so that c_j points into (R_j, r_j) for $j = 1, 2$. We say that c_1, c_2 are connectable in \mathcal{O} if (R_1, r_1) and (R_2, r_2) are connectable in \mathcal{O} .

Remark 8.2.8 (Connectable Cusps in \mathbb{R}^2) Let c_1, c_2 be two cusps on the same component of G ; then, independently of their embedding sign, c_1 and c_2 are connectable in \mathbb{R}^2 , namely c_1 and c_2 point into connectable strata in \mathbb{R}^2 . A chain of horizontally or vertically connectable strata that connect the two strata pointed by c_1 and c_2 can be found by recalling that, from Theorem 5.1.1, (G, f, d) coincides with the labelled apparent contour $(G_\Sigma, f_\Sigma, d_\Sigma)$ of a stable smooth closed surface Σ embedded in \mathbb{R}^3 . Hence, a component of the apparent contour is just the projection of a critical curve of the manifold, and the stratum pointed by a cusp corresponds to a portion of the manifold that is adjacent to this critical curve. Therefore, it is sufficient to follow the critical curve and annotate the strata corresponding to the portions of manifold adjacent to it until we reach the point of the critical curve

that projects to the target cusp c_2 . Recall that horizontally/vertically adjacent strata correspond to adjacent portions of the manifold.

Notice that two cusps may be connectable also when they belong to different components of G . Notice also that given a cusp c_1 on a component C of G , there exists at least another cusp $c_2 \in C$ with opposite embedding sign connectable with c_1 .

In this section we have not made use of the function f , which will be used in the next section.

8.3 Statement of the Elimination Theorem

In order to state the main result of this chapter (Theorem 8.3.2), we need the following concept, which defines an equivalence relation.⁸

Definition 8.3.1 (Reidemeister-Equivalence in \mathcal{O}) We say that two complete labelled contour graphs are Reidemeister-equivalent in \mathcal{O} if they can be connected by using a finite sequence of direct or inverse Reidemeister-type *moves* with compact support in \mathcal{O} , and a finite number of \mathbb{R}^2 ambient isotopies with compact support in \mathcal{O} .

The essence of the next result is, roughly speaking, that if two cusps of G in \mathcal{O} have opposite embedding signs and are connectable in \mathcal{O} , then G is Reidemeister-equivalent in \mathcal{O} to a complete labelled contour graph without c_1 and c_2 . The new apparent contour may have new crossings and new arcs: for instance, arcs originally ending in c_1 and c_2 get extended.

Theorem 8.3.2 (Cusps Elimination in \mathcal{O}) *Let (G, f, d) be a complete labelled contour graph. Let $\mathcal{O} \subseteq \mathbb{R}^2$ be an open set with $\partial\mathcal{O}$ of class \mathcal{C}^1 , in generic position with respect to G . Let $c_1, c_2 \in \mathcal{O}$ be two cusps of G . Assume that*

$$\text{embsign}(c_1) = -\text{embsign}(c_2),$$

and that c_1 and c_2 are connectable in \mathcal{O} . Then there exist

- *a complete labelled contour graph $(\tilde{G}, \tilde{f}, \tilde{d})$,*
- *a neighbourhood W of G ,*
- *two disjoint neighbourhoods U_1, U_2 of c_1 and c_2 respectively, with*

$$U_1 \cup U_2 \subset\subset \mathcal{O}, \quad U_1 \cup U_2 \cap \left(\text{nodes}(G) \setminus \{c_1, c_2\} \right) = \emptyset,$$

- *pairwise disjoint neighbourhoods A_1, \dots, A_n of portions of arcs of G , with*

⁸Reidemeister-equivalence in \mathbb{R}^2 has been already used in Chap. 6 (see Definition 6.1.5).

$$\bigcup_{i=1}^n A_i \subset\subset \mathcal{O}, \quad \bigcup_{i=1}^n A_i \cap [(U_1 \cup U_2) \cup \text{nodes}(G)] = \emptyset,$$

with the following properties:

- $(\tilde{G}, \tilde{f}, \tilde{d})$ and (G, f, d) are Reidemeister-equivalent in \mathcal{O} ;
- $(\tilde{G}, \tilde{f}, \tilde{d})$ and (G, f, d) coincide in $W \setminus D$, where $D := U_1 \cup U_2 \cup \bigcup_{i=1}^n A_i$;
- $\text{cusps}(\tilde{G}) = \text{cusps}(G) \setminus \{c_1, c_2\}$.

Notice that \tilde{G} has no cusp in $U_1 \cup U_2$, and

$$\#\text{cusps}(\tilde{G}) = \#\text{cusps}(G) - 2.$$

Moreover \tilde{G} can have in D new nodes and new arcs, but no new cusp. In addition, in \mathcal{O} and outside of W , the graph \tilde{G} can have (new) arcs and (new) nodes but no new cusp. Notice also that, if $\text{embsign}(c_1) = \text{embsign}(c_2)$, then from Lemma 8.1.4, the two cusps c_1 and c_2 cannot be eliminated, whatever the choice of \mathcal{O} . Indeed, elimination of c_1 and c_2 would then alter the total sum of the embedding sign, which is on the contrary an invariant (equal to zero).

Corollary 8.3.3 (Cusps Elimination) *Let (G, f, d) be a complete labelled contour graph. Then there exists a complete labelled contour graph $(\tilde{G}, \tilde{f}, \tilde{d})$ without cusps and Reidemeister-equivalent to (G, f, d) .*

8.4 Proof of the Elimination Theorem

The aim of this section is to prove Theorem 8.3.2 and Corollary 8.3.3.

Without loss of generality, we can assume that the embedding sign of c_1 is positive and the embedding sign of c_2 is negative. We start by selecting two points P_1 and P_2 belonging to the regions pointed by the two cusps, with P_1 sufficiently close to c_1 and P_2 sufficiently close to c_2 . Since by assumption the two cusps are connectable in \mathcal{O} , we can construct a piecewise smooth curve⁹ connecting them, and a locally constant function defined on it, indicating point by point in which stratum the curve lies. The curve can pass from a region to another region, provided the corresponding strata are either horizontally or vertically adjacent. In case of vertical adjacency, the curve “bounces” on the arc giving adjacency (remaining on the left of the arc, see Fig. 8.4). In case of horizontal adjacency, the curve crosses the arc giving adjacency. Let us state all these properties in a precise way: there exist a

⁹It may be helpful to regard this curve as a curve connecting the two corresponding points in (a connected component of) the embedded surface constructed in Chap. 5. Considered at this level, the curve does not self-intersect, and two cusps are connectable if they lie in the same connected component.

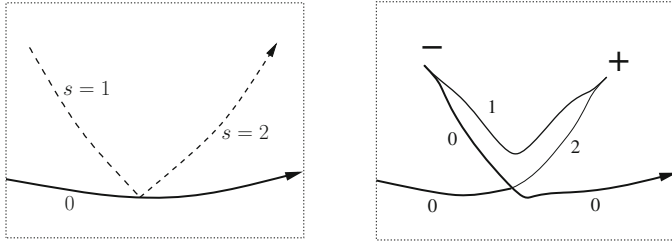


Fig. 8.4 Combined elementary moves for a vertical adjacency: the (labelled) contour on the *right* is obtained by performing one of the two inverses of an S move, which one depends on whether the value of s is increasing or decreasing. At this stage of the proof two new cusps are added to the graph. Applying this operation to the case of Fig. 8.2, the new added cusps (whose embedding sign is indicated by \pm) will be annihilated in \mathcal{O} with the preexisting cusps with corresponding opposite embedding sign, by inserting a couple of arcs of $\text{im}(\gamma)$ (see the *left* picture, where each *dashed arc* must be doubled), and using the move B. Note that the value of s coincides with the value of the labelling at a cusp increased by one. This move is applied when γ “bounces left” and the labelling of the arc is even (as in the figure) or when γ “bounces right” and the labelling of the arc is odd, otherwise we resort to the move illustrated in Fig. 8.6

continuous curve $\gamma : [0, 1] \rightarrow \mathbb{R}^2$ with $\gamma(0) = P_1$, $\gamma(1) = P_2$, and a function

$$s : \text{im}(\gamma) \rightarrow \mathbb{N},$$

(where $\text{im}(\gamma) = \gamma([0, 1])$) with the following properties:

- $\text{im}(\gamma)$ self-intersects only at a finite set S_I of points, any self-intersection is double and transverse, and $S_I \cap G = \emptyset$;
- $\text{im}(\gamma) \cap \text{nodes}(G) = \emptyset$,
- $\text{im}(\gamma) \cap G$ is finite and can be partitioned as $T_I \cup B_I$, where T_I are the *transversal* intersections, and B_I are the *bouncing* intersections, namely those points such that $\text{im}(\gamma)$ lies locally on the left of the arc;
- $\text{im}(\gamma) \setminus B_I$ is of class C^∞ up to B_I from both sides, and at a point of B_I the two tangents to $\text{im}(\gamma)$ are distinct and different from the tangent to the arc¹⁰;
- each crossing in T_I corresponds to a horizontal adjacency of strata and each crossing in B_I corresponds to a vertical adjacency of strata;
- function s is locally constant on the arcs of $\text{im}(\gamma)$ out of $T_I \cup B_I$;
- at a point of S_I the two values of s (on the two concurring arcs of $\text{im}(\gamma)$) are distinct, see Fig. 8.5;
- function s at the endpoints P_1 and P_2 has the values corresponding to the strata where the cusps point into;
- if we think of γ as taking values in the embedded surface (footnote 9), then at any point of $\text{im}(\gamma)$, the value of s would coincide with the index numerating the stratum containing that point.

¹⁰It is possible to modify the construction so that the whole of $\text{im}(\gamma)$ is C^∞ , requiring that at points of B_I the set $\text{im}(\gamma)$ is tangent to the arc. Since this smoothness is not necessary here, we do not add this requirement.

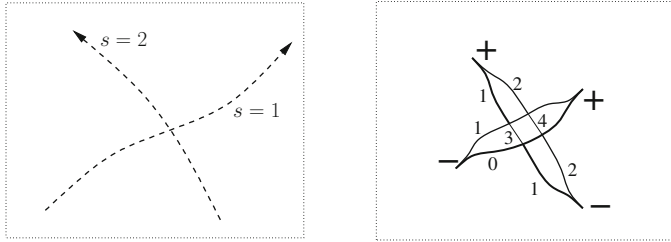


Fig. 8.5 Combined elementary moves for a self-crossing: the graph on the *right* is obtained by performing the inverse of two \mathbb{L} moves and the inverse of two \mathbb{C} moves. Again, new cusps are created, the signs \pm refer to the embedding signs, and the value of s coincides with the value of the labelling at the cusps increased by one

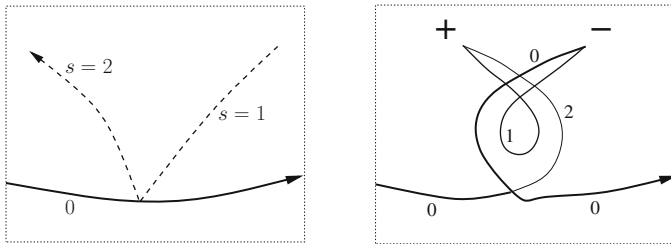


Fig. 8.6 Combined elementary moves allowing, when necessary, exchange of the embedding sign of the cusps

The next step in the proof is the following:

- for each point in T_I we perform one move as explained in Fig. 8.3;
- for each point in B_I we perform a move as explained in Fig. 8.4 or in Fig. 8.6. Which one depends on the parity of the labelling of the arc and on whether the curve γ “bounces left” or “bounces right” as explained in the caption of Fig. 8.4;
- for each point in S_I we perform a move as explained in Fig. 8.5.

Notice that several new cusps are introduced at this stage of proof. The local modifications displayed in Figs. 8.3, 8.4 and 8.6 require modifications of G in a small portion of the involved arcs. These portions are the sets

$$A_1, \dots, A_n$$

in the statement of the theorem.

Now, if we remove from $\text{im}(\gamma)$ those portions where we performed the local modifications of Figs. 8.3–8.6, we remain with a set of disjoint arcs, the closure of which is contained in the regions of G . Let us focus the attention on one of these arcs, say $\gamma([\alpha, \beta])$, for an interval $[\alpha, \beta]$ contained in $[0, 1]$. Using an inverse of an \mathbb{L} move, we double this arc into a pair of close together parallel arcs travelling along $\text{im}(\gamma)$, connecting the two newly introduced cusps (see Fig. 8.7), positioned near

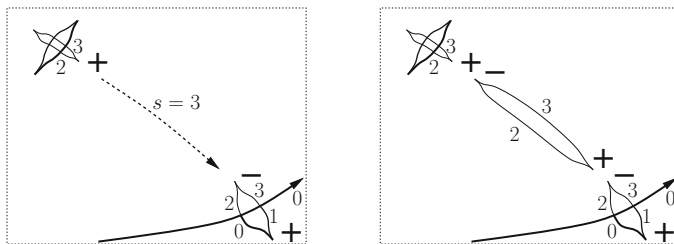


Fig. 8.7 A long arc of γ (dashed curve) is doubled, and two new cusps are created using an inverse of an L move (right picture). The embedding signs and the values of the labelling are so that these cusps can be annihilated with the preexisting nearby cusps, using a B move

the endpoints of $\gamma([\alpha, \beta])$.¹¹ A choice of the inverse of L can be done in such a way that:

- the labelling of the two arcs are $s - 1$ and s ,
- the labelling on the arc on the right (with respect to the orientation of γ) is even.

As a consequence, the cusp close to $\gamma(\alpha)$ has negative embedding sign, while the cusp close to $\gamma(\beta)$ has positive embedding sign. At this point, we have a set of new cusps paired in such a way that they can be all eliminated by B moves. The proof of Theorem 8.3.2 is concluded. \square

We can now prove Corollary 8.3.3. Let C be a component of G . From Lemma 8.1.4 it follows that the number of cusps of C having positive embedding sign is equal to the number of cusps of C having negative embedding sign. Take a cusp $c \in C$ with positive embedding sign; then, travelling along C , it follows that the subsequent cusp c' of C has negative embedding sign (Lemma 8.1.4). By Remark 8.2.8 it follows that c and c' are connectable in \mathbb{R}^2 . Then Theorem 8.3.2 provides a new complete labelled contour graph having as set of cusps the set $\text{cusps}(G) \setminus \{c, c'\}$. The assertion follows by repeating this procedure until there are no more cusps in C , and repeating the argument for any component C of G . \square

Remark 8.4.1 It is worth noticing that not all moves listed in Chap. 6 are necessary to prove Theorem 8.3.2. More precisely, in the proof we have used the move B and the inverses of the moves S, L and C, while moves K and T (and their inverses) have not been used. Note also that the constructive proof of Theorem 8.3.2 allows us to compute the number of moves which are necessary to annihilate all cusps of the apparent contour.

¹¹When G is the apparent contour of an embedded surface Σ , this operation corresponds to the creation of thin long crease (a double fold) that follows that part of γ .

8.5 Application to Closed Embedded Surfaces

From Corollary 8.3.3, Theorem 5.1.1 and the results of Chap. 6 we deduce the following result.

Theorem 8.5.1 (Cusps Elimination Applied to Shapes) *Let Σ be a stable closed (not necessarily connected) surface of class C^∞ embedded in \mathbb{R}^3 , and let G_Σ be its apparent contour. Then there exists an \mathbb{R}^3 -ambient isotopy deforming Σ into a smooth stable closed surface $\tilde{\Sigma}$ so that the apparent contour of $\tilde{\Sigma}$ has no cusps.*

Example 8.5.2 (Torus) The aim of Theorem 8.3.2 is to provide a general way to elide pairs of cusps using the Reidemeister-type moves on apparent contours, and there is no claim that such a procedure is, in some sense, the simplest one. This is clarified by the example of the torus having the apparent contour as in Fig. 8.8. In this case, an immediate way to annihilate (two by two) the four cusps is to perform the move S on each pair of (nearby) cusps, c_1 with c_2 and c_3 with c_4 (note, on the other hand, that the move B cannot be applied with the same pairing, although we could annihilate with a B move, e.g. c_1 with c_4). This is not the procedure made in the proof of Theorem 8.3.2, which, having a general validity, in this case results in a much more complicated construction. Let us consider, for instance, the positively embedded cusp c_1 : a move B cannot be applied between c_1 and the negatively embedded cusp c_2 , since c_1 and c_2 point into different strata, like in Fig. 8.4 (right). Following the strategy of proof of Theorem 8.3.2, we introduce an arc γ connecting c_1 to c_2 , and bouncing on the exterior boundary of the torus (see Fig. 8.8 (right)). We then introduce a pair of new cusps near the bouncing point arguing as in Fig. 8.4 right (using the local part of the boundary of the exterior region, which takes the role of the arc labelled by 0 in Fig. 8.2 (left)); next, we annihilate c_1 with one of these cusps, and c_2 with the other one, as in the last picture of Fig. 8.8. A similar argument can be applied to annihilate the positively embedded cusp c_3 with the negatively embedded cusp c_4 .

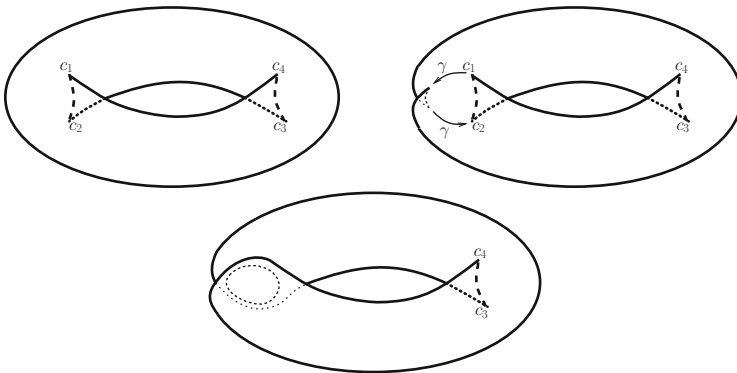


Fig. 8.8 Annihilation of cusps c_1 and c_2 in Example 8.5.2

References

1. Arnold, V.I., Goryunov, V.V., Lyashko, O.V., Vassiliev, V.A.: In: Arnold, V.I. (ed.) *Dynamical Systems VIII. Singularity Theory. II. Applications*. Encyclopaedia of Mathematical Sciences. Springer, Berlin (1993)
2. Eliashberg, J.M.: On singularities of folding type. *Izv. Akad. Nauk. SSSR* **34**, 1110–1126 (1960); *Math. USSR Izv.* **4**, 1119–1134 (1970)
3. Eliashberg, Y., Mishachev, N.M.: Wrinkling of smooth mappings and its applications. I. *Invent. Math.* **130**, 345–369 (1997)
4. Haefliger, A.: Quelques remarques sur les applications différentiables d'une surface dans le plan. *Ann. Inst. Fourier. Grenoble* **10**, 47–60 (1960)
5. Levine, H.I.: Elimination of cusps. *Topology* **3**, 263–296 (1965)
6. Martínez-Alfaro, J., Mendes de Jesus, C., Romero-Fuster, M.C.: Global classifications and graphs. In: Manoel, M., Romero-Fuster, M.C., Wall, C.T.C., (eds.) *Real and Complex Singularities*. The London Mathematical Society, pp. 246–267. Cambridge University Press, Cambridge (2010)
7. Millett, K.C.: Generic smooth maps of surfaces. *Topology Appl.* **18**, 197–215 (1984)
8. Ohmoto, T., Aicardi, F.: First order local invariants of apparent contours. *Topology* **45**, 27–45 (2006)
9. Whitney, H.: On singularities of mappings of Euclidean spaces. I. Mappings of the plane into the plane. *Ann. Math.* **62**, 374–410 (1955)

Chapter 9

The Program “Visible”

In this chapter we describe an actual implementation of the constructive proof of the completion result (Theorem 4.3.1). The corresponding software code, a *free software* program, called `visible`, is written in C language and is part of the `appcontour` project described in Chap. 10 [1]. It is hosted on *sourceforge.net*, its home page is <http://appcontour.sourceforge.net/>, from where the source code can be downloaded, compiled and installed following the standard procedure for Unix projects.

The software basically works as a *filter* that takes a description of the visible contour graph (briefly, visible contour) K as input and produces a description of a possible completion G as output.

Both input (visible contour) and output (complete labelled contour graph) are described in terms of their topological structure, by using a Morse-like description that corresponds to the one used in the proof of Theorem 4.3.1. All quantitative information, such as the actual position and shape of the contour lines in the retinal plane, are lost. In this respect the `visible` program should be considered as proof-of-concept of an implementation of the construction described in Chap. 4.

9.1 An Example

Before entering into the details we clarify the basic ideas with an example.

The drawing of Fig. 9.1 (left) is a visible contour graph that satisfies the requirements of Definition 4.1.8. Moreover, requirement (K3) is also satisfied when replacing $\text{ext}(K)$ with $\text{background}(K)$ obtained as the union of $\text{ext}(K)$ and the small internal region marked with an “e” in the picture (see also Remark 4.1.10). This allows us to apply the completion construction while forcing the region marked with “e” to have $f = f_{\Sigma} = 0$ (background) in the completed apparent contour graph, see Corollary 4.5.1. The textual description of this visible contour is shown on the right of Fig. 9.1, and is essentially based on the Morse description used in the

Fig. 9.1 The visible contour graph on the *left* is described by the text on the *right*, by listing a sequence of Morse events

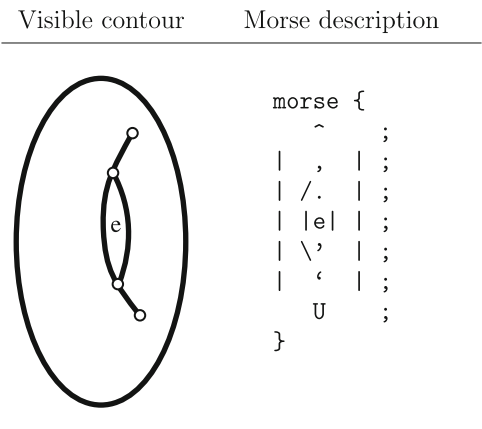


Table 9.1 To each Morse event we associate a character or a pair of characters taken from the standard set of ASCII characters

Morse event	ASCII representation	Morse event	ASCII representation
	^		U
	, (comma)		' or '
	\ '		.'/
	/.		.\

proof of Theorem 4.3.1. We need to encode the description as an ASCII¹ text file, to be used as input for the program. For this reason we substitute the list of symbols introduced in Sect. 2.5.3 (slightly modified to accommodate for T-junctions and terminal points) with similar characters from the standard ASCII chart; for example, the symbols \frown and \smile are both encoded with the caret (^) ASCII character. The complete list of Morse events and the corresponding ASCII characters are listed in Table 9.1. Information on the orientation is in most cases inferred by the program, by enforcing the validity of the assumptions in Definition 4.1.8. When needed, orientation can be given explicitly, adding one of the characters “l” (lowercase L), “r”, “u”, “d” to mean *left*, *right*, *up* and *down* respectively, to the corresponding Morse event. In this way the various typographical characters in the description of Fig. 9.1 (right) correspond to different kind of Morse events; the semicolons separate the description of different (critical) Morse lines, and the letter “e” appearing in the description is used to mark the corresponding region as part of the background background(*K*), forcing the program to reconstruct in such a way to have $f = 0$ in that region.

¹The American Standard Code for Information Interchange (ASCII) is a universally used character-encoding scheme originally based on the English alphabet. It encodes a set of 128 characters (33 are non-printing control characters) that can be easily used in a text file.

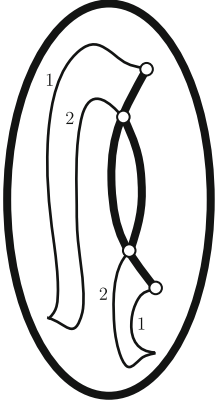
resulting Morse description	corresponding contour
<pre> morse { ~l0 ; d0 ^l1 u0 ; d0 d1 >0+ u0 ; d0 d1 ^r2 u0 u0 ; d0 d1 u2 Xu0d0 u0 ; d0 d1 u2 Xd2u0 u0 ; d0 d1 u2 d2 >1- u0 ; d0 <1+ u2 d2 u1 u0 ; d0 Ur2 d2 u1 u0 ; d0 d2 >2- u0 ; d0 Ur2 u0 ; Ur0 ; } </pre>	

Fig. 9.2 The reconstructed labelled apparent contour is described by the text shown on the *left*, with a corresponding drawing on the *right*

The Morse description of Fig. 9.1 (right) can be written into a file (with name, say, `example.morse`) and given as input to the program `visible` with the unix command

```
$ visible example.morse
```

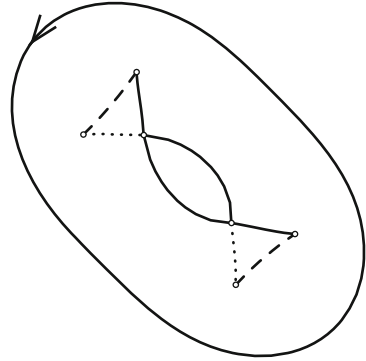
It will produce the text shown in Fig. 9.2 (left), where again the various typographical characters correspond to different Morse events and are now also equipped by additional characters carrying information about the orientation and labelling of the arcs. The drawing on the right of Fig. 9.2 is a graphic reconstruction corresponding (more or less line-by-line) to the textual description on the left. It is a (highly deformed) apparent contour of a torus. Indeed the resulting Morse description can be read by the `appcontour` program (described in Chap. 10) that can automatically produce the apparent contour shown in Fig. 9.3 with the unix command line

```
$ visible example.morse | contour printmorse | showcontour
```

where we recall that, in a unix environment, the “pipe” “|” character connects the output of the preceding command to the input of the following one. Since cusps in the Morse description resulting from the `visible` program are described with a syntax that differs from the one that `showcontour` understands, we need to interpose the `contour` filter.

The result shown in Fig. 9.3 corresponds perfectly with what we would have inferred from the `visible` contour that we selected for this starting example; however, it is important to emphasize that this is not always the case. Indeed the `visible` program does not make any effort to produce an “optimal” result, but it merely

Fig. 9.3 The Morse description obtained by the `visible` program can be processed by the `appcontour` program (Chap. 10) to produce this picture. *Solid lines* correspond to arcs with $d = 0$, *dashed lines* to $d = 1$, and *dotted lines* to $d = 2$



follows the procedure in the proof of Theorem 4.3.1. In many cases the result of the contour reconstruction, although perfectly valid, is far from optimal.

To make this clear we try to process the visible contour of the present example modified by removing the marking of the small internal region as part of the background (we remove the typographical character “e” in the Morse description). We also need to add orienting information for the arc on the right connecting the T-junctions (oriented downwards), since it can no longer be inferred by the program. This can be done by adding the “d” character right after the “|” corresponding to the transversal intersection Morse event, see Sect. 9.2. In this way we obtain exactly the visible contour in Example 4.6.4, Fig. 4.15 of Chap. 4. The resulting reconstructed apparent contour is shown in Fig. 9.15, right (Example 9.5.11) and is not immediately recognizable. Actually, it corresponds to a deformed 3D sphere.

This modified visible contour is the same presented below in Example 9.5.11. The reconstructed Morse description and a spatially corresponding graphic visualization are shown in Fig. 9.16.

9.2 Encoding a Morse Description of the Visible Contour

As already mentioned in Sect. 9.1, we first need a way to describe the visible contour using a text file, and this is done by enforcing the Morse description introduced in Sect. 4.4. To each critical value λ_i we associate a string of text terminated by a semicolon “;”. In this string we encode the various type of events occurring at that critical level, following their natural order from left to right. The overall syntax of the descriptive text is thus the following:



```
morse {
  <Morse line 1>;
  <Morse line 2>;
  ...
}
```

Each *Morse line* encodes the description of the events at the corresponding critical level. Lines with no critical events (only transversal crossings, regular levels) are allowed. Lines with more than one critical event are also allowed, although they would be forbidden in a Morse description, and are interpreted in the obvious way as two or more lines, each with a single critical event.

In the end, the syntax for the Morse description is similar to the one used by the *apcontour* software, as we shall see in Sect. 10.3.2 of Chap. 10.

9.2.1 Encoding the Morse Events

The Morse events occurring at some level are encoded using the correspondence Table 9.1, where we choose standard typographical characters from the ASCII chart (or pairs of typographical characters) resembling the local shape of the event.

Each Morse event can be optionally followed by an orientation indicator, a single letter from the set *u, d, l, r* standing respectively for *up, down, left, right*. Orientations indicated by *l* or *r* (left/right) are reserved to the two Morse events  and , with obvious meaning, whereas all other Morse events can be oriented up (*u*) or down (*d*). For a T-junction the orientation indicator refers to the transversal arc, since the occluding arc carries a natural orientation due to restriction (K4) of Definition 4.1.8.

9.2.2 Implicit Orientation

If the orientation information is omitted for some Morse event, the *visible* program tries to infer it by assuming the validity of restrictions (K3) and (K4) of Definition 4.1.8² (with $\text{background}(K)$ in place of $\text{ext}(K)$ if some region is marked as part of the background by using the letter “e”, see below). Indeed, in the example of Sect. 9.1, all orientations can be inferred by assuming the validity of the constraints of Definition 4.1.8, in particular the (downward) orientation of the short arc connecting transversally the two T-junctions is inferred by the fact that on one side there is a region marked as part of the background (letter “e”). The program complains with an error if

- it is not able to infer the orientation of all the arcs by enforcing the restrictions (K3) and (K4). In such a case it is necessary to add the orientation information at some Morse event;
- the orientations (given explicitly in the description or inferred from the constraints) are incompatible with K being a visible contour.

²See also Remark 3.6.2 of Chap. 3 and Figs. 3.15 and 3.16 in the same chapter.

9.2.3 The “e” Region Marking

It is possible to instruct the software that some internal region is actually part of the background $\text{background}(K)$. This could be, e.g., an information inferred from the image or some a priori knowledge. To this aim the character “e” can be suitably inserted between two Morse events on any Morse line that traverses the involved region. Of course more than one region can be marked in this way to form the background set $\text{background}(K)$. The external region is always part of the background and need not be marked; any non-external region that is not marked will have $f > 0$ in the reconstructed apparent contour G .

The example presented in Sect. 9.1 has the small internal region marked as part of $\text{background}(K)$.

Note that all arcs adjacent to a background region are implicitly oriented due to requirement (K3) (applied to $\text{background}(K)$).

9.3 Using the Program

The textual file (say `example.morse`) is then fed to the `visible` program with a `unix` command like

```
$ visible example.morse
```

It will implement the completion construction described in the proof of Theorem 4.3.1 (see also Corollary 4.5.1 if there are regions marked as external with the character “e”). The Morse description of the computed apparent contour is then written in the standard output.³

The syntax of the resulting textual description will be briefly described in the next Sect. 9.4 and can be used as input for the `appcontour` program (see Chap. 10) in order to obtain information on the topological structure of the reconstructed apparent contour and on the corresponding 3D surface. Automated visualization of the apparent contour can be achieved using the `showcontour` program, but the Morse description must be converted in a compatible form due to the different way in which cusps are described (see Sect. 10.3.2). This conversion can be obtained either by filtering the output through the command `contour printmorse` or using the specially crafted `morse2morse` filter. The former entails first a conversion from the initial Morse description into the region description used internally by `appcontour` and then the construction of a corresponding Morse description, which will in general be structurally different from the original one, although describing the same apparent contour up to a deformation of the plane. The latter

³In a `unix` shell the *standard output* of a program is usually displayed on the console right after the invocation command written by the user. It is however possible to redirect the standard output to another program, using the pipe character “|”, or to a file, using the redirection character “>”.

will instead maintain the structure of the Morse description and only remove the Morse events corresponding to cusps and suitably encode the information about cusps position along the so-called extended arcs.

Remark 9.3.1 We stress once more that the program `visible` makes no attempt to produce an optimized reconstruction. Moreover, due to the way it operates, quantitative information on the spatial geometry of the visible contour is lost in the process (indeed this information is lost right at the beginning when the geometry is replaced by its Morse description), making the program not suitable for the actual recovering of the precise geometry of hidden lines in an image.

9.4 Encoding a Morse Description of the Constructed Apparent Contour

The Morse description of the computed labelled apparent contour is similar to the one used for the input visible contour (see Sect. 4.4), and uses the set of ASCII characters to encode the Morse events with a few distinctions:

- T-junctions and terminal points are now respectively substituted by crossings and cusps. As explained in Chap. 4, cusps are rotated in order to have horizontal tangent and the ASCII characters “<” and “>” are conveniently used to encode them. Crossings are encoded by the character “X”;
- labelling information must now be included (see Sect. 3.4, Chap. 3).

All characters encoding a Morse event are immediately followed by orienting and labelling information. For example, the sequence “^10” at the beginning of the description in Fig. 9.2 describes a left-oriented maximum point with labelling $d_{\Sigma} = 0$. The sequence “>0+” describes a (horizontal) cusp pointing to the right; orientation is not required, since cusps are naturally oriented upwards (respectively downwards) for cusps pointing to the right (respectively left); the labelling is $d_{\Sigma} = 0$ for the arc preceding the cusp (with respect to the arcs orientation) and $d_{\Sigma} = 1$ for the arc following the cusp (the “+” sign means that the labelling is increasing). The sequence “Xu0d0” encodes a crossing. The pair “u0” refers to the arcs in the direction North–East to South–West, which are thus oriented upwards with a labelling $d_{\Sigma} = 0$ ⁴. The pair “d0” refers to the arcs in the direction NW–SE, which are oriented downwards with a labelling $d_{\Sigma} = 0$.

Remark 9.4.1 There is an important difference with the Morse description described in Chap. 10 regarding the treatment of cusps: there they do not produce

⁴Conventionally the labelling refers to the arc below the Morse line, the upper arc can have a labelling that differs by 2. Its actual value must be inherited by information on nearby Morse events. Also by convention the information on the two arcs in direction North–East to South–West (in short NE–SW) always precedes the information on the two arcs in direction NW–SE.

Morse events, but are rather regarded merely as added information (markers) on the *extended arc* to which they belong. The software `appcontour` is however capable to recognize and then read properly a Morse description that uses the syntax produced by the program `visible`. The command `contour printmorse` can be used as a tool to convert a Morse description with cusps as Morse events to the *appcontour* syntax, typically in order to feed the result to the `showcontour` visualization program. The filter `morse2morse` is provided as a more direct way to make such a conversion

9.5 Some Examples

We conclude this chapter with a number of examples. In all figures (from Fig. 9.4 to Fig. 9.17) the left picture shows a graphic representation of the visible contour, sometimes with arrows indicating explicitly given orientations. The text in the middle is a corresponding Morse description using the syntax described in Sect. 9.2; instead of displaying the textual result given by `visible` (often difficult to interpret), we prefer to show (on the right) the result of the command `contour printmorse` followed by `showcontour` (see Chap. 10). The first command transforms the Morse description resulting from `visible` into the corresponding region description (Sect. 10.2 of Chap. 10) and then back to a Morse description, in general different (although equivalent up to an \mathbb{R}^2 -ambient isotopy) from the original one. The second command produces a graphical representation of the apparent contour starting from its Morse description. As a consequence, the resulting figure spatially corresponds to the original visible contour only up to an \mathbb{R}^2 -ambient isotopy.

In a few examples we deliberately include (partially oriented) visible contours that give rise to errors from the `visible` program.

For completeness we also include a few very simple examples.

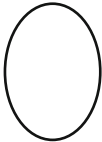
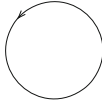
Visible contour	Morse description	Reconstructed contour
	<pre>morse { ^ ; U ; }</pre>	

Fig. 9.4 *Left:* the visible contour (a circle) is already an apparent contour. *Middle:* a possible Morse description. *Right:* the reconstructed (labelled) apparent contour using `visible`, the resulting Morse description (not shown here) is displayed using the visualization program `showcontour`

Example 9.5.1 (Sphere) The simplest possible (nonempty) visible contour consists of a (counterclockwise) circle, it can be described by the ASCII text displayed in Fig. 9.4 (middle), and happens to be a valid apparent contour (with no hidden arcs). No orienting information is required in this case. The resulting reconstructed (labelled) apparent contour is (not surprisingly) the circle itself with labelling $d_{\Sigma} = 0$.

Example 9.5.2 (Annulus 1) The visible contour of Fig. 9.5 (two concentric circles) is deliberately left unoriented. The internal circle cannot be implicitly oriented, since both orientations lead to a visible contour that satisfies the requirements of Definition 4.1.8. This leads to the error message `Insufficient orienting information` from the visible program.

Example 9.5.3 (Annulus 2) Figure 9.6 shows the same visible contour of the previous example, but with an explicit clockwise orientation of the internal circle (given in the Morse description by the “r” character after the first Morse event involving the inner circle, that is its local maximum). The resulting reconstruction (Fig. 9.6, right) does not correspond to a torus, indeed the small internal region of the visible contour is forbidden to have $f_{\Sigma} = 0$ by the reconstruction procedure, unless it is explicitly tagged as part of the background `background(K)`.

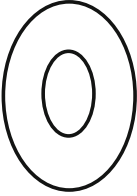
Visible contour	Morse description	Reconstructed contour
	<pre>morse { ^ ; ^ ; U ; U ; }</pre>	Insufficient orienting information

Fig. 9.5 Left: two concentric circles cannot be implicitly oriented

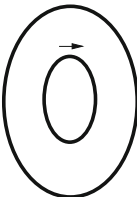
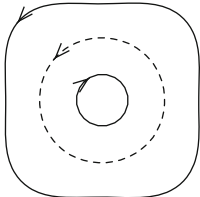
Visible contour	Morse description	Reconstructed contour
	<pre>morse { ^ ; ^r ; U ; U ; }</pre>	

Fig. 9.6 Explicit orientation (clockwise) of the inner circle

Example 9.5.4 (Annulus 3) To the Morse description of the previous Example 9.5.3 we add a marker to the small inner region in order to tag it as part of the background, thus forcing $f_{\Sigma} = 0$. This is shown in Fig. 9.7. The explicit orientation of the internal circle is not mandatory as it can be implicitly deduced by requirement (K3) of Definition 4.1.8 applied to the marked internal region. Now the `visible` program trivially reconstructs the (labelled) apparent contour of a torus with no addition of hidden arcs (where $d_{\Sigma} > 0$).

Example 9.5.5 (Annulus 4) The internal circle of the annulus of Fig. 9.8 is now oriented counterclockwise. This prevents the internal region to be part of the background. As for the previous Example 9.5.4, the program trivially reconstructs the (labelled) apparent contour with no addition of hidden arcs. The result shown on the right is the apparent contour of a small sphere in front of a bigger one.

Example 9.5.6 (Mushroom) The visible contour of Fig. 9.9 (left) reminds a mushroom. The two T-junctions necessarily give rise to crossings in the recovered apparent contour and hidden arcs with labelling $d_{\Sigma} = 2$. However, the reconstructed apparent contour corresponds to a sphere that partially occludes another sphere, and is not even a connected set. Of the infinite number of possible reconstructions of an apparent contour compatible with the given visible contour, the one selected by

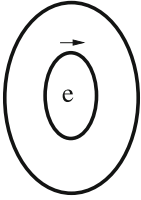
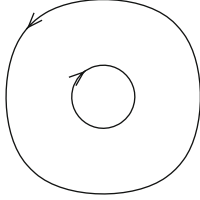
Visible contour	Morse description	Reconstructed contour
	<pre>morse { ^ ^r ; e ; U ; U; }</pre>	

Fig. 9.7 Marking the internal region as background ($f_{\Sigma} = 0$)

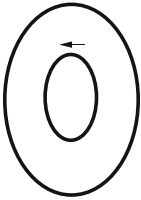
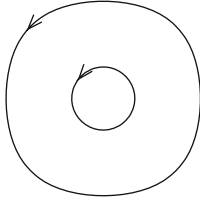
Visible contour	Morse description	Reconstructed contour
	<pre>morse { ^ ^1 ; U ; U; }</pre>	

Fig. 9.8 Orienting the inner circle counterclockwise

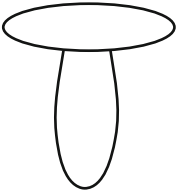
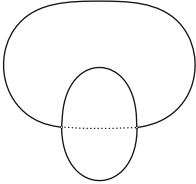
Visible contour	Morse description	Reconstructed contour
	<pre>morse { ^ . \ / . U U }</pre>	

Fig. 9.9 Visible contour of a “mushroom” (*left*). It is reconstructed as a pair of partially occluding spheres (*right*)

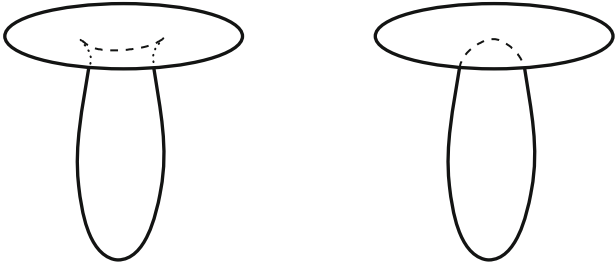


Fig. 9.10 Two possible valid reconstructions of the visible contour of Fig. 9.9

the reconstruction procedure is not necessarily the “best” one. In Fig. 9.10 we show two possible (both correct) reconstructions of this visible contour; on the left the one that actually corresponds to a mushroom, on the right a reconstruction with two spheres equivalent to the one obtained by `visible` (Fig. 9.9, right).

Example 9.5.7 (Grotto) We consider here (Fig. 9.11, left) the visible contour of the 3D scene shown in Fig. 3 of Chap. 1, reconsidered again in Chap. 3 after Definition 3.6.1. It is reconstructed “correctly” by `visible` in the labelled apparent contour on the right of Fig. 9.11, which is a deformed version of the apparent contour of Fig. 1.1, Chap. 1 (and Fig. 3.14, Chap. 3).

Example 9.5.8 (Impossible 1) The (unoriented) visible contour of Fig. 9.12 is “impossible”, in the sense that there is no way to orient its arcs to get a visible contour that satisfies requirements (K3) and (K4) of Definition 4.1.8. In particular, due to requirement (K4), the two T-junctions force incompatible orientations upon the common occluding arc. This causes `visible` to exit with the error message `Inconsistent orientation`.

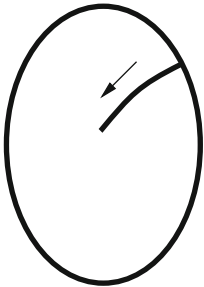
Visible contour	Morse description	Reconstructed contour
	<pre>morse { ^ ; .\ ; 'd ; ; U ; } </pre>	Wrong orientation for external boundary

Fig. 9.13 Impossible visible contour because of conflicting implicit orientation due to requirements (K3) and (K4) of Definition 4.1.8

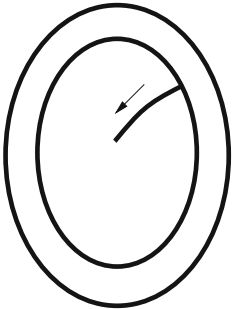
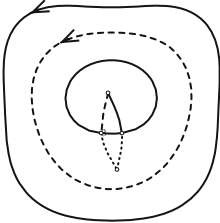
Visible contour	Morse description	Reconstructed contour
	<pre>morse { ^ ; ^ ; .\ ; 'd ; ; U ; U ; } </pre>	

Fig. 9.14 An added external circle with respect to the visible contour of Fig. 9.13

Example 9.5.11 (Not a Torus 1) This visible contour (Fig. 9.15) was already mentioned in the initial example of this chapter. By removing the “e” marking in the small internal region we force the reconstructed apparent contour to have $f_{\Sigma} > 0$ there, thus preventing the reconstruction of the torus that one would expect by looking at the visible contour. Indeed, the reconstructed 3D shape, as already mentioned in the discussion of the initial example, has the topology of the sphere, although this is not readily obvious by looking at the reconstructed apparent contour in Fig. 9.15, right. For better clarity we also include the ASCII Morse description of the reconstructed contour and a corresponding graphic drawing (Fig. 9.16). The latter graphic reconstruction is a deformation (with no change of topology) of the drawing obtained by the visualization program showcontour (Fig. 9.15, right).

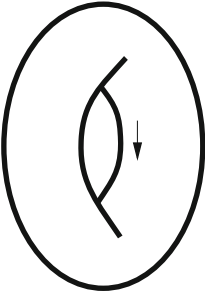
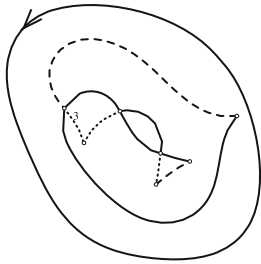
Visible contour	Morse description	Reconstructed contour
	<pre> morse { ^ , ; /. ; d ; \ ' ; ' ; U ; } </pre>	

Fig. 9.15 By omitting the “e” marking in the initial example (Fig. 9.1) the reconstructed contour corresponds to a deformed sphere. See also Fig. 9.16

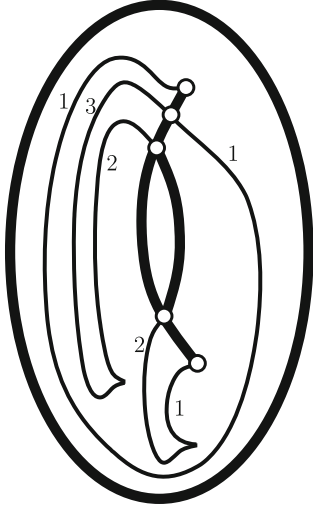
resulting Morse description	corresponding contour
<pre> morse { ^l0 ; d0 ^l1 u0 ; d0 d1 >0+ u0 ; d0 d1 ^l3 u0 u0 ; d0 d1 d3 Xu0u1 u0 ; d0 d1 d3 ^r2 u0 u1 u0 ; d0 d1 d3 u2 Xu0d0 u1 u0 ; d0 d1 d3 u2 Xd2u0 u1 u0 ; d0 d1 d3 u2 d2 >1- u1 u0 ; d0 d1 d3 >3- d2 u1 u1 u0 ; d0 d1 Ur3 d2 u1 u1 u0 ; d0 d1 d2 >2- u1 u0 ; d0 d1 Ur2 u1 u0 ; d0 Ur1 u0 ; Ur0 ; } </pre>	

Fig. 9.16 The reconstructed apparent contour corresponding to the visible contour of Fig. 9.15. Morse description (*left*) and visual, spatially corresponding, graphic reconstruction (*right*)

Example 9.5.12 (Not a Torus 2) We conclude with a visible contour obtained by vertical reflection from the one of the previous example (Fig. 9.17, left). Also we reverted the orientation of the short arc connecting the two T-junctions. After processing this visible contour with `visible` we obtain the apparent contour displayed on the right. If deformed onto the original visible contour, it turns out that the reconstructed apparent contour corresponds to a 3D shape made of two

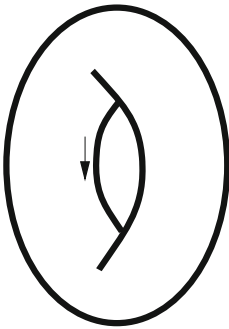
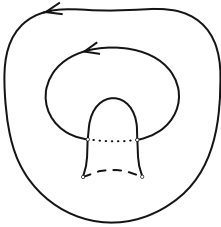
Visible contour	Morse description	Reconstructed contour
	<pre>morse { ^ , ; . \ ; d ; ' / ; ' ; U ; }</pre>	

Fig. 9.17 The visible contour on the *left* is reconstructed as a nest containing an egg

deformed spheres, a larger one deformed in the shape of a nest and a smaller one corresponding to an egg lying in the nest and partially occluded by it (rotate the image clockwise 90 degrees to have the nest and egg correctly oriented).

