

The Shapes of Things

Advances in Design and Control

SIAM's Advances in Design and Control series consists of texts and monographs dealing with all areas of design and control and their applications. Topics of interest include shape optimization, multidisciplinary design, trajectory optimization, feedback, and optimal control. The series focuses on the mathematical and computational aspects of engineering design and control that are usable in a wide variety of scientific and engineering disciplines.

Editor-in-Chief

Ralph C. Smith, North Carolina State University

Editorial Board

Athanasiadis C. Antoulas, Rice University
Siva Banda, Air Force Research Laboratory
John Betts, The Boeing Company (retired)
Stephen L. Campbell, North Carolina State University
Michel C. Delfour, University of Montreal
Fariba Fahroo, Air Force Office of Scientific Research

J. William Helton, University of California, San Diego
Arthur J. Krener, University of California, Davis
Kirsten Morris, University of Waterloo
John Singler, Missouri University of Science and Technology

Series Volumes

- Walker, Shawn W., *The Shapes of Things: A Practical Guide to Differential Geometry and the Shape Derivative*
Michiels, Wim and Niculescu, Silviu-Iulian, *Stability, Control, and Computation for Time-Delay Systems: An Eigenvalue-Based Approach, Second Edition*
Narang-Siddarth, Anshu and Valasek, John, *Nonlinear Time Scale Systems in Standard and Nonstandard Forms: Analysis and Control*
Bekiaris-Liberis, Nikolaos and Krstic, Miroslav, *Nonlinear Control Under Nonconstant Delays*
Osmolovskii, Nikolai P. and Maurer, Helmut, *Applications to Regular and Bang-Bang Control: Second-Order Necessary and Sufficient Optimality Conditions in Calculus of Variations and Optimal Control*
Biegler, Lorenz T., Campbell, Stephen L., and Mehrmann, Volker, eds., *Control and Optimization with Differential-Algebraic Constraints*
Delfour, M. C. and Zolésio, J.-P., *Shapes and Geometries: Metrics, Analysis, Differential Calculus, and Optimization, Second Edition*
Hovakimyan, Naira and Cao, Chengyu, *\mathcal{L}_1 Adaptive Control Theory: Guaranteed Robustness with Fast Adaptation*
Speyer, Jason L. and Jacobson, David H., *Primer on Optimal Control Theory*
Betts, John T., *Practical Methods for Optimal Control and Estimation Using Nonlinear Programming, Second Edition*
Shima, Tal and Rasmussen, Steven, eds., *UAV Cooperative Decision and Control: Challenges and Practical Approaches*
Speyer, Jason L. and Chung, Walter H., *Stochastic Processes, Estimation, and Control*
Krstic, Miroslav and Smyshlyaev, Andrey, *Boundary Control of PDEs: A Course on Backstepping Designs*
Ito, Kazufumi and Kunisch, Karl, *Lagrange Multiplier Approach to Variational Problems and Applications*
Xue, Dingyu, Chen, YangQuan, and Atherton, Derek P., *Linear Feedback Control: Analysis and Design with MATLAB*
Hanson, Floyd B., *Applied Stochastic Processes and Control for Jump-Diffusions: Modeling, Analysis, and Computation*
Michiels, Wim and Niculescu, Silviu-Iulian, *Stability and Stabilization of Time-Delay Systems: An Eigenvalue-Based Approach*
Ioannou, Petros and Fidan, Barış, *Adaptive Control Tutorial*
Bhaya, Amit and Kaszkurewicz, Eugenius, *Control Perspectives on Numerical Algorithms and Matrix Problems*
Robinett III, Rush D., Wilson, David G., Eisler, G. Richard, and Hurtado, John E., *Applied Dynamic Programming for Optimization of Dynamical Systems*
Huang, J., *Nonlinear Output Regulation: Theory and Applications*
Haslinger, J. and Mäkinen, R. A. E., *Introduction to Shape Optimization: Theory, Approximation, and Computation*
Antoulas, Athanasios C., *Approximation of Large-Scale Dynamical Systems*
Gunzburger, Max D., *Perspectives in Flow Control and Optimization*
Delfour, M. C. and Zolésio, J.-P., *Shapes and Geometries: Analysis, Differential Calculus, and Optimization*
Betts, John T., *Practical Methods for Optimal Control Using Nonlinear Programming*
El Ghaoui, Laurent and Niculescu, Silviu-Iulian, eds., *Advances in Linear Matrix Inequality Methods in Control*
Helton, J. William and James, Matthew R., *Extending H^∞ Control to Nonlinear Systems: Control of Nonlinear Systems to Achieve Performance Objectives*

The Shapes of Things

*A Practical Guide to Differential Geometry
and the Shape Derivative*

Shawn W. Walker

Louisiana State University
Baton Rouge, Louisiana



Society for Industrial and Applied Mathematics
Philadelphia

Copyright © 2015 by the Society for Industrial and Applied Mathematics

10 9 8 7 6 5 4 3 2 1

All rights reserved. Printed in the United States of America. No part of this book may be reproduced, stored, or transmitted in any manner without the written permission of the publisher. For information, write to the Society for Industrial and Applied Mathematics, 3600 Market Street, 6th Floor, Philadelphia, PA 19104-2688 USA.

Trademarked names may be used in this book without the inclusion of a trademark symbol. These names are used in an editorial context only; no infringement of trademark is intended.

MATLAB is a registered trademark of The MathWorks, Inc. For MATLAB product information, please contact The MathWorks, Inc., 3 Apple Hill Drive, Natick, MA 01760-2098 USA, 508-647-7000, Fax: 508-647-7001, info@mathworks.com, www.mathworks.com.

Publisher	David Marshall
Acquisitions Editor	Elizabeth Greenspan
Developmental Editor	Gina Rinelli
Managing Editor	Kelly Thomas
Production Editor	Lisa Briggeman
Copy Editor	Nicola Cutts
Production Manager	Donna Witzleben
Production Coordinator	Cally Shrader
Compositor	Techsetters, Inc.
Graphic Designer	Lois Sellers

Library of Congress Cataloging-in-Publication Data

Walker, Shawn W., 1976-

The shapes of things : a practical guide to differential geometry and the shape derivative /
Shawn W. Walker, Louisiana State University, Baton Rouge, Louisiana.

pages cm. – (Advances in design and control ; 28)

Includes bibliographical references and index.

ISBN 978-1-611973-95-2

1. Symmetry (Mathematics) 2. Geometry. 3. Shapes. 4. Surfaces (Technology) 5. Geometry,
Differential. I. Title.

QA174.7.S96W35 2015

516.3'6-dc23

2015010487

I dedicate this to Mari and Jane



Contents

Preface	ix
1 Introduction	1
1.1 Differential Equations on Surfaces	1
1.2 Differentiating with Respect to Shape	1
1.3 Abstract vs. Concrete Presentation	5
1.4 Outline	6
1.5 Prerequisites	7
1.6 Notation	7
2 Surfaces and Differential Geometry	9
2.1 Preliminaries	9
2.2 The Parametric Approach	14
2.3 Regular Surface	18
2.4 The Tangent Space	27
2.5 Minimal Regularity?	31
3 The Fundamental Forms of Differential Geometry	33
3.1 First Fundamental Form	33
3.2 Second Fundamental Form	43
3.3 Conclusion	56
4 Calculus on Surfaces	59
4.1 Functions on Surfaces	59
4.2 Differential Operators on Surfaces	61
4.3 Other Curvature Formulas	71
4.4 Integration by Parts	74
4.5 Other Identities	77
4.6 PDEs on Surfaces	82
5 Shape Differential Calculus	87
5.1 Introduction	87
5.2 General Framework	87
5.3 Derivatives of Functions with Respect to Flow Map Φ_ϵ	90
5.4 Derivatives of Functions with Respect to Flow Map X_ϵ	91
5.5 Basic Identities	93
5.6 Shape Perturbation of Functionals	101

6	Applications	105
6.1	Minimal Surfaces	105
6.2	Surface Tension	107
6.3	Gradient Flows	113
6.4	Mean Curvature Flow	115
6.5	Image Segmentation	117
6.6	Conclusion	122
7	Willmore Flow	123
7.1	The Energy	123
7.2	Perturbation Analysis	124
7.3	Gradient Flow	133
A	Vectors and Matrices	137
A.1	Vector Operations	137
A.2	Matrix Operations	137
A.3	Vector and Matrix Identities	140
B	Derivatives and Integrals	141
B.1	Differential Formulas	141
B.2	Integral Formulas	142
	Bibliography	145
	Index	153

Preface

Form follows function. This old adage from art and architecture, credited to American architect Louis H. Sullivan, holds true. The shape of an object is intimately connected to its purpose. Nature provides many examples of this: the shape of a tree and its leaves to harvest light, the wings of a bird to fly, the body of a snake to slither, and the structure of the human heart to keep us alive. So good is this rubric that it finds application in modern design principles, e.g., the shapes of tools, the profile of an automobile, and the design of a bridge.

In an 1896 essay, Sullivan wrote

... form ever follows function and this is the law.

Sullivan means that form depends completely on function. But what about the reverse? If an object's shape changes, how is its function affected? Is the object's function improved? Is the object better? In other words, does it make sense to consider function as dependent on shape? In a certain context, yes. The main purpose of this book is to explain how to differentiate a function (in the calculus sense) with respect to a "shape variable."

This book is written to be as self-contained as possible. It can be read by undergraduates who have completed the usual introductory calculus-based math courses. It can be read by experts from other fields who wish to learn the fundamentals of differential geometry and shape differential calculus and apply them in their own disciplines. It also makes a useful reference text for a variety of shape differentiation formulas. Chapter 1 gives more details on the prerequisites, framework, and overall philosophy of the book.

This book started as a set of notes I had created for my own use. Over time, I continued to refine them and used them in a special topics course I taught at Louisiana State University (LSU) in Fall 2011. Eventually, after sharing the notes I realized their potential value to others and sought to create this book to make shape derivatives accessible to a broader audience.

Acknowledgments I would like to thank the following people for reading earlier versions of the text and making useful comments: Harbir Antil, Christopher B. Davis, and Antoine Laurain. I especially want to thank Mari Walker for proofreading the entire book. I thank the anonymous reviewers; their compliments and criticisms certainly improved the manuscript. I gratefully acknowledge prior support by NSF grants DMS-1115636 and DMS-1418994.

Chapter 1

Introduction

1.1 • Differential Equations on Surfaces

The purpose of this book is to present an overview of differential geometry, which is useful for understanding mathematical models that contain geometric partial differential equations (PDEs), such as the *surface* (or manifold) version of the standard Laplace equation. In particular, this requires the development of the so-called surface gradient and surface Laplacian operators. These are nothing more than the usual gradient ∇ and Laplacian $\Delta = \nabla \cdot \nabla$ operators, except they are defined on a surface (manifold) instead of standard Euclidean space (i.e., \mathbb{R}^n).

One advantage of this approach is that it provides alternative formulas for geometric quantities, such as the summed (mean) curvature, that are much clearer than the usual presentation of texts on differential geometry.

1.2 • Differentiating with Respect to Shape

The approach to differential geometry in this book is advantageous for developing the framework of *shape differential calculus*, which is the study of how quantities change with respect to changes of an independent “shape variable.”

1.2.1 • A Simple Example

The following example requires only the tools of freshman calculus. Let $f = f(r, \theta)$ be a smooth function defined on the disk of radius R in terms of polar coordinates. Denote the disk by Ω and let \mathcal{J} be the integral of f over Ω , i.e.,

$$\mathcal{J} = \int_{\Omega} f = \int_0^{2\pi} \int_0^R f(r, \theta) r dr d\theta. \quad (1.1)$$

Clearly, \mathcal{J} depends on R . Let us assume f also depends on R , i.e., $f = f(r, \theta; R)$. A physical example could be that \mathcal{J} is the net flow rate of liquid through a pipe with cross-section Ω . In this case, f is the flow rate per unit area and could be the solution of a PDE defined on Ω , e.g., a Navier-Stokes fluid flowing in a circular pipe.

It can be advantageous to know the *sensitivity* of \mathcal{J} with respect to R , e.g., for optimization purposes. In other words, if R increases, how does \mathcal{J} change? To see this, let us

differentiate \mathcal{J} with respect to R :

$$\begin{aligned}\frac{d}{dR}\mathcal{J} &= \int_0^{2\pi} \left(\frac{d}{dR} \int_0^R f(r, \theta; R) r dr \right) d\theta \\ &= \int_0^{2\pi} \int_0^R f'(r, \theta; R) r dr d\theta + \int_0^{2\pi} f(R, \theta; R) R d\theta,\end{aligned}$$

where f' is the derivative with respect to R . The dependence of f on R can more generally be viewed as dependence on Ω , i.e., $f(\cdot; R) \equiv f(\cdot; \Omega)$. Rewriting the above formula using Cartesian coordinates \mathbf{x} , we get

$$\frac{d}{dR}\mathcal{J} = \int_{\Omega} f'(\mathbf{x}; \Omega) d\mathbf{x} + \int_{\partial\Omega} f(\mathbf{x}; \Omega) dS(\mathbf{x}), \quad (1.2)$$

where $d\mathbf{x}$ is the “volume” measure and $dS(\mathbf{x})$ is the “surface area” measure.

1.2.2 • More General Perturbations

Let ν be the unit normal vector of $\partial\Omega$ (pointing outward). We can view increasing R as a velocity field \mathbf{V} that drives points on $\partial\Omega$ in the normal direction, i.e., take $\mathbf{V} = \nu$ on $\partial\Omega$. Hence, (1.2) becomes

$$\frac{d}{dR}\mathcal{J} = \int_{\Omega} f'(\mathbf{x}; \Omega) d\mathbf{x} + \int_{\partial\Omega} f(\mathbf{x}; \Omega) \mathbf{V}(\mathbf{x}) \cdot \nu(\mathbf{x}) dS(\mathbf{x}), \quad (1.3)$$

where we view \mathbf{V} as a velocity field that **instantaneously perturbs the domain Ω** . We often call \mathbf{V} a *domain perturbation*. Let us adopt the notation $f(\mathbf{x}; \Omega) \equiv f(\Omega)$ and $f'(\mathbf{x}; \Omega) \equiv f'(\Omega; \mathbf{V})$, where f' is called the *shape derivative* of f with respect to the domain perturbation \mathbf{V} . Similarly, let us use $\delta\mathcal{J}(\Omega; \mathbf{V}) \equiv \frac{d}{dR}\mathcal{J}$ to denote the *shape perturbation* of \mathcal{J} with respect to Ω , in the direction \mathbf{V} (i.e., a directional derivative). Thus, we obtain

$$\delta\mathcal{J}(\Omega; \mathbf{V}) = \int_{\Omega} f'(\Omega; \mathbf{V}) + \int_{\partial\Omega} f(\Omega) (\mathbf{V} \cdot \nu), \quad (1.4)$$

which is formula (5.47) in Chapter 5. Hence, we have derived (5.47) for the case where Ω is a disk perturbed by a velocity field \mathbf{V} that causes Ω to uniformly expand (in the normal direction). The main purpose of this book is to derive (1.4), and other similar formulas, for general domains Ω and general choices of the perturbation \mathbf{V} .

The framework of shape differential calculus provides the tools for developing the equations of mean curvature flow and Willmore flow, which are geometric flows that occur in many applications such as fluid dynamics and biology. See Chapters 6 and 7 for examples.

1.2.3 • Sequential Optimization of Shape

Which Way Is Down?

It is obvious how to go down a hill. As long as you can see and feel the ground, it is clear which direction to move in order to *lower* your elevation. As motivation for the next section, let us view this as an optimization task. In other words, let $f = f(x, y)$ be a function describing the surface height of the hill, where (x, y) are the coordinates of our

position. Then, by using basic multivariable calculus, finding a direction that will move us downhill is equivalent to computing the gradient (vector) of f and moving in the opposite direction to the gradient. In this sense, we do not need to “see” the whole function. We just need to *locally* compute the gradient ∇f , analogous to feeling the ground beneath.

The shape perturbation in (1.4) is similar to the gradient operator. It provides information about the local slope, or the sensitivity of a quantity with respect to some parameters. In fact, (1.4) is a directional derivative, analogous to $\mathbf{V} \cdot \nabla f$, where \mathbf{V} is a given direction. This is summarized in Table 1.1.

Table 1.1. *Analogy between “standard” derivatives and shape perturbations.*

	Scalar function	Shape functional
Quantity	f	\mathcal{J}
Parameters	(x, y)	Ω
Directional derivative	$\mathbf{V} \cdot \nabla f(x, y)$	$\delta \mathcal{J}(\Omega; \mathbf{V})$

The analogies in Table 1.1 are not equivalent. For instance, it takes only two numbers to specify (x, y) , whereas an “infinite” number of coordinate pairs is needed to specify Ω . Moreover, \mathbf{V} is a two-dimensional vector in the scalar function setting; for a shape functional, \mathbf{V} is a full-blown function requiring definition at *every* point in Ω . This “infinite dimensionality” is the reason for using the notation $\delta \mathcal{J}(\Omega; \mathbf{V})$ to denote a shape perturbation.

Therefore, $\delta \mathcal{J}(\Omega; \mathbf{V})$ indicates how we should change Ω in order to decrease \mathcal{J} , similarly to how $\nabla f(x, y)$ indicates how the coordinate pair (x, y) should change in order to decrease f . This opens up the world of optimization to shape, i.e., shape optimization [3, 23, 48, 51, 54, 59, 75, 93, 106, 107]. The next section describes a classic example of engineering shape optimization.

Minimizing Drag

Although the following example is beyond the scope of this book, it gives a nice picture of the power of shape differential calculus. Consider the flow of fluid past a rigid body (see Figure 1.1). The fluid vector velocity field \mathbf{u} obeys the PDEs known as the Navier-Stokes equations [8, 79, 101] in nondimensional form:

$$\begin{aligned} (\mathbf{u} \cdot \nabla) \mathbf{u} - \nabla \cdot \boldsymbol{\sigma} &= \mathbf{0} && \text{in } \Omega, \\ \nabla \cdot \mathbf{u} &= 0 && \text{in } \Omega, \\ \mathbf{u} &= \mathbf{0} && \text{on } \Gamma_B, \\ \mathbf{u} &= \mathbf{e}_x && \text{on } \Gamma_O, \end{aligned} \tag{1.5}$$

$$\begin{aligned} \boldsymbol{\sigma}(\mathbf{u}, p) &:= -pI + \frac{1}{\text{Re}} D(\mathbf{u}), \\ D(\mathbf{u}) &:= \nabla \mathbf{u} + (\nabla \mathbf{u})^T, \end{aligned} \tag{1.6}$$

where $\mathbf{0}$ is the zero vector, $\mathbf{e}_x = (1, 0)$ is the velocity boundary condition on the outer boundary Γ_O , $\boldsymbol{\sigma}$ is the stress tensor, and p is the pressure. The Reynolds number Re is a nondimensional parameter related to the physical characteristics of the flow. It is not important at this time to completely understand (1.5) and (1.6). It is enough to know that they model how a fluid moves around a rigid body held stationary in an imposed flow field.

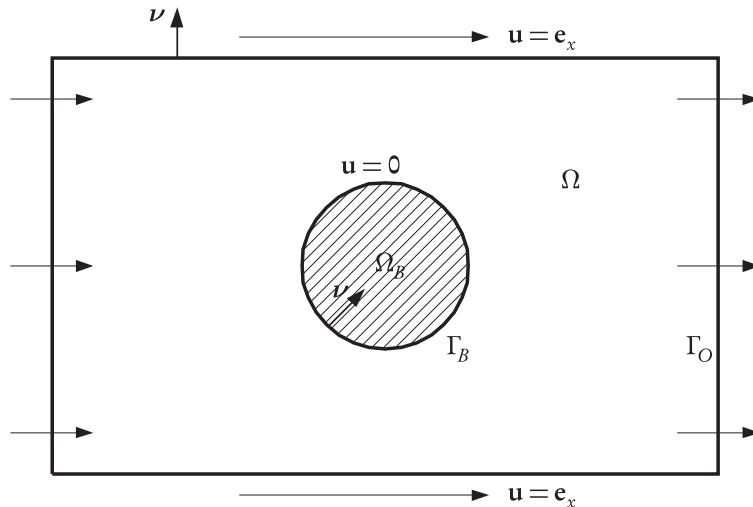


Figure 1.1. Diagram of a fluid flowing past a rigid object Ω_B . The fluid is present in Ω and is governed by the Navier–Stokes equations. The boundary of Ω , denoted $\partial\Omega$, partitions as $\partial\Omega = \Gamma_B \cup \Gamma_O$, where Γ_B is the boundary of the rigid body and Γ_O is an outer boundary far from the body. The outer unit normal vector of Ω is denoted ν .

The objective here is to find the best shape of Ω_B to minimize drag on the body; this is a classic problem in shape optimization [42, 69, 82–84]. For this, we need to specify a shape functional that represents the drag, i.e.,

$$J_d(\Omega) = -\mathbf{e}_x \cdot \int_{\Gamma_B} \boldsymbol{\sigma}(\mathbf{u}, p) \nu, \quad (1.7)$$

where we have used Ω to represent the shape of Ω_B (this is because Ω and Ω_B share the boundary Γ_B). Physically, (1.7) represents the net force that must be applied to Ω_B to keep it stationary while being acted upon by the imposed flow field. One can also show that J_d equals

$$J_d(\Omega) = \frac{1}{2\text{Re}} \int_{\Omega} |D(\mathbf{u})|^2, \quad (1.8)$$

which represents the total amount of viscous dissipation of energy (per unit of time) in the fluid domain Ω . Note that, clearly, $J_d \geq 0$. Using the machinery of shape perturbations, $\delta J_d(\Omega; \mathbf{V})$ indicates how J_d changes when we perturb Ω in the direction \mathbf{V} . Hence, we can use this information to change Ω in small steps so as to slowly deform Ω into a shape that has better (lower) drag characteristics.

A numerical computation illustrating this is shown in Figure 1.2. Let Ω^0 and Γ_B^0 be the initial guess for the shape of the body; these are shown at iteration 0. See that two large vortices appear behind the body, which indicate a large amount of viscous dissipation (i.e., large drag). The optimization process then computes $\delta J_d(\Omega^0; \mathbf{V})$ for many different choices of \mathbf{V} and chooses the one that drives down J_d the most. This choice of \mathbf{V} is used to deform Γ_B^0 into a new shape Γ_B^1 at iteration 1, with only a small difference between Γ_B^0 and Γ_B^1 . This process is repeated many times, the results of which are shown in Figure 1.2. Note how the vortices are eliminated by the more slender shape; clearly the object at iteration 60 has less drag than the initial circular shape.

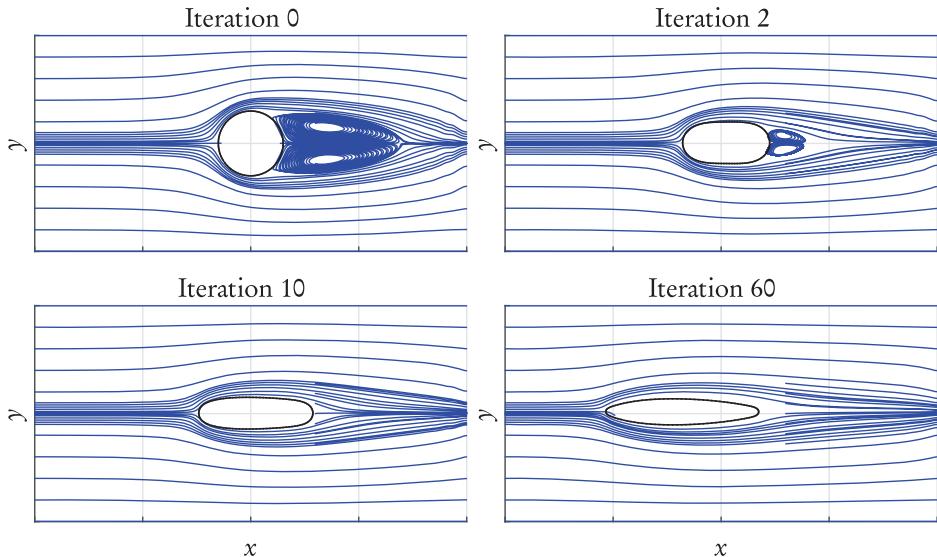


Figure 1.2. Optimizing drag through shape. Starting with a circular shape for Γ_B (not very aerodynamic), we apply a steepest descent optimization scheme to slowly evolve Γ_B toward a minimum of J_d . Blue curves are streamlines of the fluid velocity field \mathbf{u} , which satisfies (1.5) with $\text{Re} = 200$.

Hence, shape perturbations allow us to “climb down the hill” in the infinite dimensional setting of shape. This is a powerful tool for producing sophisticated engineering designs in an *automatic* way. There was no human decision involved in creating the optimized shape in Figure 1.2. The only human intervention was in creating a computer model of (1.5) and developing an optimization algorithm to generate a sequence of shapes. In fact, the same optimization machinery can be used with different PDE systems, such as elasticity. Describing the complete details of shape optimization is beyond the scope of this book, but there is a brief discussion on it in section 6.3.1.

1.3 • Abstract vs. Concrete Presentation

This book derives several formulas and identities for two-dimensional (2-D) surfaces embedded in three dimensions, as well as their shape perturbations. Many of the results also hold for one-dimensional (1-D) curves embedded in three dimensions; any discrepancies will be noted. Some results on shape perturbations of three-dimensional (3-D) domains are also given. Therefore, to make the discussion as clear as possible, we adopt the *extrinsic* point of view: curves and surfaces are assumed to lie in a Euclidean space of higher dimension. In our case, the ambient space is 3-D Euclidean space.

Alternatively, there is the *intrinsic* point of view, which means the surface is **not** assumed to lie in an ambient space. In other words, one is not allowed to reference anything “outside” of the surface when defining it. Moreover, no mathematical structures “outside” of the surface can be utilized. We do not adopt the intrinsic view or consider higher dimensional manifolds, general embedding dimensions, etc., for the following reasons:

- This book is meant to be used as background information for deriving physical models where geometry plays a critical role. Because most physical problems of interest take place in 3-D Euclidean space, the extrinsic viewpoint is sufficient.

- Many of the proofs and derivations of differential geometry relations simplify dramatically for 2-D surfaces in three dimensions and require only basic multivariable calculus and linear algebra.
- The concepts of *normal vectors* and *curvature* are harder to motivate with the intrinsic viewpoint. What does it mean for a surface to “curve through space” if you cannot talk about the ambient space?
- We want to keep in mind applications of this machinery to geometric PDEs, fluid dynamics, numerical analysis, optimization, etc. An interesting application of this methodology is for the development of numerical methods for mean curvature flow and surface tension driven fluid flow. Ergo, the extrinsic viewpoint is often more convenient for computational purposes.
- In addition, we want our framework to be useful for **analyzing and solving shape optimization problems**, i.e., **optimization problems where geometry (or shape) is the control variable.**

Therefore, this text is meant as a *practical* guide to differential geometry and shape differentiation that can be used by researchers in other fields.

Remark 1. *Despite our extrinsic point of view, many of the results we derive can be viewed in an intrinsic way, e.g., the formula will not make explicit reference to the ambient space in a fundamental way.*

1.4 ■ Outline

- Chapter 2. First, we review some preliminary background information. We then define the concept of a surface and develop the basic ideas of local charts, parameterizations, and tangent planes.
- Chapter 3. We introduce the first and second fundamental forms of differential geometry. We then motivate the concepts of summed (mean) curvature and Gaussian curvature. Chapters 2 and 3 are basically a crash course in differential geometry.
- Chapter 4. A calculus framework on surfaces is developed. We systematically build up differential operators that are defined only on a surface, i.e., the surface gradient ∇_Γ and surface Laplacian Δ_Γ . Next, we develop alternative curvature formulas in terms of ∇_Γ and Δ_Γ . We then prove an integration by parts formula on surfaces with a focus on surfaces *with boundary*. We also derive a variety of useful identities.
- Chapter 5. This chapter describes the framework of shape differential calculus, i.e., the combination of differential geometry with the calculus of variations.
- Chapter 6. We describe some applications of shape differentiation, such as finding the equilibrium shapes of droplets under surface tension. We also introduce the concept of gradient flows that can be used to derive the mean curvature flow equation.
- Chapter 7. Here, we apply the tools of shape differential calculus to derive an important geometric flow model called Willmore flow.

- Appendices. Appendix A provides basic identities on vectors and matrices. Appendix B gives background material on changes of variables for derivatives and integrals and some derivative identities involving matrices.

Most of the material in this book can be found in various forms in [5, 22–26, 60, 88, 93, 110].

1.5 • Prerequisites

When reading any mathematical text, the reader must have a certain level of mathematical “maturity” in order to efficiently learn what is in the text. The following items are necessary in this regard.

- Basic set theory. You should understand the concepts of open and closed sets, subsets, boundaries of sets, the space \mathbb{R}^n , etc. A brief review is given in section 2.1.
- Multivariable calculus. In particular, you should understand level surfaces, what the gradient is, the Hessian, the Jacobian, and line and surface integrals. You should have no problem performing computations using these concepts. A standard undergraduate course in calculus of several variables (or vector calculus) should cover this; see [61].
- Differential equations. You should be able to solve simple differential equations, i.e., first and second order constant coefficient ordinary differential equations (ODEs). A standard undergraduate course should suffice. Also, some exposure to PDEs is useful, e.g., knowledge of Laplace’s equation $\Delta u = 0$; see [112].
- Linear algebra. You should be comfortable with vectors and matrices, basic computations such as vector products, matrix-vector products, and determinants. You should understand the concept of a basis and orthogonality. Some familiarity with tensor notation is useful but not required. An upper-level undergraduate course should cover these topics; for instance, see any of [62, 97, 103].
- Geometry. You should have *some* exposure to differential geometry, such as computing the tangent and normal vectors of 3-D curves, as well as computing their curvature. For instance, see [60, pp. 17–71] or [24, Chap. 1]. Most of this is usually covered in freshman calculus and sophomore-level math classes.

It is helpful, but not required, that you be familiar with the concept of a map and mapping of sets (see section 2.1.4). Some exposure to continuum mechanics and the calculus of variations is also useful. Some concepts or terminology in the text may be unfamiliar. If so, then just skip over it. Later sections and practice should clear it up.

1.6 • Notation

We use the following notational conventions throughout the book.

1.6.1 • Vectors

All vector variables are considered **column vectors** and are denoted by **boldface** symbols. Throughout this book, we shall be rather pedantic about row vectors vs. column vectors. For instance, (q_1, \dots, q_n) is a row vector in \mathbb{R}^n with n components $\{q_i\}_{i=1}^n$. If $\mathbf{a} \in \mathbb{R}^3$ is a column vector, then we write $\mathbf{a} = (a_1, a_2, a_3)^T$, where the superscript T denotes the

transpose operator. The “dot product” of two vectors is written as $\mathbf{a} \cdot \mathbf{b} = \mathbf{a}^T \mathbf{b}$, where \mathbf{a} and \mathbf{b} are column vectors. We denote by $|\mathbf{a}|$ the Euclidean norm of the vector \mathbf{a} . See Appendix A for more information on vector/matrix notation and basic identities.

Note that we use the notation (a, b) to be the open set of real numbers contained between a and b , i.e., $x \in (a, b)$ is equivalent to $a < x < b$; thus, do not confuse this with a 1×2 row vector (the context will make it clear).

1.6.2 ■ Gradients

We use t to denote a parameter for curve parameterizations, but sometimes it may play a role similar to “physical time.” For surfaces, we use s_1, s_2 to denote the parameterization variables. The symbol ∇ is the standard spatial gradient operator. All vector derivative operators (such as ∇) are considered *row vectors*, e.g., if $f = f(x, y, z)$ is a scalar valued function, then ∇f is a 1×3 row vector. The notation $\nabla_{\mathbf{x}}$ is the gradient with respect to the variable \mathbf{x} ; the subscript is used for extra clarification. For example, if \mathbf{x} is the vector independent variable $\mathbf{x} = (x, y, z)$, then

$$\nabla_{\mathbf{x}} = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right).$$

Throughout the book, we mainly use the notation $\mathbf{x} = (x_1, x_2, x_3)$ and

$$\nabla_{\mathbf{x}} = \left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3} \right).$$

1.6.3 ■ Integrals

We usually use Ω to denote a domain in \mathbb{R}^3 with positive volume (or a domain in \mathbb{R}^2 with positive area). Likewise, we use Γ to denote a surface in \mathbb{R}^3 , and Σ denotes a curve. The identity map over a generic domain D is denoted id_D , e.g., $\text{id}_{\Omega}(\mathbf{x}) = \mathbf{x}$ for all \mathbf{x} in Ω .

When writing integrals, we shall use \int_D to denote the integral over the generic domain D . For instance, if $\Omega \subset \mathbb{R}^3$, then

$$\int_{\Omega} f(\mathbf{x}) d\mathbf{x} = \iiint_{\Omega} f(x_1, x_2, x_3) dx_1 dx_2 dx_3.$$

The differential measure is denoted $d\mathbf{x}$ for volumetric domains, $dS(\mathbf{x})$ for surfaces, and $d\alpha(\mathbf{x})$ (differential arc-length) for curves. Moreover, we will often drop the arguments of the function and the differential measure $d\mathbf{x}$, $dS(\mathbf{x})$, etc., when writing integrals, i.e.,

$$\int_{\Omega} f(\mathbf{x}) d\mathbf{x} \equiv \int_{\Omega} f, \quad \int_{\Gamma} f(\mathbf{x}) dS(\mathbf{x}) \equiv \int_{\Gamma} f, \quad \int_{\Sigma} f(\mathbf{x}) d\alpha(\mathbf{x}) \equiv \int_{\Sigma} f.$$

The appropriate differential measure to use when computing an integral is always implied by the domain of integration, i.e., if the domain is a surface, use $dS(\cdot)$, etc.

We denote the measure of a set by $|\cdot|$, i.e.,

$$|\Omega| = \int_{\Omega} 1, \quad |\Gamma| = \int_{\Gamma} 1, \quad |\Sigma| = \int_{\Sigma} 1. \quad (1.9)$$

Thus, $|\Omega|$ is the volume of Ω , $|\Gamma|$ is the surface area of Γ , and $|\Sigma|$ is the arc-length of Σ . See Appendix B for more information on derivatives and integrals.

Chapter 2

Surfaces and Differential Geometry

Differential geometry is the detailed study of the *shape* of a surface (manifold), including *local* and *global* properties. A plane in \mathbb{R}^3 (three dimensions) is a very simple surface and does not require fancy tools to characterize. On the other hand, an “arbitrarily” shaped surface, such as the hood of a car, has many distinguishing geometric features (e.g., highly curved regions, regions of near flatness, etc.). Characterizing these features quantitatively and qualitatively requires the tools of differential geometry. Moreover, geometric details are important in many physical and biological processes, such as surface tension [20, 21] and biomembranes [9, 55, 90, 114].

The framework of differential geometry is built by first defining a local map (i.e., surface parameterization) which defines the surface. Then, a calculus framework is built up on the surface analogous to the standard “Euclidean calculus.” Other approaches are also possible, such as those with implicit surfaces defined by level sets and distance functions. But parameterizations, though arbitrary, are quite useful in a variety of settings, so we will stick mostly with those. We emphasize that the geometry of a surface *does not depend* on a particular parameterization; the notion of *regular surface* in section 2.3 is introduced to deal with this (see Proposition 1).

Throughout this chapter, and for the rest of the book, we mainly focus on 2-D surfaces in 3-D space (\mathbb{R}^3). We begin by reviewing some fundamentals in order to make this text as complete as possible.

2.1 ■ Preliminaries

The following sections quickly review basic concepts that underlie much of this book. However, it is not critical to appreciate all the finer points of sets, mappings, etc., in order to read this book. But if the ideas discussed here are completely alien to you, then we encourage you to consult a good textbook such as [61, 62, 64].

2.1.1 ■ Euclidean Space

Let \mathbb{R}^n denote n -dimensional Euclidean space. Throughout the text, we mainly take $n = 3$, but sometimes we may specialize to $n = 2$. We assume the reader is well-versed in *Cartesian coordinate systems*, vector notation and vector arithmetic, vector operations (dot product, cross product), the angle between two vectors, etc. A general vector \mathbf{x} in \mathbb{R}^3 will usually have *components* denoted by $\mathbf{x} = (x_1, x_2, x_3)^T$.

Next, assume we have a given coordinate system. Any *point* P in \mathbb{R}^n has a unique *position vector*, say \mathbf{x}_P in \mathbb{R}^n , that points from the origin to P , i.e., the coordinates of the point P are just the components of the vector \mathbf{x}_P . Thus, sometimes it is convenient to *identify* the point P with the vector \mathbf{x}_P , meaning that we will refer to the point by its position vector. In this case, we will drop the subscript and just refer to the point by \mathbf{x} . We will use this “abuse” of notation when there is no possibility of ambiguity. Otherwise, we will emphasize the distinction between point and position vector.

2.1.2 ■ Sets: Open, Closed, Boundary, Neighborhood

Generally speaking, a *set* is a collection of distinct objects. For example, $\{X, Y\}$ is a set consisting of the distinct objects X and Y ; we use curly braces $\{\}, \}$ when defining a set. We often introduce another symbol, e.g., $Q = \{X, Y\}$, for convenience in referring to the set. Let S and U be sets. We assume the reader is familiar with the concept of sets, membership in sets (e.g., $\mathbf{x} \in S$), subsets (e.g., $S \subset U$), and operations on sets, such as intersection ($S \cap U$), union ($S \cup U$), set difference ($U \setminus S$), etc.

Almost all sets in this book will be sets of numbers (e.g., $\{1, \pi, \sqrt{2}, \dots\}$), sets of points in \mathbb{R}^n , or sets of vectors in \mathbb{R}^n (e.g., $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots\}$), i.e., subsets of Euclidean space. Sometimes a set is defined through a condition, e.g., $\{x \in G : \text{such that } x \text{ satisfies a condition}\}$. For example, the set $\{1, 2, 3\}$ can also be defined by $\{a \in \mathbb{Z} : a > 0 \text{ and } a < 4\}$, where \mathbb{Z} is the set of integers. The *empty set*, denoted \emptyset , is the unique set having no elements: $\{\}$.

Given a point \mathbf{x} in \mathbb{R}^n , and a positive number r , let $B_r(\mathbf{x})$ be the set of all points in \mathbb{R}^n whose distance from \mathbf{x} is *strictly less* than r . This is written more formally as

$$B_r(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^n : |\mathbf{x} - \mathbf{y}| < r\}. \quad (2.1)$$

In other words, $B_r(\mathbf{x})$ is the interior of a solid ball (in n dimensions) of radius r centered at \mathbf{x} . Next, we define the *boundary* of $B_r(\mathbf{x})$ as

$$\partial B_r(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^n : |\mathbf{x} - \mathbf{y}| = r\}, \quad (2.2)$$

i.e., $\partial B_r(\mathbf{x})$ is the *surface* of a sphere of radius r centered at \mathbf{x} .

From the above, we see that the intersection $B_r(\mathbf{x}) \cap \partial B_r(\mathbf{x})$ is empty, i.e., $B_r(\mathbf{x})$ does **not contain** any portion of $\partial B_r(\mathbf{x})$. In other words, $B_r(\mathbf{x})$ does not contain any part of its boundary. Written more formally, we have that $B_r(\mathbf{x}) \cap \partial B_r(\mathbf{x}) = \emptyset$.

We use the term *open* to indicate that a set does not contain any part of its boundary. More precisely, a subset U of \mathbb{R}^n is *open* if every point \mathbf{x} in U has a ball $B_r(\mathbf{x})$, for some radius $r > 0$, contained in U . In other words, given a point \mathbf{x} in an open set U , we can move in all directions from \mathbf{x} by a small distance and still stay within U . So $B_r(\mathbf{x})$ is an example of an open set; in fact, we often refer to $B_r(\mathbf{x})$ as an *open ball*. Another example of an open set is $(0, 1) \subset \mathbb{R}$, i.e., the set of numbers between 0 and 1, but excluding 0 and 1.

The *boundary* of a set $S \subset \mathbb{R}^n$ is the set of points \mathbf{x} in \mathbb{R}^n such that every open ball centered at \mathbf{x} contains at least one point in S and at least one point not in S . In other words, every \mathbf{x} in the boundary of S satisfies $B_r(\mathbf{x}) \cap S \neq \emptyset$ and $B_r(\mathbf{x}) \cap (\mathbb{R}^n \setminus S) \neq \emptyset$ for all $r > 0$. We denote the boundary of S by ∂S . See Figure 2.1 for a graphical illustration.

We use the term *closed* to indicate that a set contains its entire boundary. Along these lines, we denote the *closure* operation on S by \bar{S} , i.e., $\bar{S} = S \cup \partial S$ (see Figure 2.1). Thus, we have the *closed ball* $\overline{B_r(\mathbf{x})}$ of radius $r > 0$ and centered at \mathbf{x} defined by

$$\overline{B_r(\mathbf{x})} = \{\mathbf{y} \in \mathbb{R}^n : |\mathbf{x} - \mathbf{y}| \leq r\}. \quad (2.3)$$

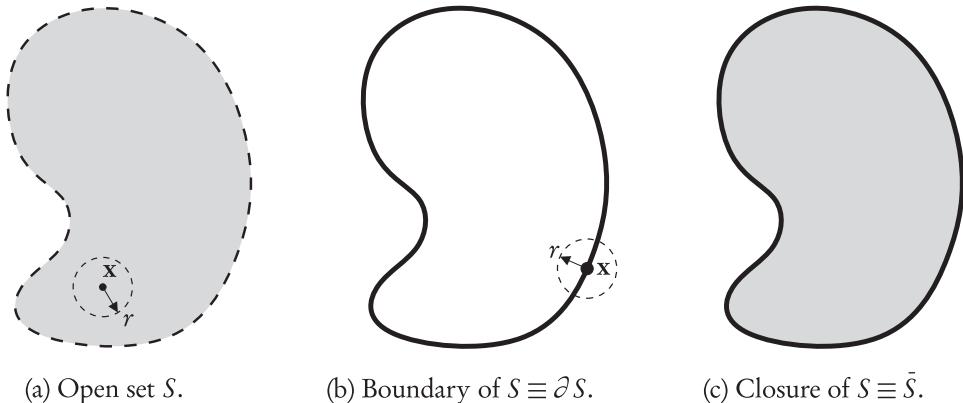


Figure 2.1. (a) An open set S (in \mathbb{R}^2) is shown (shaded). The boundary ∂S , which is not contained in S , is depicted by a dashed curve. Any point x in S can be surrounded by an open ball, of small enough radius $r > 0$, that is also contained in the set. (b) The boundary of the set is highlighted by a thick black curve. For any point x on the boundary, an open ball with any positive radius r , centered at x , will overlap the interior of S and exterior of S . (c) The closure of the set is shown, which is the union of the open set and its boundary: $\bar{S} = S \cup \partial S$.

Another example is the closed set $[0, 1] \subset \mathbb{R}$, i.e., the set of numbers between 0 and 1, and including 0 and 1.

Remark 2. This all seems pedantic, but the concept of open set is critical in multivariable calculus to properly define differentiability [61]. Furthermore, the notation we have introduced for referencing boundaries of sets, as well as the closure of sets, is practical for referencing geometric details of solid objects and their surfaces.

Throughout this book, we often make use of a “neighborhood” around a point, i.e., a neighborhood of a point x in \mathbb{R}^n , is any open set $U \subset \mathbb{R}^n$ that contains x .

2.1.3 • Compactness

A set in \mathbb{R}^n is *bounded* if it is contained in an open ball of sufficiently large, but finite, radius. Moreover, a set in \mathbb{R}^n is said to be *compact* if it is closed and bounded. The concept of compactness is actually more general than this [63, 64]. But for our purposes, the previous definition is sufficient.

We say that a (nonempty) open set S is *compactly contained* in another open set W , denoted $S \subset\subset W$, if $\bar{S} \subset W$ and \bar{S} is compact. In other words, the boundary of S cannot touch the boundary of W , i.e., there is a “little bit of room” between ∂S and ∂W . With this, we can now define the notion of a *compactly supported* function. Let S be an open subset of \mathbb{R}^n and suppose f is a function defined on S . The *support* of f is defined to be the set of points in S where f is nonzero, i.e.,

$$\text{supp}(f) = \{x \in S : f(x) \neq 0\}.$$

Furthermore, we say that f has *compact support* (in S) if $\overline{\text{supp}(f)} \subset\subset S$, i.e., f is zero outside of a compact subset of S . These definitions generalize in the obvious way for vector-valued functions f , i.e., the support of f is the set of points where f is not the zero vector 0 , etc.

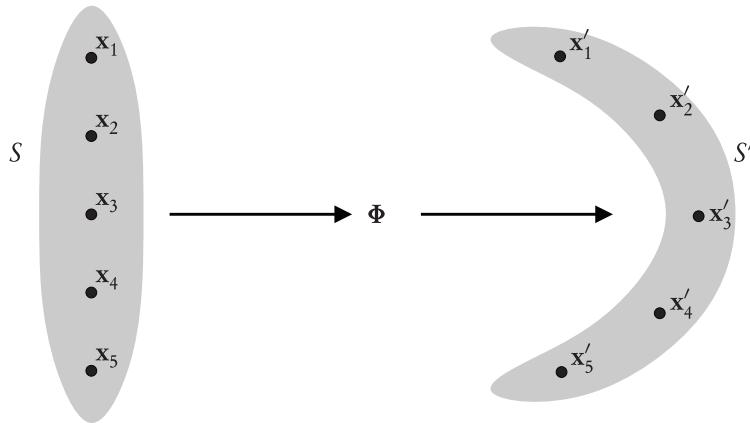


Figure 2.2. A set S in \mathbb{R}^2 that is mapped onto S' by the map Φ . The points correspond through $x'_i = \Phi(x_i)$ for $i = 1, \dots, 5$. One can interpret the action of Φ as a deformation of the set S into S' , i.e., S is “bent” into the shape of S' by the transformation Φ .

Remark 3. Compact support is useful for ignoring boundary effects. For some of the proofs in this book, we need this concept to keep the “action of a function” away from the boundary of a set, or to localize the function in a region of interest. One reason is to avoid potential difficulties with differentiating a function at its boundary of definition. Or, more commonly, we wish to ignore a quantity that depends on the value of a function at a boundary point. For example, $\int_{\partial S} f = 0$ if f has compact support in S .

2.1.4 • Mappings: Basic Definitions

Let S and S' be two sets of points. For every point x in S , if there is a “rule” (function) Φ that associates a point x' in S' to x , then we say that Φ is a *mapping* or *transformation* of the set S into S' . We use the notation $\Phi : S \rightarrow S'$ as shorthand for the previous statement. With this, we can write $x' = \Phi(x)$. We call x' the *image point*, of x , and x is called the *inverse image point* of x' .

Remark 4. In general, if $S \subset \mathbb{R}^m$ and $S' \subset \mathbb{R}^n$, then

$$\Phi = (\Phi_1, \Phi_2, \dots, \Phi_n)^T \quad (\text{recall that vector variables are column vectors}), \quad (2.4)$$

where each Φ_i is a function of m variables: $\Phi_i = \Phi_i(x_1, x_2, \dots, x_m)$ (see Appendix A).

The set of the image points of all points in S is called the *image of S* and is denoted $\Phi(S)$. If every point of S' is the image point of some point in S , then the mapping Φ maps S onto S' , i.e., $S' = \Phi(S)$. In this case, we say that Φ is *surjective*. See Figure 2.2 for an example of mapping a set of points in \mathbb{R}^2 (see Figure 2.3 for examples of mapping a set in \mathbb{R}^3).

If the image points of any pair of distinct points in S are also distinct points in S' , then we say that Φ is *injective* (the classical terminology was to call Φ a *one-to-one* map). If Φ is surjective and injective (also called *bijection*), then there exists the *inverse mapping* of Φ , denoted Φ^{-1} , that maps S' onto S such that if x, x' satisfy $x' = \Phi(x)$, then $x = \Phi^{-1}(x')$. Thus, $\Phi^{-1} : S' \rightarrow S$.

A mapping Φ of S into S' is said to be *continuous at a point \mathbf{x}* in S if, for every neighborhood \mathcal{N}' of $\mathbf{x}' = \Phi(\mathbf{x})$, there exists a neighborhood \mathcal{N} of \mathbf{x} such that $\Phi(\mathcal{N}) \subset \mathcal{N}'$. We say the mapping is *continuous* if it is continuous at every point of S .

A bijective, continuous mapping Φ whose inverse Φ^{-1} is also continuous is called a topological mapping or *homeomorphism*. Point sets that can be topologically mapped onto each other are said to be *homeomorphic*. Sets that are homeomorphic have the “same topology”, i.e., their connectedness is the same; they have the same kinds of “holes.” There is further discussion of this in section 2.3.1, where Figure 2.7 shows what can happen when a mapping is not a homeomorphism.

A mapping Φ is called a *rigid motion* if any pair of points \mathbf{a}, \mathbf{b} are the same distance apart as the corresponding pair $\Phi(\mathbf{a}), \Phi(\mathbf{b})$.

2.1.5 • Orthogonal Transformations

Let $\mathbf{b} = (b_1, b_2, b_3)^T$ be in \mathbb{R}^3 and \mathbf{A} be in $\mathbb{R}^{3 \times 3}$, i.e., a 3×3 matrix $\mathbf{A} = [a_{ij}]_{i,j=1}^3$, where a_{ij} are the entries of \mathbf{A} (see Appendix A.2.1). Define the following (affine) linear map Φ (transformation):

$$\tilde{\mathbf{x}} = \Phi(\mathbf{x}) = \mathbf{Ax} + \mathbf{b} \quad \Leftrightarrow \quad \tilde{x}_i = (\Phi(\mathbf{x}))_i = \left(\sum_{k=1}^3 a_{ik} x_k \right) + b_i, \quad (2.5)$$

where $(\Phi(\mathbf{x}))_i \equiv \Phi_i(x_1, x_2, x_3)$. If \mathbf{A} satisfies the properties

$$\mathbf{A}^{-1} = \mathbf{A}^T, \quad \det(\mathbf{A}) = 1, \quad (2.6)$$

where $\det(\mathbf{A})$ is the determinant of \mathbf{A} , then Φ represents a rigid motion. Basically, Φ consists of a rotation (represented by \mathbf{A}) followed by a translation (represented by \mathbf{b}). A rigid motion can be used to transition from one Cartesian coordinate system to another.

If $\mathbf{b} = \mathbf{0}$ and (2.6) still holds, then $\Phi(\mathbf{x}) = \mathbf{Ax}$ is a linear map known as a *direct orthogonal transformation*. This is nothing more than a rotation of the coordinate system with the origin as the center. If (2.6) is replaced by

$$\mathbf{A}^{-1} = \mathbf{A}^T, \quad \det(\mathbf{A}) = -1, \quad (2.7)$$

then $\Phi(\mathbf{x}) = \mathbf{Ax}$ is called an *opposite orthogonal transformation*, which consists of a rotation about the origin and a reflection in a plane. Both (2.6) and (2.7) are examples of *orthogonal matrices*.

Remark 5 (interpretation of transformations). We can interpret (2.5) in two different ways. Consider a point P in \mathbb{R}^3 with coordinates \mathbf{x} .

- Alias. Viewing (2.5) as a transformation of coordinates, it appears that \mathbf{x} and $\tilde{\mathbf{x}}$ are the coordinates of the *same point* with respect to two different coordinate systems. In other words, the point is referenced by two different “names” (sets of coordinates).
- Alibi. Viewing (2.5) as a mapping of sets, it appears that \mathbf{x} and $\tilde{\mathbf{x}}$ are the coordinates of two *different points* with respect to the same coordinate system. In other words, the point at $\tilde{\mathbf{x}}$ “was previously” at \mathbf{x} before applying the map.

The concept of material point is directly related to the alibi viewpoint. One can think of a “particle” of material (i.e., *material point*), initially located at \mathbf{x} , that then moves to $\tilde{\mathbf{x}}$ because of some physical process. The transformation (2.5) simply represents the

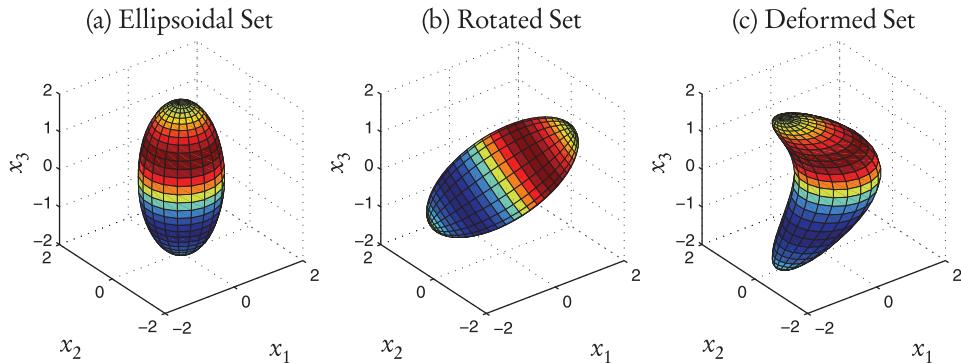


Figure 2.3. (a) Set of points S in \mathbb{R}^3 . The color and grid-lines help to visualize the shape. (b) Rotated set of points $S' = \Phi(S)$, where Φ is given by (2.5) and (2.6). (c) Deformed set of points $S' = \Phi(S)$, where Φ is given by (2.9).

kinematic outcome of that physical process. This is a standard concept in deformable continuum mechanics, especially nonlinear elasticity. See Figure 2.3 for an example of applying a rigid motion to a set of points in \mathbb{R}^3 .

2.1.6 • General Transformations

In general, transformations may not be linear. Therefore, $\Phi : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ may be written as

$$\Phi = (\Phi_1, \Phi_2, \Phi_3)^T, \quad (2.8)$$

where $\Phi_i = \Phi_i(x_1, x_2, x_3)$ ($i = 1, 2, 3$) are scalar valued (nonlinear) functions. Remark 5 applies to these transformations as well. Thus, the alias viewpoint yields a *curvilinear* coordinate system. The alibi viewpoint implies that the set S is *deformed* into the set $S' = \Phi(S)$. See Figure 2.3 for an example of a nonlinear map Φ applied to an ellipsoid shaped set of points, where Φ is defined by

$$\Phi = (x_1 - 1.2 + 1.6 \cos(x_3 \pi / 4), x_2, x_3)^T. \quad (2.9)$$

When dealing with transformations that map from \mathbb{R}^n into \mathbb{R}^n , we shall use the symbol Φ to represent the transformation (usually $n = 2$ or 3). But we will also consider transformations from \mathbb{R}^q into \mathbb{R}^n , where $q < n$. When $q = 2$ and $n = 3$, we denote the transformation by $\mathbf{X} : \mathbb{R}^2 \rightarrow \mathbb{R}^3$, where we have used a different symbol for emphasis; this is used when defining surfaces (see sections 2.2.1 and 2.3). Note that one can think of \mathbf{X} as the restriction of Φ to the x_1, x_2 plane. When $q = 1$, we use the symbol $\alpha : \mathbb{R}^1 \rightarrow \mathbb{R}^n$ ($n = 2$ or 3), which corresponds to parameterizing curves. Similarly, one can think of α as the restriction of Φ to the x_1 -axis.

2.2 • The Parametric Approach

2.2.1 • What Is a Surface?

A surface is a set of points in space that is “regular enough.” A random scattering of points in space does **not** match our intuitive notion of what a surface is, i.e., it is not regular enough. On the other hand, the boundary of a sphere **does match** our notion of a surface, i.e., it is **regular** enough to be a surface because a sphere is “smooth.”

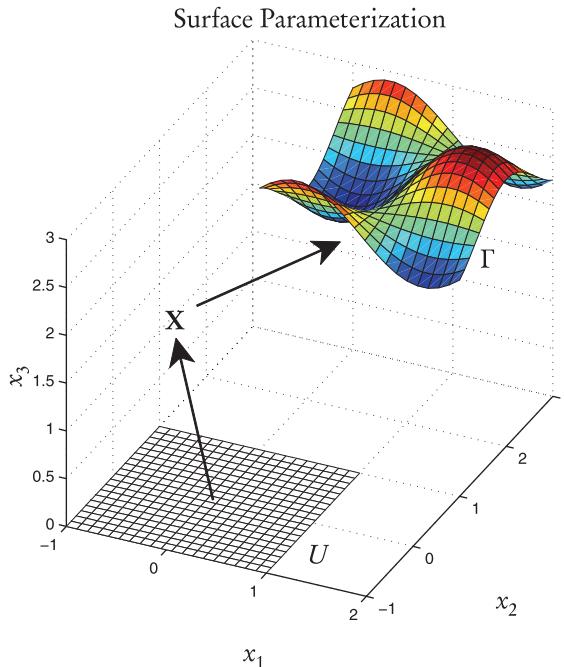


Figure 2.4. Example of a parametric representation. The reference domain U is the square shown in the x_1, x_2 plane. The map \mathbf{X} is applied to the set U to create a curved surface Γ (the surface is colored based on the x_3 coordinate).

Intuitively, one can think of creating a surface as deforming a flat rubber sheet into a curved sheet. The transformation \mathbf{X} in section 2.1.6 captures this idea. Therefore, let $U \subset \mathbb{R}^2$ be a “flat” domain and let $\mathbf{X} : U \rightarrow \mathbb{R}^3$ be this deforming transformation, i.e., for each point $(s_1, s_2)^T$ in U there is a corresponding point $\mathbf{x} = (x_1, x_2, x_3)^T$ in \mathbb{R}^3 such that

$$\mathbf{x} = \mathbf{X}(s_1, s_2). \quad (2.10)$$

Let $\Gamma = \mathbf{X}(U)$ denote the surface obtained from “deforming” U . We call (2.10) a *parametric representation* of the surface Γ , where s_1, s_2 are called the *parameters* of the representation. Sometimes, we will refer to U as a *reference domain*. See Figure 2.4 for an example of (2.10).

Allowable Parameterization

If we are going to use (2.10) to define surfaces, then we must place *assumptions* on \mathbf{X} to guarantee that $\Gamma = \mathbf{X}(U)$ is a valid surface. At the bare minimum, \mathbf{X} must be continuous to avoid “tearing” the rubber sheet. But if we want to perform calculus on Γ , we do in fact need more.

Assumption 1. We make the following regularity assumptions on \mathbf{X} .

- (A1) The function $\mathbf{X}(s_1, s_2)$ is C^∞ on U and each point $\mathbf{x} = \mathbf{X}(s_1, s_2)$ in Γ corresponds to just one point (s_1, s_2) in U , i.e., \mathbf{X} is injective.