Reference

1. Paolini, M., Pasquarelli, F.: Appcontour: a software code to interact with apparent contours. SourceForge project: <http://appcontour.sourceforge.net> (2006)

Chapter 10

The Program “Appcontour”: User’s Guide

In this chapter we describe a software code developed by the authors and capable to manipulate the topological structure of apparent contours [15].

The `appcontour` package is a *free software* project hosted on *sourceforge.net*, its home page is <http://appcontour.sourceforge.net/> from where the source code can be downloaded, compiled and installed, following the standard procedure for Unix projects. It contains the executables “`contour`”, the core of the software, and “`showcontour`”, the visualization program, a few utilities, mainly in the form of shell scripts, and many examples of apparent contours. At the time of writing, the package is at version 2.2.1.

The topological structure of an apparent contour is invariant under smooth deformations of the plane. The software code is devised in such a way to be completely insensitive to the particular embedding of G_Σ (or $\text{appcon}(\varphi)$; see Chap. 3 for the notation) in the plane, and only captures such properties as adjacency, relative position, orientation and topological structure of the apparent contour.

This choice has some drawbacks; for example, there is no way to recover the exact shape and position of the input data, so that sometimes it is not easy to identify elements (crossings, cusps, arcs or regions) on the result of some transformation (e.g., application of one of the *moves* described in Chap. 6) of the input data.

The core of the software is an engine (`contour`) that typically reads a contour, modifies it in the requested way and prints the result, usually a different apparent contour. It works as a `unix` filter, reading data as a stream of characters from `standard input` and returning the result again as a stream of characters on the `standard output`. On a `unix` (or `linux`) system this allows concatenation of many `contour` commands within a single pipe.

Internally, the software encodes the structure of an apparent contour with the so-called *region description* to be explained in Sect. 10.2. Description of an apparent contour in a manner that can be read by the software can be done mainly in two different ways: the region description itself (encoded as a stream of characters) or a

Morse description,¹ that consists in choosing some *height function* defined on the retinal plane and describing how the height function interacts with the contour. This second way of describing a contour is convenient mainly because it can be readily obtained from a drawn sketch by introducing a division of the paper in roughly horizontal stripes bounded by lines representing constant height; see, for instance, Fig. 2.5. Each stripe must contain at most one special point of the drawing: crossing, cusp or local minimum/maximum with respect to the height function.

10.1 An Overview of the Software

Before entering in any detail we shall provide here a few basic examples, just to give a rough idea of what can or cannot be done by the `appcontour` package. If correctly installed on a Unix/Linux system, the program comes with a number of prepackaged examples of apparent contours with labelling. One of these is named `torus2` and contains the description of the apparent contour of a torus when viewed from a point located far from the vertical rotation axis and somewhat above the horizontal symmetry plane. Up to a positive C^∞ deformation of the plane, the result is shown in Fig. 10.1.²

The core program is invoked from the unix shell with the command `contour`, as in this basic example:

```
$ contour info torus2
```

The `$` sign refers to the shell prompt (e.g., `bash`) and is followed by the user input: the command name (`contour`), the requested action (`info`) and the apparent contour (`torus2`). This second argument (pointing to the file containing the apparent contour description) can be omitted, in which case the program expects the description to be entered from *standard input* (the user enters the description directly with the keyboard or, more frequently, the description is the result of a previous invocation of the program as part of a Unix pipe). If necessary the exact location (unix path) of a packaged example can be obtained using a unix command like “`contour filepath torus2`”.

¹See Sect. 2.5.

²Most of the pictures of apparent contours in this chapter were obtained with the `showcontour` software, sometimes with slight additions/modifications. The command used is “`showcontour <file>.morse --ge xfig --skipptime 0.5`”, it reads a Morse description of the contour (often obtained with `contour printmorse`) and writes a picture in `xfig` format.

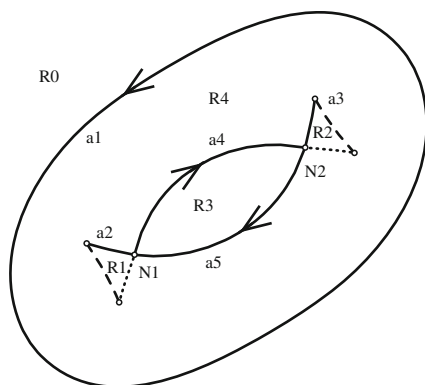


Fig. 10.1 The reference apparent contour used in the software description corresponds to an embedding of the standard torus. In this example we have five regions numbered from R0 to R4 (region R0 is the unique unbounded region), five extended arcs (a1 to a5) and two crossings (N1 and N2). The four cusps are contained in the two extended arcs a2 and a3. The labelling is 0 on the *solid lines*, 1 on *dashed lines* and 2 on *dotted lines*. Apart from R0, the enumeration of regions and arcs is chosen arbitrarily

After hitting the *enter* key, the result will be printed:

```
$ contour info torus2
This is an apparent contour with labelling

Properties of the embedded surface:
Connected comp.:    1
Total Euler ch.:    0

First order Vassiliev invariants:
Cusps:               4
Positive cusps:      2
Crossings:           2
Bennequin:           0.0

Properties of the apparent contour:
Arcs:                9
Extended arcs:       5
Link components:     2
Loops:               1
Nodes (cusps+cross): 6
Positively embedded cusps: 2
Regions:             5
Connected comp.:     2
```

The *info* action provides a display of some relevant properties of the given apparent contour. In this case the program asserts that

- `torus2` is indeed an apparent contour with labelling;
- it is the apparent contour of an embedded surface in \mathbb{R}^3 consisting of one connected component with Euler–Poincaré characteristic $\chi = 0$ (it is a torus!).

Moreover:

- a list of Vassiliev-type invariants [13] is computed, most notably the Bennequin one (see Chap. 7);
- other elementary properties readily obtainable from a sketch of the contour are displayed. In particular “Link components: 2” asserts that the critical set (see Definition 2.1.7, the critical set is a set of closed curves in the source manifold) has 2 connected components, whereas “Connected comp.: 2” refers to the number of connected components of the apparent contour itself, considered as a subset of \mathbb{R}^2 .

We can ask what direct *moves* can be applied to the apparent contour:

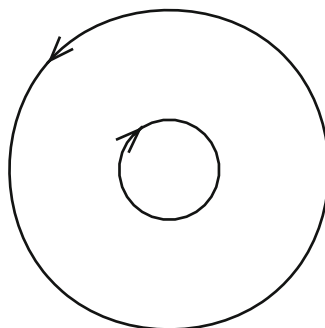
```
$ contour rules torus2
Rules that apply:
K0 B B:2 S S:2 CROL CROB CROLB CRORB
```

Action `rules` requests a list of the **simplifying** (or direct) Reidemeister-type moves (Sect. 10.4.1) or composite moves (Sect. 10.4.3) that can be legally applied to the given contour. Some of the listed rules (simple rules) are the same described in Chap. 6 under the name of Reidemeister-type moves, Definition 6.1.1, other rules are useful combinations of simple rules and their inverses. Only *simplifying* rules are listed. Inverse rules typically can always be applied, in many different places of the apparent contour; application of such rules is governed by other appcontour actions. The complete list of available rules and their description will be given in Sect. 10.4. We have:

```
$ contour rule K0 torus2
applying rule K0
sketch {
Arc 1: (0);
Arc 2: (0 1 2 1 0);
Region 0 (f = 0): () (-a1);
Region 1 (f = 4): (+a2);
Region 2 (f = 2): (+a1) (-a2);
}
```

Action `rule` followed by a rule name (K0 in the example) actually applies the rule (i.e., performs the move) and displays the result as a region description. The region description will be explained in Sect. 10.2 and the *rules* will be listed and discussed in Sect. 10.4. In this particular case the effect is to remove the two crossings by passing arc `a5` over arc `a4`, the result can be seen in Fig. 10.12 (left). It should be emphasized that `contour` refuses to apply illegal rules (rules that do not match the given contour).

Fig. 10.2 Result of double application of rule S to the contour of Fig. 10.1 as displayed by showcontour



As an example of a multiple invocation, let us apply rule S (removal of a *swallow's tail*) and then ask for the list of rules that can be applied to the result:

```
$ contour rule S torus2 | contour rules
applying rule S
Rules that apply:
S CR0R CR0LB CR2
```

Now we apply again rule S:

```
$ contour rule S torus2 | contour rule S
...
sketch {
Arc 1: (0);
Arc 2: (0);
Region 0 (f = 0): () (-a1);
Region 1 (f = 0): (-a2);
Region 2 (f = 2): (+a1) (+a2);
}
```

We can get a picture of the result by using another program, `showcontour`. This program however requires a *Morse* description of the contour (see Sects. 2.5 and 10.3.2), which is provided again by `contour` using the `printmorse` action:

```
$ contour rule S torus2 | contour rule S |
                           contour printmorse | showcontour
```

The result of this command is shown in Fig. 10.2.

We end this quick overview with

```
$ contour fg torus2
Finitely presented group with 2 generators
<a,b; abAB>
```

where `fg` stands for fundamental group.

It computes the fundamental group of the surface embedded in \mathbb{R}^3 , see Sect. 10.7.2, the result is a presentation of $\mathbb{Z} \times \mathbb{Z}$ which is indeed the fundamental group of the torus. The presentation is displayed as a list of generators (a and b in this case) followed by a list of relators where capital letters stand for the inverse of the corresponding generator, in our case the (only) relator asserts that $bab^{-1}a^{-1}$ is the identity, or equivalently that $ba = ab$. Much more interesting is the command:

```
$ contour ifg torus2
Free group of rank 1
<a; >
```

which gives the fundamental group of the solid region inside the surface: `ifg` stands for “inside fundamental group”. In this example the result is just \mathbb{Z} , the fundamental group of S^1 . Command `ofg` gives, instead, the fundamental group of the solid region outside the surface. The result is again \mathbb{Z} which essentially means that the torus is not *knotted*.

A number of `contour` commands are provided to deal with finitely presented groups (typically obtained as fundamental groups of subsets of \mathbb{R}^3 delimited by the surface), one important example is the computation of the Alexander polynomial. A description of these commands will be given in Sects. 10.7.2, 10.7.3 and 10.7.4.

10.2 Region Description

We shall use here concepts introduced in Chaps. 2 and 3 and specifically in Sect. 2.2. Basically a topological description of a labelled or unlabelled apparent contour can be obtained by focusing mainly on *regions* (the connected components of the complement of the apparent contour) and extended arcs (one-dimensional parts of the apparent contour connecting two crossings or closed loops without crossings). As a proof of concept we shall use the labelled apparent contour of Fig. 10.1 that shows a torus in its typical representation.

Neglecting cusps for the moment, we can observe that the structure of a region can be simply given by describing the boundary as an (oriented) list of arcs. A couple of caveats are:

- A region is not necessarily simply connected. In such a case its boundary has two or more connected components, one of which is the *external* boundary, the others correspond to each hole of the region. An example is region R4 of Fig. 10.1 with external boundary $a1$ and one hole described by a suitable sequence of arcs.
- Some crossing can appear more than once while travelling along the boundary of a region. This in particular happens when a crossing is adjacent to the same region on opposite quadrants. An example is given by crossings N1 and N2 that have region R4 on opposite sides (Fig. 10.1).
- Arcs can be closed (no endpoints), for instance arc $a1$ in Fig. 10.1. They can have the same crossing at both endpoints, for example the extended arc $a2$ starts and

ends at crossing N1. They can have two distinct endpoints: arc a4 starts from crossing N1 and ends at crossing N2.

10.2.1 Extended Arcs

We recall that all arcs of the apparent contour are oriented to have the higher value of f_φ (number of preimages, see Definition 2.2.5 and formula (3.5) in Chap. 3) on the left. The notion of *extended arc* was introduced in Sect. 2.5.2; in essence, extended arcs are curves connecting two possibly coincident crossings (or they are a closed curve) and may contain one or more cusps. In this way cusps are treated more as an attribute of an (extended) arc rather than as a node of the apparent contour. In Fig. 10.1 arc a2 (and similarly a3) contains two cusps; at the first cusp, the value of the labelling jumps up from 0 to 1, at the second cusp it jumps up from 1 to 2. In the end, all information for an extended arc reduce to a list of the values of the labelling d across cusps.

Each extended arc will be tagged with consecutive integers starting from 1. The numbering is arbitrary.

10.2.1.1 Streaming the Description of an Arc

We describe such an information with a string of characters as follows.

- For extended arcs that connect two distinct crossings we describe it as a list of nonnegative integers representing the values of the labelling function d , separated by spaces and enclosed in square brackets. The number of integers is the number of cusps increased by one. As an example the string “[1 2 3]”³ describes an extended arc that connects two distinct crossings with two cusps that separates the values of the labelling, listed in the order given by the arc orientation.
- For extended arcs that start and end at the same crossing we follow the same convention as in the previous case but we substitute a round parenthesis in place of the final square bracket. For example, “[0 1 2)” describes the extended arc a2 of Fig. 10.1 that connects the crossing N1 to itself. The meaning of the integer values is the same as for the previous case of distinct endpoints. This type of arc is typical when the apparent contour contains a *swallow’s tail*.
- For closed extended arcs we use round parentheses in place of the square brackets. In this case we do not have a natural starting point on the arc, so that we just choose some point (distinct from a cusp) and list the d values while traversing the arc along its orientation until we return to the chosen starting point.

³Of course the enclosing quotation marks “” and “” are *not* part of the string itself.

Note that the first and last d values must be equal. For example, the string “(0 1 0 1 0)” describes a closed arc with four cusps. The closed arc a1 of Fig. 10.1 can be described by the string “(0)”.

10.2.2 Describing a Region

Relative position of the extended arcs is recovered from a suitable description of the regions of the apparent contour. Recall that each region is a connected open set; it can be either simply connected (i.e., it has no holes) or multiply connected, in which case only a finite number of holes is allowed. For example, region R4 of Fig. 10.1 is not simply connected with one hole. Basically we list all arcs bounding a given region, in counterclockwise order, recording the arc number and an orientation. The orientation is *positive* if the arc orientation is consistent with the direction in which we are traversing the boundary (i.e., the arc is oriented counterclockwise with respect to the region), *negative* if the arc is oriented opposite with respect to the traversing direction (i.e., the arc is oriented clockwise). For simply connected regions, the boundary has a single connected component. We then select any starting arc along the boundary and traverse the boundary counterclockwise until we return to the original arc and record all encountered arcs together with their orientation (with respect to the region we are describing). The result is a circular list of pairs (ϵ, a) where the orientation ϵ belongs to $\{+, -\}$ and a is an arc. For instance, region R3 of the apparent contour of Fig. 10.1 is simply connected and bounded by arcs a4 and a5. We can select a4 as a starting arc and move along the boundary with the inside of region R3 on the left (counterclockwise). Since we are moving contrary to the orientation of a4, we have $\epsilon = -$ and start the description with $(-, a4)$ or simply by the string “-a4”. Continuing to travel along the boundary of R3 we then encounter arc a5 which is also negatively oriented and we record “-a5”. The description of R3 is then completed since we arrive again at arc a4 and the circular list describing it is “(-a4 -a5)”. A multiply connected region has a boundary consisting of two or more connected components, one of which is the *external boundary*, and the remaining components bound each hole of the region. There is no particular ordering for the holes. In this case the region description is a list of connected components of the boundary, with the external boundary being listed first and the holes listed in no particular order. The arcs of each hole are listed again counterclockwise if viewed from inside the region, which means that they are actually listed clockwise around the hole. They again form a circular list of pairs (ϵ, a) as above. An example is region R4 of the apparent contour of Fig. 10.1, its description being “(+a1) (-a2 +a4 -a3 +a5)”. Since the lists describing each connected component are circular, we could equivalently describe the region as, e.g., “(+a1) (+a4 -a3 +a5 -a2)”.

A special region is the *external region* which is the only unbounded one (e.g., region R0 in Fig. 10.1). In this case there is no external boundary, we conventionally decide that the external boundary in this case is described as an empty circular list (the empty set).

Note that any arc can appear in the description of a region at most once. Moreover each arc will be part of the description of exactly two distinct regions: the description of the region on the right of the arc will contain a reference to the arc with negative orientation, whereas the description of the left region will contain a reference to the arc with positive orientation.

As a whole, the *region description* of a labelled apparent contour is an unordered collection of descriptions of each region, one of which is the external region, with an empty external boundary. For definiteness we shall number the regions with consecutive integers starting from 0, with 0 assigned to the external region. Clearly the numbering of the regions that are not external is arbitrary.

Each region is described as explained above as a list of boundaries, the first of which is the *external boundary* and the others can be reordered arbitrarily. Each boundary is a circular list of signed arcs.

10.2.2.1 Streaming the Description of a Region

Summarizing, the region description can be encoded as a sequence of characters as follows. Each connected component of the boundary, starting with the external boundary, is described as a list of arc names prefixed by a '+' or '-' sign according to orientation, separated by spaces and enclosed in parentheses.

As an example, as we have already seen, region R4 of Fig. 10.1, which is not simply connected, will be encoded with the character string "(+a1) (-a2 +a4 -a3 +a5)". The external region R0 is described by the string "() (-a1)".

We conclude this section with an example of an apparent contour featuring three closed arcs labelled 0 and one inside the other (see Fig. 10.3). All three arcs are

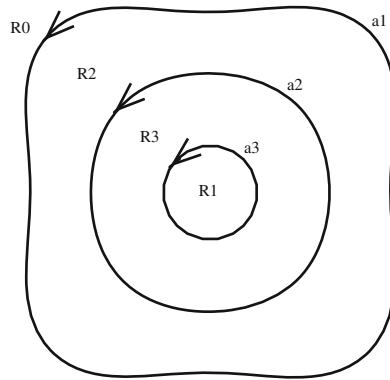


Fig. 10.3 This apparent contour represents three spheres of increasing size at increasing distance. Its region description allows to recover the relative position of each one of the three arcs, that are otherwise indistinguishable. With the exception of the external region R0, the numbering of regions and arcs is arbitrary. Reversing the orientation of the intermediate arc produces the apparent contour of a big torus and a small sphere

described with the same string “(0)”. The information about the relative position of the arcs is encoded in the description of the regions: region R0 (the external region) is described as “() (-a1)” which entails that arc a1 is the outermost one; region R2 is described as “(+a1) (-a2)” which entails that arc a2 is the boundary of the hole of a region (R2) having arc a1 as external boundary; region R3 is described as “(+a2) (-a3)” which entails that arc a3 is the boundary of the hole of R3; finally, region R1 is described as “(+a3)” which means that there is nothing inside arc a3.

10.2.3 Completeness of the Region Description

We shall say that two region descriptions are equivalent if they are the same up to:

- (1) renumbering of the extended arcs;
- (2) renumbering of the regions different from the external region;
- (3) reordering of the *holes* of each region, i.e., the connected components of the boundary of the region different from the external boundary (simply connected regions have no *holes*).
- (4) changing the extended arc used as a starting point in the description of the circular list of boundaries (see Sect. 10.2.2.1). That is, as part of a region description the string “(-a2 +a4 -a3 +a5)” is equivalent to “(+a4 -a3 +a5 -a2)”;
- (5) changing the starting point in the description of a closed extended arc (with cusps) when listing the values for the function d (see Sect. 10.2.1.1). That is, the arc description “(0 1 0 1 0)” is equivalent to “(1 0 1 0 1)”.

The region description satisfies properties similar to those satisfied by the Morse description and listed at the end of the preamble of Sect. 2.5, namely it is

- **Finite:** the description requires only a finite sequence of symbols (taken from a finite set).
- **Complete:** two apparent contours with the same region description are equivalent in the sense of Definition 2.4.2 (modified for the case of labelled contours, see Definition 3.5.1).
- **Essential:** conversely, two diffeomorphically equivalent apparent contours have equivalent region descriptions.

The *completeness* property follows from the existence of an algorithm that can recover a possible Morse description from the region description of an apparent contour. It is then clear that apparent contours with the same region description admit the same Morse description and we can conclude our claim using the completeness property of the Morse description (see Sect. 2.5.4). We shall not describe here the mentioned algorithm; it is encoded in the `contour` software (function `printmorse`), whose source code is available, and works by pushing an ideal Morse line through the region description.

The *essentiality* property is a direct consequence of the notion of equivalence of region descriptions, which in turn is suggested in a straightforward manner by the definition of the region description. Namely, if $\Phi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is a positive diffeomorphism giving the equivalence of two apparent contours G and H (see Definition 2.4.2), then Φ clearly defines a bijection between the extended arcs of G and the extended arcs of H ; corresponding extended arcs will have the same description since Φ is a positive diffeomorphism. In addition Φ also induces a bijection between the regions of G and the regions of H with the external region of G corresponding to the external region of H . Moreover, for each region R of G , Φ defines a bijection between the connected components of ∂R and those of $\partial(\Phi(R))$, the external boundary of R being mapped onto the external boundary of $\Phi(R)$. Again, Φ being a positive diffeomorphism, the description of each of the connected components of the boundary of each region is the same as the one of the corresponding boundary under Φ . The equivalence of the two region descriptions then follows.

10.3 Encoding an Apparent Contour with Labelling

10.3.1 Region Description as a Stream of Characters

The region description explained in Sect. 10.2 provides a way to encode an apparent contour to be used by our software. We first choose a numbering of the extended arcs starting from 1 and of the regions, starting from 0 (the external region). The syntax of the region description is then best understood by looking at Fig. 10.4, where we encode the apparent contour of Fig. 10.1.

```
#
# region description of the apparent contour
# of a torus in a typical position
#
sketch {
Arc 1: (0);
Arc 2: [0 1 2);
Arc 3: [2 1 0);
Arc 4: [0];
Arc 5: [0];
Region 0: () (-a1);
Region 1: (+a2);
Region 2: (+a3);
Region 3: (-a4 -a5);
Region 4: (+a1) (-a2 +a4 -a3 +a5);
}
```

Fig. 10.4 Region description corresponding to Fig. 10.1 as a stream of characters

The BNF (Backus-Naur form) of the description syntax is the following:

```

<contour-region-description> ::= "sketch" "{" <arcs> <regions> "}"
<arcs> ::= <arc> | <arc> <arcs>
<arc> ::= "Arc" <integer> ":" <arc-description> ";"
<regions> ::= <region> | <region> <regions>
<region> ::= "Region" <integer> ":" <region-description> ";"

```

where `arc-description` and `region-description` are described in Sects. 10.2.1.1 and 10.2.2.1 respectively. Their BNF is the following:

```

<arc-description> ::= "[" <d-values> "]" | "[" <d-values> "]"
                  | "(" <d-values> ")"
<d-values> ::= <integer> | <integer> <d-values>

```

```

<region-description> ::= <component> | <component> <region-description>
<component> ::= "()" | "(" <signed-arcs> ")"
<signed-arcs> ::= <signed-arc> | <signed-arc> <signed-arcs>
<signed-arc> ::= <sign>"a"<integer>
<sign> ::= "+" | "-"

```

All lines starting with the pound character # are comments and will be ignored by the software. A description can be fed to the program as standard input or (more often) saved into a file that will be read by the program.

10.3.2 Morse Description

A more useful way to describe an apparent contour follows directly from the Morse description explained in Sect. 2.5. If the descriptive map m is the identity, the horizontal Morse lines m_λ can be visualized⁴ as a single *sweep line* ℓ that moves across the whole of \mathbb{R}^2 from top to bottom traversing the apparent contour G_Σ . We shall then record all relevant events that occur during the traversal at the critical levels. One important difference with the requirements of Sect. 2.5 is that we allow here for the presence of multiple critical points at the same level. This allows for more concise Morse descriptions and yet provides a complete description of the apparent contour. Also we allow for (a finite number of) regular levels (no critical points), this is sometimes useful in order to improve the readability of the description. For a finite set of decreasing values $1 > \lambda_1 > \dots > \lambda_n > 0$, we investigate the intersection points of the sweep line with the apparent contour and list them from left to right. The chosen levels λ_i , $i \in \{1, \dots, n\}$ must include **all** the critical levels of the descriptive map. Adding regular levels to the description, i.e., levels with only transversal intersections at points that are not crossings (cusps

⁴Here $-\lambda$ has the meaning of time.

Fig. 10.5 Morse description corresponding to Fig. 10.1 as a stream of characters

```
#
# morse description of the apparent contour
# of a torus in a typical position
#
morse {
    ^10 ;
    / \ ;
    / \ ^r0 ;
    / \ / \ | ;
    | (d0++ x \ x |u2-- ) ;
    \ / \ / \ ;
    \ U U10 U ;
    \ / ;
    U ;
}
```

are treated in a special way, as we shall see shortly), is harmless and sometimes useful to obtain a description that produces a printout graphically reminiscent of the shape of the apparent contour. Thinking of the sweep line as horizontal is of course just conventional, however its orientation when listing the intersection points is important. For instance, if the sweep line is vertical and moves from left to right, events at a given level will be listed from bottom to top.

Cusps are treated as marked points along extended arcs (compare Sect. 2.5.2), hence the intersection of the sweep line with a cusp is considered as regular. Recall here that for a descriptive map we require that the Morse line (the sweep line, as we call it in this section) to be tangent to cusps.

Following the final part of Sect. 2.5.3 we now simply encode the symbols identifying the type of intersection using typographical characters that can be written into a description file.

We illustrate this in Fig. 10.5 with a Morse description of the torus of Fig. 10.1; it can be obtained by traversing the drawing of Fig. 2.8 of Chap. 3 with a vertical sweep line. Observe the preamble “morse {” which instructs the program that the apparent contour is described by means of a Morse description. Each line of text is terminated with a semicolon “;” indicating the end of the list of symbols corresponding to a given position of the sweep line. The symbols \swarrow and \searrow used in Sect. 2.5.3 are rendered here by the typographic character “^” followed by “l” (left) or “r” (right) to indicate the orientation, the following integer (“0” for the first line) gives the labelling of the corresponding arc. Similarly the typographical character “U” is used in place of \smile and \frown . The indication of the labelling is not required for all symbols, since in most cases the program can infer it from the context, and the same holds true for the orientation. Transversal intersections \uparrow and \downarrow can be rendered by a number of equivalent typographical characters: in our example, “|”, “(”, “)”, “/”, “\” all have identical meaning for the program and are differentiated only to improve the graphical rendering of the text. Each of these characters can be followed by the orientation indication, which in this case can be “u” or “d” (up or down) and by the value of the labelling.

Crossings are indicated with the “X” character and can be followed by two orientations and two labellings (with appropriate syntax) referring to the two

arcs directed downwards (for a horizontal sweep line). For instance, the sequence “Xu0d2” refers to a crossing for which the arc leaving in the south–west direction is oriented upwards with a labelling of 0 and the arc leaving in the south–east direction is oriented downwards with a labelling of 2. Information on the other two arcs towards the north cannot be given here and must be inherited from information given at other Morse events referring to the same extended arc.

Multiple spaces are inessential and only used here to give a better idea of the layout of the apparent contour directly by inspecting its Morse description.

For extended arcs containing one or more cusps, the required information must be given once (attached to any of the typographical characters involving that arc) as a list of integers or (more briefly) by the labelling before the first cusp (according to the arc orientation) followed by a sequence of “+” or “-”, one for each cusp and according to the fact that the labelling increases or decreases by one. This is illustrated in the example of Fig. 10.5 twice: the sequence “(d0++” indicates that the extended arc corresponding to this transversal intersection (“(d” marks a transversal intersection oriented downwards) has initially a labelling of 0, then there is a cusp with the labelling increasing from 0 to 1 (the first “+”), then there is a second cusp with the labelling increasing from 1 to 2. The sequence “|u2--” similarly indicates an extended arc with two cusps and a labelling with values 2, 1, 0. For apparent contours without labelling (as for $\text{appcon}(\varphi)$) we can use a sequence of “c”, each representing a cusp.

A variant of the Morse description illustrated above requires the indication of cusps individually as Morse events with the typographical characters “>” and “<” corresponding respectively to the symbols \succ and \prec . In this case cusps are treated as specific Morse events in the spirit of the Morse description as explained in Sect. 2.5. This variant was introduced mainly in order to allow to interface the program `appcontour` with the program `visible` (see Chap. 9) that implements the reconstruction procedure illustrated in Chap. 4. A reference guide for the `visible` program is included in Chap. 9.

10.3.3 Knot Description

The *knot description* is an alternative and simpler way to describe a class of solid shapes called *handlebodies* [18]. These are essentially a thin tubular neighbourhood of 1D subsets of \mathbb{R}^3 composed by a disjoint union of closed arcs and arcs that connect triple junctions and/or endpoints, see Fig. 10.6. An important special case of handlebody corresponds to a thin tubular neighbourhood of a *tame knot* (smooth embedding of \mathbb{S}^1 in \mathbb{R}^3) or of a *tame link* (smooth embedding of a finite number of copies of \mathbb{S}^1 in \mathbb{R}^3). We refer to [10] and [18] for basic terminology about knot theory.

A standard way to describe a knot or link is by means of its *diagram*, which is a projection onto a plane drawing composed by arcs and crossings. At a crossing we need information about which one of the two crossing strands passes over the other.

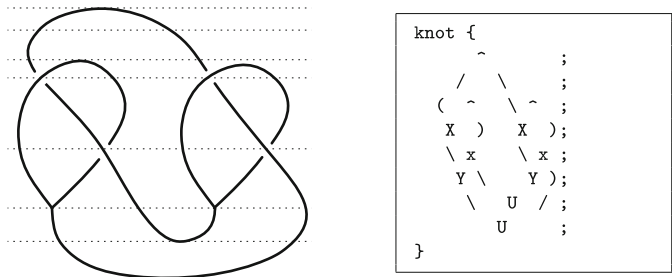
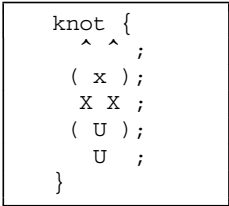


Fig. 10.6 Knot description of a tubular neighbourhood of the system of curves shown in the *left*

If S^1 is substituted by a thin pipe generically positioned in \mathbb{R}^3 and we consider the apparent contour of the resulting system of pipes, we end up with a drawing that outlines long stripes that resemble a system of *streets* that can cross over one another at *bridges*. The value of f_Σ is 2 inside the streets and 4 where one street crosses over another.

Similarly, handlebodies can be described by adding triple junctions, where three streets meet together. Although not strictly necessary to model handlebodies, we also allow for endpoints, that are a dead-end of a street.

Instead of describing the boundary of the streets, it is much more convenient to model, with a Morse-like description, the centre lines. We best explain the syntax with the simple, yet interesting, example of a solid torus knotted as in the trefoil knot. It produces the apparent contour shown in Fig. 10.18:



The typographical characters “(”, “)”, “[”, “/”, “\”, “^”, “U” have the same meaning as for the Morse description (Sect. 10.3.2) without the necessity to add information neither about orientation (knots/links are not necessarily oriented) nor about the labelling. There are two symbols that can be used for a crossing: lowercase “x” and capital “X”; this allows to distinguish between the case of the overpass going in the northwest–southeast direction (“x”) from the case of the overpass going in the northeast–southwest direction (“X”).

We allow for triple junctions, modelled with the character “Y” (two streets above the sweep line that meet one street below the sweep line) or the character “h” (one street above the sweep line that meets two streets below the sweep line).

The character “.” models a street that ends at the sweep line from above. The character “,” models a street that ends at the sweep line from below.

An interesting example is illustrated in Fig. 10.6; the solid shape E is obtained by taking a thin tubular neighbourhood of the set on the left of the figure, and its knot description is shown on the right. Note that the second row of characters of

the knot description corresponds to a noncritical Morse level and is only included in order to make the knot description more self-explanatory.

The complement in \mathbb{S}^3 of the example of Fig. 10.6 can be deformation-retracted onto the Klein bottle with a disk removed, an example that will be considered in Sect. 10.15.

10.4 The Rules (Reidemeister-Type Moves)

The main feature of the software is its ability to list and apply the Reidemeister-type moves listed in Chap. 6, together with their inverses. Within this chapter we shall call such moves **simple rules**. The **inverse rules** are a *time-reversed* version of the simple rules; the **composite rules** are the result of combining a finite sequence of simple and inverse rules.

These *rules* transform the apparent contour G_Σ into a new one $G_{\Sigma'}$, with $\Sigma' = \partial E'$, having a *different* topological structure but corresponding to an \mathbb{R}^3 -ambient isotopy of E into E' . In other words (see Chap. 6), there exists a smooth path of compactly supported diffeomorphisms of \mathbb{R}^3 that connects the identity to a smooth deformation that takes E into E' .

We have shown in Corollary 6.6.5 that two 3D shapes E and E' are \mathbb{R}^3 -ambient isotopic if and only if there exists a finite sequence of simple rules or their inverses, and a finite number of \mathbb{R}^2 -ambient isotopies, that transform the apparent contour of Σ into the apparent contour of Σ' .

It is worthwhile to notice here that, in general, for a given apparent contour, only a subset of the rules can be applied, and a given applicable rule can often be applied in different ways or in different parts of the apparent contour.

10.4.1 Simple Rules

We follow the notation of [13] and of Sect. 6.1. The presence of the labelling makes our list of moves more rich with respect to the ones in [13]; however, the type of moves is the same; see Remark 6.1.3 for more.

Rules are applied with the command `contour rule <rule>` where *<rule>* is one of the simple rules listed in the sequel. The software searches for a place in the apparent contour where the required rule can be applied, and applies it as soon as a suitable place is found. It is possible to apply the rule at the i -th suitable place by using the syntax `<rule> : i`. The name of the rule is therefore the same as that printed with the command `contour rules`, which lists all simple rules that are applicable. For example, the command `contour rules three_spheres` lists `K2 K1b K1b:2 T T:2` as the applicable rules (Fig. 10.7, left), in particular rules `K1b` and `T` can both be applied in two different places of the contour. Applying rule `K2` produces the apparent contour shown in the middle picture of Fig. 10.7: the picture displayed is produced by the software and should be interpreted, as usual, up

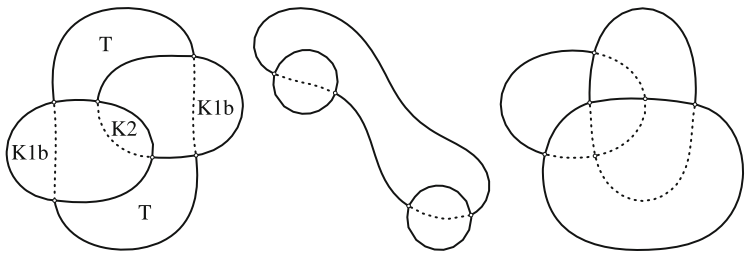


Fig. 10.7 An example with three partially occluding spheres and the applicable simple rules, written inside the region involved (*left*); result after applying rule K2 (*centre*); result after applying the lower occurrence of rule T (*right*)

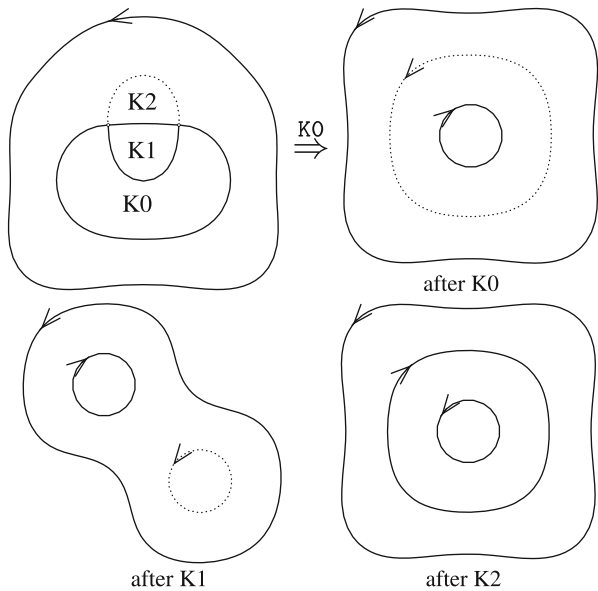


Fig. 10.8 An example with a (small) sphere behind a torus where rules K0, K1 and K2 can be applied with the results shown respectively in the *top-right*, *bottom-left*, *bottom-right* pictures

to smooth deformations of \mathbb{R}^2 . Applying the first occurrence of rule T to the contour shown on the left of the picture produces the result displayed on the right.⁵

10.4.1.1 The κ Rules (kasanie)

The κ rules (see Figs. 10.7 and 10.8 for examples involving them) are of four types: for K0 the two arcs are oriented in such a way that the lowest value of f (zero in

⁵By inspecting the resulting apparent contour it turns out that the first occurrence of rule T is the lowest one shown in Fig. 10.7, left. Applying the second occurrence of rule T (referring to it with $T : 2$) produces a result that is mirror-equivalent to Fig. 10.7, right, but not diffeomorphically equivalent in the sense of Definition 2.4.2.

the pictures) is in the internal region bounded by two arcs, i.e., the two surfaces fold in such a way that they do not occlude the internal region. Rule K2 features the highest value of f (four in the pictures) in the internal region (which is thus occluded by both folding surfaces). We have two remaining rules: K1 and K1b, where both surfaces fold on the same side, the two rules map into each other after a front–back reflection of the shape E .

10.4.1.2 The T Rules

Three folding surfaces are involved in these rules. They can be ordered on the basis of their relative distance from the point of view, using the values of the labelling. An example of application of a T rule is shown in Fig. 10.7. There is a central “triangular” region R with a boundary composed of three arcs, each arc can be oriented clockwise (the corresponding surface folds away from the region R) or counterclockwise (region R is covered twice by the corresponding folding surface) for a total of 16 different situations, also counting those obtained by a left–right reflection symmetry. However the software groups all these variants together as a T rule. As already noted in Chap. 6, the inverse of a T rule is another T rule, indeed such moves do not change the complexity of the apparent contour in terms of number of crossings, cusps, arcs and regions.

10.4.1.3 The L Rule (Lip)

An example of application of the *lip* rule L can be seen in the top-middle picture of Fig. 10.9.

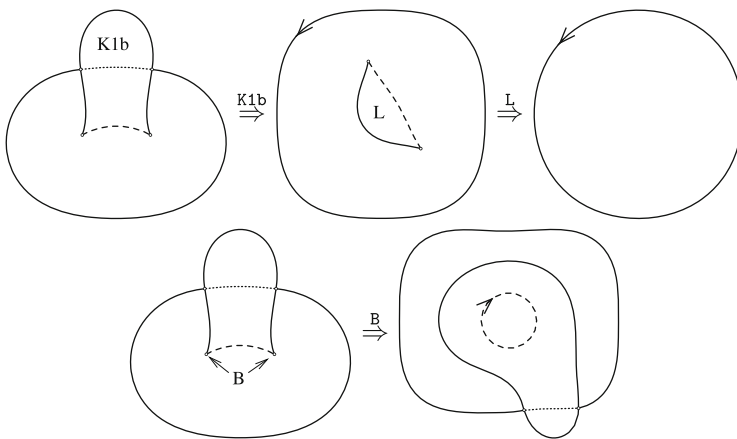


Fig. 10.9 This apparent contour allows the simple rules K1b (*top row*) and B (*bottom row*); after applying rule K1b, rule L becomes applicable, leading to the *top-right circle*

10.4.1.4 The B Rule (Beak-to-Beak)

Figure 10.9 bottom shows an example of application of the B (beak-to-beak) rule that melts together two cusps if the appropriate consistency conditions on the labelling are met.

10.4.1.5 The S Rule (Swallow’s Tail)

Two occurrences of the *swallow’s tail* rule can be applied to the apparent contour of a torus of Fig. 10.1, as already anticipated in Sect. 10.1 leading to the simpler apparent contour of Fig. 10.2.

10.4.1.6 The C Rule (Cusp-Fold)

See Fig. 10.10 for an example of this rule.

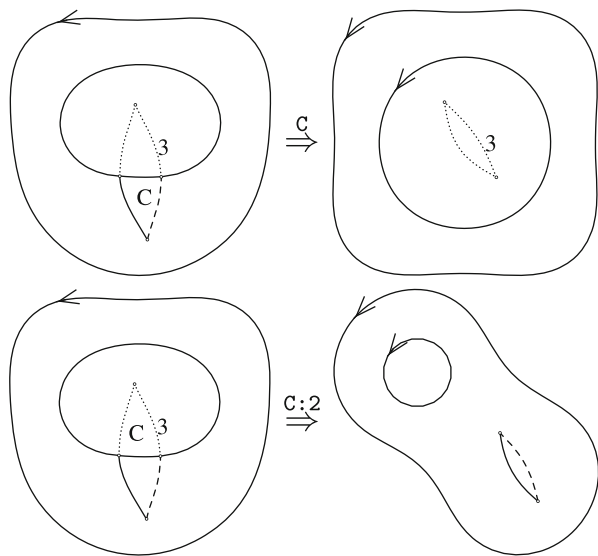


Fig. 10.10 This is the apparent contour of a small sphere in front of a larger sphere with a partially occluded wrinkle. It allows two different applications of rule C, the result of which is shown in the two *top* and *bottom* rows. In both cases rule L could be subsequently applied

10.4.2 A Nonlocal Effect of the B Rule

Suppose that a beak-to-beak rule B takes place in region R and that region R is not simply connected, its boundary being composed by $n > 1$ connected components. Suppose moreover that the two cusps involved in the B rule belong to the same connected component of the boundary of region R . This entails that region R will be divided into two regions after application of the rule.

In such a case the result of applying the rule is not univocally determined. To explain this assertion suppose for definiteness that the cusps belong to the external boundary (a similar reasoning applies in the other cases). We have $n - 1$ holes of the original region. That part of the apparent contour that is contained in a hole (including its boundary) will be called *island*; by abusing the term, we shall also refer to an *island* for the external part of the apparent contour (the external boundary of the region and everything outside of it). Each island must be placed in one of the two newly formed regions. This leads to some degree of uncertainty when applying rule B under particular circumstances and is signaled with a warning by the software. It is possible for the user to force the positioning of the islands by means of the option `--ti <i>`. The integer i is interpreted as a sequence of digits 0 or 1 in base 2, the value of each digit indicating the destination of each island of the region. Another way to control the positioning is to add an extra specification to the rule name, the complete syntax is then “`<rule>:n:i`” where n indicates which occurrence of the rule we are referring to (rules can often be applicable in more than one place) and i has the same meaning as for the `--ti` option.

An example of this situation will be encountered when applying rule B within the composite rule CR4, see Fig. 10.15.

10.4.3 Composite Rules

Composite rules are particular sequences of simple rules that include at least one inverse rule and that have an overall simplifying effect.

10.4.3.1 The CR0 Composite Rule ($B^{-1}S$)

See Fig. 10.11 for an example. There are actually four variants: CR0L is shown in the Figure, and consists in applying the inverse of B (beak-to-beak) on the two arcs with labelling 1 and 2 followed by a swallow’s tail S (the same composite rule can of course be applied if all labellings involved are incremented of the same amount). The other variants: CR0R, CR0Lb and CR0Rb are obtained with a left–right reflection, a front–back inversion or both, applied to the rule CR0L.

Notice that although the final result is simpler than the starting apparent contour, we cannot obtain that result without resorting to the inverse of a simple rule.

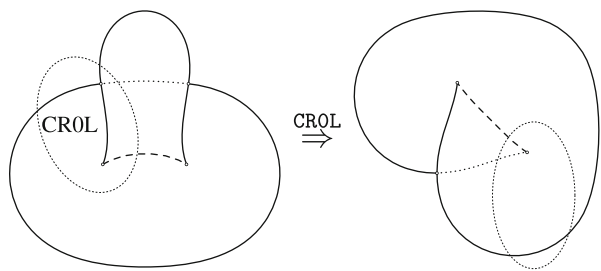


Fig. 10.11 In the indicated zone we can apply the CR0L composite rule leading to the result shown on the *right*

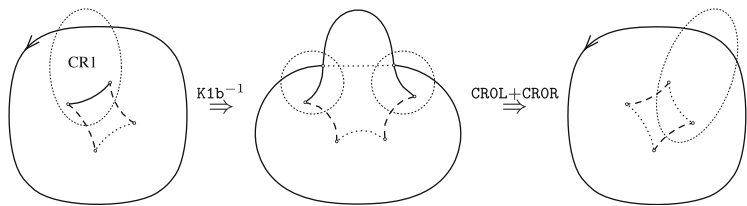


Fig. 10.12 In the indicated zone we can apply the CR1 composite rule leading to the result shown on the *right*

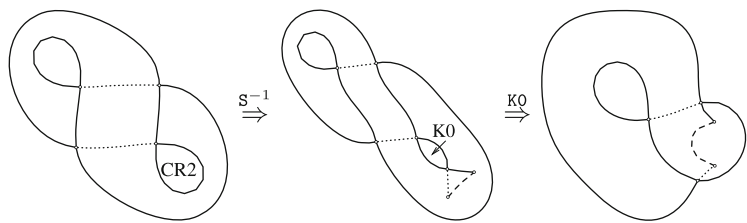


Fig. 10.13 In the indicated zone we can apply the CR2 composite rule leading to the result shown on the *right*

10.4.3.2 The CR1 Composite Rule

See Fig. 10.12 for an example. It can be obtained by applying the inverse of a K1b rule followed by two applications of the composite rule CR0 (specifically a CR0L and a CR0R).

10.4.3.3 The CR2 Composite Rule

See Fig. 10.13 for an example. It can be applied to regions surrounded by a single arc with coincident endpoints and no cusps. It is obtained by applying the inverse of a S rule followed by the rule K0. The resulting apparent contour has one less crossing and two more cusps.

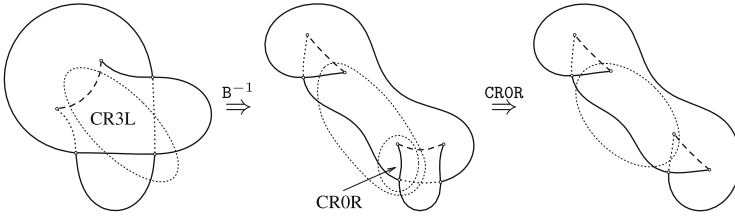


Fig. 10.14 In the indicated zone we can apply the CR3L composite rule leading to the result shown on the *right*

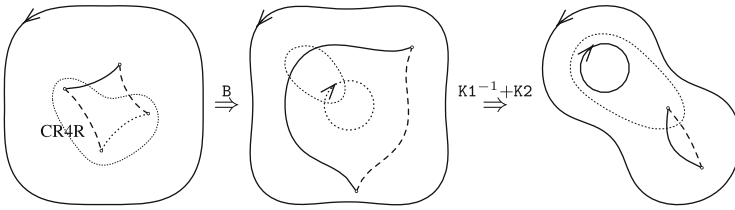


Fig. 10.15 In the indicated zone we can apply the CR4R composite rule leading to the result shown on the *right*

10.4.3.4 The CR3 Composite Rules

See Fig. 10.14 for an example. Rule CR3L is obtained by applying the inverse of a B rule followed by the composite rule CR0R. Left–right symmetry leads to rule CR3R. The resulting apparent contour has one less crossing and two more cusps.

10.4.3.5 The CR4 Composite Rules

See Fig. 10.15 for an example. Rule CR4R is obtained by applying rule B followed by the inverse of rule K1 and finally by rule K2. Left–right symmetry leads to rule CR4L, front–back symmetry leads to the remaining two rules CR4Lb and CR4Rb.

Notice that when applying rule B we are precisely in the situation described in Sect. 10.4.2, now with the two cusps belonging to the connected component of the same *hole* of the interested region. Similarly to what done in Sect. 10.4.2 we now must specify precisely how to connect the two cusps: by circumnavigating the *hole* from above or from below, so to speak.

10.4.3.6 The A1 Composite Rule

This rule can be applied whenever we have an annular region with both boundaries positively oriented (i.e., the external boundary is an S^1 oriented counterclockwise, the internal boundary is an S^1 oriented clockwise) and the labelling of the internal

\mathbb{S}^1 is of one unit larger than the labelling of the external \mathbb{S}^1 . In this situation we can apply the inverse of a B rule followed by an L rule and annihilate both \mathbb{S}^1 .

10.4.3.7 The A2 Composite Rule

Similar to rule A1 but with the internal labelling of one unit lower than the labelling of the external \mathbb{S}^1 .

10.4.3.8 The TI Composite Rule

This is obtained by applying the inverse of a B rule that melts two regions followed by the same B rule but with a different choice of the position of the *holes* that have to be placed in one of the two original regions. See the discussion in Sect. 10.4.2. The option `--ti <i>` can be used to indicate which holes have to be transferred. An alternative way is to indicate the rule as `TI : i` or `TI : n : i`; in the latter syntax, *n* indicates that we require the application of the rule at the *n*-th found location at which the rule can be applied, while *i* must be interpreted as a sequence of binary digits each of which decides the placement of a different hole.

10.4.4 Inverse Rules

With the exception of T, the inverses of the simple rules produce a more complex apparent contour; moreover, each of the inverses is typically applicable in a large number of positions for a given apparent contour. This is the reason why applicable inverse rules are not listed with the command `contour rules`.

There are specific commands to obtain a list of applicable inverse rules, also indicating the correct syntax to be used for the actual application of each occurrence.

Note that quite often we fall in the uncertainty situation described in Sect. 10.4.2 requiring the use of the `--ti <i>` option, or the more specific rule indication such as `INVK2 : 1 : i` if we want to precisely control the outcome of the application of the rule.

For convenience, the rules `INVK0`, `INVK1`, `INVK1b`, `INVK2` and `INVB` (respectively inverses of the rules `K0`, `K1`, `K1b`, `K2` and `B`) are all grouped together in the class of “mergearcs” rules because they all involve merging pairs of arcs facing the same region. For any pair of such arcs, which one of the mentioned inverse rules is applicable depends on their relative orientation and on the values of the labelling. See also the description of the `listma` command in Sect. 10.10.1.

Rule `INVL`, inverse of the *lip* rule `L`, can be applied on any region *R* with $f(R) > 0$, and in that case it can be applied to any one of the $f(R)$ strata. See also the description of the `listinvl` command in Sect. 10.10.1.

Rules `INVS` and `INVSb` are symmetric to each other by a front–back reflection and are both inverses of the *swallow’s tail* rule `S`. See also the description of the `listinvs` command in Sect. 10.10.1.

Finally `INVC` is the inverse of rule `C` and involves a cusp and an arc facing the same region with appropriate orientations and suitable values of the labelling. Application of `INVC` produces a new small region having the shape of a very sharp triangle. See also the description of the `listinvc` command in Sect. 10.10.1.

10.5 Surgeries on Apparent Contours

The structure of an apparent contour can be modified in ways that correspond to *surgery* applied to the solid shape E . A typical surgery consists in the removal of two disks in selected positions on ∂E and glueing the two boundaries together in a way that respects the surface orientation. With respect to the projection π we distinguish two types of surgery: vertical and horizontal. Other type of topological modifications like adding/removing spheres and the result of applying symmetries are listed in the reference guide, Sect. 10.10.2.

10.5.1 Vertical Surgery

The portions of the surface ∂E involved by the surgery are two consecutive strata of some region R of the apparent contour G_Σ ; the surgery acts in the projection direction and the final effect in the apparent contour is the addition of a closed arc inside R with appropriate labelling. The simplest possible application of this surgery converts a sphere into a torus, see Fig. 10.16 (left):

```
$ contour punchhole sphere -r 1:0 | contour characteristic
Euler characteristic: 0
```

If the two disks belong to different connected components of ∂E , this surgery glues them together, thus reducing by one the total number of connected

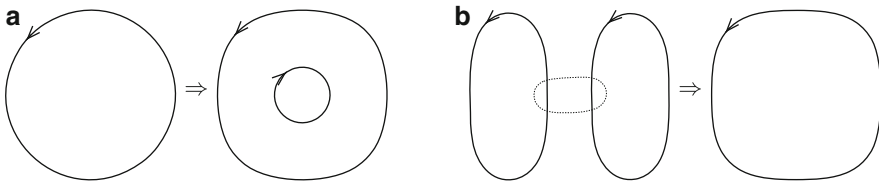


Fig. 10.16 (a): Vertical surgery (action `punchhole`) applied to the apparent contour of a sphere. (b): Horizontal surgery (action `gluearcs`) applied to the apparent contour of two disjoint sphere

components. If the two disks belong to the same connected component of ∂E , the effect of the surgery is to reduce the Euler–Poincaré characteristic by 2 units.

10.5.2 Horizontal Surgery

The surgery is done between two portions of the surface that correspond to two facing arcs of the apparent contour. The two arcs must be oriented negatively with respect to the common region and must have the same labelling. A simple example is shown in Fig. 10.16 (right) converting the apparent contour of two spheres to the contour of a single sphere:

```
$ contour gluearcs -a 1:0 -a 2:0 disjoint_spheres
sketch {
Arc 1: (0);
Region 0 (f = 0): () (-a1);
Region 1 (f = 2): (+a1);
}
```

This is the region description of the resulting contour (a single \mathbb{S}^1) of a sphere.

The topology of the resulting surface ∂E changes in the same way as for the vertical surgery.

10.6 Canonical Description and Comparison

As observed in Sect. 10.2.3 we can have many equivalent region descriptions (encoded differently as a stream of characters) corresponding to the same apparent contour, by choosing a different numbering of arcs and regions (except for the external region) and a different ordering of the *holes* of each region. Therefore it would be useful to define a procedure capable of selecting a *canonical* representative among all equivalent region descriptions: finding an optimal choice of a representative will be called *canonization procedure*.

Since each different region description is, in the end, just a string of characters, one way of selecting a representative would be simply to choose the smallest descriptive string with respect to some lexicographical ordering. However this would be inconvenient for two reasons. First, cycling among all equivalent region descriptions is computationally impractical, because the number of different reorderings of arcs and regions grows exponentially fast with respect to the number of arcs and regions; second, it would seem more natural to devise a way to compare region descriptions based on intrinsic (topological) properties and not on an arbitrary numbering.

Since an apparent contour is essentially a graph, it is natural to briefly discuss the isomorphism problem for graphs, and then the more difficult task of finding a canonical description (or canonical representation).

10.6.1 *On the Isomorphism Problem for Graphs*

The graph isomorphism problem is one of the most studied in computer science; it is in NP, no polynomial-time algorithm is known, but it is not known whether it is NP-complete [9]. For graphs with additional properties the situation is much better: for trees there are linear-time algorithms, whereas the case of planar graphs (the case we are interested in) is in Log space (a class contained in P) [4], and an $\mathcal{O}(n^2)$ algorithm for the special case of planar 3-connected graphs is known since 1966 [20]. The canonization problem is clearly harder, since the ability to find a canonical representative for a graph gives a trivial way to test for isomorphism.

However, recalling the notion of equivalence stated at the beginning of Sect. 10.2.3, we must here emphasize that the canonization problem for apparent contours can take advantage of the additional structure given by the embedding in the plane. We also note that the notion of equivalence of two apparent contours that we are interested in requires that the graph isomorphism must be consistent with the particular planar embedding of the two graphs, which makes a big difference with respect to the wider problem of graph canonization even in the case of planar graphs. This is so because planar graphs are just graphs that *admit* at least one embedding in the plane, but generally can admit more than one such (non-equivalent) embeddings. Triply connected graphs are special, as they cannot admit multiple non-equivalent planar embeddings (of course there might be none).

A planar embedding of a graph induces an additional structure. For each node, the incident arcs are organized in a circular list (obtained by making a counterclockwise turn around the node), i.e., each node is equipped with a cyclic permutation of its incident arcs. We are interested in isomorphisms of graphs that are also consistent with this additional structure.

10.6.2 *The “Regions” Graph: \mathcal{R} -Graph*

Due to the structure of the region description, it is more useful to work with a kind of *dual* of the apparent contour. Given a connected component C of the apparent contour, we consider a graph (the *regions*-graph, or \mathcal{R} -graph) having the regions (connected components of $\mathbb{R}^2 \setminus C$) as nodes. Each extended arc of C defines a link connecting the two nodes corresponding to the regions at the two sides of it; we orient this link from the region with lower value of f (the region on the right) to the region with higher value of f (the region on the left). The \mathcal{R} -graph of C is planar and can be naturally embedded in the plane simply by arbitrarily

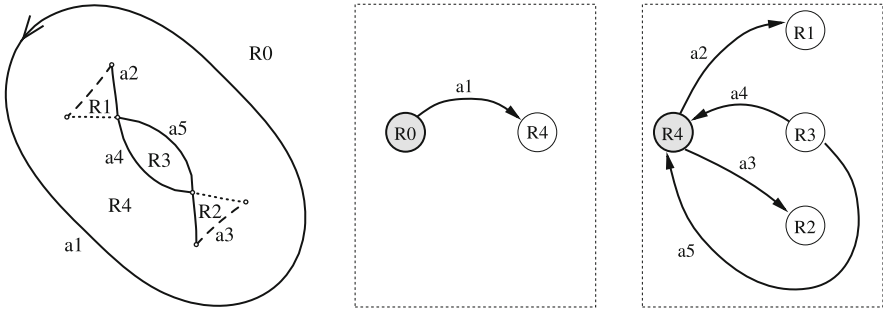


Fig. 10.17 *Left*: apparent contour of a torus. It is centrally symmetric and not equivalent to that of Fig. 10.11 due to the local structure of the two crossings. Due to the central symmetry an exchange of the two extended arcs containing the cusps produces two equivalent region descriptions. *Centre and right*: the \mathcal{R} -graphs of the two connected components of the apparent contour

choosing a point in each region and connecting the selected points with disjoint curves that transversally cross the extended arcs separating the two corresponding regions. Being C connected, it follows that its regions are all simply connected, with the exception of the *external region*⁶ that has a single *hole* and no external boundary. Traversing the boundary of a region counterclockwise and annotating the visited extended arcs allows to define an additional circular list of incident arcs at each node of the \mathcal{R} -graph. We can always identify a special node of an \mathcal{R} -graph: the one corresponding to the external region of C .

For a generic (possibly nonconnected) apparent contour we have one such \mathcal{R} -graph associated with each connected component. These graphs are naturally organized hierarchically by observing that the regions of a connected component form a partitioning of one of the regions (not the external one) of another connected component of the apparent contour. We say that the latter connected component *directly contains* the former. The most external connected components are an exception. In this way we attach \mathcal{R} -graphs to nodes of other \mathcal{R} -graphs. An example of an apparent contour (with two connected components) and the two corresponding \mathcal{R} -graphs is shown in Fig. 10.17, the dark nodes correspond to the respective external regions. Observe that region R4 appears in both graphs, but with different meanings: it is not the external region of the graph describing the outer component of the contour, whereas it is the external region of the graph describing the inner component of the contour. The whole \mathcal{R} -graph having R4 as external region should be regarded as an attribute of the node R4 of the other graph. In general each region of an \mathcal{R} -graph has its own *set of holes* as an attribute: it is a (possibly empty) set of “holes”, where each hole corresponds to a separate connected component of the

⁶Not to be confused with the external region of the whole apparent contour, it is the unbounded connected component of $\mathbb{R}^2 \setminus C$. Occasionally we shall refer to the external region of C as the subset of the unbounded connected component of $\mathbb{R}^2 \setminus C$ that coincides with the region of the apparent contour adjacent to C from the outside.

apparent contour. Simply connected regions have an empty set of holes, regions with a single hole, like the one in Fig. 10.17, have a set of holes with just one element. Since the function f jumps of two units whenever we cross an arc, we can give a natural orientation to the arcs of the \mathcal{R} -graph in the direction of increasing f .

10.6.3 The Depth-First Search of an \mathcal{R} -Graph

Depth-first search (DFS in short) is a well-known algorithm widely used to traverse trees or graphs. Basically it consists in following links as far as possible before backtracking and (in the case of graphs) it will not follow links leading to an already visited node. In the end the arcs that have been actually traversed define a *spanning tree* of the original graph: a connected subgraph that contains all nodes, but with no cycles. The result of the DFS depends on the starting node and on a total ordering of the links incident to each node.

We describe now a natural way to apply a DFS to an \mathcal{R} -graph.⁷ For each connected component of the apparent contour it depends upon the selection of a *starting* node and of one of its incident links (recall that we have an \mathcal{R} -graph associated with each connected component of the apparent contour). The external region is the obvious candidate for the starting node, whereas in general we do not have a natural way to select an incident link, that is one of the extended arcs that compose the (inner) boundary of the external region.

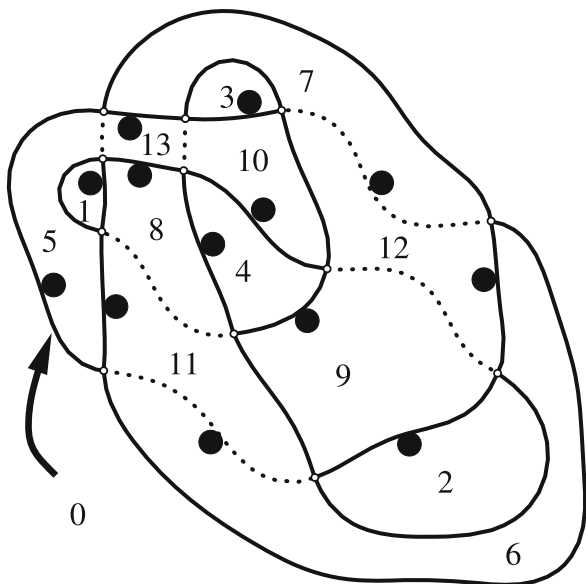
During the DFS, as soon as we visit a region (a node of the \mathcal{R} -graph) we define its *entry-point* as the extended arc (a link of the \mathcal{R} -graph) that has been crossed in order to arrive to it. This, together with the cyclic ordering of the links at the node induced by the planar embedding, allows us to define a total ordering of the incident links that in turn can be used to continue in the DFS algorithm.

To summarize, for each \mathcal{R} -graph (i.e., each connected component of the apparent contour) and for each choice of an extended arc bordering its external region, we have a procedure that allows us to define an *entry-point* for each internal region in a natural way (a way that depends on the structure of the regions graph, but not on its representation). Recalling the representation of (a connected component of) the boundary of a region given in Sect. 10.2.2.1, this amounts in deciding the starting arc of each cycle.

We illustrate the DFS procedure with an example (see Fig. 10.18). We choose the apparent contour of a tube knotted in the shape of the trefoil knot, it is connected and is composed of 14 regions and 24 arcs. We apply the DFS procedure selecting as starting arc the one pointed by the arrow in Fig. 10.18. Starting from region

⁷This procedure was recently suggested to us by Giovanni Paolini, Scuola Normale Superiore, Pisa, to whom we are indebted. It was first implemented in version 2.0.0 of the program. Previous versions of `appcontour` suffer from an imperfect canonization procedure that could lead to different regions descriptions starting from diffeomorphically equivalent apparent contours.

Fig. 10.18 The DFS procedure is applied to the apparent contour of a knotted tube by selecting the arc pointed by the arrow as the starting link from region R0. Regions are numbered from 0 (the external region) to 13; the *black circles* indicate the resulting entry points. The tube is knotted as in a trefoil knot, the simplest possible nontrivial knot



R0 the first region entered is thus R5 and consequently its entry point is the one indicated by the black circle in the Figure. From R5 the procedure first tries to enter R0 (already visited), then enters R11, crossing the next arc encountered walking counterclockwise along the boundary from the entry-point. Of course following the entry-point always leads to a just visited region, so we shall not mention such an attempt again. Then from R11 down to R6, then (after an attempt to revisit R0) down to R12, then R7, then R13, then (since R5 has already been visited) down to R8, then R1. At this point both arcs of R1 lead to visited regions, hence DFS backtracks to R8 and goes down into R4. The DFS procedures continue in this way until it is forced to backtrack to R0 with no external arcs leading to unvisited regions still to follow. The complete list of visited region in the visiting order is: 0 – 5 – 11 – 6 – 12 – 7 – 13 – 8 – 1 – 4 – 9 – 2 – 10 – 3.

Observe that the DFS also defines in a natural way a spanning tree for each \mathcal{R} -graph, in the case of Fig. 10.18 we can easily obtain the spanning tree by removing all links that correspond to extended arcs that are not entry-points for either of the two adjacent regions. The spanning tree is always rooted at the external region.

10.6.4 The Canonization Procedure

Canonization of the region description amounts in making a specific (optimal) choice in items (1)–(5) of Sect. 10.2.3, among all equivalent representations.

After a preliminary step (Sect. 10.6.4.1) for the canonification of the extended arcs description, item (5) in Sect. 10.2.3, the canonification process takes advantage

of the DFS procedure together with a lexicographic comparison procedure between different representations (Sect. 10.6.4.3). We shall refer to Fig. 10.17 as a concrete example. The overall theoretical complexity of the resulting canonization algorithm is $\mathcal{O}(n^2)$, n being the number of extended arcs. It is arguably not optimal, however it seems sufficient for the uses that we can think for the software.

10.6.4.1 Canonical Description of an Extended Arc

The streaming description of an extended arc as described in Sect. 10.2.1.1 is unique with just one exception. Namely, for a closed extended arc containing more than one cusp we have to select a starting point along the arc when listing the sequence of d values. We have thus a maximum of s (where s is the number of cusps) different ways to describe the same extended closed arc. In such a situation we simply define the *canonical* description to be the one that comes first in the lexicographical ordering of the sequences of d values. There is no possible ambiguity in this procedure. As an example, the strings $(1\ 2\ 1\ 2\ 1)$ and $(2\ 1\ 2\ 1\ 2)$ both describe the same closed extended arc with four cusps of Fig. 10.12, right (recall that conventionally for closed extended arcs the first and last values correspond to the same simple arc, the one containing the chosen starting point); only the first string gives however the (unique) canonical description.

10.6.4.2 Sorting Non-Equivalent Arcs

We can take advantage of the fact that extended arcs carry intrinsic information (cusps and labelling) in order to reduce the algorithm overhead by introducing an ordering among them based on such information. However we cannot assume that all extended arcs are distinguishable based on such information. We end up with a partition in equivalence classes of indistinguishable extended arcs, with a total ordering among the equivalence classes. This can help reduce the computational effort, yet it will not decrease the theoretical $\mathcal{O}(n^2)$ computational complexity.

For example, the apparent contour of Fig. 10.17 has one extended arc with description “ (0) ” (the external closed arc), two indistinguishable extended arcs with description “ $[0\ 1\ 2)$ ” (the two swallow’s tails) and two indistinguishable extended arcs with description “ $[0]$ ” (those bordering the inner hole of the torus).

When selecting an extended arc along the boundary of some region we can also take advantage of the orientation of the extended arc with respect to the region, allowing further discrimination.

10.6.4.3 Using DFS and Lexicographic Comparison

In view of the discussion in Sect. 10.6.3 we need a way to select (in a natural way) an incident link (bordering extended arc) to the external region of each of the

\mathcal{R} -graphs associated with each connected component of the apparent contour. For this we introduce the notion of *normalized description* of an \mathcal{R} -graph (or equivalently *normalized description* of a connected component of the apparent contour).

Assume for the moment that the apparent contour is connected, i.e., we have only one \mathcal{R} -graph. Then for any two distinct choices of a link incident to the external region we can make use of the DFS strategy described in Sect. 10.6.3 to define a complete set of *entry-points* for all internal regions and consequently two different ways to traverse the graph, leading to two different descriptions. We can then lexicographically compare the two descriptions by simultaneously traversing the graph in both ways and returning with a “winner” as soon as we find a difference. We stress that in this comparison we **must** make use of all intrinsic information attached to arcs and nodes (orientations and labellings). The “winning” selection will be subsequently compared one after the other to the descriptions (set of entry-points) obtained by varying the starting link incident to the external region. Of course we might restrict the search to one of the equivalence classes of indistinguishable extended arcs discussed in the previous Sect. 10.6.4.2, the equivalence class being selected according to some well-defined strategy, in order to reduce the cost of the procedure.

The overall winner will be a natural choice of the starting link and defines a distinguished extended arc among those bordering the external region.

We clarify the procedure by applying it to the internal connected component of the apparent contour of Fig. 10.17. The external region is named R_4 with four incident links, corresponding to arcs a_2 , a_5 , a_3 , a_4 , listed according to their circular ordering.

- Selecting e.g., the arc a_2 as the starting link, the DFS procedure enters region R_1 , and defines arc a_2 as its *entry-point* (no surprise here, since this is the only incident arc at region R_1). Following arc a_2 from region R_1 leads to the already visited region R_4 , so that the procedure backtracks to region R_4 and proceeds with the next arc following the circular ordering, namely arc a_5 . This leads to region R_3 , the procedure defines a_5 as its entry-point and tries to follow the two incident arcs at R_3 , both leading to the already visited region R_4 . Backtracking again it follows arc a_3 , defines the entry-point for region R_2 , backtracks and finally tries to follow the last arc a_4 , that however leads to the already visited region R_3 , terminating the DFS.
- The whole procedure must be repeated after changing the starting link from a_2 to a_5 , the next one in the circular list, with a new definition of entry-points for all internal regions. The two different ways to traverse the \mathcal{R} -graph, the one defined by starting with arc a_2 and the one defined by starting with arc a_5 are now compared lexicographically, they differ right at the beginning, since arcs a_2 (“[0 1 2]”) and a_5 (“[0]”) are not indistinguishable. Some comparison technique will then declare e.g., a_2 to be “better” than a_5 (we shall not enter here into the details of how extended arcs are ordered) so that the winner so far is arc a_2 .

- At this point the procedure starts from a_3 , obtains a set of new entry-points and compares the result with what obtained starting from a_2 . Not surprisingly (in view of the central symmetry of the apparent contour) the comparison will show no difference in the two descriptions, anyone between a_2 and a_3 will be declared best, say a_3 .
- The final application of the DFS strategy will start from arc a_4 , leading to a description that will compare unfavourably with respect to that starting from arc a_3 .

The final “winner” will then be arc a_3 (together with all corresponding entry-points), and this will be the *normalized description* of the inner connected component of the apparent contour of Fig. 10.17.

If there are more connected components, the procedure explained above can still be applied to the “innermost” connected components, the ones that have all internal regions with no holes (they are all simply connected), and they can all be normalized. If we now consider a connected component of the apparent contour having one or more regions with holes, but such that all contained connected components have been already normalized, we can proceed with its normalization as follows, the key-point being that already normalized apparent contours can be successfully compared using the lexicographic comparison technique already explained above.

- For each internal region possessing more than one hole we need to reorder the holes by comparing their normalized description, we end up with a “normalized” list of “normalized” components of the apparent contour;
- When comparing two sets of entry-points defined by two different starting arcs, we also must compare regions by taking into account their list of “normalized” holes. This can be done safely since everything that we compare has already been previously normalized.

By iterating the above procedure we can normalize one after the other all connected components of the apparent contour, the most exterior ones will be normalized last.

Please note that the net result of this normalization procedure allows to make a natural choice at points (3) and (4) of Sect. 10.2.3.

As an example consider the apparent contours on the left and on the right of Fig. 10.19. They are composed of many (8) very simple connected components, all made of a single closed arc. From an algorithmic point of view the two apparent contours are difficult to distinguish since each element (region or arc) of one apparent contour can be found in the other apparent contour with the same local structure: e.g., there are four empty closed arcs in both apparent contour, two closed arcs containing a region with two holes can be found in both, and so on. However, a correct application of the above strategy allows to distinguish among them.

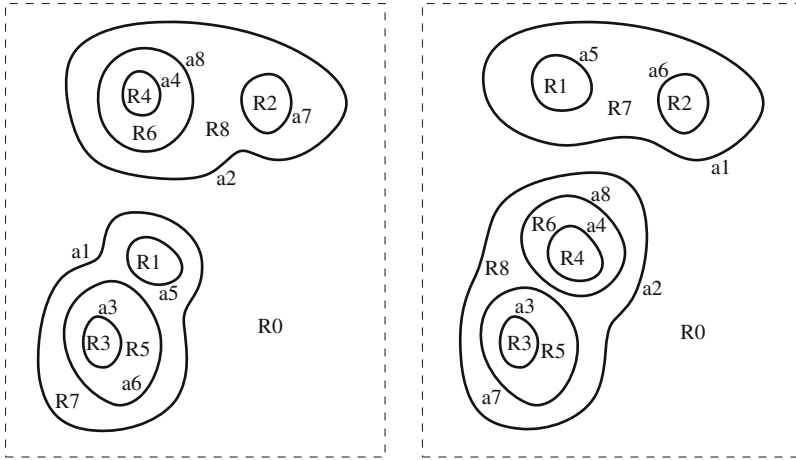


Fig. 10.19 The apparent contours on the *left* and on the *right* correspond to eight deformed spheres with different relative positions (all closed arcs are oriented counterclockwise)

10.6.4.4 Sorting Regions and Arcs

Points (1) and (2) are now simple to deal with. The overall set of selected entry-points automatically defines a transversal strategy for the overall collection of \mathcal{R} -graphs. Note that whenever we reach a region with holes we must traverse its holes, which are completely independent \mathcal{R} -graphs, in the order defined by the normalization procedure, before continuing the traversal of the original connected component. This traversal strategy naturally defines an ordering for the regions, with the external region of the apparent contour listed first.

Similarly, arcs can themselves be reordered according, e.g., to which one is used first during the traversal of the apparent contour. It is convenient however, for better presentation, to use this latter ordering strategy only to discriminate (and reorder) between indistinguishable arcs.

We have now resolved all ambiguities listed at points (1)–(5) of Sect. 10.2.3; the resulting region description can be used as a “canonical” region description. In order to decide whether two region descriptions describe diffeomorphically equivalent apparent contours it would be enough to compare (character by character) the corresponding canonical region descriptions.

10.6.4.5 Postprocessing

After we have a way to obtain a *canonical* region description we are however completely free to apply to it any well-defined reordering procedure (among those listed at points (1)–(5) of Sect. 10.2.3) that can depend both on intrinsic properties of the apparent contour and on the canonical representation, and still obtain a good

candidate for a canonical description. This is actually done by the `appcontour` program for compatibility reasons and it is harmless.

Canonization of the region description is implicitly computed by `appcontour` in most cases both before and after the application of rules, surgeries or other modifications. For this reason the outcome of “`contour print <example>`” can differ from the input region description (but will be an equivalent region description). In many cases automatic canonization can be inhibited by using the option `--nocanonify` in the command line.

10.6.5 Comparison of Apparent Contours

The canonization procedure allows to easily define a total ordering on equivalence classes of apparent contours. It is sufficient to lexicographically compare the canonical region description and have the guarantee that two region descriptions will compare as equal if and only if they represent diffeomorphically equivalent apparent contours. The ordering will moreover satisfy the usual properties of a total ordering

Comparison criteria to decide which one of two non-diffeomorphically equivalent apparent contours is better are quite subjective, so that we shall not enter into the details of the criteria used by the program (they might also change in future versions).

The command “`contour compare`” followed by one file containing the description of two apparent contours or by two files will compare the two apparent contours and print one of the strings “`s1 = s2`”, “`s1 < s2`”, “`s1 > s2`”, with the obvious meaning.

10.7 Fundamental Groups and Cell Complexes

Commands `cellcomplex`, `insidecomplex` and `outsidecomplex` of the software carry out the cell complex construction described in Sect. 7.5.1; in particular, they print the constructed CW complex, possibly after the deformation retract that reduces it to two-dimensional at most. The syntax of the description is as follows. First we have a list of the nodes (0D cells) numbered starting from 0 followed by a count of how many arcs (1D cells) have it as an endpoint. For instance, the string

```
node 0 ref 4
```

means that there are four 1D cells that concur at node 0. Then the software prints a list of all 1D cells (arcs) numbered starting from 0 with the reference to the two end nodes. This implicitly gives an orientation to the cell which is crucial when

describing the 2D cells. For instance, the string

```
arc 2 [1, 3] ref 2
```

means that arc 2 starts at node 1 and ends at node 3. The final number 2 counts the number of 2D cells that have this arc as part of the boundary. Finally all 2D cells (faces) are listed with a description of the boundary. For instance,

```
face 1 [-5 -4 +1 ]
```

means that if we traverse the boundary of face number 1, which is always an \mathbb{S}^1 , we find arcs 5, 4, 1 in this order; arcs 5 and 4 are traversed opposite to their orientation.

10.7.1 *Computing the Euler–Poincaré Characteristic and the Number of Connected Components*

From the cell complex of each of the three sets E , $\Sigma := \partial E$ and $\mathbb{R}^3 \setminus E$ constructed in Sect. 7.5.1, we can easily compute the corresponding Euler–Poincaré characteristic by means of the Euler formula: number of 2D cells minus number of 1D cells plus number of 0D cells. This information is shown respectively with the commands `contour icharacteristic`, `contour scharacteristic`, `contour ocharacteristic`, and is actually the **total** Euler–Poincaré characteristic, that is the sum of the characteristics of each connected component.

The formula constructed in Theorem 7.4.1 provides a simpler way to compute the Euler–Poincaré characteristic of Σ , and is used by the software code with the command `contour characteristic`. It gives the same result as `contour scharacteristic` and works more generally also for non-labelled apparent contours. The same type of computation is used for the Euler–Poincaré characteristic displayed with the command `contour info`.

The number of connected components of each of the three sets E , $\Sigma = \partial E$ and $\mathbb{R}^3 \setminus E$ is just the number of connected components of the respective cell complex and is displayed at the end of the corresponding cell complex description. The computation of the number of connected components of Σ displayed with `contour countcc` or as part of `contour info` is however computed directly on the labelled apparent contour without resorting to the construction of the cell complex. It should be noted, however, that we do not have a simple formula (as it happens for the Euler–Poincaré characteristic, see formula (7.11)) and the actual computation is carried on by simulating the construction of the topological manifold \mathcal{T} of Sect. 5.2. In particular there is no way to obtain the number of connected components of the source manifold in the abstract case of non-labelled apparent contours.

It is possible to obtain separately the characteristic of each connected component by combining the `contour info` command with `contour extractcc <n>` with n ranging from 1 to the number of connected components.

10.7.2 Fundamental Groups

Following the procedure described in Sect. 7.5.2, the software can easily compute, separately for each connected component, a presentation of the first fundamental group of each of the three sets E , $\Sigma := \partial E$ and $\mathbb{R}^3 \setminus E$, starting from the computed cell complex.

If this is done without any simplification, the result is however of no practical use, due to the excessive complexity of the resulting presentation.

There are two levels of simplifying procedures. The first acts on the cell complex itself, by removing pairs of 2D/1D or 1D/0D cells whenever the result is a deformation retract (the procedure is similar to the one described for the removal of the 3D cells). Another simplifying procedure glues together two 2D cells when they share a common arc that does not bound any other 2D cell; a similar glueing procedure can be applied to 1D cells.

The second level of simplification acts on the presentation of the fundamental group by applying basic simplifying Tietze transformations ([11, p. 48]).

At least for sufficiently simple apparent contours, the resulting presentation is simple enough to allow recognition of the corresponding group.

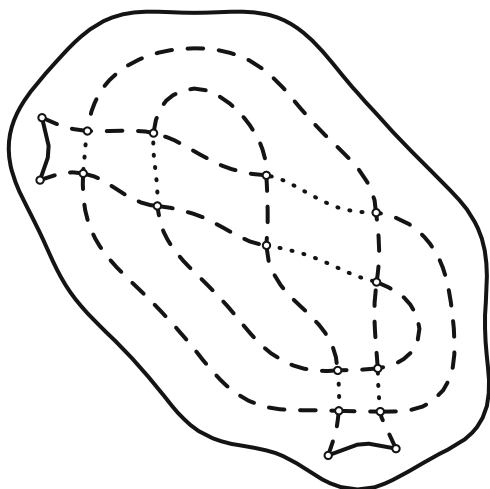
We now show how the software displays the presentation of the first fundamental group by means of an example. The first fundamental group of (the inside of) a solid shape E obtained by making a knotted tunnel into a sphere (see Fig. 10.20) is described as follows:

```
$ contour fg --in internalknot
Finitely presented group with 2 generators
<a,b; abbaB>
```

Option `--in` instructs the software to consider the solid set E *inside* the surface Σ in place of the surface itself, i.e., the set of points that can be reached from infinity traversing transversely the surface an odd number of times; command `ifg` is a shortcut for `fg --in`. The notation used to display a finitely presented group mimics the usual mathematical way to write a presentation, as we have already seen in Sect. 7.5.2. We first have a list of generators, in this case a and b , followed, after the semicolon, by a list of relators, i.e., words that must be identified with the identity of the group. In our example there is a single relator; capital letters must be understood as the inverse of the corresponding generator (i.e., B stands for b^{-1}). In the end the corresponding presentation is

$$\langle a, b; ab^2ab^{-1} \rangle.$$

Fig. 10.20 Apparent contour of a spherical shape with a knotted tunnel; *dashed lines* have labelling 1, *dotted lines* have labelling 3



Unfortunately, a presentation of a group is often far from being satisfactory, due to the inherent difficulty in deciding whether two different presentations describe the same group. In the example at hand, we can introduce the two elements $x = ba^{-1}$ and $y = b$; in this way, the original generators can be recovered from x and y as $a = x^{-1}y$, $b = y$ so that x, y form themselves a set of generators of the group. The relator now becomes $x^{-2}y^3$, that can also be written as $x^2 = y^3$. The group with presentation $\langle x, y; x^2 = y^3 \rangle$ is known to be non-free and to coincide with the knot group of the trefoil knot [8].

As a comparison, we can actually compute the knot group of the trefoil knot (see Fig. 10.18) with the command

```
$ contour fg --out trefoilknot
Finitely presented group with 2 generators
<a,b; abAbaB>
```

Here the option `--out` is used to require the computation of the fundamental group of $\mathbb{R}^3 \setminus E$, the *outside* of the surface Σ , consisting of all points that can be reached from infinity after crossing (transversely) the surface an even number of times. The combination `fg --out` can be abbreviated by using the command `ofg`. We can rewrite the presentation $\langle a, b; abAbaB \rangle$ in terms of the new elements $x = ba^{-1}b$ and $y = a^{-1}b$; they are generators since we can express $a = xy^{-2}$ and $b = xy^{-1}$. Finally, after rewriting the relator in terms of x and y , we end up again with the presentation $\langle x, y; x^2 = y^3 \rangle$.

Remark 10.7.1 The list of relators can be empty, in which case we have the free group generated by the given generators. If both the generators list and the relators list are empty, we obtain the trivial group consisting only of the identity element. A trivial first fundamental group indicates that the set is simply connected, which

unfortunately does not always imply that the set is contractible; for instance, \mathbb{S}^2 has a trivial first fundamental group, but nontrivial higher order homotopy groups.

As another example, consider the result of “`contour fg torus2`”, the first fundamental group of a torus; it is displayed as

`<a,b; abAB>`

Now, $aba^{-1}b^{-1}$ is the commutator of the two generators a and b , so that the resulting group is isomorphic to $\mathbb{Z} \times \mathbb{Z}$, as expected.

Remark 10.7.2 In the cases when the set E (or Σ , or $\mathbb{R}^3 \setminus E$) is not connected, the software will compute the fundamental group of each connected component and print a presentation of each.

Recognition of knotted/unknotted surfaces is thus related to the ability to decide whether two different presentations describe the same group. This is the well-known *group isomorphism problem*, which has been proven to be undecidable in general [11].⁸

The three fundamental groups of Σ , E and $\mathbb{R}^3 \setminus E$ are invariant under \mathbb{R}^3 -ambient isotopies, so that they can be used to prove, e.g., that there is no sequence of Reidemeister-type moves that can transform one apparent contour (with labelling) to another.

10.7.3 *Invariants of Finitely Presented Groups and the Alexander Polynomial*

The ability to compute the invariants discussed in Sect. 7.6 is essentially the only way we have to ensure that two first fundamental groups are not isomorphic, which entails that the originating 3D scenes are not \mathbb{R}^3 -ambient isotopic.

A few of the discussed invariants are actually implemented in the software code.

Computing the abelianized of the fundamental group requires an implementation of the procedure described in [11, Sect. 3.3]; this is done by implementing an iterative procedure that directly mimics the algorithm in [11]; the resulting preabelian presentation can be displayed by using the “`fg`” command in combination with the “`--preabelian`” option. As an example, for the knot group (fundamental group of the outside) of the trefoil knot we have

```
$ contour fg --out --preabelian trefoilknot
Finitely presented group with 2 generators
<a,b; abAbaBB>
```

⁸As already noted in a footnote of Sect. 7.5.2, the isomorphism problem is decidable in the special case of the fundamental groups of 3-manifolds.

If we sum up all exponents of the generator a in the relator (i.e., we count the number of occurrences of a and subtract the number of occurrences of A) we obtain the integer 1, whereas if we do the same with the generator b we obtain zero.

From the preabelian presentation the program can compute the abelianized and display its description by printing its rank and the invariant factors, if present (the torsion part). This is done with the command `abelianizedfundamental`, abbreviated as `afg` as in the following example.

```
$ contour afg --out trefoilknot
Torsion-free abelian group of rank 1
```

It should be noted that no torsion will be present for fundamental groups obtained in the context of apparent contours, the only way to obtain groups with a torsion part is to directly enter a group presentation, see Sect. 10.10.7. For example, we can ask for the abelianized of the finitely presented group $\langle a, b; a^2, b^4 \rangle$ by directly entering the group presentation as follows:

```
$ contour afg
fpgroup{<a,b;aa,bbbb>}
Abelian group of rank 0 with torsion; invariant factors: 2 4
```

showing that the group is isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_4$, the invariant factors are the nontrivial diagonal entries in the Smith normal form of the exponent-sum matrix.

The Alexander polynomial can be computed (for a knotted solid torus) with the command `alexander`, for the trefoil knot above we have

```
$ contour alexander --out trefoilknot
Computing Alexander ideal for d = 1:
Alexander polynomial (up to t -> 1/t):
+1-t+t^2;
```

The actual computation is performed following the procedure exposed in Sect. 7.6, we observe a few facts

- The $d = 1$ assertion in the software output emphasizes the fact that the Alexander polynomial computed is the generator of the elementary ideal ε_1 (Definition 7.7.2), we recall that the so-called *order ideal* ε_0 is trivially $\varepsilon_0 = \{0\}$, so that ε_1 is the first nontrivial (and generally the most informative) elementary ideal. It is always a principal ideal, generated by the Alexander polynomial. It is possible to force different values for the index d of the elementary ideal ε_d by using the option `--foxd <d>`, however elementary ideals of the outside of knotted tubes with index $d > 1$ are not at present computable by the software, excluding special cases where the result is trivially the whole ring of Laurent polynomials.
- The “up to $t \rightarrow 1/t$ ” remark emphasizes the fact that the resulting ideal is defined up to the choice of the isomorphism between G/G' and \mathbb{Z} , that in

the end corresponds to a change of base of \mathbb{Z} , i.e., $t \rightarrow 1/t$, where t is a generator of \mathbb{Z} . This remark is somewhat diminished by the known property of the Alexander polynomials of knots of having symmetric coefficients. This however is not the case for a generic finitely presented group G having infinite cyclic commutator quotient. Also recall that the Alexander polynomial is defined up to multiplication by a unit of the ring of Laurent polynomials, i.e., monomials of the form $\pm t^s$.

The printed polynomial is a canonical representative. This implies that whenever we obtained different polynomials when computing the Alexander polynomial of two scenes, the corresponding fundamental groups are **not** isomorphic, and hence the two scenes are not equivalent.

It is not a surprise that the Alexander polynomial of the “inside” of the set illustrated in Fig. 10.20, computed as

```
$ contour --in alexander internalknot
Computing Alexander ideal for d = 1:
Alexander polynomial (up to t -> 1/t):
+1-t+t^2;
```

is exactly the same as that of the `trefoilknot`.

The “Conway knot” (Fig. 10.23, left) [3, pp. 180,181] is the smallest (in terms of number of diagram crossings) nontrivial knot having trivial Alexander polynomial.⁹ The knot description corresponding to the diagram in Fig. 10.23 left allows to obtain the apparent contour of a tubular neighbourhood from which we can compute the fundamental group of the outside and the Alexander polynomial:

```
$ contour fg --out conway
Finitely presented group with 4 generators
<a,b,c,d; abAdaBAcabACabADaBAC, abAdbDBdB, abAcaBACabADDBdADbdaBAC>
```

```
$ contour alexander --out conway
Computing Alexander ideal for d = 1:
Alexander polynomial (up to t -> 1/t):
+1;
```

We conclude this section with the examples displayed in Figs. 10.21 and 10.22, four different collections of rings. The first two examples display a set of two rings, unlinked (left) and linked (middle). Here is the fundamental group of their outside.

```
$ contour fg --out two_rings
Free group of rank 2
<a,b; >
```

⁹The “Kinoshita-Terasaka” knot has the same property. Indeed the Conway and the Kinoshita-Terasaka knots form a “mutant” pair, they are non-equivalent knots with the same Alexander and Jones polynomials.

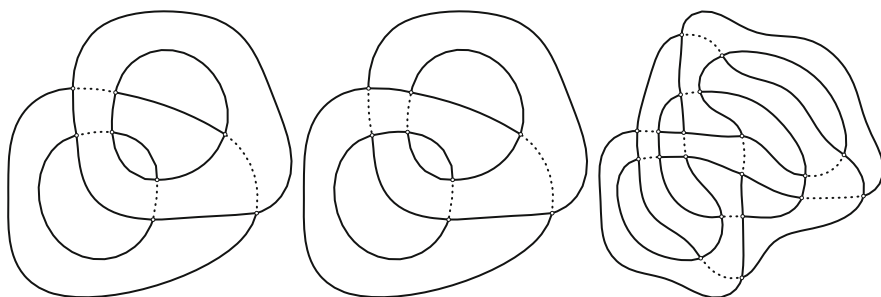


Fig. 10.21 Apparent contours of collections of rings. *Left*: two unlinked rings, *middle*: two linked rings, *right*: the “Whitehead link”, two interlaced tori that have vanishing linking number

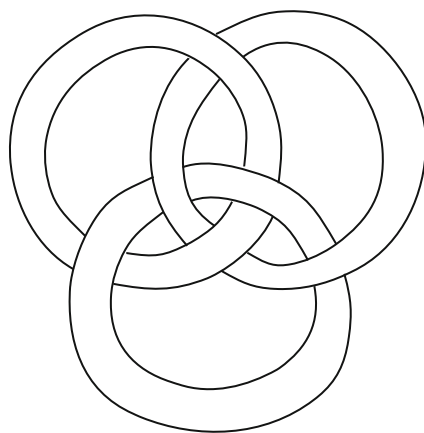


Fig. 10.22 Apparent contour of the “Borromean rings”, three interlaced tori that unlink completely when we remove any one of them

```
$ contour fg --out linked_rings
Finitely presented group with 2 generators
<a,b; abAB>
```

Since $abAB$ is the commutator between the two generators the second presentation is readily recognized as the abelian group $\mathbb{Z} \times \mathbb{Z}$, that we encountered above. These results are consistent with [8, Example 1.23, p. 46].

The third example of Fig. 10.21 gives

```
$ contour fg --out whitehead
Finitely presented group with 2 generators
<a,b; abAbAbabABaBaBAB>
```

with a quite involved group presentation. Proving that this is not a free group, or conversely, showing that it is a free group is not straightforward. In this case,

resorting to the computation of the Alexander ideal allows to conclude that the two components cannot be split, see Sect. 10.7.4.

The set of rings of Fig. 10.22 are the well-known Borromean rings. They consist of three interlaced tori with the property that they become free as soon as we remove any one of them. Here is the fundamental group of the interior.

```
$ contour fg --in borromeanrings
```

```
Connected component 0:
```

```
Free group of rank 1
```

```
<a; >
```

```
Connected component 1:
```

```
Free group of rank 1
```

```
<a; >
```

```
Connected component 2:
```

```
Free group of rank 1
```

```
<a; >
```

Since the solid interior is composed of three connected components, we also have three fundamental groups, that in this case are all equal to \mathbb{Z} , the fundamental group of the solid torus.