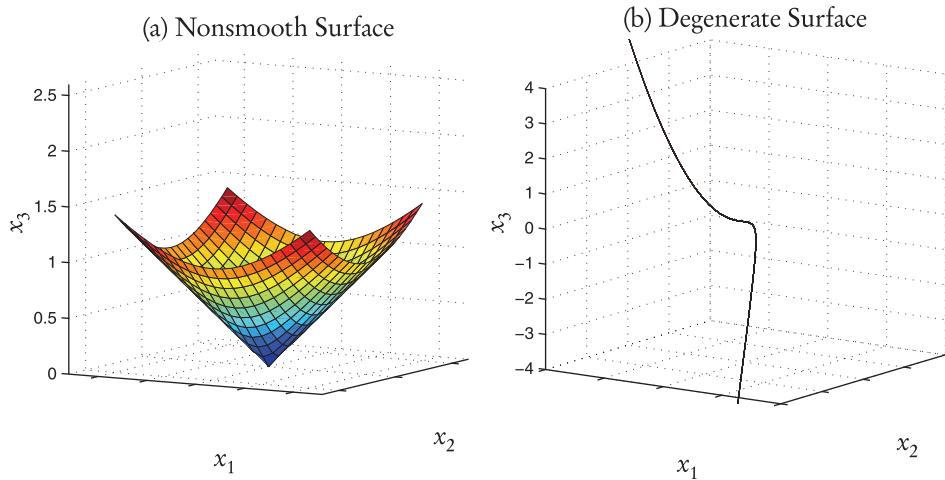


Figure 2.5. Examples where \mathbf{X} does not satisfy Assumption 1 (reference domain U is not shown). (a) Parametric representation of a cone, which has no well-defined tangent plane at the “corner” of the cone. (b) “Surface” parameterization that degenerates to a curve.

- (A2) *The Jacobian matrix*

$$J = [\partial_{s_1} \mathbf{X}, \partial_{s_2} \mathbf{X}] = \begin{bmatrix} \partial_{s_1} X_1 & \partial_{s_2} X_1 \\ \partial_{s_1} X_2 & \partial_{s_2} X_2 \\ \partial_{s_1} X_3 & \partial_{s_2} X_3 \end{bmatrix} \quad (2.11)$$

is of rank 2 on U (see rank in [62, 97, 103]), i.e., the columns of J are linearly independent.

We say that a parameterization of the form (2.10) that satisfies (A1) and (A2) is an **allowable parameterization** or, to put it in more modern terms, an *immersion* [25].

Consequences

Assumption (A1) imposes some smoothness on the surface Γ , i.e., that Γ has a **well-defined tangent plane** at every point in Γ . The precise definition of a tangent plane is given in section 2.4.3; for now, we require only an intuitive notion of a tangent plane. For example, let $U = (-1, 1) \times (-1, 1)$ and consider the map

$$\mathbf{X}(s_1, s_2) = (s_1, s_2, \sqrt{s_1^2 + s_2^2})^T \quad \text{for all } (s_1, s_2)^T \in U. \quad (2.12)$$

The surface $\Gamma = \mathbf{X}(U)$ is a cone (see Figure 2.5(a)). It is clear that (2.12) is **not** differentiable at $(0, 0)^T$, i.e., (A1) is not valid. This is realized in Figure 2.5(a) as a **sharp corner** at $(0, 0, 0)^T$ in Γ . Thus, there is **no unique plane** that passes through $(0, 0, 0)^T$ and is “tangent” to the surface Γ .

Assumption (A2) is needed to avoid the possibility that the set Γ (parameterized by (2.10)) is a curve in \mathbb{R}^3 . By linear algebra, (A2) is equivalent to $\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X} \neq 0$, which is equivalent to $\partial_{s_1} \mathbf{X}$ and $\partial_{s_2} \mathbf{X}$ being linearly independent vectors in \mathbb{R}^3 . For example, let $U = (-1, 1) \times (-1, 1)$ and consider the map

$$\mathbf{X}(s_1, s_2) = (s_1 + s_2, (s_1 + s_2)^2, (s_1 + s_2)^3)^T \quad \text{for all } (s_1, s_2)^T \in U. \quad (2.13)$$

The “surface” $\Gamma = \mathbf{X}(U)$ is just the curve described by the parameterization $\mathbf{X}(t) = (t, t^2, t^3)^T$, where t is the parameter (see Figure 2.5(b)). It is clear from (2.13) that $\partial_{s_1} \mathbf{X}$ and $\partial_{s_2} \mathbf{X}$ are linearly dependent, i.e., J has rank 1 on all of U , so (A2) is not valid. Thus, the surface “degenerates” to a curve.

Let us further characterize (A2). Let $\mathbf{q} = (q_1, q_2)^T$ be a point in U and define $J_{\mathbf{q}} = [\partial_{s_1} \mathbf{X}, \partial_{s_2} \mathbf{X}]|_{s=\mathbf{q}}$ (i.e., evaluate the Jacobian at the point \mathbf{q}). Note that $J_{\mathbf{q}}$ is a constant matrix in $\mathbb{R}^{3 \times 2}$. Next, define $\mathbf{T}_{\mathbf{q}} : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ by

$$\mathbf{T}_{\mathbf{q}}(\mathbf{p}) = J_{\mathbf{q}} \mathbf{p} \quad \Leftrightarrow \quad (\mathbf{T}_{\mathbf{q}}(\mathbf{p}))_i = \sum_{k=1}^2 (J_{\mathbf{q}})_{ik} p_k \quad \text{for } i = 1, 2, 3, \quad (2.14)$$

where $\mathbf{p} = (p_1, p_2)^T$ is any point in \mathbb{R}^2 . Then (A2) is equivalent to the map $\mathbf{T}_{\mathbf{q}}$ being injective for all \mathbf{q} in U . The set $\mathbf{T}_{\mathbf{q}}(\mathbb{R}^2)$ is a vector subspace of \mathbb{R}^3 generated by the two column vectors of $J_{\mathbf{q}}$; thus, it has dimension 2. The map $\mathbf{T}_{\mathbf{q}}$ is related to the tangent plane, which is discussed in section 2.4.3.

2.2.2 • Parametric Surface

We can now define a notion of surface.

Definition 1 (parametric surface). Let $U \subset \mathbb{R}^2$ be an open set and consider a map $\mathbf{X} : U \rightarrow \mathbb{R}^3$. We call (U, \mathbf{X}) a parametric surface if \mathbf{X} is differentiable on U . We say \mathbf{X} is regular if the map $\mathbf{T}_{\mathbf{q}}$ defined in (2.14) is injective for all \mathbf{q} in U . Moreover, if there is a point \mathbf{p} in U for which $\mathbf{T}_{\mathbf{p}}$ is not injective, or undefined, then we call \mathbf{p} a singular point of \mathbf{X} ; otherwise, it is a regular point.

Note that we refer to the pair (U, \mathbf{X}) as a parametric surface because $\Gamma = \mathbf{X}(U)$ is the set of points making up the surface and U and \mathbf{X} describe how to “draw” coordinate curves on Γ . To elaborate further, we recall Figure 2.4. The reference domain U is simply a set of points that make up a square. However, the grid-lines on U correspond to a coordinate system placed on U . And these grid-lines are mapped to Γ by \mathbf{X} (see Figure 2.4), which defines a kind of curvilinear coordinate system on Γ .

If we had chosen a different coordinate system on U , then the grid-lines would look different on U (and also on Γ). In other words, the set Γ would be the same but would be parameterized differently, i.e., there would be a different curvilinear coordinate system on Γ (see Figure 2.10 for an example).

Therefore, it is not surprising that a surface can be parameterized in multiple ways. In fact, given a parameterization (U, \mathbf{X}) of Γ , we can define another parameterization in the following way. Suppose we have a transformation in \mathbb{R}^2 given by

$$s_1 = s_1(\tilde{s}_1, \tilde{s}_2), \quad s_2 = s_2(\tilde{s}_1, \tilde{s}_2), \quad \text{where } (\tilde{s}_1, \tilde{s}_2)^T \in \tilde{U}, \quad (2.15)$$

i.e., $\mathbf{s} : \tilde{U} \rightarrow \mathbb{R}^2$ and $U = \mathbf{s}(\tilde{U})$. Next, define $\tilde{\mathbf{X}} = \mathbf{X} \circ \mathbf{s}$, meaning

$$\tilde{\mathbf{X}}(\tilde{s}_1, \tilde{s}_2) = \mathbf{X}(s_1(\tilde{s}_1, \tilde{s}_2), s_2(\tilde{s}_1, \tilde{s}_2)).$$

Then $(\tilde{U}, \tilde{\mathbf{X}})$ is also a parameterization of Γ . One can think of the map \mathbf{s} as deforming \tilde{U} into U (\mathbf{s}^{-1} deforms U into \tilde{U}).

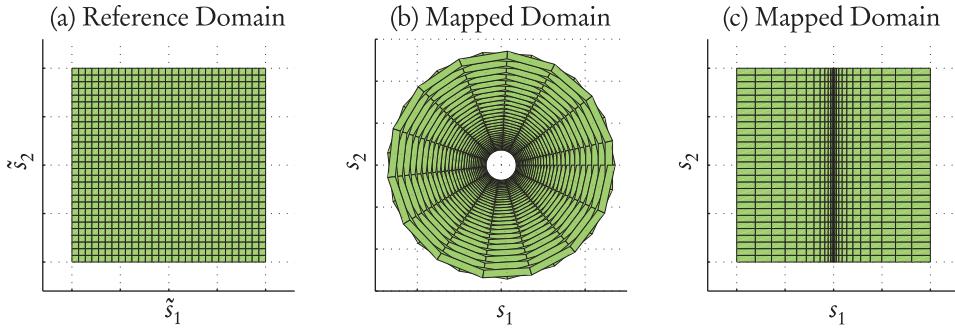


Figure 2.6. Examples where (2.15) does not satisfy Assumption 2. (a) Reference domain U . (b) Map U to an annulus using (2.17); note that the annulus gets covered twice! (c) Map U to a square using (2.18); the grid-lines get squeezed together near the $s_1 = 0$ axis.

Of course, in order to have a regular parameterization, we must have that Assumption 1 is satisfied for $(\tilde{U}, \tilde{\mathbf{X}})$. This requires the following assumptions on (2.15).

Assumption 2.

- (A0*) The functions (2.15) are defined on a domain \tilde{U} such that $U = \mathbf{s}(\tilde{U})$.
- (A1*) The functions (2.15) are C^∞ on \tilde{U} , and (2.15) is an injective transformation.
- (A2*) The Jacobian matrix

$$D = [\partial_{\tilde{s}_1} \mathbf{s}, \partial_{\tilde{s}_2} \mathbf{s}] = \begin{bmatrix} \partial_{\tilde{s}_1} s_1 & \partial_{\tilde{s}_2} s_1 \\ \partial_{\tilde{s}_1} s_2 & \partial_{\tilde{s}_2} s_2 \end{bmatrix} \quad (2.16)$$

is nonsingular for all $(\tilde{s}_1, \tilde{s}_2)^T$ in \tilde{U} , i.e., $\det(D) \neq 0$ on \tilde{U} (compare with Proposition 1).

We say that a transformation of the form (2.15) that satisfies Assumption 2 is an *allowable coordinate transformation*.

The conditions (A1*) and (A2*) are completely independent of each other. For instance,

$$s_1 = e^{\tilde{s}_1} \cos(2\pi \tilde{s}_2), \quad s_2 = e^{\tilde{s}_1} \sin(2\pi \tilde{s}_2) \quad (2.17)$$

is a transformation that is not injective when $-1 \leq \tilde{s}_2 \leq 1$; however, a simple calculation gives $\det(D) = (2\pi)^2 e^{2\tilde{s}_1}$, which is never zero in the \tilde{s}_1, \tilde{s}_2 plane (see Figure 2.6). On the other hand, the transformation

$$s_1 = \tilde{s}_1^3, \quad s_2 = \tilde{s}_2 \quad (2.18)$$

is injective everywhere, but a simple calculation gives $\det(D) = 3\tilde{s}_1^2 = 3s_1^{2/3}$, which is zero when $s_1 = 0$, i.e., on the whole s_2 -axis (see Figure 2.6).

2.3 • Regular Surface

The fundamental property that makes a set of points in \mathbb{R}^3 a surface is that it *locally looks like a plane at every point*. If you “zoom into” a surface, it should look flat. We would like a definition of a surface that reflects this fact. Definition 1 is inadequate because it defines a surface *in terms of* a parameterization. So we want to define a set, in \mathbb{R}^3 , that

is “intrinsically” 2-D, and is smooth enough so we can perform calculus on it, *without regard to a specific parameterization*. Definition 2 is what we need.

2.3.1 ■ Definition

The following is taken from [24, sect. 2-2, Def. 1].

Definition 2 (regular surface). A subset $\Gamma \subset \mathbb{R}^3$ is a regular surface if, for each \mathbf{x} in Γ , there exists a (bounded) neighborhood $\mathcal{N} \subset \mathbb{R}^3$ of \mathbf{x} , an open set $U \subset \mathbb{R}^2$, and a surjective map $\mathbf{X}: U \rightarrow \mathcal{N} \cap \Gamma$ such that the following hold (see Figure 2.8):

1. $\mathbf{X} = (X_1, X_2, X_3)^T$ is differentiable, i.e., X_i (for $i = 1, 2, 3$) have continuous partial derivatives of any order on U .
2. \mathbf{X} is a homeomorphism (recall section 2.1.4). This means \mathbf{X} is an injective, continuous mapping whose inverse \mathbf{X}^{-1} is also continuous, i.e., $\mathbf{X}^{-1}: \mathcal{N} \cap \Gamma \rightarrow U$ is a continuous function.
3. Regularity condition. For each \mathbf{q} in U , the map $\mathbf{T}_{\mathbf{q}}$ is injective; recall (2.14).

Remark 6 (local chart). The parameterization (map) \mathbf{X} in Definition 2, and associated reference domain U , is sometimes called a system of local coordinates, or local chart, in a neighborhood of \mathbf{x} . Hence, we will refer to the pair (U, \mathbf{X}) as a chart. The neighborhood $\mathcal{N} \cap \Gamma$ of \mathbf{x} in Γ is called a coordinate neighborhood.

Some comments on conditions 1, 2, and 3 are in order. Condition 1 is necessary in order to “differentiate on the surface” Γ , i.e., in order to build up the differential tools of calculus *on the surface* (see Chapter 4). Condition 3 is needed to prevent the surface from (locally) degenerating to a curve or point (see Figure 2.5(b)), which is necessary to give a well-defined tangent plane (see section 2.4.3).

The injective requirement in condition 2 is needed to prevent the surface from *intersecting itself*. Notice that Definition 1 allows for self-intersections. But a surface that intersects itself does **not** look like a plane near the region of intersection (i.e., no well-defined tangent plane), so Definition 1 is not good enough. Note that self-intersection is a *global* concept, which Definition 1 cannot handle.

The continuity of the inverse \mathbf{X}^{-1} in condition 2 is more subtle but is needed to prevent \mathbf{X} from deforming the open set U into a set with a different topology. For example, consider the following parameterization of a “curled cylinder” (see Figure 2.7):

$$U = (-1.1, 1) \times (-1, 1), \quad \mathbf{X}: U \rightarrow \mathbb{R}^3, \text{ where } \mathbf{X}(s_1, s_2) = (s_1^2 - 1, s_1(s_1^2 - 1), s_2)^T. \quad (2.19)$$

The surface $\Gamma = \mathbf{X}(U)$ is “joined to itself” along the x_3 -axis; it is *almost* a self-intersection, but technically the map **is** injective. Note that the topology of Γ is clearly different from that of U , which has no sense of being joined to itself. To be precise, U is a simply connected set, but Γ is not simply connected (only connected).

In other words, \mathbf{X} deforms U and “glues” one edge of it to itself in a nontrivial way. Clearly, there is no unique tangent plane along the x_3 -axis. Ergo, the surface in Figure 2.7 does not look like a plane near the x_3 -axis.

This is because \mathbf{X} is not a homeomorphism, i.e., \mathbf{X}^{-1} is not continuous. To see this, let $\mathcal{N} \subset \mathbb{R}^3$ be a neighborhood that contains the red, Γ_R , and green, Γ_G , parts of the surface shown in Figure 2.7, and consider $\mathbf{X}^{-1}: \mathcal{N} \cap \Gamma \rightarrow U$. Let $\mathbf{r} = (r_1, r_2, 0)^T$ be a point in Γ_R that is arbitrarily close to the origin $\mathbf{0} \in \Gamma_G$. One can show that $\mathbf{X}^{-1}(\mathbf{0}) = (-1, 0)^T \in U$

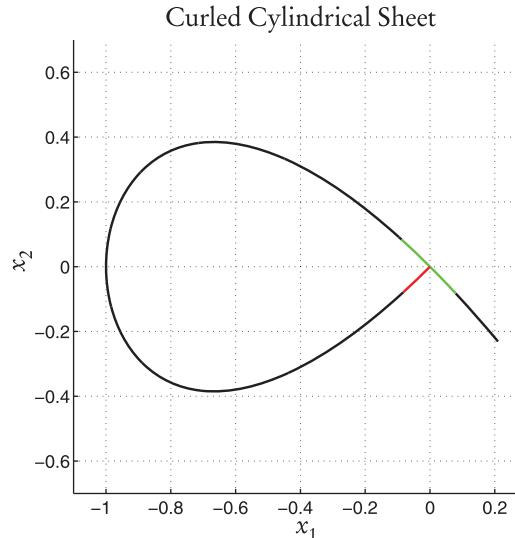


Figure 2.7. A parameterized surface Γ whose parameterization \mathbf{X} is not a homeomorphism (\mathbf{X}^{-1} is not continuous). The x_3 -axis is pointing out of the page, so the surface is viewed edge on. The union of the red and green parts of Γ form a connected set. Clearly, Γ does not have a unique tangent plane along the x_3 -axis.

and $\mathbf{X}^{-1}(\mathbf{r}) = (1 - \epsilon, 0)^T \in U$ for small $\epsilon > 0$. This implies that $|\mathbf{X}^{-1}(\mathbf{r}) - \mathbf{X}^{-1}(0)| = 2 - \epsilon$ and $|\mathbf{r} - 0| < 2\epsilon$ for all sufficiently small ϵ . So \mathbf{X}^{-1} is not continuous.

Therefore, by requiring \mathbf{X} to be a homeomorphism, we eliminate the possibility shown in Figure 2.7. Recall section 2.2.1, where we viewed the reference domain U as a rubber sheet and said that \mathbf{X} must be continuous to avoid “tearing” the rubber sheet when deforming U into the surface Γ . By requiring that \mathbf{X}^{-1} also be continuous, we disallow “inverse tearing,” i.e., we prevent the “gluing” of two disjoint pieces of the rubber sheet when applying the map \mathbf{X} .

Remark 7. Another interpretation of the problem with Figure 2.7, is that it is not possible to “smoothly” transition from the local coordinates on Γ_R to the local coordinates on Γ_G . Indeed, the homeomorphism property is needed to prove Proposition 1, i.e., that a smooth change of variables can be applied to transform between two different overlapping local coordinate systems.

2.3.2 • Atlas of Charts

The Need for Many Maps...

We now elaborate further on the implications of Definition 2. In particular, it implies that for each point \mathbf{x} in Γ , there exists a map \mathbf{X} and an open set U that parameterizes a small portion of the surface Γ that contains \mathbf{x} . So it seems that there is an infinite number of charts (U, \mathbf{X}) associated with any regular surface (recall Remark 6).

Do we really need so many charts to represent a surface? Can we get by with just one chart? If the surface is a plane, then a single map and reference domain are sufficient. But for a general surface, the answer is no. For instance, the sphere cannot be parameterized by only one chart (U, \mathbf{X}) where U is open and \mathbf{X} satisfies the requirements of Definition 2.

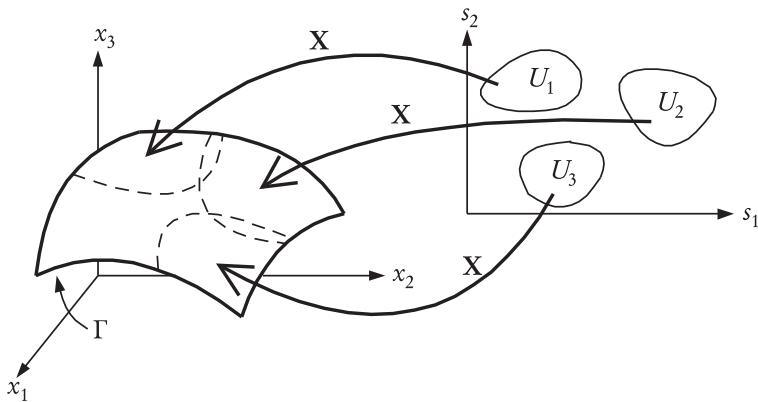


Figure 2.8. Portion of a 2-D regular surface Γ . The mapping is defined on multiple open sets (reference domains); only $\{U_1, U_2, U_3\}$ are shown (and happen to be disjoint). Each U_i is mapped to a small patch (denoted by a dashed curve) on the surface Γ . More than one reference domain is usually required to cover the entire surface Γ in order to fulfill the requirements of Definition 2.

Indeed, suppose we use spherical coordinates to parameterize the unit sphere Γ :

$$\mathbf{X}(\theta, \phi) = (\cos \theta \sin \phi, \sin \theta \sin \phi, \cos \phi)^T,$$

where (θ, ϕ) take values in the reference domain $U = [0, 2\pi] \times [0, \pi]$. The chart (U, \mathbf{X}) covers the entire sphere, but U is not an open set. Moreover, the map \mathbf{X} is not injective because $\mathbf{X}(\theta, 0) = (0, 0, 1)^T$ for any value of θ ; this violates condition 2 of Definition 2. This is sometimes called a *coordinate singularity*, i.e., when multiple sets of reference domain coordinates get mapped to the same point. If we take $U = (0, 2\pi) \times (0, \pi)$, then U and \mathbf{X} satisfy the conditions of Definition 2. But the resulting regular surface is not the entire sphere; it omits two point-sized holes from the north and south poles of the sphere (a twice-punctured sphere).

For computational purposes, it is better to use multiple charts. A general surface may be complicated, and it is easier to parameterize many small local patches than one large region. However, for some of the examples in this book, we abuse Definition 2 and let U not be open to make the definition of a particular surface simpler (see also Remark 9).

...but Not Too Many

Is it possible for a regular surface to require an infinite number of charts to be completely represented? For each point \mathbf{x}_i in Γ , let (U_i, \mathbf{X}_i) be a local chart associated with \mathbf{x}_i (in the sense of Definition 2). So $\{(U_i, \mathbf{X}_i)\}$ is an infinite set of charts. Define $V_i = \mathbf{X}_i(U_i) \subset \Gamma$, which is a nonempty open set because \mathbf{X} is continuous; furthermore, \mathbf{x}_i is in V_i .

The infinite set of open sets $\{V_i\}$ forms a cover of Γ , i.e., $\Gamma \subset \cup_i V_i$. If $\bar{\Gamma}$ is compact (see section 2.1.3), then there exists a **finite** subcover $\{V_{i_k}\}_{k=1}^M$ for some $M < \infty$, such that $\Gamma \subset \cup_{k=1}^M V_{i_k}$ and each V_{i_k} is bounded; see [64] for the technical details. Intuitively, each V_i has nonzero surface area, so we should not need an infinite number of them to cover Γ (note that the technical details are more involved than this).

Therefore, only a finite number of local charts (U_i, \mathbf{X}_i) is required to completely describe, or “map,” Γ . The set of these charts $\{(U_i, \mathbf{X}_i)\}$ forms an *atlas*; see Figure 2.8 and Remark 8. Therefore, a regular surface is usually built up using *multiple* parametric surfaces (recall Definition 1).

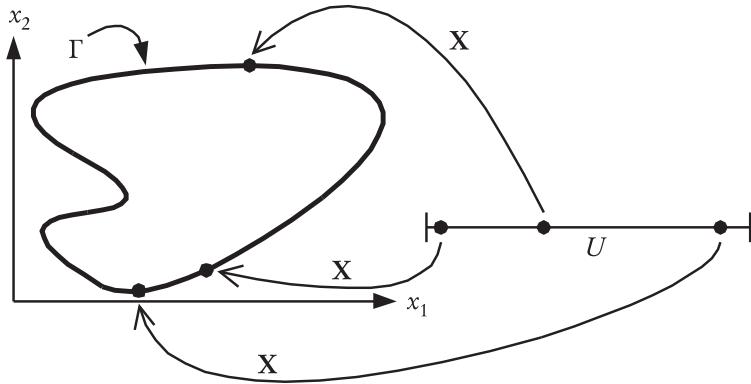


Figure 2.9. A 1-D closed curve Γ with mapping \mathbf{X} . The mapping is defined on a single (simply connected) reference domain U , which is just an interval. Only one reference domain is needed for 1-D curves.

Remark 8 (notation). The sets $\{U_i\}$ are all contained in \mathbb{R}^2 . If $\{U_i\}$ are mutually disjoint, i.e., $U_i \cap U_j = \emptyset$, then we can define \mathbf{X} by

$$\mathbf{X}(s_1, s_2) = \mathbf{X}_i(s_1, s_2) \quad \text{for all } (s_1, s_2)^T \text{ in } U_i, \quad \text{for all } i.$$

In effect, the map is denoted by \mathbf{X} without a subscript; see Figure 2.8. One can always assume the reference domains are mutually disjoint at the expense of applying suitable translations to $\{U_i\}$ and modifying the maps $\{\mathbf{X}_i\}$ accordingly (recall Assumption 2). Of course, it may be convenient to use a single map \mathbf{X} but allow the reference domains to overlap (see the torus example in section 2.3.6).

On the other hand, it may be simpler to take a single fixed reference domain U and many maps $\{\mathbf{X}_i\}$ (see the sphere example in section 2.3.6). Either way, this does not conflict with Definition 2.

The reader should compare the atlas of charts concept with the case of parameterized curves where only one map and one reference domain are required (see Figure 2.9).

2.3.3 • Parameterization by Local Charts

To summarize the above discussion, let $\Gamma \subset \mathbb{R}^3$ be a regular surface. Then there exists a mapping $\mathbf{X}_i : U_i \rightarrow \Gamma$ that parameterizes a “patch” of Γ , i.e., $\mathbf{X}_i(U_i) \subset \Gamma$ for i in some finite index set \mathcal{I} (see Figure 2.8). Each open set $U_i \subset \mathbb{R}^2$ is called a *reference domain* with local variables s_1 and s_2 , i.e., $\mathbf{X}_i(U_i) = \{\mathbf{X}_i(s_1, s_2) : (s_1, s_2)^T \in U_i\}$.

Furthermore, $\{\mathbf{X}_i\}_{i \in \mathcal{I}}$ and $\{U_i\}_{i \in \mathcal{I}}$ satisfy $\cup_{i \in \mathcal{I}} \mathbf{X}_i(U_i) = \Gamma$. Hence, $\{(U_i, \mathbf{X}_i)\}$ forms an atlas of local charts that gives a total surface parameterization of Γ [24, 60]. As convenience dictates, we may have a single map \mathbf{X} with an atlas denoted by $(\{U_i\}, \mathbf{X})$. Note that \mathbf{X} is a vector-valued function with coordinate functions denoted by $\mathbf{X} = (X_1, X_2, X_3)^T$. Alternatively, we may consider a single fixed reference domain U and a collection of maps $\{\mathbf{X}_i\}$. In this case, the atlas of local charts is denoted by $(U, \{\mathbf{X}_i\})$.

2.3.4 • Surfaces without Boundary

Let Γ be a regular surface. To say that Γ is *without boundary* means $\partial\Gamma = \emptyset$. For example, the unit sphere is a surface without boundary. But the surface S , parameterized by the

chart $\mathbf{X}(s_1, s_2) = (s_1, s_2, 0)^T$, where $(s_1, s_2)^T \in (0, 1) \times (0, 1)$, has a boundary. Indeed,

$$\partial S = [\{0\} \times (0, 1)] \cup [\{1\} \times (0, 1)] \cup [(0, 1) \times \{0\}] \cup [(0, 1) \times \{1\}].$$

A *closed* surface is a surface that is compact and without boundary.

2.3.5 • Implicit Surface Representations

A surface Γ can also be represented as the *level set* of a given function $G : \mathbb{R}^3 \rightarrow \mathbb{R}$, i.e., one can define

$$\Gamma = \{\mathbf{x} \in \mathbb{R}^3 : G(\mathbf{x}) = c\} \quad (2.20)$$

for some constant c . If G is at least C^1 in a neighborhood $\mathcal{N} \subset \mathbb{R}^3$ of some point in Γ , and $\partial_i G \neq 0$ in \mathcal{N} (for some fixed $i = 1, 2, 3$), then one can use the implicit function theorem to solve for one of the variables x_1, x_2, x_3 . For example, if $\partial_3 G \neq 0$, then there exists an open set U and an $F : U \rightarrow \mathbb{R}$ such that

$$x_3 = F(x_1, x_2), \quad \text{where } (x_1, x_2, x_3)^T \in \mathcal{N} \cap \Gamma.$$

This yields the following parameterization of Γ :

$$\mathbf{X}(s_1, s_2) = (s_1, s_2, F(s_1, s_2))^T \quad \text{for all } (s_1, s_2)^T \in U,$$

and $\mathbf{X}(U) = \mathcal{N} \cap \Gamma$. Analogous relations hold if $\partial_1 G \neq 0$ or $\partial_2 G \neq 0$.

If Γ is a given regular surface that is *closed*, then one can construct the function G by invoking the signed distance function to Γ , i.e., $\phi : \mathbb{R}^3 \rightarrow \mathbb{R}$ and

$$\phi(\mathbf{x}) = \begin{cases} \text{dist}(\mathbf{x}, \Gamma) & \text{if } \mathbf{x} \text{ is inside } \Gamma, \\ 0 & \text{if } \mathbf{x} \text{ is on } \Gamma, \\ -\text{dist}(\mathbf{x}, \Gamma) & \text{if } \mathbf{x} \text{ is outside } \Gamma, \end{cases} \quad (2.21)$$

where $\text{dist}(\mathbf{x}, \Gamma)$ denotes the minimum distance between \mathbf{x} and Γ . Thus, we can write $\Gamma = \{\mathbf{x} \in \mathbb{R}^3 : \phi(\mathbf{x}) = 0\}$. Moreover, if Γ is regular, then one can show that ϕ is C^∞ at all points sufficiently close to Γ [24, 44].

2.3.6 • Examples

We present some examples of regular surfaces and their parameterizations. We leave it to the reader to prove that these are, indeed, regular surfaces (see the sphere example).

Plane

Consider the planar surface Γ defined by $\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} \cdot \mathbf{N} = 0\}$, where \mathbf{N} is the unit normal vector with components $(N_1, N_2, N_3)^T$ (see Figure 2.10). Assuming $N_3 \neq 0$, we can solve for x_3 :

$$x_3 = \frac{-1}{N_3}(x_1 N_1 + x_2 N_2).$$

We can parameterize the entire Γ by

$$\mathbf{X}(s_1, s_2) = \left(s_1, s_2, -\frac{s_1 N_1 + s_2 N_2}{N_3} \right)^T, \quad (2.22)$$

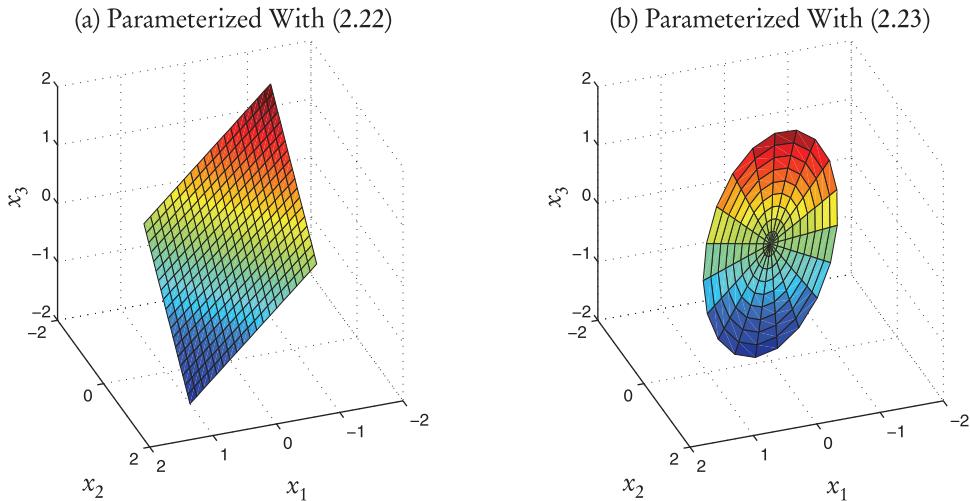


Figure 2.10. Parameterization of a plane with normal vector $\frac{1}{\sqrt{3}}(1,1,1)^T$. (a) A Cartesian coordinate system is placed on the reference domain \mathbb{R}^2 which is vertically projected to the planar surface by \mathbf{X} (see (2.22)). (b) A polar coordinate system is placed on \mathbb{R}^2 and is mapped to the surface by \mathbf{X} (see (2.23)). The surface is identical in both cases, but the local surface coordinates are “drawn” differently.

for all $(s_1, s_2)^T$ in U , where $U = \mathbb{R}^2$. Essentially, the \mathbb{R}^2 plane is projected “vertically” to the planar surface. Note that one can show that the Jacobian matrix (2.11) for this parameterization is constant with rank 2, which implies that condition 3 of Definition 2 is true. In fact, all of the conditions of Definition 2 are true.

Another parameterization is given by introducing polar coordinates on the reference domain U , i.e., consider

$$\mathbf{X}(r, \theta) = \left(r \cos \theta, r \sin \theta, -r \frac{\cos \theta N_1 + \sin \theta N_2}{N_3} \right)^T, \quad (2.23)$$

where $U = [0, \infty) \times [0, 2\pi)$ (see Figure 2.10). Here, the Jacobian is

$$J = [\partial_r \mathbf{X}, \partial_\theta \mathbf{X}] = \begin{bmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \\ -\frac{\cos \theta N_1 + \sin \theta N_2}{N_3} & r \frac{\sin \theta N_1 - \cos \theta N_2}{N_3} \end{bmatrix}.$$

Clearly, the rank of J reduces to 1 when $r = 0$, so condition 3 of Definition 2 is violated. However, this does not imply that Γ is not a regular surface. Definition 2 merely says that there exists a parameterization with the needed properties, for which (2.22) does the job. The singularity appearing in the Jacobian of (2.23) is simply an artifact of using polar coordinates. These types of singularities are often called *coordinate singularities*. In other words, a coordinate singularity has nothing to do with the actual surface, but is due to a “poor” choice of coordinates in the reference domain.

Remark 9. Note that the set U for (2.23) is not an open set; Definition 2 assumes U is open. If we instead use the pair of reference domains

$$U_1 = (0, \infty) \times (-\pi, \pi), \quad U_2 = (0, \infty) \times (0, 2\pi),$$

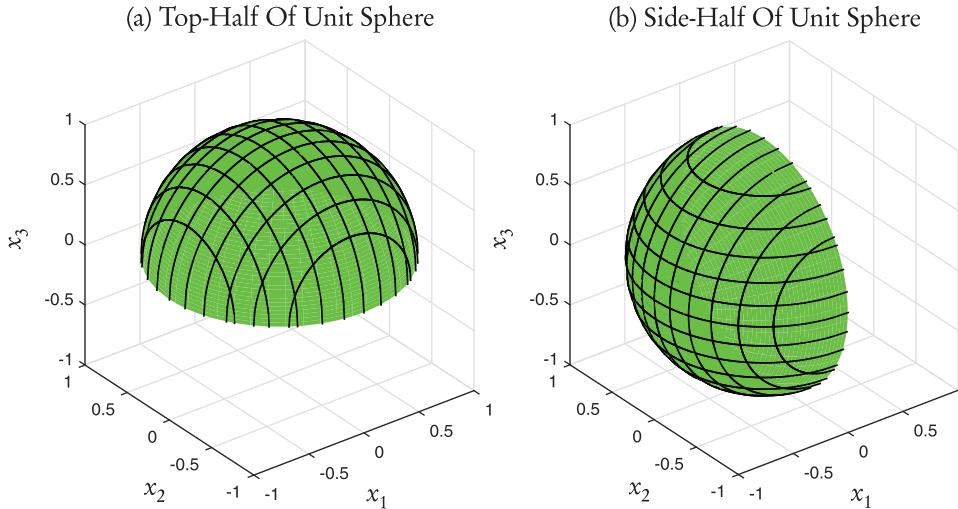


Figure 2.11. Parameterization of the unit sphere. (a) A Cartesian coordinate system is placed on the reference domain U which is mapped to the top of the sphere by \mathbf{X}_1 (see (2.24)). (b) Another parameterization \mathbf{X}_6 maps U to a side-half of the sphere (see (2.26)). Clearly, the coordinate curves are severely distorted near the boundary of the surface patch.

and the parameterization in (2.23), then we get a well-defined surface in the sense of Definition 2. However, these charts omit the origin ($r = 0$), i.e., they yield a parameterization of the plane with a hole in it. This is unavoidable because we are using polar coordinates.

However, it is sometimes convenient to allow U to not be an open set. In this example, it made the parameterization simpler (i.e., only one reference domain was needed). This is useful if you are only concerned with defining a surface. On the other hand, if you want to perform calculus on a surface (see Chapter 4), then you should adopt a set of local charts with open reference domains in the sense of Definition 2.

For simplicity of exposition, some of the examples of regular surfaces in this book have reference domains that are not open. It is implicitly understood that these surfaces can be re-parameterized with a new set of local charts with open reference domains (we leave the details to the reader).

Sphere

Consider the unit sphere Γ defined by $\{(x_1, x_2, x_3)^T \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 = 1\}$ (see Figure 2.11). We can parameterize the “top” of Γ by

$$\mathbf{X}_1(s_1, s_2) = \left(s_1, s_2, +\sqrt{1 - (s_1^2 + s_2^2)} \right)^T \quad (2.24)$$

for all $(s_1, s_2)^T$ in U , where $U = \{(s_1, s_2)^T \in \mathbb{R}^2 : s_1^2 + s_2^2 < 1\}$.

To prove that Γ is a regular surface, we must check the conditions of Definition 2. First, note that \mathbf{X}_1 is C^∞ on U , so condition 1 is true. Next, it is easy to check that \mathbf{X}_1 is injective. Moreover, the inverse map is given by

$$\mathbf{X}_1^{-1}(x_1, x_2, x_3) = (x_1, x_2)^T$$

for all $(x_1, x_2, x_3)^T$ in $\mathbf{X}(U) \subset \Gamma$. Clearly, \mathbf{X}_1^{-1} is continuous, so condition 2 is satisfied.

Condition 3 follows by verifying that $\partial_{s_1} \mathbf{X}_1(s_1, s_2)$ and $\partial_{s_2} \mathbf{X}_1(s_1, s_2)$ are linearly independent vectors in \mathbb{R}^3 for all $(s_1, s_2)^T$ in U . This proves that $\mathbf{X}_1(U) \subset \Gamma$ is a regular surface.

To prove it for all of Γ , we must **cover** Γ with parameterizations similar to (2.24). For instance, the bottom half of Γ is given by

$$\mathbf{X}_2(s_1, s_2) = (s_1, s_2, -\sqrt{1-(s_1^2 + s_2^2)})^T \quad \text{for all } (s_1, s_2)^T \in U. \quad (2.25)$$

However, the equator is not covered by $\mathbf{X}_1(U) \cup \mathbf{X}_2(U)$. Hence, we need the following additional parameterizations:

$$\begin{aligned} \mathbf{X}_3(s_1, s_2) &= (s_1, +\sqrt{1-(s_1^2 + s_2^2)}, s_2)^T, \\ \mathbf{X}_4(s_1, s_2) &= (s_1, -\sqrt{1-(s_1^2 + s_2^2)}, s_2)^T, \\ \mathbf{X}_5(s_1, s_2) &= (+\sqrt{1-(s_1^2 + s_2^2)}, s_1, s_2)^T, \\ \mathbf{X}_6(s_1, s_2) &= (-\sqrt{1-(s_1^2 + s_2^2)}, s_1, s_2)^T. \end{aligned} \quad (2.26)$$

One can check that $\Gamma = \bigcup_{i=1}^6 \mathbf{X}_i(U)$, where the same reference domain is used for each i . All of the above statements about \mathbf{X}_1 apply to $\{\mathbf{X}_i\}_{i=1}^6$ as well. Thus, Γ is a regular surface.

One can also use multiple parameterizations with spherical coordinates to show that Γ is a regular surface. We leave the details to the reader.

Torus

Consider the torus Γ parameterized by

$$\mathbf{X}(s_1, s_2) = \begin{pmatrix} (a + r \cos s_1) \cos s_2 \\ (a + r \cos s_1) \sin s_2 \\ r \sin s_1 \end{pmatrix} \quad (2.27)$$

for all $(s_1, s_2)^T$ in U , where $U = [0, 2\pi] \times [0, 2\pi]$ and $a = 1.3$, $r = 0.5$ (see Figure 2.12). This is also a regular surface. To prove this, you must introduce multiple parameterizations, each satisfying the conditions of Definition 2, to cover Γ . This can be done by restricting (2.27) to the following reference domains:

$$\begin{aligned} U_1 &= (0, 1.1\pi) \times (0, 1.1\pi), & U_2 &= (\pi, 2.1\pi) \times (0, 1.1\pi), \\ U_3 &= (0, 1.1\pi) \times (\pi, 2.1\pi), & U_4 &= (\pi, 2.1\pi) \times (\pi, 2.1\pi). \end{aligned}$$

2.3.7 ■ Change of Parameters

According to Definition 2, each point P of a regular surface Γ is surrounded by a coordinate neighborhood \mathcal{N} . All of the points in \mathcal{N} are characterized (located) by their coordinates. Thus, any local properties of Γ can be defined in terms of these coordinates.

However, a point can belong to multiple coordinate neighborhoods, e.g., recalling the sphere example in Figure 2.11, the point $(-1, -1, 1)^T$ belongs to $\mathbf{X}_1(U)$, $\mathbf{X}_4(U)$, and $\mathbf{X}_6(U)$. Moreover, other coordinate systems could also be introduced.

Therefore, it is necessary that *all* coordinate neighborhoods that contain P be “equivalent.” In other words, if P belongs to two coordinate neighborhoods with local parameters (s_1, s_2) and $(\tilde{s}_1, \tilde{s}_2)$, then it must be possible to transform from one pair of coordinates

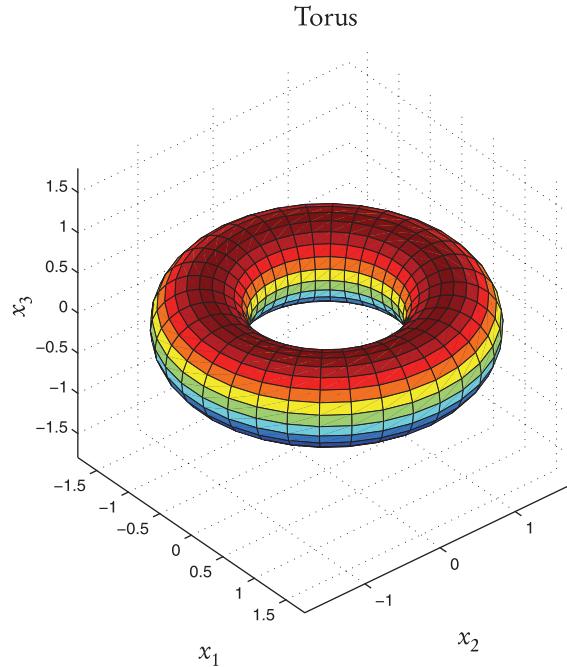


Figure 2.12. Parameterization of a torus. A Cartesian coordinate system is placed on the reference domain U which is mapped to the torus by \mathbf{X} (see (2.27)).

to the other in a “smooth” way (recall Assumption 2). The following result confirms this; see [24, sect. 2-3, Prop. 1].

Proposition 1 (change of parameters). *Let P be a point of a regular surface Γ , and let $\mathbf{X}: U \rightarrow \Gamma$ and $\tilde{\mathbf{X}}: \tilde{U} \rightarrow \Gamma$, with $U, \tilde{U} \subset \mathbb{R}^2$, be two parameterizations of Γ such that P is in $\mathbf{X}(U) \cap \tilde{\mathbf{X}}(\tilde{U}) =: W$. Then the change of variables*

$$\mathbf{Y} := \mathbf{X}^{-1} \circ \tilde{\mathbf{X}}, \quad \text{where} \quad \mathbf{Y}: \tilde{\mathbf{X}}^{-1}(W) \rightarrow \mathbf{X}^{-1}(W),$$

is differentiable and has a differentiable inverse \mathbf{Y}^{-1} . In fact, \mathbf{Y} is a diffeomorphism (see Definition 15 in Chapter 5).

Basically, Proposition 1 implies that it *does not matter* how you parameterize the surface, i.e., the parameterization is an arbitrary choice.

2.4 ■ The Tangent Space

2.4.1 ■ Flatland

If some phenomenon is inherently limited to the \mathbb{R}^2 (x_1 - x_2) plane, then it is not aware of, or affected by, any other dimensions, such as the x_3 -axis. For example, the motion of cars on a highway is unaffected (to some degree) by what happens above them. When analyzing the phenomenon, you are restricted to the tools of \mathbb{R}^2 , e.g., ∂_{x_1} and ∂_{x_2} .

Now consider a surface Γ that is a plane in \mathbb{R}^3 . Suppose there is some phenomenon that we want to analyze that *only exists within* Γ . If we want to perform calculus, differential

equations, analysis, etc., within Γ , it is not so difficult. Simply rotate and translate the coordinate system so that Γ is contained in the x_1 - x_2 plane. Then apply all the standard tools in \mathbb{R}^2 , compute the desired quantities, etc., and when all is said and done just map the results back to the original coordinate system. When the surface is a plane in \mathbb{R}^3 , it is essentially no different than working with \mathbb{R}^2 .

When Γ is a curved surface, the situation is a little more involved. But even here we can invoke the **tangent plane** to the surface (at a point) to at least locally develop a calculus on Γ , just as we did for a planar surface. In other words, we will “map” the derivative (and integral) operators of \mathbb{R}^2 onto Γ (we do this in Chapter 4). The maxim here is that *whatever happens in the surface stays in the surface*, i.e., we will avoid directly using the structure of the ambient space \mathbb{R}^3 . (Alternative approaches using level sets and distance functions on \mathbb{R}^3 can be found in [23, 26, 27, 93].) However, we emphasize that the geometry of the surface in \mathbb{R}^3 does affect “life within the surface.”

A popular interpretation of this can be found in the book *Flatland* [1], where hypothetical 2-D creatures live in a surface. None of these creatures are directly aware of the higher dimensional space that contains their “universe” Γ . Yet the curved nature of the surface does affect them indirectly [108].

2.4.2 • Curves within Surfaces

The first step in developing analytical tools for a regular surface $\Gamma \subset \mathbb{R}^3$ is to consider curves *within* the surface. Let $(\{U_i\}, \mathbf{X})$ be the atlas of charts for Γ . Any smooth curve Σ contained in Γ can be (locally) parameterized using \mathbf{X} and a particular reference domain. For instance, for some neighborhood \mathcal{N} , we have that $\mathbf{X}^{-1}(\mathcal{N} \cap \Sigma)$ is a curve \mathcal{C} contained in U_k for some k . Hence, let $\mathbf{s} : I \rightarrow U_k \subset \mathbb{R}^2$ be a parameterization of \mathcal{C} , where $I \subset \mathbb{R}$ is a finite interval, i.e.,

$$\mathbf{s} = (s_1, s_2)^T, \quad s_1 = s_1(t), \quad s_2 = s_2(t), \quad t \in I. \quad (2.28)$$

Then, a (local) parameterization of Σ is denoted $\alpha : I \rightarrow \Sigma$ and defined by

$$\alpha(t) = \mathbf{X} \circ \mathbf{s}(t) = \mathbf{X}(\mathbf{s}(t)) \quad \text{for all } t \in I. \quad (2.29)$$

A total parameterization of Σ can be built from piecing together several local parameterizations.

Helix on a Cylinder

An example of a curve on a surface is as follows. Consider the cylinder of radius R , denoted Γ , parameterized by

$$\mathbf{X}(s_1, s_2) = \begin{pmatrix} R \cos s_1 \\ R \sin s_1 \\ s_2 \end{pmatrix} \quad (2.30)$$

for all $(s_1, s_2)^T$ in U , where $U = [-\pi, \pi] \times (-1, 1)$ (see Figure 2.13).

A helix on Γ can be parameterized by first defining $\mathbf{s}(t) = (ct, t)^T$, for some constant $c \neq 0$, followed by

$$\alpha(t) = \mathbf{X}(\mathbf{s}(t)) = \begin{pmatrix} R \cos(ct) \\ R \sin(ct) \\ t \end{pmatrix}, \quad -1 < t < 1. \quad (2.31)$$

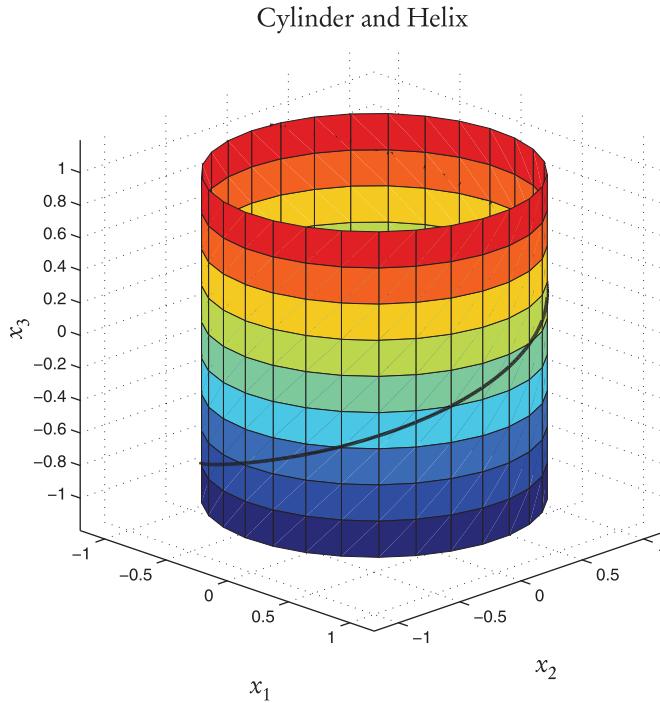


Figure 2.13. Parameterization of a cylinder and helix. A Cartesian coordinate system is placed on the reference domain U which is mapped to the cylinder by \mathbf{X} (see (2.30)). The helical (black) curve is parameterized by (2.31).

Coordinate Curves

We can also consider $s_1 = \text{constant}$ or $s_2 = \text{constant}$ curves on a surface. For example, $\mathbf{X} \circ \mathbf{s}$, with $\mathbf{s}(t) = (c_0, t)^T$ and \mathbf{X} given by (2.30), yields a parameterization of the vertical lines on the cylinder in Figure 2.13 for various values of the constant c_0 . Similarly, $\mathbf{X} \circ \mathbf{s}$, with $\mathbf{s}(t) = (t, c_0)^T$, parameterizes the circle curves around the cylinder. The curves are called *coordinate curves*, which correspond to parallels to the coordinate axes in the s_1-s_2 plane. They provide a way to “map out” the surface.

2.4.3 ■ Tangent Plane/Space

The notion of a tangent plane is related to tangent vectors of curves in a surface. Let $\alpha : I \rightarrow \mathbb{R}^3$ be a parameterized curve, where I is a bounded open interval in \mathbb{R} . Denote the set of points that make up the curve by $\Sigma = \alpha(I)$. The *tangent vector* of Σ , at a point \mathbf{x} in Σ , is defined by $\alpha'(t_x)$, where ' denotes differentiation and $\mathbf{x} = \alpha(t_x)$, i.e., $t_x = \alpha^{-1}(\mathbf{x})$.

If Σ is a subset of a surface Γ , with local chart (U, \mathbf{X}) , then α can be defined by $\alpha = \mathbf{X} \circ \mathbf{s}$ for some appropriate function $\mathbf{s} : I \rightarrow U$. The formula for the tangent vector now expands because of the chain rule as follows:

$$\alpha'(t) = \frac{d}{dt} \alpha(t) = \left(\frac{d}{dt} \mathbf{s}(t) \cdot \nabla \right) \mathbf{X}(\mathbf{s}(t)) = s'_1(t) \partial_{s_1} \mathbf{X} + s'_2(t) \partial_{s_2} \mathbf{X}. \quad (2.32)$$

The vector α' is a linear combination of the vectors $\partial_{s_1} \mathbf{X}$ and $\partial_{s_2} \mathbf{X}$. Indeed, if Σ is a coordinate curve (i.e., $\mathbf{s}(t) = (c_0, t)^T$ or $\mathbf{s}(t) = (t, c_0)^T$), then α' is simply a multiple of

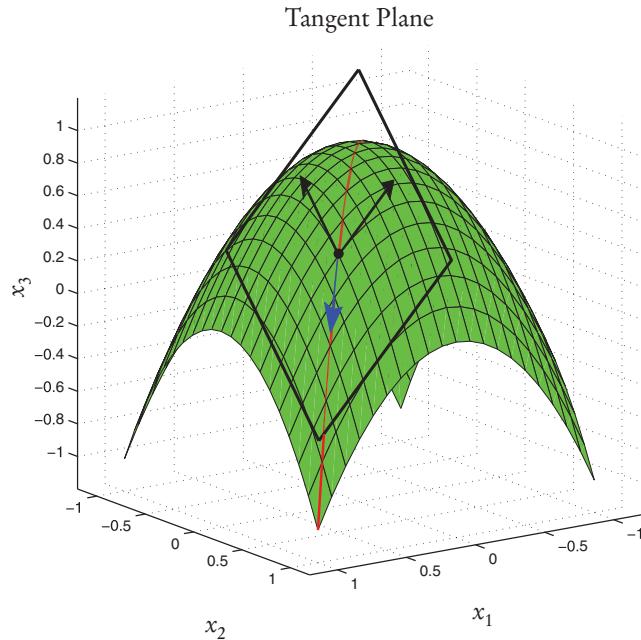


Figure 2.14. Illustration of a tangent plane to a surface Γ (green). At the point $x_0 = (0.5, 0.5, 0.5)^T$, the tangent plane $T_{x_0}(\Gamma)$ contains the vectors $\partial_{s_1}X$ and $\partial_{s_2}X$ (black arrows) emanating from the point x_0 ; $\partial_{s_1}X$, $\partial_{s_2}X$ are tangent to the coordinate curves passing through x_0 . The red curve in Γ has a tangent vector at x_0 (blue) pointing in the plane $T_{x_0}(\Gamma)$.

either $\partial_{s_1}X$ or $\partial_{s_2}X$. Ergo, if x_0 is a fixed point on Σ , such that $x_0 = X(s_0)$, then $\partial_{s_1}X(s_0)$ and $\partial_{s_2}X(s_0)$ are tangential (at x_0) to the coordinate curves passing through x_0 . See Figure 2.14 for an illustration.

Assuming that Γ is a regular surface at x_0 , $\partial_{s_1}X$ and $\partial_{s_2}X$ are then linearly independent vectors. These vectors span a plane denoted $T_{x_0}(\Gamma)$, which is called the *tangent plane* of Γ at the point x_0 . Note that sometimes we write $T_P(\Gamma)$ for the tangent plane at the point P . Hence, because of (2.32), $T_P(\Gamma)$ contains the tangent vector of *any* curve on Γ (at P) passing through P .

The tangent plane $T_{x_0}(\Gamma)$ is itself a surface that can be parameterized, say by

$$Z(p_1, p_2) = x_0 + p_1 \partial_{s_1}X(s_0) + p_2 \partial_{s_2}X(s_0), \quad \text{where } x_0 = X(s_0). \quad (2.33)$$

Referring back to (2.14), we can write (2.33) as

$$Z(p) = x_0 + T_{s_0}(p),$$

where $p = (p_1, p_2)^T$ (see Figure 2.14). Since $T_{s_0}(\mathbb{R}^2)$ is a vector space, we often refer to $T_{x_0}(\Gamma)$ as the *tangent space* of Γ at x_0 . In summary, we have the following result from [24, sect. 2-4, Prop. 1].

Proposition 2. Let $U \subset \mathbb{R}^2$ and $X : U \rightarrow \Gamma$ be a parameterization of a regular surface Γ , and let $s \in U$. The vector subspace of dimension 2,

$$T_s(\mathbb{R}^2) \subset \mathbb{R}^3,$$

is equal to the set of tangent vectors to Γ at $X(s)$.

2.5 ▪ Minimal Regularity?

The definition of regular surface requires the parameterization \mathbf{X} to be C^∞ . However, only C^1 is required to make sense of the tangent plane. And many other investigations of differential geometry, such as curvature, Willmore flow, etc., do not require more than, say, four derivatives.

The degree of differentiability of surfaces can be important, especially when studying PDE regularity theory. But it is not our purpose here to get hung up on this type of question. Henceforth, we will assume that we have enough differentiability to allow all calculations to go through in order to avoid complicating the fundamental geometric ideas.

Chapter 3

The Fundamental Forms of Differential Geometry

Quantifying the shape of surfaces requires the so-called fundamental forms. The first fundamental form is linked to the concept of a *metric* on the reference domain of the surface parameterization. The second fundamental form is related to the *curvature* of the surface.

3.1 • First Fundamental Form

Lengths, angles, and areas in the reference domain U change (or distort) when we map to the domain $\mathbf{X}(U)$. For example, the area of a region $R \subset U$ is different from the area of $\mathbf{X}(R)$. This is not surprising when the action of $\mathbf{X}(\cdot)$ is interpreted as a deformation (recall sections 2.1.6 and 2.2.1). Quantifying this distortion requires the so-called first fundamental form, which is essential to all aspects of differential geometry.

3.1.1 • Measuring the Length of Curves

Differential Arc-length

It is well-known that the arc-length of a curve Σ , parameterized by $\alpha : I \rightarrow \Sigma$, is given by

$$\text{arc-length} = \int_I |\alpha'(t)| dt.$$

Thus, denoting the *differential arc-length* by $d\alpha$, we have

$$d\alpha = |\alpha'(t)| dt = \sqrt{\alpha'(t) \cdot \alpha'(t)} dt. \quad (3.1)$$

Here, $|\cdot|$ is the standard Euclidean norm.