More interesting is the fundamental group of the outside of the surface:

```
$ contour fg --out borromeanrings
```

```
Finitely presented group with 3 generators
```

```
<a,b,c; acACbcaCAB, acAbaBCbAB>
```

The two relators can be rewritten (after rotation of the second one) using the commutator notation as $[[a, c], b]$, $[[A, b], c]$. Proving that this is not the free group of rank three (the fundamental group of three unlinked tori) is however outside the scope of this book.

The “outside” of the four examples of Figs. 10.21 and 10.22 all have a fundamental group of rank larger than one, i.e., their commutator quotient is isomorphic to \mathbb{Z}^r with $r = 2, 2, 2, 3$ respectively:

```
$ contour afg --out two_rings
```

```
Torsion-free abelian group of rank 2
```

```
$ contour afg --out linked_rings
```

```
Torsion-free abelian group of rank 2
```

```
$ contour afg --out whitehead
```

```
Torsion-free abelian group of rank 2
```

```
$ contour afg --out borromeanrings
```

```
Torsion-free abelian group of rank 3
```


Thus, they have Alexander polynomials (or, better, elementary ideals) in two or three indeterminates, see Definition 7.7.2, Sect. 7.7. The software cannot deal at present with more than two indeterminates; we shall discuss the case of two indeterminates in the next section.

10.7.4 Alexander Polynomials and Alexander Ideals in Two Indeterminates

In the context of commutator quotient groups of rank 2 we deal with Laurent polynomials in two indeterminates (which we shall indicate with the letters u and v) with integer coefficients. As already noted the ring of such polynomial is not a “principal ideals” ring, so it is more appropriate to speak of Alexander ideals instead of Alexander polynomials. There are two special cases:

- G has deficiency one. This is the case of the outside of a two-components links, and we are thus interested in the elementary ideal ε_1 . As noted in Sect. 7.8 the elementary ideal ε_1 is then given by the product $\varepsilon_1 = (\Delta)\mathcal{E}$, where (Δ) is the principal domain generated by Δ . This is not a principal ideal, however its special structure justifies the fact that the polynomial Δ in its definition is called *Alexander polynomial* (of the two-components link).
- G has deficiency two. This is the case of the outside (or the inside) of a surface of genus 2, and we are interested in the elementary ideal ε_2 . Recalling the results of Sect. 7.9, we now have a more involved situation where the Alexander ideal ε_2 is, in general, generated by a polynomial w satisfying $w(1, 1) = 1$ and the ideal $\mathcal{W} = (w_1, \dots, w_{n-2})\mathcal{E}$, product of the fundamental ideal and the ideal generated by the polynomials w_1, \dots, w_{n-2} , n being the number of generators in the group presentation. For sufficiently simple scenes we have $\mathcal{W} = \{0\}$, so that we still have a principal ideal and we can still speak of the *Alexander polynomial*.

Let us consider again the apparent contour in the left of Fig. 10.21, two unknicked rings. We can compute the corresponding Alexander ideal with

```
$ contour alexander --out two_rings --foxd 1
Alexander polynomial (special large deficiency case):
0;
```

This example requires a brief discussion. First of all in this special case we need to explicitly indicate the value of the index of the requested elementary ideal (“--foxd 1”). This is because we are in a situation where the deficiency of the fundamental group presentation is larger than expected, 2 instead of the usual value of 1 for the complement of links with two components. For this reason the software assumes that the “interesting” value for the index d is $d = 2$ instead of $d = 1$, as can be desumed by

```
$ contour alexander --out two_rings
Computing Alexander ideal for d = 2:
Trivial whole ring ideal:
1;
```

It is then not a surprise that the Alexander ideal ε_1 turns out to be trivial. It still can be written as $(\Delta)\mathcal{E}$ with the choice $\Delta(t) = 0$.

The second example, displayed in the middle of Fig. 10.21, produces

```
$ contour alexander --out linked_rings
Computing Alexander ideal for d = 1:
Alexander ideal generated by:
(+1) (u - 1);
(+1) (v - 1);
```

where the output is clearly to be interpreted as a set of generating polynomials, that are however written as the product of the Alexander polynomial Δ (in this case $\Delta(t) = 1$) and the two generators $u - 1$ and $v - 1$ of the fundamental ideal \mathcal{E} . Using option `-Q` (see Sect. 10.10.7) we obtain a description of the ideal with a different syntax:¹⁰

```
$ contour alexander --out linked_rings -Q
#
# --foxd 1
#
ideal(u,v) F: +1;
```

where the line starting with `F:` indicates that the ideal generated by the subsequent polynomial must be multiplied by the fundamental ideal \mathcal{E} .

Here is the Alexander ideal for the Whitehead link (Fig. 10.21 right):

```
$ contour alexander --out whitehead
Computing Alexander ideal for d = 1:
Alexander ideal generated by:
(+1-u-v+uv) (u - 1);
(+1-u-v+uv) (v - 1);
```

showing that the elementary ideal ε_1 is nontrivial, thus proving that the two links cannot be split.¹¹

¹⁰As we shall see in Sect. 10.10.7 this particular syntax can be used to directly feed a Laurent ideal to the software.

¹¹To conclude that the link cannot be split we also need to know that separately each component of the links is the unknot. That is clear from the picture, but could also be seen by concatenating the command “`contour --extractcc 1 whitehead`” or “`contour --extractcc 2 whitehead`” with “`contour fg --out`”.

For all three examples we can compute the linking number of the two components, the software makes the computation using the result of Corollary 7.8.3. We have

```
$ contour linkingnumber --out two_rings
Linking number is 0
```

```
$ contour linkingnumber --out linked_rings
Linking number is 1
```

```
$ contour linkingnumber --out whitehead
Linking number is 0
```

Of course, having linking number 0 by no means implies that the two components are unlinked (i.e., they can be completely separated from each other), the Whitehead link being a counterexample.

Using the Alexander ideal as an invariant of a finitely presented group raises an issue of primary importance. On the one hand, we cannot compare two ideals simply by comparing the two sets of generators: an ideal can be generated with a variety of choices for the generators. On the other hand, in view of Remark 7.7.4 the ability to decide whether two ideals are the same is not enough. What we actually need is the ability to decide whether they are *equivalent* in the sense of Definition 7.7.7.

For two important classes of ideals, namely

- (1) principal ideals, generated by the Alexander polynomial $\Delta = \Delta(u, v)$;
- (2) Alexander ideals of two-components links, of the form $\varepsilon_1 = (\Delta)\mathcal{E}$,

equivalence of the ideals reduces to base-equivalence (in the sense of Definition 7.7.8) between the defining Alexander polynomials. This is immediate for the class of principal ideals, whereas it follows from Corollary 7.7.12 for the ideals of the form $(\Delta)\mathcal{E}$.

10.7.4.1 Canonization of a Laurent Polynomial

As usual, a canonization procedure is a standard way to select a representative of an equivalence class starting from any one element of the class. Having such a procedure allows to immediately solve the problem of deciding whether two elements are equivalent: just compare the result of the canonization procedure.

In the present context the canonization process should select an “optimal” Laurent polynomial among all those equivalent to a given one. Theoretically this is quite easy to achieve, just define a well-ordering in the set L of all Laurent polynomials with integer coefficients¹² and select the minimum of the equivalence

¹²The elements of L can be represented by strings, a well-ordering on strings can e.g., be defined by first comparing the strings length and, in case of equal length, by making a lexicographic comparison.

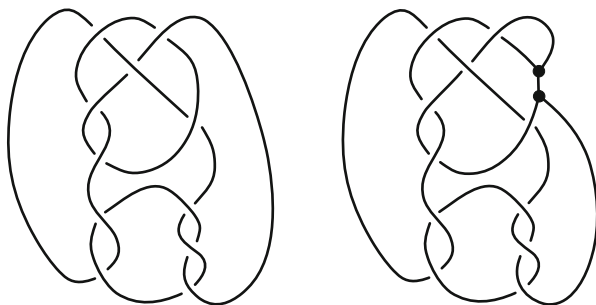


Fig. 10.23 Diagram of the Conway knot (*left*). With a suitable horizontal surgery performed on the boundary of a tubular neighbourhood of the Conway knot we obtain a surface of genus 2, boundary of a tubular neighbourhood of the diagram on the *right*

class. However we need a practical “constructive” procedure, i.e., an implementable algorithm.

The set of possible changes of base in L (Definition 7.7.5) is infinite (for $r \geq 2$), so that enumeration of all equivalent polynomials is not an option.

The software code implements an algorithm to compute a canonical representative, we postpone the description of this procedure to Appendix 10.A.

Knowing that the printed polynomial is canonical ensures that different printed polynomials indicate nonequivalent (in the sense of Definition 7.7.7) corresponding principal ideals (often the case for the fundamental group of the outside/inside of a genus two surface, but see Fig. 10.23 right for a counterexample), or indicate a nonequivalent product of the generated principal ideal by the fundamental ideal \mathcal{E} (the case of the fundamental group of the outside of two knotted tubes).

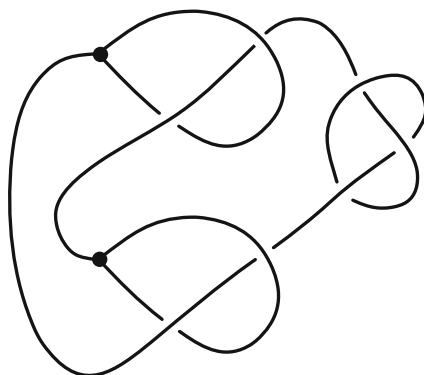
We illustrate the first situation with a few examples. The Alexander ideal of the fundamental group of the outside of Figs. 10.6 and of 10.38 are

```
$ contour alexander --out embrace
Computing Alexander ideal for d = 2:
Alexander polynomial:
+1+u-v;
```

```
$ contour alexander --out internalexternal
Computing Alexander ideal for d = 2:
Alexander polynomial:
+1-u+u^2;
```

Since the two Alexander polynomials are printed differently, we are assured that the corresponding fundamental groups are not isomorphic, so that the two scenes are not ambient isotopic.

Fig. 10.24 A diagram obtained by combining a trefoil knot and the system of curves of Fig. 10.6. As usual, this is a schematic representation of the apparent contour of the boundary of a tubular neighbourhood in 3D



Consider now the set illustrated in Fig. 10.24, obtained by combining the system of curves of Fig. 10.6 with a trefoil knot.¹³ Computation of the Alexander ideal for this example leads to

```
$ contour alexander --out embrace_trefoil
Computing Alexander ideal for d = 2:
Alexander polynomial:
+1+u-v;
```

with a result that coincides with that of the diagram without the addition of the knot. We are then in a nonconclusive situation: the Alexander ideal does not distinguish the two scenes, whereas the two fundamental groups are printed with different presentations, and we do not have sufficient information to decide whether the fundamental groups are isomorphic.

```
$ contour fg --out embrace
Finitely presented group with 3 generators
<a,b,c; abABcAC>
```

```
$ contour fg --out embrace_trefoil
Finitely presented group with 4 generators
<a,b,c,d; aBcbAdBcbDBCb, bcBcbC>
```

In view of the example of Sect. 10.13 we should not assume that the two 3D scenes (with or without the knot) are not ambient isotopic.

Concerning the Alexander ideal of scenes with two knotted tubes (tubular neighbourhood of two-component links), we already computed the Alexander ideals

¹³In the context of knot theory this would correspond to the sum of the two diagrams, however the result is different than the horizontal sum as apparent contours (of a tubular neighbourhood), indeed the horizontal sum of the apparent contour will apply surgery only on one of the two “parallel” arcs bounding each of the arcs of the diagrams involved in the sum operation of knots. Moreover one of the two “summands” is not in this case a knot, making the result of the sum not well defined.

of the three examples of Fig. 10.21. They are printed differently, so that we can conclude that the corresponding fundamental groups are not isomorphic.

10.7.4.2 Canonization of a Laurent Ideal

The more general situation of an Alexander ideal that is not necessarily principal is more involved. We can no longer resort to the canonization procedure of an Alexander polynomial and we must carefully distinguish between two different levels of equivalence between polynomials.

On the one hand, multiplication by a monomial with coefficient ± 1 corresponds to multiplication by a unit of the ring L of Laurent polynomials, and can hence be done independently on each of the polynomials that generate an ideal. On the other hand, a change of base is an action that involves the ideal as a whole, so that we must apply the same change of base (indicated by a 2×2 integral matrix B) to all generating polynomials.

Moreover, even the simpler problem of constructing a canonical set of generators for a given fixed ideal of L is not an easy task to accomplish, it requires the construction of Gröbner bases, see, for example, [16], and is not yet implemented in the *appcontour* software code.

Implementing, using, for example, a variant of the Buchberger’s algorithm, such a construction would not however give a canonical set of generators that also takes into account the notion of equivalence of ideals with respect to changes of base.

On the other hand, one would like to take advantage of the special structure possessed by the Alexander ideals of the outside/inside of a surface of genus two as generated by a principal ideal and the product of another ideal times the fundamental ideal, see Sect. 7.9.

Whenever the software code is unable to compute the canonical representative of an Alexander ideal it will make this clear by printing a warning message. Future versions of the software could implement new canonization procedures enhancing the present situation.

Recalling the computations of Sect. 7.9 it is clear that the Alexander ideal ε_2 of the outside/inside of a surface of genus two is always defined by more than one polynomial, however in many simple situations all but one of these polynomials vanish, leading in the end to a principal ideal. This fact is recognized by the software and is indeed the case for the three examples *embrace*, *internalexternal*, *embrace_internalexternal* presented in Sect. 10.7.4.1. Fig. 10.23 (right) provides an example of the general situation. This example is constructed by suitably glueing together two points of the *Conway knot*.¹⁴ The boundary of a tubular neighbourhood of the resulting set has genus two and here is the result of the computation of the fundamental group and the Alexander ideal.

¹⁴The Conway knot and its “mutant”, the *Kinoshita–Terasaka knot*, are the simplest nontrivial knots, in terms of number of crossing in their diagram, having trivial Alexander polynomial.

```
$ contour fg --out conway_pinned
Finitely presented group with 4 generators
<a,b,c,d; abADbabADBdaBAbDbdaBACabADBdADBdaBAcabADBdB,
          abADbcBdaBAbDbdaBACabADBdaBAbADbdaBAcabADBdB>
```

```
$ contour alexander --out conway_pinned
Computing Alexander ideal for d = 2:
# *** Warning: result can be noncanonical ***
Alexander ideal generated by:
+1-u-2v+2uv+u^2v+uv^2-2u^2v^2+u^3v^2;
(-1+u-u^2+u^3) (u - 1);
(-1+u-u^2+u^3) (v - 1);
(+2-2u+u^2-u^3-uv+2u^2v-u^3v+u^4v) (u - 1);
(+2-2u+u^2-u^3-uv+2u^2v-u^3v+u^4v) (v - 1);
```

The complexity of the resulting description can be greatly reduced by manually modifying the generating set. In the end we arrive at the equivalent set of two generators $-1+u+v$ and $1+u^2$, which also provides an example of a non-principal Alexander ideal.

Remark 10.7.3 It is important to emphasize here that the software cannot even compare a generic ideal generated by some polynomials w_i, \dots, w_s , $s > 1$ with the principal ideal generated by some other polynomial w , since it lacks at the moment the ability to decide whether an ideal is principal and in such a case to compute a generator.

An example of this situation can be constructed by manually modifying the fundamental group $\langle a, b, c; abABcAC \rangle$ of the example of Fig. 10.6 by means of sequence of Tietze transformations¹⁵ into the equivalent presentation $\langle a, b, c; ACBACBAbcaabccabcb \rangle$

```
$ contour alexander
fpgroupop{<a,b,c; ACBACBAbcaabccabcb>}
Computing Alexander ideal for d = 2:
# *** Warning: result can be noncanonical ***
Alexander ideal generated by:
+1-u+uv-2u^2v+u^3v+u^2v^2-u^5v^3+u^5v^4;
(-1-v-uv+u^2v+u^2v^2-u^3v^3-u^4v^3) (u - 1);
(-1-v-uv+u^2v+u^2v^2-u^3v^3-u^4v^3) (v - 1);
```

A tedious computation shows that this ideal can be equivalently generated by the single polynomial $1 - u^2v + u^3v^2$, which is base-equivalent to the polynomial $+1+u-v$ given by the software as Alexander polynomial for the example of

¹⁵Here is the list of equivalent presentations after each Tietze transformation: $\langle a, b, c; AbaBcaC \rangle$, $\langle a, b, c; BAbacabC \rangle$, $\langle a, b, c; BBAbabcabbC \rangle$, $\langle a, b, c; CBCBAbcabccabcb \rangle$, $\langle a, b, c; ACBACBAbcaabccabcb \rangle$.

Fig. 10.6. Once we have a principal ideal we can use the software to compute the canonified polynomial as follows

```
$ contour alexander
alexander(u,v) {+1-u^2v+u^3v^2;}
Alexander polynomial:
+1+u-v;
```

It turns out that the corresponding change-of-base matrix is

$$B = \begin{bmatrix} -1 & 1 \\ 2 & -3 \end{bmatrix}$$

producing the transformed polynomial $1 - u^{-1}v + u^{-1}$ that becomes $1 + u - v$ after multiplication by u .

10.8 The Mendes Graph

Recall from Sect. 2.2 that the critical set $\text{crit}(\varphi)$ is a set of m smooth closed curves in M . These curves divide M in a finite number n of connected patches M_1, \dots, M_n defining the n nodes of the Mendes graph [6]. Each of the m curves bounds two of such portions (possibly the same) and defines an arc of the Mendes graph connecting the nodes associated with those two portions.

In the context of three-dimensional shapes (Sect. 3.2) the solid set E induces a natural orientation on its boundary ∂E that plays the role of M above; this in turn induces a signature on the nodes of the Mendes graph. Arcs of the Mendes graph (components of the critical set) always connect a positive to a negative node, so that in this case we have a bipartite graph. Moreover the topology of each patch of ∂E gives a notion of *genus* [6] to be attached to the corresponding node of the graph. More precisely, the genus is $1 - e/2$ where e is the Euler–Poincaré characteristic of the patch considered as a 2D manifold with boundary. It is computed as

$$e = e_b + e_2 - e_1 + e_0,$$

where e_b is the number of connected components of the boundary of the patch, and e_i is the number of i -dimensional cells of a subdivision of the patch into a CW complex.

Our software can compute all these information and produce the Mendes graph starting from a labelled apparent contour.

10.9 Invariants

The invariants discussed in Chap. 7 can be computed by the software. We discuss here how some of these, most notably the Euler–Poincaré characteristic and the Bennequin-type invariant introduced in [13], can be computed in practice.

10.9.1 Euler–Poincaré Characteristic

This is actually a quite strong invariant, since it depends only on the source manifold M . We can however compute it using solely the apparent contour of some map $\varphi : M \rightarrow \mathbb{R}^2$ using the procedure described in Sect. 7.4. This means that we can actually infer some information on M in terms of $\text{appcon}(\varphi)$. We stress that in the general abstract setting we cannot resort to the reconstruction Theorem 5.1.1, that works only for an apparent contour with labelling, and allows essentially to “reconstruct” the source manifold M as an embedded closed surface in \mathbb{R}^3 . Also it is worth pointing out that if the source manifold M is not connected, then the result of applying the procedure of Sect. 7.4 is the sum of the Euler–Poincaré characteristics of each connected component (total characteristic), and that it is not in general possible to recover the separate characteristics. Only in some special cases this can be done: for instance, when the apparent contour has a single component, i.e., a unique closed curve with cusps that possibly intersects itself transversally, so that we can deduce that the source manifold is connected.

If $\text{appcon}(\varphi)$ is described in terms of its Morse or region description, the `appcontour` software computes the total Euler–Poincaré characteristic with a command as in the “sphere” example:

```
$ contour info sphere
This is an apparent contour with labelling

Properties of the embedded surface:
Connected comp.:    1
Total Euler ch.:    2
[...]
```

The total Euler–Poincaré characteristic is correctly given as 2. It also states that there is a single connected component, this information can be computed because we provide a labelled apparent contour.

10.9.2 Bennequin Invariant

In the context of apparent contours of maps $\varphi : M \rightarrow \mathbb{R}^2$, the Bennequin-type invariant $\mathfrak{B}(\text{appcon}(\varphi))$ studied in [13] is interesting because it is nontrivial

and, together with the number of cusps and the number of crossings, it gives a complete set of three first-order local Vassiliev-type invariants for apparent contours. It is invariant under the equivalence relation between apparent contours defined in Definition 2.4.2, whereas it changes (in a precise way) under the Reidemeister-type moves discussed in Chap. 6. We can compute this invariant also on labelled apparent contours, i.e., those originated by an orthogonal projection of an embedded surface $\Sigma = \partial E$, although we cannot take particular advantage of the additional structure. However, since an embedded surface carries a natural orientation, we have a standard notion of signature for cusps that allows to split the invariant counting the cusps into the sum of the number of positive and negative cusps, see Definition 2.2.12, and Chap. 7.

If $\text{appcon}(\varphi)$ is described in terms of its Morse or region description, the `appcontour` software computes the three (four if the contour is labelled, e.g., M is oriented) Vassiliev invariants with a command as in the example

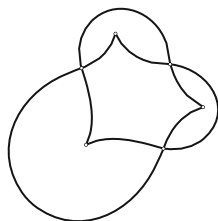
```
$ contour info sphere
This is an apparent contour with labelling
[...]
First order Vassiliev invariants:
Cusps:                0
Positive cusps:       0
Crossings:            0
Bennequin:            2.0
[...]
```

applied to the simplest possible apparent contour of a sphere. The computed value $\mathfrak{B}(\text{appcon}(\varphi)) = 2.0$ can be compared with the value of $BL(\text{appcon}(\varphi))$ as computed following the procedure of Sect. 7.2, Chap. 7. The Bennequin invariant is actually computed along the lines of the derivation of Sect. 7.1.

10.9.3 Examples of Invariants Computation

Since the Euler–Poincaré characteristic and the three Vassiliev invariants are defined on apparent contours for an M which is not necessarily orientable, the examples that we shall present hereafter are in most cases without labelling. Description of an apparent contour without labelling is possible through a Morse description that differs only slightly from that described in Sect. 10.3.2: no labelling is included and the presence of cusps along an extended arc is described by using a sequence of one or more characters ‘c’ in place of a depth value followed by a list of ‘+’ and ‘-’ (we have no longer a notion of increasing or decreasing cusp). On some figures we also include the corresponding Morse description.

Fig. 10.25 The apparent contour of the Boy surface and the corresponding Morse description. Note that multiple critical Morse events at the same level are allowed, see the beginning of Sect. 10.3.2



```
morse {
    ^l      ;
    ( ^lc ) ;
    X      X ;
    / \dc /uc \ ;
    \      X / ;
    Ur    Ur ;
}
```

10.9.3.1 Projective Plane

In this example we take the *Boy surface* (Fig. 10.25), for which Ohmoto and Aicardi in [13] computed a value of $-5/2$ for the Bennequin-type invariant. The Boy surface is a standard way to immerse the projective plane in \mathbb{R}^3 with a double curve and a triple point.

Here is the transcript of a session with the `appcontour` program, where the Morse description is included in the software package in a file with name `boysurface.morse`:

```
$ contour info boysurface
This is an apparent contour without labelling

Properties of the 2D manifold:
Total Euler ch.:    1

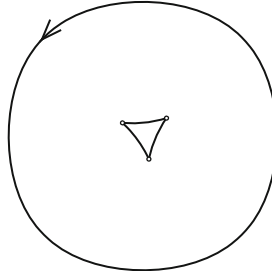
First order Vassiliev invariants:
Cusps:              3
Crossings:          3
Bennequin:          -2.5
[...]
```

We know that the source manifold is connected, since there is a single component in the apparent contour, hence the Euler–Poincaré of 1 identifies M as a projective plane.

The Bennequin invariant turns out to be evaluated correctly as $-2.5 = -5/2$. The other two first order Vassiliev invariants (3 cusps and 3 crossings) have of course an immediate interpretation in terms of the apparent contour.

As another related example, we take the apparent contour of Fig. 14a of [13], which is depicted in Fig. 10.26 with its Morse description; note that both components of the contour are oriented counterclockwise, so that the inside of the *triangle* has four preimages in M . This apparent contour is connected to that of the previous example through a number of codimension-one topological changes, the analogue of the Reidemeister moves of Chap. 6 in the context of apparent contours without labelling. Although we now have two components of the apparent contour, the source manifold is forcibly connected, since cusps pointing in the external region cannot exist.

Fig. 10.26 The apparent contour of a different map φ on the real projective plane



```
morse {
  ^1      ;
  / ^lccc \ ;
  \ U    / ;
  U      ;
}
```

Here is the output of the appcontour program:

```
$ contour info deltoid
Properties of the 2D manifold:
Total Euler ch.: 1

First order Vassiliev invariants:
Cusps: 3
Crossings: 0
Bennequin: 0.5
[...]
```

Again the resulting characteristic of 1 ensures that M is a projective plane. It is possible to construct an immersion of the projective plane in \mathbb{R}^3 such that the orthogonal projection gives this apparent contour, however there must be (at least) two pinch points (Whitney umbrellas) and a double curve connecting them. The resulting Bennequin-type invariant is in accordance with that computed in [13].

10.9.3.2 Milnor Curve and Millett Immersion

The Milnor curve [12, p. 207],[17] is the apparent contour of a particularly interesting immersion of a sphere in \mathbb{R}^3 (Fig. 10.27, left). There is a striking similarity of the Milnor curve with an example provided by Millett [12] of an immersion of the projective plane in \mathbb{R}^3 (Fig. 10.27, right). The corresponding apparent contour has only one component and a single cusp.

The result of the automated computation of the Euler–Poincaré characteristic correctly indicates that the Milnor curve must be the apparent contour of a sphere and that the Millett curve must be the apparent contour of a projective plane. The Bennequin-type invariant $\mathfrak{B}(\text{appcon}(\varphi))$ is computed for these two examples as $\mathfrak{B}(\text{appcon}(\varphi)) = 0$ for the Milnor curve and $\mathfrak{B}(\text{appcon}(\varphi)) = -\frac{5}{2}$ for the Millett example.

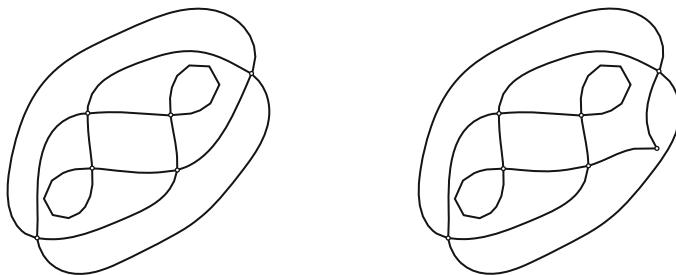


Fig. 10.27 The Milnor curve (*left*) and the Millett example for the projective plane (*right*)

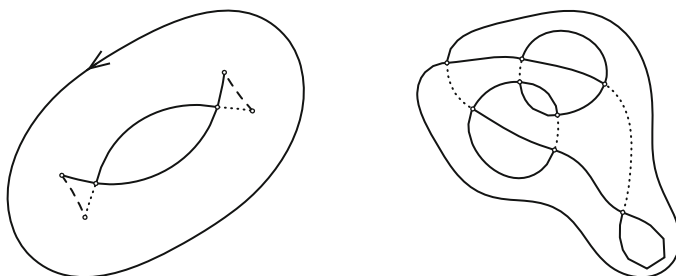


Fig. 10.28 The apparent contour of a torus (*left*) and of a knotted surface of genus 2 (*right*)

10.9.3.3 Torus

The apparent contour of Fig. 10.28 (left) corresponds to a projection of a torus.

Here is the output of the `appcontour` program:

```
$ contour info torus2
This is an apparent contour with labelling

Properties of the embedded surface:
Connected comp.:    1
Total Euler ch.:    0

First order Vassiliev invariants:
Cusps:               4
Positive cusps:      2
Crossings:           2
Bennequin:           0.0
[...]
```

Note that this example also carries information capable to reconstruct a 3D embedding of the surface (labelling); the surface is consequently oriented, and in this case there is a fourth invariant: number of positive cusps [13].

10.9.3.4 A knotted Genus-2 Surface

The apparent contour of Fig. 10.28 (right) corresponds to a surface of genus 2 (a torus with two holes), having the two holes linked together.

Here is the output of the appcontour program:

```
$ contour info genus_2_linked
This is an apparent contour with labelling

Properties of the embedded surface:
Connected comp.:    1
Total Euler ch.:    -2

First order Vassiliev invariants:
Cusps:              0
Positive cusps:     0
Crossings:          8
Bennequin:          -2.0
[...]
```

10.10 contour Reference Guide

The typical contour command is of the type:

`contour [options] <command> <description-file>`,

where the description file contains the apparent contour data in one of the forms described above. If missing, the contour description is taken from the standard input, typically as output of a previous command in a pipe chain. We shall describe here all commands understood by `contour`; they can be grouped as informational commands, commands that operate and modify the given contour, commands that convert from and to different types of contour descriptions (most notably from the region description to the Morse description), and commands related to the cell complexes of the 3D shape associated with a labelled apparent contour. Options are prefixed by a single dash character (single-letter options) as in `-q` or a double dash (long options) as in `--nocanonify`, they will be listed at the end of this section.

10.10.1 Informational Commands

These are commands that print various kinds of information on the given apparent contour.

`info`. Prints a list of properties and invariants of the given apparent contour, in particular the Euler–Poincaré characteristic, and some 2D Vassiliev-type invariants: the Bennequin-type invariant, the number of crossings, the number of cusps and

(if the apparent contour is labelled, and hence the source surface is oriented) the number of positive cusps [19], see Sect. 10.9. Finally it prints a few simple 2D properties such as the number of extended arcs, crossings, cusps and regions.

A different notion of sign of a cusp was introduced in Sect. 8.1, and the number of “positively embedded” cusps using this notion is also printed. We refer to Sect. 10.9 for more details about the computation of invariants.

`characteristic`. Prints just the Euler–Poincaré characteristic of the embedded surface associated with the apparent contour (not necessarily including a labelling). In case of a labelled apparent contour, or more generally if the manifold M is orientable, the Euler–Poincaré characteristic is always an even integral number.

`iscontour/islabelled`. Checks if the given description is an apparent contour or if it is a *labelled* apparent contour. In the latter case the program checks all conditions listed in Sect. 3.4 for the labelling function d_Σ . If the option `-q` is also given no text is printed and the program has a nonzero exit code if the apparent contour does not have a labelling or if the labelling is not consistent.

`rules`. Prints a list of all simple and composite rules that can be applied to the given apparent contour, see Sects. 10.4.1 and 10.4.3 for a complete list and a description of each.

For instance, on the apparent contour of Fig. 10.1 we obtain the following list:

```
$ contour rules torus2
Rules that apply:
K0 B B:2 S S:2 CR0L CR0R CR0LB CR0RB
```

where each string indicates an applicable rule and can be used as an argument to the command `contour rule` described in Sect. 10.10.2.

`listma`. Prints all positions where an inverse simple rule of the type *mergearcs* (the inverses of K0, K1, K1b, K2 and B) can be applied. The program loops through all pairs of arcs that bound the same region and searches the *mergearcs* rules for an applicable one, based on the orientation relative to the common region and the two values of the labelling. Here with *arc* we refer to the portions of extended arcs between consecutive cusps, or the entire extended arc if there are no cusps. Note that the pairing of an arc with itself is also legitimate. Even for quite simple apparent contours there is typically a large number of possible applications of inverse rules of the *mergearcs* type; for instance, the command `contour listma torus2` lists a total of 47 different positions for the contour of Fig. 10.1, one of which involving the external region R0:

```
$ contour listma torus2
Region 0:
-r 0 -a 1:0 -a 1:0 (INVK2:1)
Region 1:
[...]
```

The output contains one line for each possible application with the following information (for definiteness we refer to the only position relative to region R_0 in the example): “-r 0” indicates the region number 0, “-a 1:0” is the first involved arc, it is the arc number 1 (the external arc), the 0 after the colon indicates what portion of extended arc between cusps we are referring to, starting from 0; it is always 0 for arcs without cusps as in this case; the second occurrence of “-a 1:0” similarly identifies the second arc involved in the inverse rule, which in this case coincides with the first arc. This set of informations is presented in the form of options to be passed to `contour mergearcs` in order to apply the described instance of the inverse rule, for instance “`contour mergearcs -r 0 -a 1:0 -a 1:0 torus2`” actually applies the indicated inverse rule, see also command `mergearcs` in Sect. 10.10.2. There is an alternative way to apply each instance of the inverse rule which uses the same syntax as for simple and composite rules by using the string indicated in parentheses: “`contour rule INVK2:1 torus2`” meaning that we require application of the inverse of the simple rule K2 in the first position found by the software (the 1 after the colon) which is guaranteed to be the same as that indicated with the “`mergearcs`” syntax.

In order to control the exact result it might be necessary to use the “-ti i ” option or make the rule indication more specific such as in “`INVK2:1: i` ”, see Sect. 10.4.2.

`listinvl`. Prints all positions where the inverse of the simple rule L can be applied. This is possible in all regions R having a nonzero number $f(R)$ of preimages, for each of such region the inverse rule can be applied to anyone of the $f(R)$ strata. For instance, the command `contour listinvl torus2` lists a total of ten different positions for the contour of Fig. 10.1, here is an excerpt of the result:

```
$ contour listinvl torus2
[...]
Region 4:
-r 4 --stratum 0 (INVL:9)
-r 4 --stratum 1 (INVL:10)
```

relative to the annular region R_4 . Similarly to the command `listma` there are two different ways to apply one of the listed inverse rules (we show here the command that applies the last of the listed rules):

- `contour rule INVL:10 torus2`
- `contour wrinkle -r 4 --stratum 1 torus2`

see also command `wrinkle` in Sect. 10.10.2.

`listinvs`. Prints all positions where the inverse of the simple rule S can be applied. This is possible on all arcs a and there are two variants related by front-back inversion. Here arcs are intended as that portion of an extended arc between two consecutive cusps. For instance, the command `contour listinvs torus2`

lists 9 arcs for a total of 18 different applications of the inverse rule on the contour of Fig. 10.1, here is an excerpt of the result:

```
$ contour listinvs torus2
[...]
-a 5:0 (INVS:9 INVCB:9)
```

relative to the arc a5 (the number 0 after the colon indicates which portion of the extended arc 5 is to be selected, in this case there are no cusps, so that 0 is the only possible value. Similarly to the command `listma` there are two different ways to apply one of the listed inverse rules, here is an example:

- `contour rule INVS:9 torus2`
- `contour swallowtail -a 5:0 torus2`

see also command `swallowtail` in Sect. 10.10.2.

`listinvc`. Prints all positions where the inverse of the simple rule C can be applied. This is possible for suitable pairs of a cusp and an arc. Arcs are intended as that portion of an extended arc between two consecutive cusps. For instance, the command `contour listinvc torus2` lists 24 possibilities, one of which reads as:

```
-a 2:0 -a 3:1 (INVC:5)
```

where the first `-a` option indicates the cusp (2:0 and 2:1 identify the two cusps of the extended arc a2) and the second `-a` option indicates the arc involved in the inverse rule. Similarly to the command `listma` there are two different ways to apply the inverse rule, here is an example related to the information line shown above:

- `contour rule INVC:5 torus2`
- `contour puncture -a 2:0 -a 3:1 torus2`

see also command `puncture` in Sect. 10.10.2, a picture of the result is shown in Fig. 10.29.

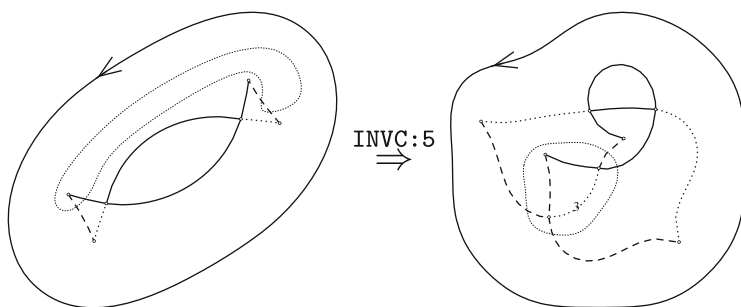


Fig. 10.29 In the indicated zone we can apply an inverse of the C simple rule leading to the result shown on the *right*

`liststrata`. Prints the number of strata (the value of f_{Σ}) for each region.

`countcc`. Prints the number of connected components of the surface ∂E .

`ccordering`. Gives the containment relation among the connected components of ∂E . A connected component ∂E_1 is contained in the connected component ∂E_2 if it is not possible to connect ∂E_1 with infinity without having to cross ∂E_2 , it is *directly* contained if there is no other components in between. A component is external if it is not contained in any other component. For instance, a result of “1{2 3}” means that the external connected component 1 directly contains the connected components 2 and 3 (contour `ccordering twohollowsphere`). A result of “1 2 3” means that the three connected components (example `three_spheres`) are all external.

`ccparent <n>`. Shows which connected component of ∂E directly contains the connected component n , a result of None means that the component n is external (see command `ccordering` above).

`ccorientation <n>`. Positive or negative orientation of the connected component n of ∂E . The orientation of a connected component is positive if it is external, otherwise it is the opposite of the connected component that directly contains it. The terminology comes from the fact that ∂E is the boundary of the solid shape E and this induces an orientation on each connected component of ∂E , that can be compared to the orientation of the connected component when considered alone.

`compare`. This command requires two apparent contours as input. They can be given in two input files (contour `compare sphere torus2`) or, equivalently, in a single file or a stream passed to the standard input as in the example `cat sphere.morse torus2.morse | contour compare`, here the two files must be in the current directory. The result shows which apparent contour (after canonification) is *simpler* in the sense explained in Sect. 10.6. The result `s1 = s2` implies that the two apparent contours are equivalent, the converse is also true unless we disable canonification, see Sect. 10.6.

`mendes`. Prints the *Mendes* graph of the apparent contour discussed in Sect. 10.8 [6]. The result is printed in text form as “mendes { <Nodes information> <Arcs information> }.” The <Nodes information> section lists each node of the Mendes graph and gives its sign (“+” or “-”) and its genus. The <Arcs information> section lists each arc and displays the id number of the positive and the negative node connected by that arc.

10.10.2 Operating Commands

These are commands that modify the given apparent contour in various ways according to the requested action. Some of the modifications correspond to 3D isotopic deformations of the surface associated with the apparent contour (with labelling), see Sect. 10.4; other commands correspond to the result of applying some surgery to the contour, see Sect. 10.5.

`rule <rule>:n:i`. Applies the indicated rule to the given apparent contour. The rules are all listed in Sect. 10.4, they are K0, K1, K1b, K2, T, L, B, S, C (simple rules), CR0L, CR0R, CR0Lb, CR0Rb, CR1, CR2, CR3L, CR3R, CR4L, CR4R, CR4Lb, CR4Rb (composite rules), INVK0, INVK1, INVK1b, INVK2, INVB, INVL, INVS, INVSb, INVC (inverse rules). The complete syntax includes the indication of a number n corresponding to a numbering (starting from 1) of all positions in the apparent contour where the given rule is applicable (often there is more than one position where a given rule is applicable) and the indication of an integer i whose digits in base 2 tell where to place the *islands* when a region is split into two by the rule. Both n and i are optional; if i is not present, the second colon should be omitted; if both n and i are not present, both colons should be omitted. A complete description of the rules is given in Sect. 10.4.

`mergearcs`. This applies the inverse of one of the rules K0, K1, K1b, K2, B. Options `-r` and `-a` (twice) are mandatory and allow to specify where the inverse rule must be applied, Which one of the listed rules should be applied depends on the relative orientation of the arcs and on their labelling. See also command `listma` in Sect. 10.10.1.

`wrinkle`. This applies the inverse of rule L. Options `-r` and `--stratum` are to be used to indicate where the inverse rule must be applied. See also command `listinvl` in Sect. 10.10.1.

`swallowtail`. This applies the inverse of rule S, thus creating a swallow's tail. Option `-a` must be used to indicate where the inverse rule must be applied. See also command `listinvs` in Sect. 10.10.1. Note that command `swallowtail` only creates one of the two variants listed by the command `listinvs`.

`puncture`. This applies the inverse of rule C, thus creating a thin triangle-shaped new region. Option `-a` must be used twice to indicate both the cusp and the arc involved in the rule. See also command `listinvc` in Sect. 10.10.1.

`punchhole`. Applies a vertical surgery to the apparent contour, see Sect. 10.5.1. It is mandatory to indicate the region involved and the two strata to be glued using the option `"-r < r > : < s > "`. The strata s and $s + 1$ of region r are glued together by removing a small disk from both strata and glueing the resulting boundaries together. For instance, the command `"contour punchhole hollowsphere -r 1:0"` creates a tube that connects the first two (of a total of four) strata of region 1.

`removehole`. Applies the inverse of the `punchhole` surgery. It can be applied only to simply connected regions bounded by an \mathbb{S}^1 (closed arc without cusps) oriented clockwise.

`gluearcs`. Applies a horizontal surgery to the apparent contour, see Sect. 10.5.2. The user must indicate the two arcs to be glued using the option `"-a"` twice like for instance in `"-a 1:0 -a 2:0"` which refers to the extended arcs 1 and 2. The `" : 0 "` substring refers to the portion between cusps of the extended arc (before the first cusp if 0).

`pinchneck`. Applies the inverse of the `gluearcs` surgery. The user indicates the two arcs where to apply the surgery using the `"-a"` option twice.

`addsphere`. Computes the apparent contour resulting from the addition of a small spherical surface to ∂E , this amounts to adding an \mathbb{S}^1 in one of the regions with the appropriate labelling. For instance, we obtain a hollow sphere from a sphere with the command

```
$ contour addsphere sphere -r 1:1
sketch {
Arc 2: (1);
Arc 1: (0);
Region 0 (f = 0): () (-a1);
Region 2 (f = 4): (+a2);
Region 1 (f = 2): (+a1) (-a2);
}
```

The “-r 1:1” option indicates the region of the apparent contour (the 1 before the colon) and requires that the sphere be added between stratum 0 and stratum 1 (the 1 after the colon). The option “-r 1:0” in the command before would add a small sphere in front of the original one. Although similar in syntax and application to the action `punchhole`, the effect on the structure of the solid shape and its apparent contour is quite different.

`removesphere`. Removes a small sphere indicated by a region of the apparent contour having an \mathbb{S}^1 oriented counterclockwise as boundary. The region is indicated using option “-r”.

`wrap`. Has the effect of wrapping the whole solid set E in a big sphere. Note that points in the inside of E become external and vice versa points not in E but inside the big sphere become internal.

`extractcc <int>`. Extracts the indicated connected component of the surface ∂E by removing everything else. The result is of course the apparent contour of a connected surface.

`removecc <int>`. Removes the indicated connected component from the surface ∂E . The result is the apparent contour of a surface with one less connected components.

`leftright`. Applies a left–right symmetry to the apparent contour and changes orientations of the arcs appropriately. The resulting apparent contour corresponds to a left–right reflection of the solid shape E with respect to a plane orthogonal to the projection plane.

`frontback`. Applies a front–back symmetry to the solid shape originating the apparent contour with respect to the projection direction. This has the effect of reversing the labelling.

`3devert <int>`. This action corresponds to an *eversion* of the surface ∂E . It works by selecting an external portion of the surface, stratum 0 of the region indicated by the argument *int*, cutting out a disk in that position and attaching a large wrapping sphere at the boundary of a similarly cut disk. The same result can be achieved by wrapping ∂E with the action `wrap` and then applying a `punchhole` action.

`evert <int>`. The selected region (indicated by the argument *int*) becomes the unbounded region. This is obtained by a projective transformation of the retinal

plane that moves some point internal to the region to infinity. Since the value of f_Σ is not necessarily 0 in the selected region we end up with an apparent contour with a nonstandard value of f_Σ at infinity. The software has partial support for such apparent contours and the use of option “`--infinity < f >`” allows to force a nonzero value of f_Σ at infinity for an apparent contour to be given in input.

`union`. This gives the result of the union of two solid shapes E_1 and E_2 , given by their apparent contours, and placed in such a way not to occlude each other. The description of the two apparent contours is given as one or two arguments or passed to the standard input (see also the action “`compare`” above).

`sum`. This is similar to the “`union`” command, but performs a horizontal surgery afterwards to glue the two apparent contours in a manner corresponding to the connected sum of the boundaries of the two solid shapes.

10.10.3 Conversion and Standardization Commands

These are commands that convert between different types of contour descriptions.

`print`. Prints the region description of the given apparent contour. The description is first canonified as explained in Sect. 10.6 unless the option “`--nocanonify`” is present. This is almost a trivial action since apparent contours are internally described using their region description. Most of the contour commands print the canonified region description of the resulting apparent contour after the required action, so this command is equivalent to a *no operation* on the input data.

`printmorse`. Computes and prints a Morse description of the apparent contour. This is extremely useful in order to interface the `contour` program with the `showcontour` program that displays a graphic picture of the apparent contour based on a Morse description. The result is obtained by letting an imaginary sweep line traverse the apparent contour which is described in terms of its region description. Note that the Morse description obtained with this action is often more involved than necessary and in particular it is different from the Morse description of the apparent contour given in input. This is because the original Morse description is converted into a corresponding region description and then back into an equivalent Morse description. For this reason, to display a picture of an apparent contour described by a Morse description (most of the examples are described in this way) it is usually better to feed the description file directly to the `showcontour` program rather than use `contour printmorse` as a filter.

`knot2morse`. Converts a *knot-type* contour, or more generally the knot description of a handlebody [18], into the corresponding Morse description, see Sect. 10.3.3. This is done typographically without the usual internal conversion through the region description. Each typographical character of the knot description is substituted by a small matrix of typographical symbols that describes locally the apparent contour by means of the Morse description.

`canonify`. Prints the region description after canonization (see Sect. 10.6). This actually produces the same result as the `print` command.

10.10.4 Cell Complex and Fundamental Group Commands

Commands related to the cell complexes of the 3D sets associated with the apparent contour (with labelling).

`cellcomplex`. Computes a cell complex corresponding to the surface $\Sigma = \partial E$ as described in Sect. 7.5.1. Let us give a very simple example when E is a solid sphere; the corresponding apparent contour is described in file `sphere.morse` and consists in just one closed arc bounding a region with the shape of a disk.

```
$ contour cellcomplex sphere
node 0 ref 2
arc 0[0,0] ref 2
face 0 [+0 ]
face 1 [-0 ]
Found 1 connected component
```

The constructed cell complex consists in a single node (numbered 0), a single arc that starts and ends at node 0 and two faces, both having arc 0 as their boundary oriented in opposite directions. This implicitly gives an orientation to the two faces (floor and ceiling of the sphere E). We can note here that the orientation of the faces chosen by the software is **not** consistent across arcs, although we know that a globally consistent orientation do exist since Σ is oriented.

`insidecomplex`. Computes a cell complex corresponding to the solid set E as described in Sect. 7.5.1. If we take a sphere as the set E as before, we simply get

```
$ contour insidecomplex sphere
node 0 ref 0
```

This is the trivial cell complex consisting of a single 0D cell, which is the result of the application of a sequence of simplifying deformation retractions.

`outsidecomplex`. Computes a cell complex corresponding to the complement $\mathbb{R}^3 \setminus E$ as described in Sect. 7.5.1. If we take a sphere as the set E as before, we obtain

```
$ contour outsidecomplex sphere
node 1 ref 3
node 2 ref 3
arc 1[1,1] ref 2
arc 2[2,2] ref 2
arc 5[1,2] ref 2
face 1 [+1 ]
face 2 [+2 ]
face 3 [+1 +5 -2 -5 ]
```

This is more involved than the other two cell complexes, and derives from the way the software actually works, i.e., by placing everything in a big bounding ball B . Indeed the result is exactly the same as that obtained by asking the inside cell

complex of a solid hollow sphere. In particular face 3 corresponds to an annulus (arc 5 is the cutting arc added to make the annulus simply connected) which derives from the external region of the apparent contour, bounded by the apparent contour of the ball B .

`fundamental` (or `fg`). Computes the fundamental group of the surface $\Sigma = \partial E$. The group is printed by means of a *presentation* as described in Sect. 10.7.2. We stress that different presentations can refer to the same group.

`insidefundamental` (or `ifg`). Computes the fundamental group of the solid shape E . The group is printed by means of a *presentation* as described in Sect. 10.7.2. We stress that different presentations can refer to the same group. This command is equivalent to the combination `fg --in`.

`outsidefundamental` (or `ofg`). Computes the fundamental group of the complement $\mathbb{R}^3 \setminus E$ of the solid shape E . This command is equivalent to the combination `fg --out`.

`abelianizedfundamental` (or `afg`). Computes the quotient of the fundamental group by its commutator, i.e., its abelianization. The resulting group is described in terms of its rank and its invariant factors (torsion numbers). As already observed in Sect. 10.7.3 fundamental groups constructed in the context of apparent contours are always torsion-free, see also Sect. 10.10.7.

`alexander`. Computes the Alexander polynomial, or more generally one of the Alexander ideals, from a group presentation. At the moment this computation is only performed when the abelianized is the infinite cyclic group \mathbb{Z} , in which case the Alexander polynomial is a Laurent polynomial in one indeterminate t , or in the case where the abelianized is $\mathbb{Z} \times \mathbb{Z}$ (torsion free with rank 2), in which case the Alexander polynomial is a Laurent polynomial in two indeterminates u and v . In the case of one indeterminate, the Alexander polynomial $p(t)$ always satisfies $p(1) = 1$ and is defined up to multiplication by powers of t or the substitution $t \rightarrow \frac{1}{t}$. The resulting polynomial is hence canonified such that $p(0) \in \mathbb{R} \setminus \{0\}$. Invariance with respect to $t \rightarrow \frac{1}{t}$ is ineffectual since it is known that the Alexander polynomial of knots has symmetric coefficients. In the case of two indeterminates, the Alexander polynomial is only partially canonified (see Sect. 10.7.4 for details). Option `--foxd <i>` can be used to force the index of the Alexander ideal that should be computed in some special cases; details on this can be found in Sect. 10.7.4.

`linkingnumber`. This is valid only when the preabelian presentation has rank 2 and the number of relators is one less the number of generators. This happens for the complement of a two-components link (two solid tori) or the equivalent situation obtained after a 3D eversion (see examples in Sect. 10.14). See Sect. 10.7.3 for details.

10.10.5 Options Specific to Fundamental Group Computations

They are only meaningful for commands `cellcomplex` (only options `--in` and `--out`), `fg` (or `fundamental`), `alexander`, `linkingnumber`.

`--in`. Apply the given command to the internal set E . `--out`. Apply the given command to the external set $\mathbb{R}^3 \setminus E$.

`--preabelian`. Applies the procedure illustrated in [11, Sect. 3.3] to produce a *preabelian* presentation.

`--nosimplify`. Inhibits the automatic simplification process of the presentation of the fundamental group as computed from the cell complex.

`--foxd <i>`. Indicates the index of the elementary ideal to be computed (see Definition 7.7.2). If missing, the implied value depends on the deficiency of the finitely presented group and is generally the smallest value that produces a nontrivial result ($d = 1$ for the outside of a knotted solid torus or two solid tori, $d = 2$ for the inside/outside of a surface of genus 2). In order to avoid misinterpretations of the result, whenever the value of d is implied it will be displayed in the output of the computation.

`--nobasecanonify`. Inhibits the canonization process for the Alexander polynomial/ideal. A warning message will be printed to alert the user that the result might be non-canonical.

`--shuffle`. The Alexander polynomial/ideal is subjected to a random change-of-base before application of the base canonization. This might be useful for debugging purposes to test the computational complexity of the base-canonification procedure. Option `--seed` followed by a number can be used to initialize the pseudo-random generator.

`-Q`. Format the result using a syntax that can be understood in input by the program, see Sect. 10.10.7.

If neither of the two options `--in` or `--out` is given, the requested action is applied to the surface itself. In some cases there are abbreviated commands that imply option `--in`, like `ifg` or option `--out`, like `ofg`.

10.10.6 Common Options

Options can alter the way that commands behave, or provide additional information.

`--help`. Prints a brief list of all available commands and options.

`--version`. Prints the version number of the program.

`-q`. Reduces the amount of text displayed to a minimum. This is useful if the output needs to be parsed by automatic scripts or other programs.

`-v` (or `--verbose`). Increases the verbosity of the output for some actions.

`--nocanonify`. Disables in most cases the automatic canonification of the region description before printing it.

`--transfer_islands <int-coded-flags>` (or `--ti <int-coded-flags>`). Application of some rules (most notable the B, beak-to-beak, move) to an apparent contour has an indeterminate behaviour in presence of non-simply connected regions, as described in Sect. 10.4.2. This option allows to precisely control the result. The integer argument is interpreted as a sequence of *bits* (digits in base 2), each bit controlling the final position of an island (part of the apparent contour contained in a hole of the region).

`--finfinity <int>`. There is partial support for apparent contours with a nonzero value of f_Σ at infinity. They can arise if the target space is \mathbb{S}^2 instead of \mathbb{R}^2 with the North pole of \mathbb{S}^2 identified with the *point at infinity* of \mathbb{R}^2 . Such apparent contours are not allowed in the definition of Sect. 3.2, however it might be useful to use them in particular circumstances. The `evert` action (which in some sense corresponds to a motion of the point at infinity into one of the regions) typically produces this kind of apparent contours.

`--mendes_ge <type>`. Allows for a rough graphical presentation of the Mendes graph, to be used in conjunction with action `mendes` described in Sects. 10.8 and 10.10.1. `type` can be one of

`text`: This is the default, and produces a textual display of the Mendes graph as described in Sect. 10.10.1;

`kig`: Produces a graphical representation using the syntax of a save file for the `kig` program for interactive geometry (<http://kig.kde.org/>);

`pykig`: Similar to the above, but produces a file suitable to be executed by the `pykig` frontend to `kig`.

The graphical position of the nodes is obtained with the help of a random number generator.

`--seed <int>`. Allows to give a seed for the random number generator used to produce a graphical presentation of the Mendes graph.

`-r <int>` (or `--region <int>`). Some actions require the indication of a region where to operate, this can be done using this option. The syntax “`-r n:i`” can be used to indicate stratum i (starting from 0) of region n .

`-a <int>` (or `--arc <int>`). Some actions require the indication of an extended arc where to operate, this can be done with this option. The syntax “`-a n:i`” can be used to indicate a specific arc of an extended arc when cusps are present ($i = 0$ refers to the arc from the starting node to the first cusp) or to indicate a specific cusp ($i = 0$ refers to the first cusp). If two arcs are required (e.g., for action `mergearcs`) this option must be used twice.

10.10.7 Direct Input of a Finitely Presented Group or an Alexander Ideal

All `appcontour` commands related to fundamental groups (more generally, finitely presented groups) allow the user to directly input a group presentation, on standard input or from a properly formatted file. As an example the file (named `trefoilknot.fpgroup`) containing

```
fpgroup {
<x, y; xxYYY>
}
```

describes one of the possible presentations of the fundamental group of the trefoil knot.

We can use the description contained in `trefoilknot.fpgroup` in an `appcontour` command as in

```
$ contour fg --preabelian trefoilknot.fpgroup
Finitely presented group with 2 generators
<a,b; bAbaBBa>
```

or we can enter directly the group description as in

```
$ contour afg
fpgroup { <x, y; xxYYY> }
Torsion-free abelian group of rank 1
```

or

```
$ contour alexander
fpgroup { <x, y; xxYYY> }
Computing Alexander ideal for d = 1:
Alexander polynomial (up to t -> 1/t):
+1-t+t^2;
```

where the lines with gray background are entered by the user as input for the software code.

Since the fundamental groups of the three sets Σ , E , $\mathbb{R}^3 \setminus E$ are always of a very special nature, and in particular are always torsion-free, direct feed of a group presentation can be the only way to force the software to operate on them. In this case the two options (see below) `--in` and `--out` have no effect, and option `--nosimplify` can be very useful to exclude the preliminary automatic simplification attempted by the software.

Similarly, it is possible to directly enter an Alexander polynomial or Alexander ideal, for example in order to take advantage of the canonification postprocess. Here are two examples.

```
$ contour alexander
alexander(t) {
t^(-1)+t-1;
}
Alexander polynomial (up to t -> 1/t):
+1-t+t^2;
```

```
$ contour alexander
alexander(u,v) {-1+u+uv;}
Alexander polynomial:
+1+u-v;
```

In the latter example the difference between the input polynomial and the resulting one is a consequence of the change of base of \mathbb{Z}^2 performed in the

base-canonization process. Option `--nobasecanonify` inhibits the base-canonization and the resulting polynomial will be equal to that given in input.

The syntax to be used should be clear from the above examples, that cover the case of finitely presented groups and Alexander polynomials. It requires a brief explanation for the description of an Alexander ideal. The description starts with `“ideal(u,v) {”`, and must be closed by a `“}”`. The name of the variables can be changed and must match those used in the polynomial description, they must consist of a single lowercase letter. The description of the ideal itself consists of one or more polynomials (the generators), each one terminated by a semicolon `“;”`. The polynomial description can optionally be preceded by the keyword `“F:”`, in which case the subsequent polynomial is to be multiplied by the fundamental ideal \mathcal{E} and the resulting ideal will act as a generator.

For all contour commands that produce a finitely presented group or a Laurent polynomial/ideal (in one or two indeterminates), option `-Q` can be used to obtain a formatted printout in a syntax that can be directly used as input. Here is an example:

```
$ contour -Q --out alexander whitehead
#
# --foxd 1
#
ideal(u,v) {
F: +1-u-v+uv;
}
```

10.11 showcontour Reference Guide

`showcontour` is a utility that produces a graphical display of an apparent contour. As for the `contour` program, we are only interested in the topological structure of the apparent contour, so that the aim is to produce a simple drawing up to an \mathbb{R}^2 -ambient isotopy. Moreover the program must be able to display apparent contours obtained by `contour`.

The Morse description seems the most appropriate way to translate a contour description into a drawing, by somehow *inverting* the process of producing a Morse description from a drawing.

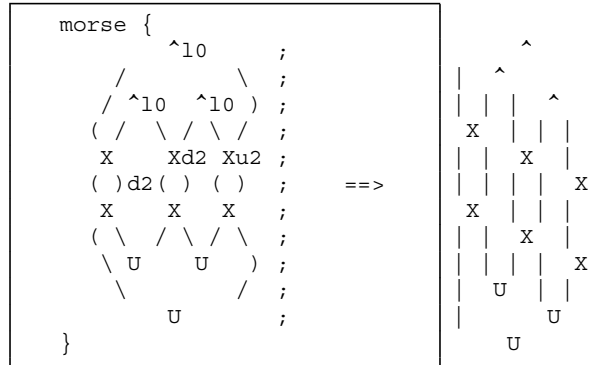
There are four stages in the process; hereafter, we will shortly describe them, by using as an example the apparent contour of three spheres with partial occlusions.

10.11.1 Producing a Proper Morse Description

The Morse description produced, e.g., by `contour printmorse`, or used by the user to describe an apparent contour, allows for more than one Morse event (other than transversal crossings, that are not, properly speaking, Morse events) at each

critical level; it also allows for levels with only transversal crossings. It is convenient to convert such a description into one where there is exactly one (proper) Morse event per line.

We illustrate this with the following example:



where on the left we have a Morse description that follows the syntactic rules of Sect. 10.3.2 (the same also used by `contour printmorse`), and on the right we have a schematic representation of the same Morse description, where lines with only transversal crossings have been removed, and lines with more than one Morse event are split in more lines with one Morse event each. In the example there are a total of 12 Morse events: three maxima, six crossings and three minima.

10.11.2 From the Morse description to a polygonal drawing

From the new Morse description we can produce a polygonal drawing made with unit segments oriented horizontally or vertically, where the Morse line is substituted by a straight line tilted at 45 degrees with critical levels at regular intervals (see Fig. 10.30). It is not difficult to procedurally obtain the polygonal, by associating with each Morse event a set of two (for *maximal/minima* events) or four (for *crossings* events) unit segments. By adding other small portions corresponding to the transversal crossings, we obtain in this way a (nonsmooth) drawing that already has the required properties.

10.11.3 Discrete Optimization of the Polygonal Drawing

As it happens, the Morse description obtained by `contour printmorse` is often quite involved. As a result, the polygonal drawing obtained with the previous step can be quite complex and difficult to understand. The third step then performs a *discrete* optimization consisting in subsequent application of a few local rules that

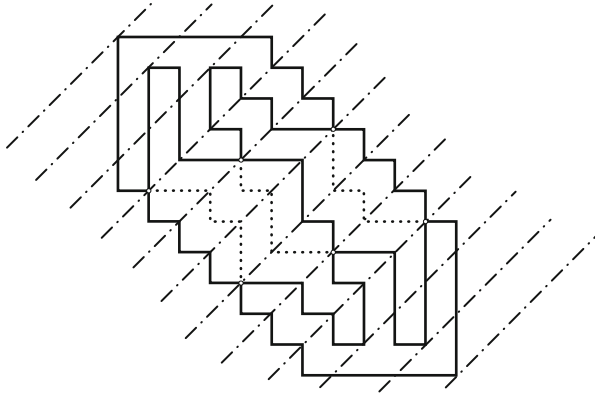


Fig. 10.30 The starting polygonal, directly obtained from the Morse description. The tilted lines are Morse lines corresponding to proper Morse events

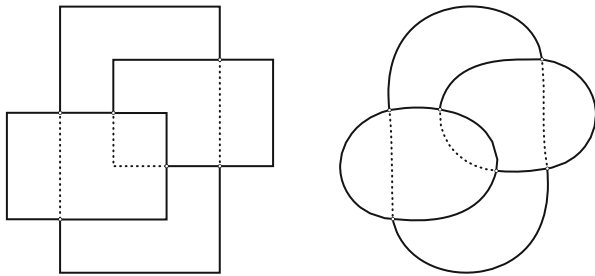


Fig. 10.31 The result of the discrete optimization (*left*) and the effect of the smoothing evolution (*right*)

simplify the drawing by still using only unit horizontal or vertical segments. The result of such optimization is illustrated in Fig. 10.31, left.

10.11.4 *Dynamic Smoothing of the Polygonal*

Finally, we need to smooth out the polygonal. This is done by replacing the unit segments with smaller segments that are no longer constrained to be horizontal or vertical. The resulting new polygonal is interpreted as a discretized curve, that we subject to a smoothing evolution process. The evolution is actually just a gradient flow with respect to a suitable action functional.

It is in this stage that cusps are introduced, originally just as distinguished points in the polygonal curve.

The action functional contains a length term and also a curvature term, the latter having a strong smoothing effect and preventing in particular the vanishing of a circle.

Unfortunately, the local nature of the functional does not prevent the curve from self-intersecting (which is of course undesirable) which can sometimes happen in complex examples. Moreover the presence of the curvature term leads to a fourth-order evolution equation, which is sometimes slow to converge.

In most circumstances, however, this technique produces acceptable results.

10.12 Using contour in Scripts

Some useful scripts included in the appcontour package can simplify the use of the program for specific tasks.

10.12.1 *contour_interact.sh*

This script is tailored for application of successive Reidemeister-type rules. It is invoked as “*contour_interact.sh <example>*” where *example* refers to a file containing the description of the apparent contour in any of the accepted forms: region description, Morse description, knot description. The prompt “Contour>” then accepts a small set of commands that in particular include the set of applicable rules, including inverse rules and composite rules. The apparent contour is modified during the session according to the commands given. At any time the command “show” makes a call to the showcontour program in order to graphically display the current apparent contour. We illustrate the basic usage with excerpts from a sample session.

```
$ contour_interact.sh genus_2_linked
Found matching file: /usr/local/share/appcontour/examples/genus_2...
found showcontour [gtk] in /usr/local/bin
exemplename: genus_2_linked
This is an apparent contour with labelling
[...]
Applicable rules: CR2
Contour> cr2
OK, applying rule cr2
Applied rules: cr2
Applicable rules: CR0L CR0RB CR3L CR3R CR3R:2
Contour> cr0rb
OK, applying rule cr0rb
Applied rules: cr2 cr0rb
Applicable rules: B CR0L CR0R CR2 CR3L CR3R
Contour> show
[A window with a drawing of the current apparent contour is displayed]
Applied rules: cr2 cr0rb
Applicable rules: B CR0L CR0R CR2 CR3L CR3R
```

```

Contour> back
Applied rules:  cr2
Applicable rules: CR0L CR0RB CR3L CR3R CR3R:2
Contour> cr0l
OK, applying rule cr0l
Applied rules:  cr2 cr0l
Applicable rules: B CR0LB CR0RB CR2 CR3L CR3R
Contour> quit
$

```

The accepted commands (besides the applicable rules) are:

- **info**. Displays the result of “contour info” for the current apparent contour;
- **print**. Prints the region description of the contour;
- **morse**. Prints the Morse description of the contour;
- **back**. Goes back one step in the history of applied rules (undoes the effect of the last rule requested);
- **show [rule]**. Invokes showcontour to graphically display the apparent contour, if a rule is specified, then the contour displayed corresponds to the result of applying the rule;
- **quit**. Ends the session;
- **wrinkle**, **mergearcs**, **swallowtail**, **puncture**. Each of these commands displays a list of applicable inverse rules in the classes respectively: INVL, INVB or INVK, INVS, INVC.

10.12.2 *contour_describe.sh*

This script extracts information about the embedded surface originating an apparent contour (with labelling) and describes it in english words, here is an example

```

$ contour_describe.sh -n sphere_behind_torus
Total number of connected components: 2
External components:
  component #1, genus 0
  component #2, genus 1
$ contour_describe.sh -n three_spheres
Total number of connected components: 3
External components:
  component #1, genus 0
  component #2, genus 0
  component #3, genus 0
$ contour_describe.sh -n hollowsphere
Total number of connected components: 2
External components:
  component #1, genus 0
Component #1 contains:
  component #2, genus 0

```

As can be seen, the Euler–Poincaré characteristic allows to compute the genus of each connected component of the embedded surface. Furthermore, the labelling allows to recover the relative relationship among the connected components in terms of containment. This is because an embedded closed surface always divides the ambient space \mathbb{R}^3 in an *inside* and in an *outside*. A hollow solid sphere is a 3D subset of \mathbb{R}^3 the boundary of which is composed of two concentric spheres, the external surface and the internal surface of the hole. In the description of this surface the boundary of the hole (a sphere) is described as contained in the external boundary (a sphere). Option “-n” is used to ask for a rather formal and essential description in terms of containment relationship and genus of each connected component of $\Sigma = \partial E$.

If no option is given, or if we use option “-z”, then we obtain an amusing verbal description in the spirit of the old *zork* textual adventure games. This latter description also contains a larger amount of information since it also takes into account the result of computing the fundamental groups and Alexander polynomials. Here are the examples above as obtained with this type of description.

```
$ contour_describe.sh sphere_behind_torus
You are in a clearing, with a forest surrounding you on all sides.
There is a total of two objects here.

You can see a sphere and a torus.

$ contour_describe.sh three_spheres
You are in a clearing, with a forest surrounding you on all sides.
There is a total of three objects here.

You can see a white sphere, a black sphere and a red sphere.

$ contour_describe.sh hollowsphere
You are in a clearing, with a forest surrounding you on all sides.
There is a total of two objects here.

You can see a white sphere.
The white sphere contains a black sphere.

$ contour_describe.sh genus_2_linked
You are in a clearing, with a forest surrounding you on all sides.
There is one object here.

You can see a double torus.
```

A “double torus” indicates a sphere with two handles (genus 2). The indicated color of the objects is of course completely fictitious and allows to identify the same object in different parts of the description. Here are a few examples involving the computation of the Alexander polynomial.

```
$ contour_describe.sh internalknot
You are in a clearing, with a forest surrounding you on all sides.
```



```

There is one object here.

You can see a torus with a knotted hole, you can read "Alexander:
1-t+t^2" written in the inside.

$ contour_describe.sh painted_klein_bottle
You are in a clearing, with a forest surrounding you on all sides.
There is one object here.

You can see a double torus with knotted holes, you can read "Alexander:
-1+u+v" written in the inside.

$ contour_describe.sh links/link_2_6_1
You are in a clearing, with a forest surrounding you on all sides.
There is a total of two objects here.

You can see a white torus and a black torus.
The white torus is linked three times to the black torus.

$ contour_describe.sh whitehead
You are in a clearing, with a forest surrounding you on all sides.
There is a total of two objects here.

You can see a white torus and a black torus.
The white torus is linked with the black torus.

```

In the last example the `contour_describe.sh` script uses the phrasing “is linked with” to describe a situation where the presentation of some fundamental group (the fundamental group of the outside $\mathbb{R}^3 \setminus E$ in this case) is nontrivial with a nontrivial Alexander ideal, meaning that the link is not splittable. However, the linking number turns out to be zero. The Whitehead link is indeed the simplest example of an unsplittable link (with two components) where the two components have vanishing linking number.

10.12.3 *contour_transform.sh*

The script “`contour_transform.sh <example> <n>`” automates the task of applying all possible sequences of a maximum of n successive moves (direct rules, simple or composite). Using large values of n is typically impractical, since the number of possible sequences grows exponentially. Using the “`contour compare`” action the script is able to sort all obtained apparent contours, so that it can avoid applying rules to apparent contours that were already obtained with a different path of subsequent rules. This script was used during the writing of this book in order to find a simplifying set of moves for the examples of Sect. 10.13 and

Sect. 10.15. As a very simple example, we can check the result of two successive rules applied to the apparent contour of the torus displayed in Fig. 10.1.

```
$ contour_transform.sh torus2 2
Transforming torus2 with rule K0
Transforming torus2 with rule B
Transforming torus2 with rule B:2
[...]
```

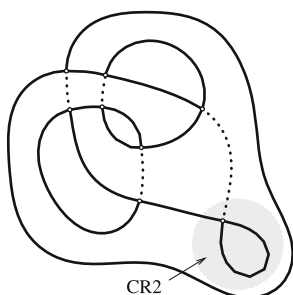
A directory named `torus2.transformations` is created containing all resulting apparent contours; they are numbered in such a way that equal numbers correspond to diffeomorphically equivalent apparent contours. In the example at hand we obtain a total of 30 different apparent contours; some of them can be obtained with a different set of rules, for example applying `CR0Lb` and then `S` gives the same result as first applying `K0` and then `CR4L`.

10.13 Example: knotted Surface of Genus 2

Figure 10.32 (left) shows a knotted surface of genus 2, a sphere with two (linked) handles. A simple way to describe this surface is by means of a knot description, as shown on the right of Fig. 10.32.

Surprisingly, it is possible to find a continuous deformation of the ambient space that “unknots” this surface. We start by asking the program to compute the fundamental group of the inside and outside of the surface:

```
$ contour ifg genus_2_linked
Free group of rank 2
<a,b; >
$ contour ofg genus_2_linked
Free group of rank 2
<a,b; >
```



```
knot {
  ^ ^ ;
  | x | ;
  | x | ;
  y y ;
  U ;
}
```

Fig. 10.32 An apparently knotted surface of genus 2. The *shaded region* shows where the first rule (CR2) of the unlinking chain of rules is applied

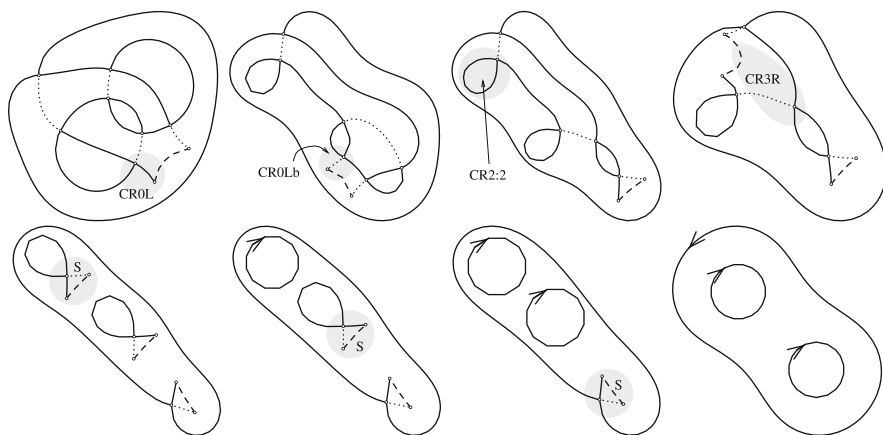


Fig. 10.33 Intermediate results of application in sequence of moves CR2 CR0L CR0Lb CR2:2 CR3R S S S S to the apparent contour of Fig. 10.32

That the “inside” is the free group of rank 2 is not surprising, since the solid inside can clearly be deformation retracted to a one-dimensional CW complex equivalent to a bouquet of two circles. More interestingly, the “outside” has the same fundamental group, hinting that the surface can be unknotted. Of course this is by no means a proof of equivalence: having the same fundamental group is a necessary condition for isotopic equivalence, but in general it is *not* sufficient.

We can though try to apply Reidemeister-type moves to the apparent contour to obtain the apparent contour of an unknotted sphere with two handles (torus with two holes).

It turns out that this is indeed possible, and the minimum number of moves required to achieve the goal is eight (counting composite moves as one), it can be done in more than one way, here is one possible sequence of moves: CR2 CR0L CR0Lb CR2:2 CR3R S S S S.

Figure 10.33 shows the result of the application in sequence of the list of moves mentioned above, with a final apparent contour of an unknotted torus with two holes.

In order to find the optimal simplifying sequence we used the script `contour_transform.sh` to obtain the result of all sequences of eight subsequent moves. This is a quite heavy computation and produced a total of 13, 560 nonequivalent apparent contours, ordered from simplest to most complex using the “`contour compare`” action. According to “`contour compare`” the simplest result is indeed the unknotted torus with two holes appearing last in Fig. 10.33.

10.14 Example: Knots in a Solid Torus

In Fig. 10.34 we show the diagram of four knots, three of them can be readily seen to be \mathbb{R}^3 -ambient isotopic (smoothly deformable as an embedded closed curve) to the trivial knot (the “unknot”), whereas the lower-left diagram is a realization

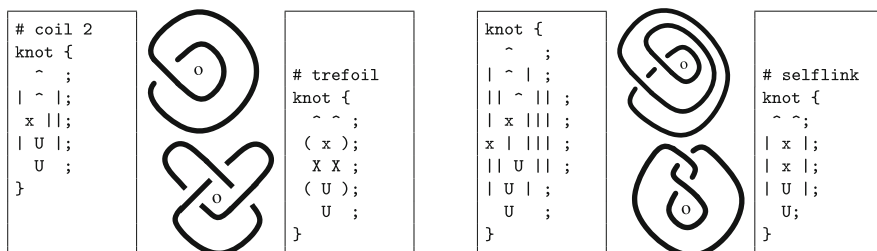


Fig. 10.34 Diagram of four knots with the corresponding “knot” description. They have to be interpreted as knotted tubular surfaces which we shall subsequently embed inside a solid torus (the small letter ‘o’ indicating the hole of the torus). The *lower-left* diagram corresponds to the trefoil knot, the others are all *unknots*, they can be deformed to the trivial knot

of the trefoil knot. Using the “knot” description we can regard these diagrams as representing a knotted tubular surface diffeomorphic to a torus and bounding a solid torus.

We already used the `appcontour` program to compute the fundamental group of the trefoil knot (Sect. 7.5.2, Fig. 10.18) and its Alexander polynomial; in the three remaining cases, the fundamental group is computed as $\langle a \rangle$, the free abelian group of rank 1, which provides sufficient evidence that the three knots are the unknot, thanks to the unknotting theorem [18, page 103].

We shall now enclose each of the four knotted tubes inside a torus in such a way that the “hole” of the wrapping torus is positioned at the small letter “o” shown in the diagrams. This can be achieved using the `contour` command twice:

- `contour wrap`: first places the apparent contour inside a big sphere;
- `contour punchhole -r n:0`: performs a vertical surgery in the region indicated by the integer n . The appropriate value for n can be obtained by examining the region description resulting after the application of `contour wrap`; in all cases, the region of interest is easily identified.

On the final result we then compute the “internal fundamental group”, which is actually the fundamental group of the original knots *constrained to live inside the wrapping torus*.

As an example, all these operations can be carried out for the trefoil knot as follows:

```
$ contour wrap trefoilknot | contour punchhole -r 4:0 | contour ifg
Finitely presented group with 2 generators
<a,b; ababABAB>
```

That this is not isomorphic to the fundamental group of the “unconstrained” trefoil knot can be seen by computing the abelianization (e.g., using the command

`contour afg --in`) which turns out to be the free abelian group of rank 2: $\mathbb{Z} \times \mathbb{Z}$ ¹⁶.

The result of the computation of the internal fundamental group for all four examples is the following:

- (upper-left diagram): $\langle a, b; \text{abbABB} \rangle$
- (lower-left diagram, trefoil): $\langle a, b; \text{ababABAB} \rangle$
- (upper-right diagram): $\langle a, b; \text{abbbABBB} \rangle$
- (lower-right diagram): $\langle a, b; \text{abbaBabABBAbAB} \rangle$

By using Tietze transformations it can be shown that the first two presentations give the same finitely presented group where the relation says that the square of one generator commutes with the other generator.

The third presentation can be interpreted as two generators, the cube of one commuting with the other. The fourth presentation is rather long, however this fact does not imply that the corresponding group is different from the others. In all four cases the abelianization is always the free abelian group of rank 2.

By performing a 3D eversion that takes the point at infinity inside the wrapping torus the effect is to turn the wrapping torus “inside-out”, the internal solid torus is now external and we obtain a scene with two solid tori possibly linked with each other. This suggest to compute the linking number between the two tori; an example for the lower-left diagram is

```
$ contour wrap trefoilknot | contour punchhole -r 4:0
                             | contour linkingnumber --in
Linking number is 2
```

The same linking number 2 is obtained for the upper-left diagram, whereas we obtain 3 and 0 respectively for the upper-right and the lower-right diagrams.

The software code can distinguish the two scenes corresponding to the left diagrams by computing the Alexander polynomial of the scene obtained by removing the external torus: command `contour cparent` can be used to identify the external connected component, to be removed.

The `contour-describe.sh` script described in Sect. 10.12.1 can be used to do just this:

```
$ contour wrap trefoilknot | contour punchhole -r 4:0
                             > trefoil_in_torus.sketch
$ contour_describe.sh -z trefoil_in_torus.sketch
You are in a clearing, with a forest surrounding you on all sides.
There is a total of two objects here.
```

¹⁶This is not unexpected: the abelianized of the fundamental group coincides with the first homology group of the inside set E , the rank of which is the Betti number b_1 and can be computed by using the Euler–Poincaré characteristic $\chi = 0$ of E as $b_1 = b_0 + b_2 - \chi$ where the Betti number $b_0 = 1$ is the number of connected components of E and the Betti number $b_2 = 1$ is the number of “voids” of E , which in our context is the number of connected components of ∂E decreased by 1.

You can see a white torus.
 The white torus contains a black knotted torus, you can read "Alexander:
 $1-t+t^2$ " written on it.
 The black torus is linked twice into the white torus.

The lower-right diagram corresponds to a tricky situation, indeed a 3D eversion transforms the scene into the famous Whitehead link, a link with two components that has vanishing linking number, but the two components cannot be separated.

The appcontour software code can distinguish the 3D structure corresponding to the lower-right diagram (everted Whitehead link) from a scene where the inner torus does not circle around the hole of the wrapping torus by computing the Alexander ideal of the “inside”:

```
$ contour --in alexander contour
Computing Alexander ideal for d = 1:
Alexander ideal generated by:
(+1-u-v+uv) (u - 1);
(+1-u-v+uv) (v - 1);
```

where *contour* is the name of a file containing a description of the lower-right diagram.

10.15 Example: Klein Bottle and the “House with Two Rooms”

As a more complex example, we construct an embedded surface $\Sigma = \partial E$ inspired by the Klein bottle, the famous nonorientable closed surface. We start from the typical immersion of the Klein bottle in \mathbb{R}^3 displayed in Fig. 10.35 a). Since it is nonorientable, any immersion of the Klein bottle in the three-dimensional space must have self-intersections. In the immersion of Fig. 10.35 the self-intersection reduces to a single circle γ of double points. It bounds a small disk-like portion of the surface, which is shown in dark in Fig. 10.35 a).

We denote by F the set that we obtain from the immersed surface by removing this small disk. This is not a manifold, because three sheets of the surface meet at the curve γ . However γ is the only singular part of F ; if we remove a small tubular neighbourhood of γ , we obtain a 2D manifold with a boundary composed by three small deformations of γ , intersection of F with the boundary of the tubular neighbourhood.

The set F has two interesting properties:

- (1) it cannot be deformation-retracted into a smaller subset. Indeed we cannot remove any portion of F without changing its homotopy type;
- (2) the complement of F is connected. Indeed we can reach the *inside* of the bottle from the outside by following the dotted curve in Fig. 10.35b.

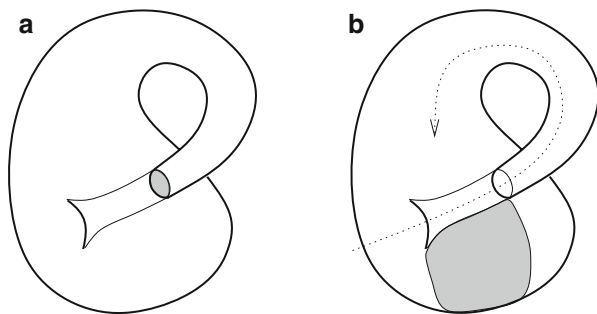


Fig. 10.35 Standard immersion of the Klein bottle in \mathbb{R}^3 . The small shaded disk in the left picture is removed to obtain an interesting surface with a ramification line (the boundary of the removed disk) where three sheets meet together

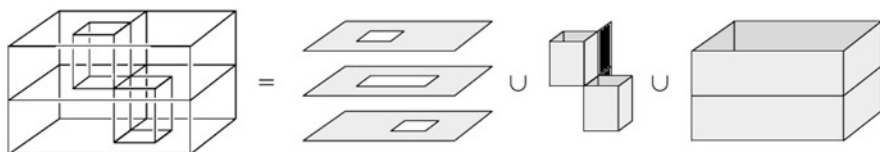


Fig. 10.36 The house with two rooms of Bing, original image taken from [8]. The two *black vertical panels* in the middle picture of the union sequence are added with the explicit purpose of preventing a complete tour around each of the two vertical tubes

Another famous example devised by Bing [1], and called *the house with two rooms*, is displayed in Fig. 10.36. It is a set that is contractible but not in an obvious way, see [8, page 4]. In that example the upper room can only be accessed through a tunnel traversing the lower room and vice versa.

Two of the panels forming the house with two rooms (coloured black in Fig. 10.36) are added with the explicit purpose of preventing the construction of closed curves around the “tunnels” in the complement of the house. If we remove those two panels, we obtain a set that satisfies the two properties above and also has the same (nontrivial) fundamental group of our example. It is described as the finitely presented group $\langle x, y, z; xyzx^{-1}z^{-1}xy^{-1} \rangle$.

This is exactly the construction of Exercise 11 in [8, Chap. 1, p. 53], where in particular it is proved that the constructed fundamental group is not isomorphic to the free group with two generators.

We note that the complement of F is not simply connected, since we can encircle the tube-like portion of F that lies inside of the bottle and also the neck of the bottle on the outside with a closed curve that cannot be deformed to a point without crossing F .

Now, we slightly enlarge F by considering the set of points with distance smaller than or equal to ϵ from F for a sufficiently small $\epsilon > 0$. We can smooth the boundary of this set and call E the result (a solid set). The boundary $\Sigma = \partial E$ is the image of a smooth embedding in \mathbb{R}^3 of a closed 2D manifold.

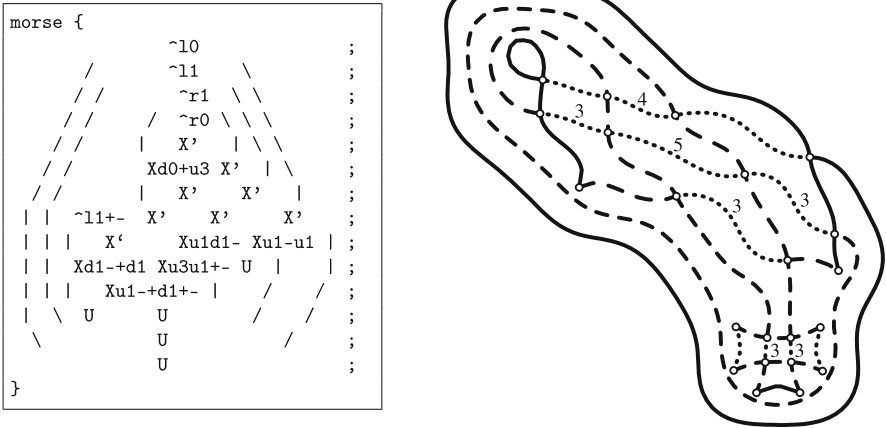


Fig. 10.37 Morse description of the “painted Klein bottle” obtained as boundary of a thickened Klein bottle with a small disc removed, as described in Sect. 10.15 (left). The corresponding apparent contour is displayed on the right

A possible Morse description of the apparent contour of this surface is shown in Fig. 10.37, left, which we write into a file called, say, `painted_klein_bottle.morse`. In this description there are a few complex references to Morse events of *crossing* type. For example, the sequence `Xd1-+d1` refers to a crossing where the extended arc in the south–west direction is oriented downwards with initial labelling 1 that decreases to 0 after the first cusp, then increases to 1 after the second cusp; the extended arc in the south–east direction is also oriented downwards, does not contain cusps and is labelled 1. The sequence `X'` describes a crossing with the north–east and south–west arcs passing “in front” of the other two (the labelling is the same), orientation is inherited from information on other Morse events.

This description reveals quite complex, which is due to the fact that we have locally two sheets of the surface on the two sides of the set F (globally there is a single sheet, due to the nonorientability of the Klein bottle).

Basic properties of the apparent contour can be obtained with

```
$ contour info painted_klein_bottle.morse
This is an apparent contour with labelling

Properties of the embedded surface:
Connected comp.: 1
Total Euler ch.: -2
[...]
```

that tells us that indeed the surface Σ is connected, and has genus 2, i.e., has the topology of a sphere with two handles (Euler–Poincaré characteristic -2), or equivalently a torus with two holes.

The fundamental group of the surface is

```
$ contour fg painted_klein_bottle
Finitely presented group with 4 generators
<a,b,c,d; abADCdB>
```

that, after a simple change of generators, can be readily transformed into the presentation $\langle a, b, c, d; [a, b][c, d] \rangle$ which is the same for all oriented surfaces of genus 2. Here the notation $[a, b]$ denotes the commutator $aba^{-1}b^{-1}$.

The command

```
$ contour --in fg painted_klein_bottle
Finitely presented group with 3 generators
<a,b,c; acACbAB>
```

computes the more interesting fundamental group of the inside E of Σ , that is the same as the fundamental group of the set F . The abelianized of this group is clearly $\mathbb{Z} \times \mathbb{Z}$, however, as shown in [8], at the end of example 1B.13, it is not a free group and in particular is different from the free group with two generators $\mathbb{Z} * \mathbb{Z}$. We can confirm this conclusion by computing the Alexander polynomial (in two indeterminates) of the inside with the command

```
$ contour --in alexander painted_klein_bottle
Computing Alexander ideal for d = 2:
Alexander polynomial:
+1+u-v;
```

which is nontrivial (see Sect. 10.7.3).

Here is the fundamental group of the complement of E :

```
$ contour --out fg painted_klein_bottle
Free group of rank 2
<a,b; >
```

the same as for a solid torus with two holes (a handlebody with two handles). As expected, the corresponding Alexander polynomial is trivial:

```
$ contour --out alexander painted_klein_bottle
Computing Alexander ideal for d = 2:
Trivial whole ring ideal:
1;
```

There is a strong relationship between the example that we are discussing and the surface described (with its knot description) in Fig. 10.6; it turns out that the fundamental group of the inside of one surface is the same as the fundamental group of the outside of the other and the other way around. Indeed the outside fundamental group of the surface of Fig. 10.6 is printed as $\langle a, b, c; abABCAC \rangle$, the same

as that of the inside of Σ after interchanging generators b , and c . The Alexander polynomial turns out to be exactly the same: $1 + u - v$.

Such correspondence suggests that if we “evert” in 3D the set described in 10.6, i.e., we move the point at infinity inside the solid set, we have a shape that can be ambiently deformed into the set E . However equality of the fundamental groups does not prove this equivalence, which is guaranteed if we can find a sequence of Reidemeister-type *moves* that transforms one apparent contour into the other, and this is the case. The set shown in 10.6, left, can be seen to be a deformation retract of the complement of the closed set F , which is another way to assess the stated equivalence.

The command

```
$ contour 3devert 7 embrace.knot
```

can be used to perform the eversion; the number 7 indicates what region has to be used for the surgery, regions from 0 to 6 all have $f_\Sigma = 0$ and cannot be used for the eversion process.

Let us now go back to the “2D” set F with the closed curve γ as critical set (the Klein bottle with the small disk removed). Recall that the complement of F is not simply connected. We can modify F by adding a disk-like *panel* inside the bottle whose boundary consists of a closed curve in F that starts from a point of γ , travels longitudinally along the internal tube and continues on the external part of F until it meets the starting point on γ again. This added panel is shown in gray in Fig. 10.35b. Similarly we can add another panel that prevents the construction of a closed curve around the neck of the bottle.

These two added panels serve exactly the same purpose as the two panels of the house with two rooms of Fig. 10.36.

After this modification the resulting set will be called \hat{F} , it “almost” satisfies the two properties mentioned at the beginning of the section, the first property is now true in a local sense: we cannot remove any small portion of \hat{F} without changing its homotopy type. Moreover its complement is now simply connected. Surprisingly it turns out that \hat{F} can be deformation-retracted to a point, indeed \hat{F} has the same structure as the house with two rooms of Fig. 10.36 and the discussion in [1] and [8] applies.

The apparent contour of the enlarged set obtained from \hat{F} can be computed from the apparent contour of E by applying suitable surgeries. As it happens addition of the first panel can be achieved by successively applying to `painted_klein_bottle.morse` the commands `contour rule B:4` (perform a beak-to-beak) followed by the inverse vertical surgery `contour removehole -r 1`. The resulting surface after the surgery has the topology of a torus (we can confirm this by computing the Euler–Poincaré characteristic with `contour characteristic`). It is known that a surface embedded in S^3 (\mathbb{R}^3 compactified with the addition of the point at infinity) with the topology of a torus borders a solid torus on either side [18]. We check this by asking `appcontour` to compute the fundamental group of the inside and of the outside (“`contour`

--in fg" and "contour --out fg") that in this case both result as $\langle a; \rangle$, the free group of rank 1. This is a hint that we have a solid torus on both sides, both unknotted.¹⁷

Addition of the other panel is more elaborated and we must prepare the apparent contour for the surgery with a sequence of Reidemeister-type moves.

- `contour INVK0:2` (inverse of the `K0` rule, second application possible). This is equivalent to `contour mergearcs -r 5 -a 25:1 -a 19:0` as can be seen from the result of `contour listma` where we explicitly list the region and arcs involved;
- `contour rule C`, followed by `contour rule T`, then `contour rule T:2`, and again `contour rule C`;
- `contour rule CR0L`, one of the composite rules described in Sect. 10.4.3.

We are now in a position to apply as before the vertical surgery `contour removehole -r 1` equivalent to adding the second panel.

The resulting apparent contour has the topology of the sphere, as can be seen with the command `contour info` that asserts that the embedded surface has one connected component (is connected) and has Euler–Poincaré characteristic 2.

As stated by the 3D version of the Schoenflies theorem [2], the embedded surface so obtained must bound a solid sphere, and thus there must exist a sequence of Reidemeister-type moves that transforms the apparent contour into that of a round sphere. As it turns out one such sequence is: `C`, `C`, `CR2`, `B:3` (third possible application of the beak-to-beak rule), `K1b`, `K2`, `CR4Lb`, `S`, `S`, `A2`, for a total of 10 moves.

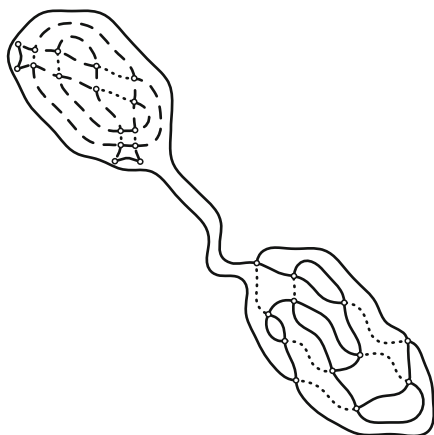
This sequence was automatically found by using the `contour_transform.sh` shell script described in Sect. 10.12, which simply tries out all possible sequences of moves of given maximal length. Such task, however, has a complexity that grows exponentially with the number of moves applied sequentially, and even our relatively simple example took a very long computational time to be completed with the construction of a total of 101,687 distinct apparent contours, obtainable with ten or less moves (simple or composite). We do not know at present a constructive way to produce in polynomial time a simplifying sequence of this sort for any connected surface with the topology of the sphere.