Recalling the discussion in section 2.4.2, when Σ lies in a regular surface Γ , with parameterization $\mathbf{X} : U \rightarrow \Gamma$, we can think of Σ as being obtained by deforming an underlying curve in U . So, plugging $\alpha = \mathbf{X} \circ s$ into $|\alpha'(t)|$, and using the chain rule as in (2.32), we get

$$\begin{aligned} |\alpha'(t)|^2 &= \alpha'(t) \cdot \alpha'(t) \\ &= (s'_1(t) \partial_{s_1} \mathbf{X} + s'_2(t) \partial_{s_2} \mathbf{X}) \cdot (s'_1(t) \partial_{s_1} \mathbf{X} + s'_2(t) \partial_{s_2} \mathbf{X}) \\ &= (s'_1)^2 \partial_{s_1} \mathbf{X} \cdot \partial_{s_1} \mathbf{X} + 2s'_1 s'_2 \partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \mathbf{X} + (s'_2)^2 \partial_{s_2} \mathbf{X} \cdot \partial_{s_2} \mathbf{X}, \end{aligned}$$

where we suppressed the t notation in the last line. Using $ds_i = s'_i(t)dt$ ($i = 1, 2$), we can see how \mathbf{X} affects the differential arc-length, i.e.,

$$d\alpha^2 = \alpha'(t) \cdot \alpha'(t) dt^2 = ds_1^2 \partial_{s_1} \mathbf{X} \cdot \partial_{s_1} \mathbf{X} + 2ds_1 ds_2 \partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \mathbf{X} + ds_2^2 \partial_{s_2} \mathbf{X} \cdot \partial_{s_2} \mathbf{X}.$$

Introducing the notation

$$g_{ij} = \partial_{s_i} \mathbf{X} \cdot \partial_{s_j} \mathbf{X} \quad \text{for } 1 \leq i, j \leq 2, \quad (3.2)$$

where clearly $g_{12} = g_{21}$, we have

$$d\alpha^2 = ds_1^2 g_{11} + 2ds_1 ds_2 g_{12} + ds_2^2 g_{22} = \sum_{i,j=1}^2 g_{ij} ds_i ds_j. \quad (3.3)$$

This quadratic form is called the *first fundamental form*. Equivalently, we can write

$$|\alpha'(t)|^2 = \sum_{i,j=1}^2 g_{ij} s'_i s'_j. \quad (3.4)$$

The first fundamental form is actually defined on the tangent space of Γ , which the following definition formalizes.

Definition 3. *The first fundamental form of the surface $\Gamma \subset \mathbb{R}^3$, at the point P in Γ , is a quadratic form $I_P(\cdot)$ defined on $T_P(\Gamma)$ by*

$$I_P(\mathbf{v}) = \sum_{i,j=1}^2 g_{ij} a_i a_j, \quad \text{where } a_1, a_2 \in \mathbb{R} \quad \text{and} \quad \mathbf{v} = a_1 \partial_{s_1} \mathbf{X} + a_2 \partial_{s_2} \mathbf{X}, \quad (3.5)$$

where $\mathbf{v} \in T_P(\Gamma)$ is expressed in the tangent basis $\{\partial_{s_1} \mathbf{X}, \partial_{s_2} \mathbf{X}\}$ evaluated at P .

For instance, if $P = \alpha(0)$, the reader can verify that $I_P(\alpha'(0)) = |\alpha'(0)|^2$. Hence, $I_P(\mathbf{v})$ computes the Euclidean norm squared (with respect to \mathbb{R}^3 !) of \mathbf{v} when \mathbf{v} is in the tangent space.

Deforming the Reference Domain

The coefficients in (3.3) give a “measure” of how much each differential element in the reference domain U is “deformed” from applying the map \mathbf{X} . Recall the parameterization of the plane given in (2.22). A straightforward calculation shows that the coefficients of the first fundamental form are

$$g_{11} = 1 + (N_1/N_3)^2, \quad g_{12} = g_{21} = (N_1/N_3)(N_2/N_3), \quad g_{22} = 1 + (N_2/N_3)^2,$$

which are all constant. This means the amount of deformation or distortion is constant over the whole surface, i.e., in Figure 2.10(a), one can see that the reference domain is stretched vertically in a uniform manner.

Next, consider the parameterization of the sphere in (2.24), i.e.,

$$\begin{aligned} \mathbf{X}(s_1, s_2) &= (s_1, s_2, \sqrt{1 - (s_1^2 + s_2^2)})^T, \quad \text{where } s_1^2 + s_2^2 < 1, \\ \partial_{s_1} \mathbf{X}(s_1, s_2) &= (1, 0, -s_1(1 - (s_1^2 + s_2^2))^{-1/2})^T, \\ \partial_{s_2} \mathbf{X}(s_1, s_2) &= (0, 1, -s_2(1 - (s_1^2 + s_2^2))^{-1/2})^T. \end{aligned} \quad (3.6)$$

The coefficients of the first fundamental form are

$$g_{11} = \frac{1 - s_2^2}{1 - (s_1^2 + s_2^2)}, \quad g_{12} = g_{21} = \frac{s_1 s_2}{1 - (s_1^2 + s_2^2)}, \quad g_{22} = \frac{1 - s_1^2}{1 - (s_1^2 + s_2^2)}, \quad (3.7)$$

which all blow up for $|(\dot{s}_1, \dot{s}_2)| \rightarrow 1$. This means the amount of deformation or distortion increases dramatically near the boundary (equator) of the hemisphere (note the coordinate curves in Figure 2.11(a)).

Computing Arc-length

The arc-length of $\alpha(I) = \mathbf{X} \circ \mathbf{s}(I)$ is computed in the following way. Using (3.3) and $ds_i = s'_i(t) dt$ ($i = 1, 2$), we obtain

$$\text{arc-length} = \int_I d\alpha = \int_I \left(\sum_{i,j=1}^2 g_{ij} s'_i(t) s'_j(t) \right)^{1/2} dt. \quad (3.8)$$

Next, we consider *approximating* the arc-length. This is accomplished by decomposing I into many subintervals $\{I_k\}_{k=1}^N$, where $I_k := [t_{k-1}, t_k]$, and approximating (3.8) over each I_k . Since the integrand of (3.8) is positive (see Proposition 3), we can use the mean value theorem for integrals on I_k to get

$$\begin{aligned} \Delta\alpha_k &:= \int_{I_k} d\alpha = \int_{I_k} \left(\sum_{i,j=1}^2 g_{ij} s'_i(t) s'_j(t) \right)^{1/2} dt = \left(\sum_{i,j=1}^2 g_{ij} s'_i(t) s'_j(t) \right)^{1/2} \Big|_{t=\xi_k} \int_{I_k} dt \\ &= \left(\sum_{i,j=1}^2 g_{ij}(\mathbf{s}(\xi_k)) s'_i(\xi_k) s'_j(\xi_k) \right)^{1/2} |I_k|, \end{aligned} \quad (3.9)$$

where $\xi_k \in I_k$. For convenience, let us assume that $|I_k| = t_k - t_{k-1} = \Delta t = |I|/N$ for all k . If Δt is sufficiently small, we can approximate $s'_i(\xi_k)$ by

$$s'_i(\xi_k) = \frac{\Delta s_{i,k}}{\Delta t} + O(\Delta t) \quad \text{for } i = 1, 2,$$

where $\Delta s_{i,k} = s_i(t_k) - s_i(t_{k-1})$ for $i = 1, 2$. Since we do not know ξ_k precisely, it is convenient to replace ξ_k with t_k in $g_{ij}(\mathbf{s}(\xi_k))$. By using multiple Taylor expansions (we omit the details), one can show

$$\begin{aligned} \Delta\alpha_k &= \left(\sum_{i,j=1}^2 g_{ij}(\mathbf{s}(t_k)) \Delta s_{i,k} \Delta s_{j,k} \right)^{1/2} + O(\Delta t^2) \\ &\approx \left(\sum_{i,j=1}^2 g_{ij}(\mathbf{s}(t_k)) \Delta s_{i,k} \Delta s_{j,k} \right)^{1/2}. \end{aligned} \quad (3.10)$$

Therefore, we can approximate the arc-length by

$$\text{approximate arc-length} = \sum_{k=1}^N \Delta\alpha_k = \sum_{k=1}^N \left(\sum_{i,j=1}^2 g_{ij}(\mathbf{s}(t_k)) \Delta s_{i,k} \Delta s_{j,k} \right)^{1/2}. \quad (3.11)$$

The error between (3.8) and (3.11) is order Δt . So as long as the partition is fine enough, the approximation is good.

3.1.2 • The Metric Tensor

The first fundamental form provides a way of *measuring* arc-length, angles, and areas on a surface (sections 3.1.4 and 3.1.5). Therefore, one can say that it defines a “metric” on the surface.

The coefficients g_{ij} can be grouped into a 2×2 matrix denoted g :

$$g = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}, \quad (3.12)$$

which is a **symmetric** matrix because $g_{12} = g_{21}$. We often refer to g as the *metric tensor* for the surface. Actually, g is the matrix realization of the metric tensor (but we will not belabor this point).

Remark 10. When dealing with surfaces of higher dimension the matrix g has a larger size. For example, if $\mathbf{X}: \mathbb{R}^n \rightarrow \mathbb{R}^m$, then g is an $n \times n$ matrix.

The inverse of the matrix g is given by

$$g^{-1} = \begin{bmatrix} g^{11} & g^{12} \\ g^{21} & g^{22} \end{bmatrix} = \frac{1}{\det(g)} \begin{bmatrix} g_{22} & -g_{12} \\ -g_{21} & g_{11} \end{bmatrix}, \quad (3.13)$$

where we denote the coefficients of the inverse with a superscript for the indices. Of course, we have the following property because $g g^{-1} = I$:

$$\delta_{ij} = \sum_{k=1}^2 g_{ik} g^{kj} = \sum_{k=1}^2 g_{ik} g^{jk}, \quad (3.14)$$

where δ_{ij} is the “Kronecker delta”:

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j. \end{cases} \quad (3.15)$$

A metric defined by a quadratic differential form (e.g., (3.3)) is called a *Riemannian metric*. The corresponding geometry is a *Riemannian geometry*, and the space where such a metric has been introduced is a *Riemannian space*. Therefore, surfaces with a metric defined by (3.3) are 2-D *Riemannian spaces*, and the geometry in these surfaces is a *Riemannian geometry*.

We can rewrite the quadratic form (3.5) in terms of matrix notation, i.e.,

$$I_p(\mathbf{v}) = \mathbf{a}^T g \mathbf{a} = \mathbf{a} \cdot g \mathbf{a}, \quad \text{where } \mathbf{a} = (a_1, a_2)^T \in \mathbb{R}^2, \quad \text{and } \mathbf{v} = a_1 \partial_{s_1} \mathbf{X} + a_2 \partial_{s_2} \mathbf{X}. \quad (3.16)$$

Since $I_p(\mathbf{v})$ is the squared Euclidean \mathbb{R}^3 norm of \mathbf{v} , this suggests that g is a positive definite matrix. The next result confirms this property.

Proposition 3. At regular points of a surface, the metric tensor (matrix) g (3.12) is positive definite, i.e., $\mathbf{a}^T g \mathbf{a} > 0$ for all $\mathbf{a} \in \mathbb{R}^2$ with $\mathbf{a} \neq 0$.

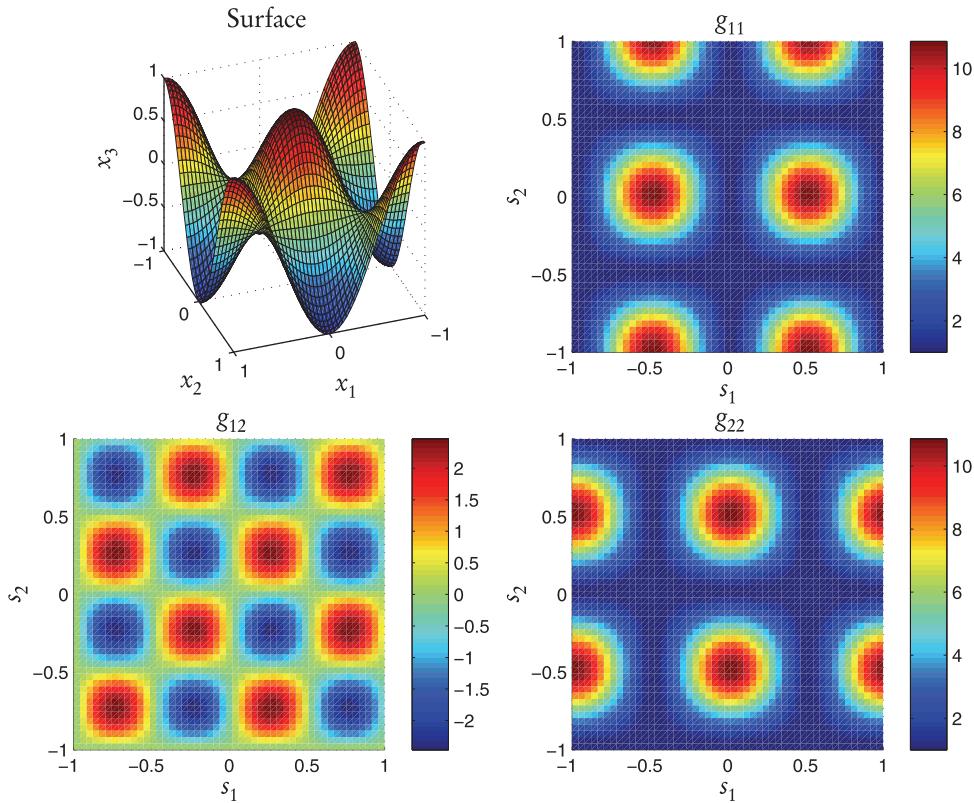


Figure 3.1. Illustration of the components of the first fundamental form for the surface given by (3.18). Note that g_{11} is largest when the surface is most stretched in the s_1 direction; similarly, g_{22} is largest when the surface is most stretched in the s_2 direction. g_{12} is close to zero when the coordinate curves are least distorted (see Lemma 3.3).

Proof. Recalling the Jacobian matrix (2.11), it is easy to see that

$$g = J^T J = (\nabla_s \mathbf{X})^T (\nabla_s \mathbf{X}). \quad (3.17)$$

Thus, we must show that $\mathbf{a}^T J^T J \mathbf{a} > 0$ for all $\mathbf{a} \in \mathbb{R}^2$ with $\mathbf{a} \neq 0$. Since the surface is regular, we know that J has rank 2. This means that $\mathbf{b} := J \mathbf{a} \in \mathbb{R}^3$ and $\mathbf{b} \neq 0$. Therefore, the result follows from

$$\mathbf{a}^T J^T J \mathbf{a} = \mathbf{b}^T \mathbf{b} = |\mathbf{b}|^2 > 0. \quad \square$$

Remark 11. Proposition 3 implies that g has real positive eigenvalues and $\det g > 0$ for any regular surface.

A visualization of the components of g is shown in Figure 3.1, where the surface parameterization is

$$\begin{aligned} \mathbf{X}(s_1, s_2) &= (s_1, s_2, \cos(\pi s_1) \cos(\pi s_2))^T \quad \text{on} \quad U = (-1, 1) \times (-1, 1), \\ \partial_{s_1} \mathbf{X}(s_1, s_2) &= (1, 0, -\pi \sin(\pi s_1) \cos(\pi s_2))^T, \\ \partial_{s_2} \mathbf{X}(s_1, s_2) &= (0, 1, -\pi \cos(\pi s_1) \sin(\pi s_2))^T. \end{aligned} \quad (3.18)$$

The components of g for this surface are

$$\begin{aligned} g_{11} &= 1 + (\pi \sin(\pi s_1) \cos(\pi s_2))^2, & g_{12} = g_{21} &= \pi^2 \sin(\pi s_1) \cos(\pi s_2) \cos(\pi s_1) \sin(\pi s_2), \\ g_{22} &= 1 + (\pi \cos(\pi s_1) \sin(\pi s_2))^2. \end{aligned} \quad (3.19)$$

We close with a result on how g transforms under a change of parameters.

Lemma 3.1. *Let Γ be a regular surface. Consider two local charts (U, \mathbf{X}) and $(\tilde{U}, \tilde{\mathbf{X}})$ that parameterize a region R such that $\mathbf{X}(U) = \tilde{\mathbf{X}}(\tilde{U}) = R \subset \Gamma$. Let $\mathbf{Y} := \mathbf{X}^{-1} \circ \tilde{\mathbf{X}}$ (recall Proposition 1 in Chapter 2), where $\mathbf{Y}(\tilde{U}) = U$. Then,*

$$g \circ \mathbf{Y}(\tilde{s}) = [\mathbf{M}^{-1}]^T \tilde{g}(\tilde{s}) \mathbf{M}^{-1}, \quad \mathbf{M} = [\nabla_{\tilde{s}} \mathbf{Y}(\tilde{s})], \quad (3.20)$$

where g and \tilde{g} are the first fundamental form matrices associated with (U, \mathbf{X}) and $(\tilde{U}, \tilde{\mathbf{X}})$, respectively.

Proof. Note that $\tilde{U} = \tilde{\mathbf{X}}^{-1}(R)$. Referring to Appendix B.1.2, we have that

$$(\nabla_s \mathbf{X}) \circ \mathbf{Y}(\tilde{s}) = \underbrace{\nabla_{\tilde{s}} \tilde{\mathbf{X}}(\tilde{s})}_{3 \times 2 \text{ matrix}} \mathbf{M}^{-1}, \quad \text{where } \mathbf{M} = [\nabla_{\tilde{s}} \mathbf{Y}(\tilde{s})] \text{ is a } 2 \times 2 \text{ matrix.} \quad (3.21)$$

In (3.17), we noted that $g = J^T J$, where $J = \nabla_s \mathbf{X}$. Thus, using (3.21), we see that

$$\begin{aligned} g \circ \mathbf{Y}(\tilde{s}) &= (\nabla_{\tilde{s}} \tilde{\mathbf{X}}(\tilde{s}) \mathbf{M}^{-1})^T (\nabla_{\tilde{s}} \tilde{\mathbf{X}}(\tilde{s}) \mathbf{M}^{-1}) \\ &= [\mathbf{M}^{-1}]^T \tilde{J}^T \tilde{J} \mathbf{M}^{-1} = [\mathbf{M}^{-1}]^T \tilde{g}(\tilde{s}) \mathbf{M}^{-1}. \quad \square \end{aligned}$$

3.1.3 ■ Normal Vector

Definition

From section 2.4.3, we know that for any point \mathbf{x} of a regular surface Γ , $\partial_{s_1} \mathbf{X}(s)$ and $\partial_{s_2} \mathbf{X}(s)$ (where $\mathbf{x} = \mathbf{X}(s)$) are linearly independent vectors and they span the tangent space $T_x(\Gamma)$. Since $\partial_{s_1} \mathbf{X}$, $\partial_{s_2} \mathbf{X}$ are in \mathbb{R}^3 , there exists another vector orthogonal to both of them.

Hence, we denote the unit normal vector to Γ (at \mathbf{x}) by ν and define it in terms of local reference domain coordinates $s = (s_1, s_2)^T$:

$$\nu(s) = \frac{(\partial_{s_1} \mathbf{X}(s) \times \partial_{s_2} \mathbf{X}(s))}{|\partial_{s_1} \mathbf{X}(s) \times \partial_{s_2} \mathbf{X}(s)|} = \frac{(\partial_{s_1} \mathbf{X}(s) \times \partial_{s_2} \mathbf{X}(s))}{\sqrt{\det g}}, \quad (3.22)$$

where the second equality follows by

$$\begin{aligned} |\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}|^2 &= |\partial_{s_1} \mathbf{X}|^2 |\partial_{s_2} \mathbf{X}|^2 \sin^2 \theta = |\partial_{s_1} \mathbf{X}|^2 |\partial_{s_2} \mathbf{X}|^2 - |\partial_{s_1} \mathbf{X}|^2 |\partial_{s_2} \mathbf{X}|^2 \cos^2 \theta \\ &= g_{11} g_{22} - |\partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \mathbf{X}|^2 = g_{11} g_{22} - g_{12} g_{21} = \det g, \end{aligned} \quad (3.23)$$

where θ is the angle between $\partial_{s_1} \mathbf{X}$ and $\partial_{s_2} \mathbf{X}$.

Remark 12. *Sometimes, we will consider ν as being defined on Γ . In this case, we have $\nu \circ \mathbf{X}^{-1} : \Gamma \rightarrow \mathbb{R}^3$. For convenience, i.e., abuse of notation, we may omit \mathbf{X}^{-1} and simply refer to ν as being defined on Γ .*

The next result says, essentially, that ν depends only on Γ .

Lemma 3.2. *Assume the hypothesis of Lemma 3.1. The formula (3.22) is invariant, up to a sign, with respect to the change of parameters $s = Y(\tilde{s})$, i.e.,*

$$\nu \circ Y(\tilde{s}) = \operatorname{sgn}(\det(\nabla_{\tilde{s}} Y(\tilde{s}))) \tilde{\nu}(\tilde{s}). \quad (3.24)$$

Proof. By (3.20) (Lemma 3.1), and properties of determinants, we have

$$\det g \circ Y(\tilde{s}) = \det \tilde{g}(\tilde{s})(\det M)^{-2}, \quad \text{where } M = \nabla_{\tilde{s}} Y(\tilde{s}),$$

so the denominator of (3.22) transforms as

$$\sqrt{\det g \circ Y(\tilde{s})} = \sqrt{\det \tilde{g}(\tilde{s})} |\det M|^{-1}. \quad (3.25)$$

As for the numerator in (3.22), let m^{ij} be the i, j components of M^{-1} . Recalling (3.21), note that

$$(\partial_{s_i} X) \circ Y(\tilde{s}) = [\nabla_{\tilde{s}} \tilde{X}(\tilde{s})] M^{-1} e_i \quad \text{for } i = 1, 2, \quad (3.26)$$

where $e_1 = (1, 0)^T$, $e_2 = (0, 1)^T$. Expanding (3.26) out, we get

$$(\partial_{s_i} X) \circ Y(\tilde{s}) = m^{1i} \partial_{\tilde{s}_1} \tilde{X} + m^{2i} \partial_{\tilde{s}_2} \tilde{X} \quad \text{for } i = 1, 2.$$

Thus, evaluating the numerator yields

$$\begin{aligned} (\partial_{s_1} X \times \partial_{s_2} X) \circ Y(\tilde{s}) &= (m^{11} \partial_{\tilde{s}_1} \tilde{X} + m^{21} \partial_{\tilde{s}_2} \tilde{X}) \times (m^{12} \partial_{\tilde{s}_1} \tilde{X} + m^{22} \partial_{\tilde{s}_2} \tilde{X}) \\ &= (m^{11} m^{22} - m^{12} m^{21})(\partial_{\tilde{s}_1} \tilde{X} \times \partial_{\tilde{s}_2} \tilde{X}) \\ &= \det(M^{-1})(\partial_{\tilde{s}_1} \tilde{X} \times \partial_{\tilde{s}_2} \tilde{X}) = (\det M)^{-1}(\partial_{\tilde{s}_1} \tilde{X} \times \partial_{\tilde{s}_2} \tilde{X}), \end{aligned} \quad (3.27)$$

where we have used (A.1) and (A.2). Therefore, we obtain

$$\begin{aligned} \nu \circ Y(\tilde{s}) &= \frac{(\partial_{s_1} X(s) \times \partial_{s_2} X(s)) \circ Y(\tilde{s})}{\sqrt{\det g \circ Y(\tilde{s})}} = \frac{(\det M)^{-1}(\partial_{\tilde{s}_1} \tilde{X} \times \partial_{\tilde{s}_2} \tilde{X})}{|\det M|^{-1} \sqrt{\det \tilde{g}(\tilde{s})}} \\ &= \operatorname{sgn}(\det(\nabla_{\tilde{s}} Y(\tilde{s}))) \frac{(\partial_{\tilde{s}_1} \tilde{X}(\tilde{s}) \times \partial_{\tilde{s}_2} \tilde{X}(\tilde{s}))}{\sqrt{\det \tilde{g}(\tilde{s})}} = \operatorname{sgn}(\det(\nabla_{\tilde{s}} Y(\tilde{s}))) \tilde{\nu}(\tilde{s}). \quad \square \end{aligned}$$

From Lemma 3.2, we see that the sense, or *orientation*, of the normal vector does depend on the parameterization, i.e., on our choice of coordinates in the reference domain. It is clear from (3.22) that if we swap the roles of s_1 and s_2 , then ν will point in the opposite direction; of course, in this case, $\det(\nabla_{\tilde{s}} Y(\tilde{s}))$ is negative. The orientation of ν , defined on a surface Γ , is linked to the *orientation* of Γ (see below).

Orientation

Imagine walking along the simplest surface of all: the plane \mathbb{R}^2 . As you move forward, there is a clear sense of *left* and *right*, i.e., if you move in closed circles, it is clear whether you are turning to the left or to the right. Therefore, it is always possible to “orient” yourself while walking on the \mathbb{R}^2 plane. Thus, we say, the plane \mathbb{R}^2 is *orientable*. In fact, any plane is orientable, with *orientation* determined by selecting one of the two normal

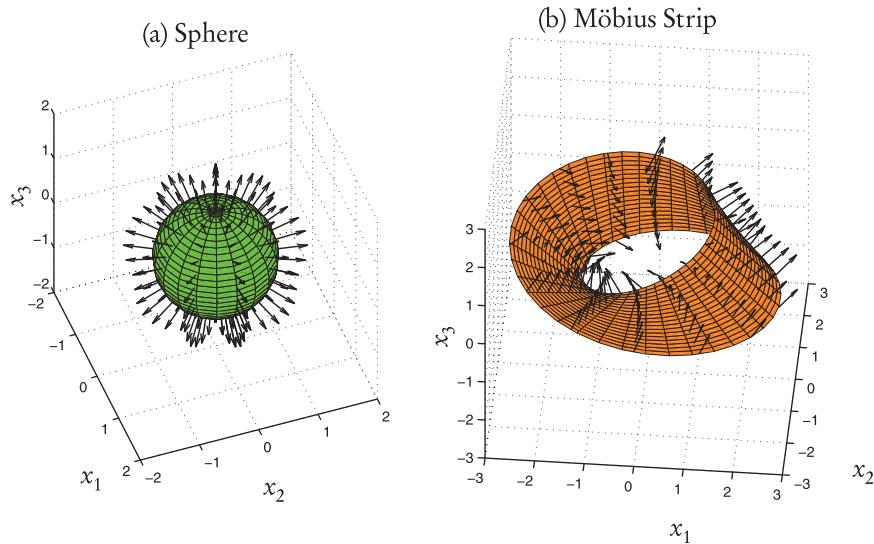


Figure 3.2. Illustration of orientability (arrows depict unit normal vector ν). (a) The unit sphere can be covered with a differentiable unit normal vector field (here a positive orientation is chosen). (b) The Möbius strip is a nonorientable surface. It is *not* possible to choose a normal vector field that is differentiable (or even continuous) over the entire surface.

vectors of the plane (e.g., ν or $-\nu$); choosing the normal vector decides which “side” of the plane you walk on.

Is this always possible for arbitrary surfaces? Let Γ be a regular surface. The choice of an orientation of the tangent plane $T_P(\Gamma)$ (section 2.4.3) induces an orientation of Γ in a neighborhood of P . If we can make this choice for each P in Γ so that in the intersection of any two neighborhoods the orientations coincide, then Γ is *orientable*. Otherwise, Γ is called *nonorientable*. The following definition makes this more precise.

Definition 4. Let Γ be a regular surface that is completely defined by an atlas of charts $\{(U_i, \mathbf{X}_i)\}$, i.e., $\cup_i \mathbf{X}_i(U_i) = \Gamma$. We say Γ is *orientable* if for each point P in Γ such that $P \in \mathbf{X}_i(U_i) \cap \mathbf{X}_j(U_j)$ and $i \neq j$, then the change of coordinates,

$$\mathbf{Y} := \mathbf{X}_i^{-1} \circ \mathbf{X}_j \quad (\text{recall Proposition 1 in Chapter 2}),$$

has positive Jacobian at P . The choice of such an atlas is called an *orientation* of Γ and we say that Γ is *oriented*. If such a choice is not possible, then Γ is called *nonorientable*.

The following result gives an equivalent condition to check for orientability.

Proposition 4. A regular surface $\Gamma \subset \mathbb{R}^3$ is orientable if and only if there exists a differentiable unit normal vector field $\nu : \Gamma \rightarrow \mathbb{R}^3$ on Γ .

Proof. See [24, sect. 2-6, Prop. 1]. □

Figure 3.2 shows an example of an orientable and a nonorientable surface. When dealing with closed, orientable surfaces, we usually choose ν to be the **outward** pointing normal vector. In this case, we say that the closed surface is *positively oriented*.

3.1.4 • Computing Angles in a Surface

Fix a point s_0 in U and evaluate $\partial_{s_1} \mathbf{X}$ and $\partial_{s_2} \mathbf{X}$ at s_0 . Suppose \mathbf{a} and \mathbf{b} are vectors in \mathbb{R}^3 that can be written as

$$\mathbf{a} = \hat{a}_1 \partial_{s_1} \mathbf{X} + \hat{a}_2 \partial_{s_2} \mathbf{X}, \quad \mathbf{b} = \hat{b}_1 \partial_{s_1} \mathbf{X} + \hat{b}_2 \partial_{s_2} \mathbf{X}, \quad (3.28)$$

where $\hat{a}_1, \hat{a}_2, \hat{b}_1, \hat{b}_2$ are numbers. This means that \mathbf{a} and \mathbf{b} lie in the tangent space of Γ at $\mathbf{x}_0 = \mathbf{X}(s_0)$. In fact, one can view $\hat{\mathbf{a}} = (\hat{a}_1, \hat{a}_2)^T$ and $\hat{\mathbf{b}} = (\hat{b}_1, \hat{b}_2)^T$ as vectors in $U \subset \mathbb{R}^2$ that are mapped to \mathbf{a} and \mathbf{b} by (3.28).

The dot product of \mathbf{a} and \mathbf{b} , in terms of the reference domain vectors, is

$$\mathbf{a} \cdot \mathbf{b} = (\hat{a}_1 \partial_{s_1} \mathbf{X} + \hat{a}_2 \partial_{s_2} \mathbf{X}) \cdot (\hat{b}_1 \partial_{s_1} \mathbf{X} + \hat{b}_2 \partial_{s_2} \mathbf{X}) = \sum_{i,j=1}^2 \hat{a}_i \hat{b}_j g_{ij}. \quad (3.29)$$

Clearly, the lengths of \mathbf{a} and \mathbf{b} are

$$|\mathbf{a}| = (\mathbf{a} \cdot \mathbf{a})^{1/2} = \left(\sum_{p,q=1}^2 \hat{a}_p \hat{a}_q g_{pq} \right)^{1/2}, \quad |\mathbf{b}| = \left(\sum_{p,q=1}^2 \hat{b}_p \hat{b}_q g_{pq} \right)^{1/2}, \quad (3.30)$$

so the angle θ between \mathbf{a} and \mathbf{b} is

$$\cos \theta = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}| |\mathbf{b}|} = \frac{\sum_{i,j=1}^2 \hat{a}_i \hat{b}_j g_{ij}}{\left(\sum_{p,q=1}^2 \hat{a}_p \hat{a}_q g_{pq} \right)^{1/2} \left(\sum_{p,q=1}^2 \hat{b}_p \hat{b}_q g_{pq} \right)^{1/2}}. \quad (3.31)$$

Now consider two curves \mathcal{C} and $\tilde{\mathcal{C}}$ in U that are parameterized by $\mathbf{s}(t)$ and $\tilde{\mathbf{s}}(t)$. The corresponding curves Σ and $\tilde{\Sigma}$ in Γ are parameterized by $\mathbf{X} \circ \mathbf{s}$ and $\mathbf{X} \circ \tilde{\mathbf{s}}$. Assume Σ and $\tilde{\Sigma}$ intersect at a point P in Γ . The angle of intersection θ between Σ and $\tilde{\Sigma}$ at P is the angle between their respective velocity vectors, i.e., the angle between

$$\begin{aligned} \mathbf{v} &= \frac{d}{dt} \mathbf{X} \circ \mathbf{s}(t) = (\mathbf{s}'(t) \cdot \nabla) \mathbf{X} = s'_1(t) \partial_{s_1} \mathbf{X} + s'_2(t) \partial_{s_2} \mathbf{X}, \\ \tilde{\mathbf{v}} &= \frac{d}{dt} \mathbf{X} \circ \tilde{\mathbf{s}}(t) = (\tilde{\mathbf{s}}'(t) \cdot \nabla) \mathbf{X} = \tilde{s}'_1(t) \partial_{s_1} \mathbf{X} + \tilde{s}'_2(t) \partial_{s_2} \mathbf{X}, \end{aligned}$$

at some appropriate value of t . So, from (3.31), we get

$$\cos \theta = \frac{\mathbf{v} \cdot \tilde{\mathbf{v}}}{|\mathbf{v}| |\tilde{\mathbf{v}}|} = \frac{\sum_{i,j=1}^2 s'_i \tilde{s}'_j g_{ij}}{\left(\sum_{p,q=1}^2 s'_p s'_q g_{pq} \right)^{1/2} \left(\sum_{p,q=1}^2 \tilde{s}'_p \tilde{s}'_q g_{pq} \right)^{1/2}}. \quad (3.32)$$

Thus, Σ and $\tilde{\Sigma}$ intersect orthogonally at P if and only if $\sum_{i,j=1}^2 s'_i \tilde{s}'_j g_{ij} = 0$.

Clearly, the vectors $\partial_{s_1} \mathbf{X}$ and $\partial_{s_2} \mathbf{X}$ are tangent to the coordinate curves $s_2 = \text{constant}$ and $s_1 = \text{constant}$, respectively. If ω is the angle of intersection of these curves, then

$$\cos \omega = \frac{\partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \mathbf{X}}{|\partial_{s_1} \mathbf{X}| |\partial_{s_2} \mathbf{X}|} = \frac{g_{12}}{\sqrt{g_{11} g_{22}}}. \quad (3.33)$$

A coordinate system on Γ is said to be orthogonal if at any point of Γ , the coordinate curves intersect orthogonally. Therefore, (3.33) implies the following.

Lemma 3.3. *A coordinate system on a regular surface Γ is orthogonal if and only if at every point of Γ , the first fundamental form g satisfies $g_{12} = g_{21} = 0$.*

For an example, see Figure 2.13 and the parameterization of the cylinder in (2.30). The components of the first fundamental form for (2.30) are

$$g_{11} = R^2, \quad g_{12} = g_{21} = 0, \quad g_{22} = 1.$$

where R is the fixed radius of the cylinder. It is clear from Figure 2.13 that the coordinate curves are orthogonal.

Remark 13. *Orthogonal coordinates can be convenient because they simplify various differential geometry formulas.*

3.1.5 ■ Computing Surface Area

Computing the area of a regular surface Γ is accomplished by the first fundamental form, which the next theorem describes.

Theorem 3.4. *Let (U, \mathbf{X}) be a local chart (coordinate system) of the regular surface Γ . Suppose $R = \mathbf{X}(Q)$ is a bounded set contained in Γ where $Q \subset U$. Then the area of R is given by*

$$\text{area}(R) := \int_R 1 = \iint_Q |\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}| ds_1 ds_2 = \iint_Q \sqrt{\det g(s_1, s_2)} ds_1 ds_2, \quad (3.34)$$

where $g = g(s_1, s_2)$ is the first fundamental form defined by (3.2).

Proof. From standard multivariable calculus, it is known that (see (B.11))

$$\text{area}(R) = \iint_Q |\partial_{s_1} \mathbf{X}(s_1, s_2) \times \partial_{s_2} \mathbf{X}(s_1, s_2)| ds_1 ds_2.$$

See [24, sect. 2-8] for a proof of this. The assertion then follows by (3.23). \square

Corollary 1. *The formula (3.34) does not depend on the parameterization \mathbf{X} .*

Proof. Let $(\tilde{U}, \tilde{\mathbf{X}})$ be another local chart such that $R \subset \tilde{\mathbf{X}}(\tilde{U})$ and set $\tilde{Q} = \tilde{\mathbf{X}}^{-1}(R)$. Define the change of parameter function $\mathbf{Y} := \mathbf{X}^{-1} \circ \tilde{\mathbf{X}}$, where \mathbf{Y} maps between subsets of \mathbb{R}^2 , i.e., $\mathbf{Y}(\tilde{Q}) = Q$.

From (3.25), we know that

$$\sqrt{\det g \circ \mathbf{Y}(\tilde{s}_1, \tilde{s}_2)} |\det \mathbf{M}| = \sqrt{\det \tilde{g}(\tilde{s}_1, \tilde{s}_2)}, \quad \text{where } \mathbf{M} = [\nabla_{\tilde{s}} \mathbf{Y}(\tilde{s})].$$

Applying the change of variable formula (B.10), we get

$$\begin{aligned} \iint_Q \sqrt{\det g(s_1, s_2)} ds_1 ds_2 &= \iint_{\tilde{Q}} \sqrt{\det g \circ \mathbf{Y}(\tilde{s}_1, \tilde{s}_2)} |\det \mathbf{M}| d\tilde{s}_1 d\tilde{s}_2 \\ &= \iint_{\tilde{Q}} \sqrt{\det \tilde{g}(\tilde{s}_1, \tilde{s}_2)} d\tilde{s}_1 d\tilde{s}_2. \end{aligned}$$

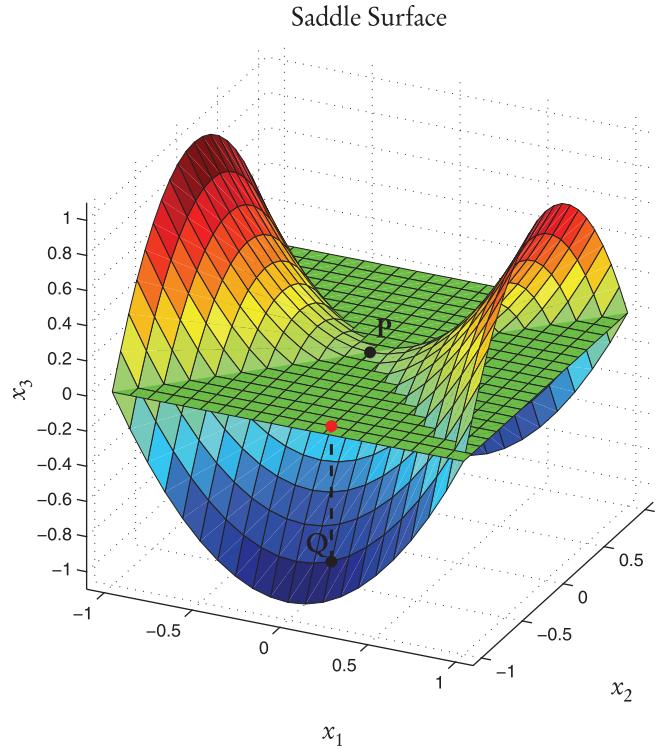


Figure 3.3. A saddle surface Γ and its tangent plane $T_p(\Gamma)$ (solid green). Here, the point P is the origin. As one moves away from P (toward Q) the surface Γ “pulls away” from the tangent plane (at P) in a way that depends on its curvature. The red dot is the closest point in $T_p(\Gamma)$ to Q , with the minimum distance given by the length of the dashed line.

Therefore, (3.34) is unaffected by the choice of parameterization (except for a trivial renaming of the variables). \square

3.2 ■ Second Fundamental Form

Most surfaces of interest are “curvy”, i.e., they are not planes in \mathbb{R}^3 . In fact, viewing a surface Γ and its tangent plane $T_p(\Gamma)$ at a point P , one can see that the surface “pulls away” from $T_p(\Gamma)$ as you move away from P (see Figure 3.3). Quantifying this curving or “pulling away” requires the so-called second fundamental form, which is crucial for characterizing the shape of surfaces.

3.2.1 ■ Deviating from the Tangent Plane

Theorem 3.5. Let $\Gamma \subset \mathbb{R}^3$ be a regular surface parameterized by $\mathbf{X}: U \rightarrow \Gamma$. Suppose P is a fixed point in Γ with reference coordinates $(s_1, s_2)^T \in U$; similarly, let Q in Γ have reference coordinates $(s_1 + \ell_1, s_2 + \ell_2)^T \in U$, where $|\ell_1|, |\ell_2| \ll 1$. Then the minimum signed distance between Q and the tangent plane $T_p(\Gamma)$ (see Figure 3.3) is

$$\text{dist}(Q, T_p(\Gamma)) = \frac{1}{2} \sum_{i,j=1}^2 \ell_i \ell_j \partial_{s_i} \partial_{s_j} \mathbf{X} \cdot \boldsymbol{\nu} + O((|\ell_1| + |\ell_2|)^3), \quad (3.35)$$

where $\boldsymbol{\nu}$ is the normal vector of $T_p(\Gamma)$ which is given by (3.22) evaluated at P .

Proof. Start with a Taylor expansion:

$$\mathbf{X}(s_1 + \ell_1, s_2 + \ell_2) = \mathbf{X}(s_1, s_2) + \sum_{i=1}^2 \ell_i \partial_{s_i} \mathbf{X} + \frac{1}{2!} \sum_{i,j=1}^2 \ell_i \ell_j \partial_{s_i} \partial_{s_j} \mathbf{X} + \text{h.o.t.},$$

where $\partial_{s_i} \mathbf{X}$ and $\partial_{s_i} \partial_{s_j} \mathbf{X}$ are evaluated at $(s_1, s_2)^T$, and h.o.t. are higher order terms (third order and above). The minimum distance between Q and the tangent plane at P , denoted $\text{dist}(Q, T_P(\Gamma))$, is simply the projection of $\mathbf{X}(s_1 + \ell_1, s_2 + \ell_2) - \mathbf{X}(s_1, s_2)$ onto the unit normal vector of $T_P(\Gamma)$. Since ν is orthogonal to $\partial_{s_1} \mathbf{X}$ and $\partial_{s_2} \mathbf{X}$, we get the assertion

$$\begin{aligned} \text{dist}(Q, T_P(\Gamma)) &= [\mathbf{X}(s_1 + \ell_1, s_2 + \ell_2) - \mathbf{X}(s_1, s_2)] \cdot \nu \\ &= \frac{1}{2} \sum_{i,j=1}^2 \ell_i \ell_j \partial_{s_i} \partial_{s_j} \mathbf{X} \cdot \nu + O((|\ell_1| + |\ell_2|)^3). \quad \square \end{aligned}$$

Remark 14. The sign of $\text{dist}(Q, T_P(\Gamma))$ depends on the orientation of ν that is induced by the chosen parameterization \mathbf{X} .

From (3.35), we can see that $\partial_{s_i} \partial_{s_j} \mathbf{X} \cdot \nu$ quantifies (locally) how the surface “curves away” from the tangent plane. Therefore, we introduce the notation

$$h_{ij} = \nu \cdot \partial_{s_i} \partial_{s_j} \mathbf{X} \quad \text{for } 1 \leq i, j \leq 2,$$

where clearly $h_{12} = h_{21}$. Since $\nu \cdot \partial_{s_j} \mathbf{X} = 0$ for $j = 1, 2$, we can differentiate to see that

$$\partial_{s_i} \nu \cdot \partial_{s_j} \mathbf{X} + \nu \cdot \partial_{s_i} \partial_{s_j} \mathbf{X} = 0 \quad \text{for } 1 \leq i, j \leq 2.$$

Hence, we can write

$$h_{ij} = -\partial_{s_i} \nu \cdot \partial_{s_j} \mathbf{X} = \nu \cdot \partial_{s_i} \partial_{s_j} \mathbf{X} \quad \text{for } 1 \leq i, j \leq 2. \quad (3.36)$$

These are the **coefficients** of the *second fundamental form*.

3.2.2 • Examples

Let us compute the coefficients h_{ij} for some example surfaces.

Sphere

Recall the sphere parameterization in (3.6). Computing the normal vector and its derivatives gives

$$\begin{aligned} \nu(s_1, s_2) &= (s_1, s_2, \sqrt{1 - (s_1^2 + s_2^2)})^T, \\ \partial_{s_1} \nu(s_1, s_2) &= (1, 0, -s_1(1 - (s_1^2 + s_2^2))^{-1/2})^T, \\ \partial_{s_2} \nu(s_1, s_2) &= (0, 1, -s_2(1 - (s_1^2 + s_2^2))^{-1/2})^T. \end{aligned}$$

From (3.36) and (3.6), we find that the coefficients are

$$\begin{aligned} h_{11} &= -[1 + s_1^2(1 - (s_1^2 + s_2^2))^{-1}] = -(1 - s_2^2)(1 - (s_1^2 + s_2^2))^{-1}, \\ h_{12} &= h_{21} = s_1 s_2 (1 - (s_1^2 + s_2^2))^{-1}, \\ h_{22} &= -[1 + s_2^2(1 - (s_1^2 + s_2^2))^{-1}] = -(1 - s_1^2)(1 - (s_1^2 + s_2^2))^{-1}. \end{aligned} \quad (3.37)$$

We can see that h_{ij} blows up as $|(\mathbf{s}_1, \mathbf{s}_2)| \rightarrow 1$. This also happens in (3.7) (recall Figure 2.11). The blowup is caused by the way the surface is parameterized. It is not due to the geometric shape of the surface.

Wavy Surface

Recall the parameterization (3.18). Computing the normal vector and the second order derivatives of \mathbf{X} gives

$$\begin{aligned}\boldsymbol{\nu}(\mathbf{s}_1, \mathbf{s}_2) &= \frac{(\pi \sin(\pi s_1) \cos(\pi s_2), \pi \cos(\pi s_1) \sin(\pi s_2), 1)^T}{\sqrt{1 + \pi^2(\sin^2(\pi s_1) \cos^2(\pi s_2) + \cos^2(\pi s_1) \sin^2(\pi s_2))}}, \\ \partial_{s_1} \partial_{s_1} \mathbf{X}(\mathbf{s}_1, \mathbf{s}_2) &= (0, 0, -\pi^2 \cos(\pi s_1) \cos(\pi s_2))^T, \\ \partial_{s_1} \partial_{s_2} \mathbf{X}(\mathbf{s}_1, \mathbf{s}_2) &= (0, 0, +\pi^2 \sin(\pi s_1) \sin(\pi s_2))^T, \\ \partial_{s_2} \partial_{s_2} \mathbf{X}(\mathbf{s}_1, \mathbf{s}_2) &= (0, 0, -\pi^2 \cos(\pi s_1) \cos(\pi s_2))^T.\end{aligned}$$

From (3.36), we find that the coefficients are

$$\begin{aligned}b_{11} &= -\pi^2 \cos(\pi s_1) \cos(\pi s_2) W^{-1}, \\ b_{12} = b_{21} &= +\pi^2 \sin(\pi s_1) \sin(\pi s_2) W^{-1}, \\ b_{22} &= -\pi^2 \cos(\pi s_1) \cos(\pi s_2) W^{-1}, \\ W &= \sqrt{1 + \pi^2(\sin^2(\pi s_1) \cos^2(\pi s_2) + \cos^2(\pi s_1) \sin^2(\pi s_2))}.\end{aligned}\tag{3.38}$$

A visualization of the components b_{ij} are shown in Figure 3.4. It is clear that the components represent the “curviness” of the surface. But b_{ij} also depends on the parameterization. Eventually, we will derive another object (shape operator) that correctly represents the “curvature” of the surface and does **not** depend on the parameterization (see section 3.2.5).

3.2.3 • Tensor Notation

The coefficients b_{ij} can be grouped into a 2×2 matrix denoted \mathbf{h} ,

$$\mathbf{h} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix},\tag{3.39}$$

which is a **symmetric** matrix because $b_{12} = b_{21}$. \mathbf{h} is a tensor that encodes the “*curvature*” of the surface—in fact, \mathbf{h} is the matrix realization of the tensor. Just as with \mathbf{g} , if $\mathbf{X}: \mathbb{R}^n \rightarrow \mathbb{R}^m$, then \mathbf{h} is an $n \times n$ matrix. However, unlike \mathbf{g} , \mathbf{h} may **not** be positive definite. For example, if Γ is a plane in \mathbb{R}^3 , then \mathbf{h} is the zero matrix (the reader can check this). Note that from (3.36) we can write \mathbf{h} as a matrix-matrix product

$$\mathbf{h} = -(\nabla_s \boldsymbol{\nu})^T (\nabla_s \mathbf{X}).\tag{3.40}$$

Similar to the first fundamental form, the second fundamental form is a quadratic form that is defined on the tangent space. To see this, consider the unit normal vector $\boldsymbol{\nu}$

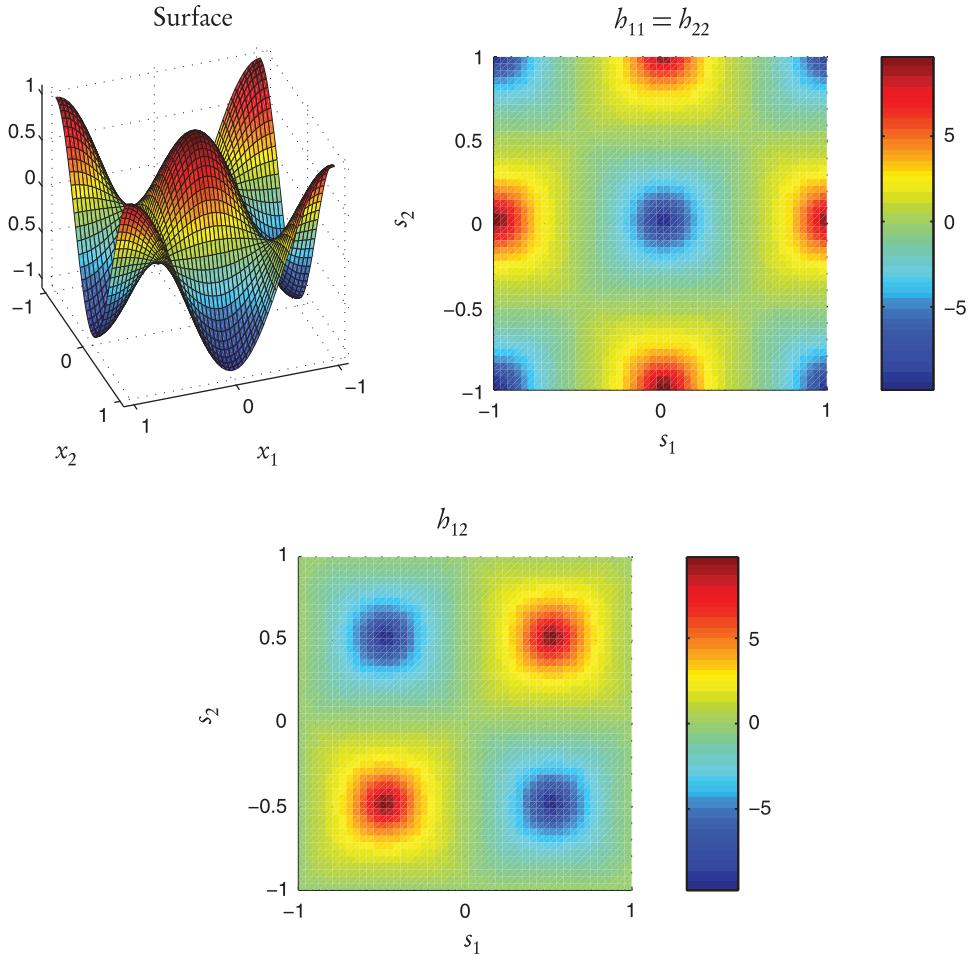


Figure 3.4. Illustration of the components of the second fundamental form for the surface given by (3.18) (see also Figure 3.1). Note that \$h_{11}\$ (\$h_{22}\$) is most negative at a peak of the surface; similarly, it is most positive at a “valley” of the surface. \$h_{12}\$ is close to zero near the peaks and valleys.

of \$\Gamma\$ evaluated along a curve \$\Sigma \subset \Gamma\$ parameterized by \$\alpha = \mathbf{X} \circ \mathbf{s}\$, i.e., \$\nu(t) = \nu(s_1(t), s_2(t))\$, where \$\mathbf{s}(t) = (s_1(t), s_2(t))^T\$. Differentiating, we see that

$$\nu'(t) = (\mathbf{s}'(t) \cdot \nabla) \nu(\mathbf{s}(t)), \quad \text{where } \nabla = (\partial_{s_1}, \partial_{s_2}).$$

Since \$\nu(t) \cdot \nu(t) = 1\$, we know \$\nu(t) \cdot \nu'(t) = 0\$, so \$\nu'(t)\$ lies in the tangent space \$T_{\alpha(t)}(\Gamma)\$ for all \$t\$. So the quadratic form given by the dot product of \$\nu'(t)\$ with \$\alpha'(t) \in T_{\alpha(t)}(\Gamma)\$ measures the rate of change of \$\nu\$ along the tangent direction \$\alpha'(t)\$. A preliminary calculation gives

$$\begin{aligned} -\nu'(t) \cdot \alpha'(t) &= -(s'_1(t) \partial_{s_1} \nu + s'_2(t) \partial_{s_2} \nu) \cdot (s'_1(t) \partial_{s_1} \mathbf{X} + s'_2(t) \partial_{s_2} \mathbf{X}) \\ &= -((s'_1)^2 \partial_{s_1} \nu \cdot \partial_{s_1} \mathbf{X} + s'_1 s'_2 \partial_{s_1} \nu \cdot \partial_{s_2} \mathbf{X} + s'_2 s'_1 \partial_{s_2} \nu \cdot \partial_{s_1} \mathbf{X} + (s'_2)^2 \partial_{s_2} \nu \cdot \partial_{s_2} \mathbf{X}) \\ &= \sum_{i,j=1}^2 h_{ij} s'_i s'_j, \end{aligned}$$

where we used (3.36); note that a sign was introduced because of convention. One should compare this formula with (3.4). The following definition formalizes this concept.

Definition 5. *The second fundamental form of the surface $\Gamma \subset \mathbb{R}^3$, at the point P in Γ , is a quadratic form $\Pi_P(\cdot)$ defined on $T_P(\Gamma)$ by*

$$\Pi_P(\mathbf{v}) = \sum_{i,j=1}^2 h_{ij} a_i a_j, \quad \text{where } a_1, a_2 \in \mathbb{R} \quad \text{and} \quad \mathbf{v} = a_1 \partial_{s_1} \mathbf{X} + a_2 \partial_{s_2} \mathbf{X}, \quad (3.41)$$

where $\mathbf{v} \in T_P(\Gamma)$ is expressed in the tangent basis $\{\partial_{s_1} \mathbf{X}, \partial_{s_2} \mathbf{X}\}$ evaluated at P .

The quadratic form (3.41) is key to understanding the “shape,” or curvature, of Γ . We finish this section with a result on how h transforms under a change of parameters.

Lemma 3.6. *Assume the hypothesis of Lemma 3.1. Then,*

$$h \circ \mathbf{Y}(\tilde{s}) = \operatorname{sgn}(\det \mathbf{M})(\mathbf{M}^{-1})^T \tilde{h}(\tilde{s}) \mathbf{M}^{-1}, \quad \text{where } \mathbf{M} = \nabla_{\tilde{s}} \mathbf{Y}(\tilde{s}), \quad (3.42)$$

and h and \tilde{h} are the second fundamental form matrices associated with (U, \mathbf{X}) and $((\tilde{U}, \tilde{\mathbf{X}}))$, respectively.

Proof. Note that $\tilde{U} = \tilde{\mathbf{X}}^{-1}(R)$. Using (3.24), we obtain a relation similar to (3.21),

$$(\nabla_s \boldsymbol{\nu}) \circ \mathbf{Y}(\tilde{s}) = \operatorname{sgn}(\det \mathbf{M}) \underbrace{\nabla_{\tilde{s}} \tilde{\boldsymbol{\nu}}(\tilde{s})}_{3 \times 2 \text{ matrix}} \mathbf{M}^{-1}, \quad (3.43)$$

where $\boldsymbol{\nu}$ ($\tilde{\boldsymbol{\nu}}$) is the unit normal vector with respect to \mathbf{X} ($\tilde{\mathbf{X}}$). By (3.36), we know that $h = -(\nabla_s \boldsymbol{\nu})^T \nabla_s \mathbf{X}$. Therefore,

$$\begin{aligned} h \circ \mathbf{Y}(\tilde{s}) &= -(\operatorname{sgn}(\det \mathbf{M}) \nabla_{\tilde{s}} \tilde{\boldsymbol{\nu}}(\tilde{s}) \mathbf{M}^{-1})^T (\nabla_{\tilde{s}} \tilde{\mathbf{X}}(\tilde{s}) \mathbf{M}^{-1}) \\ &= \operatorname{sgn}(\det \mathbf{M}) (\mathbf{M}^{-1})^T (-(\nabla_{\tilde{s}} \tilde{\boldsymbol{\nu}})^T \nabla_{\tilde{s}} \tilde{\mathbf{X}}) \mathbf{M}^{-1} \\ &= \operatorname{sgn}(\det \mathbf{M}) (\mathbf{M}^{-1})^T \tilde{h} \mathbf{M}^{-1}. \quad \square \end{aligned}$$

Remark 15. *The presence of $\operatorname{sgn}(\det \mathbf{M})$ in (3.42) implies that h changes sign if we swap the local coordinates, i.e., set $s_1 = \tilde{s}_2$ and $s_2 = \tilde{s}_1$. This is clear from the definition of h_{ij} (3.36) because the normal vector $\boldsymbol{\nu}$ appears there, and we know that the orientation of $\boldsymbol{\nu}$ “flips” if we swap s_1 and s_2 (recall Lemma 3.2).*

3.2.4 • Investigation of h

Curvature of 1-D Curves

To better understand the meaning of h , we first review the definition of curvature for curves. Let $\alpha : I \rightarrow \mathbb{R}^3$ be a parameterization of a 3-D curve $\Sigma = \alpha(I)$. In addition, we allow for $\alpha(t)$ to not be an arc-length parameterization with respect to t , i.e., $|\alpha'(t)| \neq 1$.