10.16 Example: Mixed Internal/External Knot

The generalized Schoenflies theorem [2] implies that if $\Sigma = \partial E$ is a topological sphere (i.e., Σ is connected and its Euler–Poincaré characteristic is $\chi = 2$), then it bounds a 3D ball. A consequence of this fact is that both E and $\mathbb{R}^3 \setminus E$ have trivial

¹⁷ Actually the *unknotting theorem* in [18, page 103] implies that indeed both the inside and outside are solid tori (in \mathbb{S}^3) and that they are both unknotted.

Fig. 10.38 Apparent contour of the connected sum of a trefoil-knotted solid torus and a three-ball with a knotted hole. The *dotted lines* in the *upper-left* portion are labelled with $d = 3$



fundamental group, which is confirmed by `appcontour`. The simplest possible example of this situation is

```
$ contour --in fg sphere
Trivial group
<>
$ contour --out fg sphere
Trivial group
<>
```

For a surface $\Sigma = \partial E$ with the topology of a torus (i.e., connected and $\chi = 0$) it can be shown [18] that either the internal set E or the external set $\mathbb{R}^3 \setminus E$ (or both) has \mathbb{Z} (free group of rank one) as fundamental group:

```
$ contour --in fg trefoilknot
Free group of rank 1
<a; >
$ contour --out fg internalknot
Free group of rank 1
<a; >
```

Here we recall that `internalknot` contains the Morse description of the apparent contour of a sphere traversed by a knotted tunnel.

We shall now construct a surface $\Sigma = \partial E$ having both internal and external fundamental groups that are non-free [18, Exercise 4, page 108]. Command `contour sum trefoilknot internalknot` constructs the connected sum of the two surfaces just mentioned, resulting in a surface with genus two (Euler–Poincaré characteristic $\chi = -2$), its apparent contour is displayed in Fig. 10.38, we can ask for the fundamental groups of the resulting surface:

```
$ contour sum trefoilknot internalknot | contour --in fg
Finitely presented group with 3 generators
<a,b,c; bcbCC>
$ contour sum trefoilknot internalknot | contour --out fg
Finitely presented group with 3 generators
<a,b,c; bcBcbC>
```

We can ensure that these are not free groups by computing the Alexander polynomial which, for surfaces of genus two, is a polynomial in two indeterminates (say u, v):

```
$ contour sum trefoilknot internalknot | contour alexander --in
Computing Alexander ideal for d = 2:
Alexander polynomial:
+1-u+u^2;
$ contour sum trefoilknot internalknot | contour alexander --out
Computing Alexander ideal for d = 2:
Alexander polynomial:
+1-u+u^2;
```

The fact that each Alexander polynomial ultimately only depends on one of the two indeterminates is not surprising and is related to the particular structure of the surface, also note that $1 - t + t^2$ is the Alexander polynomial of the trefoil knot. The *base-canonicalization* process discussed in Sect. 10.7.4 can be disabled by using the option `--nobasecanonicalify`, for this example we obtain

```
$ contour sum trefoilknot internalknot | contour alexander --in
--nobasecanonicalify
Computing Alexander ideal for d = 2:
# *** Warning: result can be noncanonical ***
Alexander polynomial:
+1-u+u^2;
$ contour sum trefoilknot internalknot | contour alexander --out
--nobasecanonicalify
Computing Alexander ideal for d = 2:
# *** Warning: result can be noncanonical ***
Alexander polynomial:
+1-v+v^2;
```

The difference in the indeterminate name, although intriguing, is just incidental and depends on the specific construction of the cell complex and on the details of the various simplification procedures.

10.17 Using appcontour on Apparent Contours Without Labelling

We conclude this chapter with a few examples of apparent contours in the context of stable maps $\varphi : M \rightarrow \mathbb{R}^2$ (Sect. 2.2) where M is a smooth closed manifold of dimension two. Most of the examples that we gather here were already mentioned previously in this chapter (Boy surface, Sect. 10.9.3; Milnor and Millett curves, Sect. 10.9.3.2) or in previous chapters (Haefliger sphere, Chap. 2, Example 2.2.11).

The apparent contour $\text{appcon}(\varphi)$ alone does not carry, in general, enough information to reconstruct the topological structure of the source manifold M , and indeed we often have topologically distinct possible “realizations” of a map $\varphi : M \rightarrow \mathbb{R}^2$ with the same apparent contour. A natural problem is then to investigate for a given apparent contour (without labelling) what are all possible topologically distinct such realizations. Some of them might admit a factorization through an embedding in \mathbb{R}^3 followed by a projection; these correspond exactly to a choice of a labelling that makes the apparent contour a “complete labelled contour graph” in the sense of Definition 4.2.6. Finding all possible (possibly none) such labellings is a finite combinatorial problem that can be readily solved for not too complex apparent contours.

However, there exist maps $\varphi : M \rightarrow \mathbb{R}^2$ having an apparent contour that does not admit any labelling consistent with Definition 4.2.6, hence providing examples of apparent contours of truly non-embeddable combinations of a 2D manifold and a projection onto the plane; a fascinating example of this situation is illustrated in the next Sect. 10.17.1.

10.17.1 Haefliger Sphere

Figure 2.4 shows a famous example devised by Haefliger in [7]. There cannot be any consistent labelling simply because for any complete labelled contour graph any component of the contour must contain an even number of cusps (see Remark 4.2.7) whereas in this example each of the two components has exactly one cusp. Nevertheless Haefliger shows how to realize this apparent contour with a particular mapping φ from the sphere \mathbb{S}^2 to \mathbb{R}^2 . Of course this mapping cannot be factorized through an embedding in \mathbb{R}^3 and a projection onto \mathbb{R}^2 . Haefliger in fact proves the stronger result that the mapping cannot be factorized through a *smooth immersion* in \mathbb{R}^3 (thus allowing for self-intersections and triple points) followed by a projection onto \mathbb{R}^2 . Although not given in explicit form, in [7] the mapping φ is qualitatively described by first cutting the sphere in three pieces, two polar caps and an equatorial band, using two parallels, the two parallels coinciding with the critical lines, and then describing the behaviour of φ in the three regions.

The `appcontour` program can be used to compute the Euler–Poincaré characteristic of M , resulting to be $\chi = 2$, which, provided we know that M is connected,

leaves the topological sphere \mathbb{S}^2 as the unique candidate for M . Connectedness of M cannot be given as granted; it is indeed true for this particular example, because a single component of the apparent contour (a closed curve with a single protruding cusp) cannot be an apparent contour by itself. This in turn is a consequence of the fact that the function f_φ cannot vanish in the vicinity of a cusp.

10.17.2 Boy Surface

The apparent contour of Fig. 10.25 corresponds to the Boy surface, a well-known smooth immersion of the real projective plane in \mathbb{R}^3 . It has self-intersections in the form of double curves and a single triple point. The `appcontour` program can be used on the apparent contour to obtain the Euler–Poincaré characteristic $\chi = 1$ for M ; moreover, as already remarked in Sect. 10.9.3.1, M is forcibly connected since the apparent contour has a single component, which in turn forces M to be diffeomorphic to the real projective plane. If we want to factorize the map φ through a generic immersion in \mathbb{R}^3 and a subsequent projection on \mathbb{R}^2 , it turns out after a tedious check that the Boy surface is the only possibility up to a diffeomorphism of \mathbb{R}^3 . Since the Boy surface is chiral (it cannot be rotated onto its mirror image), we actually have two distinct (not ambient isotopic) realizations of the Boy surface with opposite orientation.

Figure 10.26 shows the apparent contour of a different choice of a mapping of the real projective plane on \mathbb{R}^2 , and again it can be shown that the given apparent contour comes necessarily from a map of the real projective plane. We again can try to factorize the map through an immersion, however we are now forced to allow for at least two “pinch point” singularities [21], as already noted in Sect. 10.9.3.1, obtaining a “crosscapped disk” [5].

10.17.3 Milnor Curve

Figure 10.27 (left) shows the apparent contour corresponding to an example provided by Milnor (see [12, p. 207] and [17]). The single component of the apparent contour is a curve that bounds a disk in two different ways. If we glue together these two disks along the curve, we obtain a map of the sphere \mathbb{S}^2 onto the plane that cannot be factorized through an embedding and a projection. It is possible however to factorize this map through a regular immersion with no triple points, as shown in Fig. 10.39. The numbers appearing in the figure give the labelling of the corresponding curves, defined according to Definition 3.4.1 applied to the case of immersions and extended in the obvious way to the image of double curves.

It is of course also possible to glue two copies of the same disk (choosing one of the two disks bounded by the Milnor curve) to form a sphere \mathbb{S}^2 that is now

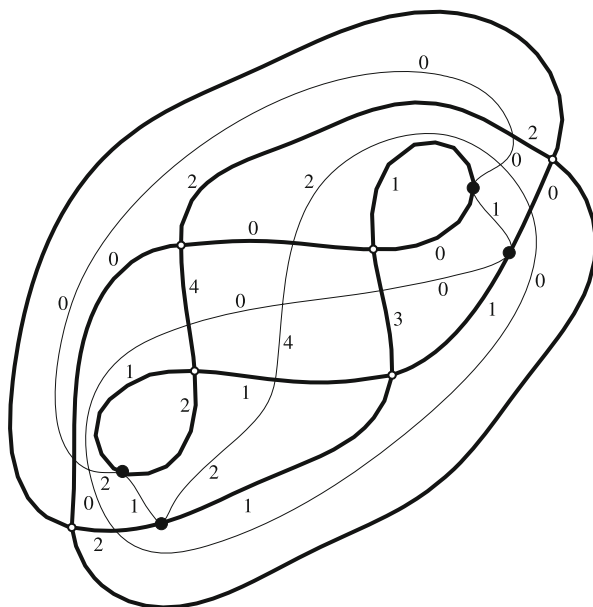


Fig. 10.39 The *thin curves* in the picture correspond to the projection of the self-intersection line of a possible immersion of the sphere \mathbb{S}^2 in \mathbb{R}^3 having the Milnor curve as its apparent contour. There are no pinch points (stable singularities of the immersion), hence the immersion is regular. Moreover there are no triple points

embeddable in \mathbb{R}^3 , thus allowing the construction of a labelling making the apparent contour a *complete labelled contour graph*. Figure 10.40 shows one of the two possibilities, corresponding to the choice of one of the two disks bounded by the curve. We also apparently have the freedom of choosing in which way the sphere occludes itself, the constraints on the labelling leaves us two possibilities, they are however the same up to 180 degrees rotation of the apparent contour. The two possible embeddings, the other one obtained by choosing the other disk bounded by the Milnor curve, are not equivalent in the sense of Definition 3.5.2, however they are mirror images of each other, hence equivalent in the sense of Definition 2.1.1.

It turns out that there are no other structurally different realizations of the Milnor curve as apparent contour.

10.17.4 Millett curve

Figure 10.27 right shows an apparent contour that differs from the apparent contour of the Milnor curve of the previous example only for the presence of a single cusp. It was provided by Millett [12] as the simplest possible example (in the appropriate sense) having the projective plane as source manifold. Indeed the appcontour

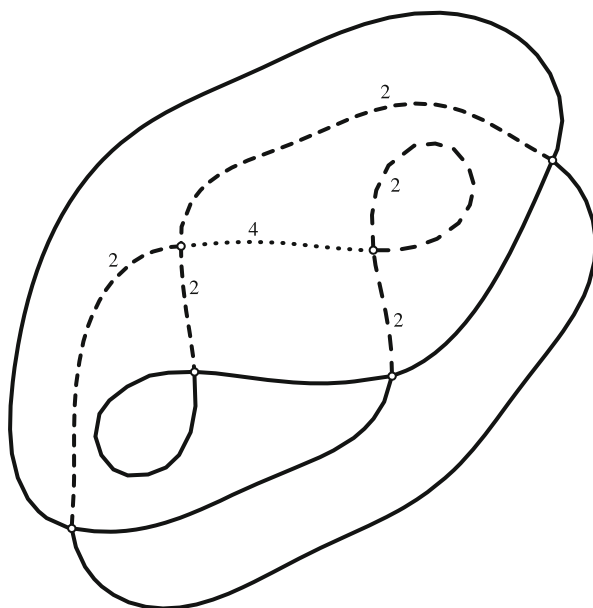


Fig. 10.40 The Milnor curve can also be the apparent contour of an embedding of the sphere. This can be done by enriching the apparent contour with a labelling that makes it a *complete labelled contour graph*. Solid lines have labelling 0, dashed lines have labelling 2, dotted lines have labelling 4

program computes the Euler–Poincaré characteristic of the unlabelled apparent contour as $\chi = 1$. Non-orientability implies that the map cannot be factorized through an embedding in \mathbb{R}^3 , however we can factorize it (in more than one way) through an immersion. Fig. 10.41 shows one possibility where the immersion is stable (in the sense of 2.1.2) but exhibits two pinch point singularities. The two singularities can be removed by sliding them along the contour onto each other at the expense of creating a triple point.

10.17.5 Klein bottle

The apparent contour of Fig. 10.43 (left) is generally recognized as the standard picture of the well-known Klein bottle; imagination is facilitated by adding the double curve of the usual immersion of the Klein bottle in \mathbb{R}^3 , as shown in Fig. 10.43 (right), where the *neck* of the bottle crosses the lateral wall.

This is a non-orientable surface having Euler–Poincaré characteristic $\chi = 0$, as confirmed by the appcontour program:

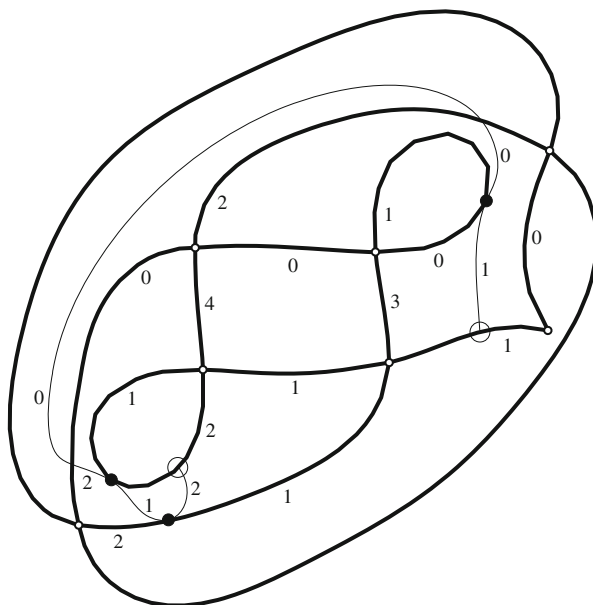


Fig. 10.41 The thin curves in the picture correspond to the projection of the self-intersection line of a possible immersion of the projective plane in \mathbb{R}^3 having the Millett curve as its apparent contour. There are two pinch points (stable singularities of the immersion) shown as empty circles, that can be removed at the expense of creating a triple point

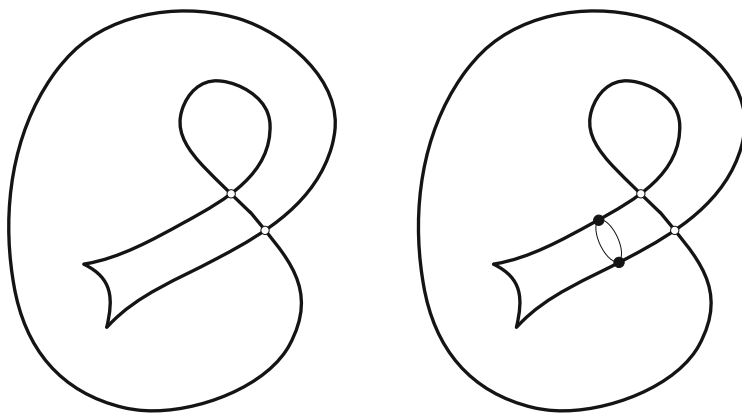


Fig. 10.42 The unlabelled apparent contour on the left corresponds to a possible map from the Klein bottle to the plane. This map can be factorized through an immersion in \mathbb{R}^3 (an embedding is of course not possible since the Klein bottle is non-orientable). On the right the thin line corresponds to the self-intersection of such an immersion where the neck of the bottle enters through the lateral wall

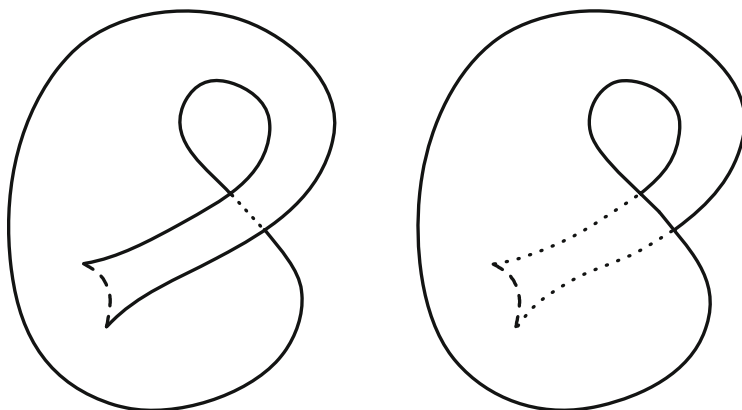


Fig. 10.43 The apparent contour of Fig. 10.42 can also be obtained from an embedded torus, with two possible labellings

```
$ contour characteristic kleinbottle
Euler characteristic: 0
```

where the file `kleinbottle.morse` contains an unlabelled Morse description of the apparent contour in the left of Fig. 10.42.

This is however an intriguing example, since $\chi = 0$ is also the Euler–Poincaré characteristic of a torus, and indeed it is possible to adjoin a consistent labelling to the given apparent contour obtaining as a result a non-obvious embedding of the torus in \mathbb{R}^3 , see Fig. 10.43.

There are actually two different choices of the labelling, related through a `contour frontback` reflection. We leave the actual reconstruction to the reader as a simple exercise.

10.A Appendix: Practical Canonization of Laurent Polynomials

Given a Laurent polynomial $p = p(u, v) \in L$ with integer coefficients we need a practical way to canonically select a representative of the equivalence class $[p]$ with respect to the equivalence relation of Definition 7.7.8.

It is useful to interpret a Laurent polynomial as a function (the *coefficient function*) $\mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{Z}$ giving the coefficient of the monomial having given exponents of the unknowns u and v . Clearly, Laurent polynomials correspond to those functions having finite support (or equivalently *bounded* support, since $\mathbb{Z} \times \mathbb{Z}$ is a discrete set) $S = \text{Supp}(p)$. If we neglect the possible change of sign, the notion of base-equivalency is then interpreted as the composition of the coefficient function

with an invertible affine transformation of $\mathbb{Z} \times \mathbb{Z}$, and we are led to the problem of finding a canonical representation of a bounded subset of $\mathbb{Z} \times \mathbb{Z}$, the support of the coefficient function, up to such affine transformations. Canonification of the support $S = \text{Supp}(p)$ is not equivalent to canonifying the Laurent polynomial p ; however, it allows to restrict the canonization procedure to only those few polynomials equivalent to p and having support equal to the canonified support. Selecting the canonical polynomial among these few polynomials can be done using any lexicographic ordering.

Similarly, a possible change of sign can be taken into account by separately canonifying both p and $-p$ and then selecting an optimal one by lexicographic comparison.

If the support is empty or consists of a single point, the canonization procedure is trivial and leads to a nonnegative constant canonical polynomial; hence, we shall suppose that p contains at least two nonzero distinct terms. We distinguish two cases:

- the support S is at most one-dimensional, i.e., it consists of a finite number of *aligned* points in $\mathbb{Z} \times \mathbb{Z}$;
- the support S is two-dimensional, i.e., there exist three points $P_A, P_B, P_C \in S$ that are vertices of a nondegenerate triangle.

10.A.1 One-Dimensional Support

Let P_A and $P_B \in \mathbb{Z} \times \mathbb{Z}$ be the two extremal points of its support. A straightforward computation shows that there exists an affine invertible transformation that takes P_A and P_B onto the origin and the point $(d, 0)$ respectively, with d the greatest common divisor (GCD) between the coordinates of $P_B - P_A$. We have actually two choices, upon exchange of P_A and P_B , and we can select one of these using a lexicographic type of comparison between the two corresponding polynomials. In the end we obtain a polynomial in only one variable u .

10.A.2 Two-Dimensional Support

Let $p = p(u, v) \in L$; we first observe that a suitable “translation”, i.e., multiplication by a monomial with unit coefficient, allows to transform it into a polynomial with no negative exponent and such that its support intersects both the “ u ” axis and the “ v ” axis. We shall call this polynomial the “stem” of p and denote it by $\text{stem}(p)$. Equivalently $\text{stem}(p)(u, v) = p(u, v)u^\alpha v^\beta$ where $\alpha, \beta \in \mathbb{Z}$ are the smallest possible values leading to a polynomial with no negative exponent. Two polynomials have the same stem if and only if they differ by multiplication of a unit monomial and in particular they are base-equivalent; we can thus reduce the

canonification problem to a selection process of a canonical representative among base-equivalent stems.

The canonification procedure presently implemented selects the candidate among the set of those base-equivalent stems that minimize their total degree (the maximal sum of the two exponents of all terms). By extension, from now on we shall call “total degree” of a Laurent polynomial the total degree of its stem. Let $\delta \in \mathbb{N}$ denote this minimal total degree.

If $M \geq \delta$ is some estimate of the minimal total degree, we want to find a corresponding estimate on the elements of the change of base matrix B in order to exclude changes of base leading to a polynomial p_B with total degree larger than M . A starting value for M is of course the total degree of p ; then during the process of selection of the canonical representative we shall decrease the bound M whenever we find a base-equivalent polynomial with lower total degree.

We need to find a bound on the elements of a change of base matrix B that is guaranteed to be satisfied whenever p_B has total degree not larger than M . The software will then scan all matrices satisfying such bound in search of the optimal representative among those having minimal total degree.

We proceed as follows. Let $[\alpha_i, \beta_i], i = 1, 2, 3$ denote three points in the support of p , vertices of a nondegenerate triangle and listed in a counterclockwise order. Define the (3×3) matrix

$$A = \begin{bmatrix} \alpha_1 & \beta_1 & 1 \\ \alpha_2 & \beta_2 & 1 \\ \alpha_3 & \beta_3 & 1 \end{bmatrix}, \quad (10.1)$$

whose determinant δ turns out to be twice the area of the triangle.

The three vectors $\hat{\xi}_i \in \mathbb{Z}^3, i = 1, 2, 3$ are defined as

$$\begin{aligned} \hat{\xi}_1 &= [\beta_2 - \beta_3, \alpha_3 - \alpha_2, \alpha_2\beta_3 - \alpha_3\beta_2]^T \\ \hat{\xi}_2 &= [\beta_3 - \beta_1, \alpha_1 - \alpha_3, \alpha_3\beta_1 - \alpha_1\beta_3]^T \\ \hat{\xi}_3 &= [\beta_1 - \beta_2, \alpha_2 - \alpha_1, \alpha_1\beta_2 - \alpha_2\beta_1]^T. \end{aligned}$$

A direct computation shows that they satisfy

$$A\hat{\xi}_i = \delta e_i, \quad i = 1, 2, 3 \quad (10.2)$$

where $e_i, i = 1, 2, 3$ denote the standard basis of \mathbb{Z}^3 . We define the rombohedron $\hat{K} \subset \mathbb{R}^3$ as

$$\hat{K} := \left\{ \sum_{i=1}^3 \lambda_i \hat{\xi}_i : 0 \leq \lambda_i \leq 1, i = 1, 2, 3 \right\}$$

and set

$$K := \left(\frac{M}{\delta} \right) \hat{K} \cap \mathbb{Z}^3.$$

Finally, we define K^e as the smallest subset of \mathbb{Z}^3 of the form $K^e = ([a_{\min}, a_{\max}] \times [b_{\min}, b_{\max}] \times \mathbb{R}) \cap \mathbb{Z}^3$ such that $K \subset K^e$.

Proposition 10.1 *If B is such that p_B has total degree not larger than M , then the elements of $B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ necessarily satisfy the bounds*

$$a_{\min} \leq a, c \leq a_{\max} \quad \text{and} \quad b_{\min} \leq b, d \leq b_{\max}.$$

Proof There exists a translation $t = [e, f]^T \in \mathbb{Z}^2$ such that the polynomial $p_B u^e v^f$ is the stem of p_B and has total degree not larger than M and in particular all points in its support are contained in the square $[0, M] \times [0, M]$. Among these points we find the result of the affine transformation $z \rightarrow Bz + t$ applied to the three points $[\alpha_i, \beta_i]$. They turn out to be the three columns of the matrix

$$\begin{bmatrix} a & b & e \\ c & d & f \end{bmatrix} A^T,$$

where A is the matrix defined in (10.1). The constraint on the total degree of p_B entails in particular that each component of each of the three vectors is contained in the interval $[0, M]$. This means that there are six values λ_i and μ_i , $i = 1, 2, 3$ such that $0 \leq \lambda_i, \mu_i \leq 1$, $i = 1, 2, 3$ and

$$A\xi = M \sum_{i=1}^3 \lambda_i e_i, \quad A\eta = M \sum_{i=1}^3 \mu_i e_i,$$

where $\xi = [a, b, e]^T$ and $\eta = [c, d, f]^T$. Using (10.2) this is equivalent to

$$\xi = \frac{M}{\delta} \sum_{i=1}^3 \lambda_i \hat{\xi}_i, \quad \eta = \frac{M}{\delta} \sum_{i=1}^3 \mu_i \hat{\xi}_i \quad \implies \quad \xi, \eta \in K$$

and we can conclude recalling that $K \subset K^e$. □

The algorithm that can be constructed using this bound, actually implemented in the software, is not very efficient, although it seems to work well for scenes that are not too complex. It is possible to achieve a much lower computational complexity by using the strategy described in [14]. However that procedure will produce a canonical representative that does not minimize, in general, the total degree.

References

1. Bing, R.H.: Some aspects of the topology of 3-manifolds related to the Poincaré Conjecture. In: Saaty, T.L. (ed.) *Lectures on Modern Mathematics II*, pp. 93–128. Wiley, Boston (1964)
2. Brown, M.: A proof of the generalized Schoenflies theorem. *Bull. Amer. Math. Soc.* **66**, 74–76 (1960)
3. Cromwell, P.: *Knots and Links*. Cambridge University Press, Cambridge (2004)
4. Datta, S., Limaye, N., Nimbhorkar, P., Thierauf, T., Wagner, F.: Planar Graph Isomorphism is in Log-Space. 24th Annual IEEE Conference on Computational Complexity (2009)
5. Francis, G.k, Weeks, J.R.: Conway's ZIP proof. *Am. Math. Mon.* **106**, 393–399 (1999)
6. Hacon, D., Mendes de Jesus, C., Romero Fuster, M.C.: Global topological invariants of stable maps from a surface to the plane. *Proceedings of the 6th Workshop on Real and Complex Singularities*, 2001. *Lecture Notes in Pure Applied Mathematics*, vol. 232, pp. 227–235 (2003)
7. Haefliger, A.: Quelques remarques sur les applications différentiables d'une surface dans le plan. *Ann. Inst. Fourier. Grenoble* **10**, 47–60 (1960)
8. Hatcher, A.: *Algebraic Topology* online book. Cambridge University Press, Cambridge (2002)
9. Köbler, J., Schöning, U., Torán, J.: *The Graph Isomorphism Problem: Its Structural Complexity*. Birkhäuser Verlag, Berlin (1993)
10. Lickorish, W.B.R.: *An Introduction to Knot Theory*. Springer, New York (1997)
11. Magnus, W., Karrass, A., Solitar, D.: *Combinatorial Group Theory: Presentations of Groups in Terms of Generators and Relations*. Dover Publications, New York (1976)
12. Millett, K.C.: Generic smooth maps of surfaces. *Topology Appl.* **18**, 197–215 (1984)
13. Ohmoto, T., Aicardi, F.: First order local invariants of apparent contours. *Topology* **45**, 27–45 (2006)
14. Paolini, G.: An algorithm for a canonical form of finite subsets of \mathbb{Z}^d up to affinities, arXiv:1408.3310 [cs.DS]
15. Paolini, M., Pasquarelli, F.: Appcontour: a software code to interact with apparent contours, SourceForge project: <http://appcontour.sourceforge.net> (2006)
16. Pauer, F., Unterkircher, A.: Gröbner bases for ideals in Laurent polynomial rings and their application to systems of difference equations. *Appl. Algebra Eng. Commun. Comput.* **9**, 271–291 (1999)
17. Poenaru, V.: Extension des Immersions en Codimension 1 (d'après Samuel Blank). *Séminaire Bourbaki* **342**, 473–505 (1967/68)
18. Rolfsen, D.: *Knots and Links*, AMS Chelsea Publishing, New York (2003)
19. Vassiliev, V.A.: Cohomology of knot spaces. *Adv. Sov. Math.* **21**, 23–69 (1990)
20. Weinberg, L.: A simple and efficient algorithm for determining isomorphism of planar triply connected graphs. *Circuit Theory* **13**, 142–148 (1966)
21. Whitney, H.: On singularities of mappings of Euclidean spaces. I. Mappings of the plane into the plane. *Ann. Math.* **62**, 374–410 (1955)

Chapter 11

Variational Analysis of the Model on Labelled Graphs

In this chapter, essentially following [2],¹ we discuss some coerciveness and semicontinuity properties of the functional \mathcal{F} introduced in Sect. 1.5 and motivating our study of apparent contours and three-dimensional shapes.

Notation As in Chap. 1, $g \in L^\infty(\Omega, [0, 1])$ is a given function, which stands for the grey-level intensity of an image. Given a bounded open connected subset I of \mathbb{R} or of the unit circle \mathbb{S}^1 , and²

$$p > 1, \quad (11.1)$$

we denote by $W^{1,p}(I)$ (respectively $W^{2,p}(I)$) the Sobolev space consisting of all functions of $L^p(I)$ having first (respectively first and second) distributional derivative in $L^p(I)$. Recall³ that $W^{1,p}(I)$ embeds compactly in $C^0(\bar{I})$, and even in the space of $(1 - 1/p)$ -Hölder functions in I .

Definition 11.0.2 (Curve of Class $W^{2,p}$) We say that a continuous plane curve is in $W^{2,p}$ if its first and second distributional derivatives, with respect to an arc length parameter, belong to L^p .

Depending on the situation, a continuous closed plane curve shall be considered equivalently either as a map $\gamma : \mathbb{S}^1 \rightarrow \bar{\Omega}$ (a map $\gamma : [0, \mathcal{L}] \rightarrow \bar{\Omega}$ with 0 and $\mathcal{L} > 0$

¹With kind permission from Springer Science+Business Media, in this chapter we report some of the results and figures from the quoted paper [2].

²We recall that the model described in Sect. 1.5 requires $p \in (1, 2)$. The results of the present chapter hold under the less stringent assumption (11.1).

³See, e.g., [4, Theorem 8.8] and [5, Corollary 8.31].

identified) or as a map defined on the one-dimensional torus $\mathbb{R}/(\mathcal{L}\mathbb{Z})$ (namely, as an \mathcal{L} -periodic map).⁴ We denote by

$$\text{im}(\gamma) := \{\gamma(t) : t \in [0, \mathcal{L}]\} = \gamma([0, \mathcal{L}])$$

the image of γ .

Given a closed set $C \subset \Omega := (0, 1) \times (0, 1)$, we let $H^1(\Omega \setminus C)$ be the space of functions in $L^2(\Omega \setminus C)$ having square integrable distributional gradient on the open set $\Omega \setminus C$ (see [4]). Given $u \in H^1(\Omega \setminus C)$, we set

$$\mathcal{M}(u, \Omega \setminus C) := \int_{\Omega} (u - g)^2 dx + \alpha \int_{\Omega \setminus C} |\nabla u|^2 dx, \quad (11.2)$$

where α is a given positive parameter. The functional \mathcal{M} is one of the addenda of the functional \mathcal{F} , defined in the next section.

Following some notation given in previous chapters, we denote by $BV(\Omega, 2\mathbb{N})$ the space of functions with bounded variation in Ω taking values in the nonnegative even natural numbers. If $f \in BV(\Omega, 2\mathbb{N})$, we denote by J_f the jump set of f and by f^\pm the two traces of f on J_f . We refer the reader to [1] for the theory of BV functions.

11.1 The Action Functional

Our first task is to give a rigorous definition of the domain of the functional \mathcal{F} . We start by defining the class of plane graphs on which the functional \mathcal{F} is naturally defined.

11.1.1 Graphs with Cusps and Curvature in L^p

Since one of the terms in the expression of \mathcal{F} involves the L^p norm of the curvature of a graph, we need to slightly generalize Definitions 4.2.2 and 4.2.6: the generalization concerns only the regularity of the graph and the regularity at a junction between two arcs.

Definition 11.1.1 (Complete Contour Graph of Class $\mathcal{W}_{\text{graph}}^{2,p}$) Let G be a contour graph.⁵ We say that G is a complete contour graph of class $\mathcal{W}_{\text{graph}}^{2,p}$ if conditions (G2), (G3) of Definition 4.2.2 hold, and if in addition:

⁴Accordingly, a subinterval of $[0, \mathcal{L}]$ corresponds to a connected subset of $\mathbb{R}/(\mathcal{L}\mathbb{Z})$; we shall adopt this convention, for instance, when integrating functions on such subintervals.

⁵See Definition 4.1.4.

- (G1') the arcs of G are of class $W^{2,p}$;
- (G4') at a crossing the opposite arcs join in a $W^{2,p}$ way⁶ and the two pairs of opposite arcs cross transversally;
- (G5') at a cusp the two tangent vectors have the same direction and opposite orientations. Moreover, the orientation is so that when travelling along a cusp, the cusp lies on the left during the 180 degrees turn.

Notice that a complete contour graph is a complete contour graph of class $\mathcal{W}_{\text{graph}}^{2,p}$.

Remark 11.1.2 A component of a complete contour graph of class $\mathcal{W}_{\text{graph}}^{2,p}$ having some cusp is not in $W^{2,p}$, since the tangent vector jumps at a cusp, and therefore its distributional derivative with respect to an arc length parameter does not belong to L^p .

Let G be a complete contour graph of class $\mathcal{W}_{\text{graph}}^{2,p}$ and let $f : \mathbb{R}^2 \setminus G \rightarrow 2\mathbb{N}$ be twice the winding number of G (see Definition 4.2.1, and recall condition (G3) of Definition 4.2.2). A labelling of G is a function

$$d : G \setminus \text{nodes}(G) \rightarrow \mathbb{N}$$

satisfying properties (L1), (L2) and (L3) of Definition 4.2.5.

Definition 11.1.3 (Complete Labelled Contour Graph of Class $\mathcal{W}_{\text{graph}}^{2,p}$) A complete contour graph of class $\mathcal{W}_{\text{graph}}^{2,p}$ endowed with a labelling is called a complete labelled contour graph of class $\mathcal{W}_{\text{graph}}^{2,p}$.

Following Definition 3.6.1, the closure of the set $\{d = 0\}$ is called the *visible* part of J_f , and it is denoted by $\text{vis}(J_f)$.

11.1.2 The Functional

We start with the definition of the domain of the action functional.

Definition 11.1.4 (Domain of \mathcal{F}) We denote by $\text{Dom}(\mathcal{F})$ the set of all triplets (f, d, u) such that

- (J_f, d) is a complete labelled contour graph of class $\mathcal{W}_{\text{graph}}^{2,p}$ contained in Ω ,
- $f(x) = 2w(x, J_f)$ for any $x \in \Omega \setminus J_f$,
- $u \in H^1(\Omega \setminus \text{vis}(J_f))$.

⁶We could equivalently require that the opposite arcs join in a C^1 way: indeed, if $0 < \mathcal{L}_1 < \mathcal{L}_2$, and $\gamma \in W^{2,p}((0, \mathcal{L}_1), \mathbb{R}^2) \cap W^{2,p}((\mathcal{L}_1, \mathcal{L}_2), \mathbb{R}^2) \cap C^1((0, \mathcal{L}_2), \mathbb{R}^2)$, an integration by parts shows that $\gamma \in W^{2,p}((0, \mathcal{L}_2), \mathbb{R}^2)$.

If $(f, d, u) \in \text{Dom}(\mathcal{F})$, then $f \in BV(\Omega, 2\mathbb{N})$ and it vanishes near $\partial\Omega$. In addition,

$$|f^+(x) - f^-(x)| = 2, \quad x \in J_f \setminus \text{nodes}(J_f), \quad (11.3)$$

Given a function $f \in BV(\Omega, 2\mathbb{N})$ whose jump J_f is a complete contour graph of class $\mathcal{W}_{\text{graph}}^{2,p}$ contained in Ω , we set

$$\mathcal{W}(f) := \int_{J_f \setminus \text{nodes}(J_f)} \frac{|f^+ - f^-|}{2} (\beta + \phi_p(\kappa)) d\mathcal{H}^1,$$

where $\kappa \in L^p(J_f \setminus \text{nodes}(J_f))$ is the \mathcal{H}^1 -almost everywhere defined curvature (a real number) of $J_f \setminus \text{nodes}(J_f)$, $\phi_p(\xi) := \lambda|\xi|^p$ for any $\xi \in \mathbb{R}$, and β, λ are positive parameters.

Notice that if f satisfies the properties listed in Definition 11.1.4, then $\mathcal{W}(f)$ has the simpler expression

$$\mathcal{W}(f) = \int_{J_f \setminus \text{nodes}(J_f)} (\beta + \lambda|\kappa|^p) d\mathcal{H}^1. \quad (11.4)$$

Making use of (11.2) and (11.4), we are now in a position to define the action functional, whose minimization describes our variational model (see Sect. 1.5⁷). Let σ be a positive parameter.

Definition 11.1.5 (The Functional \mathcal{F}) The functional $\mathcal{F} : \text{Dom}(\mathcal{F}) \rightarrow [0, +\infty)$ is defined as follows:

$$\mathcal{F}(f, d, u) := \mathcal{M}(u, \Omega \setminus \text{vis}(J_f)) + \mathcal{W}(f) + \sigma \# \text{nodes}(J_f), \quad (f, d, u) \in \text{Dom}(\mathcal{F}).$$

11.1.3 A Notion of Convergence

In the study of the coercivity properties of \mathcal{F} it is necessary to inspect the compactness properties of a sequence $((f_n, d_n, u_n)) \subset \text{Dom}(\mathcal{F})$ satisfying the uniform bound

$$\sup_{n \in \mathbb{N}} \mathcal{F}(f_n, d_n, u_n) < +\infty. \quad (11.5)$$

⁷Referring to the final discussion in Sect. 1.5, suppose that the infimization of \mathcal{F} has a solution; then condition (G5) in Definition 4.2.2 is not necessarily satisfied. As a consequence, if we adapt the proofs of Theorems 5.1.1 and 5.1.4 to this case, the reconstructed three-dimensional scene E is not necessarily of class \mathcal{C}^∞ anymore.

This inspection suggests a suitable notion of convergence on $\text{Dom}(\mathcal{F})$ which makes \mathcal{F} lower semicontinuous. The characterization of the closure of the space of sequences satisfying (11.5) is, on the other hand, a problem related to the relaxation of \mathcal{F} , and this study is not carried on here in its full generality; in Sect. 11.3, we shall discuss some related issues.

Definition 11.1.6 (Convergence on $\text{Dom}(\mathcal{F})$) Let $(f_n, d_n, u_n) \in \text{Dom}(\mathcal{F})$ and $(f, d, u) \in \text{Dom}(\mathcal{F})$. We say that the sequence $((f_n, d_n, u_n))$ converges to (f, d, u) if

- $\lim_{n \rightarrow +\infty} f_n = f$ in $L^1(\Omega)$,
- $\lim_{n \rightarrow +\infty} \text{vis}(J_{f_n}) = \text{vis}(J_f)$ in the sense of Kuratowski,⁸
- $\lim_{n \rightarrow +\infty} u_n = u$ in $L^1(\Omega)$.

Note that, concerning the sequence (d_n) , only the convergence of the visible part of J_{f_n} to the visible part of J_f is required. It would be natural to require some convergence on (d_n) , however the actual value of \mathcal{F} would be unaffected, and we prefer to impose as few restrictions as possible at this stage.

11.2 Lower Semicontinuity

We start to prove the sequential lower semicontinuity of the functional \mathcal{M} introduced in (11.2).

Lemma 11.2.1 (Lower Semicontinuity of \mathcal{M}) Let $((f_n, d_n, u_n)) \subset \text{Dom}(\mathcal{F})$ be a sequence converging to $(f, d, u) \in \text{Dom}(\mathcal{F})$. Then

$$\mathcal{M}(u, \Omega \setminus \text{vis}(J_f)) \leq \liminf_{n \rightarrow +\infty} \mathcal{M}(u_n, \Omega \setminus \text{vis}(J_{f_n})). \quad (11.6)$$

Proof Since all J_{f_n} are contained in Ω , it is possible to prove⁹ that $\text{vis}(J_{f_n})$ converges to $\text{vis}(J_f)$ in the Hausdorff distance (see [1, Section 6.1]).

⁸A sequence (K_n) of compact subsets of the plane converges to K in the sense of Kuratowski, if the following two conditions hold:

- any $x \in K$ is the limit of a sequence (x_n) with $x_n \in K_n$ for any $n \in \mathbb{N}$,
- if $x_n \in K_n$ for any $n \in \mathbb{N}$, then any limit point of (x_n) belongs to K .

See [8, Chapter 2, par. 20, Section VI] and [6, Definition 4.10] for more information.

⁹See also [1, Theorem 6.1].

Therefore, for any $\varepsilon > 0$ we can take an ε -tubular neighbourhood N_ε of $\{d = 0\}$ so that, for n sufficiently large, the closure of $\{d_n = 0\}$ is contained in N_ε . From the lower semicontinuity of the Dirichlet integral functional in $\Omega \setminus N_\varepsilon$, we have

$$\mathcal{M}(u, \Omega \setminus N_\varepsilon) \leq \liminf_{n \rightarrow \infty} \mathcal{M}(u_n, \Omega \setminus N_\varepsilon) \leq \liminf_{n \rightarrow \infty} \mathcal{M}(u_n, \Omega \setminus \text{vis}(J_{f_n})).$$

Letting $\varepsilon \rightarrow 0^+$ we then get (11.6). \square

Now, let us show that the functional \mathcal{F} is lower semicontinuous on its domain.

Theorem 11.2.2 (Sequential Lower Semicontinuity of \mathcal{F}) *Let $((f_n, d_n, u_n)) \subset \text{Dom}(\mathcal{F})$ be a sequence converging to $(f, d, u) \in \text{Dom}(\mathcal{F})$. Then*

$$\mathcal{F}(f, d, u) \leq \liminf_{n \rightarrow +\infty} \mathcal{F}(f_n, d_n, u_n). \quad (11.7)$$

Proof We can suppose that the right-hand side of (11.7) is finite, since when it is $+\infty$ the thesis is trivially satisfied. Take a (not relabelled) subsequence $((f_n, d_n, u_n))$ in such a way that the right-hand side of (11.7) is a limit, and also that the three sequences $(\mathcal{M}(u_n, \Omega \setminus \text{vis}(J_{f_n})))$, $(\mathcal{W}(f_n))$ and $(\# \text{nodes}(J_{f_n}))$ admit a limit.

Let

$$C := \sup_{n \in \mathbb{N}} \mathcal{F}(f_n, d_n, u_n) < +\infty. \quad (11.8)$$

From (11.6) it follows that (11.7) is proven once we show that

$$\mathcal{W}(f) \leq \lim_{n \rightarrow +\infty} \mathcal{W}(f_n) \quad (11.9)$$

and

$$\# \text{nodes}(J_f) \leq \lim_{n \rightarrow +\infty} \# \text{nodes}(J_{f_n}). \quad (11.10)$$

To show inequality (11.9), it is convenient to parametrize the graphs J_{f_n} . Given $n \in \mathbb{N}$, using Definition 11.A.6, we consider an oriented parametrization

$$\hat{\Gamma}_n = \{\hat{\gamma}_1^n, \dots, \hat{\gamma}_m^n\} \quad (11.11)$$

of J_{f_n} . The bound

$$\sup_{n \in \mathbb{N}} \left(\int_{J_{f_n} \setminus \text{nodes}(J_{f_n})} (\beta + \lambda |\kappa|^p) d\mathcal{H}^1 + \sigma \# \text{nodes}(J_{f_n}) \right) \leq C, \quad (11.12)$$

which follows from (11.8), implies, together with Theorem 11.B.6 that, possibly passing to a not relabelled subsequence, we can suppose that:

- m is independent of n , the number of cusps of $J_{f_n} = \text{im}(\hat{\Gamma}_n)$ is independent of n , and the number of crossings of J_{f_n} is independent of n ,
- there is a (possibly empty) proper subset \mathcal{I} of $\{1, \dots, m\}$, independent of n , such that

$$i \in \mathcal{I} \iff \lim_{n \rightarrow +\infty} \ell(\hat{\gamma}_i^n) = 0;$$

note that, in view of Lemmas 3.4.6 and 11.B.1, if $i \in \mathcal{I}$ then $\text{im}(\hat{\gamma}_i^n)$ has at least two cusps.

We now focus our attention on the components of J_{f_n} having length that does not go to zero as $n \rightarrow +\infty$. Since we want to take the limits of these components, it is preferable to have a common parameter space independent of n .

Set

$$h := m - \#\mathcal{I}.$$

If $h = 0$, we have $f = 0$ in Ω , hence $\mathcal{F}(f, d, u) = \mathcal{M}(u, \Omega)$ and (11.7) follows from Lemma 11.2.1. Hence, we can suppose $h \geq 1$.

Up to renumbering the curves, we can assume that $\hat{\gamma}_1^n, \dots, \hat{\gamma}_h^n$ are the curves of $\hat{\Gamma}_n$ having length that does not tend to zero as $n \rightarrow +\infty$; therefore, there are two real constants ℓ_1, ℓ_2 such that

$$0 < \ell_1 \leq \ell(\hat{\gamma}_i^n) \leq \ell_2, \quad n \in \mathbb{N}, i \in \{1, \dots, h\}. \quad (11.13)$$

Since h is independent of n , from (11.13) it follows that we can regularly parametrize $\hat{\gamma}_i^n$ on $[0, 1]$ (0 and 1 identified) for any $n \in \mathbb{N}$. For any $i \in \{1, \dots, h\}$ let us denote by $\gamma_i^n \in \text{pwr}_c^p([0, 1], \Omega)$ (Definition 11.A.1) such a reparametrized curve, and set

$$\Gamma_n := \{\gamma_1^n, \dots, \gamma_h^n\},$$

which is a system of $\text{pwr}_c^p([0, 1], \Omega)$ curves, defined on the disjoint union of h copies of $[0, 1]$. From Theorem 11.B.6 it follows that, possibly passing to a not relabelled subsequence, there exist h curves $\gamma_1, \dots, \gamma_h$ of class $\text{pwr}_c^p([0, 1], \overline{\Omega})$, such that (Γ_n) converges weakly in pwr_c^p to

$$\Gamma := \{\gamma_1, \dots, \gamma_h\}.$$

Let us prove that

$$\text{im}(\Gamma) \supseteq J_f. \quad (11.14)$$

Since the number of curves composing each system Γ_n does not depend on n , remembering that f_n is twice the winding number of Γ_n , we have¹⁰

$$\sup_{n \in \mathbb{N}} \|f_n\|_{L^\infty(\Omega)} < +\infty. \quad (11.15)$$

From (11.12) it follows $\mathcal{H}^1(J_{f_n}) \leq C$ hence, using (11.15) we get

$$\sup_{n \in \mathbb{N}} \int_{\Omega} |Df_n| < +\infty, \quad (11.16)$$

where $\int_{\Omega} |Df_n|$ denotes the total variation of f_n in Ω .

Let now $x \in J_f$, $\rho > 0$, and denote by $B_\rho(x) \subset \mathbb{R}^2$ the open ball centred at x with radius ρ . From the bound (11.16) it follows (possibly passing to a subsequence) that the sequence $(|Df_n|)$ weakly converges in the sense of measures to a Radon measure λ , which turns out¹¹ to be supported on $\text{im}(\Gamma)$. Hence, for any $\rho > 0$ such that $\lambda(\partial B_\rho(x)) = 0$ (in particular, for almost every $\rho > 0$) we have¹²

$$\lambda(B_\rho(x)) = \lim_{n \rightarrow +\infty} |Df_n|(B_\rho(x)) \geq |Df|(B_\rho(x)),$$

where the last inequality follows from the L^1 -lower semicontinuity of the total variation in $B_\rho(x)$. Since $|Df|(B_\rho(x)) = \int_{J_f \cap B_\rho(x)} |f^+ - f^-| d\mathcal{H}^1 \geq 2\rho$, it follows that x belongs to the support of λ , hence $x \in \text{im}(\Gamma)$, and (11.14) is proven.

Now, let $i \in \{1, \dots, h\}$; since (γ_i^n) weakly converges to γ_i in pwr_c^p , we have $\ell(\gamma_i) = \lim_{n \rightarrow +\infty} \ell(\gamma_i^n)$, hence

$$\ell(\Gamma) = \lim_{n \rightarrow +\infty} \ell(\Gamma_n) = \lim_{n \rightarrow +\infty} \ell(J_{f_n}).$$

Remembering inclusion (11.14), we deduce

$$\mathcal{H}^1(J_f) \leq \ell(\Gamma) = \lim_{n \rightarrow +\infty} \ell(J_{f_n}). \quad (11.17)$$

We observe now that, if $x \in \text{crossings}(J_f)$ then, for any $n \in \mathbb{N}$, there exists $x_n \in \text{crossings}(J_{f_n})$ such that $\lim_{n \rightarrow +\infty} x_n = x$. Indeed, if by contradiction x has a neighbourhood U where there are no crossings of J_{f_n} for n sufficiently large, by the

¹⁰The bound (11.15) can be proven with an inequality similar to that used in (11.37) below, which implies that a complete turn without cusps around a point has a fixed cost in terms of the action. As a consequence, an unbounded number (as $n \rightarrow +\infty$) of complete turns is forbidden, in view of (11.8). If cusps are present, it is sufficient to recall that their number must be uniformly bounded with respect to n , again due to assumption (11.8).

¹¹See [3, Lemma 3.3 (iii)] for similar arguments.

¹²See, for instance, [7, Theorem 1, Section 1.9].

convergence¹³ of $\text{im}(\Gamma_n)$ to $\text{im}(\Gamma)$, it follows that $\text{im}(\Gamma)$ has in U no points of self-intersections with transverse tangent vectors, and this contradicts inclusion (11.14). We deduce that

$$\#\text{crossings}(J_f) \leq \#\text{crossings}(J_{f_n}), \quad n \in \mathbb{N}. \quad (11.18)$$

We also notice that, if $x \in \text{cusps}(J_f)$ then, for any $n \in \mathbb{N}$, there exists $x_n \in \text{cusps}(J_{f_n})$ such that $\lim_{n \rightarrow +\infty} x_n = x$. Indeed, from (11.14), we have $x = \gamma_i(t)$ for some $i \in \{1, \dots, h\}$ and $t \in [0, 1]$. If by contradiction x has a neighbourhood U where there are no cusps of J_{f_n} for n sufficiently large, it follows that there is a neighbourhood of t , independent of n , where γ_i^n is of class \mathcal{C}^1 . Remembering (ii) of Definition 11.B.2, it then follows that γ_i is of class \mathcal{C}^1 in a neighbourhood of t , which is a contradiction. We deduce that

$$\#\text{cusps}(J_f) \leq \#\text{cusps}(J_{f_n}), \quad n \in \mathbb{N}. \quad (11.19)$$

From (11.18) and (11.19) we obtain

$$\#\text{nodes}(J_f) \leq \#\text{nodes}(J_{f_n}), \quad n \in \mathbb{N},$$

and (11.10) follows.

Reasoning as in [3, Lemma 3.4], from (11.14) and the regularity properties of J_f and Γ , it follows

$$\int_{J_f \setminus \text{cusps}(J_f)} |\kappa|^p d\mathcal{H}^1 \leq \mathcal{K}(\Gamma). \quad (11.20)$$

Arguing as in [3, Theorem 3.2], one can prove that, if $i \in \{1, \dots, h\}$ and if I is a compact subset contained in $[0, 1] \setminus \text{par}_{\text{cusp}}(\gamma_i)$, we have

$$\int_1 |\kappa_{\gamma_i}|^p |\gamma_i'| dt \leq \liminf_{n \rightarrow \infty} \int_1 |\kappa_{\gamma_i^n}|^p |\gamma_i^{n'}| dt \leq \liminf_{n \rightarrow \infty} \int_{[0,1] \setminus \text{par}_{\text{cusp}}(\gamma_i^n)} |\kappa_{\gamma_i^n}|^p |\gamma_i^{n'}| dt \leq C.$$

Passing to the supremum with respect to I gives

$$\int_{[0,1] \setminus \text{par}_{\text{cusp}}(\gamma_i)} |\kappa_{\gamma_i}|^p ds \leq \liminf_{n \rightarrow \infty} \int_{[0,1] \setminus \text{par}_{\text{cusp}}(\gamma_i^n)} |\kappa_{\gamma_i^n}|^p |\gamma_i^{n'}| dt.$$

Summing over i and using (11.20) we deduce

$$\int_{J_f \setminus \text{cusps}(J_f)} |\kappa|^p d\mathcal{H}^1 \leq \lim_{n \rightarrow +\infty} \mathcal{K}(\Gamma_n). \quad (11.21)$$

¹³For instance, in the Hausdorff distance.

From (11.17), (11.20) and (11.21) it follows

$$\mathcal{W}(f) \leq \beta \ell(\Gamma) + \lambda \mathcal{K}(\Gamma) \leq \lim_{n \rightarrow +\infty} \mathcal{W}(f_n),$$

which gives (11.9) and concludes the proof of the theorem. \square

11.3 On the Lower Semicontinuous Envelope of the Action

The minimization of the functional \mathcal{F} requires to relax the formulation of the problem (see point (1)) in the summarizing discussion in the end of Chap. 1 and to introduce the sequential lower semicontinuous envelope¹⁴ $\overline{\mathcal{F}}$ of \mathcal{F} , which is defined as follows. Denote by $\mathcal{P}(\mathbb{N})$ the set of all subsets of \mathbb{N} .

Definition 11.3.1 (Convergence) Let $(f_n, d_n, u_n) \in \text{Dom}(\mathcal{F})$, $f \in BV(\Omega, 2\mathbb{N})$, let $d : J_f \rightarrow \mathcal{P}(\mathbb{N})$ be a multifunction defined \mathcal{H}^1 almost everywhere, and $u \in H^1(\Omega \setminus \overline{\{d \ni 0\}})$. We say that the sequence $((f_n, d_n, u_n))$ converges to (f, d, u) , and we write $\lim_{n \rightarrow +\infty} (f_n, d_n, u_n) = (f, d, u)$, if

- $\lim_{n \rightarrow +\infty} f_n = f$ in $L^1(\Omega)$,
- $\lim_{n \rightarrow +\infty} \text{vis}(J_{f_n}) = \overline{\{d \ni 0\}}$ in the sense of Kuratowski,
- $\lim_{n \rightarrow +\infty} u_n = u$ in $L^1(\Omega)$.

Here we have used the notation $\{d \ni 0\} = \{x \in J_f : 0 \in d(x)\}$, and $\overline{\{d \ni 0\}}$ denotes the closure of $\{d \ni 0\}$.

The actual value of the multivalued function d is intentionally left essentially free except for the set $\{0 \in d\}$, and indeed the value of the functional is insensitive to such modifications. Defining d as some limit of the sequence (d_n) seems natural; however, the complexity of what can happen in the limit (see Sect. 11.3.2) suggests to impose as few restrictions as possible.

Definition 11.3.2 (Sequential Lower Semicontinuous Envelope) For any $f \in BV(\Omega, 2\mathbb{N})$, any multifunction $d : J_f \rightarrow \mathcal{P}(\mathbb{N})$, and any $u \in H^1(\Omega \setminus \overline{\{d \ni 0\}})$, we set

$$\overline{\mathcal{F}}(f, d, u)$$

$$:= \inf \left\{ \liminf_{n \rightarrow +\infty} \mathcal{F}(f_n, d_n, u_n) : (f_n, d_n, u_n) \in \text{Dom}(\mathcal{F}), \lim_{n \rightarrow +\infty} (f_n, d_n, u_n) = (f, d, u) \right\}.$$

¹⁴Sometimes called also (sequential) relaxation of \mathcal{F} .

The aim of this section is to study some preliminary properties of $\overline{\mathcal{F}}$. In connection with the analysis of the domain of $\overline{\mathcal{F}}$, the first problem to solve is maybe to understand what compactness properties are inherited by a sequence $(f_n, d_n, u_n) \in \text{Dom}(\mathcal{F})$ satisfying

$$\sup_{n \in \mathbb{N}} \mathcal{F}(f_n, d_n, u_n) < +\infty, \quad (11.22)$$

and to recognize the qualitative behaviour of the limit points. As we shall see, such sequences admit converging subsequences, but the limit does not belong, in general, to $\text{Dom}(\mathcal{F})$. Such limits consist, roughly speaking, of triplets (f, d, u) where f is a $BV(\Omega, 2\mathbb{N})$ function having jump contained in the image $\text{im}(\Gamma)$ of a limit system $\Gamma \in \text{sys}(\text{pwr}_c^p)$ of complete labelled contour graphs of class $\mathcal{W}_{\text{graph}}^{2,p}$. Twice the winding number of Γ with respect to a point $x \in \Omega \setminus \text{im}(\Gamma)$ turns out to coincide with $f(x)$. In contrast with what happens in the domain of \mathcal{F} , here:

- J_f may have an infinite number of singular points;
- J_f may have singular points not of the form of crossings or cusps [see Fig. 11.1 (right)]; in general, a cusp (or a crossing) of J_f may collide with another arc, or with another cusp and/or crossing (or many of them). Moreover, the aperture of a crossing may vanish in the limit, so that the transversality properties of crossings are lost; in addition, cusps are not necessarily ordinary cusps;
- f may jump of an even number along an arc of J_f , and not necessarily by two units [see again Fig. 11.1 (right)]; moreover, f may not jump along some parts of $\text{im}(\Gamma)$ (see Fig. 11.2).

In addition, d is now an \mathbb{N} -multivalued function defined on $\text{im}(\Gamma)$ (and $u \in H^1(\Omega \setminus \{x \in \text{im}(\Gamma) : d(x) \ni 0\})$). The multivalued character of d shows up along self-intersections of $\text{im}(\Gamma)$ of positive one-dimensional Hausdorff measure: an example of this behaviour is given in Fig. 11.2. As a consequence, it seems not easy to derive all compatibility conditions between the values of f and d , standing for a possible generalization of the rules in Fig. 3.11.

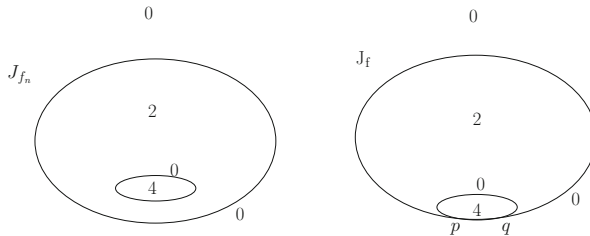


Fig. 11.1 The limit function f has a jump of 4 (right picture) along the arc (p, q) , hence (11.3) fails in the limit. Note that p and q are singularities of J_f which are neither cusps nor crossings in the usual sense that we have considered for apparent contours. Original image taken from [2]

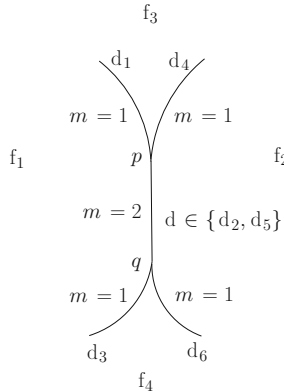


Fig. 11.2 Locally, the image $\text{im}(\Gamma)$ of the limit system Γ and the values of f . Note that if $f_1 = f_2$ then $\text{im}(\Gamma)$ strictly contains J_f . We display also the values of d out of the vertical segment $[p, q]$. On $[p, q]$ we have to retain the possible limits of (d_n) (for instance, two values d_2, d_5). This expresses the fact that limits of labellings may be, in general, \mathbb{N} -valued multifunctions. The symbol m stands for the multiplicity m_γ . Original image taken from [2]

We can conclude that the triplets in the domain of $\overline{\mathcal{F}}$ do not share, in general, the nice properties of the triplets in the domain of \mathcal{F} , which were a consequence of certain stability assumptions analysed in the previous chapters.

Let us start by finding some necessary conditions that arise in the limit when enforcing the bound (11.22).

Lemma 11.3.3 (Limits of Twice Winding Numbers) *Let $((f_n, d_n, u_n)) \subset \text{Dom}(\mathcal{F})$ be a sequence satisfying the bound (11.22). Then (f_n) has a subsequence converging in $L^1(\Omega)$ to a function*

$$f \in BV(\Omega, 2\mathbb{N}) \cap L^\infty(\Omega).$$

Proof Since (11.22) implies the bound (11.12), we can choose a (re)parametrized system $\hat{\Gamma}_n$ of pwr_c^p curves as in the proof of Theorem 11.2.2 [see formula (11.11)]. Arguing as in the proof of that theorem, we have that (11.15) and (11.16) hold; the assertion of the lemma follows.¹⁵ \square

Remark 11.3.4 (Even Heights of Jumps) In general, if f is as in Lemma 11.3.3, then

$$|f^+(x) - f^-(x)| \geq 2 \quad \text{for } \mathcal{H}^1 - \text{almost every } x \in J_f,$$

¹⁵See, for instance, [1, Theorem 3.23].

and the equality is not guaranteed. An example of this behaviour is shown in Fig. 11.1 (right). This is a difference between f and the function f appearing in Definition 11.1.4; compare with formula (11.3).

The next lemma informs us on the regularity of the jump of a function f : it must be contained in the image of a limit system.

Lemma 11.3.5 (Limit System and Jump of the Limit) *Let (f_n, d_n, u_n) and f be as in the statement of Lemma 11.3.3. If Γ_n is an oriented parametrization of J_{f_n} , then the sequence (Γ_n) admits a subsequence weakly converging in pwr_c^p to a limit system $\Gamma \in \text{pwr}_c^p([0, \mathcal{L}], \Omega)$, such that*

$$\text{im}(\Gamma) \supseteq J_f. \quad (11.23)$$

Moreover

$$f(x) = 2w(x, \Gamma), \quad x \in \Omega \setminus \text{im}(\Gamma). \quad (11.24)$$

Proof Inclusion (11.23) follows as in the proof of (11.14). Let us now consider a (not relabelled) subsequence of (f_n) converging almost everywhere to f in Ω ; then, equality (11.24) follows by passing to the limit in the equality

$$f_n(x) = 2w(x, \Gamma_n), \quad x \in \Omega \setminus \text{im}(\Gamma_n),$$

as $n \rightarrow +\infty$. □

Proposition 11.3.6 (Further Properties of Cusp Parameters) *Let $\Gamma = \{\gamma_1, \dots, \gamma_h\}$ be the limit system given by Lemma 11.3.5 and let $i \in \{1, \dots, h\}$. Then*

- $\#\text{par}_{\text{cusp}}(\gamma_i) \in 2\mathbb{N}$;
- for any $t \in \text{par}_{\text{cusp}}(\gamma)$ there exists a neighbourhood N of t such that $\gamma|_N$ is injective.

Proof The first conclusion follows invoking Proposition 11.B.4, while the second conclusion follows from (ii) of Definition 11.B.2. □

11.3.1 Limits of Labellings

Assume that, on the image of a limit system Γ of curves as in Lemma 11.3.5, it is defined an \mathbb{N} -valued multifunction d , i.e.,

$$d : \text{im}(\Gamma) \rightarrow \mathcal{P}(\mathbb{N}).$$

For such a map d to be “limit” of a sequence (d_n) , when $(f_n, d_n, u_n) \in \text{Dom}(\mathcal{F})$ satisfy (11.22), it is reasonable to expect that:

- (i) $\{x \in \text{im}(\Gamma) : d(x) \ni 0\} \neq \emptyset$;
- (ii) d is single valued on points $x \in \text{im}(\Gamma)$ such that $\#\Gamma^{-1}(x) = 1$;
- (iii) if $x \in \text{im}(\Gamma)$, then¹⁶

$$d(x) \leq \#\Gamma^{-1}(x), \quad (11.25)$$

see Fig. 11.2;

- (iv) the limit system Γ admits a reparametrization, still denoted by Γ , such that, if $x \in \text{im}(\Gamma)$, then

$$\max d(x) \leq 2 \min \left\{ \lim_{h \rightarrow +\infty} w(\Gamma, x_h) : (x_h) \subset \Omega \setminus \text{im}(\Gamma), \lim_{h \rightarrow +\infty} x_h = x \right\}.$$

The latter condition is a “relaxed” version of the compatibility condition (L2) of Definition 4.2.5 between d and f on the relative interior of the arcs; if $\text{im}(\Gamma)$ is a complete labelled contour graph of class $\mathcal{W}_{\text{graph}}^{2,p}$, it reduces to that compatibility condition.

Furthermore, one can expect that:

- (v) there exists a locally constant function \tilde{d} defined on the domain of Γ , and taking values in \mathbb{N} , such that

$$d(x) = \bigcup_{s \in \Gamma^{-1}(x)} \tilde{d}(s), \quad x \in \text{im}(\Gamma),$$

and

$$|\tilde{d}(t^+) - \tilde{d}(t^-)| = 1, \quad t \in \text{par}_{\text{cusp}}(\Gamma). \quad (11.26)$$

Condition (11.26) is a “relaxed” version of one of the conditions in (L3) of Definition 4.2.5, concerning the behaviour of d around a cusp of an apparent contour, see the second and third pictures of Fig. 3.11.

- (vi) For a function u which is the limit of a sequence (u_n) with $(f_n, d_n, u_n) \in \text{Dom}(\mathcal{F})$ satisfying (11.22), it is reasonable to expect that u belongs to the space of special functions of bounded variation in Ω [1], and moreover that

$$u \in H^1 \left(\Omega \setminus \overline{\{d \ni 0\}} \right). \quad (11.27)$$

¹⁶Equality in (11.25) in general does not hold, as in the case of Fig. 11.2 with the choices $d_1 = d_2 = d_3 = d_4 = d_5 = d_6 = 0$.

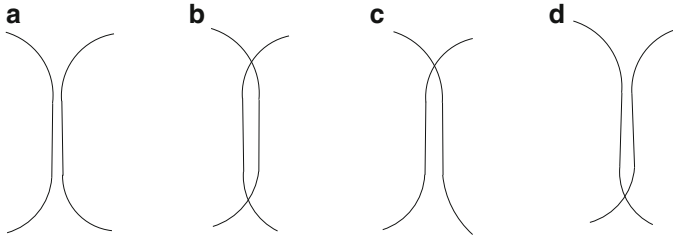


Fig. 11.3 Only (c) and (d) should correspond to a crossing in the limit system of curves Γ of Fig. 11.2. Original image taken from [2]

The above remarks enlighten a difficulty, related to the definition of the character of self-intersection points (which may even lie on Cantor-type sets) of an immersion (or of two or more immersions) of \mathbb{S}^1 : in particular, understanding whether or not such points must be considered as “crossings”. Moreover, once one decides whether self-intersections must be considered a crossing or not then, in order to describe the functional $\overline{\mathcal{F}}$, one has to understand which compatibility conditions should be satisfied between the local values of f and d . For instance, in the simple case of Fig. 11.2, understanding whether or not the *whole* vertical segment $[p, q]$ must be considered as a “crossing” requires to decide whether or not Γ is the limit of the third or the fourth configuration of Fig. 11.3. Note however that the first two pictures in the same figure are other possible approximations, that cannot be neglected a priori. A detailed discussion of this example is made in the next section where we show, in particular, that certain choices of f_i and d_i in Fig. 11.2 are not compatible for (f, d) to be a limit of a sequence $((f_n, d_n))$ with $((f_n, d_n, u_n)) \subset \text{Dom}(\mathcal{F})$ satisfying (11.22).

11.3.2 Sufficient Conditions: An Example

The characterization of the domain of $\overline{\mathcal{F}}$ is related to the description of the compatibility conditions between the values of the multifunction d and of the function f , locally around self-intersections of $\text{im}(\Gamma)$ having positive \mathcal{H}^1 measure. The problem is the following. Let be given f, Γ, d, u satisfying conditions (i)–(vi) of Sect. 11.3.1: find under which further conditions there exists a sequence $((f_n, d_n, u_n)) \subset \text{Dom}(\mathcal{F})$ converging to (f, d, u) and satisfying (11.22). In this section we want to show that the answer to this problem is involved, even in the simplest cases: we will exhibit several situations where keeping (11.22) is not possible. For simplicity, we shall concentrate only on f and d , for a given function u as in (11.27).

Example 11.3.7 (Approximating a Whole Segment of Self-Intersections) Let γ be given locally a curve γ , the image of which is depicted in Fig. 11.2, having the following multiplicity¹⁷ $m = m_\gamma$:

- $m = 1$ on the (relatively open) curved arcs,
- $m = 2$ on the vertical segment $[p, q]$.

Assume also that the parametrization of $\text{im}(\gamma)$ is everywhere regular and smooth, so that $\text{par}_{\text{cusp}}(\gamma) = \emptyset$.

Let $f_{\min} \in 2\mathbb{N}$ and let be given numbers

$$f_1, f_2, f_3, f_4 \in \{f_{\min}, f_{\min} + 2, f_{\min} + 4\},$$

which are the values of f out of $\text{im}(\gamma)$, such that locally the values of f are as in Fig. 11.2.

Finally, let be given integer numbers

$$d_1, d_2, d_3, d_4, d_5, d_6 \in \{0, \dots, f_{\min}\}, \quad (11.28)$$

which are the values of the limit labelling d . Suppose that

- d is single valued on the relatively open curved arcs of $\text{im}(\gamma)$,
- d takes two (possibly equal) values $\{d_2, d_5\}$ on $[p, q]$ (recall inequality (11.25)).

The set $\text{im}(\gamma)$ is obtained as a limit of (part of) the boundary of the two regions corresponding to $f = f_i$, $i = 1, 2$, which locally collide on $[p, q]$ keeping (11.22) satisfied (see Fig. 11.3).

Note that

- J_f may be strictly contained in $\text{im}(\gamma)$, for instance when $f_1 = f_2$, since in this case f does not jump along the vertical segment $[p, q]$;
- it may happen that $|f^+(x) - f^-(x)| = 4$ on $[p, q] \subset J_f$, for instance when $f_2 = f_1 + 4$;
- J_f may not contain $\{d \ni 0\}$ (while $\{d \ni 0\} \subseteq \text{im}(\gamma)$), for instance when $d_1 = d_2 = d_3 = d_4 \neq 0$ and $d_5 = d_6 = 0$ and $f_1 = f_2$.

Now, we show that for some values of f_i and d_j there cannot be any sequence $((f_n, d_n))$ converging to (f, d) and keeping (11.22) valid. We shall consider configurations up to obvious symmetries (such as, in some cases, that of pictures (c) and (d) of Fig. 11.3).

Case 1. f takes two values.

The approximating sequence (f_n) of the last picture of Fig. 11.4 violates one of the compatibility conditions in Fig. 3.11 at the crossing. Similarly, this violation happens when inserting $f_n = f_{\min}$ in the thin region between the two crossings in

¹⁷See Definition 11.A.3, below.

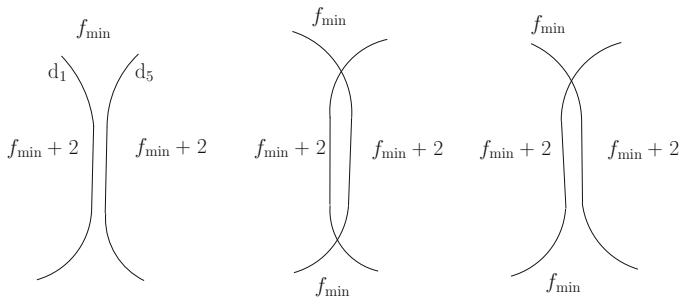


Fig. 11.4 A case in which f takes two values, a situation similar to the one in Fig. 3.2 of Chap. 3. We depict the values of possible approximations f_n of f . The first two pictures represent approximating sequences in the domain of \mathcal{F} satisfying the bound (11.22); the first configuration does not introduce crossings, hence the corresponding value of \mathcal{F} is lower. Original image taken from [2]

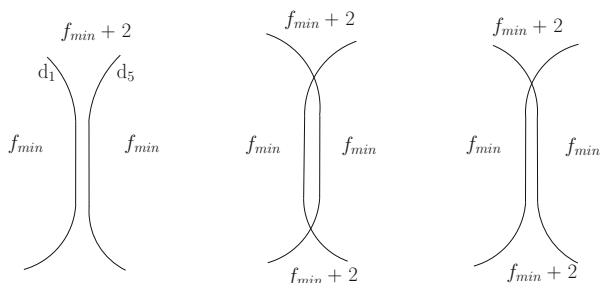
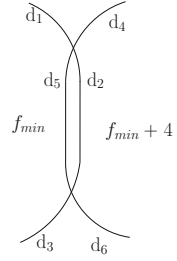


Fig. 11.5 A case in which f takes two values, corresponding locally, for instance, to the limit of a thin dumbbell. We depict the values of possible approximations f_n of f . Original image taken from [2]

the central picture of Fig. 11.4. On the other hand, inserting $f = f_{\min} + 4$ in the thin region is admissible, but in terms of the functional \mathcal{F} two crossings are penalized. The other way to approximate (f, d) with (f_n, d_n) is displayed in the first picture Fig. 11.4, giving no crossings contribution to \mathcal{F} . In this case the limit segment $[p, q]$ in Fig. 11.2 should *not* be considered as a crossing, and therefore the two singular points p and q should not be penalized in the computation of the value of $\overline{\mathcal{F}}$.

The other case when f takes two values is depicted in Fig. 11.5. The approximating sequence (f_n) of the third picture in Fig. 11.5 violates one of the conditions of Fig. 3.11, at the crossing. Similarly, inserting $f_n = f_{\min} + 2$ or $f_n = f_{\min} + 4$ in the thin region between the two crossings in the central picture of Fig. 11.5 violates that condition, while inserting $f_{\min} - 2$ is admissible, provided $f_{\min} \geq 2$. The other way to approximate (f, d) with (f_n, d_n) is shown in the left picture of Fig. 11.5. As in the situation of Fig. 11.4, also in this case the limit segment $[p, q]$ in Fig. 11.2 should *not* be considered as a crossing.

Fig. 11.6 A case in which f takes three values. This configuration represents an approximating sequence in the domain of \mathcal{F} only if $d_1 = d_3$ and $d_4 = d_6$.
Original image taken from [2]



The conclusion is the following:

If f takes two values $\{f_{\min}, f_{\min+2}\}$ around $[p, q]$, then the system Γ in Fig. 11.2 can be parametrized (as the limit of the parametrizations corresponding to the left pictures of Figs. 11.4 or 11.5) in such a way that it is possible to find a sequence $((f_n, d_n))$ (presumably locally \mathcal{F} -optimal) satisfying (11.22) and converging to (f, d) , provided $d_1 = d_2 = d_3, d_4 = d_5 = d_6$ satisfy (11.28).

Case 2. f takes three values.

We shall consider only the interesting case when $f_1 = f_{\min}, f_2 = f_{\min} + 4$ and $f_3 = f_4 = f_{\min} + 2$.

Proposition 11.3.8 (Admissible Approximating Sequences) *The configuration of Fig. 11.6 represents a sequence $((f_n, d_n))$ in the domain of \mathcal{F} only if*

$$d_1 = d_3 \quad \text{and} \quad d_4 = d_6. \quad (11.29)$$

Proof Assume by contradiction that $d_1 \neq d_3$. From the compatibility conditions in Fig. 3.11 it follows $d_2 \geq d_1$ and $d_2 \geq d_3$, more precisely

$$\min(d_1, d_3) + 2 = d_2 = \max(d_1, d_3). \quad (11.30)$$

Suppose $d_1 < d_3$, the case $d_1 > d_3$ being similar. From (11.30) it follows

$$d_1 + 2 = d_2 = d_3. \quad (11.31)$$

From Fig. 3.11 we also have $d_4 = d_5$, and

$$d_5 \leq d_1. \quad (11.32)$$

Moreover the only possibility which is not ruled out by (11.31) is $d_5 = d_6 - 2$. This, together with Fig. 3.11, implies $d_5 \geq d_3$. Using (11.32) we deduce $d_3 \leq d_5 \leq d_1$, a contradiction. The first relation in (11.29) is proved.

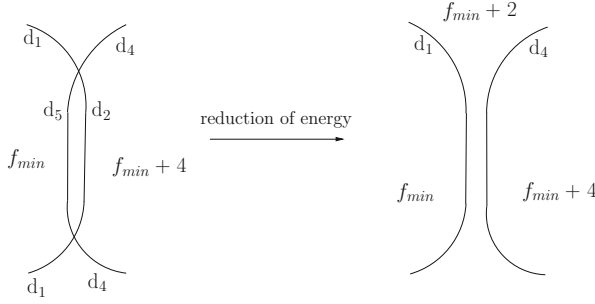


Fig. 11.7 In these two configurations we depict the values of (f_n, d_n) , where f_n converge to the winding number of a parametrization of the curve in Fig. 11.2, and d_n converge to the multifunction d in Fig. 11.2. The value of \mathcal{F} is lower for the right configuration, since no singular points are present. Original image taken from [2]

It remains to show that $d_4 = d_6$. We have two cases: either $d_2 = d_1 + 2$ or $d_2 = d_1$. If $d_2 = d_1 + 2$, then $d_2 = d_3 + 2$ by the equality $d_1 = d_3$ proved above. Hence from Fig. 3.11 we deduce $d_4 = d_5 = d_6$.

If $d_2 = d_1$, we have $d_2 = d_3$. Then $d_4 = d_5 + 2 = d_6$. The proof of (11.29) is complete. \square

The conclusion is the following:

If f takes three values $\{f_{\min} = f_1, f_{\min+2} = f_3 = f_4, f_{\min+4} = f_2\}$ around $[p, q]$, and if (11.28) and (11.29) hold, provided we reparametrize the system Γ of Fig. 11.2 (as the limit of the parametrizations corresponding to one of the two pictures in Fig. 11.7), then it is possible to find a sequence $((f_n, d_n))$ converging to (f, d) and satisfying (11.22). The sequence corresponding to the right picture of Fig. 11.7 has no singular points, and therefore its value in terms of \mathcal{F} is lower than the one in the left picture; in this case, obviously $d_1 = d_2$ and $d_4 = d_5$. It follows that also in this situation the limit segment $[p, q]$ *must not* be considered as a “crossing”.

If f takes three values $\{f_{\min} = f_1, f_{\min+2} = f_3 = f_4, f_{\min+4} = f_2\}$ around $[p, q]$, if (11.28) holds, and if

$$d_1 \neq d_3 \quad \text{or} \quad d_4 \neq d_6, \quad (11.33)$$

then the approximating sequence in Fig. 11.6 does not belong to $\text{Dom}(\mathcal{F})$ by Proposition 11.3.8. Still the value of $\overline{\mathcal{F}}$ is finite, since Fig. 11.8 (with either $d_2 = d_1 + 2$ and $d_4 = d_5$, or $d_1 = d_2$ and $d_4 = d_5 + 2$) shows a possible approximating sequence in $\text{Dom}(\mathcal{F})$ satisfying (11.22). In this case the *whole* segment $[p, q]$ should be considered as one crossing, and therefore correspondingly penalized in the computation of $\overline{\mathcal{F}}$.

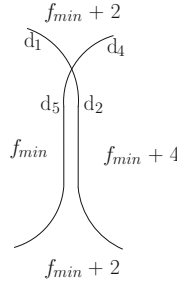


Fig. 11.8 A possible approximation of the configuration of Fig. 11.2 under condition (11.33). The limit parametrization has a crossing in correspondence of the vertical overlapping segment, that should presumably contribute in the computation of \mathcal{F} . Original image taken from [2]

11.A Appendix A: Systems of Curves

In this section we collect some useful definitions and results on parametrizations of curves (and of finite family of curves) with cusps, in particular for parametrization of complete labelled contour graphs. These definitions are needed, in particular, when passing to the limit along sequences consisting of connected components of the graphs.

11.A.1 Curves of Class pwr_c^p

In the presence of cusps, a component of a complete labelled contour graph of class $\mathcal{W}_{\text{graph}}^{2,p}$ cannot be parametrized with a single curve of class $W^{2,p}$ (Remark 11.1.2), and a finite number of such parametrizations is required. These latter maps fall within the class of parametrizations of the next definition which, in addition, contains the limits of converging sequences of components of complete contour graphs of class $\mathcal{W}_{\text{graph}}^{2,p}$ having a uniform bound on the functional \mathcal{F} .

Definition 11.A.1 (The Class pwr_c^p) Let $\mathcal{L} > 0$ and $\gamma : [0, \mathcal{L}] \subset \mathbb{R} \rightarrow \overline{\Omega}$ be a continuous closed curve. We say that γ is a piecewise regular $W^{2,p}$ curve with cusps, and we write

$$\gamma \in \text{pwr}_c^p([0, \mathcal{L}], \overline{\Omega}),$$

if there exists a (possibly empty) finite set

$$\text{par}_{\text{cusp}}(\gamma)$$

of points of $[0, \mathcal{L}]$, called cusp parameters of γ , such that the following properties hold: if I is any connected component of $[0, \mathcal{L}] \setminus \text{par}_{\text{cusp}}(\gamma)$, then

- (i) $\gamma \in W^{2,p}(I, \overline{\Omega})$,
- (ii) there exists a constant $c > 0$ independent of I , such that

$$|\gamma'(t)| = c, \quad t \in I,$$

where $|\gamma'(t)|$ denotes the euclidean norm of the vector $\gamma'(t)$,

- (iii) for any $t \in \text{par}_{\text{cusp}}(\gamma)$, the angle between the two unit vectors $\lim_{\tau \rightarrow t^-} \frac{\gamma'(\tau)}{|\gamma(\tau)|}$ and $\lim_{\tau \rightarrow t^+} \frac{\gamma'(\tau)}{|\gamma(\tau)|}$ belongs to $\{-\pi, \pi\}$.

When the image of γ is contained in Ω , we write $\gamma \in \text{pwr}_c^p([0, \mathcal{L}], \Omega)$. When it is not necessary to specify the domain and the codomain of γ , we simply write $\gamma \in \text{pwr}_c^p$.

Remark 11.A.2 (Behaviours of Components) A component of a complete contour graph of class $\mathcal{W}_{\text{graph}}^{2,p}$ is the image of a map $\gamma \in \text{pwr}_c^p$. However, in general the image of a map $\gamma \in \text{pwr}_c^p$ is a much more general object. Indeed:

- $\text{im}(\gamma)$ may have a large set (for instance, a continuum set) of self-intersections,
- no transversality conditions on isolated self-intersections of $\text{im}(\gamma)$ is imposed,
- no condition on the number of arcs concurring at an isolated self-intersection of $\text{im}(\gamma)$ is imposed,
- if t is a cusp parameter of γ , then $\gamma(t)$ could be a self-intersection of $\text{im}(\gamma)$,
- a local situation like the one displayed in the right picture of Fig. 3.7 is allowed,
- the winding number of a point in $\Omega \setminus \text{im}(\gamma)$ with respect to γ is not necessarily nonnegative.

Definition 11.A.3 (Multiplicity) Given $\gamma \in \text{pwr}_c^p$, we set

$$m_\gamma(x) := \#\gamma^{-1}(x), \quad x \in \text{im}(\gamma),$$

and we call m_γ the multiplicity of γ .

We denote by $\ell(\gamma) \in (0, +\infty)$ the length of γ , and we set

$$\mathcal{K}(\gamma) := \int_{(0, \mathcal{L}) \setminus \text{par}_{\text{cusp}}(\gamma)} |\kappa_\gamma|^p |\gamma'| dt,$$

where κ_γ is the curvature (a real number) of γ . Notice that $\kappa_\gamma \in L^p(I)$ for any interval $I \subseteq [0, \mathcal{L}] \setminus \text{par}_{\text{cusp}}(\gamma)$, and in particular

$$\mathcal{K}(\gamma) < +\infty.$$

A regular reparametrization of $\gamma \in \text{pwr}_c^p([0, \mathcal{L}], \mathbb{R}^2)$ is a map $\hat{\gamma}$ of the form $\hat{\gamma} = \gamma \circ \varphi$, where $\varphi : [0, L] \rightarrow [0, \mathcal{L}]$ is a Lipschitz¹⁸ strictly monotone surjective function, in such a way that $\hat{\gamma} \in \text{pwr}_c^p([0, L], \mathbb{R}^2)$.

The map γ can be regularly reparametrized¹⁹ by an arc length parameter s , i.e., with unit speed, on each connected component of $[0, \ell(\gamma)] \setminus \text{par}_{\text{cusp}}(\gamma)$, where $\text{par}_{\text{cusp}}(\gamma)$ consists now of the set of cusp parameters in the new parametrization, which is the image via the reparametrizing map of the set of cusp parameters in the original parametrization. We denote by $\dot{\gamma}$ the derivative of γ with respect to s , and we have $|\kappa_\gamma| = |\ddot{\gamma}|$ almost everywhere in $[0, \ell(\gamma)] \setminus \text{par}_{\text{cusp}}(\gamma)$. If $x \in \overline{\Omega} \setminus (\gamma)$,

$$w(\gamma, x)$$

denotes the winding number (or index) of x with respect to γ .

11.A.2 Systems of Curves

Since in general the graphs that we consider consist of more than one component, we need the following definition.

Definition 11.A.4 (System of pwr_c^p Curves) By a system of $\text{pwr}_c^p([0, \mathcal{L}], \overline{\Omega})$ curves we mean a finite family $\{\gamma_1, \dots, \gamma_h\}$, where $\gamma_i \in \text{pwr}_c^p([0, \mathcal{L}], \overline{\Omega})$ for any $i \in \{1, \dots, h\}$.

The set of all systems of $\text{pwr}_c^p([0, \mathcal{L}], \overline{\Omega})$ curves is denoted by

$$\text{sys}(\text{pwr}_c^p)([0, \mathcal{L}], \overline{\Omega}).$$

When it is not necessary to specify the domain and the codomain, we shall use the symbol $\text{sys}(\text{pwr}_c^p)$; when the image of all curves composing the system is contained in Ω , we shall write $\text{sys}(\text{pwr}_c^p)([0, \mathcal{L}], \Omega)$.

The parameter space of the system $\Gamma = \{\gamma_1, \dots, \gamma_h\} \in \text{sys}(\text{pwr}_c^p)$ is the disjoint union of h copies of \mathbb{S}^1 . A regular reparametrization of $\Gamma = \{\gamma_1, \dots, \gamma_h\}$ is a system $\{\hat{\gamma}_1, \dots, \hat{\gamma}_h\} \in \text{sys}(\text{pwr}_c^p)$ where $\hat{\gamma}_i$ is a regular reparametrization of γ_i for any $i \in \{1, \dots, h\}$.

¹⁸It is also of class $C^1(\bar{I})$, where I is any connected component of the complement of cusp parameters.

¹⁹For notational simplicity, we shall keep the symbol γ to denote an arc-length reparametrization.

We set

$$\begin{aligned}\mathrm{im}(\Gamma) &:= \bigcup_{i=1}^h \mathrm{im}(\gamma_i), \\ \mathrm{par}_{\mathrm{cusp}}(\Gamma) &:= \bigcup_{i=1}^h \mathrm{par}_{\mathrm{cusp}}(\gamma_i),\end{aligned}$$

and

$$\ell(\Gamma) := \sum_{i=1}^h \ell(\gamma_i), \quad \mathcal{K}(\Gamma) := \sum_{i=1}^h \mathcal{K}(\gamma_i).$$

We define the winding number of Γ as

$$\mathrm{w}(\Gamma, x) := \sum_{i=1}^h \mathrm{w}(\gamma_i, x), \quad x \in \overline{\Omega} \setminus \mathrm{im}(\Gamma).$$

11.A.3 Parametrizations of Complete Contour Graphs

Let $f \in BV(\Omega, 2\mathbb{N})$ be such that J_f is a complete contour graph of class $\mathcal{W}_{\mathrm{graph}}^{2,p}$. Then J_f can be parametrized by a system of pwr_c^p curves as follows. Let C_1, \dots, C_h be the components of J_f . For any $i \in \{1, \dots, h\}$ we take a curve $\gamma_i \in \mathrm{pwr}_c^p$ such that

$$C_i = \mathrm{im}(\gamma_i) \quad \text{and} \quad \gamma_i^{-1}(\mathrm{cusps}(C_i)) = \mathrm{par}_{\mathrm{cusp}}(\gamma_i).$$

Accordingly, we say that the system $\Gamma := \{\gamma_1, \dots, \gamma_h\} \in \mathrm{sys}(\mathrm{pwr}_c^p)$ parametrizes J_f . Observe that

$$\mathcal{W}(f) + \sigma \# \mathrm{cusps}(J_f) = \beta \ell(\Gamma) + \lambda \mathcal{K}(\Gamma) + \sigma \# \mathrm{par}_{\mathrm{cusp}}(\Gamma).$$

The next lemma says that f can be described with the winding number of a “canonical” parametrization of its jump.

Lemma 11.A.5 (Existence of Oriented Parametrizations) *Let (G, f, d) be a complete labelled contour graph of class $\mathcal{W}_{\mathrm{graph}}^{2,p}$. Then there exists a system Γ of pwr_c^p curves parametrizing G and satisfying*

$$f(x) = 2\mathrm{w}(\Gamma, x), \quad x \in \Omega \setminus \mathrm{im}(\Gamma). \quad (11.34)$$

Proof Let $\{\gamma_1, \dots, \gamma_h\}$ be any system of pwr_c^p curves parametrizing G . Then, in order to fulfill condition (11.34), it is sufficient to orient $\gamma_1, \dots, \gamma_h$ in such a way that at each point out of G , the higher value of f is taken locally on the left. \square

Definition 11.A.6 (Oriented Parametrization of G) Let (G, f, d) be a complete labelled contour graph of class $\mathcal{W}_{\text{graph}}^{2,p}$. A system $\Gamma \in \text{sys}(\text{pwr}_c^p)$ parametrizing G and satisfying condition (11.34) is called an oriented parametrization of G .

11.B Appendix B: Convergence and Compactness of Systems of Curves

Let $(\gamma_n) \subset \text{pwr}_c^p([0, \mathcal{L}], \Omega)$ be a sequence of curves such that

$$\sup_{n \in \mathbb{N}} \{\beta \ell(\gamma_n) + \lambda \mathcal{K}(\gamma_n) + \sigma \# \text{par}_{\text{cusp}}(\gamma_n)\} < +\infty. \quad (11.35)$$

The uniform bound (11.35) does not exclude, in principle, that

$$\lim_{n \rightarrow +\infty} \ell(\gamma_n) = 0. \quad (11.36)$$

For instance, let $\text{im}(\gamma)$ be as in the right picture of Fig. 3.12, so that $\text{im}(\gamma)$ has two cusps. By homothety and aligning²⁰ at the same time the arcs connecting the cusps, we can shrink $\text{im}(\gamma)$ to a point, and produce a sequence of curves γ_n satisfying (11.35) and (11.36).