The derivative with respect to arc-length $\frac{d}{d\alpha}$ is related to $\frac{d}{dt}$ by the formula

$$\frac{d}{d\alpha} = \frac{1}{|\alpha'(t)|} \frac{d}{dt}, \quad \text{where } \frac{d}{dt} = '.$$

The (oriented) unit tangent vector τ of Σ is defined by

$$\tau(t) = \frac{\alpha'(t)}{|\alpha'(t)|} \quad \text{or} \quad \tau = \frac{d\alpha}{d\alpha}. \quad (3.44)$$

The vector curvature of Σ is defined by

$$k\mathbf{n} = -\frac{d^2\alpha}{d\alpha^2}, \quad (3.45)$$

where k is the signed (scalar) curvature of Σ and $\mathbf{n} \in \mathbb{R}^3$ is the unit normal vector of Σ . We use the sign convention that k is positive and \mathbf{n} points outside of Σ when Σ is a convex, closed, positively oriented curve in \mathbb{R}^2 . In terms of the t variable, (3.45) becomes

$$k(t)\mathbf{n}(t) = -\frac{1}{|\alpha'(t)|} \left(\frac{\alpha'(t)}{|\alpha'(t)|} \right)' = -\frac{\alpha''(t)}{\alpha'(t) \cdot \alpha'(t)} - \left(\frac{1}{|\alpha'(t)|} \right)' \tau(t). \quad (3.46)$$

Normal Curvature

Now let $\Gamma \subset \mathbb{R}^3$ be a regular surface parameterized by $\mathbf{X} : U \rightarrow \Gamma$, suppose $\Sigma \subset \Gamma$, and $\alpha = \mathbf{X} \circ s$, where $s : I \rightarrow U$ is the parameterization of a “reference” curve in U . From (3.4) and (3.16), we know $|\alpha'(t)| = \sqrt{s' \cdot gs'}$. So the formula for the tangent vector becomes

$$\tau(t) = \frac{(s' \cdot \nabla)\mathbf{X}}{\sqrt{s' \cdot gs'}}, \quad (3.47)$$

where $\nabla = (\partial_{s_1}, \partial_{s_2})$. The curvature vector is then given by

$$k(t)\mathbf{n}(t) = -\frac{1}{\sqrt{s' \cdot gs'}} (\tau(t))' = -\frac{(s' \cdot \nabla)^2 \mathbf{X}}{s' \cdot gs'} - \left(\frac{1}{\sqrt{s' \cdot gs'}} \right)' \tau(t). \quad (3.48)$$

To link this discussion back to the second fundamental form, we introduce the *normal curvature* of curves embedded in surfaces. It is defined by

$$\begin{aligned} k_n(t) &= k(t)\mathbf{n}(t) \cdot \nu(s(t)) = -\frac{1}{s' \cdot gs'} \nu(s(t)) \cdot \sum_{i,j=1}^2 s'_i s'_j \partial_{s_i} \partial_{s_j} \mathbf{X} \\ &= -\frac{1}{s' \cdot gs'} \sum_{i,j=1}^2 s'_i s'_j h_{ij} = -\frac{s' \cdot hs'}{s' \cdot gs'} = -\frac{\Pi_{\alpha(t)}(\alpha'(t))}{\mathcal{I}_{\alpha(t)}(\alpha'(t))}, \end{aligned} \quad (3.49)$$

where ν is the unit normal vector of Γ and we have used the fact that $\nu \cdot \tau = 0$. Clearly, k_n is the component of the curve’s curvature vector $k\mathbf{n}$ along the unit normal of the surface

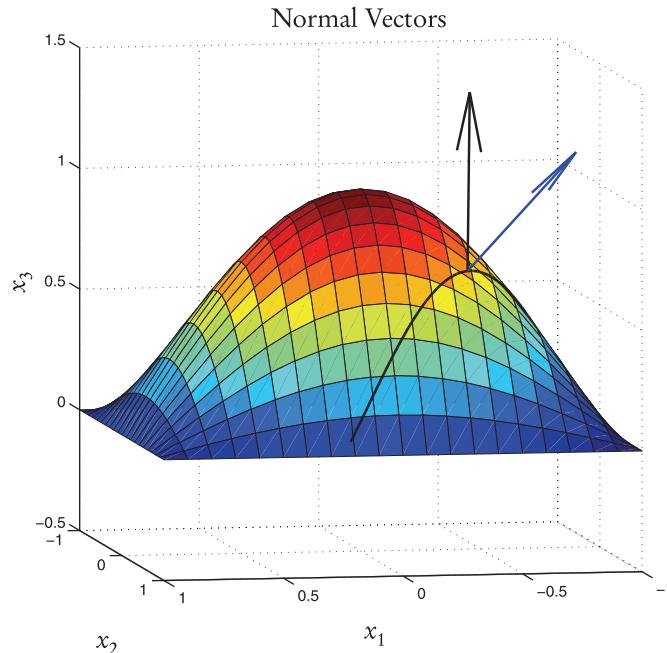


Figure 3.5. Comparison of the normal vector of a 1-D curve and the normal vector of a surface. The black arrow is the unit normal \mathbf{n} of the black curve embedded in the surface. The blue arrow is the unit normal ν of the surface.

(see Figure 3.5). It measures the part of the curvature of Σ that is due to the “bending” of the surface.

We immediately observe the following in (3.49).

- If $P = \alpha(0)$, then the normal curvature at P depends only on the tangent vector $\alpha'(0)$.
- Multiplying $\alpha'(0)$ by a nonzero constant, or equivalently scaling $s'(0)$, does not change the normal curvature.

Thus, the normal curvature depends only on the direction in which one moves in the tangent plane. We have shown the following result.

Proposition 5 (Meusnier). All curves lying on a surface Γ , and having at a fixed point P in Γ the same tangent direction, have the same normal curvature at P .

Therefore, it makes sense to talk about the normal curvature along a given direction at a point.

Principal Curvatures and Principal Directions

In view of (3.49), we can define the normal curvature at a point \mathbf{x} in Γ , in the direction \mathbf{v} in the tangent space $T_{\mathbf{x}}(\Gamma)$, by

$$k_n(\mathbf{x}, \mathbf{v}) = -\frac{\Pi_{\mathbf{x}}(\mathbf{v})}{I_{\mathbf{x}}(\mathbf{v})}, \quad (3.50)$$

or one can define it in terms of local coordinates s ,

$$k_n(s, q) = -\frac{q \cdot h(s)q}{q \cdot g(s)q}, \quad (3.51)$$

where $q \in \mathbb{R}^2$. The correspondence between (3.50) and (3.51) is

$$x = X(s), \quad v = (q \cdot \nabla)X(s) = q_1 \partial_{s_1} X(s_1, s_2) + q_2 \partial_{s_2} X(s_1, s_2), \quad (3.52)$$

where v is the vector in the tangent space corresponding to q .

By choosing different directions in (3.50) (or (3.51)), you obtain different pieces of information on how the surface curves away from the tangent plane. Alternatively, one can find the direction in which the curvature is a maximum (or minimum). This can be done by simply maximizing or minimizing (3.51) with respect to q (while keeping s fixed). Because $k_n(s, q) = k_n(s, Cq)$ for any nonzero constant C , we can just maximize (minimize) over all q such that $q^T g q = 1$. Since $k_n(s, q)$ is continuous (recall that the denominator is positive definite), from basic analysis we know that there exists a q_1 and q_2 that achieves the maximum and minimum, respectively, of $k_n(s, q)$.

Let κ_1 and κ_2 be the maximum and minimum normal curvatures, i.e.,

$$\kappa_1 = \max_{\substack{q \in \mathbb{R}^2, \\ q^T g q = 1}} -q^T h q, \quad \kappa_2 = \min_{\substack{q \in \mathbb{R}^2, \\ q^T g q = 1}} -q^T h q. \quad (3.53)$$

Because g and h are symmetric and g is positive definite, it can be shown that κ_1 and κ_2 are the maximum and minimum eigenvalues of the *generalized eigenvalue problem*: $-hq = \lambda gq$ [12, 68, 70, 81]. In other words,

$$-h q_1 = \kappa_1 g q_1, \quad -h q_2 = \kappa_2 g q_2, \quad \text{and} \quad q_1^T g q_1 = q_2^T g q_2 = 1, \quad (3.54)$$

where q_1 (q_2) is an eigenvector with eigenvalue κ_1 (κ_2). Moreover, it can be shown [70, 81] that q_1 and q_2 are g -orthogonal, i.e., $q_1^T g q_2 = 0$. Using (3.52), the corresponding vectors in the tangent space are d_1 and d_2 which are orthonormal in the usual \mathbb{R}^3 sense, i.e., $d_1^T d_2 = 0$ and $|d_1| = |d_2| = 1$ (the reader can verify this). Thus, $\{d_1, d_2\}$ form an orthonormal basis of $T_x(\Gamma)$.

The extreme values of the normal curvature have a special significance, which leads to the following definition.

Definition 6. Assume (3.53) is computed with h and g evaluated at a point $x = X(s)$ in Γ . The maximum normal curvature κ_1 and minimum normal curvature κ_2 are called the *principal curvatures* at x . The associated directions in the tangent space $T_x(\Gamma)$ given by d_1 and d_2 are called the *principal directions* at x .

Notice that we can compute the normal curvature along any direction in the tangent space in terms of κ_1 and κ_2 . Since d_1 and d_2 are orthogonal unit length vectors, we can represent an arbitrary direction in $T_x(\Gamma)$ by the unit vector d defined by

$$d = \cos \theta d_1 + \sin \theta d_2,$$

where θ is the angle from the \mathbf{d}_1 -axis measured counterclockwise in the positive direction. Recall the corresponding vectors $\mathbf{q}_1, \mathbf{q}_2$ to $\mathbf{d}_1, \mathbf{d}_2$. By (3.52) and linearity, the corresponding vector to \mathbf{d} is

$$\mathbf{q} = \cos \theta \mathbf{q}_1 + \sin \theta \mathbf{q}_2.$$

Hence, the normal curvature at \mathbf{x} , in the direction \mathbf{d} , using (3.51), is given by

$$\begin{aligned} k_n(\mathbf{x}, \mathbf{d}) &= -\mathbf{q}^T h \mathbf{q} = -(\cos \theta \mathbf{q}_1 + \sin \theta \mathbf{q}_2)^T h (\cos \theta \mathbf{q}_1 + \sin \theta \mathbf{q}_2) \\ &= \cos^2 \theta \mathbf{q}_1^T (-h \mathbf{q}_1) + 2 \sin \theta \cos \theta \mathbf{q}_1^T (-h \mathbf{q}_2) + \sin^2 \theta \mathbf{q}_2^T (-h \mathbf{q}_2) \\ &= \kappa_1 \cos^2 \theta \mathbf{q}_1^T g \mathbf{q}_1 + 2 \sin \theta \cos \theta \kappa_2 \underbrace{\mathbf{q}_1^T g \mathbf{q}_2}_{=0} + \kappa_2 \sin^2 \theta \mathbf{q}_2^T g \mathbf{q}_2 \\ &= \kappa_1 \cos^2 \theta + \kappa_2 \sin^2 \theta; \end{aligned}$$

ergo, we arrive at

$$k_n(\mathbf{x}, \mathbf{d}) = \kappa_1 (\mathbf{d}_1 \cdot \mathbf{d})^2 + \kappa_2 (\mathbf{d}_2 \cdot \mathbf{d})^2. \quad (3.55)$$

3.2.5 • Shape Operator

Defining $\hat{\mathbf{q}}_i = g \mathbf{q}_i$ (for $i = 1, 2$), (3.54) implies that

$$-h g^{-1} \hat{\mathbf{q}}_1 = \kappa_1 \hat{\mathbf{q}}_1, \quad -h g^{-1} \hat{\mathbf{q}}_2 = \kappa_2 \hat{\mathbf{q}}_2.$$

Thus, κ_1 and κ_2 are the maximum and minimum eigenvalues of the matrix $-h g^{-1}$. This leads to the following definition.

Definition 7. Let g and h be the 2×2 matrices corresponding to the first and second fundamental forms of the surface Γ , and define $S = -h g^{-1}$. We call S the **shape operator** of the surface Γ , in terms of the local coordinates $\mathbf{s} = (s_1, s_2)^T$ in \mathbb{R}^2 . The eigenvalues of S are the principal curvatures κ_1 and κ_2 . Note that, in general, S is not symmetric.

Remark 16. For surfaces of higher dimension, the matrix S has a larger size. For example, if $\mathbf{X}: \mathbb{R}^n \rightarrow \mathbb{R}^m$, then S is an $n \times n$ matrix. Moreover, there are n principal curvatures.

This next result gives a transformation rule for S when changing parameters.

Lemma 3.7. Assume the hypothesis of Lemma 3.1. Then,

$$S \circ \mathbf{Y}(\tilde{\mathbf{s}}) = \text{sgn}(\det \mathbf{M})(\mathbf{M}^T)^{-1} \tilde{S}(\tilde{\mathbf{s}}) \mathbf{M}^T, \quad \mathbf{M} = [\nabla_{\tilde{\mathbf{s}}} \mathbf{Y}(\tilde{\mathbf{s}})], \quad (3.56)$$

where S (\tilde{S}) is the shape operator with respect to (U, \mathbf{X}) ($(\tilde{U}, \tilde{\mathbf{X}})$).

Proof. Use Lemmas 3.1 and 3.6. We leave the details to the reader. \square

The shape operator encodes information about the local shape of the surface. Furthermore, the right-hand side of (3.56) involves a similarity transformation of the 2×2 matrix \tilde{S} [62, 97] with a possible sign change. Similarity transformations do not affect the eigenvalues, so the eigenvalues of $S \circ Y(\tilde{s})$ and $\tilde{S}(\tilde{s})$ are the same (up to a change in sign). This means that the principal curvatures, κ_1 and κ_2 , are fundamental quantities that describe the local shape of the surface, i.e., they do not depend on the parameterization. Moreover, the principal directions d_1, d_2 in \mathbb{R}^3 are independent of the parameterization (we leave this as an exercise for the reader).

Remark 17. The term $\text{sgn}(\det M)$ in (3.56) is negative if the change of parameters Y reverses the orientation of the unit normal vector ν . Thus, the sign of the principal curvatures, depends on the orientation of ν .

3.2.6 • Curvature of Surfaces

Definitions of Curvature

It is clear from the previous section that anything we define in terms of the principal curvatures (and principal directions) will be invariant with respect to the parameterization, i.e., they will depend only on the surface Γ .

The following concepts of curvature are useful in differential geometry. First is the sum of the principal curvatures

$$\kappa := \kappa_1 + \kappa_2, \quad (3.57)$$

then the product

$$\kappa_G := \kappa_1 \kappa_2. \quad (3.58)$$

In addition, we have the product of κ with the unit normal vector

$$\kappa := \kappa \nu. \quad (3.59)$$

Definition 8. Equation (3.57) is called the *summed curvature*, (3.58) is called the *Gaussian curvature* (or *Gauss curvature*), and (3.59) is called the *vector curvature* (or *summed curvature vector*).

Remark 18. Most differential geometry texts alternatively define κ as the average of the principal curvatures, i.e., $\kappa = \frac{1}{2}(\kappa_1 + \kappa_2)$. In this case, they refer to it as the *mean curvature*. We will not use this definition in this book. It turns out that the summed curvature is the more natural quantity that appears in various formulas.

The following result gives more information on the sign of the curvatures.

Proposition 6. Let $\Gamma \subset \mathbb{R}^3$ be a regular surface with unit normal vector ν . Moreover, let κ and κ_G be the summed and Gaussian curvatures of Γ , respectively. We then have the following facts.

1. The sign of κ depends on the chosen orientation of ν .
2. The sign of κ_G does not depend on the chosen orientation of ν .
3. The vector curvature $\kappa \nu$ is invariant with respect to a change of parameters.
4. If Γ is the unit sphere with ν chosen to point outward, then $\kappa_1 = \kappa_2 = 1$ and so $\kappa = 2$.

Proof. We leave it to the reader. See the sphere example below for fact 4 above. \square

In fact, for any closed, convex surface with an **outward** pointing normal vector, both principal curvatures are nonnegative, and hence $\kappa \geq 0$.

Explicit Formulas

The following results are left as exercises for the reader. Since κ_1 and κ_2 are the eigenvalues of S , we have that the summed curvature is $\kappa = \text{trace } S = -\text{trace}(h g^{-1})$ which yields the formula

$$\kappa = \kappa_1 + \kappa_2 = -\sum_{i,j=1}^2 g^{ij} h_{ij}. \quad (3.60)$$

As for the Gaussian curvature, it is given by $\kappa_G = \det S$ which leads to

$$\kappa_G = \kappa_1 \kappa_2 = \frac{\det(-h)}{\det g} = \frac{\det h}{\det g}, \quad (3.61)$$

where the last equality follows because h is a 2×2 matrix.

3.2.7 • Classification of Local Surface Shape

The local geometry of a surface can be characterized by the principal curvatures or, alternatively, by the summed and Gaussian curvatures. Suppose Γ is described by a simple surface parameterization:

$$\mathbf{X}(s_1, s_2) = (s_1, s_2, 1 - (s_1^2 + \rho s_2^2))^T, \quad (3.62)$$

where ρ is a tunable parameter. We leave it to the reader to show that

$$\mathbf{v}(s_1, s_2) = \frac{1}{\sqrt{4s_1^2 + 4\rho^2 s_2^2 + 1}} (2s_1, 2\rho s_2, 1)^T, \quad (3.63)$$

and that the principal curvatures and directions, evaluated at $(0, 0, 1)^T$ in Γ , satisfy

$$\begin{aligned} \kappa_1 &= 2, & \mathbf{d}_1 &= (1, 0, 0)^T, \\ \kappa_2 &= 2\rho, & \mathbf{d}_2 &= (0, 1, 0)^T. \end{aligned} \quad (3.64)$$

In Figure 3.6, we plot the surface for different choices of ρ . Each choice shows a distinct surface type. This motivates the following definition.

Definition 9 (surface types). A point of a surface Γ is called

1. *elliptic* if $\kappa_G > 0$ (principal curvatures have the same sign).
2. *hyperbolic* if $\kappa_G < 0$ (principal curvatures have opposite sign).
3. *parabolic* if $\kappa_G = 0$ but $\kappa \neq 0$ (one principal curvature is zero).
4. *planar* if $\kappa_G = 0$ and $\kappa = 0$ (both principal curvatures are zero).

We say Γ is *elliptic* if all points in Γ are elliptic. A similar convention holds for the other surface types.

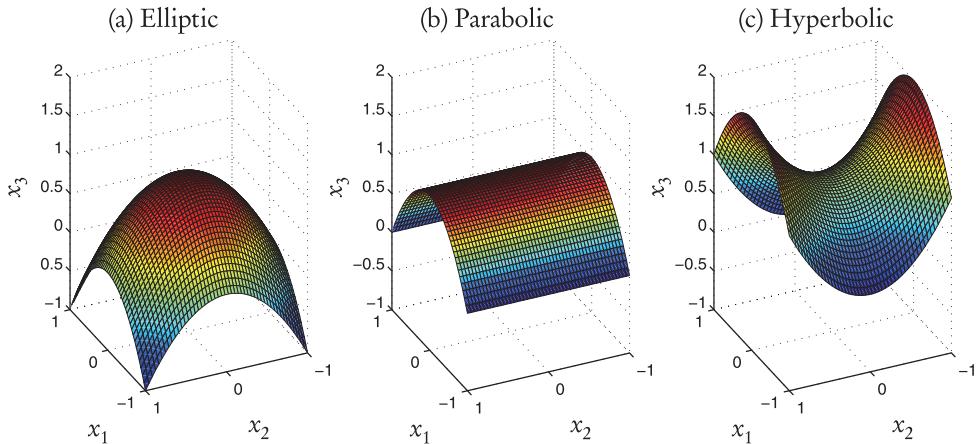


Figure 3.6. Illustration of canonical surface types using the parameterization (3.62). (a) Elliptic surface: $\rho = 1$ produces a “sphere” like surface patch with $\kappa_G > 0$. (b) Parabolic surface: $\rho = 0$ produces a “cylinder” like surface patch with $\kappa_G = 0$ (but $\kappa \neq 0$). (c) Hyperbolic surface: $\rho = -1$ produces a “saddle” like surface patch with $\kappa_G < 0$.

Some classic surface examples are as follows:

- Spheres and ellipsoids are elliptic surfaces.
- Cones and cylinders are parabolic surfaces.
- A helicoid and catenoid are hyperbolic surfaces.
- Of course, planes are planar surfaces.

In general, a surface may contain points of all surface types (see Figure 3.7).

3.2.8 • Examples

Sphere

Recall the sphere parameterization (3.6) and the associated first and second fundamental form coefficients (3.7) and (3.37). Computing $\det g$ and g^{-1} , we get

$$\det g = \frac{1}{1 - (s_1^2 + s_2^2)}, \quad (3.65)$$

$$g^{11} = 1 - s_1^2, \quad g^{12} = g^{21} = s_1 s_2, \quad g^{22} = 1 - s_2^2. \quad (3.66)$$

Applying (3.60) yields

$$\kappa = - \sum_{i,j=1}^2 g^{ij} h_{ij} = 2. \quad (3.67)$$

The vector curvature is then

$$\kappa = \kappa \nu = 2\nu = 2(s_1, s_2, \sqrt{1 - (s_1^2 + s_2^2)})^T. \quad (3.68)$$

Next, compute $\det h$:

$$\det h = \frac{1}{1 - (s_1^2 + s_2^2)}. \quad (3.69)$$

Applying (3.61) gives

$$\kappa_G = \frac{\det h}{\det g} = 1. \quad (3.70)$$

Hence, we see that $\kappa_1 = \kappa_2 = 1$ over the entire unit sphere, which makes sense because curvature and radius are inversely related.

Wavy Surface

Recall the wavy surface parameterization (3.18) and the associated first and second fundamental form coefficients (3.19) and (3.38). Computing $\det g$ and g^{-1} , we get

$$\begin{aligned} \det g &= [1 + (\pi \sin(\pi s_1) \cos(\pi s_2))^2][1 + (\pi \cos(\pi s_1) \sin(\pi s_2))^2] \\ &\quad - [\pi^2 \sin(\pi s_1) \cos(\pi s_2) \cos(\pi s_1) \sin(\pi s_2)]^2, \\ &= 1 + (\pi \sin(\pi s_1) \cos(\pi s_2))^2 + (\pi \cos(\pi s_1) \sin(\pi s_2))^2, \\ g^{11} &= \frac{1 + (\pi \cos(\pi s_1) \sin(\pi s_2))^2}{\det g}, \end{aligned} \quad (3.71)$$

$$\begin{aligned} g^{12} = g^{21} &= -\frac{\pi^2 \sin(\pi s_1) \cos(\pi s_2) \cos(\pi s_1) \sin(\pi s_2)}{\det g}, \\ g^{22} &= \frac{1 + (\pi \sin(\pi s_1) \cos(\pi s_2))^2}{\det g}. \end{aligned} \quad (3.72)$$

Applying (3.60) yields

$$\kappa = - \sum_{i,j=1}^2 g^{ij} h_{ij} = \frac{\pi^2 [2 + \pi^2 (\sin^2(\pi s_1) + \sin^2(\pi s_2))] [\cos(\pi s_1) \cos(\pi s_2)]}{[1 + (\pi \sin(\pi s_1) \cos(\pi s_2))^2 + (\pi \cos(\pi s_1) \sin(\pi s_2))^2]^{3/2}}. \quad (3.73)$$

The vector curvature is then

$$\kappa = \kappa \nu = \kappa \frac{(\pi \sin(\pi s_1) \cos(\pi s_2), \pi \cos(\pi s_1) \sin(\pi s_2), 1)^T}{\sqrt{1 + (\pi \sin(\pi s_1) \cos(\pi s_2))^2 + (\pi \cos(\pi s_1) \sin(\pi s_2))^2}}. \quad (3.74)$$

Next, compute $\det h$:

$$\det h = \pi^4 \frac{(\cos(\pi s_1) \cos(\pi s_2))^2 - (\sin(\pi s_1) \sin(\pi s_2))^2}{1 + (\pi \sin(\pi s_1) \cos(\pi s_2))^2 + (\pi \cos(\pi s_1) \sin(\pi s_2))^2}. \quad (3.75)$$

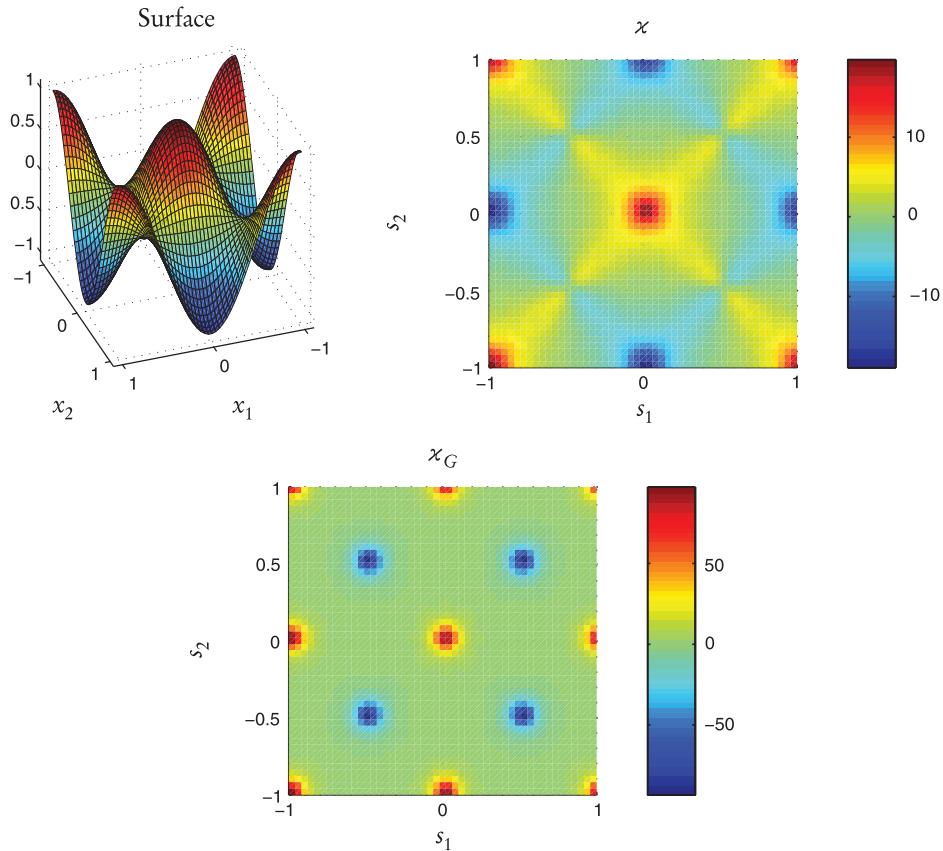


Figure 3.7. Illustration of the curvatures of the surface given by (3.18) (also see Figures 3.1 and 3.4). Notice that the peaks (valleys) of the surface have a positive (negative) summed curvature; the sign of κ is because ν points in the positive z direction. However, the Gaussian curvature is positive at both peaks and valleys. Moreover, the summed curvature is close to zero near the saddle-points of the surface, yet the Gaussian curvature is negative at the saddle-points.

Applying (3.61) gives

$$\kappa_G = \frac{\det h}{\det g} = \pi^4 \frac{(\cos(\pi s_1)\cos(\pi s_2))^2 - (\sin(\pi s_1)\sin(\pi s_2))^2}{[1 + (\pi \sin(\pi s_1)\cos(\pi s_2))^2 + (\pi \cos(\pi s_1)\sin(\pi s_2))^2]^2}. \quad (3.76)$$

A visualization of the summed and Gaussian curvatures is shown in Figure 3.7. The local behavior of the surface matches the classifications in Definition 9 in this chapter.

3.3 ■ Conclusion

This finishes the basic overview of the fundamentals of differential geometry. Of course, there are many other things one could discuss, such as the third fundamental form, geodesics, isometries, etc. The interested reader should consult [24, 25, 60] for more information. For our purposes, we do not need to go into detail on these topics.

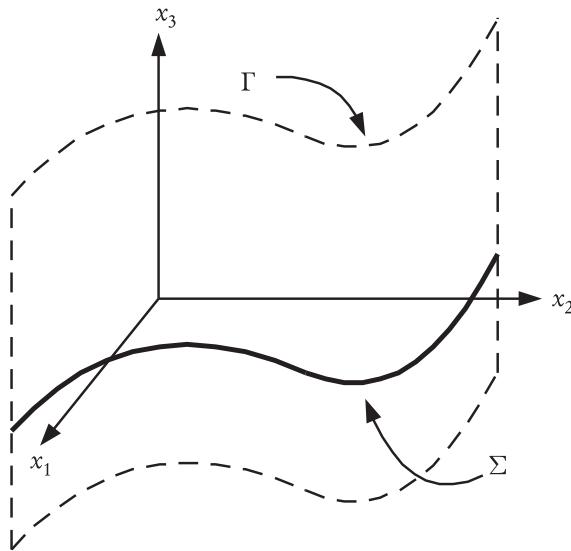


Figure 3.8. A 1-D curve Σ (in the x - y plane) shown as a cross-section of a 2-D surface Γ . Many of the differential geometric formulas derived in later chapters are for a 2-D surface in an ambient 3-D space. Most of these formulas (though not all) also hold for a 1-D curve in a 2-D ambient space. This can be seen by noting that a 1-D curve can be interpreted as a cross-section of a cylindrical surface.

The remainder of the book will formulate most relations and formulas for a 2-D surface (embedded in 3-D space), but many of the derived formulas are also true for a 1-D curve. This is because a 1-D curve can be interpreted as a cross-section of a cylindrical 2-D surface (see Figure 3.8). However, there are exceptions to this rule.

Chapter 4

Calculus on Surfaces

We define and develop the fundamental tools of calculus on a regular surface. We start with the notion of differentiability of functions defined only on a surface. Next, we define the concept of vector fields in a surface. Then we proceed to develop the gradient and Laplacian operators with respect to a surface. These operators allow for alternative expressions of the summed and Gaussian curvatures (discussed earlier). Next, we derive integration by parts on surfaces (i.e., the domain of integration is a surface). We conclude with some identities and inequalities which will be useful in later chapters.

Throughout this whole chapter, we always take Γ to be a regular surface, either with or without a boundary.

4.1 • Functions on Surfaces

If a function $f : \Gamma \rightarrow \mathbb{R}$ is defined only on a regular surface Γ , what does it mean for f to be differentiable? Since f only “lives” in the surface (recall the Flatland discussion in section 2.4.1), we need a notion of derivative *within the surface* (actually in the tangent space).

4.1.1 • Differentiability

Differentiability is defined through the use of the surface parameterization \mathbf{X} .

Definition 10 (differentiable function on a surface). Let $f : \Gamma \rightarrow \mathbb{R}$ be a function defined on a regular surface Γ . We say f is **differentiable at \mathbf{a} in Γ** if, for some open set U and parameterization $\mathbf{X} : U \subset \mathbb{R}^2 \rightarrow \Gamma$ with \mathbf{a} in $\mathbf{X}(U)$, the composite map $f \circ \mathbf{X} : U \rightarrow \mathbb{R}$ is differentiable at $\mathbf{X}^{-1}(\mathbf{a})$. f is **differentiable in Γ** if it is differentiable at all points of Γ .

From Proposition 1 in Chapter 2, we see that this definition does not depend on the parameterization. Indeed, if $\tilde{\mathbf{X}} : \tilde{U} \subset \mathbb{R}^2 \rightarrow \Gamma$ is another parameterization with \mathbf{a} in $\tilde{\mathbf{X}}(\tilde{U})$, and if $\mathbf{Y} = \mathbf{X}^{-1} \circ \tilde{\mathbf{X}}$, then $f \circ \tilde{\mathbf{X}} = f \circ \mathbf{X} \circ \mathbf{Y}$ is also differentiable by the chain rule. Thus, f is differentiable regardless of which parameterization is used to verify its differentiability.

Remark 19. In the following sections, it will be necessary to emphasize the difference between a function defined on Γ and a composite function defined on the reference domain U . Hence,

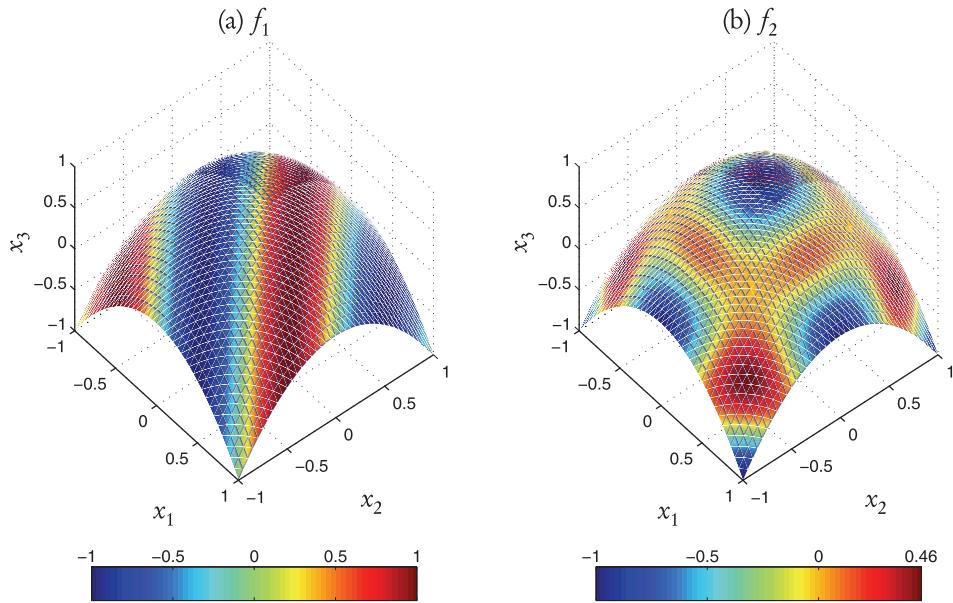


Figure 4.1. Illustration of differentiable functions defined on a surface; see (4.1) and (4.2). Color corresponds to function value.

if $f : \Gamma \rightarrow \mathbb{R}$ and (U, \mathbf{X}) is a local chart, then $\tilde{f} = f \circ \mathbf{X}$ is the associated function defined in terms of the local coordinates in U . In other words, we use \sim to denote functions defined on a reference domain.

4.1.2 • Examples

In the following, let Γ be a regular surface parameterized by $\mathbf{X} : U \rightarrow \mathbb{R}^3$ on the reference domain U . In particular, $U = \{(s_1, s_2)^T \in \mathbb{R}^2 : -1 < s_1, s_2 < 1\}$ and Γ is the paraboloid parameterized by

$$\begin{aligned}\mathbf{X}(s_1, s_2) &= (s_1, s_2, 1 - (s_1^2 + s_2^2))^T, \quad \text{where } (s_1, s_2)^T \in U, \\ \mathbf{X}^{-1}(x_1, x_2, x_3) &= (x_1, x_2)^T \quad \text{for all } (x_1, x_2, x_3)^T \in \Gamma.\end{aligned}$$

Defining Functions on the Reference Domain

A trivial way to define a differentiable function on Γ is to first define a differentiable function on U . For instance, let $\tilde{f} : U \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ be a differentiable function in the usual sense. Then, defining $f := \tilde{f} \circ \mathbf{X}^{-1}$, we have $f : \Gamma \rightarrow \mathbb{R}$, and f is differentiable in the sense of Definition 10.

For example, let $\tilde{f}_1(s_1, s_2) = \sin(\pi(s_1 + s_2))$ which is a differentiable function. Then $f_1 = \tilde{f}_1 \circ \mathbf{X}^{-1}$ is given by

$$f_1(x_1, x_2, x_3) = \sin(\pi(x_1 + x_2)) \quad \text{for all } (x_1, x_2, x_3)^T \in \Gamma. \quad (4.1)$$

See Figure 4.1(a) for a visualization of this function.

Functions Defined in the Ambient Space

Alternatively, one can first define a function $F : \Omega \rightarrow \mathbb{R}$ on a domain $\Omega \subset \mathbb{R}^3$ such that $\Gamma \subset \Omega$. Then, define $f : \Gamma \rightarrow \mathbb{R}$ as the restriction, i.e., $f = F|_{\Gamma}$. For example, let

$$F(x_1, x_2, x_3) = \cos(\pi x_1) \cos(\pi x_2) \cos(\pi x_3) \quad \text{for all } (x_1, x_2, x_3)^T \in \mathbb{R}^3,$$

then define

$$f_2(x_1, x_2, x_3) = F(x_1, x_2, x_3) \quad \text{for all } (x_1, x_2, x_3)^T \in \Gamma. \quad (4.2)$$

See Figure 4.1(b) for a visualization of this function. In order to check the differentiability of f_2 , we must map back to the reference domain. So let $\tilde{f}_2 = f_2 \circ \mathbf{X}$, which is given by

$$\tilde{f}_2(s_1, s_2) = \cos(\pi s_1) \cos(\pi s_2) \cos(\pi(1 - (s_1^2 + s_2^2))) \quad \text{for all } (s_1, s_2)^T \in U.$$

Clearly, \tilde{f}_2 is differentiable with respect to s_1 and s_2 , so f_2 is then differentiable on Γ .

4.1.3 ■ Vector Fields

Recall the usual notion of a (3-D) vector field, i.e., a function, or rule, that assigns to each point in space a vector (in \mathbb{R}^3). A similar definition of differentiability applies to vector-valued functions. In other words, we say that $\mathbf{f} : \Gamma \rightarrow \mathbb{R}^3$, where $\mathbf{f} = (f_1, f_2, f_3)^T$, is differentiable on Γ if all of its component functions f_1, f_2, f_3 are differentiable on Γ . No other restriction is made on \mathbf{f} .

For later sections, we will need the concept of a tangential vector field, which is a slight restriction of a general (3-D) vector field.

Definition 11 (Vector fields in a surface). A tangential vector field $\mathbf{v} : \Gamma \rightarrow \mathbb{R}^3$ in a regular surface Γ is a function which assigns to each \mathbf{x} in Γ a vector $\mathbf{v}(\mathbf{x})$ in $T_{\mathbf{x}}(\Gamma)$. We say the vector field \mathbf{v} is differentiable at \mathbf{y} in Γ if there is an open set U and parameterization \mathbf{X} , and functions $q_1(s_1, s_2), q_2(s_1, s_2)$, such that

$$\mathbf{v} = q_1(s_1, s_2) \partial_{s_1} \mathbf{X}(s_1, s_2) + q_2(s_1, s_2) \partial_{s_2} \mathbf{X}(s_1, s_2) \quad \text{for all } (s_1, s_2)^T \text{ in } U,$$

where $q_1(s_1, s_2), q_2(s_1, s_2)$ are differentiable at \mathbf{y} , i.e., if $(a_1, a_2)^T = \mathbf{X}^{-1}(\mathbf{y})$; then

$$\mathbf{v}(\mathbf{y}) = q_1(a_1, a_2) \partial_{s_1} \mathbf{X}(a_1, a_2) + q_2(a_1, a_2) \partial_{s_2} \mathbf{X}(a_1, a_2),$$

and q_1, q_2 are differentiable at $(a_1, a_2)^T$.

4.2 ■ Differential Operators on Surfaces

We define operators on the surface Γ that are directly analogous to the gradient, divergence, and Laplace operators in standard calculus.

4.2.1 ■ Tangential Directional Derivative

Before defining the notion of a “gradient on a surface” we need to develop a simpler concept: directional derivatives.

Definition 12. Let $\omega : \Gamma \rightarrow \mathbb{R}$ be a function defined on a regular surface Γ . We denote the tangential directional derivative of ω at P in Γ , in the direction \mathbf{v} in $T_P(\Gamma)$, by

$$(D_{\mathbf{v}} \omega)(P) \equiv D_{\mathbf{v}} \omega(P).$$

If $\alpha(t)$ is a parameterization of a curve contained in Γ such that $\alpha(0) = P$ and $\alpha'(0) = \mathbf{v}$, then

$$D_{\mathbf{v}}\omega(P) = \frac{d}{dt}\omega(\alpha(t))\Big|_{t=0}. \quad (4.3)$$

Note that usually \mathbf{v} is a unit vector, but we do not require it here.

If Γ is parameterized by the local chart (U, \mathbf{X}) , then we can say more about (4.3). Recall section 2.4.2. Suppose we can write α as $\alpha(t) = \mathbf{X} \circ \mathbf{s}(t)$, where $\mathbf{s} : I \rightarrow U$ parameterizes a curve in U . Let $\tilde{\omega} = \omega \circ \mathbf{X}$ be the function in local coordinates, i.e., $\tilde{\omega} : U \rightarrow \mathbb{R}$. Then the chain rule gives

$$\frac{d}{dt}\omega(\alpha(t)) = \frac{d}{dt}\tilde{\omega}(\mathbf{s}(t)) = (\mathbf{s}'(t) \cdot \nabla_{\mathbf{s}})\tilde{\omega}(\mathbf{s}) = \nabla \tilde{\omega} \mathbf{s}',$$

where $\nabla \tilde{\omega}$ is a 1×2 row vector and \mathbf{s}' is a 2×1 column vector. Expanding further with the metric tensor $g = (\nabla_{\mathbf{s}} \mathbf{X})^T (\nabla_{\mathbf{s}} \mathbf{X})$, we obtain

$$\begin{aligned} \frac{d}{dt}\omega(\alpha(t)) &= \nabla \tilde{\omega} g^{-1} g \mathbf{s}', \\ &= \nabla \tilde{\omega} g^{-1} (\nabla_{\mathbf{s}} \mathbf{X})^T (\nabla_{\mathbf{s}} \mathbf{X}) \mathbf{s}' \\ &= \nabla \tilde{\omega} g^{-1} (\nabla_{\mathbf{s}} \mathbf{X})^T \alpha'(t), \end{aligned}$$

where the last follows from the chain rule applied to $\alpha'(t)$. Therefore, we arrive at an explicit formula (in local coordinates) for the tangential directional derivative

$$D_{\mathbf{v}}\omega \circ \mathbf{X} = \underbrace{\nabla \tilde{\omega} g^{-1} (\nabla_{\mathbf{s}} \mathbf{X})^T}_{1 \times 3 \text{ row vector}} \cdot \mathbf{v} \quad (4.4)$$

evaluated at \mathbf{s} with $P = \mathbf{X}(\mathbf{s})$.

4.2.2 • Surface Gradient Operator

Adopting the notation in the previous section, we define the gradient operator “on surfaces.”

Definition 13. Let Γ be a regular surface and suppose $\omega : \Gamma \rightarrow \mathbb{R}$ is a differentiable function. We call $\nabla_{\Gamma}\omega : \Gamma \rightarrow \mathbb{R}^3$ the **surface gradient** which assigns to each point P in Γ a vector $\nabla_{\Gamma}\omega(P)$ in $T_P(\Gamma) \subset \mathbb{R}^3$ such that

$$\nabla_{\Gamma}\omega(P) \cdot \mathbf{v} = D_{\mathbf{v}}\omega(P) \quad \text{for all } \mathbf{v} \in T_P(\Gamma). \quad (4.5)$$

Note that this is also referred to as the **tangential gradient**.

From (4.4), we see that

$$(\nabla_{\Gamma}\omega \circ \mathbf{X}) \cdot \mathbf{v} = D_{\mathbf{v}}\omega \circ \mathbf{X} = \nabla \tilde{\omega} g^{-1} (\nabla_{\mathbf{s}} \mathbf{X})^T \cdot \mathbf{v} \quad \text{for all } \mathbf{v} \in T_P(\Gamma).$$

From this, we see that the surface gradient $\nabla_{\Gamma}\omega$ of ω , in local coordinates, is given by

$$(\nabla_{\Gamma}\omega) \circ \mathbf{X} = \sum_{i,j=1}^2 g^{ij} \partial_{s_i} \tilde{\omega} \partial_{s_j} \mathbf{X}^T, \quad (4.6)$$

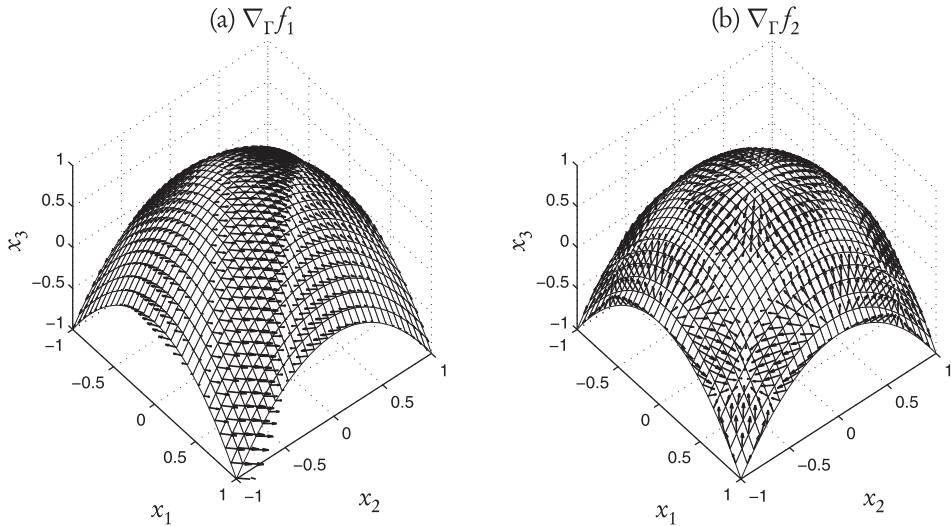


Figure 4.2. Illustration of the surface gradient of the functions depicted in Figure 4.1; see (4.1), (4.2), and (4.6). Arrow lengths are proportional to $|\nabla_{\Gamma}f_i|$ ($i = 1, 2$). Note that the vector field is tangential to the surface.

where $\tilde{\omega} = \omega \circ \mathbf{X}$. Note that $(\nabla_{\Gamma}\omega) \circ \mathbf{X}$ is a 1×3 row vector. Clearly, $(\nabla_{\Gamma}\omega) \circ \mathbf{X}$ is in the tangent space of Γ because $(\nabla_{\Gamma}\omega) \circ \mathbf{X} = a(s_1, s_2)\partial_{s_1}\mathbf{X} + b(s_1, s_2)\partial_{s_2}\mathbf{X}$ for some functions a and b . In other words, $(\nabla_{\Gamma}\omega) \circ \mathbf{X}$ is a tangential vector field. We leave it to the reader to show that (4.6) does not depend on the parameterization. See Figure 4.2 for a visualization of the surface gradient of the functions (4.1) and (4.2).

We can also compute the surface gradient of vector-valued functions. Let $\varphi : \Gamma \rightarrow \mathbb{R}^3$ be a vector function on Γ , and let φ_k denote the coordinate functions of φ (i.e., $\varphi = (\varphi_1, \varphi_2, \varphi_3)$). Let $\tilde{\varphi} = \varphi \circ \mathbf{X}$ and $\tilde{\varphi}_k = \varphi_k \circ \mathbf{X}$ denote the functions in local coordinates. Then we define the surface gradient of a vector function (in local coordinates) by

$$(\nabla_{\Gamma}\varphi) \circ \mathbf{X} = \begin{bmatrix} (\nabla_{\Gamma}\varphi_1) \circ \mathbf{X} \\ (\nabla_{\Gamma}\varphi_2) \circ \mathbf{X} \\ (\nabla_{\Gamma}\varphi_3) \circ \mathbf{X} \end{bmatrix}, \quad \text{where } (\nabla_{\Gamma}\varphi) \circ \mathbf{X} \text{ is a } 3 \times 3 \text{ matrix,} \quad (4.7)$$

and

$$(\nabla_{\Gamma}\varphi_k) \circ \mathbf{X} = \sum_{i,j=1}^2 g^{ij} \partial_{s_i} \tilde{\varphi}_k \partial_{s_j} \mathbf{X}^T \Leftrightarrow (\nabla_{\Gamma}\varphi) \circ \mathbf{X} = (\nabla_s \tilde{\varphi}) g^{-1} (\nabla_s \mathbf{X})^T. \quad (4.8)$$

Surface Gradient Operator on a Sphere

Let $(s_1, s_2) \equiv (\theta, \phi)$ and suppose $\mathbf{X} = \mathbf{X}(\theta, \phi)$ parameterizes a sphere, of radius R , by standard spherical coordinates:

$$\mathbf{X}(\theta, \phi) = R \begin{bmatrix} \cos \theta \sin \phi \\ \sin \theta \sin \phi \\ \cos \phi \end{bmatrix}, \quad \partial_{\theta} \mathbf{X} = R \begin{bmatrix} -\sin \theta \sin \phi \\ \cos \theta \sin \phi \\ 0 \end{bmatrix}, \quad \partial_{\phi} \mathbf{X} = R \begin{bmatrix} \cos \theta \cos \phi \\ \sin \theta \cos \phi \\ -\sin \phi \end{bmatrix}, \quad (4.9)$$

which gives the following metric and inverse metric:

$$g = R^2 \begin{bmatrix} \sin^2 \phi & 0 \\ 0 & 1 \end{bmatrix}, \quad g^{-1} = \frac{1}{R^2} \begin{bmatrix} \frac{1}{\sin^2 \phi} & 0 \\ 0 & 1 \end{bmatrix}. \quad (4.10)$$

For the sphere, (4.6) reduces to

$$(\nabla_\Gamma \omega)^T \circ \mathbf{X} = \frac{1}{R^2} \left\{ \frac{1}{\sin^2 \phi} \partial_\theta \tilde{\omega} \partial_\theta \mathbf{X} + \partial_\phi \tilde{\omega} \partial_\phi \mathbf{X} \right\},$$

which in terms of the tangent basis

$$\tau_\theta = \frac{1}{R^2} \frac{1}{\sin^2 \phi} \partial_\theta \mathbf{X}, \quad \tau_\phi = \frac{1}{R^2} \partial_\phi \mathbf{X} \quad (4.11)$$

is

$$(\nabla_\Gamma \omega)^T \circ \mathbf{X} = \partial_\theta \tilde{\omega} \tau_\theta + \partial_\phi \tilde{\omega} \tau_\phi. \quad (4.12)$$

Surface Gradient Operator on a Curve

Let $s_1 \equiv t$ and suppose $\alpha = \alpha(t)$ parameterizes a curve Σ . So the metric is a 1×1 matrix:

$$g = |\partial_t \alpha|^2, \quad g^{-1} = |\partial_t \alpha|^{-2} \quad (4.13)$$

and (4.6) reduces to

$$(\nabla_\Sigma \omega)^T \circ \alpha = \frac{1}{|\partial_t \alpha|^2} \partial_t \tilde{\omega} \partial_t \alpha = \frac{1}{|\partial_t \alpha|} \partial_t \tilde{\omega} \frac{\partial_t \alpha}{|\partial_t \alpha|}.$$

In terms of arc-length coordinates, we obtain

$$\nabla_\Sigma \omega(\alpha) = \tau^T \frac{d\omega}{d\alpha},$$

where τ is the unit (oriented) tangent vector of Σ and $d/d\alpha = \partial_t / |\partial_t \alpha|$. Thus, on a curve, the surface gradient operator is (in terms of arc-length coordinates)

$$\nabla_\Sigma = \tau^T \frac{d}{d\alpha}. \quad (4.14)$$

Remark 20. We use the subscript Γ in ∇_Γ to denote the surface on which the surface gradient is defined. Similarly, if Σ is a curve, then ∇_Σ is the tangential gradient on Σ .

Projection to the Tangent Space

A common quantity that shows up in differential geometry and shape differential calculus is the surface gradient of the coordinate function on Γ . Let $\text{id}_\Gamma : \Gamma \rightarrow \Gamma$ denote the identity map on Γ , i.e., $\text{id}_\Gamma(x) = x$ for all x in Γ . Note that $\text{id}_\Gamma = \mathbf{X} \circ \mathbf{X}^{-1}$. Using (4.7) and (4.8), we have that $\nabla_\Gamma \text{id}_\Gamma$ is a symmetric 3×3 matrix:

$$\begin{aligned} (\nabla_\Gamma \text{id}_\Gamma) \circ \mathbf{X} &\equiv (\nabla_\Gamma (\mathbf{X} \circ \mathbf{X}^{-1})) \circ \mathbf{X} = \sum_{i,j=1}^2 g^{ij} \partial_{s_i} \mathbf{X} \otimes \partial_{s_j} \mathbf{X} \\ &= \sum_{i,j=1}^2 \partial_{s_i} \mathbf{X} g^{ij} (\partial_{s_j} \mathbf{X})^T = (\nabla_s \mathbf{X}) g^{-1} (\nabla_s \mathbf{X})^T, \end{aligned} \quad (4.15)$$

where the last equality is a product of three matrices. Furthermore, we can write $\nabla_\Gamma \text{id}_\Gamma$ in a way that does **not** involve a parameterization.

Proposition 7. *If Γ is a 2-D surface, and $\text{id}_\Gamma : \Gamma \rightarrow \Gamma$ is the identity map on Γ , then*

$$\nabla_\Gamma \text{id}_\Gamma = \mathbf{I} - \boldsymbol{\nu} \otimes \boldsymbol{\nu}, \quad (4.16)$$

where $\boldsymbol{\nu}$ is the unit normal vector of Γ .

If Σ is a 1-D curve, and $\text{id}_\Sigma : \Sigma \rightarrow \Sigma$ is the identity map on Σ , then

$$\nabla_\Sigma \text{id}_\Sigma = \boldsymbol{\tau} \otimes \boldsymbol{\tau}, \quad (4.17)$$

where $\boldsymbol{\tau}$ is the unit tangent vector of Σ .

Proof. Clearly, (4.16) is a 3×3 projection matrix onto the tangent space of Γ . Since a projection matrix is unique, we just need to show that (4.15) leaves any vectors in the tangent space unchanged and maps vectors parallel to $\boldsymbol{\nu}$ to the zero vector.

Let \mathbf{v} be an arbitrary vector in the tangent space of Γ , i.e., \mathbf{v} can be written as

$$\mathbf{v} = q_1(s_1, s_2) \partial_{s_1} \mathbf{X} + q_2(s_1, s_2) \partial_{s_2} \mathbf{X} = (\nabla_s \mathbf{X}) \mathbf{q}, \quad \text{where } \mathbf{q} = (q_1, q_2)^T \in \mathbb{R}^2.$$

Computing the matrix-vector product, we get

$$(\nabla_\Gamma \text{id}_\Gamma) \circ \mathbf{X} \cdot \mathbf{v} = (\nabla_s \mathbf{X}) g^{-1} (\nabla_s \mathbf{X})^T (\nabla_s \mathbf{X}) \mathbf{q} = (\nabla_s \mathbf{X}) g^{-1} g \mathbf{q} = (\nabla_s \mathbf{X}) \mathbf{q} = \mathbf{v},$$

where we have used (3.17) (i.e., $g = (\nabla_s \mathbf{X})^T (\nabla_s \mathbf{X})$). Clearly, $[(\nabla_\Gamma \text{id}_\Gamma) \circ \mathbf{X}] \cdot \boldsymbol{\nu} = 0$ because $\boldsymbol{\nu} \cdot \partial_{s_i} \mathbf{X} = 0$ ($i = 1, 2$). This proves (4.16). We leave the proof of (4.17) to the reader. \square

4.2.3 ■ Surface Gradient and the Euclidean Gradient

Let us consider f defined on the ambient space, i.e., $f : \mathbb{R}^3 \rightarrow \mathbb{R}$. Then one can write $\nabla_\Gamma f$ in terms of the standard gradient $\nabla = (\partial_{x_1}, \partial_{x_2}, \partial_{x_3})$:

$$\nabla_\Gamma f = \nabla f - [(\boldsymbol{\nu} \cdot \nabla) f] \boldsymbol{\nu}^T \Leftrightarrow (\nabla_\Gamma f)_i = \partial_{x_i} f - (\boldsymbol{\nu})_i \sum_{j=1}^3 (\boldsymbol{\nu})_j \partial_{x_j} f. \quad (4.18)$$

To see this, let $\boldsymbol{\alpha}(t)$ be a parameterization of a curve Σ contained in Γ such that $\boldsymbol{\alpha}(0) = P$ and $\boldsymbol{\alpha}'(0) = \mathbf{v}$. From (4.3), this implies

$$D_{\mathbf{v}} f(P) = \frac{d}{dt} f(\boldsymbol{\alpha}(t)) \Big|_{t=0} = (\boldsymbol{\alpha}'(t) \cdot \nabla) f \Big|_{t=0} = (\mathbf{v} \cdot \nabla) f(P)$$

by the standard chain rule. This holds for all curves contained in Γ , and thus for any \mathbf{v} in $T_P(\Gamma)$. Thus, the definition of the surface gradient (4.5) yields

$$\nabla_\Gamma f(P) \cdot \mathbf{v} = \nabla f(P) \cdot \mathbf{v} \quad \text{for all } \mathbf{v} \in T_P(\Gamma).$$

Moreover, one can show $\nabla_\Gamma f \cdot \boldsymbol{\nu} = 0$ (using (4.6)), as well as $(\nabla f - [(\boldsymbol{\nu} \cdot \nabla) f] \boldsymbol{\nu}^T) \cdot \boldsymbol{\nu} = 0$, which proves (4.18). Therefore, we can write

$$\nabla_\Gamma = \nabla_\Gamma \text{id}_\Gamma \cdot \nabla = (\mathbf{I} - \boldsymbol{\nu} \otimes \boldsymbol{\nu}) \cdot \nabla. \quad (4.19)$$

4.2.4 ■ Surface Divergence Operator

The usual divergence operator is defined as $\nabla \cdot \mathbf{v} = \text{trace}(\nabla \mathbf{v})$. We define the surface divergence operator in a similar way (again in local coordinates):

$$(\nabla_\Gamma \cdot \varphi) \circ \mathbf{X} = [\text{trace}(\nabla_\Gamma \varphi)] \circ \mathbf{X} = \sum_{i,j=1}^2 g^{ij} \partial_{s_i} \tilde{\varphi} \cdot \partial_{s_j} \mathbf{X} = \sum_{i,j=1}^2 g^{ij} \sum_{k=1}^3 \partial_{s_i} \tilde{\varphi}_k \partial_{s_j} X_k, \quad (4.20)$$

where $(\nabla_\Gamma \cdot \varphi) \circ \mathbf{X}$ is a scalar function.

Surface Divergence of the Identity Map

Viewing “ \mathbf{x} ” as the identity map in \mathbb{R}^n (Euclidean space), from standard multivariable calculus, we know that $\nabla \cdot \mathbf{x} = n$. We show a similar result for the surface divergence of id_Γ .

Proposition 8. *If Γ is a 2-D surface, and id_Γ is the identity map on Γ , then*

$$\nabla_\Gamma \cdot \text{id}_\Gamma = 2. \quad (4.21)$$

If Σ is a 1-D curve, and id_Σ is the identity map on Σ , then

$$\nabla_\Sigma \cdot \text{id}_\Sigma = 1. \quad (4.22)$$

Proof. From (4.16), we have $\nabla_\Gamma \text{id}_\Gamma = \mathbf{I} - \boldsymbol{\nu} \otimes \boldsymbol{\nu}$. Since $\nabla_\Gamma \cdot \text{id}_\Gamma = \text{trace}(\nabla_\Gamma \text{id}_\Gamma)$ we see that

$$\nabla_\Gamma \cdot \text{id}_\Gamma = \text{trace}(\mathbf{I}) - \text{trace}(\boldsymbol{\nu} \otimes \boldsymbol{\nu}) = 3 - \boldsymbol{\nu} \cdot \boldsymbol{\nu} = 3 - 1 = 2.$$

Next, from (4.17), we have $\nabla_\Sigma \text{id}_\Sigma = \boldsymbol{\tau} \otimes \boldsymbol{\tau}$. Again, we obtain

$$\nabla_\Sigma \cdot \text{id}_\Sigma = \text{trace}(\boldsymbol{\tau} \otimes \boldsymbol{\tau}) = \boldsymbol{\tau} \cdot \boldsymbol{\tau} = 1. \quad \square$$

4.2.5 ■ Surface Laplacian Operator

Before defining the surface Laplacian operator, we compute some useful formulas.

Intermediate Formulas

Proposition 9.

$$\frac{1}{\sqrt{\det(g)}} \partial_{s_q} \sqrt{\det(g)} = \frac{1}{2} \sum_{i,j=1}^2 g^{ij} \partial_{s_q} (g_{ij}).$$

Proof. Recall that $\det(g) = g_{11}g_{22} - g_{12}g_{21}$; thus so we have

$$\partial_{s_q} \sqrt{\det(g)} = \frac{1}{2\sqrt{\det(g)}} (g_{22} \partial_{s_q} g_{11} + g_{11} \partial_{s_q} g_{22} - g_{21} \partial_{s_q} g_{12} - g_{12} \partial_{s_q} g_{21}).$$

Then, a slight manipulation gives

$$\partial_{s_q} \sqrt{\det(g)} = \frac{\sqrt{\det(g)}}{2} \left(\frac{g_{22}}{\det(g)} \partial_{s_q} g_{11} + \frac{g_{11}}{\det(g)} \partial_{s_q} g_{22} - \frac{g_{21}}{\det(g)} \partial_{s_q} g_{12} - \frac{g_{12}}{\det(g)} \partial_{s_q} g_{21} \right).$$

After using (3.13), and the fact that g is symmetric, we get

$$\begin{aligned}\partial_{s_q} \sqrt{\det(g)} &= \frac{\sqrt{\det(g)}}{2} (g^{11} \partial_{s_q} g_{11} + g^{22} \partial_{s_q} g_{22} + g^{12} \partial_{s_q} g_{12} + g^{21} \partial_{s_q} g_{21}) \\ &= \frac{\sqrt{\det(g)}}{2} \sum_{i,j=1}^2 g^{ij} \partial_{s_q} (g_{ij}),\end{aligned}$$

which implies the assertion. \square

Proposition 10.

$$\sum_{i,j=1}^2 g^{ij} (\partial_{s_q} \partial_{s_i} \mathbf{X}) \cdot (\partial_{s_j} \mathbf{X}) = \frac{1}{\sqrt{\det(g)}} \partial_{s_q} \sqrt{\det(g)}.$$

Proof. Start by writing out the sum

$$\begin{aligned}\sum_{i,j=1}^2 g^{ij} (\partial_{s_q} \partial_{s_i} \mathbf{X}) \cdot (\partial_{s_j} \mathbf{X}) &= g^{11} (\partial_{s_q} \partial_{s_1} \mathbf{X}) \cdot (\partial_{s_1} \mathbf{X}) + g^{12} (\partial_{s_q} \partial_{s_1} \mathbf{X}) \cdot (\partial_{s_2} \mathbf{X}) \\ &\quad + g^{21} (\partial_{s_q} \partial_{s_2} \mathbf{X}) \cdot (\partial_{s_1} \mathbf{X}) + g^{22} (\partial_{s_q} \partial_{s_2} \mathbf{X}) \cdot (\partial_{s_2} \mathbf{X}).\end{aligned}$$

Using the standard product rule and the fact that $g^{12} = g^{21}$, we get

$$\begin{aligned}\sum_{i,j=1}^2 g^{ij} (\partial_{s_q} \partial_{s_i} \mathbf{X}) \cdot (\partial_{s_j} \mathbf{X}) &= \frac{g^{11}}{2} \partial_{s_q} (\partial_{s_1} \mathbf{X} \cdot \partial_{s_1} \mathbf{X}) + \frac{g^{12}}{2} \partial_{s_q} (\partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \mathbf{X}) \\ &\quad + \frac{g^{21}}{2} \partial_{s_q} (\partial_{s_2} \mathbf{X} \cdot \partial_{s_1} \mathbf{X}) + \frac{g^{22}}{2} \partial_{s_q} (\partial_{s_2} \mathbf{X} \cdot \partial_{s_2} \mathbf{X}).\end{aligned}$$

Plugging in (3.2) gives

$$\sum_{i,j=1}^2 g^{ij} (\partial_{s_q} \partial_{s_i} \mathbf{X}) \cdot (\partial_{s_j} \mathbf{X}) = \frac{1}{2} \sum_{i,j=1}^2 g^{ij} \partial_{s_q} (g_{ij}) = \frac{1}{\sqrt{\det(g)}} \partial_{s_q} \sqrt{\det(g)},$$

where the last equality follows from Proposition 9. \square

Define Laplace–Beltrami Operator

The standard Laplacian is written as $\Delta = \nabla \cdot \nabla$ (i.e., divergence of the gradient). We define the surface Laplacian (or Laplace–Beltrami) operator in a similar way:

$$\Delta_\Gamma \omega := \nabla_\Gamma \cdot \nabla_\Gamma \omega, \tag{4.23}$$

where ω is a scalar function on Γ . Applying the surface divergence (4.20) to the surface gradient of ω (4.6) gives (in local coordinates)

$$(\Delta_\Gamma \omega) \circ \mathbf{X} = \sum_{i,j=1}^2 g^{ij} \partial_{s_i} \left\{ \sum_{p,q=1}^2 g^{pq} \partial_{s_p} \tilde{\omega} \partial_{s_q} \mathbf{X} \right\} \cdot \partial_{s_j} \mathbf{X}. \tag{4.24}$$

However, there is a more convenient form in which we can put (4.24). Expanding slightly and applying the product rule gives

$$\begin{aligned} (\Delta_\Gamma \omega) \circ \mathbf{X} = & \sum_{i,j=1}^2 \sum_{p,q=1}^2 g^{ij} \{ (\partial_{s_i} g^{pq}) (\partial_{s_p} \tilde{\omega}) \partial_{s_q} \mathbf{X} + g^{pq} (\partial_{s_i} \partial_{s_p} \tilde{\omega}) \partial_{s_q} \mathbf{X} \\ & + g^{pq} (\partial_{s_p} \tilde{\omega}) (\partial_{s_i} \partial_{s_q} \mathbf{X}) \} \cdot \partial_{s_j} \mathbf{X}. \end{aligned}$$

This can be simplified slightly using (3.2):

$$\begin{aligned} (\Delta_\Gamma \omega) \circ \mathbf{X} = & \sum_{i,j=1}^2 \sum_{p,q=1}^2 \{ (\partial_{s_i} g^{pq}) (\partial_{s_p} \tilde{\omega}) g^{ij} g_{qj} + g^{pq} (\partial_{s_i} \partial_{s_p} \tilde{\omega}) g^{ij} g_{qj} \\ & + g^{pq} g^{ij} (\partial_{s_p} \tilde{\omega}) (\partial_{s_i} \partial_{s_q} \mathbf{X}) \cdot \partial_{s_j} \mathbf{X} \}, \end{aligned}$$

followed by plugging in (3.14) to get

$$\begin{aligned} (\Delta_\Gamma \omega) \circ \mathbf{X} = & \sum_{i=1}^2 \sum_{p,q=1}^2 \{ (\partial_{s_i} g^{pq}) (\partial_{s_p} \tilde{\omega}) \delta_q^i + g^{pq} (\partial_{s_i} \partial_{s_p} \tilde{\omega}) \delta_q^i \} \\ & + \sum_{i,j=1}^2 \sum_{p,q=1}^2 \{ g^{pq} g^{ij} (\partial_{s_p} \tilde{\omega}) (\partial_{s_i} \partial_{s_q} \mathbf{X}) \cdot \partial_{s_j} \mathbf{X} \}. \end{aligned}$$

After using the definition of the “Kronecker delta” (3.15), renaming certain indices, and rearranging slightly, we get

$$\begin{aligned} (\Delta_\Gamma \omega) \circ \mathbf{X} = & \sum_{i,j=1}^2 \{ (\partial_{s_j} g^{ij}) (\partial_{s_i} \tilde{\omega}) + g^{ij} (\partial_{s_j} \partial_{s_i} \tilde{\omega}) \} \\ & + \sum_{p,q=1}^2 g^{pq} (\partial_{s_p} \tilde{\omega}) \sum_{i,j=1}^2 g^{ij} (\partial_{s_q} \partial_{s_i} \mathbf{X}) \cdot \partial_{s_j} \mathbf{X}. \end{aligned}$$

Plugging in Proposition 10 simplifies to

$$(\Delta_\Gamma \omega) \circ \mathbf{X} = \sum_{i,j=1}^2 \{ (\partial_{s_j} g^{ij}) (\partial_{s_i} \tilde{\omega}) + g^{ij} (\partial_{s_j} \partial_{s_i} \tilde{\omega}) \} + \sum_{p,q=1}^2 g^{pq} (\partial_{s_p} \tilde{\omega}) \frac{1}{\sqrt{\det(g)}} \partial_{s_q} \sqrt{\det(g)}.$$

Relabeling p, q as i, j and factoring out $\frac{1}{\sqrt{\det(g)}}$ gives

$$\begin{aligned} (\Delta_\Gamma \omega) \circ \mathbf{X} = & \frac{1}{\sqrt{\det(g)}} \sum_{i,j=1}^2 \{ \sqrt{\det(g)} (\partial_{s_j} g^{ij}) (\partial_{s_i} \tilde{\omega}) + \sqrt{\det(g)} g^{ij} (\partial_{s_j} \partial_{s_i} \tilde{\omega}) \\ & + (\partial_{s_j} \sqrt{\det(g)}) g^{ij} (\partial_{s_i} \tilde{\omega}) \}. \end{aligned}$$

Applying the product rule in reverse gives the alternate form of the surface Laplacian applied to a scalar function in local coordinates:

$$(\Delta_\Gamma \omega) \circ \mathbf{X} = \frac{1}{\sqrt{\det(g)}} \sum_{i,j=1}^2 \partial_{s_j} \left\{ \sqrt{\det(g)} g^{ij} (\partial_{s_i} \tilde{\omega}) \right\}. \quad (4.25)$$

And the surface Laplacian of a 3×1 column vector φ is computed by applying Δ_Γ to each component of the vector,

$$(\Delta_\Gamma \varphi) \circ \mathbf{X} = \begin{bmatrix} (\Delta_\Gamma \varphi_1) \circ \mathbf{X} \\ (\Delta_\Gamma \varphi_2) \circ \mathbf{X} \\ (\Delta_\Gamma \varphi_3) \circ \mathbf{X} \end{bmatrix}, \quad (4.26)$$

which is also a 3×1 vector.

Alternative Formula for the Surface Laplacian

We give another formula for computing the surface Laplacian of a general vector function on a 2-D surface Γ in local coordinates (see Appendix A).

Proposition 11.

$$\begin{aligned} (\Delta_\Gamma \varphi) \circ \mathbf{X} = & \frac{1}{\sqrt{\det(g)}} \left\{ -\partial_{s_1} \left[\frac{[(\partial_{s_1} \tilde{\varphi} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \tilde{\varphi})]}{\sqrt{\det(g)}} \right] \times \partial_{s_2} \mathbf{X} \right. \\ & - \partial_{s_1} \mathbf{X} \times \partial_{s_2} \left[\frac{[(\partial_{s_1} \tilde{\varphi} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \tilde{\varphi})]}{\sqrt{\det(g)}} \right] \\ & + \partial_{s_1} \left[\frac{[(\partial_{s_1} \tilde{\varphi} \cdot \partial_{s_2} \mathbf{X}) - (\partial_{s_2} \tilde{\varphi} \cdot \partial_{s_1} \mathbf{X})]}{\sqrt{\det(g)}} \right] \partial_{s_2} \mathbf{X} \\ & - \partial_{s_2} \left[\frac{[(\partial_{s_1} \tilde{\varphi} \cdot \partial_{s_2} \mathbf{X}) - (\partial_{s_2} \tilde{\varphi} \cdot \partial_{s_1} \mathbf{X})]}{\sqrt{\det(g)}} \right] \partial_{s_1} \mathbf{X} \\ & \left. + (\partial_{s_1} \mathbf{\nu} \times \partial_{s_2} \tilde{\varphi}) + (\partial_{s_1} \tilde{\varphi} \times \partial_{s_2} \mathbf{\nu}) \right\} \end{aligned} \quad (4.27)$$

for all (s_1, s_2) in a reference domain U_i of the atlas $(\{U_i\}, \mathbf{X})$, where $\tilde{\varphi} = \varphi \circ \mathbf{X}$.

Proof. For notational convenience, let $Q = \sqrt{\det(g)}$. Writing out the sum in (4.25) gives

$$\begin{aligned} (\Delta_\Gamma \varphi) \circ \mathbf{X} = & \frac{1}{Q} \left\{ \partial_{s_1} [Q g^{11}(\partial_{s_1} \tilde{\varphi})] + \partial_{s_2} [Q g^{12}(\partial_{s_1} \tilde{\varphi})] \right. \\ & + \partial_{s_1} [Q g^{21}(\partial_{s_2} \tilde{\varphi})] + \partial_{s_2} [Q g^{22}(\partial_{s_2} \tilde{\varphi})] \Big\} \\ = & \frac{1}{Q} \left\{ \partial_{s_1} \left[\frac{g_{22}(\partial_{s_1} \tilde{\varphi}) - g_{21}(\partial_{s_2} \tilde{\varphi})}{Q} \right] + \partial_{s_2} \left[\frac{g_{11}(\partial_{s_2} \tilde{\varphi}) - g_{12}(\partial_{s_1} \tilde{\varphi})}{Q} \right] \right\}, \end{aligned}$$

where $\tilde{\varphi} = \varphi \circ \mathbf{X}$. Writing (A.4) as $(\mathbf{a} \cdot \mathbf{c})\mathbf{b} = \mathbf{a} \times (\mathbf{b} \times \mathbf{c}) + (\mathbf{a} \cdot \mathbf{b})\mathbf{c}$ and noting

$$\begin{aligned} g_{22}(\partial_{s_1} \tilde{\varphi}) &= (\partial_{s_2} \mathbf{X} \cdot \partial_{s_2} \mathbf{X})(\partial_{s_1} \tilde{\varphi}) = \partial_{s_2} \mathbf{X} \times (\partial_{s_1} \tilde{\varphi} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_2} \mathbf{X} \cdot \partial_{s_1} \tilde{\varphi})\partial_{s_2} \mathbf{X}, \\ g_{21}(\partial_{s_2} \tilde{\varphi}) &= (\partial_{s_2} \mathbf{X} \cdot \partial_{s_1} \mathbf{X})(\partial_{s_2} \tilde{\varphi}) = \partial_{s_2} \mathbf{X} \times (\partial_{s_2} \tilde{\varphi} \times \partial_{s_1} \mathbf{X}) + (\partial_{s_2} \mathbf{X} \cdot \partial_{s_2} \tilde{\varphi})\partial_{s_1} \mathbf{X}, \\ g_{11}(\partial_{s_2} \tilde{\varphi}) &= (\partial_{s_1} \mathbf{X} \cdot \partial_{s_1} \mathbf{X})(\partial_{s_2} \tilde{\varphi}) = \partial_{s_1} \mathbf{X} \times (\partial_{s_2} \tilde{\varphi} \times \partial_{s_1} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \tilde{\varphi})\partial_{s_1} \mathbf{X}, \\ g_{12}(\partial_{s_1} \tilde{\varphi}) &= (\partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \mathbf{X})(\partial_{s_1} \tilde{\varphi}) = \partial_{s_1} \mathbf{X} \times (\partial_{s_1} \tilde{\varphi} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \cdot \partial_{s_1} \tilde{\varphi})\partial_{s_2} \mathbf{X}, \end{aligned}$$

then plugging it in, we get

$$\begin{aligned} (\Delta_\Gamma \varphi) \circ \mathbf{X} = \frac{1}{Q} \left\{ \partial_{s_1} \left[\frac{\partial_{s_2} \mathbf{X} \times (\partial_{s_1} \tilde{\varphi} \times \partial_{s_2} \mathbf{X}) - \partial_{s_2} \mathbf{X} \times (\partial_{s_2} \tilde{\varphi} \times \partial_{s_1} \mathbf{X})}{Q} \right] \right. \\ + \partial_{s_1} \left[\frac{(\partial_{s_2} \mathbf{X} \cdot \partial_{s_1} \tilde{\varphi}) \partial_{s_2} \mathbf{X} - (\partial_{s_2} \mathbf{X} \cdot \partial_{s_2} \tilde{\varphi}) \partial_{s_1} \mathbf{X}}{Q} \right] \\ + \partial_{s_2} \left[\frac{\partial_{s_1} \mathbf{X} \times (\partial_{s_2} \tilde{\varphi} \times \partial_{s_1} \mathbf{X}) - \partial_{s_1} \mathbf{X} \times (\partial_{s_1} \tilde{\varphi} \times \partial_{s_2} \mathbf{X})}{Q} \right] \\ \left. + \partial_{s_2} \left[\frac{(\partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \tilde{\varphi}) \partial_{s_1} \mathbf{X} - (\partial_{s_1} \mathbf{X} \cdot \partial_{s_1} \tilde{\varphi}) \partial_{s_2} \mathbf{X}}{Q} \right] \right\}. \end{aligned}$$

More rearranging yields

$$\begin{aligned} (\Delta_\Gamma \varphi) \circ \mathbf{X} = \frac{1}{Q} \left\{ -\partial_{s_1} \left[\frac{[(\partial_{s_1} \tilde{\varphi} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \tilde{\varphi})] \times \partial_{s_2} \mathbf{X}}{Q} \right] \right. \\ - \partial_{s_2} \left[\frac{\partial_{s_1} \mathbf{X} \times [(\partial_{s_1} \mathbf{X} \times \partial_{s_2} \tilde{\varphi}) + (\partial_{s_1} \tilde{\varphi} \times \partial_{s_2} \mathbf{X})]}{Q} \right] \\ + \partial_{s_1} \left[\frac{(\partial_{s_2} \mathbf{X} \cdot \partial_{s_1} \tilde{\varphi}) \partial_{s_2} \mathbf{X} - (\partial_{s_2} \mathbf{X} \cdot \partial_{s_2} \tilde{\varphi}) \partial_{s_1} \mathbf{X}}{Q} \right] \\ \left. + \partial_{s_2} \left[\frac{(\partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \tilde{\varphi}) \partial_{s_1} \mathbf{X} - (\partial_{s_1} \mathbf{X} \cdot \partial_{s_1} \tilde{\varphi}) \partial_{s_2} \mathbf{X}}{Q} \right] \right\}. \end{aligned}$$

Next, use the same vector identity to rewrite $(\partial_{s_2} \mathbf{X} \cdot \partial_{s_2} \tilde{\varphi}) \partial_{s_1} \mathbf{X}$ and $(\partial_{s_1} \mathbf{X} \cdot \partial_{s_1} \tilde{\varphi}) \partial_{s_2} \mathbf{X}$:

$$\begin{aligned} (\partial_{s_2} \tilde{\varphi} \cdot \partial_{s_2} \mathbf{X}) \partial_{s_1} \mathbf{X} &= \partial_{s_2} \tilde{\varphi} \times (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_2} \tilde{\varphi} \cdot \partial_{s_1} \mathbf{X}) \partial_{s_2} \mathbf{X}, \\ (\partial_{s_1} \tilde{\varphi} \cdot \partial_{s_1} \mathbf{X}) \partial_{s_2} \mathbf{X} &= \partial_{s_1} \tilde{\varphi} \times (\partial_{s_2} \mathbf{X} \times \partial_{s_1} \mathbf{X}) + (\partial_{s_1} \tilde{\varphi} \cdot \partial_{s_2} \mathbf{X}) \partial_{s_1} \mathbf{X}, \\ &= -\partial_{s_1} \tilde{\varphi} \times (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \tilde{\varphi} \cdot \partial_{s_2} \mathbf{X}) \partial_{s_1} \mathbf{X}, \end{aligned}$$

and plug it in:

$$\begin{aligned} (\Delta_\Gamma \varphi) \circ \mathbf{X} = \frac{1}{Q} \left\{ -\partial_{s_1} \left[\frac{[(\partial_{s_1} \tilde{\varphi} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \tilde{\varphi})] \times \partial_{s_2} \mathbf{X}}{Q} \right] \right. \\ - \partial_{s_2} \left[\frac{\partial_{s_1} \mathbf{X} \times [(\partial_{s_1} \mathbf{X} \times \partial_{s_2} \tilde{\varphi}) + (\partial_{s_1} \tilde{\varphi} \times \partial_{s_2} \mathbf{X})]}{Q} \right] \\ + \partial_{s_1} \left[\frac{[(\partial_{s_2} \mathbf{X} \cdot \partial_{s_1} \tilde{\varphi}) - (\partial_{s_2} \tilde{\varphi} \cdot \partial_{s_1} \mathbf{X})] \partial_{s_2} \mathbf{X}}{Q} \right] \\ + \partial_{s_2} \left[\frac{[(\partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \tilde{\varphi}) - (\partial_{s_1} \tilde{\varphi} \cdot \partial_{s_2} \mathbf{X})] \partial_{s_1} \mathbf{X}}{Q} \right] \\ \left. - \partial_{s_1} \left[\frac{\partial_{s_2} \tilde{\varphi} \times (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X})}{Q} \right] - \partial_{s_2} \left[\frac{-\partial_{s_1} \tilde{\varphi} \times (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X})}{Q} \right] \right\}. \end{aligned}$$

Next, we simplify:

$$\begin{aligned}
(\Delta_\Gamma \varphi) \circ \mathbf{X} = & \frac{1}{Q} \left\{ -\partial_{s_1} \left[\frac{[(\partial_{s_1} \tilde{\varphi} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \tilde{\varphi})] \times \partial_{s_2} \mathbf{X}}{Q} \right] \right. \\
& - \partial_{s_2} \left[\frac{\partial_{s_1} \mathbf{X} \times [(\partial_{s_1} \tilde{\varphi} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \tilde{\varphi})]}{Q} \right] \\
& + \partial_{s_1} \left[\frac{[(\partial_{s_2} \mathbf{X} \cdot \partial_{s_1} \tilde{\varphi}) - (\partial_{s_2} \tilde{\varphi} \cdot \partial_{s_1} \mathbf{X})]}{Q} \right] \partial_{s_2} \mathbf{X} \\
& + \partial_{s_2} \left[\frac{[(\partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \tilde{\varphi}) - (\partial_{s_1} \tilde{\varphi} \cdot \partial_{s_2} \mathbf{X})]}{Q} \right] \partial_{s_1} \mathbf{X} \\
& + \underbrace{\left[\frac{[(\partial_{s_2} \mathbf{X} \cdot \partial_{s_1} \tilde{\varphi}) - (\partial_{s_2} \tilde{\varphi} \cdot \partial_{s_1} \mathbf{X})]}{Q} \right] \partial_{s_1} \partial_{s_2} \mathbf{X}}_{\text{these terms cancel}} \\
& + \underbrace{\left[\frac{[(\partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \tilde{\varphi}) - (\partial_{s_1} \tilde{\varphi} \cdot \partial_{s_2} \mathbf{X})]}{Q} \right] \partial_{s_2} \partial_{s_1} \mathbf{X}}_{\text{these terms cancel}} \\
& \left. + \partial_{s_1} [\boldsymbol{\nu} \times \partial_{s_2} \tilde{\varphi}] + \partial_{s_2} [\partial_{s_1} \tilde{\varphi} \times \boldsymbol{\nu}] \right\}.
\end{aligned}$$

After noting similar cancelations in the other terms, we obtain (4.27). \square

4.3 ■ Other Curvature Formulas

We would like to have another formula to compute the summed curvature κ and Gaussian curvature κ_G .

4.3.1 ■ Alternative Summed Curvature

Proposition 12.

$$(-\Delta_\Gamma \text{id}_\Gamma) \circ \mathbf{X} = \frac{(\partial_{s_1} \boldsymbol{\nu} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \boldsymbol{\nu})}{\sqrt{\det(g)}}, \quad (4.28)$$

$$\begin{aligned}
(-\Delta_\Gamma \text{id}_\Gamma) &= \kappa \boldsymbol{\nu}, \\
(-\Delta_\Gamma \text{id}_\Gamma) \cdot \boldsymbol{\nu} &= \kappa.
\end{aligned} \quad (4.29)$$

Proof. First, we compute $-\Delta_\Gamma(\mathbf{X} \circ \mathbf{X}^{-1})$ in local coordinates. According to (4.25) and (4.26), and using the definition (3.13) for the inverse metric, we have

$$\begin{aligned}
(-\Delta_\Gamma \text{id}_\Gamma) \circ \mathbf{X} = & \frac{-1}{\sqrt{\det(g)}} \left\{ \partial_{s_1} \left(\frac{(\partial_{s_2} \mathbf{X} \cdot \partial_{s_2} \mathbf{X}) \partial_{s_1} \mathbf{X} - (\partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \mathbf{X}) \partial_{s_2} \mathbf{X}}{\sqrt{\det(g)}} \right) \right. \\
& \left. + \partial_{s_2} \left(\frac{(\partial_{s_1} \mathbf{X} \cdot \partial_{s_1} \mathbf{X}) \partial_{s_2} \mathbf{X} - (\partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \mathbf{X}) \partial_{s_1} \mathbf{X}}{\sqrt{\det(g)}} \right) \right\}.
\end{aligned}$$

Now, by the vector identity (A.4), this becomes

$$(-\Delta_\Gamma \text{id}_\Gamma) \circ \mathbf{X} = \frac{-1}{\sqrt{\det(g)}} \left\{ \partial_{s_1} \left(\frac{\partial_{s_2} \mathbf{X} \times (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X})}{\sqrt{\det(g)}} \right) \right. \\ \left. + \partial_{s_2} \left(\frac{\partial_{s_1} \mathbf{X} \times (\partial_{s_2} \mathbf{X} \times \partial_{s_1} \mathbf{X})}{\sqrt{\det(g)}} \right) \right\},$$

which, after using the definition of the normal vector (3.22), simplifies to

$$(-\Delta_\Gamma \text{id}_\Gamma) \circ \mathbf{X} = \frac{-1}{\sqrt{\det(g)}} \{ \partial_{s_1} (\partial_{s_2} \mathbf{X} \times \boldsymbol{\nu}) - \partial_{s_2} (\partial_{s_1} \mathbf{X} \times \boldsymbol{\nu}) \}.$$

Expanding this further gives

$$(-\Delta_\Gamma \text{id}_\Gamma) \circ \mathbf{X} = \frac{-1}{\sqrt{\det(g)}} \{ (\partial_{s_1} \partial_{s_2} \mathbf{X} \times \boldsymbol{\nu}) + (\partial_{s_2} \mathbf{X} \times \partial_{s_1} \boldsymbol{\nu}) \\ - (\partial_{s_2} \partial_{s_1} \mathbf{X} \times \boldsymbol{\nu}) - (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \boldsymbol{\nu}) \},$$

followed by canceling and rearranging which proves (4.28):

$$(-\Delta_\Gamma \text{id}_\Gamma) \circ \mathbf{X} = \frac{(\partial_{s_1} \boldsymbol{\nu} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \boldsymbol{\nu})}{\sqrt{\det(g)}}.$$

Note that the vector field given by (4.28) is orthogonal to the tangent vectors $\partial_{s_1} \mathbf{X}$ and $\partial_{s_2} \mathbf{X}$; hence (4.28) is parallel to the normal vector $\boldsymbol{\nu}$. We use this to simplify (4.28) further:

$$(-\Delta_\Gamma \text{id}_\Gamma) \circ \mathbf{X} \cdot \boldsymbol{\nu} = \frac{(\partial_{s_1} \boldsymbol{\nu} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \boldsymbol{\nu})}{\sqrt{\det(g)}} \cdot \boldsymbol{\nu} \\ = \frac{1}{\det(g)} \{ (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}) \cdot (\partial_{s_1} \boldsymbol{\nu} \times \partial_{s_2} \mathbf{X}) \\ + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}) \cdot (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \boldsymbol{\nu}) \},$$

where we have used (3.22). By using the vector identity (A.3), we can transform this into

$$(-\Delta_\Gamma \text{id}_\Gamma) \circ \mathbf{X} \cdot \boldsymbol{\nu} = \frac{1}{\det(g)} \{ (\partial_{s_1} \mathbf{X} \cdot \partial_{s_1} \boldsymbol{\nu}) (\partial_{s_2} \mathbf{X} \cdot \partial_{s_2} \mathbf{X}) - (\partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \mathbf{X}) (\partial_{s_2} \mathbf{X} \cdot \partial_{s_1} \boldsymbol{\nu}) \\ + (\partial_{s_1} \mathbf{X} \cdot \partial_{s_1} \mathbf{X}) (\partial_{s_2} \mathbf{X} \cdot \partial_{s_2} \boldsymbol{\nu}) - (\partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \boldsymbol{\nu}) (\partial_{s_2} \mathbf{X} \cdot \partial_{s_1} \mathbf{X}) \},$$

and by definitions (3.2) and (3.36) this reduces to

$$(-\Delta_\Gamma \text{id}_\Gamma) \circ \mathbf{X} \cdot \boldsymbol{\nu} = \frac{-h_{11}g_{22} + g_{12}h_{12} - g_{11}h_{22} + h_{21}g_{21}}{\det(g)}.$$

Finally, using the formula for the inverse metric (3.13) gives

$$(-\Delta_\Gamma \text{id}_\Gamma) \circ \mathbf{X} \cdot \boldsymbol{\nu} = -(g^{11}h_{11} + g^{12}h_{12} + g^{22}h_{22} + g^{21}h_{21}),$$

which, by the equation for the summed curvature (3.60), proves the second line of (4.29):

$$[(-\Delta_\Gamma \text{id}_\Gamma) \cdot \nu] \circ \mathbf{X} = \kappa \circ \mathbf{X}.$$

Note that we have abused notation by using the identification $\nu \equiv \nu \circ \mathbf{X}$. Because we know that $(-\Delta_\Gamma \text{id}_\Gamma) \circ \mathbf{X}$ is parallel to ν , we have that the summed curvature vector (in terms of local coordinates) is given by the first equation in (4.29). \square

Remark 21. *The formula (4.29) is the analogue of the curvature formula for curves, i.e.,*

$$(-\Delta_\Sigma \text{id}_\Sigma) \circ \alpha = -\partial_s^2 \alpha = k \mathbf{n}, \quad (4.30)$$

where Σ is a curve parameterized by arc-length by $\alpha : I \rightarrow \mathbb{R}^3$, ∂_s is the derivative with respect to the arc-length, k is the curvature of Σ , and \mathbf{n} is the normal vector of Σ (recall section 3.2.4).

4.3.2 • Alternative Gaussian Curvature

This section derives another way of writing the Gaussian curvature that is similar in spirit to Proposition 12.

Proposition 13. *The Gaussian curvature κ_G (in local coordinates) can be written as*

$$\kappa_G = \nu \cdot \frac{\partial_{s_1} \nu \times \partial_{s_2} \nu}{\sqrt{\det(g)}}, \quad \kappa_G \nu = \frac{\partial_{s_1} \nu \times \partial_{s_2} \nu}{\sqrt{\det(g)}}. \quad (4.31)$$

Proof. By the definition of the normal vector (3.22) and the vector identity (A.3), we have

$$\begin{aligned} \nu \cdot \frac{\partial_{s_1} \nu \times \partial_{s_2} \nu}{\sqrt{\det(g)}} &= \frac{(\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}) \cdot (\partial_{s_1} \nu \times \partial_{s_2} \nu)}{\det(g)}, \\ &= \frac{(\partial_{s_1} \mathbf{X} \cdot \partial_{s_1} \nu)(\partial_{s_2} \mathbf{X} \cdot \partial_{s_2} \nu) - (\partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \nu)(\partial_{s_2} \mathbf{X} \cdot \partial_{s_1} \nu)}{\det(g)}. \end{aligned}$$

Then, by the definition of the second fundamental form (3.36),

$$\nu \cdot \frac{\partial_{s_1} \nu \times \partial_{s_2} \nu}{\sqrt{\det(g)}} = \frac{b_{11} b_{22} - b_{21} b_{12}}{\det(g)},$$

which is just (by (3.61))

$$\nu \cdot \frac{\partial_{s_1} \nu \times \partial_{s_2} \nu}{\sqrt{\det(g)}} = \frac{\det(h)}{\det(g)} =: \kappa_G.$$

This proves the first part of the assertion.

The second part follows by showing that $\partial_{s_k} \mathbf{X} \cdot (\partial_{s_1} \nu \times \partial_{s_2} \nu) = 0$ for $k = 1, 2$. First, note that by Proposition 12,

$$(\partial_{s_1} \mathbf{X} \times \partial_{s_2} \nu) = \kappa \nu \sqrt{\det(g)} - (\partial_{s_1} \nu \times \partial_{s_2} \mathbf{X}). \quad (4.32)$$

Thus, using vector identities and (4.32), we get

$$\begin{aligned}\partial_{s_1} \mathbf{X} \cdot (\partial_{s_1} \boldsymbol{\nu} \times \partial_{s_2} \boldsymbol{\nu}) &= -(\partial_{s_1} \mathbf{X} \times \partial_{s_2} \boldsymbol{\nu}) \cdot \partial_{s_1} \boldsymbol{\nu} \\ &= [(\partial_{s_1} \boldsymbol{\nu} \times \partial_{s_2} \mathbf{X}) - \kappa \boldsymbol{\nu} \sqrt{\det(g)}] \cdot \partial_{s_1} \boldsymbol{\nu} \\ &= \partial_{s_1} \boldsymbol{\nu} \cdot (\partial_{s_1} \boldsymbol{\nu} \times \partial_{s_2} \mathbf{X}) - (\partial_{s_1} \boldsymbol{\nu}) \cdot \boldsymbol{\nu} \kappa \sqrt{\det(g)} = 0,\end{aligned}$$

because $\mathbf{a} \cdot (\mathbf{a} \times \mathbf{b}) = 0$ for any \mathbf{a} , and $2(\partial_{s_1} \boldsymbol{\nu}) \cdot \boldsymbol{\nu} = \partial_{s_1}(\boldsymbol{\nu} \cdot \boldsymbol{\nu}) = 0$. Similarly,

$$\partial_{s_2} \mathbf{X} \cdot (\partial_{s_1} \boldsymbol{\nu} \times \partial_{s_2} \boldsymbol{\nu}) = 0. \quad \square$$

4.4 ■ Integration by Parts

In this section, we develop a basic calculus identity on surfaces: integration by parts. Note that in writing the integrals, we will often drop the measure dS (refer to section 1.6.3).

4.4.1 ■ Preliminary Formulas

Proposition 14. *Let Γ be a smooth, oriented surface with boundary $\partial\Gamma$. Let κ be the summed curvature (see (3.60)) of Γ . Then the following relation is true:*

$$\int_{\Gamma} \nabla_{\Gamma} \omega = \int_{\Gamma} \omega \kappa \boldsymbol{\nu} + \int_{\partial\Gamma} \omega (\boldsymbol{\tau} \times \boldsymbol{\nu}) \quad (4.33)$$

for all smooth scalar $\omega : \Gamma \rightarrow \mathbb{R}$, where $\boldsymbol{\nu}$ is the unit (oriented) normal vector of Γ and $\boldsymbol{\tau}$ is the unit tangent vector of $\partial\Gamma$ such that $\boldsymbol{\tau}$ is positively oriented with respect to $\boldsymbol{\nu}$. Note that $\boldsymbol{\tau} \times \boldsymbol{\nu}$ points out of $\partial\Gamma$. If Γ is a surface without boundary, then the term $\int_{\partial\Gamma}$ vanishes.

Proof. Let ω have compact support on Γ (see section 2.1.3) such that $\tilde{\omega} := \omega \circ \mathbf{X}$ has compact support in some open set U_i (recall that $\{U_i\}$ are reference domains). We start on the left side of (4.33) and do a change of variables, followed by plugging in the definition of the surface gradient operator (4.6):

$$\begin{aligned}\int_{\Gamma} \nabla_{\Gamma} \omega &= \int_{U_i} (\nabla_{\Gamma} \omega) \circ \mathbf{X} \sqrt{\det(g)} ds_1 ds_2 \\ &= \int_{U_i} \sqrt{\det(g)} \sum_{i,j=1}^2 g^{ij} \partial_{s_i} \tilde{\omega} \partial_{s_j} \mathbf{X} ds_1 ds_2.\end{aligned}$$

Writing out the sum gives

$$\begin{aligned}\int_{\Gamma} \nabla_{\Gamma} \omega &= \int_{U_i} \sqrt{\det(g)} [\partial_{s_1} \tilde{\omega} (g^{11} \partial_{s_1} \mathbf{X} + g^{12} \partial_{s_2} \mathbf{X}) \\ &\quad + \partial_{s_2} \tilde{\omega} (g^{21} \partial_{s_1} \mathbf{X} + g^{22} \partial_{s_2} \mathbf{X})] ds_1 ds_2,\end{aligned}$$

and after using the definition of the inverse metric (3.13) and rearranging slightly, we get

$$\begin{aligned}\int_{\Gamma} \nabla_{\Gamma} \omega &= \int_{U_i} \frac{1}{\sqrt{\det(g)}} \left[\partial_{s_1} \tilde{\omega} ((\partial_{s_2} \mathbf{X} \cdot \partial_{s_2} \mathbf{X}) \partial_{s_1} \mathbf{X} - (\partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \mathbf{X}) \partial_{s_2} \mathbf{X}) \right. \\ &\quad \left. + \partial_{s_2} \tilde{\omega} ((\partial_{s_1} \mathbf{X} \cdot \partial_{s_1} \mathbf{X}) \partial_{s_2} \mathbf{X} - (\partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \mathbf{X}) \partial_{s_1} \mathbf{X}) \right] ds_1 ds_2.\end{aligned}$$

Applying the vector identity (A.4) simplifies this to

$$\begin{aligned}\int_{\Gamma} \nabla_{\Gamma} \omega = & \int_{U_i} \frac{1}{\sqrt{\det(g)}} \left[\partial_{s_1} \tilde{\omega} (\partial_{s_2} \mathbf{X} \times (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X})) \right. \\ & \left. + \partial_{s_2} \tilde{\omega} (\partial_{s_1} \mathbf{X} \times (\partial_{s_2} \mathbf{X} \times \partial_{s_1} \mathbf{X})) \right] ds_1 ds_2,\end{aligned}$$

and using the definition of the normal vector ν (3.22) and making note of (A.2) reduces this further:

$$\int_{\Gamma} \nabla_{\Gamma} \omega = \int_{U_i} \left[\partial_{s_1} \tilde{\omega} (\partial_{s_2} \mathbf{X} \times \nu) - \partial_{s_2} \tilde{\omega} (\partial_{s_1} \mathbf{X} \times \nu) \right] ds_1 ds_2.$$

Next, use standard integration by parts on the two terms in the integrand,

$$\int_{\Gamma} \nabla_{\Gamma} \omega = \int_{U_i} \left[-\tilde{\omega} \partial_{s_1} (\partial_{s_2} \mathbf{X} \times \nu) + \tilde{\omega} \partial_{s_2} (\partial_{s_1} \mathbf{X} \times \nu) \right] ds_1 ds_2,$$

where there are no boundary terms because $\tilde{\omega}$ has compact support in U_i . After applying the product rule and rearranging, we get

$$\int_{\Gamma} \nabla_{\Gamma} \omega = \int_{U_i} \tilde{\omega} \frac{[(\partial_{s_1} \nu \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \nu)]}{\sqrt{\det(g)}} \sqrt{\det(g)} ds_1 ds_2.$$

Using (4.29) for the summed curvature vector in local coordinates gives

$$\int_{\Gamma} \nabla_{\Gamma} \omega = \int_{U_i} [\omega \circ \mathbf{X}] [(\kappa \nu) \circ \mathbf{X}] \sqrt{\det(g)} ds_1 ds_2.$$

Finally, we apply a change of variables back to the surface Γ ,

$$\int_{\Gamma} \nabla_{\Gamma} \omega = \int_{\Gamma} \omega \kappa \nu,$$

which holds for any ω with compact support on Γ .

Now, suppose $\omega \neq 0$ on $\partial\Gamma$. Specifically, assume $\tilde{\omega}$ has support in U_i and $\tilde{\omega} \neq 0$ on ∂U_i . Then the integration by parts step above yields

$$\begin{aligned}\int_{\Gamma} \nabla_{\Gamma} \omega = & \int_{U_i} \left[-\tilde{\omega} \partial_{s_1} (\partial_{s_2} \mathbf{X} \times \nu) + \tilde{\omega} \partial_{s_2} (\partial_{s_1} \mathbf{X} \times \nu) \right] ds_1 ds_2 \\ & + \underbrace{\int_{\partial U_i} \left[\tilde{\omega} (\partial_{s_2} \mathbf{X} \times \nu) q_1 - \tilde{\omega} (\partial_{s_1} \mathbf{X} \times \nu) q_2 \right] dt}_{=: BI},\end{aligned}$$

where dt is the arc-length measure on ∂U_i and $\mathbf{q} = (q_1, q_2)^T$ is the outward unit normal vector on ∂U_i . Let the oriented unit tangent vector of ∂U_i be $\mathbf{r} = \mathbf{q}^\perp$ (i.e., $(r_1, r_2)^T = (-q_2, q_1)^T$). Focusing on the boundary term, we have

$$BI = \int_{\partial U_i} \tilde{\omega} \left[(r_1 \partial_{s_1} \mathbf{X} \times \nu) + (r_2 \partial_{s_2} \mathbf{X} \times \nu) \right] dt = \int_{\partial U_i} \tilde{\omega} (r_1 \partial_{s_1} \mathbf{X} + r_2 \partial_{s_2} \mathbf{X}) \times \nu dt.$$

Next, introduce a parameterization $\mathbf{w}(t)$ of ∂U_i using the t arc-length variable. Thus, $\mathbf{r}(t) = d\mathbf{w}(t)/dt$. By the standard chain rule, we get

$$r_1 \partial_{s_1} \mathbf{X} + r_2 \partial_{s_2} \mathbf{X} = [(\mathbf{r}(t) \cdot \nabla_s) \mathbf{X}] \circ \mathbf{w}(t) = \frac{d}{dt} \mathbf{X}(\mathbf{w}(t)).$$

Now, consider a parameterization $\alpha(t)$ of $\partial\Gamma \cap \mathbf{X}(\partial U_i)$ also using the t arc-length variable. Clearly, $\alpha(t) = \mathbf{X}(\mathbf{w}(t))$, so then

$$\begin{aligned} BI &= \int_{\partial U_i} \tilde{\omega} \left(\frac{d}{dt} \alpha(t) \right) \times \nu dt = \int_{\partial U_i} \tilde{\omega} \left(\frac{1}{|\frac{d}{dt} \alpha(t)|} \frac{d}{dt} \alpha(t) \right) \times \nu \left| \frac{d}{dt} \alpha(t) \right| dt \\ &= \int_{\partial U_i} \tilde{\omega}(\tau \times \nu) \left| \frac{d}{dt} \alpha(t) \right| dt = \int_{\partial\Gamma} \omega(\tau \times \nu) d\alpha, \end{aligned}$$

where τ is the oriented tangent vector of $\partial\Gamma$, and $d\alpha$ is the differential arc-length variable on $\partial\Gamma$. Therefore, we obtain

$$\int_{\Gamma} \nabla_{\Gamma} \omega = \int_{\Gamma} \omega \kappa \nu + \int_{\partial\Gamma} \omega(\tau \times \nu),$$

which holds for any ω such that $\omega \circ \mathbf{X}$ has support in some U_i .

This formula is true for any scalar ω because it is always possible to decompose ω using a **partition of unity** (with respect to $\{U_i\}$) (see [25, p. 30] or [37]). This allows ω to be written as a sum of scalar functions ω_i , each of which has $\text{supp}(\omega_i \circ \mathbf{X}) = U_i$, and the above formula is true for each individual ω_i . The assertion then follows (for general ω) by adding up the integrals. \square

Proposition 14 can be extended to vector functions.

Proposition 15. *Under the same assumptions as in Proposition 14, we have the following relations:*

$$\int_{\Gamma} \nabla_{\Gamma} \varphi = \int_{\Gamma} \kappa \nu \otimes \varphi + \int_{\partial\Gamma} (\tau \times \nu) \otimes \varphi, \quad (4.34)$$

$$\int_{\Gamma} \nabla_{\Gamma} \cdot \varphi = \int_{\Gamma} \kappa \nu \cdot \varphi + \int_{\partial\Gamma} (\tau \times \nu) \cdot \varphi \quad (4.35)$$

for all smooth vector functions $\varphi : \Gamma \rightarrow \mathbb{R}^3$. See (A.13) for the definition of \otimes .

Proof. Equation (4.34) is obtained by applying (4.33) to each component of φ , and (4.35) comes from taking the trace of the matrix equation (4.34). \square

4.4.2 • Main Integration by Parts Formula

Proposition 16. *Assume the hypothesis of Proposition 14. Let φ and η be smooth scalar functions on Γ (i.e., $\varphi, \eta : \Gamma \rightarrow \mathbb{R}$). Then the following integration by parts formula is true:*

$$\int_{\Gamma} \nabla_{\Gamma} \varphi \cdot \nabla_{\Gamma} \eta = - \int_{\Gamma} \varphi \Delta_{\Gamma} \eta + \int_{\partial\Gamma} \varphi (\tau \times \nu) \cdot \nabla_{\Gamma} \eta. \quad (4.36)$$

Proof. Because $\nabla_{\Gamma} \cdot (\varphi \nabla_{\Gamma} \eta) = \nabla_{\Gamma} \varphi \cdot \nabla_{\Gamma} \eta + \varphi \Delta_{\Gamma} \eta$, we have

$$\int_{\Gamma} (\nabla_{\Gamma} \varphi \cdot \nabla_{\Gamma} \eta + \varphi \Delta_{\Gamma} \eta) = \int_{\Gamma} \nabla_{\Gamma} \cdot (\varphi \nabla_{\Gamma} \eta) = \int_{\Gamma} \kappa(\nu \cdot \nabla_{\Gamma} \eta) \varphi + \int_{\partial\Gamma} \varphi (\tau \times \nu) \cdot \nabla_{\Gamma} \eta,$$

where the last equality follows from (4.35) applied to the vector $\varphi \nabla_{\Gamma} \eta$. But $\nabla_{\Gamma} \eta$ is tangent to the surface Γ , so $\nu \cdot \nabla_{\Gamma} \eta = 0$. This proves the assertion. \square

Proposition 16 also extends to vector functions.

Proposition 17. Assume the hypothesis of Proposition 14. Let $\varphi = (\varphi_1, \varphi_2, \varphi_3)^T$ and $\eta = (\eta_1, \eta_2, \eta_3)^T$ be smooth vector functions on Γ (i.e., $\varphi, \eta : \Gamma \rightarrow \mathbb{R}^3$). Then (4.36) generalizes to

$$\int_{\Gamma} \nabla_{\Gamma} \varphi : \nabla_{\Gamma} \eta = - \int_{\Gamma} \varphi \cdot \Delta_{\Gamma} \eta + \int_{\partial\Gamma} \varphi \cdot \{[(\tau \times \nu) \cdot \nabla_{\Gamma}] \eta\}, \quad (4.37)$$

where the “:” is the matrix dot product in (A.11).

Proof. Clearly, we can apply (4.36) to the vector components

$$\int_{\Gamma} \nabla_{\Gamma} \varphi_k \cdot \nabla_{\Gamma} \eta_k = - \int_{\Gamma} \varphi_k \Delta_{\Gamma} \eta_k + \int_{\partial\Gamma} \varphi_k (\tau \times \nu) \cdot \nabla_{\Gamma} \eta_k$$

for $k = 1, 2, 3$. Then by summing over k , we get the assertion. \square

Proposition 18. Assume the hypothesis of Propositions 14 and 17. Moreover, suppose $\eta = (\nabla_{\Gamma} \text{id}_{\Gamma}) \eta$ (i.e., η is tangential only). Then,

$$\int_{\Gamma} \nabla_{\Gamma} \phi \cdot \eta = - \int_{\Gamma} \phi \nabla_{\Gamma} \cdot \eta + \int_{\partial\Gamma} (\tau \times \nu) \cdot (\phi \eta). \quad (4.38)$$

Proof. This is a straightforward application of the previous results:

$$\int_{\Gamma} (\nabla_{\Gamma} \phi \cdot \eta + \phi \nabla_{\Gamma} \cdot \eta) = \int_{\Gamma} \nabla_{\Gamma} \cdot (\phi \eta) = \int_{\Gamma} \phi \underbrace{\eta \cdot \nu}_{=0} \kappa + \int_{\partial\Gamma} (\tau \times \nu) \cdot (\phi \eta). \quad \square$$

4.5 • Other Identities

We derive some formulas that will be useful later in this book.

4.5.1 • Preliminary Calculations

Proposition 19. Let $\varphi = (\varphi_1, \varphi_2, \varphi_3)^T$ and $\eta = (\eta_1, \eta_2, \eta_3)^T$ be smooth vector functions on Γ (i.e., $\varphi, \eta : \Gamma \rightarrow \mathbb{R}^3$), and let $\tilde{\varphi} = \varphi \circ X$, $\tilde{\eta} = \eta \circ X$ be the vector functions in local coordinates. Then the following is true:

$$\{(\nabla_{\Gamma} \varphi) \circ X\} : \{(\nabla_{\Gamma} \eta) \circ X\} = \sum_{i,j=1}^2 g^{ij} \partial_{s_i} \tilde{\varphi} \cdot \partial_{s_j} \tilde{\eta}. \quad (4.39)$$

Proof. By equations (4.7) and (4.8), we have

$$\begin{aligned} \{(\nabla_{\Gamma}\varphi)\circ\mathbf{X}\} : \{(\nabla_{\Gamma}\eta)\circ\mathbf{X}\} &= \sum_{k=1}^3 \{(\nabla_{\Gamma}\varphi_k)\circ\mathbf{X}\} \cdot \{(\nabla_{\Gamma}\eta_k)\circ\mathbf{X}\} \\ &= \sum_{k=1}^3 \sum_{i,j=1}^2 \sum_{p,q=1}^2 g^{ij} g^{pq} \partial_{s_i} \tilde{\varphi}_k \partial_{s_p} \tilde{\eta}_k \partial_{s_j} \mathbf{X} \cdot \partial_{s_q} \mathbf{X} \\ &= \sum_{i,j=1}^2 \sum_{p,q=1}^2 g^{ij} g^{pq} g_{jq} \sum_{k=1}^3 \partial_{s_i} \tilde{\varphi}_k \partial_{s_p} \tilde{\eta}_k, \end{aligned}$$

where we have used the definition of the first fundamental form (3.2) in the last step. By properties (3.14) and (3.15), the above equation simplifies to (4.39). \square

4.5.2 • Surface Gradient of the Normal Vector

Shape Operator

In section 3.2.5, we defined the shape operator $S = -h g^{-1}$. This is actually an intrinsic definition, because one could define g and h by abstractly setting the entries g_{ij} and h_{ij} without invoking a parameterization \mathbf{X} . It is also possible to define the shape operator in an extrinsic way.

Definition 14. Let ν be the unit normal vector field of an orientable surface $\Gamma \subset \mathbb{R}^3$. We call $\nabla_{\Gamma}\nu$ the (extrinsic) shape operator of Γ , in terms of coordinates in Γ .

In the next section, we relate $\nabla_{\Gamma}\nu$ to the intrinsic shape operator S . But first, we must show that $\nabla_{\Gamma}\nu$ is a **symmetric** matrix.

Proposition 20. Assume the hypothesis of Definition 14. Then, $\nabla_{\Gamma}\nu$ is a symmetric 3×3 matrix.

Proof. It is clear that $\nabla_{\Gamma}\nu$ is a 3×3 matrix. Let $\mathbf{M} = \nabla_{\Gamma}\nu - (\nabla_{\Gamma}\nu)^T$. We will show that $\mathbf{M} = \mathbf{0}$ by showing $\mathbf{M}\mathbf{v} = \mathbf{0}$ for three linearly independent choices of the vector \mathbf{v} . Let $\mathbf{e}_1 = (1, 0)^T$ and $\mathbf{e}_2 = (0, 1)^T$. Since Γ is a regular surface, we know $\partial_{s_1} \mathbf{X} = (\nabla_s \mathbf{X})\mathbf{e}_1$ and $\partial_{s_2} \mathbf{X} = (\nabla_s \mathbf{X})\mathbf{e}_2$ are linearly independent vectors in the tangent space of Γ . Thus, $\{(\nabla_s \mathbf{X})\mathbf{e}_1, (\nabla_s \mathbf{X})\mathbf{e}_2, \tilde{\nu}\}$, where $\tilde{\nu} = \nu \circ \mathbf{X}$, forms a linearly independent set of vectors in \mathbb{R}^3 .

Using local coordinates, and (4.8), we have $(\nabla_{\Gamma}\nu)\circ\mathbf{X} = (\nabla_s \tilde{\nu})g^{-1}(\nabla_s \mathbf{X})^T$. Computing the matrix-vector product with $(\nabla_s \mathbf{X})\mathbf{e}_k$ (for $k = 1, 2$), we get

$$\begin{aligned} [(\nabla_{\Gamma}\nu)\circ\mathbf{X}] (\nabla_s \mathbf{X})\mathbf{e}_k &= (\nabla_s \tilde{\nu})g^{-1}(\nabla_s \mathbf{X})^T (\nabla_s \mathbf{X})\mathbf{e}_k \\ &= (\nabla_s \tilde{\nu})g^{-1}g \mathbf{e}_k = (\nabla_s \tilde{\nu})\mathbf{e}_k = \partial_{s_k} \tilde{\nu}, \end{aligned}$$

where we have used $g = (\nabla_s \mathbf{X})^T (\nabla_s \mathbf{X})$. We also have

$$[(\nabla_{\Gamma}\nu)\circ\mathbf{X}] \tilde{\nu} = (\nabla_s \tilde{\nu})g^{-1}(\nabla_s \mathbf{X})^T \tilde{\nu} = \mathbf{0},$$

because $\tilde{\nu}$ is orthogonal to $\partial_{s_1} \mathbf{X}$ and $\partial_{s_2} \mathbf{X}$. Next, since g^{-1} is symmetric, we obtain

$$\begin{aligned} [(\nabla_\Gamma \nu) \circ \mathbf{X}]^T (\nabla_s \mathbf{X}) \mathbf{e}_k &= (\nabla_s \mathbf{X}) g^{-1} (\nabla_s \tilde{\nu})^T (\nabla_s \mathbf{X}) \mathbf{e}_k \\ &= (\nabla_s \mathbf{X}) g^{-1} (-h) \mathbf{e}_k = (\nabla_s \mathbf{X}) g^{-1} (-h)^T \mathbf{e}_k \\ &= (\nabla_s \mathbf{X}) g^{-1} (\nabla_s \mathbf{X})^T (\nabla_s \tilde{\nu}) \mathbf{e}_k \end{aligned} \quad (4.40)$$

where we have used (3.40) (i.e., $h = -(\nabla_s \tilde{\nu})^T (\nabla_s \mathbf{X})$) and the fact that h is symmetric. Combining (4.40) with (4.15), we get

$$\begin{aligned} [(\nabla_\Gamma \nu) \circ \mathbf{X}]^T (\nabla_s \mathbf{X}) \mathbf{e}_k &= [(\nabla_\Gamma \text{id}_\Gamma) \circ \mathbf{X}] (\nabla_s \tilde{\nu}) \mathbf{e}_k = [(\nabla_\Gamma \text{id}_\Gamma) \circ \mathbf{X}] \partial_{s_k} \tilde{\nu} \\ &= [\mathbf{I} - \nu \otimes \nu] \partial_{s_k} \tilde{\nu} = \partial_{s_k} \tilde{\nu} - (\tilde{\nu} \cdot \partial_{s_k} \tilde{\nu}) \tilde{\nu} \\ &= \partial_{s_k} \tilde{\nu}, \end{aligned} \quad (4.41)$$

where we have used (4.16) and the fact that $\tilde{\nu} \cdot \tilde{\nu} = 1$ implies $\tilde{\nu} \cdot \partial_{s_k} \tilde{\nu} = 0$. Similarly, one can show that $[(\nabla_\Gamma \nu) \circ \mathbf{X}]^T \tilde{\nu} = 0$. Therefore,

$$[\mathbf{M} \circ \mathbf{X}] (\nabla_s \mathbf{X}) \mathbf{e}_1 = 0, \quad [\mathbf{M} \circ \mathbf{X}] (\nabla_s \mathbf{X}) \mathbf{e}_2 = 0, \quad [\mathbf{M} \circ \mathbf{X}] \tilde{\nu} = 0,$$

so $\mathbf{M} = 0$ and we obtain the assertion. \square

Relation to Curvature

From the previous discussion, we have that $(\nabla_\Gamma \nu) \nu = 0$. So the extrinsic shape operator is a symmetric operator that acts on the tangent space of Γ . Moreover, the eigenvalues and eigenvectors of $\nabla_\Gamma \nu$ are related to the principal curvatures and principal directions of Γ (see Definition 6 of Chapter 3), which we now prove.