Lemma 11.B.1 (Cusps on a Component with Infinitesimal Length) *Let $(\gamma_n) \subset \text{pwr}_c^p([0, \mathcal{L}], \Omega)$ be a sequence satisfying the bound (11.35), and suppose that $\# \text{par}_{\text{cusp}}(\gamma_n) \in 2\mathbb{N}$ for any $n \in \mathbb{N}$. If (11.36) holds, then there exists $\bar{n} \in \mathbb{N}$ such that*

$$\# \text{par}_{\text{cusp}}(\gamma_n) \geq 2, \quad n \in \mathbb{N}, \quad n \geq \bar{n}.$$

Proof Suppose by contradiction that, for an arbitrarily large $n \in \mathbb{N}$, we can find a curve $\gamma_n \in \text{pwr}_c^p([0, \mathcal{L}], \Omega)$ so that $\# \text{par}_{\text{cusp}}(\gamma_n) < 2$, and therefore $\text{par}_{\text{cusp}}(\gamma_n) = \emptyset$. We choose a continuous function $\theta_n : [0, \mathcal{L}] \rightarrow \mathbb{R}$ as follows: given $t \in [0, \mathcal{L}]$, we define $\theta_n(t)$ as the oriented angle between the first coordinate axis and the tangent

²⁰For example, if the common tangent line at the two cusps is $\{x_2 = 0\}$, then one can use a transformation of the form $(x_1, x_2) \rightarrow (\varepsilon x_1, \varepsilon^2 x_2)$, and then let $\varepsilon \rightarrow 0$.

vector of γ_n at t . Let us write, with obvious notation, $\dot{\gamma}_n(s) = (\cos \theta_n(s), \sin \theta_n(s))$ at the differentiability points of γ_n . Then, $\theta_n \in W^{1,p}((0, \mathcal{L}))$, and using Hölder's inequality, it follows

$$\int_{(0, \ell(\gamma_n))} |\kappa_{\gamma_n}|^p ds = \int_{(0, \ell(\gamma_n))} |\dot{\theta}_n|^p ds \geq \ell(\gamma_n)^{-p/p'} \left(\int_{(0, \ell(\gamma_n))} |\dot{\theta}_n| ds \right)^p,$$

where $\frac{1}{p} + \frac{1}{p'} = 1$. Hence

$$\begin{aligned} \ell(\gamma_n) &\geq \left(\int_{(0, \ell(\gamma_n))} |\dot{\theta}_n| ds \right)^{p'} \left(\int_{(0, \ell(\gamma_n))} |\kappa_{\gamma_n}|^p ds \right)^{-\frac{p'}{p}} \\ &\geq (2\pi)^{p'} \left(\int_{(0, \ell(\gamma_n))} |\kappa_{\gamma_n}|^p ds \right)^{-\frac{p'}{p}} \geq (2\pi)^{p'} \lambda^{\frac{p'}{p}} C^{-\frac{p'}{p}}, \end{aligned} \quad (11.37)$$

where $C \in (0, +\infty)$ is an upper bound for the left-hand side of (11.35). The proof is concluded, because formula (11.37) contradicts assumption (11.36). \square

11.B.1 Convergence

Now, we give the notion of convergence for a sequence of elements of pwr_c^p ; this convergence is related to the coercivity properties of \mathcal{F} (see Theorem 11.B.3).

For simplicity, we start with sequences of curves, the systems being considered later.

Definition 11.B.2 (Weak Convergence of pwr_c^p Curves) Let $(\gamma_n) \subset \text{pwr}_c^p([0, \mathcal{L}], \Omega)$ and $\gamma \in \text{pwr}_c^p([0, \mathcal{L}], \overline{\Omega})$. We say that the sequence (γ_n) weakly converges to γ in pwr_c^p if:

- there exist two real constants ℓ_1, ℓ_2 such that

$$0 < \ell_1 \leq \ell(\gamma_n) \leq \ell_2, \quad n \in \mathbb{N}; \quad (11.38)$$

- $\#\text{par}_{\text{cusp}}(\gamma_n)$ is independent of n ;
- the curve γ and each curve γ_n admit a regular reparametrization on $[0, 1]$, for simplicity still denoted by γ and γ_n respectively, such that $\lim_{n \rightarrow +\infty} \gamma_n = \gamma$ in $C^0([0, 1]; \overline{\Omega})$, and the following properties hold:

- (i) the sequence $(\text{par}_{\text{cusp}}(\gamma_n))$ converges in the sense of Kuratowski to a finite set K of points of $[0, 1]$, with

$$K \supseteq \text{par}_{\text{cusp}}(\gamma),$$

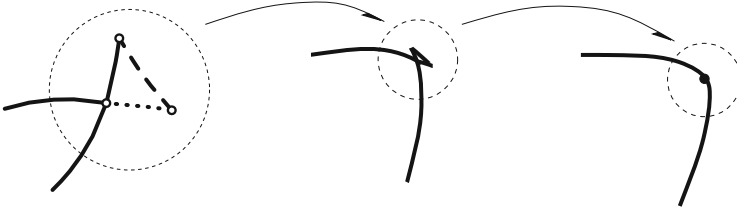


Fig. 11.9 The three points making up a swallow's tail (two cusps and one crossing) can collapse to a single point as shown in this sequence. Observe that, from the discussion after Definition 11.B.2, the bound (11.35) remains valid [as well as (11.38)]. The curve of the last picture is of class $W^{2,p}$; in particular, the tangent vector is continuous in a neighbourhood of the marked point

- (ii) the sequence (γ_n) converges to γ as $n \rightarrow +\infty$ weakly in $W^{2,p}([0, 1] \setminus K, \overline{\Omega})$, and in $C^1(\overline{I}, \overline{\Omega})$ for any connected component I of $[0, 1] \setminus K$.

The set $\#\text{par}_{\text{cusp}}(\gamma)$ of cusp parameters²¹ of γ can be strictly contained in the set K , because of the possible collision, in the limit, of (two or more) distinct cusp parameters of γ_n . To construct an example of this behaviour, it is sufficient to take a sequence (γ_n) having as image the torus in Fig. 3.6 and satisfying (11.35), in such a way that, as $n \rightarrow +\infty$, one of the two swallow's tails tends to disappear (Fig. 11.9), forcing the two involved cusps to collide at the crossing. The construction must be made so that, as $n \rightarrow +\infty$, also the crossing disappears, in the sense that the curve becomes locally of class C^1 , and hence of class $W^{2,p}$. In this case the parameter corresponding to the collision points belongs to $K \setminus \text{par}_{\text{cusp}}(\gamma)$.

Theorem 11.B.3 (Compactness in pwr_c^p) *Let $(\gamma_n) \subset \text{pwr}_c^p([0, \mathcal{L}], \Omega)$ be a sequence such that $\#\text{par}_{\text{cusp}}(\gamma_n) \in 2\mathbb{N}$ for any $n \in \mathbb{N}$. Suppose that the bound (11.35) holds. Then there exists a (not relabelled) subsequence such that one of the two following alternatives holds: either (11.36) holds, and in this case $\#\text{par}_{\text{cusp}}(\gamma_n) \geq 2$, or there exists $\gamma \in \text{pwr}_c^p([0, \mathcal{L}], \overline{\Omega})$ such that (γ_n) weakly converges to γ in pwr_c^p .*

Proof Take a (not relabelled) subsequence such that any addendum on the left-hand side of (11.35) has a limit as $n \rightarrow +\infty$, and so that the sum of the three limits is finite. If (11.36) holds, then the thesis follows from Lemma 11.B.1. Hence, we can suppose that

$$\inf_{n \in \mathbb{N}} \ell(\gamma_n) > 0. \quad (11.39)$$

The bound (11.35) implies that

$$\sup_{n \in \mathbb{N}} \ell(\gamma_n) < +\infty, \quad (11.40)$$

²¹The structure of $\text{par}_{\text{cusp}}(\gamma)$ can be described (see (11.44), below).

and that, possibly extracting a not relabelled subsequence,

$$\#\text{par}_{\text{cusp}}(\gamma_n) = l \text{ is independent of } n.$$

We have to construct a curve $\gamma \in \text{pwr}_c^p([0, \mathcal{L}], \overline{\Omega})$ for which all requirements in Definition 11.B.2 are satisfied.

From (11.39) and (11.40) it follows that, for any $n \in \mathbb{N}$, we can regularly reparametrize γ_n with a map still denoted, for simplicity, by $t \in [0, 1] \rightarrow \gamma_n(t)$ (recall that 0 and 1 are identified). If

$$\text{par}_{\text{cusp}}(\gamma_n) = \{t_{1,n}, \dots, t_{l,n}\} \subset [0, 1],$$

possibly passing to a further not relabelled subsequence, we can assume that

$$\exists \lim_{n \rightarrow +\infty} t_{j,n} =: t_j \in [0, 1], \quad j \in \{1, \dots, l\}, \quad (11.41)$$

where the t_j are not necessarily distinct (see Fig. 11.9 for an example). We define

$$T := \{t_1, \dots, t_l\}. \quad (11.42)$$

Observe that the sequence $(\text{par}_{\text{cusp}}(\gamma_n))$ of sets converges to T in the sense of Kuratowski.

Let $I \subseteq [0, 1]$ be a connected component of $[0, 1] \setminus T$. Let I_n be a connected component of $[0, 1] \setminus \text{par}_{\text{cusp}}(\gamma_n)$, chosen so that the left (respectively right) extremum of I_n converges, as $n \rightarrow +\infty$, to the left (respectively right) extremum of I . Let $\delta > 0$ be sufficiently small, and let $I_\delta^- := \{t \in I : \text{dist}(t, T) > \delta\}$, for $\delta > 0$ small enough. Notice that $I_n \supset I_\delta^-$, for $n \in \mathbb{N}$ large enough. Denote by \mathcal{J}_n the connected component, in an arc length parametrization of γ_n , corresponding to I_n , and let $\ell(\gamma_n; I_n)$ be the length of $\gamma_n(I_n)$. Using the equality

$$\int_{\mathcal{J}_n} |\kappa_{\gamma_n}|^p ds = \ell(\gamma_n; I_n)^{1-2p} \int_{I_n} |\gamma_n''|^p dt,$$

from (11.40) and (11.35) we obtain that

$$\sup_{n \in \mathbb{N}} \int_{I_n} |\gamma_n''|^p dt < +\infty.$$

Since the image of each γ_n is contained in Ω , we deduce that

$$\sup_{n \in \mathbb{N}} \|\gamma_n\|_{W^{2,p}(I_n, \mathbb{R}^2)} < +\infty.$$

Therefore, from (11.41), it follows that there exist a not relabelled subsequence²² of (γ_n) and a map $\gamma_{I_\delta^-} \in W^{2,p}(I_\delta^-, \mathbb{R}^2)$, such that

$$\lim_{n \rightarrow +\infty} \gamma_n|_{I_\delta^-} = \gamma_{I_\delta^-} \quad \text{in } C^1(\overline{I_\delta^-}, \mathbb{R}^2) \text{ and weakly in } W^{2,p}(I_\delta^-, \mathbb{R}^2), \quad (11.43)$$

and $\|\gamma_{I_\delta^-}\|_{W^{2,p}(I_\delta^-, \mathbb{R}^2)} \leq \liminf_{n \rightarrow +\infty} \|\gamma_n\|_{W^{2,p}(I_\delta^-, \mathbb{R}^2)}$. Taking a suitable sequence (δ_n) of positive numbers converging to zero as $n \rightarrow +\infty$, we obtain a limit map $\gamma_1 \in W^{2,p}(I, \mathbb{R}^2)$. Now, we glue together the various γ_1 , and define $\gamma : [0, 1] \setminus T \rightarrow \mathbb{R}^2$ as:

$$\gamma(t) := \gamma_1(t) \quad \text{if } t \in I.$$

Since γ_n is Lipschitz continuous in $[0, 1]$, with a Lipschitz constant independent of n (recall (11.39) and (11.40)), observing that $\lim_{\varepsilon \rightarrow 0^+} \gamma(t_j - \varepsilon) = \lim_{\varepsilon \rightarrow 0^+} \gamma(t_j + \varepsilon)$ for any $j \in \{1, \dots, l\}$, we have

$$\gamma \in \text{Lip}([0, 1], \overline{\Omega}) \cap W^{2,p}([0, 1] \setminus T, \mathbb{R}^2).$$

In addition, for any I we have $\gamma \in C^1(\overline{I}, \overline{\Omega})$, and there exists a constant $c \neq 0$ independent of I such that $|\gamma'| = c$ in I .

Now, let I_1 and I_2 be two consecutive intervals of $[0, 1] \setminus T$, hence with $\sup I_1 = \inf I_2$, and let $I_{1,n}, I_{2,n}$ be the corresponding intervals considered above, so that $\sup I_{1,n} \leq \inf I_{2,n}$. We notice that it may happen $\sup I_{1,n} < \inf I_{2,n}$, a case which may occur in case of collision of two or more cusp parameters of γ_n . Now, we recall that the total number of cusp parameters is uniformly bounded; in addition, if $\tau_n < \sigma_n$ are two distinct consecutive cusp parameters belonging to $\text{par}_{\text{cusp}}(\gamma_n) \cap [\sup I_{1,n}, \inf I_{2,n}]$, we have that $\lim_{n \rightarrow +\infty} |\tau_n - \sigma_n| = 0$. The same argument leading to formula (11.45), below, shows that the variation of the angle made by $\frac{\gamma'_n}{|\gamma'_n|}$ in the interval (τ_n, σ_n) tends to zero as $n \rightarrow +\infty$.

In conclusion, as a consequence of (11.43), at a point of T either the unit tangent vector to γ is continuous (and in this case this point can be eliminated from T), or it jumps of an angle belonging to $\{\pi, -\pi\}$. \square

Proposition 11.B.4 (Characterization of the Limit Cusp Parameters Set) *Following the notation of (11.42), for any $t \in \{t_1, \dots, t_l\} = T$ define²³*

$$\mathcal{J}_t := \{j \in \{1, \dots, l\} : \lim_{n \rightarrow +\infty} t_{j,n} = t\}.$$

Then

$$\text{par}_{\text{cusp}}(\gamma) = \{t \in T : \#\mathcal{J}_t \text{ is odd}\}. \quad (11.44)$$

²²See [4, Theorem 8.8 and Proposition 3.5].

²³That is, at t there is a collision of $\#\mathcal{J}_t$ cusp parameters of γ_n as $n \rightarrow +\infty$.

In particular, if $t \in T$ is such that $\#\mathcal{J}_t$ is even, then γ is of class $W^{2,p}$ (and hence C^1) locally around t : in this case, an even number of cusp parameters collide, leaving a point $t \in T$ around which γ is $W^{2,p}$, as in Fig. 11.9.

Proof In order to show (11.44), we take $\varepsilon > 0$ small enough; we want to compute the rotation angle between $\frac{\gamma'(t-\varepsilon)}{|\gamma'(t-\varepsilon)|}$ and $\frac{\gamma'(t+\varepsilon)}{|\gamma'(t+\varepsilon)|}$. From (ii) of Definition 11.B.2, it is sufficient to estimate the rotation angle between $\frac{\gamma'_n(t-\varepsilon)}{|\gamma'_n(t-\varepsilon)|}$ and $\frac{\gamma'_n(t+\varepsilon)}{|\gamma'_n(t+\varepsilon)|}$ for large values of $n \in \mathbb{N}$. This angle is the sum of three contributions: the first one, denoted by $C_{1,n}$, is due to the cusps, the second one, denoted by $C_{2,n}$, due to each of the arcs between two consecutive parameters $t_{i,n}, t_{i+1,n}$ belonging to \mathcal{J}_t , and the third one, originated from the rotation angle between $t - \varepsilon$ and the first cusp parameter in $(t - \varepsilon, t + \varepsilon)$, and the rotation angle between the last cusp parameter in $(t - \varepsilon, t + \varepsilon)$ and $t + \varepsilon$. This latter contribution turns out to be of order $\mathcal{O}(\varepsilon)$, uniformly with respect to n . In addition, the first contribution is an integer multiple of π , more precisely

$$C_{1,n} = \pi \#\mathcal{J}_t.$$

Let us show that

$$\lim_{n \rightarrow +\infty} C_{2,n} = 0. \quad (11.45)$$

Choose, as usual, a continuous function $\theta_n : [0, \mathcal{L}] \rightarrow \mathbb{R}$ in such a way that $\theta_n(t)$ is the oriented angle between the first coordinate axis and the tangent vector of γ_n at $t \in [0, \mathcal{L}]$. Denoting by $\ell(\gamma_n; (t_{i,n}, t_{i+1,n}))$ the length of $\gamma_n((t_{i,n}, t_{i+1,n}))$, using the Hölder inequality we have

$$\left(\int_{l(\gamma_n; (t_{i,n}, t_{i+1,n}))(t_{i,n}, t_{i+1,n})} |\dot{\theta}_n| ds \right)^{p'} \leq \ell(\gamma_n; (t_{i,n}, t_{i+1,n})) \left(\int_{(0, \ell(\gamma_n))} |\kappa_{\gamma_n}|^p ds \right)^{\frac{p'}{p}}. \quad (11.46)$$

Observe now that, from (11.35),

$$\sup_{n \in \mathbb{N}} \left(\int_{(0, \ell(\gamma_n))} |\kappa_{\gamma_n}|^p ds \right)^{\frac{p'}{p}} < +\infty, \quad (11.47)$$

and that

$$\lim_{n \rightarrow +\infty} \ell(\gamma_n; (t_{i,n}, t_{i+1,n})) = 0.$$

Hence, being

$$C_{2,n} \leq \left(\int_{(t_i, n, t_{i+1}, n)} |\dot{\theta}_n| ds \right)^{p'},$$

formula (11.45) follows from (11.46) and 11.47. We then conclude the proof of our claim (11.44), observing that if $\#\mathcal{J}_t$ is even, then in view of the previous arguments, the unit tangent vector to γ does not jump at t , namely γ is locally of class \mathcal{C}^1 in a neighbourhood of t (and therefore also of class $W^{2,p}$). \square

We are now in a position to generalize the previous definitions and results to the case of systems of curves.

Definition 11.B.5 (Weak Convergence of Systems of pwr_c^p Curves) We say that the sequence $(\Gamma_n) \subset \text{sys}(\text{pwr}_c^p)([0, \mathcal{L}], \Omega)$ weakly converges to $\Gamma = \{\gamma_1, \dots, \gamma_h\} \in \text{sys}(\text{pwr}_c^p)([0, \mathcal{L}], \bar{\Omega})$ in pwr_c^p if the number of curves of each system Γ_n equals the number of curves of Γ , i.e., $\Gamma_n = \{\gamma_1^n, \dots, \gamma_h^n\}$ for any $n \in \mathbb{N}$, and for any $i \in \{1, \dots, h\}$ the sequence (γ_i^n) weakly converges to γ_i in pwr_c^p .

The next theorem generalizes Theorem 11.B.3; we omit the proof, which is the same as that of Theorem 11.B.3, and is obtained reasoning separately for the sequences of curves composing the approximating systems, possibly taking into account Lemma 11.B.1.

Theorem 11.B.6 (Compactness in $\text{sys}(\text{pwr}_c^p)$) Let $m \in \mathbb{N}$ be given. For any $n \in \mathbb{N}$ let $\Gamma_n = \{\gamma_1^n, \dots, \gamma_m^n\} \in \text{sys}(\text{pwr}_c^p)([0, \mathcal{L}], \Omega)$ be a system of pwr_c^p curves parametrizing a complete labelled contour graph of class $\mathcal{W}_{\text{graph}}^{2,p}$. Suppose that

$$\sup_{n \in \mathbb{N}} \left(\beta \ell(\Gamma_n) + \lambda \mathcal{K}(\Gamma_n) + \sigma \# \text{par}_{\text{cusp}}(\Gamma_n) \right) < +\infty.$$

Then there exists $l \in \{1, \dots, m\}$ independent of n such that, possibly reordering the curves composing Γ_n , the system $\{\gamma_1^n, \dots, \gamma_l^n\}$ admits a subsequence weakly converging in pwr_c^p to a system $\{\gamma_1, \dots, \gamma_l\} \in \text{sys}(\text{pwr}_c^p)([0, \mathcal{L}], \bar{\Omega})$, and, if $l < m$, $\lim_{n \rightarrow +\infty} \ell(\gamma_j^n) = 0$ for any $j \in \{l+1, \dots, m\}$.

The system $\{\gamma_1, \dots, \gamma_l\}$ consists of a number l of curves in general smaller than the number m of curves composing Γ_n , because of the possible presence of curves in Γ_n having infinitesimal length as $n \rightarrow +\infty$. Observe also that there is no claim that the limit system is a parametrization of a complete labelled contour graph of class $\mathcal{W}_{\text{graph}}^{2,p}$.

We conclude this appendix with a definition, useful in the study of the functional $\bar{\mathcal{F}}$.

Definition 11.B.7 (Limit System of Complete Contour Graphs) We say that Γ is a limit system of complete labelled contour graphs of class $\mathcal{W}_{\text{graph}}^{2,p}$ if Γ is the weak limit of a sequence (Γ_n) of oriented parametrizations of complete labelled contour graphs of class $\mathcal{W}_{\text{graph}}^{2,p}$.

References

1. Ambrosio, L., Fusco, N., Pallara, D.: Functions of Bounded Variation and Free Discontinuity Problems. Clarendon Press, Oxford (2000)
2. Bellettini, G., Beorchia, V., Paolini, M.: Topological and variational properties of a model for the reconstruction of three-dimensional transparent images with self-occlusions. J. Math. Imaging Vision **32**, 265–291 (2008)
3. Bellettini, G., Dal Maso, G., Paolini, M.: Semicontinuity and relaxation properties of a curvature depending functional in 2D. Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **20**, 247–299 (1993)
4. Brezis, H.: Functional Analysis and Partial Differential Equations. Springer, New York (2010)
5. Bressan, A.: Lecture Notes on Functional Analysis. With Applications to Linear Partial Differential Equations. Graduate Studies in Mathematics, vol. 143. American Mathematical Society, Providence (2013)
6. Dal Maso, G.: An Introduction to Γ -Convergence. Birkhäuser, Boston (1993)
7. Evans, L.C., Gariepy, R.F.: Measure Theory and Fine Properties of Functions. CRC Press, Boca Raton (1992)
8. Kuratowski, K.: Topology. Academic Press, New York (1968)

Nomenclature

$\#$	Function measuring the cardinality of a set, page 33
$<$	Order relation in the Nitzberg–Mumford model, page 6
$\int_{\Omega} Df $	Total variation of the BV function f , page 330
$2\mathbb{N}$	Nonnegative even natural numbers, page 16
a_a^{\pm}	Arcs “above” in a crossing, page 60
a_b^{\pm}	Arcs “below” in a crossing, page 60
$A(u, C, P)$	A curvature depending functional with penalization of singular points, page 11
$\text{appcon}(\varphi)$	Apparent contour of the map φ , page 31
$\text{arcs}(\text{appcon}(\varphi))$	Arcs of an apparent contour, page 32
$\text{arcs}(G_{\Sigma})$	Arcs of G_{Σ} , Sect. 3.2, page 58
\mathfrak{b}	Definition 7.1.1, page 159
B	Move B, Definition 6.1.1, page 133
$\text{background}(K)$	Remark 4.1.10, page 76
$\mathfrak{B}(\text{appcon}(\varphi))$	Bennequin’s type invariant, Definition 7.1.2, page 159
BM_1	A curvature depending functional for occlusions, II, page 13
$BM(u, \chi)$	A curvature depending functional for occlusions, I, page 12
$BV(\Omega, 2\mathbb{N})$	Functions of bounded variation in Ω taking values in the even natural numbers (zero included), page 15
C	Move C, Definition 6.1.1, page 133
$\mathcal{C}^{\infty}(\mathcal{X}, \mathcal{Y})$	Section 2.1, page 26
c^{-}	Incoming arc at cusp c , page 59
c^{+}	Outgoing arc at cusp c , page 59

$\text{crit}(\varphi)$	Critical set of the map φ , page 28
$\text{crit}_{\text{or}}(\text{appcon}(\varphi))$	Oriented critical points of $\text{appcon}(\varphi)$, page 159
$\text{crossings}(\text{appcon}(\varphi))$	Crossings of the apparent contour $\text{appcon}(\varphi)$, page 32
$\text{crossings}(G_{\Sigma})$	Crossings of G_{Σ} , Sect. 3.2, page 58
$\text{crossings}_{\text{or}}(\text{appcon}(\varphi))$	Symbols for oriented crossings of $\text{appcon}(\varphi)$, page 158
$\text{cusps}(\text{appcon}(\varphi))$	Cusps of the apparent contour $\text{appcon}(\varphi)$, page 32
$\text{cusps}(G_{\Sigma})$	Cusps of G_{Σ} , Sect. 3.2, page 58
$\text{cusps}_{\text{or}}(\text{appcon}(\varphi))$	Symbols for oriented cusps of $\text{appcon}(\varphi)$, page 158
d	Limit of a sequence of labellings, page 332
$\deg(p)$	Degree of the node p of the plane graph H , page 74
$\text{Diff}(\mathcal{Z})$	Section 2.1, page 25
$\text{Diff}_c(\mathbb{R}^n)$	Diffeomorphisms of \mathbb{R}^n with compact support, Sect. 2.3, page 25
$\text{Diff}^+(\mathbb{S})$	Set of all positive diffeomorphisms of \mathbb{S}^2 , page 37
$\text{Dom}(\mathcal{F})$	Domain of the action functional \mathcal{F} , page 325
d_{Σ}	Value of the labelling on the arc a , page 65
Δ	Alexander polynomial, page 182
$[E]$	Depth-equivalence class of the shape E , page 102
\mathcal{E}	Fundamental ideal, page 188
E	3D scene, page 53
e	Embedding, usually from a two-manifold in \mathbb{R}^3 , page 125
$\text{Emb}(\mathcal{X}, \mathcal{Y})$	Section 2.1, page 30
$\text{embsign}(c)$	Embedding sign of a cusp, Definition 8.1.2, page 197
$\text{ext}(H)$	External region of the plane graph H , page 74
$\text{ext}(G_{\Sigma})$	External region of G_{Σ} , Sect. 3.2, page 58
ε_d	Elementary ideals, Definition 7.7.2, page 186
f	Limit of a sequence of twice winding numbers, page 332
$\mathcal{F} = \mathcal{F}(f, d, u)$	Functional on apparent contours, page 17
f_{φ}	Definition 2.2.5, see equation (2.2), page 33
f_{inn}	Value of f in R_{max} , page 66
f_{max}	Value of f in R_{max} , page 66
f_{min}	Value of f in R_{min} , page 66
$\overline{\mathcal{F}}$	Sequential lower semicontinuous envelope of \mathcal{F} , page 332

f^\pm	The two traces of the BV function f on J_f , page 324
f_Σ	Function counting the total number of intersections between Σ and a light ray, Sect. 3.3, page 59
$f_\Sigma(a)$	Value of f_Σ on arc a , Sect. 3.3, page 61
$f_\Sigma(R)$	Value of f_Σ on region R , Sect. 3.3, page 59
ϕ_2	Quadratic integrand in the elastica functional, page 6
ϕ_{NM}	Integrand in the Nitzberg–Mumford functional, page 7
ϕ_p	p -growing integrand, page 12
$\varphi(\text{crit}(\varphi))$	Critical value set, page 29
g	Grey level, page 1
G	A generic finitely presented group, page 180
G_Σ	Apparent contour of a factorized map, Sect. 3.2, page 58
G'	Commutator normal subgroup of G , page 181
(G, d)	Complete labelled contour graph, page 79
(G, f, d)	Complete labelled contour graph, page 79
g'_i	Mean value of g on R'_i in the Nitzberg–Mumford model, page 7
$(G_\Sigma, f_\Sigma, d_\Sigma)$	Complete labelled contour graph originated from Σ , page 79
H	\mathbb{R}^n -ambient isotopy, page 36
h	\mathbb{R}^2 -ambient isotopy, page 37
\mathcal{H}^1	One-dimensional Hausdorff measure in \mathbb{R}^2 , page 4
H_t	Time-slice of H , page 36
h_t	Time-slice of h , page 37
$\text{im}(\gamma)$	Image of the curve γ , page 324
$\text{im}(\Gamma)$	Image of the system Γ , page 345
INVB	Inverse of move B, page 247
INVC	Inverse of move C, page 248
INVK0	Inverse of move K0, page 247
INVK1b	Inverse of move K1b, page 247
INVK1	Inverse of move K1, page 247
INVK2	Inverse of move K2, page 247
INVL	Inverse of move L, page 247
INVS	Inverse of move S, page 248
J	Jacobian of a presentation, page 184
J_f	Jump set of the function f , page 324
J_u	Jump of the BV function u , page 5
K	Move K, Definition 6.1.1, page 133

$K0, K1, K1b, K2$	Simple moves of type K , page 134
$\mathcal{K}(\gamma)$	Curvature-depending functional on the curve γ , page 343
$\mathcal{K}(\Gamma)$	Curvature-depending functional on the system Γ , page 345
K_μ	Localization of K in the half-plane U_μ^+ , page 84
κ	Curvature, page 7
κ_γ	Curvature of the curve γ , page 343
L	Move L , Definition 6.1.1, page 133
L	Ring of Laurent polynomials in one or more indeterminates and integer coefficients, page 185
$\ell(\gamma)$	Length of the curve γ , page 343
$\ell(\Gamma)$	Length of the system Γ , page 345
m	Descriptive map, page 46
M	A closed two-dimensional manifold, page 31
$\text{maxmin}_{\text{or}}(\text{appcon}(\varphi))$	Symbols for oriented local maxima/minima of $\text{appcon}(\varphi)$, page 158
m_γ	Multiplicity of γ , page 343
m_λ	λ -slices of the descriptive map m , page 44
$MS(u)$	Mumford–Shah functional, weak form, page 5
$MS(u, K)$	Mumford–Shah functional, strong form, page 4
$\mathcal{M}(u, \Omega \setminus C)$	Part of the functional \mathcal{F} containing the Dirichlet integral, page 324
\mathbb{N}	Nonnegative integer numbers, page 33
\mathbb{N}_+	Positive natural numbers, page 104
NM	Nitzberg–Mumford functional, page 7
\widetilde{NM}	Modification of the Nitzberg–Mumford functional, page 8
$\text{nodes}(\text{appcon}(\varphi))$	Nodes (or vertices) of an apparent contour, page 32
$\text{nodes}(H)$	Nodes (or vertices) of the plane graph H , page 74
$\text{nodes}(\text{appcon}(\varphi))$	Nodes (or vertices) of an apparent contour, page 32
$\text{nodes}(G_\Sigma)$	Nodes of G_Σ , Sect. 3.2, page 58
o	Trivial map $o : G \rightarrow \{1\}$, page 185
Ω	Screen, typically $\Omega = (0, 1) \times (0, 1)$, page 1
$\text{par}_{\text{cusp}}(\gamma)$	Cusp parameters, Definition 11.A.1, page 343
$\text{par}_{\text{cusp}}(\Gamma)$	Cusp parameters of the system Γ , page 345
$\text{Part}_{\text{overlap}}(\mathbb{R}^2)$	Set of all finite overlapping partitions of \mathbb{R}^2 , page 6
$\mathcal{P}(\mathbb{N})$	Set of all subsets of \mathbb{N} , page 332
$\mathcal{P}(\mathbb{R})$	Set of all subsets of \mathbb{R} , page 113

pwr_c^p	Definition 11.A.1, page 342
$\text{pwr}_c^p([0, \mathcal{L}], \Omega)$	Definition 11.A.1, page 342
$\text{pwr}_c^p([0, \mathcal{L}], \overline{\Omega})$	Definition 11.A.1, page 342
π	Orthogonal projection $\mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}^2$ on the first factor, Sect. 3.2, page 56
\mathcal{Q}	Container of all 3D scenes, $\Omega \times I$, page 2
(R, r)	Stratum, Definition 3.3.1, page 61
R_{inn}	Region inside a cusp, page 66
R'_i	Visible part of R_i in the Nitzberg–Mumford model, page 6
$ R_i $	Lebesgue measure of R_i , page 7
R_{max}	Region inside a cusp, page 66
R_{min}	Region outside a cusp, page 66
R_{out}	Region outside a cusp, page 66
\mathcal{S}	Source three-manifold in the completeness theorem, see also (6.7), page 136
S	Move S, Definition 6.1.1, page 133
$S_1(\varphi)$	Section 2.1, page 28
\mathbb{S}^1	Unit circle, page 27
$S_{1_2}(\varphi)$	Section 2.1, page 28
$S_{1_3}(\varphi)$	Section 2.1, page 29
\mathbb{S}^2	Unit two-dimensional sphere, sometimes considered as the compactification of \mathbb{R}^2 , page 37
\mathbb{S}^3	Compactification of \mathbb{R}^3 , page 40
$SBV(\Omega)$	Special functions of bounded variation in Ω , page 5
$\text{Stable}(\mathcal{X}, \mathcal{Y})$	Stable maps from \mathcal{X} to \mathcal{Y} , page 26
$\text{sys}(\text{pwr}_c^p)$	Systems of pwr_c^p curves, page 344
$\text{sys}(\text{pwr}_c^p)([0, \mathcal{L}], \Omega)$	Systems of $\text{pwr}_c^p([0, \mathcal{L}], \Omega)$ curves, page 344
$\text{sys}(\text{pwr}_c^p)([0, \mathcal{L}], \overline{\Omega})$	Systems of $\text{pwr}_c^p([0, \mathcal{L}], \overline{\Omega})$ curves, page 344
Σ	Boundary of a 3D shape, Sect. 3.2, page 56
\mathcal{T}	Topological manifold, page 108
\mathcal{T}	Target three-manifold in the completeness theorem, see also (6.8), page 136
T	Move T, Definition 6.1.1, page 133
Θ	Homeomorphism preserving the order on the fibres, Definition 5.1.2, page 102
U_λ^+	Half-plane above the line at height λ , page 84
$\text{Unstable}(M, \mathbb{R}^2)$	Unstable maps in $\mathcal{C}^\infty(M, \mathbb{R}^2)$, page 162
$\text{vis}(G)$	Visible part of a complete labelled contour graph, page 79
$\text{vis}(J_f)$	Visible part of J_f , page 325
$W^{1,p}(\mathbf{I})$	Sobolev space, page 323
$W^{2,p}(\mathbf{I})$	Sobolev space, page 323

$w(\text{appcon}(\varphi), \cdot)$	Winding number of $\text{appcon}(\varphi)$, Definition 2.2.8, page 34
\mathcal{W}	Curvature-depending part of the functional \mathcal{F} , page 326
$w(\gamma, \cdot)$	Winding number (or index) of x with respect to γ , page 344
$w(\Gamma, \cdot)$	Winding number of the system Γ , page 345
$w(H, \cdot)$	Winding number of the contour graph H , page 77
w_{left}	Definition 7.3.2, page 170
w_{right}	Definition 7.3.2, page 170
X	Free group with n generators for some $n \in \mathbb{N}$, page 183
$\{X_0(F), X_1(F), X_2(F), X_3(F)\}$	Stratification of \mathcal{S} , Sect. 6.2.1, page 137
$\{Y_1(F), Y_2(F), Y_3(F)\}$	Stratification of \mathcal{T} , Sect. 6.2.1, page 138
\mathbb{Z}	Ring of integer numbers, page 33
$\mathbb{Z}X$	Group ring, page 183
$[\xi^-, \xi^+]$	Smooth two-valued function, page 113

Index

- Abelianized, 181
- Abelianizing map, 185
- Adjacency, 198
- Adjacency of strata in an open set, 199
- Adjacent arc, 74
- Admissible background, 76
- Alexander polynomial, 181, 182, 262
- Ambient isotopic apparent contours, 41
- Ambient isotopic embeddings, 132
- Ambient isotopic labelled apparent contours, 69
- Ambient isotopic surfaces, 132
- Ambient isotopy, 36
- Ambient isotopy with compact support, 36
- Apparent contour, 31
 - without labelling, or unlabelled apparent contour, 31
 - orientation, 34
- Arc
 - of an apparent contour, 32
 - incoming at a cusp, 196
 - outgoing at a cusp, 196
- Arnold, V.I., xvi

- Background, 76
- Beak-to-beak (rule B), 133
- Bean, 96
- Bennequin type invariant, 159, 275, 280
- Betti numbers, 181
- Bing's example, 305
- Body, 53
- Borromean rings, 266
- Boy surface, 277, 313

- Canonization procedure, 249
- Cell complex, 176
- Closed arc, 32
- Closed arcs of a graph, 74
- Closed manifold, 26
- Command
 - abelianizedfundamental, 263, 289
 - addsphere, 286
 - afg, 263, 289
 - alexander, 263, 289
 - canonify, 287
 - ccordering, 284
 - ccparent, 284
 - cellcomplex, 258, 288
 - characteristic, 259, 281
 - compare, 258, 284
 - contour rule, 240
 - coorientation, 284
 - countcc, 284
 - 3devert, 286
 - evert, 286
 - extractcc, 260, 286
 - fg, 229, 289
 - frontback, 286
 - fundamental, 289
 - gluearcs, 285
 - icharacteristic, 259
 - ifg, 230, 260, 289
 - info, 226, 280
 - insidecomplex, 258, 288
 - insidefundamental, 289
 - iscontour, 281
 - islabelled, 281
 - knot2morse, 287

- leftright, 286
- linkingnumber, 289
- listinv, 283
- listinvl, 282
- listinvs, 282
- listma, 281
- liststrata, 284
- mendes, 284
- mergearcs, 247, 285
- ocharacteristic, 259
- ofg, 261, 289
- outsidecomplex, 258, 288
- outsidefundamental, 289
- pinchneck, 285
- print, 287
- printmorse, 287
- punchole, 285
- puncture, 285
- removecc, 286
- removehole, 285
- removesphere, 286
- rule, 228
- rules, 228, 281
- scharacteristic, 259
- sum, 287
- swallowtail, 285
- union, 287
- wrap, 286
- wrinkle, 285
- Commutator, 181
- Completable contour graph, 80
- Complete contour graph, 77
- Complete contour graph of class $\mathcal{W}_{\text{graph}}^{2,p}$, 324
- Complete labelled contour graph, 79
- Complete labelled contour graph of class $\mathcal{W}_{\text{graph}}^{2,p}$, 325
- Completeness of a Morse description of an apparent contour, 43
- Completeness theorem, xix, 154
- Completion, 80
- Completion theorem, xv, 80
- Component of a complete contour graph of class $\mathcal{W}_{\text{graph}}^{2,p}$, 343
- Component of an apparent contour, 31
- Connectable cusps in an open set, 200
- Connectable strata in an open set, 199
- Consecutive arc, 74
- Consistent labelling, 78
- Contour graph, 75
- Convergence on $\text{Dom}(\mathcal{F})$, 327
- Critical curve, 58
- Critical set, 28
- Critical value set, 29
- Crossing of a contour graph, 75
- Crossing of an apparent contour, 32
- Curvature-depending functionals, 11
- Cusp
 - of an apparent contour, 32
 - decreasing, 196
 - increasing, 196
 - negative, 35, 68
 - pointing into a stratum, 200
 - positive, 35, 68
 - simple, or ordinary, or semicubic, 32
- Cusp parameter, 343
- Cusp-fold (rule C), 133
- Cutting arc, 178, 289
- Cutting node, 178
- Dangling arc, 84
- Deficiency, 182
- Degree of a node of a graph, 74
- Depth equivalence, 102
- Depth-first search, 252
- Descriptive map
 - of an apparent contour, 46
 - of a visible contour graph, 82
- Diffeomorphically equivalent apparent contours, 42
- Diffeomorphism supported in a compact set, 25
- Discriminant hypersurface in $\mathcal{C}^\infty(M, \mathbb{R}^2)$, 162
- Domain of the action functional \mathcal{F} on labelled graphs, 325
- Double curve, 28
- Elementary ideals, 186, 263
- Elementary moves, xix
- Embedding sign of a cusp, 197
- Emerging (or transversal) arc at a T-junction, 76
- Ending terminal point, 83
- Ending T-junction, 83
- Equivalent labelled apparent contours, 69
- Equivalent maps, 26
- Euler-Poincaré characteristic, 275
- Excellent map, 28
- Exponent sum matrix, 181
- Extended arc, 47
- External region of an apparent contour, 33
- Fiber, 64
- Fiber of E , 102
- Fidelity term, 4

- Finitely presented group, 180
- Finiteness of a Morse description of an
 - apparent contour, 43
- Fold, 28, 29, 137
- Fold curves, 28
- Fold map, 195
- Fox, 183
- Free derivative, 183
- Free differential calculus, 183
- Functional \mathcal{F} , 326
- Fundamental ideal, 188

- General position of an embedded surface with
 - respect to a projection, 56
- Generic family of horizontal Morse lines, 45
- Generic Morse line, 44, 46, 82
- Generic position of an open set with respect to
 - an apparent contour, 199
- Group ring, 183

- Haefliger sphere, 35, 312
- Handlebody, 238
- Head of an arc, 74
- Height function, 43, 44
- Horizontal adjacency of a stratum, 198
- Horizontal and vertical adjacency between
 - strata, 198
- Horizontal glueing along an arc, 105

- Image of a curve, 324
- Impossible graph, 67
- Interior terminal point, 75
- Inverse moves, see rules, inverse, 133
- Invisible arc, 79
- Invisible part of a complete labelled contour
 - graph, 79
- Island, 244
- Isomorphism problem, 180
- Isotopic embeddings, 142
- Isotopic surfaces, 142
- Isotopy, 142
 - \mathbb{R}^n -ambient, 36

- Kanizsa triangle, 18
- Kasane (rule K), 133
- Klein bottle, 315
- Knot description, 238
- Knot group, 180
- Knots in a solid torus, 301
- Kuratowski convergence, 327

- Labelled apparent contour, 79
- Labelling, 65
- Labelling on a complete contour graph, 78
- Laurent polynomials, 185, 263
- Legendrian lift, 160
- Level-preserving map, 144
- Limit system of complete contour graphs of
 - class $\mathcal{W}_{\text{graph}}^{2,p}$, 353
- Link group, 180
- Lip (rule L), 133
- List of Reidemeister-type moves, 133
- Localized completion of a contour graph in a
 - halfplane, 84
- Loop of a graph, 74
- Loop of an apparent contour, 32
- Lower semicontinuous envelope (or relaxation)
 - of \mathcal{F} , 332

- Marked point, 47
- Marker, 47
- Mendes graph, 274
- Millett curve, 278, 314
- Milnor curve, 278, 313
- Morse description, 226, 236
 - of the visible contour, 212
- Morse event, 43
- Morse function, 26
- Morse lines, 44
- Moves, 133. *See also* Rules
- Multiplicity of a curve, 343
- Mumford-Shah functional, 4

- Nitzberg-Mumford functional, 6

- Occluding arc, 75
- Opposite arc, 74
- Option
 - finfinity, 287
 - foxd, 263
 - help, 290
 - in, 260, 290
 - mendes_ge, 291
 - nobasecanonify, 290
 - nocanonify, 258
 - nosimplify, 290
 - out, 290
 - preabelian, 262, 290
 - Q, 268
 - seed, 290, 291
 - shuffle, 290
 - stratum, 285

- ti, 244, 247
- transfer_islands, 290
- verbose, 290
- version, 290
- a, 285
- n, 298
- q, 290
- Q, 290, 293
- r, 285
- v, 290
- z, 298
- out, 261
- Outline of a scene, 55
- Parity of cusps, 67
- Piecewise regular $W^{2,p}$ curves with cusps, 342
- Pinch point, 29
- Pleat, 29, 137
- Preabelian presentation, 181
- Rank, 181
- Reconstruction theorem, xvi, 102
- Recovery problem, 14
- Region description, 233
 - canonization, 253
- Region of an apparent contour region, 33
- Regions graph, 250
- Regular point, 28
- Regular reparametrization, 344
- Reidemeister-equivalence, 136
- Reidemeister equivalence in an open set, 201
- Reidemeister-equivalence of apparent
 - contours, 201
- Reidemeister-type moves, xix, 133
- Retinal plane, 54
- Rule
 - B, 133
 - C, 133, 243
 - composite A1, 246
 - composite A2, 247
 - composite CR0L, 244
 - composite CR0Lb, 244
 - composite CR0R, 244
 - composite CR0Rb, 244
 - composite CR0, 244
 - composite CR1, 245
 - composite CR2, 245
 - composite CR3, 246
 - composite CR3L, 246
 - composite CR3R, 246
 - a composite CR4, 246
 - composite CR4L, 246
 - composite CR4Lb, 246
 - composite CR4R, 246
 - composite CR4Rb, 246
 - inverse INVB, 247
 - inverse INVK0, 247
 - inverse INVK1, 247
 - inverse INVK1b, 247
 - inverse INVK2, 247
 - K, 133, 241
 - K0, 134, 241
 - K1, 134, 242
 - K1b, 134, 242
 - K2, 134, 242
 - L, 133, 242
 - S, 133, 243
 - T, 133, 242
- Rules, xix
- Schoenflies, 309
- Script
 - contour_describe.sh, 297
 - contour_interact.sh, 296
 - contour_transform.sh, 299
- Showcontour, 293
- Singular set. *See* Critical set
- Smith normal form, 181
- Sphere with a cave, I: reconstruction from the
 - visible part, 98
- Stable map, 26
- Stable scene, 56
- Standard output: redirect, 214
- Starting terminal point, 83
- Starting T-junction, 83
- Stratification of an apparent contour, 43
- Stratification of the plane induced by an
 - apparent contour, 43
- Stratum of a region, 61
- Swallow's tail (rule S), 29, 133, 137
- System of pwr_c^p curves, 344
- Tail of an arc, 74
- Tame knot, 238
- Tame link, 238
- Terminal point, 75
- Theorem
 - \mathbb{R}^2 -ambient isotopies and diffeomorphisms, 42, 70
 - \mathbb{R}^3 -ambient isotopies and diffeomorphisms, 132
 - coincidence of the two Bennequin's type invariants, 163
 - compactness in $\text{sys}(\text{pwr}_c^p)$, 352

- compactness in $\mathcal{W}_{\text{graph}}^{2,p}$, 348
- completeness, 145
- completion, 80
- density of stable maps, 31
- elimination of cusps, 206
- elimination of cusps in an open set, 201
- equivalence to stability in the two-dimensional case, 29
- existence of a 3D-shape with given apparent contour, 102
- plane or space isotopies with compact support and diffeomorphisms, 37
- reconstruction in terms of maps, 125
- sequential lower semicontinuity of \mathcal{F} , 328
- total characteristic from the apparent contour, 172
- uniqueness of a 3D-shape with given apparent contour, 103
- Thom, xvi, 26
- Three-dimensional scene, 53
- Tietze transformation, 180
- Torsion, 181
- Torus, I: reconstruction from the visible part, 97
- Torus, II: reconstruction from the visible part, 98
- Transversal crossing (or crossing), 78
- Transversal (or emerging) arc at a T-junction, 76
- Triple point, 29
- Triple point (rule T), 133
- Trivial map, 185
- Unknotting theorem, 176, 302
- Unlabelled apparent contour, i.e., an apparent contour without labelling, 230
- Vertical adjacency of a stratum, 198
- Vertical glueing along an arc, 105
- Visible arc, 79
- Visible contour graph, 76
- Visible map, 214
- Wall, xvi
- Weak convergence of systems of $\mathcal{W}_{\text{graph}}^{2,p}$ curves, 352
- Weak convergence of $\mathcal{W}_{\text{graph}}^{2,p}$ curves, 347
- Whitehead link, 304
- Whitney, xvi, 28
- Whitney umbrella, 154, 278
- Winding number, 34, 77