Proposition 21. *Assume the hypothesis of Definition 14. Let κ_1, κ_2 be the principal curvatures of Γ , and let $\mathbf{d}_1, \mathbf{d}_2$ be the corresponding principal directions. Then $\{0, \kappa_1, \kappa_2\}$ are the eigenvalues of $\nabla_\Gamma \nu$, i.e.,*

$$(\nabla_\Gamma \nu) \nu = 0, \quad (\nabla_\Gamma \nu) \mathbf{d}_1 = \kappa_1 \mathbf{d}_1, \quad (\nabla_\Gamma \nu) \mathbf{d}_2 = \kappa_2 \mathbf{d}_2. \quad (4.42)$$

Proof. We already know that $(\nabla_\Gamma \nu) \nu = 0$. From section 3.2.4 (see Definition 6), and recalling (3.52) and (3.54), we have that

$$-h \mathbf{q}_1 = \kappa_1 g \mathbf{q}_1, \quad -h \mathbf{q}_2 = \kappa_2 g \mathbf{q}_2, \quad \mathbf{d}_1 = (\nabla_s \mathbf{X}) \mathbf{q}_1, \quad \mathbf{d}_2 = (\nabla_s \mathbf{X}) \mathbf{q}_2, \quad (4.43)$$

where g and h are the first and second fundamental form matrices. Next, note $h = -(\nabla_s \tilde{\nu})^T (\nabla_s \mathbf{X})$, and compute the following for $i = 1, 2$:

$$\begin{aligned} [(\nabla_\Gamma \nu) \circ \mathbf{X}] \mathbf{d}_i &= [(\nabla_\Gamma \nu) \circ \mathbf{X}]^T \mathbf{d}_i = (\nabla_s \mathbf{X}) g^{-1} (\nabla_s \tilde{\nu})^T \mathbf{d}_i \\ &= (\nabla_s \mathbf{X}) g^{-1} (\nabla_s \tilde{\nu})^T (\nabla_s \mathbf{X}) \mathbf{q}_i \\ &= (\nabla_s \mathbf{X}) g^{-1} (-h) \mathbf{q}_i = \kappa_i (\nabla_s \mathbf{X}) g^{-1} g \mathbf{q}_i \\ &= \kappa_i (\nabla_s \mathbf{X}) \mathbf{q}_i = \kappa_i \mathbf{d}_i, \end{aligned} \quad (4.44)$$

which gives the assertion. \square

We thus obtain a useful formula relating the principal curvatures to the surface gradient of the normal vector.

Proposition 22. Assume the hypothesis of Definition 14. Then the following is true:

$$|\nabla_{\Gamma}\nu|^2 = \kappa^2 - 2\kappa_G = \kappa_1^2 + \kappa_2^2. \quad (4.45)$$

In the case of a curve $\Sigma \subset \mathbb{R}^2$ with unit normal ν , $\kappa_G = 0$ and

$$|\nabla_{\Sigma}\nu| = |\kappa|. \quad (4.46)$$

Proof. From (A.11) and (A.12), we have

$$|\nabla_{\Gamma}\nu|^2 = (\nabla_{\Gamma}\nu) : (\nabla_{\Gamma}\nu) = \text{trace}((\nabla_{\Gamma}\nu)^T(\nabla_{\Gamma}\nu)) = \text{trace}((\nabla_{\Gamma}\nu)(\nabla_{\Gamma}\nu)) = \sum_{i=1}^3 \lambda_i^2,$$

where λ_i are the eigenvalues of $\nabla_{\Gamma}\nu$ [97, 103]. Together with Proposition 21, we obtain the result. \square

Before continuing, we make note of a modified version of (4.46) for the tangent vector. Suppose Σ is a 1-D curve in \mathbb{R}^2 , and let $\tau = (t_1, t_2)^T$ be the unit tangent vector with components t_1, t_2 . Then the outward pointing normal vector $\nu = (n_1, n_2)^T$ is related to the tangent vector by a 90 degree rotation:

$$(n_1, n_2)^T = (t_2, -t_1)^T.$$

Then, by (4.46), we have the following identity:

$$|\nabla_{\Sigma}\tau|^2 = |\nabla_{\Sigma}t_1|^2 + |\nabla_{\Sigma}t_2|^2 = |\nabla_{\Sigma}n_2|^2 + |\nabla_{\Sigma}n_1|^2 = |\nabla_{\Sigma}\nu|^2 = \kappa^2.$$

Hence,

$$|\nabla_{\Sigma}\tau| = \kappa. \quad (4.47)$$

4.5.3 ■ Surface Divergence of the Normal Vector

Here is another relation between the curvature and the normal vector.

Proposition 23. Assume the hypothesis of Definition 14. Then the following is true:

$$\nabla_{\Gamma} \cdot \nu = \kappa. \quad (4.48)$$

Proof. In local coordinates, by (4.20), (3.36), and the definition of summed curvature (3.60), we have

$$(\nabla_{\Gamma} \cdot \nu) \circ \mathbf{X} = \sum_{i,j=1}^2 g^{ij} \partial_{s_i} \nu \cdot \partial_{s_j} \mathbf{X} = - \sum_{i,j=1}^2 g^{ij} h_{ij} = \kappa \circ \mathbf{X}. \quad \square$$

4.5.4 ■ Surface Laplacian of the Normal Vector

Here is a relation between the second fundamental form and the normal vector.

Proposition 24. Let ν be the unit normal vector of Γ . Then the following is true:

$$\Delta_{\Gamma}\nu = -|\nabla_{\Gamma}\nu|^2\nu + (\nabla_{\Gamma}\kappa)^T. \quad (4.49)$$

Proof. By (4.27), we have

$$(\Delta_\Gamma \boldsymbol{\nu}) \circ \mathbf{X} = \frac{1}{\sqrt{\det(g)}} \left\{ -\partial_{s_1} \left[\frac{[(\partial_{s_1} \boldsymbol{\nu} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \boldsymbol{\nu})]}{\sqrt{\det(g)}} \right] \times \partial_{s_2} \mathbf{X} \right. \\ - \partial_{s_1} \mathbf{X} \times \partial_{s_2} \left[\frac{[(\partial_{s_1} \boldsymbol{\nu} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \boldsymbol{\nu})]}{\sqrt{\det(g)}} \right] \\ + \partial_{s_1} \left[\frac{[(\partial_{s_1} \boldsymbol{\nu} \cdot \partial_{s_2} \mathbf{X}) - (\partial_{s_2} \boldsymbol{\nu} \cdot \partial_{s_1} \mathbf{X})]}{\sqrt{\det(g)}} \right] \partial_{s_2} \mathbf{X} \\ - \partial_{s_2} \left[\frac{[(\partial_{s_1} \boldsymbol{\nu} \cdot \partial_{s_2} \mathbf{X}) - (\partial_{s_2} \boldsymbol{\nu} \cdot \partial_{s_1} \mathbf{X})]}{\sqrt{\det(g)}} \right] \partial_{s_1} \mathbf{X} \\ \left. + 2(\partial_{s_1} \boldsymbol{\nu} \times \partial_{s_2} \boldsymbol{\nu}) \right\}$$

for all (s_1, s_2) in any reference domain U_i . Next, plug in (3.36) and use Proposition 12 to get

$$(\Delta_\Gamma \boldsymbol{\nu}) \circ \mathbf{X} = \frac{1}{\sqrt{\det(g)}} \left\{ -\partial_{s_1}(\kappa \boldsymbol{\nu}) \times \partial_{s_2} \mathbf{X} - \partial_{s_1} \mathbf{X} \times \partial_{s_2}(\kappa \boldsymbol{\nu}) + 2(\partial_{s_1} \boldsymbol{\nu} \times \partial_{s_2} \boldsymbol{\nu}) \right. \\ \left. + \partial_{s_1} \left[\frac{-b_{12} + b_{21}}{\sqrt{\det(g)}} \right] \partial_{s_2} \mathbf{X} - \partial_{s_2} \left[\frac{-b_{12} + b_{21}}{\sqrt{\det(g)}} \right] \partial_{s_1} \mathbf{X} \right\},$$

and use the symmetry of b :

$$(\Delta_\Gamma \boldsymbol{\nu}) \circ \mathbf{X} = \frac{1}{\sqrt{\det(g)}} \left[-\partial_{s_1}(\kappa \boldsymbol{\nu}) \times \partial_{s_2} \mathbf{X} - \partial_{s_1} \mathbf{X} \times \partial_{s_2}(\kappa \boldsymbol{\nu}) + 2(\partial_{s_1} \boldsymbol{\nu} \times \partial_{s_2} \boldsymbol{\nu}) \right].$$

By (4.31), this becomes

$$(\Delta_\Gamma \boldsymbol{\nu}) \circ \mathbf{X} = 2\kappa_G \boldsymbol{\nu} + \frac{1}{\sqrt{\det(g)}} \left[-\partial_{s_1}(\kappa \boldsymbol{\nu}) \times \partial_{s_2} \mathbf{X} - \partial_{s_1} \mathbf{X} \times \partial_{s_2}(\kappa \boldsymbol{\nu}) \right].$$

Expanding the derivatives leads to:

$$(\Delta_\Gamma \boldsymbol{\nu}) \circ \mathbf{X} = 2\kappa_G \boldsymbol{\nu} - \frac{1}{\sqrt{\det(g)}} \left[(\partial_{s_1} \kappa)(\boldsymbol{\nu} \times \partial_{s_2} \mathbf{X}) + \kappa(\partial_{s_1} \boldsymbol{\nu} \times \partial_{s_2} \mathbf{X}) \right. \\ \left. + (\partial_{s_2} \kappa)(\partial_{s_1} \mathbf{X} \times \boldsymbol{\nu}) + \kappa(\partial_{s_1} \mathbf{X} \times \partial_{s_2} \boldsymbol{\nu}) \right] \\ = 2\kappa_G \boldsymbol{\nu} - \kappa \left[\frac{(\partial_{s_1} \boldsymbol{\nu} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \boldsymbol{\nu})}{\sqrt{\det(g)}} \right] \\ - \frac{1}{\sqrt{\det(g)}} \left[(\partial_{s_1} \kappa)(\boldsymbol{\nu} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_2} \kappa)(\partial_{s_1} \mathbf{X} \times \boldsymbol{\nu}) \right] \\ = 2\kappa_G \boldsymbol{\nu} - \kappa(\boldsymbol{\nu} \boldsymbol{\nu}) - \frac{1}{\sqrt{\det(g)}} \left[(\partial_{s_1} \kappa)(\boldsymbol{\nu} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_2} \kappa)(\partial_{s_1} \mathbf{X} \times \boldsymbol{\nu}) \right].$$

Next, plug in $\nu = (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}) / \sqrt{\det(g)}$, and use (4.45) and (A.4):

$$\begin{aligned} (\Delta_\Gamma \nu) \circ \mathbf{X} &= -(\kappa^2 - 2\kappa_G)\nu + \frac{1}{\det(g)} \left[(\partial_{s_1} \kappa)(\partial_{s_2} \mathbf{X} \times (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X})) \right. \\ &\quad \left. - (\partial_{s_2} \kappa)(\partial_{s_1} \mathbf{X} \times (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X})) \right] \\ &= -|\nabla_\Gamma \nu|^2 \nu + \frac{1}{\det(g)} \left[(\partial_{s_1} \kappa)(\partial_{s_2} \mathbf{X} \cdot \partial_{s_2} \mathbf{X}) \partial_{s_1} \mathbf{X} - (\partial_{s_1} \kappa)(\partial_{s_2} \mathbf{X} \cdot \partial_{s_1} \mathbf{X}) \partial_{s_2} \mathbf{X} \right. \\ &\quad \left. - (\partial_{s_2} \kappa)(\partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \mathbf{X}) \partial_{s_1} \mathbf{X} + (\partial_{s_2} \kappa)(\partial_{s_1} \mathbf{X} \cdot \partial_{s_1} \mathbf{X}) \partial_{s_2} \mathbf{X} \right] \\ &= -|\nabla_\Gamma \nu|^2 \nu + \frac{1}{\det(g)} \left[(\partial_{s_1} \kappa) g_{22} \partial_{s_1} \mathbf{X} - (\partial_{s_1} \kappa) g_{21} \partial_{s_2} \mathbf{X} \right. \\ &\quad \left. - (\partial_{s_2} \kappa) g_{12} \partial_{s_1} \mathbf{X} + (\partial_{s_2} \kappa) g_{11} \partial_{s_2} \mathbf{X} \right]. \end{aligned}$$

Then, applying (3.13) for the inverse metric g^{-1} (and using its symmetry), we get

$$\begin{aligned} (\Delta_\Gamma \nu) \circ \mathbf{X} &= -|\nabla_\Gamma \nu|^2 \nu + \left[(\partial_{s_1} \kappa) g^{11} \partial_{s_1} \mathbf{X} + (\partial_{s_1} \kappa) g^{12} \partial_{s_2} \mathbf{X} \right. \\ &\quad \left. + (\partial_{s_2} \kappa) g^{21} \partial_{s_1} \mathbf{X} + (\partial_{s_2} \kappa) g^{22} \partial_{s_2} \mathbf{X} \right] \\ &= -|\nabla_\Gamma \nu|^2 \nu + \sum_{i,j=1}^2 g^{ij} (\partial_{s_i} \kappa) \partial_{s_j} \mathbf{X}, \end{aligned}$$

and using (4.6) gives the assertion. \square

4.6 ■ PDEs on Surfaces

So far, this chapter has been concerned with defining and building up the tools of calculus on surfaces. Now that we have the concept of differentiation and integration on surfaces under our belts, along with some useful identities such as integration by parts, what can we do with them?

A central goal of calculus is the study of differential equations. In fact, students often encounter a course on ordinary differential equations (ODEs) after taking the standard freshman calculus sequence. Most multivariable calculus courses give some small exposure to partial differential equations (PDEs), such as Laplace's equation, the heat equation, and the wave equation.

Continuing this progression, we can define “surface versions” of these equations using the surface calculus tools we have developed. The following sections illustrate this for the heat equation. Since PDE theory is outside our scope of study (our goal is to keep the prerequisites of this book to a minimum), our discussion here will be brief. However, we hope it gives the reader some food for thought on how these tools can be used.

4.6.1 ■ The Heat Equation in Euclidean Space

Let Ω be a volumetric domain in \mathbb{R}^3 , e.g., a rectangular block of material. The standard heat equation, on Ω , takes the form

$$\begin{aligned} \partial_t u(\mathbf{x}, t) - \Delta u(\mathbf{x}, t) &= 0 \quad \text{for all } \mathbf{x} \in \Omega \quad \text{and } t > 0, \\ u(\mathbf{x}, t) &= g(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \partial\Omega \quad \text{and } t > 0, \\ u(\mathbf{x}, 0) &= u_0(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \Omega, \end{aligned} \tag{4.50}$$

where $\mathbf{x} = (x, y, z)^T$, $\Delta = \partial_x^2 + \partial_y^2 + \partial_z^2$, $u(\cdot, t) : \Omega \rightarrow \mathbb{R}$ is the temperature field, $t \geq 0$ is time, $u_0 : \Omega \rightarrow \mathbb{R}$ is the initial temperature, and $g : \partial\Omega \rightarrow \mathbb{R}$ is the boundary condition (i.e., g is the temperature on $\partial\Omega$ that is given and held fixed) [37, 112]. The heat equation models how an initial temperature distribution evolves in time, i.e., how the heat energy redistributes itself over Ω .

Remark 22. *In physics, there are always material parameters to consider. Here, we take a simple approach and set all constants to 1 (unitless).*

Given a specific domain Ω and data u_0 , g , the goal is usually to find a solution $u(\mathbf{x}, t)$ of (4.50). In other words, find a function u such that $\partial_t u - \Delta u = 0$, $u = g$ on $\partial\Omega$, and $u(\cdot, 0) = u_0(\cdot)$. PDE theory is quite vast; the novice should see [112] for a readable introduction; the first three chapters of [37] are also an option.

For illustration, let us derive an identity that connects the heat equation to the law of conservation of energy. The total thermal energy in Ω is given by $\int_{\Omega} u$, where a constant of proportionality is usually included to make the units work out (remember, we take all constants equal to 1). Differentiating with respect to time, and assuming u is sufficiently smooth in both variables \mathbf{x} and t , we have

$$\partial_t \int_{\Omega} u = \int_{\Omega} \partial_t u = \int_{\Omega} \Delta u = \int_{\Omega} \nabla \cdot \nabla u = \int_{\partial\Omega} \boldsymbol{\nu} \cdot \nabla u, \quad (4.51)$$

where $\boldsymbol{\nu}$ is the outer unit normal of $\partial\Omega$. Thus, the rate of change of the total thermal energy is equal to the total flux of heat energy through $\partial\Omega$ (i.e., ∇u is the heat flux vector). In other words, heat energy is being exchanged between Ω and the “outside” through the boundary $\partial\Omega$.

If we changed the boundary condition $u = g$ in (4.50) to $\boldsymbol{\nu} \cdot \nabla u = 0$ on $\partial\Omega$, then $\int_{\Omega} u$ equals a constant for all time. So, the total thermal energy is conserved for all time. This alternative boundary condition corresponds to having an “insulating layer” surrounding Ω that prevents any thermal energy from entering or escaping.

4.6.2 ■ The Heat Equation on a Curved Surface

Now let us consider the heat equation on a closed regular surface $\Gamma \subset \mathbb{R}^3$, i.e., let u solve

$$\begin{aligned} \partial_t u(\mathbf{x}, t) - \Delta_{\Gamma} u(\mathbf{x}, t) &= 0 \quad \text{for all } \mathbf{x} \in \Gamma \quad \text{and } t > 0, \\ u(\mathbf{x}, 0) &= u_0(\mathbf{x}) \quad \text{for all } \mathbf{x} \in \Gamma, \end{aligned} \quad (4.52)$$

where Δ_{Γ} is the surface Laplacian, $u(\cdot, t) : \Gamma \rightarrow \mathbb{R}$ is the temperature field, $t \geq 0$ is time, and $u_0 : \Omega \rightarrow \mathbb{R}$ is the initial temperature. Note that there is no boundary condition here because we have assumed that Γ is closed (i.e., Γ has no boundary).

Deriving the law of conservation of energy is similar to (4.51). Differentiating with respect to time gives

$$\partial_t \int_{\Gamma} u = \int_{\Gamma} \partial_t u = \int_{\Gamma} \Delta_{\Gamma} u = \int_{\Gamma} \nabla_{\Gamma} \cdot \nabla_{\Gamma} u = \int_{\Gamma} \kappa \boldsymbol{\nu} \cdot \nabla_{\Gamma} u + \int_{\partial\Gamma} (\boldsymbol{\tau} \times \boldsymbol{\nu}) \cdot \nabla_{\Gamma} u = 0, \quad (4.53)$$

where we have used (4.35). The last equality follows because $\boldsymbol{\nu} \cdot \nabla_{\Gamma} u = 0$ (because $\nabla_{\Gamma} u$ lies in the tangent space of Γ), and $\partial\Gamma = \emptyset$. Hence, the total thermal energy is conserved (stays constant in time). This is because there is nowhere for the energy to go. The PDE in

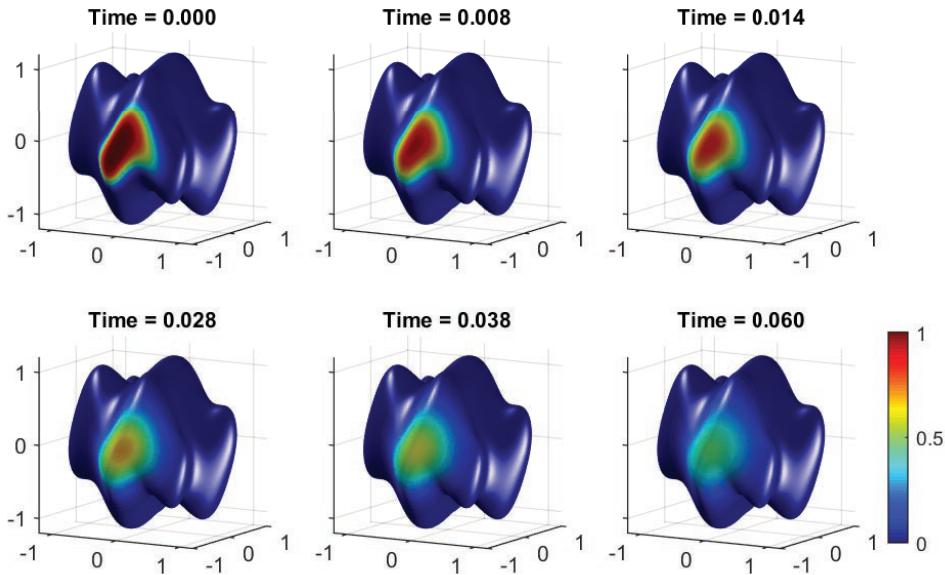


Figure 4.3. Illustration of the heat equation on a closed surface Γ . The solution u of (4.52) represents temperature and is denoted by color at each point on Γ . The time t is indicated above each plot. The temperature starts out concentrated in a small region and eventually diffuses over the entire surface. As $t \rightarrow \infty$, u converges to a constant over Γ .

(4.52) is posed entirely in the tangent space of Γ ; the temperature u is completely unaware of the “outside” world. In Figure 4.3, we show the solution of (4.52) on a generic closed surface Γ at several instances of time.

4.6.3 ■ Conclusion

The numerical solution of PDEs, either in flat or curved domains, requires a good bit of background knowledge, such as PDE theory, real analysis, numerical analysis and methods, linear algebra, etc. Hence, the path to mastering the numerical solution of PDEs can take many turns. One possible path, after having read this book, is given in the following (in chronological order).

- Get some exposure to analysis. A particularly good option is [64], which has a solution manual.
- Take an introductory course on PDEs, or find a good self-study book (one good option is [112]). Alternatively, if your background in analysis and linear algebra/vector spaces is sufficient, the first three chapters of [37] are quite good.
- Learn some basic numerical analysis. A very readable introduction is [95]. Another option is [86], but it is a bit outdated. A more extensive and modern presentation can be found in [85] (it also includes some MATLAB code). A graduate level text is [96]. Other options are [49, 99].
- Some programming experience, especially with an eye toward linear algebra, is a must. An easy-to-read introduction to numerical linear algebra is [103]. As for programming, there are too many options to list, so we refrain from making a specific recommendation.

- Finally, learn how to solve differential equations numerically. There are many options here. Some basic introductory texts are [36, 72, 92]. But there are many methodologies for solving PDEs *on surfaces*. Further investigation is needed before making an informed choice on the best method for a particular problem.

Depending on your mathematical background, you may want to modify this list accordingly.

Chapter 5

Shape Differential Calculus

This is the point where differential geometry meets the calculus of variations. Standard references for this material can be found in [23, 84, 93]. The reader may want to simply skim the results, skip most of the gory details, and go to the next chapter.

5.1 • Introduction

There are many functions and functionals that depend on the shape of a set or domain. One example is the volume functional,

$$V(\Omega) = \int_{\Omega} 1, \quad \text{where } \Omega \subset \mathbb{R}^3, \quad (5.1)$$

and another is the surface area (or perimeter) functional,

$$A(\Omega) = \int_{\partial\Omega} 1. \quad (5.2)$$

These kinds of quantities appear in a variety of optimization problems (along with many other functionals). In other words, you may want to minimize a cost functional (that depends on shape) over an admissible set of domains. This type of optimization is called *shape optimization*.

Therefore, solving these optimization problems requires computing the *sensitivity* of these quantities to changes (deformations) of the domain. This gives information on how to locally perturb a domain so as to lower its associated cost (recall the example from section 1.2.3). With this, one can build a gradient-based optimization method to solve the minimization problem. This chapter develops the tools needed for computing these *shape sensitivities*.

5.2 • General Framework

5.2.1 • Preliminaries

Assumption 3. Throughout this chapter, we assume $\Omega \subset \mathbb{R}^3$ is an open set with positive volume and smooth boundary $\partial\Omega$ (i.e., $\partial\Omega$ is a regular surface). Moreover, we assume that Γ is another regular surface, with positive area, contained in Ω . If desired, we can choose $\Gamma \equiv \partial\Omega$. Note that these assumptions can be weakened at the expense of more technical details.

For the sake of generality, let $\mathcal{E}(\cdot)$ denote a generic shape functional of Ω ,

$$\mathcal{E}(\Omega) = \int_{\Omega} f, \quad (5.3)$$

where $f : \Omega \rightarrow \mathbb{R}$ may also depend on Ω in an indirect way, i.e., $f \equiv f(\Omega, \mathbf{x})$; see the introductory example in section 1.2. For example, f may be the solution of a PDE defined on Ω . Obviously, changing Ω changes f . We also consider a shape functional of Γ ,

$$\mathcal{J}(\Gamma) = \int_{\Gamma} f, \quad \mathcal{B}(\Gamma) = \int_{\Gamma} g, \quad (5.4)$$

where $g : \Gamma \rightarrow \mathbb{R}$ may also depend on Γ in an indirect way, i.e., $g \equiv g(\Gamma, \mathbf{x})$ (i.e., g may be the solution of a PDE on Γ). Hence, an optimization problem for these functionals could be stated as

$$\Omega^* = \arg \min_{\Omega} \mathcal{E}(\Omega), \quad \Gamma^* = \arg \min_{\Gamma} \mathcal{J}(\Gamma), \quad \Gamma^* = \arg \min_{\Gamma} \mathcal{B}(\Gamma). \quad (5.5)$$

Remark 23. Solving (5.5) can be accomplished via an L^2 -gradient flow [3, 13, 28]. This is basically a gradient descent method that, for example, deforms the domain Ω in a direction that minimizes the cost \mathcal{E} (see section 6.3). The “gradient” in this case is called the **shape gradient**.

5.2.2 • Domain Perturbations

Computing derivatives (sensitivities) involves *perturbing* certain quantities. In the case of shape derivatives, we must perturb the domain and look at how the functionals (and functions) are correspondingly perturbed.

Diffeomorphism

The concept of a differentiable map “between domains” is useful for perturbing domains. First, we give an explanation for maps between surfaces. Let Γ_1 and Γ_2 be regular surfaces and suppose that $\varphi : V_1 \rightarrow \Gamma_2$ is a map from an open subset $V_1 \subset \Gamma_1$ to the other surface $\varphi(V_1) \subset \Gamma_2$.

We say that φ is *differentiable at P in V_1* if, given parameterizations

$$\mathbf{X}_1 : U_1 \rightarrow \Gamma_1, \quad \mathbf{X}_2 : U_2 \rightarrow \Gamma_2, \quad \text{with } U_1, U_2 \subset \mathbb{R}^2 \quad (5.6)$$

with P in $\mathbf{X}_1(U_1)$ and $\varphi(\mathbf{X}_1(U_1)) \subset \mathbf{X}_2(U_2)$, the map

$$\mathbf{X}_2^{-1} \circ \varphi \circ \mathbf{X}_1 : U_1 \rightarrow U_2 \quad (5.7)$$

is differentiable at $Q = \mathbf{X}_1^{-1}(P)$ in the usual sense, i.e., if $\mathbf{q} := \mathbf{X}_2^{-1} \circ \varphi \circ \mathbf{X}_1$ and $\mathbf{q} = (q_1, q_2)^T$, then q_1 and q_2 are C^∞ functions at the point Q in U_1 . If $\varphi : V_1 \rightarrow \Gamma_2$ is differentiable at all points in V_1 , then we say that φ is a *differentiable map*. The definition of differentiability of a map between surfaces does not depend on the parameterization (we leave this for the reader to prove).

With differentiability comes the notion of diffeomorphism.

Definition 15. Two regular surfaces Γ_1 and Γ_2 are *diffeomorphic* if there exists a differentiable map $\varphi : \Gamma_1 \rightarrow \Gamma_2$ with a differentiable inverse $\varphi^{-1} : \Gamma_2 \rightarrow \Gamma_1$. Such a φ is called a *diffeomorphism* from Γ_1 to Γ_2 .

Hence, a diffeomorphism applied to a regular surface preserves the “smoothness” of the surface. It is one step up from a homeomorphism. An example is given as follows. Suppose $\mathbf{X} : U \rightarrow \Gamma$ is a parameterization of a regular surface Γ . Then we can show that $\mathbf{X}^{-1} : \mathbf{X}(U) \rightarrow \mathbb{R}^2$ is not only continuous but differentiable. Indeed, let P be in $\mathbf{X}(U)$ and consider another parameterization $\tilde{\mathbf{X}} : \tilde{U} \rightarrow \Gamma$ such that P is in $W = \mathbf{X}(U) \cap \tilde{\mathbf{X}}(\tilde{U})$. By Proposition 1 in Chapter 2, we know that

$$\mathbf{X}^{-1} \circ \tilde{\mathbf{X}} : \tilde{\mathbf{X}}^{-1}(W) \rightarrow \mathbf{X}^{-1}(W)$$

is differentiable at P . Referring back to (5.6) and (5.7), if we set $\mathbf{X}_2 := \text{id}_U$, $\varphi := \mathbf{X}^{-1}$, and $\mathbf{X}_1 := \tilde{\mathbf{X}}$, we see that

$$\mathbf{X}_2^{-1} \circ \varphi \circ \mathbf{X}_1 = \text{id}_U^{-1} \circ \varphi \circ \tilde{\mathbf{X}} = \mathbf{X}^{-1} \circ \tilde{\mathbf{X}},$$

which means that φ is differentiable at P in $\Gamma = \mathbf{X}(U)$. Since P is arbitrary in Γ , we have that $\varphi = \mathbf{X}^{-1}$ is a differentiable map.

Since $\varphi^{-1} = \mathbf{X}$ is also differentiable (because \mathbf{X} parameterizes a regular surface), we find that \mathbf{X} is a diffeomorphism. This proves that U and $\Gamma = \mathbf{X}(U)$ are diffeomorphic, i.e., every regular surface is locally diffeomorphic to a plane. In fact, we can characterize regular surfaces as those subsets $\Gamma \subset \mathbb{R}^3$ that are locally diffeomorphic to \mathbb{R}^2 . This fits with the comment in section 2.3 that a regular surface locally looks like a plane.

The same concepts apply to maps $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ between volumetric domains, e.g., if $\Phi : \Omega_1 \rightarrow \Omega_2$ is differentiable with a differentiable inverse $\Phi^{-1} : \Omega_2 \rightarrow \Omega_1$, then Ω_1 and Ω_2 are diffeomorphic.

Domain Mappings

We introduce a mapping function that represents the domain perturbation (recall Figure 2.2). Similarly to previous chapters, this allows calculations to be performed on a *reference* domain. Let Ω_ϵ be a perturbed domain, and let Ω_{all} be a fixed “hold-all” domain that contains $\overline{\Omega}_\epsilon$ for all ϵ in $[0, \epsilon_{\max}]$ (for some $\epsilon_{\max} > 0$). Let $\Phi_\epsilon : \Omega_{\text{all}} \rightarrow \Omega_{\text{all}}$ be a mapping for bulk domains Ω such that $\Phi_\epsilon(\Omega) = \Omega_\epsilon$ and

$$\Phi_0(\mathbf{x}) := \text{id}_\Omega(\mathbf{x}) = \mathbf{x} \quad \text{for all } \mathbf{x} \in \Omega \quad \Rightarrow \quad \Omega_0 \equiv \Omega.$$

The parameter ϵ is a flow parameter which parameterizes the deformation of Ω into Ω_ϵ .

For boundaries (i.e., surfaces) Γ , first let $\mathbf{X}(\cdot)$ be a total surface parameterization of Γ using local charts and reference domains $\{U_i\}$ (see section 2.3.3). This represents the *reference* surface. Next, let $\mathbf{X}_\epsilon : \cup_i U_i \rightarrow \Gamma_\epsilon$ be a total surface parameterization of a new (perturbed) surface Γ_ϵ (using the same reference domains $\{U_i\}$) such that $\mathbf{X}_\epsilon(\cup_i U_i) = \Gamma_\epsilon$, $\overline{\Gamma}_\epsilon \subset \Omega_{\text{all}}$ for all ϵ in $[0, \epsilon_{\max}]$, $\mathbf{X}_0 \equiv \mathbf{X}$, and

$$\mathbf{X}_0 \circ \mathbf{X}^{-1}(\mathbf{x}) := \text{id}_\Gamma(\mathbf{x}) = \mathbf{x} \quad \text{for all } \mathbf{x} \in \Gamma \quad \Rightarrow \quad \Gamma_0 \equiv \Gamma.$$

So, \mathbf{X}_ϵ is a mapping for **boundary** or surface domains Γ (i.e., domains of codimension 1). Similar considerations can be made for curve domains (i.e., domains of codimension 2).

For our purposes, the maps Φ_ϵ and \mathbf{X}_ϵ must be diffeomorphisms for all ϵ in $[0, \epsilon_{\max}]$ for some $\epsilon_{\max} > 0$.

Perturbations of the Identity

The deformation map can be defined for any $\epsilon > 0$ in many different ways. A straightforward way is to use a *perturbation of the identity*, i.e.,

$$\Phi_\epsilon(\mathbf{x}) := \text{id}_{\Omega_{\text{all}}}(\mathbf{x}) + \epsilon \mathbf{V}(\mathbf{x}) \quad \text{for all } \mathbf{x} \text{ in } \Omega_{\text{all}}, \tag{5.8}$$

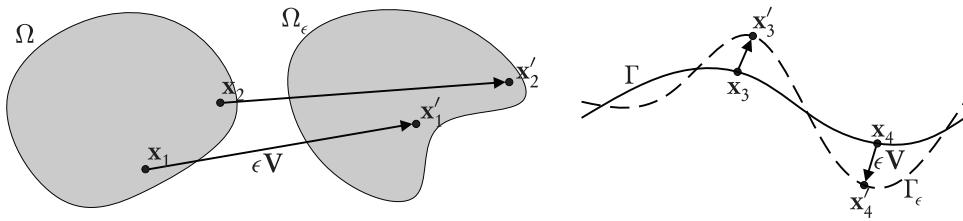


Figure 5.1. *Deformation of domains by perturbation of the identity. Using (5.8) and (5.9), we have $\Omega_\epsilon = \Phi_\epsilon(\Omega)$ and $\Gamma_\epsilon = X_\epsilon \circ X^{-1}(\Gamma)$. Each arrow corresponds to the vector $x'_i - x_i = \epsilon V(x_i)$ for $i = 1, 2, 3, 4$.*

where $V : \Omega_{\text{all}} \rightarrow \mathbb{R}^d$ and d is the dimension of Ω_{all} (here, we assume $d = 3$). Hence, one can think of V as a vector velocity field over Ω_{all} and ϵ as a “time” parameter; so $\epsilon V(x)$ is the net displacement of the point x . See Figure 5.1.

Similarly for surfaces, we have

$$X_\epsilon \circ X^{-1}(x) := \text{id}_\Gamma(x) + \epsilon V(x) \quad \text{for all } x \text{ in } \Gamma \subset \Omega_{\text{all}}. \quad (5.9)$$

In other words, $X_\epsilon = X + \epsilon(V \circ X)$, and $\overline{\Gamma}_\epsilon \subset \Omega_{\text{all}}$ for all ϵ in $[0, \epsilon_{\max}]$. Note that one can extend the definition of $X_\epsilon \circ X^{-1}$ to all of Ω_{all} .

If V is sufficiently smooth (e.g., V is C^∞) and $\epsilon_{\max} > 0$ is sufficiently small, then Φ_ϵ and X_ϵ are diffeomorphisms [23].

5.3 • Derivatives of Functions with Respect to Flow Map Φ_ϵ

Let $f_\epsilon : \Omega_\epsilon \rightarrow \mathbb{R}$ be a smooth function on Ω_ϵ that depends smoothly on ϵ . In other words, f_ϵ is a function defined on the deforming domain Ω_ϵ that changes with the flow parameter ϵ . So, not only is f_ϵ defined on Ω_ϵ , but f_ϵ also depends on Ω_ϵ . Note the identification $f \equiv f_0$.

5.3.1 • Material Derivative

The *material (Lagrangian) derivative* of f , with respect to the flow map Φ_ϵ , is defined as

$$\dot{f}(\Omega; V)(x) \equiv \dot{f}(x) := \lim_{\epsilon \rightarrow 0} \frac{f_\epsilon(\Phi_\epsilon(x)) - f(x)}{\epsilon} \quad \text{for all } x \text{ in } \Omega, \quad (5.10)$$

where the mapping Φ_ϵ is used to make the subtraction meaningful. Sometimes we use $D_V f \equiv \dot{f}(\Omega; V)$ to denote the material derivative.

Note that Φ_ϵ depends directly on V , which is the “direction” of differentiation (directional derivative). This operator obeys the usual rules of differential operators (e.g., product rule, chain rule, etc.), but does **not** commute with the standard time and space derivatives (see the counterexample below).

Let $\hat{f} : [0, \epsilon_{\max}] \times \Omega_{\text{all}} \rightarrow \mathbb{R}$ be a smooth function. Suppose that $f_\epsilon : \Omega_\epsilon \rightarrow \mathbb{R}$ can be viewed as the restriction of \hat{f} :

$$f_\epsilon(\Phi_\epsilon(x)) = \hat{f}(\epsilon, \Phi_\epsilon(x)) \quad \text{for all } x \in \Omega, \quad \epsilon \in [0, \epsilon_{\max}]. \quad (5.11)$$

Then the material derivative of f , with respect to the flow map Φ_ϵ , can be written more

explicitly:

$$\begin{aligned}\dot{f}(\mathbf{x}) &= \frac{d}{d\epsilon} \hat{f}(\epsilon, \Phi_\epsilon(\mathbf{x})) \Big|_{\epsilon=0} \\ &= \frac{\partial}{\partial \epsilon} \hat{f}(0, \mathbf{x}) + \nabla \hat{f}(0, \mathbf{x}) \cdot \frac{d}{d\epsilon} \Phi_\epsilon(\mathbf{x}) \Big|_{\epsilon=0} \\ &= \frac{\partial}{\partial \epsilon} \hat{f}(0, \mathbf{x}) + (\mathbf{V}(\mathbf{x}) \cdot \nabla) \underbrace{\hat{f}(0, \mathbf{x})}_{=f(\mathbf{x})}.\end{aligned}\quad (5.12)$$

The noncommutativity of the material derivative is demonstrated by simply comparing the following two formulas:

$$\begin{aligned}D_{\mathbf{V}}(\partial_{x_1} f) &= \frac{\partial}{\partial \epsilon} (\partial_{x_1} \hat{f}) + (\mathbf{V} \cdot \nabla)(\partial_{x_1} \hat{f}), \\ \partial_{x_1}(D_{\mathbf{V}} f) &= \partial_{x_1} \left(\frac{\partial}{\partial \epsilon} \hat{f} + (\mathbf{V} \cdot \nabla) \hat{f} \right) \\ &= \frac{\partial}{\partial \epsilon} (\partial_{x_1} \hat{f}) + (\mathbf{V} \cdot \nabla)(\partial_{x_1} \hat{f}) + ((\partial_{x_1} \mathbf{V}) \cdot \nabla) \hat{f}.\end{aligned}$$

5.3.2 ■ Shape Derivative

The *shape derivative* of f , with respect to the flow map Φ_ϵ , is defined as

$$f'(\Omega; \mathbf{V})(\mathbf{x}) \equiv f'(\mathbf{x}) := \frac{\partial}{\partial \epsilon} \hat{f}(0, \mathbf{x}).\quad (5.13)$$

In other words, the shape derivative is just the change in the function due to deforming the domain but without the convective effect of moving the domain. This is especially important if f is the solution of a PDE in the domain Ω . This operator is just a standard partial derivative, i.e., it obeys the usual rules of differential operators (e.g., product rule, chain rule, etc.), and it also commutes with the standard time and space derivatives (∂_t or ∂_{x_i}). One way to interpret this is that the standard time and space derivatives do not depend on the shape; hence, they can be factored through the shape derivative like a constant.

We conclude by writing the material derivative in terms of the shape derivative:

$$\dot{f}(\mathbf{x}) = f'(\mathbf{x}) + (\mathbf{V}(\mathbf{x}) \cdot \nabla) f(\mathbf{x}),\quad (5.14)$$

which makes no explicit mention of the extended function \hat{f} .

5.4 ■ Derivatives of Functions with Respect to Flow Map X_ϵ

The main difference here is that g_ϵ is defined on a surface Γ_ϵ , i.e., let $g_\epsilon : \Gamma_\epsilon \rightarrow \mathbb{R}$ be a smooth function on Γ_ϵ that depends smoothly on ϵ . Note the identification $g \equiv g_0$.

5.4.1 ■ Constant Normal Extension

It will behoove us to now describe the notion of a constant normal extension. Suppose Γ is a regular surface with a differentiable function $f : \Gamma \rightarrow \mathbb{R}$ defined on it. It is convenient

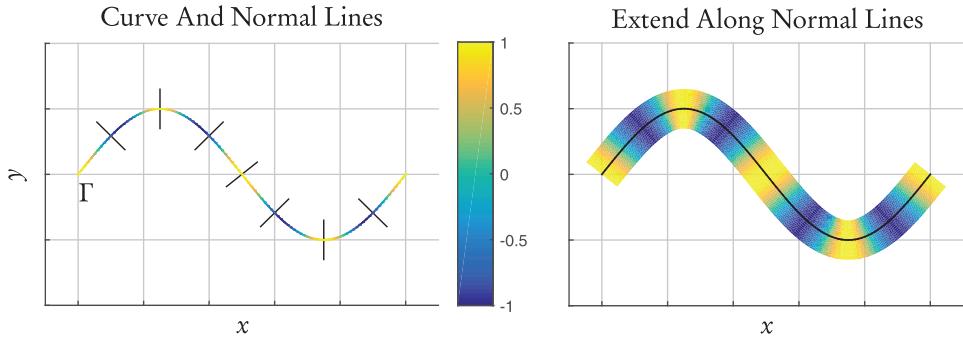


Figure 5.2. Constant normal extension. Left: Γ is a curve in the plane (normal lines are black). A function $f : \Gamma \rightarrow \mathbb{R}$ is given with values shown by color along Γ . Right: f is extended into a tubular neighborhood of Γ ; extension is constant along the normal lines.

to define an extension of f , i.e., $f_{\text{ext}} : \Omega_{\text{all}} \rightarrow \mathbb{R}$, where $f(\mathbf{x}) = f_{\text{ext}}(\mathbf{x})$ for all \mathbf{x} in Γ , with $\Gamma \subset \Omega_{\text{all}}$.

Of course, there are many choices of extensions. A useful one is the so-called **constant normal extension**, defined as follows. Let $\text{cl}(\Gamma, \mathbf{y})$ denote the closest point \mathbf{x} in Γ to the given point \mathbf{y} in Ω_{all} , i.e., $\text{dist}(\Gamma, \mathbf{y}) = |\text{cl}(\Gamma, \mathbf{y}) - \mathbf{y}|$. Next, define

$$f_{\text{ext}}(\mathbf{y}) := f(\text{cl}(\Gamma, \mathbf{y})) \quad \text{for all } \mathbf{y} \text{ in } \Omega_{\text{all}}. \quad (5.15)$$

Compare this extension with what was done in (5.11).

Now let \mathbf{x} be a fixed point in Γ and let $\mathbf{y}(t) = \mathbf{x} + t\nu(\mathbf{x})$, where $\nu(\mathbf{x})$ is the unit normal vector of Γ at \mathbf{x} . From (5.15), it can be shown that $f_{\text{ext}}(\mathbf{y}(t)) = f(\mathbf{x})$ for all $|t| < \epsilon$ for some $\epsilon > 0$ [44]. In other words, near Γ , f_{ext} is constant along normal lines through Γ ; Figure 5.2 gives an illustration. Thus, the constant normal extension satisfies the following property:

$$(\nu(\mathbf{x}) \cdot \nabla) f_{\text{ext}}(\mathbf{x}) = 0 \quad \text{for all } \mathbf{x} \text{ in } \Gamma. \quad (5.16)$$

By section 4.2.3, this implies that

$$\nabla f_{\text{ext}}(\mathbf{x}) = \nabla_{\Gamma} f_{\text{ext}}(\mathbf{x}) \quad \text{for all } \mathbf{x} \text{ in } \Gamma. \quad (5.17)$$

The constant normal extension can be done smoothly in a region near Γ [44]. In other words, there exists $\epsilon > 0$ such that if f is C^∞ , then f_{ext} is C^∞ in the open set

$$T_b(\Gamma) := \{\mathbf{x} \in \Omega_{\text{all}} : \text{dist}(\Gamma, \mathbf{x}) < \epsilon\}; \quad \text{note that } \overline{\Gamma} \subset T_b(\Gamma). \quad (5.18)$$

We call $T_b(\Gamma)$ a *tubular neighborhood* of Γ (see Figure 5.2).

5.4.2 • Material Derivative

Similar to (5.10), we have that the *material (Lagrangian) derivative* of g , with respect to the flow map \mathbf{X}_ϵ , is defined as

$$\dot{g}(\Gamma; \mathbf{V})(\mathbf{x}) \equiv \dot{g}(\mathbf{x}) := \lim_{\epsilon \rightarrow 0} \frac{g_\epsilon(\mathbf{X}_\epsilon \circ \mathbf{X}^{-1}(\mathbf{x})) - g(\mathbf{x})}{\epsilon} \quad \text{for all } \mathbf{x} \text{ in } \Gamma. \quad (5.19)$$

Let $\hat{g} : [0, \epsilon_{\max}] \times \Omega_{\text{all}} \rightarrow \mathbb{R}$ be a smooth function. Suppose that $g_\epsilon : \Gamma_\epsilon \rightarrow \mathbb{R}$ can be viewed as the restriction of \hat{g} :

$$g_\epsilon(\mathbf{X}_\epsilon \circ \mathbf{X}^{-1}(\mathbf{x})) = \hat{g}(\epsilon, \mathbf{X}_\epsilon \circ \mathbf{X}^{-1}(\mathbf{x})) \quad \text{for all } \mathbf{x} \in \Gamma, \quad \epsilon \in [0, \epsilon_{\max}]. \quad (5.20)$$

In other words, \hat{g} is an extension of g_ϵ from Γ_ϵ to Ω_{all} . Moreover, we choose this extension away from Γ_ϵ to be *constant in the normal direction* (section 5.4.1). Then the material derivative of g , with respect to the flow map \mathbf{X}_ϵ , can be written more explicitly:

$$\begin{aligned} \dot{g}(\mathbf{x}) &= \frac{d}{d\epsilon} \hat{g}(\epsilon, \mathbf{X}_\epsilon \circ \mathbf{X}^{-1}(\mathbf{x})) \Big|_{\epsilon=0} \\ &= \frac{\partial}{\partial \epsilon} \hat{g}(0, \mathbf{x}) + \nabla \hat{g}(0, \mathbf{x}) \cdot \frac{d}{d\epsilon} \mathbf{X}_\epsilon \circ \mathbf{X}^{-1}(\mathbf{x}) \Big|_{\epsilon=0} \\ &= \frac{\partial}{\partial \epsilon} \hat{g}(0, \mathbf{x}) + (\mathbf{V}(\mathbf{x}) \cdot \nabla) \hat{g}(0, \mathbf{x}), \\ &= \frac{\partial}{\partial \epsilon} \hat{g}(0, \mathbf{x}) + (\mathbf{V}(\mathbf{x}) \cdot \nabla_\Gamma) \hat{g}(0, \mathbf{x}), \end{aligned} \quad (5.21)$$

where the last line follows because \hat{g} is constant in the normal direction.

5.4.3 ■ Shape Derivative

The *shape derivative* of g , with respect to the flow map \mathbf{X}_ϵ , is defined as

$$g'(\Gamma; \mathbf{V})(\mathbf{x}) \equiv g'(\mathbf{x}) := \frac{\partial}{\partial \epsilon} \hat{g}(0, \mathbf{x}). \quad (5.22)$$

We conclude by writing the material derivative in terms of the shape derivative:

$$\dot{g}(\mathbf{x}) = g'(\mathbf{x}) + (\mathbf{V}(\mathbf{x}) \cdot \nabla_\Gamma) g(\mathbf{x}), \quad (5.23)$$

which makes no explicit mention of the extended function \hat{g} .

5.5 ■ Basic Identities

5.5.1 ■ Preliminaries

Clearly, $(\Phi_\epsilon^{-1} \circ \Phi_\epsilon)(\mathbf{x}) = (\Phi_\epsilon \circ \Phi_\epsilon^{-1})(\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x} \in \Omega_{\text{all}}$ and all ϵ in $[0, \epsilon_{\max}]$. Let $\mathbf{y} = \Phi_\epsilon(\mathbf{x})$ (for a fixed ϵ) so that $\Phi_\epsilon^{-1}(\mathbf{y}) = \mathbf{x}$. Then,

$$\begin{aligned} \mathbf{I} &= \nabla_{\mathbf{x}}(\Phi_\epsilon^{-1} \circ \Phi_\epsilon)(\mathbf{x}) = (\nabla_{\mathbf{y}} \Phi_\epsilon^{-1}(\mathbf{y}))(\nabla_{\mathbf{x}} \Phi_\epsilon(\mathbf{x})) \\ &\implies (\nabla_{\mathbf{y}} \Phi_\epsilon^{-1} \circ \Phi_\epsilon)(\mathbf{x}) = (\nabla_{\mathbf{x}} \Phi_\epsilon(\mathbf{x}))^{-1}, \end{aligned} \quad (5.24)$$

where the term on the end is the matrix inverse of $\nabla_{\mathbf{x}} \Phi_\epsilon(\mathbf{x})$. We use the symbol $\nabla_{\mathbf{x}}$ to emphasize that we are differentiating with respect to \mathbf{x} .

5.5.2 ■ Perturbations of Mappings

Lemma 5.1 (identities for Ω mappings). *Let Φ_ϵ be given by (5.8). Then Φ_ϵ satisfies*

1. $\nabla \Phi_\epsilon(\mathbf{x}) \Big|_{\epsilon=0} = \mathbf{I}$ (i.e., the identity matrix),
2. $\frac{d}{d\epsilon} \Phi_\epsilon(\mathbf{x}) \Big|_{\epsilon=0} = \mathbf{V}(\mathbf{x}),$

3. $\frac{d}{d\epsilon} \nabla \Phi_\epsilon(\mathbf{x})|_{\epsilon=0} = \nabla \mathbf{V}(\mathbf{x}),$
4. $\frac{d}{d\epsilon} \nabla (\Phi_\epsilon^{-1})(\mathbf{x})|_{\epsilon=0} = -\nabla \mathbf{V}(\mathbf{x}),$
5. $\frac{d}{d\epsilon} \det(\nabla \Phi_\epsilon(\mathbf{x}))|_{\epsilon=0} = \nabla \cdot \mathbf{V}(\mathbf{x})$

for all \mathbf{x} in Ω .

Proof. Identities (1), (2), and (3) follow from (5.8). Identity (4) follows by computing the derivative of (5.24), using (1) and (3), and noting that $\mathbf{y} = \Phi_\epsilon(\mathbf{x})|_{\epsilon=0} = \mathbf{x}$. Identity (5) follows from (B.2):

$$\begin{aligned} \frac{d}{d\epsilon} \det(\nabla_x \Phi_\epsilon(\mathbf{x})) &= \det(\nabla_x \Phi_\epsilon(\mathbf{x})) \cdot \text{trace} \left[\left(\frac{d}{d\epsilon} (\nabla_x \Phi_\epsilon(\mathbf{x})) \right) (\nabla_x \Phi_\epsilon(\mathbf{x}))^{-1} \right], \\ \frac{d}{d\epsilon} \det(\nabla_x \Phi_\epsilon(\mathbf{x}))|_{\epsilon=0} &= \det(\mathbf{I}) \cdot \text{trace} [(\nabla_x \mathbf{V}(\mathbf{x})) (\mathbf{I})^{-1}] = \nabla_x \cdot \mathbf{V}(\mathbf{x}). \quad \square \end{aligned}$$

Lemma 5.2 (identities for Γ mappings). Let \mathbf{X}_ϵ be given by (5.9). In particular, assume that $\mathbf{X}_\epsilon : \cup_i U_i \rightarrow \Gamma_\epsilon \subset \mathbb{R}^3$ (i.e., a 2-D surface in three dimensions), with local coordinates in the reference domain U_i denoted (s_1, s_2) . Then \mathbf{X}_ϵ satisfies

1. $\nabla_\Gamma (\mathbf{X}_\epsilon \circ \mathbf{X}^{-1})(\mathbf{x})|_{\epsilon=0} = \nabla_\Gamma \text{id}_\Gamma(\mathbf{x}) := \mathbf{I} - \boldsymbol{\nu} \otimes \boldsymbol{\nu}$ (i.e., tangent space projection),
2. $\frac{d}{d\epsilon} (\mathbf{X}_\epsilon \circ \mathbf{X}^{-1})(\mathbf{x})|_{\epsilon=0} = \mathbf{V}(\mathbf{x}),$
3. $\frac{d}{d\epsilon} \nabla_\Gamma (\mathbf{X}_\epsilon \circ \mathbf{X}^{-1})(\mathbf{x})|_{\epsilon=0} = \nabla_\Gamma \mathbf{V}(\mathbf{x}),$
4. $\frac{d}{d\epsilon} \nabla_\Gamma ((\mathbf{X}_\epsilon \circ \mathbf{X}^{-1})^{-1})(\mathbf{x})|_{\epsilon=0} = -\nabla_\Gamma \mathbf{V}(\mathbf{x})$

for all \mathbf{x} in Γ . In addition, \mathbf{X}_ϵ satisfies

$$5. \frac{d}{d\epsilon} \left| \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon \right|_{\epsilon=0} = \{(\nabla_\Gamma \text{id}_\Gamma) \circ \mathbf{X}\} : \{(\nabla_\Gamma \mathbf{V}) \circ \mathbf{X}\} | \partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}|$$

for all (s_1, s_2) in every reference domain U_i .

Proof. Identities (1), (2), and (3) follow from (5.9).

Show identity (4): first, define $F_\epsilon(\mathbf{x}) = \mathbf{X}_\epsilon \circ \mathbf{X}^{-1}(\mathbf{x})$ (recall (5.9)). Let $\widehat{\mathbf{V}}(\mathbf{x})$ be an extension of $\mathbf{V}(\mathbf{x})$ to a “hold-all” domain Ω_{all} that compactly contains (see section 2.1.3) Γ_ϵ for all $\epsilon \in [0, \epsilon_{\max}]$. Now define the extension of $F_\epsilon(\mathbf{x})$:

$$\widehat{F}_\epsilon(\mathbf{x}) = \text{id}_{\Omega_{\text{all}}}(\mathbf{x}) + \epsilon \widehat{\mathbf{V}}(\mathbf{x}) \quad \text{for all } \mathbf{x} \text{ in } \Omega_{\text{all}},$$

which is an extension of (5.9) to Ω_{all} . Assuming sufficient smoothness, $\widehat{F}_\epsilon(\mathbf{x})$ satisfies identity (4) of Lemma 5.1: $\frac{d}{d\epsilon} \nabla \widehat{F}_\epsilon^{-1}(\mathbf{x})|_{\epsilon=0} = -\nabla \widehat{\mathbf{V}}(\mathbf{x})$. The identity then follows by restricting \mathbf{x} to Γ and multiplying by $\nabla_\Gamma \text{id}_\Gamma(\mathbf{x})$:

$$\begin{aligned} \left(\frac{d}{d\epsilon} \nabla F_\epsilon^{-1}(\mathbf{x}) \right) \nabla_\Gamma \text{id}_\Gamma(\mathbf{x}) &= -\nabla \mathbf{V}(\mathbf{x}) \nabla_\Gamma \text{id}_\Gamma(\mathbf{x}) = -(\nabla_\Gamma \text{id}_\Gamma \cdot \nabla) \mathbf{V}(\mathbf{x}), \\ \frac{d}{d\epsilon} [(\nabla_\Gamma \text{id}_\Gamma \cdot \nabla) F_\epsilon^{-1}(\mathbf{x})]|_{\epsilon=0} &= -\nabla_\Gamma \mathbf{V}(\mathbf{x}) \quad \Rightarrow \quad \frac{d}{d\epsilon} [\nabla_\Gamma (\mathbf{X}_\epsilon \circ \mathbf{X}^{-1})^{-1}(\mathbf{x})]|_{\epsilon=0} = -\nabla_\Gamma \mathbf{V}(\mathbf{x}), \end{aligned}$$

because $(\nabla_\Gamma \text{id}_\Gamma) \cdot \nabla = \nabla_\Gamma$.

Show identity (5): by Proposition 19 of Chapter 4, we have

$$\{(\nabla_\Gamma \mathbf{V}) \circ \mathbf{X}\} : \{(\nabla_\Gamma \text{id}_\Gamma) \circ \mathbf{X}\} = \sum_{i,j=1}^2 g^{ij} \partial_{s_i} \tilde{\mathbf{V}} \cdot \partial_{s_j} \mathbf{X},$$

where $\tilde{\mathbf{V}} = \mathbf{V} \circ \mathbf{X}$. Next, write this more explicitly for later use:

$$\begin{aligned} \{(\nabla_\Gamma \mathbf{V}) \circ \mathbf{X}\} : \{(\nabla_\Gamma \text{id}_\Gamma) \circ \mathbf{X}\} &= \frac{1}{\det(g)} [g_{22}(\partial_{s_1} \tilde{\mathbf{V}} \cdot \partial_{s_1} \mathbf{X}) - g_{12}(\partial_{s_1} \tilde{\mathbf{V}} \cdot \partial_{s_2} \mathbf{X}) \\ &\quad - g_{21}(\partial_{s_2} \tilde{\mathbf{V}} \cdot \partial_{s_1} \mathbf{X}) + g_{11}(\partial_{s_2} \tilde{\mathbf{V}} \cdot \partial_{s_2} \mathbf{X})]. \end{aligned} \quad (5.25)$$

Now compute the left-hand side of (5):

$$\begin{aligned} \frac{d}{d\epsilon} \left| \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon \right| \Big|_{\epsilon=0} &= \frac{d}{d\epsilon} [(\partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon) \cdot (\partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon)]^{1/2} \Big|_{\epsilon=0} \\ &= \frac{(\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X})}{|\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}|} \cdot \frac{d}{d\epsilon} (\partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon) \Big|_{\epsilon=0}. \end{aligned} \quad (5.26)$$

Since $\frac{d}{d\epsilon} \mathbf{X}_\epsilon \Big|_{\epsilon=0} = \mathbf{V} \circ \mathbf{X} = \tilde{\mathbf{V}}$, we have that

$$\frac{d}{d\epsilon} \partial_{s_1} \mathbf{X}_\epsilon \Big|_{\epsilon=0} = \partial_{s_1} \tilde{\mathbf{V}}, \quad \frac{d}{d\epsilon} \partial_{s_2} \mathbf{X}_\epsilon \Big|_{\epsilon=0} = \partial_{s_2} \tilde{\mathbf{V}}.$$

Plugging this into (5.26) and applying the product rule gives

$$\frac{d}{d\epsilon} \left| \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon \right| \Big|_{\epsilon=0} = \frac{(\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X})}{|\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}|} \cdot [(\partial_{s_1} \tilde{\mathbf{V}} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \tilde{\mathbf{V}})]. \quad (5.27)$$

Using $|\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}| = \sqrt{\det(g)}$ and the vector identity (A.3), we rewrite (5.27) as

$$\begin{aligned} \frac{d}{d\epsilon} \left| \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon \right| \Big|_{\epsilon=0} &= \frac{1}{\sqrt{\det(g)}} [+(\partial_{s_2} \mathbf{X} \cdot \partial_{s_2} \mathbf{X})(\partial_{s_1} \tilde{\mathbf{V}} \cdot \partial_{s_1} \mathbf{X}) \\ &\quad - (\partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \mathbf{X})(\partial_{s_1} \tilde{\mathbf{V}} \cdot \partial_{s_2} \mathbf{X}) \\ &\quad - (\partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \mathbf{X})(\partial_{s_2} \tilde{\mathbf{V}} \cdot \partial_{s_1} \mathbf{X}) \\ &\quad + (\partial_{s_1} \mathbf{X} \cdot \partial_{s_1} \mathbf{X})(\partial_{s_2} \tilde{\mathbf{V}} \cdot \partial_{s_2} \mathbf{X})], \end{aligned}$$

which, by (3.2), becomes

$$\begin{aligned} \frac{d}{d\epsilon} \left| \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon \right| \Big|_{\epsilon=0} &= \frac{1}{\sqrt{\det(g)}} [g_{22}(\partial_{s_1} \tilde{\mathbf{V}} \cdot \partial_{s_1} \mathbf{X}) - g_{12}(\partial_{s_1} \tilde{\mathbf{V}} \cdot \partial_{s_2} \mathbf{X}) \\ &\quad - g_{21}(\partial_{s_2} \tilde{\mathbf{V}} \cdot \partial_{s_1} \mathbf{X}) + g_{11}(\partial_{s_2} \tilde{\mathbf{V}} \cdot \partial_{s_2} \mathbf{X})]. \end{aligned} \quad (5.28)$$

Now notice that (5.28) is almost exactly (5.25). Thus, we get the identity (5) that we wanted:

$$\begin{aligned} \frac{d}{d\epsilon} \left| \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon \right| \Big|_{\epsilon=0} &= \{(\nabla_\Gamma \mathbf{V}) \circ \mathbf{X}\} : \{(\nabla_\Gamma \text{id}_\Gamma) \circ \mathbf{X}\} \sqrt{\det(g)} \\ &= \{(\nabla_\Gamma \mathbf{V}) \circ \mathbf{X}\} : \{(\nabla_\Gamma \text{id}_\Gamma) \circ \mathbf{X}\} |\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}|. \quad \square \end{aligned}$$

5.5.3 ▪ Perturbations of Geometric Quantities (Local Charts)

Proposition 25 (material derivative of surface Jacobian). *Assume the hypothesis in Lemma 5.2. Then \mathbf{X}_ϵ satisfies*

$$\frac{d}{d\epsilon} \left| \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon \right|^2 \Big|_{\epsilon=0^+} = 2(\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}) \cdot [\partial_{s_1} \tilde{\mathbf{V}} \times \partial_{s_2} \mathbf{X} + \partial_{s_1} \mathbf{X} \times \partial_{s_2} \tilde{\mathbf{V}}], \quad (5.29)$$

$$\frac{d}{d\epsilon} \left| \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon \right| \Big|_{\epsilon=0^+} = \nu \cdot [\partial_{s_1} \tilde{\mathbf{V}} \times \partial_{s_2} \mathbf{X} + \partial_{s_1} \mathbf{X} \times \partial_{s_2} \tilde{\mathbf{V}}] \quad (5.30)$$

for all (s_1, s_2) in any reference domain U_i , where $\tilde{\mathbf{V}} = \mathbf{V} \circ \mathbf{X}$.

Proof. The first result follows from straight differentiation:

$$\begin{aligned} \frac{d}{d\epsilon} \left| \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon \right|^2 &= \frac{d}{d\epsilon} (\partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon) \cdot (\partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon) \\ &= 2(\partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon) \cdot \left[\left(\partial_{s_1} \frac{d}{d\epsilon} \mathbf{X}_\epsilon \right) \times \partial_{s_2} \mathbf{X}_\epsilon + \partial_{s_1} \mathbf{X}_\epsilon \times \left(\partial_{s_2} \frac{d}{d\epsilon} \mathbf{X}_\epsilon \right) \right] \\ &= 2(\partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon) \cdot [\partial_{s_1} \tilde{\mathbf{V}} \times \partial_{s_2} \mathbf{X}_\epsilon + \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \tilde{\mathbf{V}}]. \end{aligned}$$

Letting $\epsilon \rightarrow 0^+$, we obtain (5.29).

As for (5.30),

$$\begin{aligned} \frac{d}{d\epsilon} \left| \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon \right| &= \frac{d}{d\epsilon} \sqrt{(\partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon) \cdot (\partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon)} \\ &= \frac{1}{2} \frac{1}{\left| \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon \right|} \frac{d}{d\epsilon} \left| \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon \right|^2 \\ &= \frac{\partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon}{\left| \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon \right|} \cdot [\partial_{s_1} \tilde{\mathbf{V}} \times \partial_{s_2} \mathbf{X}_\epsilon + \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \tilde{\mathbf{V}}] \\ &= \nu_\epsilon \cdot [\partial_{s_1} \tilde{\mathbf{V}} \times \partial_{s_2} \mathbf{X}_\epsilon + \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \tilde{\mathbf{V}}], \end{aligned}$$

where ν_ϵ is the normal vector of Γ_ϵ . Letting $\epsilon \rightarrow 0^+$, we obtain (5.30). \square

Lemma 5.3 (material derivative of normal vector). *Assume the hypothesis in Lemma 5.2. Then*

$$\dot{\nu} = \frac{d}{d\epsilon} \nu_\epsilon \Big|_{\epsilon=0^+} = -\nu \{ [\nabla_\Gamma \text{id}_\Gamma \circ \mathbf{X}] : [\nabla_\Gamma \mathbf{V} \circ \mathbf{X}] \} + \frac{[\partial_{s_1} \tilde{\mathbf{V}} \times \partial_{s_2} \mathbf{X} + \partial_{s_1} \mathbf{X} \times \partial_{s_2} \tilde{\mathbf{V}}]}{\left| \partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X} \right|} \quad (5.31)$$

for all (s_1, s_2) in any reference domain U_i , where $\tilde{\mathbf{V}} = \mathbf{V} \circ \mathbf{X}$.

Proof. Note that

$$\nu_\epsilon = \frac{\partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon}{\left| \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon \right|} \quad (5.32)$$

and differentiate

$$\begin{aligned}\frac{d}{d\epsilon}\nu_\epsilon &= \left(\frac{d}{d\epsilon} \left| \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon \right|^{-1}\right) (\partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon) \\ &\quad + \left| \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon \right|^{-1} \frac{d}{d\epsilon} (\partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon).\end{aligned}$$

This simplifies to

$$\frac{d}{d\epsilon}\nu_\epsilon = -\frac{(\partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon)}{\left| \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon \right|^2} \left(\frac{d}{d\epsilon} \left| \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon \right| \right) + \frac{\left[\partial_{s_1} \tilde{\mathbf{V}} \times \partial_{s_2} \mathbf{X}_\epsilon + \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \tilde{\mathbf{V}} \right]}{\left| \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon \right|},$$

and then taking $\epsilon \rightarrow 0^+$ and using identity (5) from Lemma 5.2 yields (5.31). \square

Lemma 5.4 (material derivative of summed curvature vector). *Assume the hypothesis in Lemma 5.2. Let $\kappa = \kappa\nu$, where κ is the summed curvature of Γ and ν is the oriented normal vector. Then*

$$\begin{aligned}\dot{\kappa} &= \frac{d}{d\epsilon}\kappa_\epsilon \Big|_{\epsilon=0^+} = -\{[\nabla_\Gamma \text{id}_\Gamma \circ \mathbf{X}] : [\nabla_\Gamma \mathbf{V} \circ \mathbf{X}]\} \kappa \\ &\quad + \left| \partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X} \right|^{-1} \left\{ (\partial_{s_1} \dot{\nu} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \dot{\nu}) \right. \\ &\quad \left. + (\partial_{s_1} \nu \times \partial_{s_2} \tilde{\mathbf{V}}) + (\partial_{s_1} \tilde{\mathbf{V}} \times \partial_{s_2} \nu) \right\} \end{aligned}\quad (5.33)$$

for all (s_1, s_2) in any reference domain U_i , where $\tilde{\mathbf{V}} = \mathbf{V} \circ \mathbf{X}$.

Proof. By Proposition 12 of Chapter 4, we have

$$\kappa_\epsilon = \frac{(\partial_{s_1} \nu_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon) + (\partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \nu_\epsilon)}{\left| \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon \right|}.$$

Differentiating, we get

$$\begin{aligned}\frac{d}{d\epsilon}\kappa_\epsilon \Big|_{\epsilon=0^+} &= -\frac{d}{d\epsilon} \left| \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon \right| \Big|_{\epsilon=0^+} \frac{(\partial_{s_1} \nu \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \nu)}{\left| \partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X} \right|^2} \\ &\quad + \frac{(\partial_{s_1} \dot{\nu} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \dot{\nu})}{\left| \partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X} \right|} + \frac{(\partial_{s_1} \nu \times \partial_{s_2} \dot{\mathbf{X}}) + (\partial_{s_1} \dot{\mathbf{X}} \times \partial_{s_2} \nu)}{\left| \partial_{s_1} \dot{\mathbf{X}} \times \partial_{s_2} \mathbf{X} \right|}.\end{aligned}$$

Next, simplify with identity (5) of Lemma 5.2 and $\dot{\mathbf{X}} = \tilde{\mathbf{V}}$ (see Proposition 27):

$$\begin{aligned}\frac{d}{d\epsilon}\kappa_\epsilon \Big|_{\epsilon=0^+} &= -\{[\nabla_\Gamma \text{id}_\Gamma \circ \mathbf{X}] : [\nabla_\Gamma \mathbf{V} \circ \mathbf{X}]\} \frac{(\partial_{s_1} \nu \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \nu)}{\left| \partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X} \right|} \\ &\quad + \frac{(\partial_{s_1} \dot{\nu} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \dot{\nu})}{\left| \partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X} \right|} + \frac{(\partial_{s_1} \nu \times \partial_{s_2} \tilde{\mathbf{V}}) + (\partial_{s_1} \tilde{\mathbf{V}} \times \partial_{s_2} \nu)}{\left| \partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X} \right|}.\end{aligned}$$

Using (4.28), we obtain (5.33). \square

5.5.4 • Perturbations of Geometric Quantities (Intrinsic Formulas)

We derive other formulas for perturbing the geometry, e.g., for $\dot{\text{id}}_\Gamma$, $\dot{\nu}$, and $\dot{\kappa}$, where the formulas clearly do **not** depend on the parameterization.

Surface Measure

One can view (5) of Lemma 5.2 as a deformation of the local area of the surface Γ .

Proposition 26 (shape perturbation of surface measure). *Assume the hypothesis in Lemma 5.2. Suppose $S \subset \Gamma$ is a surface patch of Γ and let U_S be a reference domain for the surface patch S , i.e., $S = \mathbf{X}(U_S)$. Define $S_\epsilon = \mathbf{X}_\epsilon(U_S)$, where $S \equiv S_0$. Then*

$$|\dot{S}| := \frac{d}{d\epsilon} |S_\epsilon| \Big|_{\epsilon=0^+} = \int_S \nabla_\Gamma \cdot \mathbf{V}, \quad (5.34)$$

where $|S|$ denotes the surface area of S .

Proof. Applying the standard change of variable formula for surface integrals gives

$$|S_\epsilon| = \int_{S_\epsilon} 1 = \int_{U_S} \left| \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon \right|.$$

Therefore, by standard calculus of variations and (5) of Lemma 5.2, we have

$$\begin{aligned} |\dot{S}| &= \frac{d}{d\epsilon} |S_\epsilon| \Big|_{\epsilon=0^+} = \int_{U_S} \frac{d}{d\epsilon} \left| \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon \right| \Big|_{\epsilon=0^+} \\ &= \int_{U_S} [(\mathbf{I} - \boldsymbol{\nu} \otimes \boldsymbol{\nu}) : (\nabla_\Gamma \mathbf{V})] \circ \mathbf{X} \left| \partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X} \right| \\ &= \int_{U_S} (\nabla_\Gamma \cdot \mathbf{V}) \circ \mathbf{X} \left| \partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X} \right| = \int_S \nabla_\Gamma \cdot \mathbf{V}, \end{aligned}$$

where we have applied the change of variable formula again (in reverse). \square

Position Vector

Proposition 27 (shape derivative of id_Γ). *Assume the hypothesis in Lemma 5.2. Then*

$$\begin{aligned} \dot{\text{id}}_\Gamma &= \mathbf{V}, \\ \dot{\text{id}}'_\Gamma &= (\mathbf{V} \cdot \boldsymbol{\nu}) \boldsymbol{\nu} \quad (2\text{-D surface } \Gamma), \\ \dot{\text{id}}''_\Sigma &= \mathbf{V} - (\mathbf{V} \cdot \boldsymbol{\tau}) \boldsymbol{\tau} \quad (1\text{-D curve } \Sigma), \end{aligned} \quad (5.35)$$

evaluated on Γ . Note that $\dot{\text{id}}_\Gamma = \dot{\mathbf{X}} \circ \mathbf{X}^{-1}$, $\dot{\text{id}}'_\Gamma = \mathbf{X}' \circ \mathbf{X}^{-1}$, and $\dot{\text{id}}''_\Sigma = \boldsymbol{\alpha}' \circ \boldsymbol{\alpha}^{-1}$.

Proof. Recall that $\text{id}_\Gamma = \mathbf{X} \circ \mathbf{X}^{-1}$. The fact that $\dot{\text{id}}_\Gamma = \mathbf{V}$ follows directly from (5.9) and the definition (5.19). For the other formula, by (5.23), we see that

$$\begin{aligned}\text{id}'_\Gamma &= \dot{\text{id}}_\Gamma - (\mathbf{V} \cdot \nabla_\Gamma) \text{id}_\Gamma = \mathbf{V} - (\nabla_\Gamma \text{id}_\Gamma) \mathbf{V} \\ &= [\mathbf{I} - (\mathbf{I} - \boldsymbol{\nu} \otimes \boldsymbol{\nu})] \mathbf{V} = (\mathbf{V} \cdot \boldsymbol{\nu}) \boldsymbol{\nu},\end{aligned}$$

or

$$\begin{aligned}\text{id}'_\Sigma &= \dot{\text{id}}_\Sigma - (\mathbf{V} \cdot \nabla_\Gamma) \text{id}_\Sigma = \mathbf{V} - (\nabla_\Sigma \text{id}_\Sigma) \mathbf{V} \\ &= [\mathbf{I} - \boldsymbol{\tau} \otimes \boldsymbol{\tau}] \mathbf{V} = \mathbf{V} - (\mathbf{V} \cdot \boldsymbol{\tau}) \boldsymbol{\tau},\end{aligned}$$

where we have used Proposition 7 of Chapter 4. \square

Remark 24. Suppose $\mathbf{V}|_\Gamma$ points in the tangent space of Γ (i.e., $\mathbf{V} \cdot \boldsymbol{\nu} = 0$). Then (5.9) says that points in Γ are displaced tangentially on Γ , i.e., the shape of Γ does not (significantly) change for small ϵ . According to (5.35), $\dot{\text{id}}'_\Gamma = (\mathbf{V} \cdot \boldsymbol{\nu}) \boldsymbol{\nu} = 0$. This is an example of how the shape derivative is insensitive to tangential velocity perturbations.

Normal Vector

We need a formula for transporting the normal vector of a surface when the surface is deformed. Recall that $\mathbf{X}(s_1, s_2)$ parameterizes the surface $\Gamma \equiv \Gamma_0$, with normal vector $\boldsymbol{\nu}$ given by (3.22). So $\boldsymbol{\nu}$ is transported to $\boldsymbol{\nu}_\epsilon$ (defined on Γ_ϵ) by using $\mathbf{X}_\epsilon(s_1, s_2) := \Phi_\epsilon(\mathbf{X}(s_1, s_2))$:

$$\begin{aligned}\boldsymbol{\nu}_\epsilon \circ \Phi_\epsilon(\mathbf{X}) &= \frac{\partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon}{\left| \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon \right|} = \frac{[(\nabla_{\mathbf{X}} \Phi_\epsilon(\mathbf{X})) \partial_{s_1} \mathbf{X}] \times [(\nabla_{\mathbf{X}} \Phi_\epsilon(\mathbf{X})) \partial_{s_2} \mathbf{X}]}{\left| [(\nabla_{\mathbf{X}} \Phi_\epsilon(\mathbf{X})) \partial_{s_1} \mathbf{X}] \times [(\nabla_{\mathbf{X}} \Phi_\epsilon(\mathbf{X})) \partial_{s_2} \mathbf{X}] \right|} \\ (A.18) \rightarrow &= \frac{(\det \nabla_{\mathbf{X}} \Phi_\epsilon(\mathbf{X})) (\nabla_{\mathbf{X}} \Phi_\epsilon(\mathbf{X})^T)^{-1} (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X})}{\left| (\det \nabla_{\mathbf{X}} \Phi_\epsilon(\mathbf{X})) (\nabla_{\mathbf{X}} \Phi_\epsilon(\mathbf{X})^T)^{-1} (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}) \right|} \\ (3.22) \rightarrow &= \frac{(\nabla_{\mathbf{X}} \Phi_\epsilon(\mathbf{X})^T)^{-1} \boldsymbol{\nu}}{\left| (\nabla_{\mathbf{X}} \Phi_\epsilon(\mathbf{X})^T)^{-1} \boldsymbol{\nu} \right|}.\end{aligned}\tag{5.36}$$

Note that the Jacobian is positive for sufficiently small ϵ . In (5.36), we may interpret \mathbf{X} as the coordinates of a point on the surface Γ , and $\boldsymbol{\nu}$ is evaluated at the point \mathbf{X} . To avoid confusing notation, we rewrite (5.36) as

$$\boldsymbol{\nu}_\epsilon \circ \Phi_\epsilon(\mathbf{a}) = \frac{(\nabla_{\mathbf{a}} \Phi_\epsilon(\mathbf{a})^T)^{-1} \boldsymbol{\nu}(\mathbf{a})}{\left| (\nabla_{\mathbf{a}} \Phi_\epsilon(\mathbf{a})^T)^{-1} \boldsymbol{\nu}(\mathbf{a}) \right|} \quad \text{for all } \mathbf{a} \in \Gamma.\tag{5.37}$$

Lemma 5.5 (shape derivative of $\boldsymbol{\nu}$). Assume the hypothesis in Lemma 5.2. Then

$$\dot{\boldsymbol{\nu}} = -(\nabla_\Gamma \mathbf{V})^T \boldsymbol{\nu}, \quad \boldsymbol{\nu}' = -\nabla_\Gamma (\mathbf{V} \cdot \boldsymbol{\nu})^T,\tag{5.38}$$

evaluated on Γ .

Proof. We proceed to differentiate (5.37), i.e.,

$$\begin{aligned}\dot{\nu} &= \frac{d}{d\epsilon} \nu_\epsilon \circ \Phi_\epsilon(\mathbf{a}) \Big|_{\epsilon=0} \\ &= \frac{d}{d\epsilon} (\nabla_{\mathbf{a}} \Phi_\epsilon(\mathbf{a})^T)^{-1} \nu \Big|_{\epsilon=0} - \left[\nu \cdot \left(\frac{d}{d\epsilon} (\nabla_{\mathbf{a}} \Phi_\epsilon(\mathbf{a})^T)^{-1} \nu \right) \Big|_{\epsilon=0} \right] \nu, \\ &= (\mathbf{I} - \nu \otimes \nu) \left(\frac{d}{d\epsilon} (\nabla_{\mathbf{a}} \Phi_\epsilon(\mathbf{a})^T)^{-1} \Big|_{\epsilon=0} \right) \nu,\end{aligned}\quad (5.39)$$

where we have taken advantage of $\epsilon \rightarrow 0$. Evaluating the derivative in parentheses gives

$$\begin{aligned}\frac{d}{d\epsilon} (\nabla_{\mathbf{a}} \Phi_\epsilon(\mathbf{a})^T)^{-1} \Big|_{\epsilon=0} &= -(\nabla_{\mathbf{a}} \Phi_\epsilon(\mathbf{a})^T)^{-2} \frac{d}{d\epsilon} (\nabla_{\mathbf{a}} \Phi_\epsilon(\mathbf{a})^T) \Big|_{\epsilon=0} \\ &= -\mathbf{I} (\nabla \mathbf{V}(\mathbf{a})^T) = -(\nabla \mathbf{V}(\mathbf{a}))^T,\end{aligned}\quad (5.40)$$

where we have used Lemma 5.1. Combining (5.39) and (5.40), we get the first part of (5.38).

As for the second part in (5.38), we make use of (5.23):

$$\begin{aligned}\nu' &= \dot{\nu} - (\mathbf{V} \cdot \nabla_{\Gamma}) \nu = -[(\nabla_{\Gamma} \mathbf{V})^T \nu + (\mathbf{V} \cdot \nabla_{\Gamma}) \nu] \\ &= -[(\nabla_{\Gamma} \mathbf{V})^T \nu + (\nabla_{\Gamma} \nu) \mathbf{V}] = -[(\nabla_{\Gamma} \mathbf{V})^T \nu + (\nabla_{\Gamma} \nu)^T \mathbf{V}] \\ &= -\nabla_{\Gamma} (\mathbf{V} \cdot \nu)^T,\end{aligned}\quad (5.41)$$

where we have used the fact that $\nabla_{\Gamma} \nu$ is a symmetric matrix. \square

Summed Curvature

Lemma 5.6 (shape derivative of κ). *Assume the hypothesis in Lemma 5.2. Then*

$$\dot{\kappa} = -\nu \cdot (\Delta_{\Gamma} \mathbf{V}) - 2(\nabla_{\Gamma} \mathbf{V}) : (\nabla_{\Gamma} \nu), \quad \kappa' = -\Delta_{\Gamma} (\mathbf{V} \cdot \nu), \quad (5.42)$$

evaluated on Γ .

Proof. We shall take advantage of Lemma 5.5. Recall from Proposition 23 of Chapter 4 that $\kappa = \nabla_{\Gamma} \cdot \nu$. By extending the normal vector from Γ to Ω_{all} by using a constant normal extension (section 5.4.1), we can write the previous formula as $\kappa = \nabla \cdot \nu$. The advantage here is that ∇ does not depend on the shape! Therefore,

$$\begin{aligned}\kappa' &= \nabla \cdot \nu' = \nabla \cdot (-\nabla_{\Gamma} (\mathbf{V} \cdot \nu)^T) \\ &= -\nabla \cdot \nabla_{\Gamma} (\mathbf{V} \cdot \nu) = -\nabla_{\Gamma} \cdot \nabla_{\Gamma} (\mathbf{V} \cdot \nu) \\ &= -\Delta_{\Gamma} (\mathbf{V} \cdot \nu).\end{aligned}$$

Next, let us expand κ' :

$$\begin{aligned}\kappa' &= -\nabla_{\Gamma} \cdot \nabla_{\Gamma} (\mathbf{V} \cdot \nu) = -\nabla_{\Gamma} \cdot [(\nabla_{\Gamma} \mathbf{V})^T \nu + (\nabla_{\Gamma} \nu)^T \mathbf{V}] \\ &= -[(\Delta_{\Gamma} \mathbf{V}) \cdot \nu + 2(\nabla_{\Gamma} \mathbf{V}) : (\nabla_{\Gamma} \nu) + (\nabla_{\Gamma} \cdot (\nabla_{\Gamma} \nu)^T) \cdot \mathbf{V}] \\ &= -[(\Delta_{\Gamma} \mathbf{V}) \cdot \nu + 2(\nabla_{\Gamma} \mathbf{V}) : (\nabla_{\Gamma} \nu) + (\mathbf{V} \cdot \nabla_{\Gamma}) (\nabla_{\Gamma} \cdot \nu)],\end{aligned}$$

where we have used $[\nabla_{\Gamma} \boldsymbol{\nu}] = [\nabla_{\Gamma} \boldsymbol{\nu}]^T$. Thus, by (5.23), we have

$$\begin{aligned}\dot{\kappa} &= \kappa' + (\mathbf{V} \cdot \nabla_{\Gamma}) \kappa \\ &= -[\boldsymbol{\nu} \cdot (\Delta_{\Gamma} \mathbf{V}) + 2(\nabla_{\Gamma} \mathbf{V}) : (\nabla_{\Gamma} \boldsymbol{\nu})] - (\mathbf{V} \cdot \nabla_{\Gamma})(\nabla_{\Gamma} \cdot \boldsymbol{\nu}) + (\mathbf{V} \cdot \nabla_{\Gamma}) \kappa \\ &= -[\boldsymbol{\nu} \cdot (\Delta_{\Gamma} \mathbf{V}) + 2(\nabla_{\Gamma} \mathbf{V}) : (\nabla_{\Gamma} \boldsymbol{\nu})],\end{aligned}$$

which is the assertion. \square

5.6 • Shape Perturbation of Functionals

Let D be a general domain and \mathcal{J} be a functional that depends on D . Moreover, let D_{ϵ} be a variation of D that depends smoothly on $\epsilon \geq 0$ such that $D = D_0$ (see earlier sections). Define \mathcal{J}_{ϵ} by

$$\mathcal{J}_{\epsilon} \equiv \mathcal{J}(D_{\epsilon}) := \int_{D_{\epsilon}} f(\epsilon, \mathbf{x}) d\mathbf{x}, \quad (5.43)$$

where $\mathcal{J} = \mathcal{J}_0$ and $f : [0, \epsilon_{\max}] \times D_{\epsilon} \rightarrow \mathbb{R}$; f could also be defined over a fixed domain containing D_{ϵ} . Sometimes we write $f(D) \equiv f$ to emphasize that f depends on the domain (shape) D (i.e., D is an *independent* variable), or we write $f(D_{\epsilon}, \mathbf{x}) \equiv f(\epsilon, \mathbf{x})$ to emphasize dependence on the perturbed domain D_{ϵ} .

With the above considerations, we define the **shape perturbation** (or **shape sensitivity**) of $\mathcal{J} \equiv \mathcal{J}(D)$ as

$$\delta \mathcal{J}(D) := \frac{d}{d\epsilon} \mathcal{J}(D_{\epsilon}) \Big|_{\epsilon=0^+}. \quad (5.44)$$

5.6.1 • Shape Perturbation Formulas

Recall the domains and functionals from section 5.2.1. In particular, consider the following bulk and boundary functionals:

$$\mathcal{E}_{\epsilon} = \int_{\Omega_{\epsilon}} f(\Omega_{\epsilon}), \quad \mathcal{J}_{\epsilon} = \int_{\Gamma_{\epsilon}} f(\Omega_{\epsilon}), \quad \mathcal{B}_{\epsilon} = \int_{\Gamma_{\epsilon}} g(\Gamma_{\epsilon}) \quad \text{for all } \epsilon \geq 0, \quad (5.45)$$

where $f(\Omega_{\epsilon})(\cdot) : \Omega_{\epsilon} \rightarrow \mathbb{R}$, $g(\Gamma_{\epsilon})(\cdot) : \Gamma_{\epsilon} \rightarrow \mathbb{R}$, and Ω_{ϵ} and Γ_{ϵ} are defined in terms of the velocity perturbation \mathbf{V} in (5.8) and (5.9). Note that $\Gamma_{\epsilon} \subset \bar{\Omega}_{\epsilon}$ for all $\epsilon \geq 0$.

Let the shape perturbations of $\mathcal{E} \equiv \mathcal{E}_0$, $\mathcal{J} \equiv \mathcal{J}_0$, and $\mathcal{B} \equiv \mathcal{B}_0$, in the direction \mathbf{V} , be defined by

$$\begin{aligned}\delta \mathcal{E}(\Omega) \cdot \mathbf{V} &\equiv \delta \mathcal{E}(\Omega; \mathbf{V}) := \frac{d}{d\epsilon} \mathcal{E}_{\epsilon} \Big|_{\epsilon=0^+}, \\ \delta \mathcal{J}(\Gamma) \cdot \mathbf{V} &\equiv \delta \mathcal{J}(\Gamma; \mathbf{V}) := \frac{d}{d\epsilon} \mathcal{J}_{\epsilon} \Big|_{\epsilon=0^+}, \\ \delta \mathcal{B}(\Gamma) \cdot \mathbf{V} &\equiv \delta \mathcal{B}(\Gamma; \mathbf{V}) := \frac{d}{d\epsilon} \mathcal{B}_{\epsilon} \Big|_{\epsilon=0^+}.\end{aligned} \quad (5.46)$$

Thus, we obtain the main result of this chapter.

Lemma 5.7 (shape perturbations of functionals). *Adopt Assumption 3 and consider the functionals given in (5.45). Then the shape perturbations defined by (5.46) can be written*

more explicitly as

$$\begin{aligned}\delta \mathcal{E}(\Omega; \mathbf{V}) &= \int_{\Omega} \dot{f}(\Omega; \mathbf{V}) + \int_{\Omega} f(\Omega)(\nabla \cdot \mathbf{V}) \\ &= \int_{\Omega} f'(\Omega; \mathbf{V}) + \int_{\partial\Omega} f(\Omega)(\mathbf{V} \cdot \boldsymbol{\nu}),\end{aligned}\tag{5.47}$$

$$\begin{aligned}\delta \mathcal{J}(\Gamma; \mathbf{V}) &= \int_{\Gamma} \dot{f}(\Omega; \mathbf{V}) + f(\nabla_{\Gamma} \cdot \mathbf{V}) \\ &= \int_{\Gamma} f'(\Omega; \mathbf{V}) + (\mathbf{V} \cdot \nabla)f + f(\nabla_{\Gamma} \cdot \mathbf{V}) \\ &= \int_{\Gamma} f'(\Omega; \mathbf{V}) + [(\boldsymbol{\nu} \cdot \nabla)f + f\kappa](\mathbf{V} \cdot \boldsymbol{\nu}) + \int_{\partial\Gamma} f(\boldsymbol{\tau} \times \boldsymbol{\nu}) \cdot \mathbf{V},\end{aligned}\tag{5.48}$$

$$\begin{aligned}\delta \mathcal{B}(\Gamma; \mathbf{V}) &= \int_{\Gamma} \dot{g}(\Gamma; \mathbf{V}) + g(\nabla_{\Gamma} \cdot \mathbf{V}) \\ &= \int_{\Gamma} g'(\Gamma; \mathbf{V}) + (\mathbf{V} \cdot \nabla_{\Gamma})g + g(\nabla_{\Gamma} \cdot \mathbf{V}) \\ &= \int_{\Gamma} g'(\Gamma; \mathbf{V}) + g\kappa(\mathbf{V} \cdot \boldsymbol{\nu}) + \int_{\partial\Gamma} g(\boldsymbol{\tau} \times \boldsymbol{\nu}) \cdot \mathbf{V},\end{aligned}\tag{5.49}$$

where $\boldsymbol{\nu}$ is the unit normal vector of Γ and $\boldsymbol{\tau}$ is the unit tangent vector of $\partial\Gamma$ (see Proposition 14).

Proof. To show (5.47), we apply the change of variables $\mathbf{y} = \Phi_{\epsilon}(\mathbf{x})$:

$$\int_{\Omega_{\epsilon}} f(\epsilon, \mathbf{y}) d\mathbf{y} = \int_{\Omega} f(\epsilon, \Phi_{\epsilon}(\mathbf{x})) \det(\nabla \Phi_{\epsilon}(\mathbf{x})) d\mathbf{x},$$

and differentiate

$$\begin{aligned}\frac{d}{d\epsilon} \left(\int_{\Omega_{\epsilon}} f(\epsilon, \mathbf{y}) d\mathbf{y} \right) &= \int_{\Omega} \frac{d}{d\epsilon} [f(\epsilon, \Phi_{\epsilon}(\mathbf{x}))] \det(\nabla \Phi_{\epsilon}(\mathbf{x})) d\mathbf{x} \\ &\quad + \int_{\Omega} f(\epsilon, \Phi_{\epsilon}(\mathbf{x})) \frac{d}{d\epsilon} \det(\nabla \Phi_{\epsilon}(\mathbf{x})) d\mathbf{x}.\end{aligned}$$

Then,

$$\delta \mathcal{E}(\Omega; \mathbf{V}) = \frac{d}{d\epsilon} \left(\int_{\Omega_{\epsilon}} f(\epsilon, \mathbf{y}) d\mathbf{y} \right) \Big|_{\epsilon=0^+} = \int_{\Omega} \left\{ \dot{f}(\mathbf{x}) + f(\mathbf{x})(\nabla \cdot \mathbf{V}(\mathbf{x})) \right\} d\mathbf{x},$$

where we have used Lemma 5.1. Using (5.14) and the divergence theorem, we obtain

$$\begin{aligned}\delta \mathcal{E}(\Omega; \mathbf{V}) &= \int_{\Omega} \left\{ f'(\mathbf{x}) + (\mathbf{V}(\mathbf{x}) \cdot \nabla)f(\mathbf{x}) + f(\mathbf{x})(\nabla \cdot \mathbf{V}(\mathbf{x})) \right\} d\mathbf{x} \\ &= \int_{\Omega} f'(\mathbf{x}) d\mathbf{x} + \int_{\Omega} \nabla \cdot (f\mathbf{V}) d\mathbf{x} = \int_{\Omega} f'(\mathbf{x}) d\mathbf{x} + \int_{\partial\Omega} f(\mathbf{x})(\mathbf{V} \cdot \boldsymbol{\nu}) dS(\mathbf{x}).\end{aligned}$$

Next, we derive (5.48). Let \mathbf{X}_ϵ be given by (5.9). Assume that $\mathbf{X}_\epsilon : \cup_i U_i \rightarrow \mathbb{R}^3$ (i.e., a 2-D surface in three dimensions), with local coordinates in the reference domain U_i denoted (s_1, s_2) . Suppose \mathbf{V} has compact support on Γ , i.e., $\mathbf{V} \circ \mathbf{X}$ has compact support (see section 2.1.3) in U_i for some i . Then, the change of variables $\mathbf{x} = \mathbf{X}_\epsilon(s_1, s_2)$ gives

$$\mathcal{J}_\epsilon = \int_{\Gamma_\epsilon} f(\epsilon, \mathbf{x}) dS(\mathbf{x}) = \int_{U_i} f(\epsilon, \mathbf{X}_\epsilon(s_1, s_2)) |\partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon| ds_1 ds_2.$$

Differentiating yields

$$\begin{aligned} \frac{d}{d\epsilon} \mathcal{J}_\epsilon &= \int_{U_i} \left(\frac{d}{d\epsilon} f(\epsilon, \mathbf{X}_\epsilon(s_1, s_2)) \right) |\partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon| ds_1 ds_2 \\ &\quad + \int_{U_i} f(\epsilon, \mathbf{X}_\epsilon(s_1, s_2)) \frac{d}{d\epsilon} |\partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon| ds_1 ds_2. \end{aligned}$$

The first term gives

$$\frac{d}{d\epsilon} f(\epsilon, \mathbf{X}_\epsilon(s_1, s_2)) \Big|_{\epsilon=0} = \dot{f} \circ \mathbf{X}(s_1, s_2),$$

and the second term follows from identity (5) of Lemma 5.2:

$$\frac{d}{d\epsilon} \left| \partial_{s_1} \mathbf{X}_\epsilon \times \partial_{s_2} \mathbf{X}_\epsilon \right| \Big|_{\epsilon=0} = \{\nabla_\Gamma \text{id}_\Gamma \circ \mathbf{X}\} : \{\nabla_\Gamma \mathbf{V} \circ \mathbf{X}\} |\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}|.$$

Taking $\epsilon \rightarrow 0^+$, combining these results, and changing the variables back to Γ leads to

$$\delta \mathcal{J}(\Gamma; \mathbf{V}) := \frac{d}{d\epsilon} \mathcal{J}_\epsilon \Big|_{\epsilon=0^+} = \int_\Gamma \dot{f}(\mathbf{x}) dS(\mathbf{x}) + \int_\Gamma f(\mathbf{x}) \underbrace{(\nabla_\Gamma \text{id}_\Gamma(\mathbf{x})) : (\nabla_\Gamma \mathbf{V})}_{=\nabla_\Gamma \cdot \mathbf{V}} dS(\mathbf{x}),$$

which gives the first line of (5.48) for a compact \mathbf{V} . The same formula holds for an arbitrary \mathbf{V} by a partition of unity argument; see the proof of Proposition 14 in Chapter 4.

The second line of (5.48) follows from (5.14). The third line follows by expanding $(\mathbf{V} \cdot \nabla) f$ into normal and tangential components: $(\mathbf{V} \cdot \nabla) f = (\mathbf{V} \cdot \boldsymbol{\nu})(\boldsymbol{\nu} \cdot \nabla) f + (\mathbf{V} \cdot \nabla_\Gamma) f$ and

$$\begin{aligned} \int_\Gamma (\mathbf{V} \cdot \nabla) f + f(\nabla_\Gamma \cdot \mathbf{V}) &= \int_\Gamma (\mathbf{V} \cdot \boldsymbol{\nu})(\boldsymbol{\nu} \cdot \nabla) f + (\mathbf{V} \cdot \nabla_\Gamma) f + f(\nabla_\Gamma \cdot \mathbf{V}) \\ &= \int_\Gamma (\mathbf{V} \cdot \boldsymbol{\nu})(\boldsymbol{\nu} \cdot \nabla) f + \int_\Gamma \nabla_\Gamma \cdot (f \mathbf{V}) \\ &= \int_\Gamma (\mathbf{V} \cdot \boldsymbol{\nu})(\boldsymbol{\nu} \cdot \nabla) f + \int_\Gamma f \kappa(\mathbf{V} \cdot \boldsymbol{\nu}) + \int_{\partial\Gamma} f(\boldsymbol{\tau} \times \boldsymbol{\nu}) \cdot \mathbf{V}, \end{aligned}$$

where we have used (4.35); note that the surface Γ may have a nonempty boundary $\partial\Gamma$.

Finally, (5.49) follows directly from (5.48) by letting f be an extension of g such that the extension is *constant* in the normal direction (see section 5.4.1). \square

5.6.2 • Structure of Shape Perturbations

The following theorem states that shape perturbations of bulk functionals depend only on boundary quantities.

Theorem 5.8. Assume Ω is C^{k+1} for $k \geq 0$. Let $\mathcal{J}(\Omega)$ be a shape functional, whose shape perturbation (5.46) exists, such that

$$\delta \mathcal{J}(\Omega; \mathbf{V}) = \langle G(\Omega), \mathbf{V} \rangle_{M^*, M} \quad \text{for all } \mathbf{V} \text{ in } M,$$

where $M = D^k(\mathbb{R}^n, \mathbb{R}^n)$ (smooth test functions), and $G(\Omega)$ is in the dual space M^* (space of distributions). Then there exists $g(\partial\Omega)$ in E^* , where $E = D^k(\partial\Omega)$, such that

$$\delta \mathcal{J}(\Omega; \mathbf{V}) = \langle g, \mathbf{V} \cdot \boldsymbol{\nu} \rangle_{E^*, E},$$

i.e., the shape gradient g is a measure concentrated on the boundary $\partial\Omega$. Moreover, suppose g is more regular, i.e., g is in $L^1(\partial\Omega)$. Then,

$$\delta \mathcal{J}(\Omega; \mathbf{V}) = \int_{\partial\Omega} g(\mathbf{V} \cdot \boldsymbol{\nu}).$$

Proof. See [23, 93]. □

Remark 25. By (5.47), we have $\delta \mathcal{J}(\Omega; \mathbf{V}) = \int_{\Omega} f'(\Omega; \mathbf{V}) + \int_{\partial\Omega} f(\Omega)(\mathbf{V} \cdot \boldsymbol{\nu})$. Theorem 5.8 implies that one can rewrite the term $\int_{\Omega} f'(\Omega; \mathbf{V})$ into an integral over $\partial\Omega$. Formally, this entails integration by parts (and other manipulations). For problems where f is the solution of a PDE on Ω , and f' is its shape derivative, solutions of adjoint PDEs are necessary for the manipulations [46, 51, 54, 59, 84, 106, 107].

Chapter 6

Applications

In this chapter, we give some applications of shape differential calculus and describe a strategy for minimizing shape functionals. Some applications exhibit some kind of *free energy*. In this case, physical forces can be obtained by differentiating the energy with respect to the variable of interest, which may be the domain itself.

6.1 • Minimal Surfaces

Minimal surfaces are perhaps the most famous example of differential geometry in combination with the calculus of variations [18, 24] (see Figure 6.1 for some examples of minimal surfaces). The concept arises in soap films, bubbles, surface tension, membranes, and more [53, 76]. The tools we have developed with shape differential calculus make the derivation of the minimal surface equation trivial.

6.1.1 • Problem Formulation

There are many versions of minimal surfaces. We shall consider only the following. Let $\Gamma \subset \mathbb{R}^3$ be a regular surface with smooth boundary $\partial\Gamma$, and consider the area functional

$$\mathcal{J}(\Gamma) = \int_{\Gamma} 1 dS. \quad (6.1)$$

The surface Γ can be arbitrary, but we assume that $\partial\Gamma$ is fixed. In other words, we define a fixed closed curve $\Sigma \subset \mathbb{R}^3$ and consider all regular surfaces Γ such that $\partial\Gamma \equiv \Sigma$. One can think of a closed loop of wire (i.e., Σ) with a soap film solution draped over it (Γ is the soap film).

6.1.2 • Shape Perturbation

Imagine we compute the surface area of every possible surface Γ , such that its boundary matches Σ , and we find the one with the *smallest surface area*. More formally, we want to find a minimizer of (6.1) amongst all smooth surfaces with boundary given by Σ . Recalling basic calculus, the rule for finding the minimum of a differentiable function is that the slope of the function is zero. In the present context, finding the minimum of $\mathcal{J}(\Gamma)$ means that the shape perturbation of $\mathcal{J}(\Gamma)$ is zero for any smooth perturbation that keeps the boundary Σ fixed.

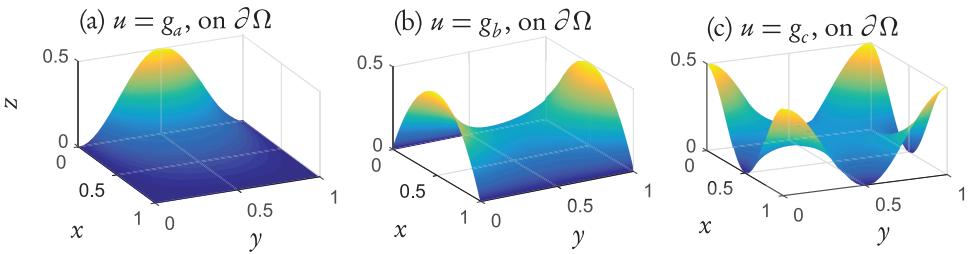


Figure 6.1. Minimal surfaces from given boundary data. Each surface is a solution of (6.5) on $\Omega = [0, 1] \times [0, 1]$ (unit square). (a) $g_a(x, y) = (-\cos(2\pi y) + 1)/4$ for $x = 0$, and $g_a(x, y) = 0$ otherwise. (b) $g_b(x, y) = 2x(1-x)$. (c) $g_c(x, y) = (\cos(2\pi x)\cos(2\pi y) + 1)/4$.

To this end, let $\mathbf{V} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ be a smooth vector field that vanishes on Σ . Using (5.48) and setting $f = 1$, we compute the shape perturbation of \mathcal{J} ,

$$\begin{aligned}\delta \mathcal{J}(\Gamma; \mathbf{V}) &= \int_{\Gamma} f' + [(\boldsymbol{\nu} \cdot \nabla) f + f \kappa](\mathbf{V} \cdot \boldsymbol{\nu}) + \int_{\Sigma} f(\boldsymbol{\tau} \times \boldsymbol{\nu}) \cdot \mathbf{V}, \\ &= \int_{\Gamma} \kappa \boldsymbol{\nu} \cdot \mathbf{V} = \int_{\Gamma} \kappa \cdot \mathbf{V},\end{aligned}\tag{6.2}$$

where $\boldsymbol{\nu}$ is the unit normal of Γ . (Note that $1' = 0$, because clearly the function 1 does not depend on shape!) Thus, if Γ is a surface with *minimal surface area*, then $\delta \mathcal{J}(\Gamma; \mathbf{V}) = 0$ for all smooth vector fields \mathbf{V} that vanish on Σ .

6.1.3 • Characterization

We can extract a bit more information. Let us choose a particular form for \mathbf{V} . Let $\phi : \mathbb{R}^3 \rightarrow \mathbb{R}$ be arbitrary but satisfy $\phi \geq 0$ with $\phi = 0$ on Σ . Assuming we know Γ , choose \mathbf{V} such that $\mathbf{V} = \phi \kappa$ on Γ . Away from Γ , \mathbf{V} is smoothly extended; for simplicity, we will not detail this extension. Plugging this into (6.2) and setting (6.2) to zero, we obtain

$$\int_{\Gamma} \phi |\kappa|^2 = 0 \quad \text{for all } \phi \geq 0.\tag{6.3}$$

Choosing $\phi > 0$ away from Σ , we see that the condition (6.3) implies that $|\kappa| = |\kappa| = 0$.

Therefore, if Γ is a smooth surface with smooth boundary Σ , and Γ has minimal surface area while keeping Σ fixed, then the sum of the principal curvatures must be zero, i.e., $\kappa = 0$ everywhere on Γ . The converse is not necessarily true. Given a surface whose $\kappa = 0$, we cannot conclude that the surface has minimal area; it may be a maximum or some other extremal point. Just like in basic calculus, one must check second order conditions. This involves computing a **second** shape perturbation of (6.2) and analyzing the resultant bilinear form. But, in order to keep the development here as simple as possible, we will not do this.

6.1.4 • Summary

Despite the previous discussion, for historical reasons, we define a minimal surface in terms of $\kappa = 0$.

Definition 16. A regular surface $\Gamma \subset \mathbb{R}^3$ is called a *minimal surface* if and only if the summed curvature $\kappa = 0$ everywhere on Γ .

We have barely touched on the theory of minimal surfaces (see [18] for an extensive introduction). When looking more deeply, many issues arise. The surface may not be smooth but piecewise smooth. Simple experiments with soap films demonstrate this [53, 76]. Furthermore, it is not obvious that any shape of closed curve Σ will always yield a well-defined minimal surface.

We close this section with some exercises.

- Suppose we parameterize a surface by the graph of a smooth function $u = u(x, y)$ defined on $\Omega \subset \mathbb{R}^2$. Show that the summed curvature is given by the formula

$$\kappa = -\nabla \cdot \left(\frac{\nabla u}{\sqrt{1 + |\nabla u|^2}} \right), \quad (6.4)$$

where $\nabla = (\partial_x, \partial_y)$. The minus sign is chosen so that $u(x, y) = 1 - x^2 - y^2$ has $\kappa > 0$. Thus, a minimal surface satisfies the nonlinear second order PDE

$$\begin{aligned} -\nabla \cdot \left(\frac{\nabla u}{\sqrt{1 + |\nabla u|^2}} \right) &= 0 \quad \text{in } \Omega, \\ u &= g \quad \text{on } \partial\Omega, \end{aligned} \quad (6.5)$$

where $g : \Omega \rightarrow \mathbb{R}$ is a function that specifies the height of the boundary. This is known as the *prescribed mean curvature equation*. Note that the surfaces in Figure 6.1 are solutions of this equation.

- Show that the catenoid, parameterized by

$$\mathbf{X}(s_1, s_2) = a(\cos s_1 \cosh s_2, \sin s_1 \cosh s_2, s_2)^T, \quad 0 < s_1 < 2\pi, \quad -\infty < s_2 < \infty, \quad (6.6)$$

is a minimal surface for any value of the constant $a \neq 0$.

- Show that the helicoid, parameterized by

$$\mathbf{X}(s_1, s_2) = a(\cos s_1 \sinh s_2, \sin s_1 \sinh s_2, s_1)^T, \quad 0 < s_1 < 2\pi, \quad -\infty < s_2 < \infty, \quad (6.7)$$

is a minimal surface for any value of the constant $a \neq 0$.

6.2 ■ Surface Tension

Assume Γ is a regular surface. Consider the surface energy functional [20, 21],

$$\mathcal{J}(\Gamma) = \int_{\Gamma} \gamma(\mathbf{x}) dS(\mathbf{x}), \quad (6.8)$$

which arises in problems involving surface tension. The coefficient γ is the surface tension coefficient and may be variable.

The surface tension force is obtained by computing the shape perturbation of \mathcal{J} . For simplicity, assume $\gamma = c$ (constant). By (5.49), we have

$$\delta \mathcal{J}(\Gamma; \mathbf{V}) = \int_{\Gamma} \gamma \kappa \boldsymbol{\nu} \cdot \mathbf{V} + \int_{\partial\Gamma} \gamma (\boldsymbol{\tau} \times \boldsymbol{\nu}) \cdot \mathbf{V}, \quad (6.9)$$

where \mathbf{V} is a perturbation of the surface Γ . The quantity $\delta \mathcal{J}(\Gamma; \mathbf{V})$ is the relative force exerted by the surface tension against deforming the surface by \mathbf{V} . Thus a flat surface (i.e., $\kappa = 0$) is easy to deform (to first order) with a compact perturbation \mathbf{V} .

In continuum mechanics, the term $\gamma \kappa \nu$ is a surface stress that is usually balanced against quantities such as elastic or fluid stresses that are in contact with the surface. If the boundary $\partial\Gamma$ is not empty, then the second term in (6.9) is balanced against some kind of boundary effect. For instance, if $\partial\Gamma$ is “pinned” (not free to move), then $\mathbf{V} = \mathbf{0}$ on $\partial\Gamma$ and the second term vanishes. If it is free to move, then $\gamma(\tau \times \nu)$ is usually balanced by some additional forces, such as substrate surface tensions (see section 6.2.2).

6.2.1 • Floating Droplets

Lagrange Multipliers

In the sole presence of surface tension forces, the equilibrium shape of a free-floating droplet is captured by minimizers of (6.8). But one must include a volume constraint when minimizing \mathcal{J} because many liquids (e.g., water) are incompressible. This is simple to do using Lagrange multipliers.

Let Ω be the domain of the droplet with boundary $\Gamma = \partial\Omega$; note that $\partial\Gamma = \emptyset$. For simplicity, take $\gamma(\mathbf{x}) = 1$, and define a Lagrangian shape functional,

$$\mathcal{L}(\Omega, p) = \int_{\Gamma} 1 dS - p \left(\int_{\Omega} 1 d\mathbf{x} - C \right), \quad (6.10)$$

where $|\Omega| \equiv \int_{\Omega} 1 d\mathbf{x} = C$ represents the volume constraint for some constant $C > 0$. The Lagrange multiplier is $p \in \mathbb{R}$ (constant) and represents the equilibrium physical pressure inside the droplet.

In order for Ω to be a minimizer of $\int_{\Gamma} 1 dS$, subject to the volume constraint, the following first order optimality conditions must be satisfied [54]:

$$\begin{aligned} \delta_{\Omega} \mathcal{L}(\Omega, p; \mathbf{V}) &= \int_{\Gamma} \kappa \nu \cdot \mathbf{V} - p \int_{\Gamma} \mathbf{V} \cdot \nu \equiv 0, \\ \partial_p \mathcal{L}(\Omega, p) &= \int_{\Omega} 1 d\mathbf{x} - C \equiv 0, \end{aligned} \quad (6.11)$$

where we have used (5.47) and \mathbf{V} is an arbitrary smooth perturbation of Ω . We use δ_{Ω} to emphasize on which argument the shape perturbation is acting; note that ∂_p is a standard partial derivative.

In particular, we have

$$\int_{\Gamma} (\kappa - p) \mathbf{V} \cdot \nu = 0 \quad \text{for all smooth perturbations } \mathbf{V}.$$

Proceeding as in section 6.1.3, choose $\mathbf{V} = (\kappa - p)\nu$. This gives $\int_{\Gamma} |\kappa - p|^2 = 0$, which means

$$\kappa = p, \quad \text{i.e., the summed curvature is a constant.} \quad (6.12)$$

Since Γ is a surface without boundary, and p is assumed positive, there is only one shape that satisfies this: a sphere of radius R . The particular value of R comes from the volume constraint, i.e., $(4/3)\pi R^3 = C$.

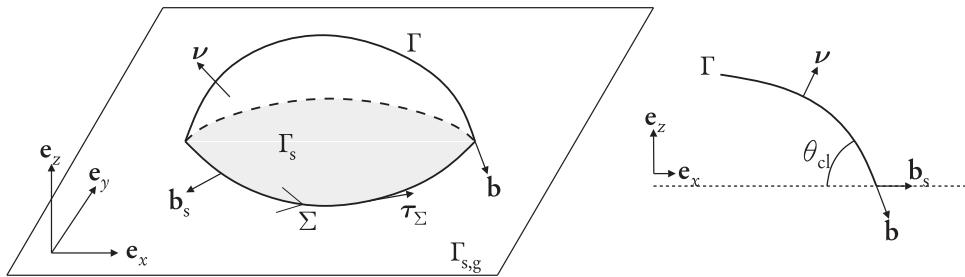


Figure 6.2. Illustration of a droplet on a flat solid substrate [66]. The volume region of the droplet is denoted Ω , and the boundary decomposes as $\partial\bar{\Omega} = \bar{\Gamma} \cup \bar{\Gamma}_s$, where Γ is the liquid-gas interface and Γ_s is the liquid-solid interface (shaded). The solid-gas interface is labeled $\Gamma_{s,g}$. The plane of the substrate is denoted \mathcal{P} , where $\bar{\Gamma}_s, \bar{\Gamma}_{s,g} \subset \mathcal{P}$. The contact line is denoted Σ , defined by $\Sigma = \bar{\Gamma} \cap \bar{\Gamma}_s$, and oriented with unit tangent vector τ_Σ . The unit outer normal to Ω is denoted ν on Γ and $\nu_s \equiv -\mathbf{e}_z$ on Γ_s . The open surface Γ has outward unit boundary vector \mathbf{b} defined only on $\Sigma \equiv \partial\Gamma$; likewise, Γ_s has boundary vector \mathbf{b}_s which points in the \mathcal{P} plane. The contact angle at Σ is denoted θ_{cl} , where $\cos\theta_{cl} = \mathbf{b} \cdot \mathbf{b}_s$.

Isoperimetric Inequality

We have essentially deduced the isoperimetric inequality [24].

Theorem 6.1. Let \mathcal{A} be the set of all domains $\Omega \subset \mathbb{R}^3$, with smooth boundary, such that $|\Omega| = (4/3)\pi R^3$ for some fixed constant $R > 0$. Then the unique shape Ω in \mathcal{A} with the smallest surface area is the sphere of radius R .

A more in-depth look at the mathematical details can be found in [4, 45, 109]. This result is consistent with experimental observations of droplets in microgravity environments.

6.2.2 • Droplets on Substrates

The origin of surface tension is when two different material phases meet at an interface (e.g., oil and water). In section 6.2.1, we had a droplet of liquid (e.g., water) surrounded by a gas (e.g., air). Now we consider droplets in contact with a rigid substrate (see Figure 6.2), which have different equilibrium shapes than floating droplets. This is because the substrate is another material phase that introduces additional interfaces, each with its own surface tension coefficient [11, 20, 21, 78, 80, 111]. Hence, the free energy functional that governs the equilibrium shape includes additional terms relative to (6.8). We now deduce the equations that govern the equilibrium shape of droplets bound to a substrate.

Free Energy

Let Ω be a 3-D droplet sitting on a flat substrate as shown in Figure 6.2. The entire plane of the substrate is denoted $\mathcal{P} := \mathbb{R}^2 \times \{z = 0\}$. The relevant free energy for this problem is [89, Chap. 2]

$$\mathcal{A}(\Omega) = \int_{\Gamma_s} \gamma_s + \int_{\Gamma} \gamma + \int_{\Gamma_{s,g}} \gamma_{s,g} - \rho \mathbf{g} \cdot \int_{\Omega} (\mathbf{x} - \mathbf{x}_0) d\mathbf{x}, \quad (6.13)$$

where \mathbf{g} is the vector acceleration due to gravity and $\gamma_s, \gamma_{s,g} : \mathcal{P} \rightarrow \mathbb{R}$ and $\gamma : \Gamma \rightarrow \mathbb{R}$ are surface tension coefficients for the different interfaces (note that $\Gamma_{s,g}$ is a bounded set). The first three terms in (6.13) are surface tension terms, while the last term is the potential energy of the droplet due to gravity. Note that ρ is the density of the liquid, and \mathbf{x}_0 is an arbitrary reference point.

The “dry” part of \mathcal{P} is denoted by $\Gamma_{s,g}$, and the “wet” part of \mathcal{P} is Γ_s . Most material surfaces have a surface tension coefficient that depends on the adjoining material [20, 21], i.e., the surface tension in the solid-gas region, $\gamma_{s,g}$, is different from the surface tension in the liquid-solid region, γ_s . Note that $\gamma_{s,g}$ and γ_s are *functions* defined in the entire plane but are active only in $\Gamma_{s,g}$ and Γ_s , respectively. The presence of $\gamma_{s,g}$ and γ_s plays an important role in determining the equilibrium shape of Ω (see the boundary condition in (6.21)).

Equilibrium Conditions

To facilitate deriving the equilibrium equations for the *shape* of the droplet, we follow section 6.2.1 and introduce the Lagrangian

$$\mathcal{L}(\Omega, p) = \mathcal{A}(\Omega) - p \left(\int_{\Omega} 1 d\mathbf{x} - C \right), \quad (6.14)$$

which includes the volume constraint $|\Omega| = C$ via the Lagrange multiplier $p \in \mathbb{R}$.

Let $\mathbf{V} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ be a vector field that vanishes at some distance from Ω . In particular, we set $\mathbf{V} = \mathbf{0}$ on $\partial(\overline{\Gamma_s} \cup \overline{\Gamma_{s,g}})$ (the outer boundary of $\Gamma_{s,g}$). We will perturb the domain Ω with \mathbf{V} . Furthermore, we restrict \mathbf{V} such that $\mathbf{V} \cdot \mathbf{e}_z = -\mathbf{V} \cdot \boldsymbol{\nu}_s = 0$ on the plane \mathcal{P} . This is because the substrate is rigid and cannot be deformed. Next, apply the second lines of (5.47) and (5.49) to compute the shape derivative of $\mathcal{L}(\Omega, p)$:

$$\begin{aligned} \delta_{\Omega} \mathcal{L}(\Omega, p; \mathbf{V}) &= \int_{\Gamma_s} \nabla_{\Gamma_s} \cdot (\gamma_s \mathbf{V}) + \int_{\Gamma} \nabla_{\Gamma} \cdot (\gamma \mathbf{V}) + \int_{\Gamma_{s,g}} \nabla_{\Gamma_{s,g}} \cdot (\gamma_{s,g} \mathbf{V}) \\ &\quad - \rho \mathbf{g} \cdot \int_{\partial\Omega} (\mathbf{x} - \mathbf{x}_0) (\mathbf{V} \cdot \boldsymbol{\nu}) - p \int_{\Gamma} \mathbf{V} \cdot \boldsymbol{\nu}, \end{aligned} \quad (6.15)$$

where we have used the fact that $\gamma_s, \gamma, \gamma_{s,g}$, and \mathbf{g} are independent of the shape Ω . For simplicity, assume $\gamma \in \mathbb{R}$ (constant surface tension); this is typical of most liquids without surfactants. Next, after integration by parts (applying Proposition 15) and using $\mathbf{V} \cdot \mathbf{e}_z = 0$, (6.15) reduces to

$$\begin{aligned} \delta_{\Omega} \mathcal{L}(\Omega, p; \mathbf{V}) &= \int_{\partial\Gamma_s} \gamma_s \mathbf{V} \cdot \mathbf{b}_s + \int_{\partial\Gamma} \gamma \mathbf{V} \cdot \mathbf{b} + \int_{\partial\Gamma_{s,g}} \gamma_{s,g} \mathbf{V} \cdot \widehat{\mathbf{b}_{s,g}} \\ &\quad + \int_{\Gamma_s} \gamma_s \kappa_s \boldsymbol{\nu}_s \cdot \mathbf{V} + \int_{\Gamma} \gamma \kappa \boldsymbol{\nu} \cdot \mathbf{V} + \int_{\Gamma_{s,g}} \gamma_{s,g} \kappa_{s,g} \boldsymbol{\nu}_{s,g} \cdot \mathbf{V} \\ &\quad - \rho \mathbf{g} \cdot \int_{\Gamma} (\mathbf{x} - \mathbf{x}_0) (\mathbf{V} \cdot \boldsymbol{\nu}) - p \int_{\Gamma} \mathbf{V} \cdot \boldsymbol{\nu}, \end{aligned} \quad (6.16)$$

where κ_j is the summed curvature of Γ_j ($j = s$ or (s, g)), and $\kappa_j \boldsymbol{\nu}_j = -\Delta_{\Gamma_j} \text{id}_{\Gamma_j}$, where $\text{id}_{\Gamma_j} : \Gamma_j \rightarrow \Gamma_j$ is the identity map on Γ_j (recall (4.29)); we reserve κ to refer to the summed curvature of Γ . Note that \mathbf{V} satisfies $\mathbf{V} \cdot \mathbf{e}_z = 0$ on \mathcal{P} , so then $\boldsymbol{\nu}_{s,g} \cdot \mathbf{V} = \boldsymbol{\nu}_s \cdot \mathbf{V} = 0$ on $\Gamma_{s,g}$

and Γ_s . Since $\Sigma \equiv \partial\Gamma_s \equiv \partial\Gamma \equiv \partial\Gamma_{s,g} \cap \overline{\Omega}$, and $\mathbf{V} = \mathbf{0}$ at the outer boundary of $\Gamma_{s,g}$, we arrive at

$$\delta_\Omega \mathcal{L}(\Omega, p; \mathbf{V}) = \int_\Sigma (\gamma_s - \gamma_{s,g}) \mathbf{V} \cdot \mathbf{b}_s + \int_\Sigma \gamma \mathbf{V} \cdot \mathbf{b} + \int_\Gamma (\gamma \kappa - \rho \mathbf{g} \cdot (\mathbf{x} - \mathbf{x}_0) - p) \mathbf{V} \cdot \boldsymbol{\nu}. \quad (6.17)$$

At equilibrium, we must have $\delta_\Omega \mathcal{L}(\Omega, p; \mathbf{V}) = 0$ for all admissible \mathbf{V} . So take $\mathbf{V} = \phi \boldsymbol{\nu}$, where $\phi : \Gamma \rightarrow \mathbb{R}$ is smooth with compact support on Γ (i.e., $\phi = 0$ on Σ) and plug into (6.17) to obtain

$$\int_\Gamma (\gamma \kappa - \rho \mathbf{g} \cdot (\mathbf{x} - \mathbf{x}_0) - p) \phi = 0 \quad \text{for all smooth, compact } \phi. \quad (6.18)$$

So, by reasoning similar to that in section 6.1.3 [37, 41, 58], we obtain

$$\gamma \kappa - \rho \mathbf{g} \cdot (\mathbf{x} - \mathbf{x}_0) - p = 0 \quad \text{on } \Gamma, \quad (6.19)$$

which is a PDE that determines the shape of Γ . The multiplier $p \in \mathbb{R}$ is determined by the volume constraint $\int_\Omega 1 d\mathbf{x} = C$ (recall (6.11)). Thus, $\gamma \kappa - \rho \mathbf{g} \cdot (\mathbf{x} - \mathbf{x}_0)$ must be a constant (a generalization of (6.12)).

Next, we derive the boundary condition associated with (6.19). Choose \mathbf{V} such that $\mathbf{V} = \phi \mathbf{b}_s$ on Σ , with $\phi : \mathbb{R}^3 \rightarrow \mathbb{R}$ being an arbitrary smooth function. Then, using the equilibrium condition (6.19), equation (6.17) reduces to

$$\delta_\Omega \mathcal{L}(\Omega, p; \mathbf{V}) = \int_\Sigma (\gamma \cos \theta_{cl} + \gamma_s - \gamma_{s,g}) \phi, \quad (6.20)$$

where we have used the fact that $\cos \theta_{cl} = \mathbf{b} \cdot \mathbf{b}_s$. Again, at equilibrium, (6.20) must vanish for all smooth ϕ . Therefore,

$$\gamma \cos \theta_{cl} + \gamma_s - \gamma_{s,g} = 0 \quad \text{on } \Sigma, \quad (6.21)$$

which is the boundary condition we are after; it is known as the classic *Young–Laplace equation* [20, 21, 89]. Figure 6.3 shows the equilibrium shape of a droplet on a substrate where γ_s and $\gamma_{s,g}$ are nonconstant over the substrate. Figure 6.4 plots the contact angle θ_{cl} along the contact line Σ .

Well-posedness

The pair of equations (6.19) and (6.21) forms a closed set of PDEs. Their solution determines Γ because $\kappa \boldsymbol{\nu} \circ \mathbf{X} = -\Delta_\Gamma \mathbf{X}$, and \mathbf{X} parameterizes Γ . However, the equations are *posed on* Γ , which we do not know a priori. In other words, the system (6.19), (6.21) is *geometrically nonlinear*. Therefore, finding a solution of (6.19), (6.21) let alone proving that it is unique, is nontrivial. Several investigations have been done on this type of problem; see, for instance, [4, 14, 19, 39, 45, 57, 71, 87, 94, 98, 109].

Exercises

We close this section with some exercises.

- Verify all the details in deriving (6.19) and (6.21). How would the derivation change if the solid substrate was not flat but still rigid?

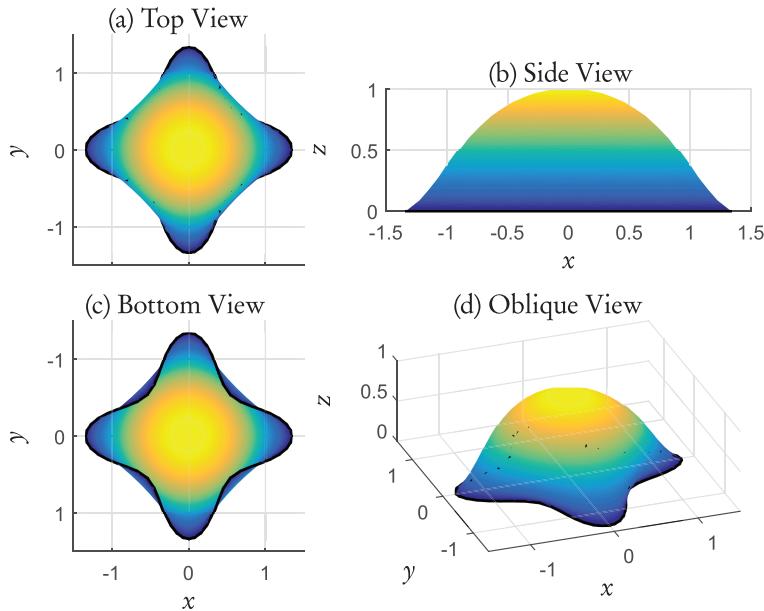


Figure 6.3. A droplet shape that solves (6.19) and (6.21). Color represents height above $z = 0$. The black curve in the x - y plane is the contact line Σ . The liquid-solid interface Γ_s is a four-pointed star. The liquid-gas interface Γ is not a spherical cap because γ_s and $\gamma_{s,g}$ are variable on Σ .

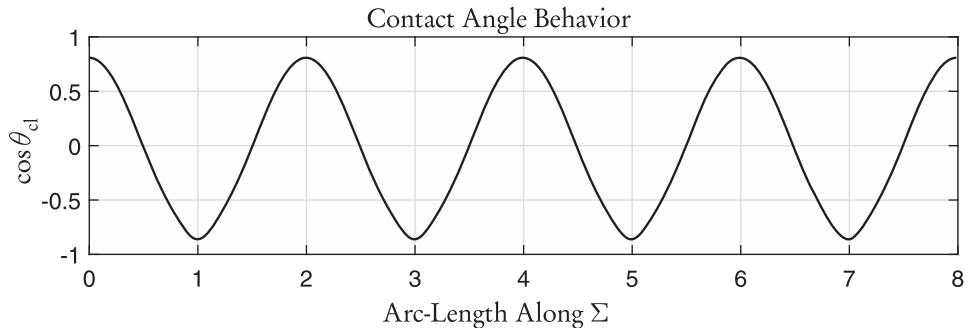


Figure 6.4. Plot of $\cos \theta_{cl}$ vs. arc-length of Σ for the droplet in Figure 6.3. Note that the contact angle θ_{cl} is the angle through the liquid (recall Figure 6.2). The peaks and valleys match the undulations of the contact line Σ .

- Write out the formulas for computing τ_Σ , \mathbf{b}_s , and \mathbf{b} in terms of the parameterization \mathbf{X} .
- Suppose that γ , γ_s , and $\gamma_{s,g}$ are constants (not necessarily the same), and set $\mathbf{g} = \mathbf{0}$ (i.e., ignore gravity). Show that a spherical cap (subset of a sphere) is a valid solution of (6.19) and (6.21). What determines the radius and the contact angle θ_{cl} ?

6.2.3 • Fluids and Surface Tension

Surface tension plays an important role in the dynamics of liquids. In this case, curvature stresses are balanced by fluid normal stresses and are coupled through the fluid PDE's

boundary condition [8]. Furthermore, the movement of droplets on substrates means that the contact line Σ must also move, which leads to a serious theoretical issue [31, 52, 56, 74, 102]. The literature on moving contact lines is enormous (the few resources we cite are not exhaustive). The interested reader is referred to [105] for a more extensive literature review and a finite element method for simulating Stokes fluids with moving contact lines.

6.3 • Gradient Flows

Many geometric problems that exhibit fine differential geometric details, such as curvatures, cannot be solved directly because the underlying equations are nonlinear. Recall (6.19) and (6.21) and comments on their geometric nonlinearity. Thus, more tools are required to actually solve the resulting system of equations.

A powerful tool in that regard is the concept of a *gradient flow*. It is beyond the scope of this book to consider a full treatise on this topic; see [3, 13, 27, 28, 50, 106, 107] for further details. However, a very basic introduction follows.

6.3.1 • Shape Optimization

In general, gradient flows are used to solve optimization problems. In finite dimensional settings, they are known as gradient descent methods [75]. Many optimization problems fit the following framework. Let $\mathcal{J} : \mathcal{U} \rightarrow \mathbb{R}$ be an objective functional, where \mathcal{U} is an admissible set of optimization parameters. The goal is to search through \mathcal{U} (either by brute force or a more intelligent way) so as to find a particular optimization parameter u in \mathcal{U} that optimizes \mathcal{J} . Suppose we want to minimize \mathcal{J} . Then we want to find u^* in \mathcal{U} such that $\mathcal{J}(u^*) \leq \mathcal{J}(u)$ for all u in \mathcal{U} .

In the context of *shape optimization*, the problem takes the following form. Let $\mathcal{J} = \mathcal{J}(\Gamma)$ be a shape functional defined over a set of admissible shapes \mathcal{U} . Then the formal statement of the optimization problem is

$$\text{find } \Gamma^* \in \mathcal{U} \quad \text{such that } \mathcal{J}(\Gamma^*) = \min_{\Gamma \in \mathcal{U}} \mathcal{J}(\Gamma). \quad (6.22)$$

6.3.2 • Flow Velocity

We alluded to a way of solving this in section 1.2.3. Indeed, the shape perturbation $\delta \mathcal{J}(\Gamma; V)$ tells us how \mathcal{J} changes for any perturbation field V . Recall that V can be interpreted as a velocity field that deforms Γ . Hence, given an initial shape $\Gamma(0)$, we want to find a flow velocity V that induces a “time-dependent” shape $\Gamma(t)$ such that

$$\mathcal{J}(\Gamma(t_2)) < \mathcal{J}(\Gamma(t_1)) \quad \text{whenever } t_1 < t_2.$$

Recall section 5.2.2 about domain perturbations and perturbations of the identity. Note that the t variable is not really “time,” but a pseudotime to accompany the flow “velocity” V .

How can one find this V ? Let $L(\Gamma, \mathbb{R}^3)$ denote the set of vector-valued functions on \mathbb{R}^3 , i.e., V in $L(\Gamma, \mathbb{R}^3)$ means $V : \Gamma \rightarrow \mathbb{R}^3$. Now define $\mathbb{V}(\Gamma)$ to be a space of velocity fields, e.g.,

$$\mathbb{V}(\Gamma) = \left\{ V \in L(\Gamma, \mathbb{R}^3) : \int_{\Gamma} |V|^2 < \infty \right\}, \quad (6.23)$$

which is a Hilbert space also known as the space of $L^2(\Gamma)$ vector-valued functions [2, 63, 65, 100]. We measure the size of functions in $\mathbb{V}(\Gamma)$ by its norm:

$$\|\mathbf{V}\|_{\mathbb{V}(\Gamma)} \equiv \|\mathbf{V}\|_{L^2(\Gamma)} := \left(\int_{\Gamma} |\mathbf{V}|^2 \right)^{1/2}. \quad (6.24)$$

6.3.3 ■ Variational Form

Next, define a bilinear form $b : \mathbb{V}(\Gamma) \times \mathbb{V}(\Gamma) \rightarrow \mathbb{R}$:

$$b(\mathbf{V}, \mathbf{Y}) = \int_{\Gamma} \mathbf{V} \cdot \mathbf{Y}. \quad (6.25)$$

Note that b is related to the norm of the space $\mathbb{V}(\Gamma)$:

$$\|\mathbf{V}\|_{\mathbb{V}(\Gamma)} = \sqrt{b(\mathbf{V}, \mathbf{V})}.$$

With these preliminaries, we compute the flow velocity \mathbf{V} in the following way. At each “time” $t \geq 0$ of the flow, we find $\mathbf{V}(t)$ in $\mathbb{V}(\Gamma(t))$ that solves the following variational problem [13, 28, 37, 107]:

$$b(\mathbf{V}(t), \mathbf{Y}) = -\delta \mathcal{J}(\Gamma(t); \mathbf{Y}) \quad \text{for all } \mathbf{Y} \in \mathbb{V}(\Gamma(t)). \quad (6.26)$$

We call \mathbf{Y} a “test” perturbation. This problem is uniquely solvable, provided appropriate conditions are satisfied [37]; we do not go into details here.

The solution of (6.26) is precisely what we need because choosing $\mathbf{Y} = \mathbf{V}(t)$ yields

$$\delta \mathcal{J}(\Gamma(t); \mathbf{V}(t)) = -b(\mathbf{V}(t), \mathbf{V}(t)) = -\|\mathbf{V}(t)\|_{\mathbb{V}(\Gamma)}^2 = -\int_{\Gamma} |\mathbf{V}(t, \mathbf{x})|^2 dS(\mathbf{x}) < 0, \quad (6.27)$$

i.e., $\delta \mathcal{J}(\Gamma(t); \mathbf{V}(t))$ is negative, so \mathcal{J} is guaranteed to decrease if we instantaneously deform $\Gamma(t)$ using $\mathbf{V}(t)$. Hence, we call $\mathbf{V}(t)$ a *descent direction* for \mathcal{J} at time t . Assuming the optimization problem is well-posed [51, 54], letting $\Gamma(t)$ evolve according to $\mathbf{V}(t)$, and taking $t \rightarrow \infty$, $\Gamma(\infty)$ will be a local minimizer of \mathcal{J} .

6.3.4 ■ Choice of Space

The choice of $\mathbb{V}(\Gamma)$ and b affects the solution \mathbf{V} when solving (6.26). A different choice will yield a different velocity field \mathbf{V} . For instance, solving (6.26), with $\mathbb{V}(\Gamma)$ replaced by

$$\mathbb{V}(\Gamma) = \left\{ \mathbf{V} \in L(\Gamma, \mathbb{R}^3) : \int_{\Gamma} |\mathbf{V}|^2 + \int_{\Gamma} |\nabla_{\Gamma} \mathbf{V}|^2 < \infty \right\}, \quad (6.28)$$

and the bilinear form b replaced by

$$b(\mathbf{V}, \mathbf{Y}) = \int_{\Gamma} \mathbf{V} \cdot \mathbf{Y} + \int_{\Gamma} \nabla_{\Gamma} \mathbf{V} : \nabla_{\Gamma} \mathbf{Y}, \quad (6.29)$$

gives a “smoother” solution \mathbf{V} . So the choice of the space and bilinear form affects the descent direction. It can also affect the speed of convergence to the local minimizer. The “best” choice of \mathbb{V} and b depends on the nature of the particular optimization problem. See [13, 28, 51] for more details.

6.3.5 ■ Discretizing “Time”

As a practical matter, we cannot solve (6.26) for *every time* $t \geq 0$. Typically, one discretizes the gradient flow in time [13, 28]. First, define a sequence of time points $\{t_i\}_{i \geq 0}$, such that $t_i = i\Delta t$, where $\Delta t > 0$ is a fixed constant that represents the time step. Then the time-discrete gradient flow algorithm is as follows. Given t_i and $\Gamma(t_i)$, identify $\Gamma^i \equiv \Gamma(t_i)$ and $\mathbf{V}^i \equiv \mathbf{V}(t_i)$. Now find \mathbf{V}^i in $\mathbb{V}(\Gamma^i)$ such that

$$b(\mathbf{V}^i, \mathbf{Y}) = -\delta \mathcal{J}(\Gamma^i; \mathbf{Y}) \quad \text{for all } \mathbf{Y} \in \mathbb{V}(\Gamma^i). \quad (6.30)$$

Next, recalling (5.9), define a map

$$\mathbf{X}_{\Delta t} \circ \mathbf{X}^{-1}(\mathbf{x}) := \text{id}_{\Gamma^i}(\mathbf{x}) + \Delta t \mathbf{V}^i(\mathbf{x}) \quad \text{for all } \mathbf{x} \text{ in } \Gamma^i, \quad (6.31)$$

where \mathbf{X} is a parameterization of Γ^i . Then define the domain at the next time point $\Gamma^{i+1} = \mathbf{X}_{\Delta t} \circ \mathbf{X}^{-1}(\Gamma^i)$. Repeating this procedure gives a sequence of domains $\{\Gamma^i\}_{i \geq 0}$.

If the time step Δt is small enough, then the sequence $\{\mathcal{J}(\Gamma^i)\}_{i \geq 0}$ will decrease, i.e.,

$$\mathcal{J}(t_0) > \mathcal{J}(t_1) > \mathcal{J}(t_2) > \dots > \mathcal{J}(t_i) > \mathcal{J}(t_{i+1}) > \dots$$

Again, the choice of \mathbb{V} and b affects the speed of convergence of this minimizing sequence. See [13, 28, 48] for more details.

6.3.6 ■ Summary

This is not the end of the story for gradient flows. Some issues we have omitted are as follows.

- *Explicit, implicit, and semi-implicit time-stepping.* For instance, the procedure in section 6.3.5 uses explicit time-stepping. This is not always the best strategy [28].
- *Solving (6.30).* This is still an infinite dimensional problem. In practice, one must discretize space as well [3, 28, 48, 50] in order to turn (6.30) into a finite dimensional problem.
- *Well-posedness and stability.* Proving that the optimization problem makes sense in the first place can be a significant task [23, 47, 84, 104, 106]. Moreover, guaranteeing that the sequence of shapes generated by the method in section 6.3.5 actually yields a minimizing sequence is also not trivial.

The interested reader should look into the references mentioned here for more details on gradient flows. The following sections describe some applications of gradient flows to motivate their use.

6.4 ■ Mean Curvature Flow

A classic example of a flow driven purely by geometry is the mean curvature flow [32, 38, 40, 43, 67]. Given an initial regular, closed surface Γ_0 at time $t = 0$, we evolve $\Gamma = \Gamma(t)$ in the following way. Each point \mathbf{x} in Γ moves with velocity proportional to the summed curvature vector $\kappa \nu$. A goal in this field is to study the geometric properties of $\Gamma(t)$ for $t > 0$.

6.4.1 • Derivation

Mean curvature flow is simply a gradient flow applied to the perimeter functional

$$\mathcal{J}(\Gamma) = \int_{\Gamma} 1 dS(\mathbf{x}), \quad (6.32)$$

when Γ is a closed surface (see section 2.3.4). To see this, apply (6.26) to (6.32). For each $t \geq 0$, find $\mathbf{V}(t)$ such that

$$\int_{\Gamma(t)} \mathbf{V}(t) \cdot \mathbf{Y} = -\delta \mathcal{J}(\Gamma(t); \mathbf{Y}) = -\int_{\Gamma(t)} \kappa(t) \boldsymbol{\nu}(t) \cdot \mathbf{Y}$$

for all smooth perturbations \mathbf{Y} (for simplicity, we ignore the function space that \mathbf{Y} lives in). Note that the orientation of $\boldsymbol{\nu}$ is taken to point outside of the volume enclosed by Γ . Thus, if Γ is convex, then $\kappa \geq 0$.

Continuing, we get

$$\int_{\Gamma(t)} [\mathbf{V}(t) + \kappa(t) \boldsymbol{\nu}(t)] \cdot \mathbf{Y} = 0$$

for all smooth \mathbf{Y} . Again, a standard argument [37, 41, 58] shows

$$\mathbf{V}(t, \mathbf{x}) = -\kappa(t, \mathbf{x}) \boldsymbol{\nu}(t, \mathbf{x}) \quad \text{for all } \mathbf{x} \in \Gamma(t). \quad (6.33)$$

If Γ is convex, then \mathbf{V} points toward the interior of Γ (meaning that points on Γ move toward the interior). Thus, a convex surface *shrinks* under mean curvature flow.

If $\mathbf{X}(t)$ is a parameterization of $\Gamma(t)$, then $\partial_t \mathbf{X}(t) = \mathbf{V}(t, \mathbf{X}(t))$ and

$$\partial_t \mathbf{X} = -(\kappa \boldsymbol{\nu}) \circ \mathbf{X} = (\Delta_{\Gamma} \text{id}_{\Gamma}) \circ \mathbf{X}, \quad \mathbf{X}(0) = \mathbf{X}_0, \quad (6.34)$$

where we have used (4.29), and \mathbf{X}_0 is the initial parameterization (i.e., the initial surface Γ_0). Note that $\text{id}_{\Gamma} = \mathbf{X} \circ \mathbf{X}^{-1}$. Equation (6.34) is the parametric form of the *mean curvature flow equation*. Often, an abuse of notation is made, and we write (6.34) as

$$\partial_t \mathbf{X} = \Delta_{\Gamma} \mathbf{X}, \quad \mathbf{X}(0) = \mathbf{X}_0,$$

which is a geometric version of the heat equation [37, 112] applied to the surface parameterization \mathbf{X} .

Remark 26. *If the initial surface $\Gamma(0)$ is a sphere, then $\Gamma(t)$ is also a sphere for all $t \in [0, T_f]$, where $T_f > 0$ is some final time. At $t = T_f$, the surface collapses to a point, and the geometric flow is no longer defined. If $\Gamma(0)$ is merely convex, then the flow drives Γ toward a spherical shape while still (eventually) collapsing to a point. In \mathbb{R}^3 (or higher dimensions), a nonconvex shape can possibly lead to changes in topology of the surface Γ at later times; see [38] for further discussion of this. In \mathbb{R}^2 (i.e., Γ is a curve in the plane), no topological changes are possible, and mean curvature flow is referred to as “curve-shortening flow.”*

6.4.2 • Volume Constraint

Mean curvature flow is interesting, but for some applications it is not so convenient for the surface to collapse to a point. We can prevent this by imposing a volume constraint

on the gradient flow. This follows by applying (6.26) to (6.10). For each $t \geq 0$, find $\mathbf{V}(t)$ such that

$$\int_{\Gamma(t)} \mathbf{V}(t) \cdot \mathbf{Y} = -\delta \mathcal{L}(\Gamma(t), p; \mathbf{Y}) = - \int_{\Gamma(t)} \kappa(t) \nu(t) \cdot \mathbf{Y} + p \int_{\Gamma} \mathbf{Y} \cdot \nu \quad (6.35)$$

for all smooth perturbations \mathbf{Y} . Here, $p > 0$ is set to a constant that controls the volume enclosed by $\Gamma(t)$ as $t \rightarrow \infty$.

Proceeding as we did in the previous section, we get

$$\partial_t \mathbf{X} = -[(\kappa - p)\nu] \circ \mathbf{X}, \quad \mathbf{X}(0) = \mathbf{X}_0. \quad (6.36)$$

At equilibrium, $\partial_t \mathbf{X} = 0$, so then $\kappa = p$. Thus, if the initial shape is close to a sphere, then the flow determined by (6.36) converges to the solution of (6.11). Since the final shape is a sphere, $\kappa = 2/R$, where R is the radius of the sphere. Therefore, p fixes the size (and volume) of the sphere. A numerical example of time-discrete mean curvature flow is shown in Figure 6.5.

6.4.3 ■ Summary

A large amount of work has been done on mean curvature flow, from both theoretical and applied perspectives. Moreover, there is extensive literature on the numerical simulation of mean curvature flow; for instance, see [6, 15, 22, 28, 91]. We close this section with some exercises.

- How does mean curvature flow change if we replace \mathcal{J} with $\mathcal{J} = \int_{\Gamma} \gamma(\mathbf{x}) dS(\mathbf{x})$? How about for $\mathcal{J} = \int_{\Gamma} \gamma(\nu) dS(\mathbf{x})$, where ν is the unit normal vector of Γ ? (This last case leads to *anisotropic* mean curvature flow [7].)
- Derive a gradient flow to find the equilibrium shape of a droplet on a rigid substrate (refer to section 6.2.2).

6.5 ■ Image Segmentation

Image processing is any process that takes an image, such as a photograph, and outputs another image or a set of characteristics or properties related to the image. Often, the images are digital and 2-D; 3-D images are also possible (for instance, MRI data is a 3-D image). Many image-processing methods involve standard signal-processing techniques such as filtering. An advanced introduction to this diverse field can be found in [16]. The following sections explain how to *segment an image* using a gradient flow; the method we describe follows the one discussed in [17, 73]. Image segmentation is the process of isolating parts of an image for further inquiry or analysis; applications are in computer vision, where a common goal is to automate the processing of digital images, e.g., detecting intruders in surveillance video.

6.5.1 ■ Representing an Image

Let $\mathcal{D} \subset \mathbb{R}^2$ be the *image domain*. In practice, \mathcal{D} is a rectangle; see Figure 6.6. Next, let $I : \mathcal{D} \rightarrow [0, 1]$ be a function that represents the intensity of an image, i.e., $I(\mathbf{x})$ is a number where $I(\mathbf{x}) = 0$ means the image is black at \mathbf{x} and $I(\mathbf{x}) = 1$ means the image is white at \mathbf{x} . Intermediate values are different shades of gray. Usually, I is referred to as a *grayscale* image. Another type of image is a “black and white” image, where $I : \mathcal{D} \rightarrow \{0, 1\}$, i.e., the intensity values are either 0 or 1.

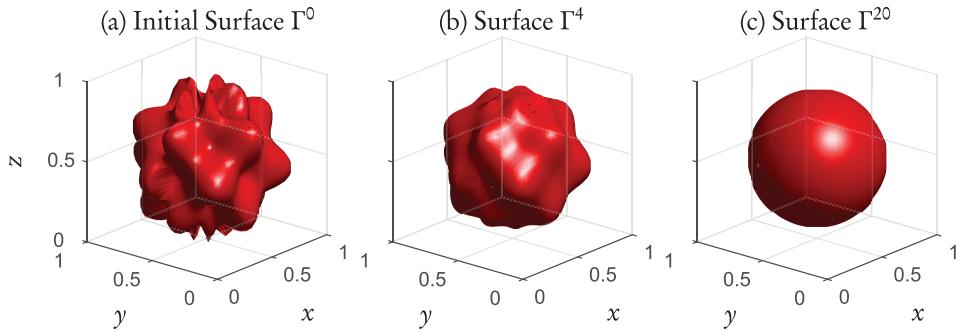


Figure 6.5. Time-discrete mean curvature flow with a volume constraint. (a) The initial shape is a “bumpy sphere.” (b) Applying four steps of a time-discrete gradient flow (see section 6.3.5) smoothes out the bumps (note that $\Delta t = 10^{-3}$ in nondimensional units). (c) At iteration 20, the surface is indistinguishable from a sphere.

6.5.2 • Modeling Features; Object Detection

Features (or objects) in an image are represented by variations in the image intensity I . For instance, a dark object on a light background is essentially represented by the dark points in the image. For such a simple image, identifying the location of the object is just a matter of thresholding the image intensity. In other words, we define $\mathcal{O} \subset \mathcal{D}$ to be the set of points that represent the object, i.e.,

$$\mathcal{O}(c) = \{\mathbf{x} \in \mathcal{D} : I(\mathbf{x}) \leq c\} \quad \text{where } c \in (0, 1). \quad (6.37)$$

The value of c is the *threshold value* we use to define the object. The problem of identifying objects or features in an image is known as *image segmentation*.

After some careful thought, it should be clear that the method in (6.37) is a very crude way to identify an object in an image. Unless the image is black and white (black object on a white background), different values of c will give different objects. The technique described in the following sections is a smarter way to segment images, although it is not completely foolproof.

6.5.3 • Fitting an Energy Functional

Let Ω_i , Ω_o be subsets of \mathcal{D} such that $\mathcal{D} = \overline{\Omega_i} \cup \overline{\Omega_o}$, $\Gamma = \overline{\Omega_i} \cap \overline{\Omega_o}$, $\partial\Omega_i = \Gamma$, and $\partial\Omega_o = \Gamma \cup \partial\mathcal{D}$ (see Figure 6.6). We use the sets Ω_i (Ω_o) to identify the inside (outside) of the object or feature of interest. Since Γ is a closed curve and \mathcal{D} is fixed, knowing Γ completely determines the partition $\{\Omega_i, \Omega_o, \Gamma\}$ of \mathcal{D} .

The Chan–Vese Functional

With these definitions in hand, we now define a shape functional that measures how well our choice of Γ “fits” the feature of interest (i.e., how well the image is segmented). For any smooth closed curve $\Gamma \subset \mathcal{D}$, let [17]

$$\mathcal{J}(\Gamma, c_i, c_o) = \mu \int_{\Gamma} 1 dS(\mathbf{x}) + \omega \int_{\Omega_i} 1 d\mathbf{x} + \lambda_i \int_{\Omega_i} |I(\mathbf{x}) - c_i|^2 d\mathbf{x} + \lambda_o \int_{\Omega_o} |I(\mathbf{x}) - c_o|^2 d\mathbf{x}, \quad (6.38)$$

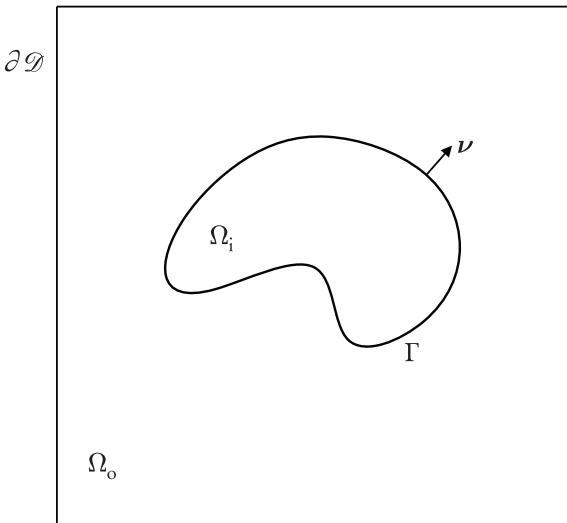


Figure 6.6. Domain of an image \mathcal{D} (with subdomains). An image function $I : \mathcal{D} \rightarrow [0, 1]$ represents the “intensity” of an image at each point in \mathcal{D} . In most applications, \mathcal{D} is a rectangle, as shown here. The image contains features represented by the subdomains Ω_i and Ω_o that are separated by the contour Γ . The unit normal vector ν of Γ is assumed to point into Ω_o .

where $\mu, \omega \geq 0$ and $\lambda_i, \lambda_o > 0$ are parameters that the “user” specifies (these will depend on the particular image I being segmented). For a grayscale image, $0 \leq c_i, c_o \leq 1$. Note that $\mathcal{J} \geq 0$.

To see how (6.38) measures the fit of the image feature, first let $\mu = \omega = 0$. Assume I is a black and white image, and set $c_i = 0$ (the object’s image intensity) and $c_o = 1$ (the intensity outside the object). If Γ corresponds to the boundary of the object, then Ω_i corresponds exactly to the object, and so $\mathcal{J}(\Gamma, c_i, c_o) = 0$. Thus, we have fitted the object perfectly. Therefore, we can view image segmentation as the process of finding a contour Γ and numbers $c_i, c_o \in [0, 1]$, such that $\mathcal{J}(\Gamma, c_i, c_o)$ is minimized.

Regularization

Of course, reality is not so simple. Usually the image is not black and white, and it may contain extraneous features such as multiple objects, large intensity variations, noise, etc. However, the domain integrals in (6.38) provide some “averaging” that helps us to ignore some of these extraneous effects, especially noise. Because our goal is to minimize \mathcal{J} , the first two terms in (6.38) can be viewed as penalty terms for the length of Γ and area of Ω_i .

Many optimization problems are *ill-posed*, i.e., their minimizers are hard to find or are not even well-defined. Image segmentation is one example. So penalty terms are often used to *regularize* the problem, i.e., to make the optimization problem smoother and the minimizers easier to find. Choosing $\mu, \omega > 0$ helps us to regularize the image segmentation problem.

6.5.4 • Active Contours

From section 6.3, we already know how to find a minimizer of (6.38). Applying the gradient flow methodology to (6.38) is sometimes called the *active contour algorithm*—active because the gradient flow induces a motion of the contour Γ .

Let us adopt the time-discrete gradient flow described in section 6.3.5. For the perturbation space, we use

$$\mathbb{V}(\Gamma) = \left\{ \mathbf{V} \in L(\Gamma, \mathbb{R}^2) : \int_{\Gamma} |\mathbf{V}|^2 + \int_{\Gamma} |\nabla_{\Gamma} \mathbf{V}|^2 < \infty \right\}, \quad (6.39)$$

with the bilinear form $b(\mathbf{V}, \mathbf{Y}) = \int_{\Gamma} \mathbf{V} \cdot \mathbf{Y} + \int_{\Gamma} \nabla_{\Gamma} \mathbf{V} : \nabla_{\Gamma} \mathbf{Y}$. Given a contour Γ^k , and holding c_i, c_o fixed, find \mathbf{V}^k in $\mathbb{V}(\Gamma^k)$ such that

$$b(\mathbf{V}^k, \mathbf{Y}) = -\delta_{\Gamma^k} \mathcal{J}(\Gamma^k, c_i, c_o; \mathbf{Y}) \quad \text{for all } \mathbf{Y} \in \mathbb{V}(\Gamma^k), \quad (6.40)$$

where

$$\begin{aligned} \delta_{\Gamma^k} \mathcal{J}(\Gamma^k, c_i, c_o; \mathbf{Y}) &= \mu \int_{\Gamma^k} \kappa^k \boldsymbol{\nu}^k \cdot \mathbf{Y} + \omega \int_{\Gamma^k} \mathbf{Y} \cdot \boldsymbol{\nu}^k \\ &\quad + \lambda_i \int_{\Gamma^k} |I - c_i|^2 \mathbf{Y} \cdot \boldsymbol{\nu}^k - \lambda_o \int_{\Gamma^k} |I - c_o|^2 \mathbf{Y} \cdot \boldsymbol{\nu}^k. \end{aligned} \quad (6.41)$$

Remark 27. The calculation of (6.41) requires some comments. First, note that I, c_i, c_o do not depend on shape.

Furthermore, the space \mathbb{V} is defined only on Γ . How do we perturb Ω_i, Ω_o ? Formally, we must extend \mathbf{V}^k from $\Gamma^k \subset \mathcal{D}$ to all of \mathcal{D} such that $\mathbf{V}^k = \mathbf{0}$ on $\partial \mathcal{D}$ (because the image domain is not allowed to change). There are many ways to extend a function. The details are not important because (6.41) depends only on Γ^k ; the domains Ω_i^k and Ω_o^k do not appear. In practice, we can simply track the movement of a polygonal curve that represents Γ^k [28]. In this case, an extension is not even needed; this is because Ω_i^k and Ω_o^k are essentially defined by Γ^k .

Lastly, we note the minus sign in front of λ_o ; this is due to the assumed orientation of $\boldsymbol{\nu}$ (see Figure 6.6).

After solving (6.40), we use (6.31) to obtain the next contour Γ^{k+1} . Given an initial guess for the contour Γ^0 , iterating this procedure gives a sequence of contours $\{\Gamma^k\}_{k \geq 0}$. If the parameters $\mu, \omega, \lambda_i, \lambda_o$ are chosen well, then $\Gamma^k \rightarrow \Gamma^\infty$, where Γ^∞ gives the desired segmentation of the image. In practice, some trial and error on the choice of parameters must be done to give reasonable results.

The choice of c_i, c_o can be done in different ways. If the user has prior knowledge of the features in the image, then c_i, c_o may be set constant to their “known” values. On the other hand, c_i and c_o can also be updated during the gradient flow. Holding the curve Γ fixed, $\mathcal{J}(\Gamma, c_i, c_o)$ is quadratic in c_i and c_o . Hence, we can compute the values of c_i, c_o that minimize $\mathcal{J}(\Gamma, c_i, c_o)$ by just setting the derivative to zero:

$$\begin{aligned} \partial_{c_i} \mathcal{J}(\Gamma, c_i, c_o) &= -2\lambda_i \int_{\Omega_i} (I - c_i) = 0, \quad \Rightarrow \quad c_i = \frac{1}{|\Omega_i|} \int_{\Omega_i} I, \\ \partial_{c_o} \mathcal{J}(\Gamma, c_i, c_o) &= 2\lambda_o \int_{\Omega_o} (I - c_o) = 0, \quad \Rightarrow \quad c_o = \frac{1}{|\Omega_o|} \int_{\Omega_o} I. \end{aligned} \quad (6.42)$$

In other words, just set c_i (c_o) to be the average image intensity in Ω_i (Ω_o). Adopting this method, an algorithm could be as follows. Starting with initial guesses Γ^0, c_i^0, c_o^0 , set $k = 0$, and work through the following procedure.



Figure 6.7. Image segmentation example (“ I ” in the text means this image). The contours correspond to different iterations of the optimization algorithm in section 6.5.4: Γ^0 (white), Γ^1 (cyan), Γ^3 (green), and Γ^{110} (red). The final contour (red) is a local minimizer of the Chan–Vese functional, with $\mu = 5 \cdot 10^{-4}$, $\omega = 0$, and $\lambda_i = \lambda_o = 1$.

1. Compute c_i^{k+1}, c_o^{k+1} using (6.42), with $\Omega_i (\Omega_o)$ replaced by $\Omega_i^k (\Omega_o^k)$.
2. Solve (6.40) for \mathbf{V}^k , i.e., find \mathbf{V}^k such that
$$b(\mathbf{V}^k, \mathbf{Y}) = -\delta_{\Gamma^k} \mathcal{J}(\Gamma^k, c_i^{k+1}, c_o^{k+1}; \mathbf{Y}) \text{ for all } \mathbf{Y} \in \mathbb{V}(\Gamma^k).$$
3. Update Γ^k to the next contour Γ^{k+1} using (6.31).
4. Replace k with $k + 1$ and return to step 1.

As an alternative, you can choose to update c_i and c_o every few iterations instead of every iteration.

The choice of \mathbb{V} and the bilinear form b can have a dramatic effect on the speed with which Γ^k approaches the minimizing shape [13, 28, 51]. It will also depend on the image I as well as the parameters. The interested reader can refer to [16] for an extensive introduction. Figure 6.7 shows an example of a gradient flow applied to (6.38). The initial, circular contour flows toward an “optimal” contour which segments the image into two regions: Ω_i = the child’s face, Ω_o = outside the child’s face.

6.5.5 • Summary

We have only given a brief introduction to image segmentation. For example, images are usually digital, meaning that I is really a piecewise constant function over a Cartesian

grid (i.e., an image is a collection of pixels). Furthermore, the gradient flow method can be reformulated as a PDE over \mathcal{D} using level set functions. In addition, the same methodology can be used for 3-D images, such as medical MRI data. More details can be found in [16, 17, 73]. We close this section with some exercises for the reader.

- What is the solution \mathbf{V}^k in (6.40) actually equal to? Follow the arguments in section 6.4.1 to derive an equation analogous to (6.33).
- How does the choice of $\mathbb{V}(\Gamma^k)$ or b affect \mathbf{V}^k ? For instance, what if we used $b(\mathbf{V}, \mathbf{Y}) = \int_{\Gamma} \mathbf{V} \cdot \mathbf{Y}$?

6.6 • Conclusion

Applications of shape differentiation are extremely diverse. Some examples are drag minimization (see section 1.2.3), optimal design of structures (bridges), acoustic optimization, topology optimization, etc., with connections to many other areas such as solid mechanics (deformable shells), fluid dynamics (surface tension), biomembranes (see Chapter 7), and computer graphics (surface restoration/smoothing). A full exploration of these areas is beyond the scope of this book; we encourage the reader to peruse the references for pointers to some of these topics.

Many of these applications involve PDEs. In keeping with the level of this book, we have chosen not to delve too deeply into PDE theory. In particular, we have not discussed *weak formulations*. This is necessary in order to better understand how to actually solve equations such as (6.26) and (6.30). Indeed, a spatial discretization of the problem is required for the practical implementation of gradient flows.

Moreover, we have not described how to discretize these problems. This involves replacing an idealized surface by a simpler approximation, such as a triangulated surface. In addition, the space \mathbb{V} must be replaced by a finite dimensional approximation. For example, a finite element method may be used to compute gradient flows. There are also other alternatives to discretizing the surface, such as the level set method [77].

Chapter 7

Willmore Flow

In this chapter, we consider a particular shape functional that represents the total bending energy of a surface. Warning to the reader: the presentation here is a bit more advanced than in previous chapters.

7.1 • The Energy

Let Γ be a regular surface with an atlas of charts, etc., and consider the *bending energy functional* (with variable coefficient $\beta : \Gamma \rightarrow \mathbb{R}$):

$$J_b(\Gamma) = \frac{1}{2} \int_{\Gamma} \beta \kappa^2. \quad (7.1)$$

This functional is also known as the *Willmore functional* when $\beta \equiv 1$ [110].

In this chapter, we derive the Euler–Lagrange equation (i.e., the first order optimality condition) for this functional with respect to perturbations of Γ . This can be considered as a geometrically nonlinear plate or shell model. It is also a model of biological membranes (biomembranes) and lipid vesicles [9, 10, 29, 30, 33, 35, 55, 89, 90, 113, 114]. We conclude with an application of the gradient flow methodology (see section 6.3) for computing local minimizers of (7.1). Note that a gradient flow applied to $J_b(\Gamma)$ is known as *Willmore flow*.

7.1.1 • Identities for Perturbing the Willmore Functional

We need some basic identities to facilitate the perturbation analysis.

Lemma 7.1 (material derivative of summed curvature squared).

$$\begin{aligned} \frac{1}{2} \dot{\kappa}^2 &= \kappa \cdot \dot{\kappa} = -2\kappa^2 \{ (\nabla_{\Gamma} \text{id}_{\Gamma} \circ \mathbf{X}) : (\nabla_{\Gamma} \mathbf{V} \circ \mathbf{X}) \} \\ &+ \frac{\kappa}{|\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}|} \cdot \left\{ \partial_{s_1} \left(\frac{(\partial_{s_1} \tilde{\mathbf{V}} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \tilde{\mathbf{V}})}{|\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}|} \right) \times \partial_{s_2} \mathbf{X} \right. \\ &\quad \left. + \partial_{s_1} \mathbf{X} \times \partial_{s_2} \left(\frac{(\partial_{s_1} \tilde{\mathbf{V}} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \tilde{\mathbf{V}})}{|\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}|} \right) \right\} \\ &+ \kappa \cdot \frac{(\partial_{s_1} \nu \times \partial_{s_2} \tilde{\mathbf{V}}) + (\partial_{s_1} \tilde{\mathbf{V}} \times \partial_{s_2} \nu)}{|\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}|} \end{aligned} \quad (7.2)$$

for all (s_1, s_2) in any reference domain U_i , where $\tilde{\mathbf{V}} = \mathbf{V} \circ \mathbf{X}$.

Proof. From (5.33), we have

$$\begin{aligned} \frac{1}{2} \frac{d}{d\epsilon} (\kappa_\epsilon \cdot \kappa_\epsilon) \Big|_{\epsilon=0^+} &= \kappa \cdot \dot{\kappa} \\ &= -\{(\nabla_\Gamma \text{id}_\Gamma \circ \mathbf{X}) : (\nabla_\Gamma \mathbf{V} \circ \mathbf{X})\} \kappa^2 \\ &\quad + \left| \partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X} \right|^{-1} \underbrace{\left\{ (\partial_{s_1} \dot{\nu} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \dot{\nu}) \right\} \cdot \kappa}_{=:T_1} \\ &\quad + \left| \partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X} \right|^{-1} \left\{ (\partial_{s_1} \nu \times \partial_{s_2} \tilde{\mathbf{V}}) + (\partial_{s_1} \tilde{\mathbf{V}} \times \partial_{s_2} \nu) \right\} \cdot \kappa. \end{aligned} \quad (7.3)$$

Now focus on T_1 . Using (5.31), we get

$$\begin{aligned} T_1 &= (\partial_{s_1} \dot{\nu} \times \partial_{s_2} \mathbf{X}) \cdot \kappa + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \dot{\nu}) \cdot \kappa \\ &= -(\partial_{s_1} (W \nu) \times \partial_{s_2} \mathbf{X}) \cdot \kappa + \left(\partial_{s_1} \left\{ \frac{[\partial_{s_1} \tilde{\mathbf{V}} \times \partial_{s_2} \mathbf{X} + \partial_{s_1} \mathbf{X} \times \partial_{s_2} \tilde{\mathbf{V}}]}{\left| \partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X} \right|} \right\} \times \partial_{s_2} \mathbf{X} \right) \cdot \kappa \\ &\quad - (\partial_{s_1} \mathbf{X} \times \partial_{s_2} (W \nu)) \cdot \kappa + \left(\partial_{s_1} \mathbf{X} \times \partial_{s_2} \left\{ \frac{[\partial_{s_1} \tilde{\mathbf{V}} \times \partial_{s_2} \mathbf{X} + \partial_{s_1} \mathbf{X} \times \partial_{s_2} \tilde{\mathbf{V}}]}{\left| \partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X} \right|} \right\} \right) \cdot \kappa, \end{aligned}$$

where $W = (\nabla_\Gamma \text{id}_\Gamma \circ \mathbf{X}) : (\nabla_\Gamma \mathbf{V} \circ \mathbf{X})$. Next, manipulate the terms with W :

$$\begin{aligned} &- (\partial_{s_1} (W \nu) \times \partial_{s_2} \mathbf{X}) \cdot \kappa - (\partial_{s_1} \mathbf{X} \times \partial_{s_2} (W \nu)) \cdot \kappa \\ &= -(\partial_{s_1} W) \kappa (\nu \times \partial_{s_2} \mathbf{X}) \cdot \nu - W (\partial_{s_1} \nu \times \partial_{s_2} \mathbf{X}) \cdot \kappa \\ &\quad - (\partial_{s_2} W) \kappa (\partial_{s_1} \mathbf{X} \times \nu) \cdot \nu - W (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \nu) \cdot \kappa \\ &= -W \left| \partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X} \right| \left[\frac{(\partial_{s_1} \nu \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \nu)}{\left| \partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X} \right|} \right] \cdot \kappa \\ &= -\left| \partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X} \right| \{(\nabla_\Gamma \text{id}_\Gamma \circ \mathbf{X}) : (\nabla_\Gamma \mathbf{V} \circ \mathbf{X})\} \kappa^2, \end{aligned}$$

where we used Proposition 12 from Chapter 4. Combining this with T_1 and (7.3) yields (7.2). \square

7.2 ■ Perturbation Analysis

We compute the shape perturbation of the Willmore (bending energy) functional.

7.2.1 ■ Preliminaries

We shall use some basic results from Appendix A.

Proposition 28 (normal vector identities).

$$(\partial_{s_i} \nu \times \partial_{s_j} \mathbf{X}) \cdot \partial_{s_k} \mathbf{X} = 0 \quad (7.4)$$

for $i, j = 1, 2$, such that $i \neq j$, and $k = 1, 2$.

Proof. If $j = k$, then

$$(\partial_{s_i} \boldsymbol{\nu} \times \partial_{s_j} \mathbf{X}) \cdot \partial_{s_k} \mathbf{X} = (\partial_{s_i} \boldsymbol{\nu}) \cdot (\partial_{s_j} \mathbf{X} \times \partial_{s_k} \mathbf{X}) = 0.$$

Otherwise,

$$(\partial_{s_i} \boldsymbol{\nu} \times \partial_{s_j} \mathbf{X}) \cdot \partial_{s_k} \mathbf{X} = \pm (\partial_{s_i} \boldsymbol{\nu}) \cdot \boldsymbol{\nu} \sqrt{\det(g)} = 0,$$

because $2(\partial_{s_i} \boldsymbol{\nu}) \cdot \boldsymbol{\nu} = \partial_{s_i} (\boldsymbol{\nu} \cdot \boldsymbol{\nu}) = \partial_{s_i} (1) = 0$. \square

Lemma 7.2 (pointwise identity).

$$\begin{aligned} \boldsymbol{\nu} \cdot \{ & (\partial_{s_1} \boldsymbol{\nu} \times \partial_{s_2} \tilde{\mathbf{V}}) + (\partial_{s_1} \tilde{\mathbf{V}} \times \partial_{s_2} \boldsymbol{\nu}) \} \\ &= \{\kappa(\nabla_\Gamma \cdot \mathbf{V}) \circ \mathbf{X} - [(\nabla_\Gamma \boldsymbol{\nu}) : (\nabla_\Gamma \mathbf{V})] \circ \mathbf{X}\} \Big| \partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X} \end{aligned} \quad (7.5)$$

for all (s_1, s_2) in any reference domain U_i , where $\tilde{\mathbf{V}} = \mathbf{V} \circ \mathbf{X}$.

Proof. Start with part of the right-hand-side and note that $\kappa = \nabla_\Gamma \cdot \boldsymbol{\nu}$:

$$\begin{aligned} R_1 &:= [(\nabla_\Gamma \cdot \boldsymbol{\nu}) \circ \mathbf{X}] [(\nabla_\Gamma \cdot \mathbf{V}) \circ \mathbf{X}] - [(\nabla_\Gamma \boldsymbol{\nu}) : (\nabla_\Gamma \mathbf{V})] \circ \mathbf{X} \\ &= \left(\sum_{i,j=1}^2 g^{ij} \partial_{s_i} \boldsymbol{\nu} \cdot \partial_{s_j} \mathbf{X} \right) \left(\sum_{p,q=1}^2 g^{pq} \partial_{s_p} \tilde{\mathbf{V}} \cdot \partial_{s_q} \mathbf{X} \right) - \left(\sum_{i,j=1}^2 g^{ij} \partial_{s_i} \boldsymbol{\nu} \cdot \partial_{s_j} \tilde{\mathbf{V}} \right), \end{aligned} \quad (7.6)$$

where we have used (4.20) and (4.39). Since $(\nabla_\Gamma \text{id}_\Gamma \circ \mathbf{X}) \partial_{s_i} \boldsymbol{\nu} = \partial_{s_i} \boldsymbol{\nu}$, we make this replacement in the last term of (7.6) using (4.15), i.e.,

$$\partial_{s_i} \boldsymbol{\nu} = \sum_{p,q=1}^2 g^{pq} [\partial_{s_p} \mathbf{X} \otimes \partial_{s_q} \mathbf{X}] \partial_{s_i} \boldsymbol{\nu} = \sum_{p,q=1}^2 g^{pq} (\partial_{s_i} \boldsymbol{\nu} \cdot \partial_{s_q} \mathbf{X}) \partial_{s_p} \mathbf{X} = \sum_{p,q=1}^2 g^{pq} (-h_{iq}) \partial_{s_p} \mathbf{X},$$

because $h_{ij} = -\partial_{s_i} \boldsymbol{\nu} \cdot \partial_{s_j} \mathbf{X}$ (see (3.36)). For convenience, let $r_{ij} = \partial_{s_i} \tilde{\mathbf{V}} \cdot \partial_{s_j} \mathbf{X}$. Then (7.6) becomes

$$\begin{aligned} R_1 &= - \left(\sum_{i,j=1}^2 \sum_{p,q=1}^2 g^{ij} g^{pq} h_{ij} r_{pq} \right) + \left(\sum_{i,j=1}^2 \sum_{p,q=1}^2 g^{ij} g^{pq} h_{iq} r_{jp} \right) \\ (\text{use symmetry of } g^{pq}) &= \sum_{i,j=1}^2 \sum_{p,q=1}^2 g^{ij} g^{pq} (h_{ip} r_{jq} - h_{ij} r_{pq}). \end{aligned}$$

Next, expand the sum, cancel terms, and get

$$\begin{aligned} R_1 &= [g^{11} g^{21} (h_{12} r_{11} - h_{11} r_{21}) + g^{12} g^{11} (h_{11} r_{21} - h_{12} r_{11}) \\ &\quad + g^{11} g^{22} (h_{12} r_{12} - h_{11} r_{22}) + g^{12} g^{12} (h_{11} r_{22} - h_{12} r_{12}) \\ &\quad + g^{21} g^{21} (h_{22} r_{11} - h_{21} r_{21}) + g^{22} g^{11} (h_{21} r_{21} - h_{22} r_{11}) \\ &\quad + g^{21} g^{22} (h_{22} r_{12} - h_{21} r_{22}) + g^{22} g^{12} (h_{21} r_{22} - h_{22} r_{12})]. \end{aligned}$$

Next, use (3.13), the symmetry of g , and the Kronecker delta property (3.14) to obtain

$$\begin{aligned}
R_1 &= [(g^{11}g^{21}-g^{12}g^{11})(b_{12}r_{11}-b_{11}r_{21}) \\
&\quad +(g^{11}g^{22}-g^{12}g^{12})(b_{12}r_{12}-b_{11}r_{22}) \\
&\quad +(g^{21}g^{21}-g^{22}g^{11})(b_{22}r_{11}-b_{21}r_{21}) \\
&\quad +(g^{21}g^{22}-g^{22}g^{12})(b_{22}r_{12}-b_{21}r_{22})] \\
&= \frac{1}{\det(g)} [(g_{22}g^{21}+g_{21}g^{11})(b_{12}r_{11}-b_{11}r_{21}) \\
&\quad +(g_{22}g^{22}+g_{21}g^{12})(b_{12}r_{12}-b_{11}r_{22}) \\
&\quad -(g_{12}g^{21}+g_{11}g^{11})(b_{22}r_{11}-b_{21}r_{21}) \\
&\quad -(g_{12}g^{22}+g_{11}g^{12})(b_{22}r_{12}-b_{21}r_{22})] \\
&= \frac{1}{\det(g)} [\delta_2^1(b_{12}r_{11}-b_{11}r_{21})+\delta_2^2(b_{12}r_{12}-b_{11}r_{22}) \\
&\quad -\delta_1^1(b_{22}r_{11}-b_{21}r_{21})-\delta_1^2(b_{22}r_{12}-b_{21}r_{22})] \\
&= \frac{1}{\det(g)} [b_{12}r_{12}-b_{11}r_{22}-b_{22}r_{11}+b_{21}r_{21}].
\end{aligned}$$

Now expand the left-hand side of (7.5) using (A.3):

$$\begin{aligned}
L_1 &:= \nu \cdot [(\partial_{s_1} \nu \times \partial_{s_2} \tilde{\mathbf{V}}) + (\partial_{s_1} \tilde{\mathbf{V}} \times \partial_{s_2} \nu)] \\
&= \frac{1}{\sqrt{\det(g)}} (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}) \cdot [(\partial_{s_1} \nu \times \partial_{s_2} \tilde{\mathbf{V}}) + (\partial_{s_1} \tilde{\mathbf{V}} \times \partial_{s_2} \nu)] \\
&= \frac{1}{\sqrt{\det(g)}} [(\partial_{s_1} \mathbf{X} \cdot \partial_{s_1} \nu)(\partial_{s_2} \mathbf{X} \cdot \partial_{s_2} \tilde{\mathbf{V}}) - (\partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \tilde{\mathbf{V}})(\partial_{s_2} \mathbf{X} \cdot \partial_{s_1} \nu) \\
&\quad + (\partial_{s_1} \mathbf{X} \cdot \partial_{s_1} \tilde{\mathbf{V}})(\partial_{s_2} \mathbf{X} \cdot \partial_{s_2} \nu) - (\partial_{s_1} \mathbf{X} \cdot \partial_{s_2} \nu)(\partial_{s_2} \mathbf{X} \cdot \partial_{s_1} \tilde{\mathbf{V}})].
\end{aligned}$$

Then, using the definitions of b_{ij} , r_{ij} and the symmetry of b_{ij} (recall (3.36)), we get

$$L_1 = \frac{1}{\sqrt{\det(g)}} [-b_{11}r_{22} + b_{21}r_{21} - b_{22}r_{11} + b_{12}r_{12}]. \quad (7.7)$$

Clearly, $L_1 = R_1 \sqrt{\det(g)}$. Since $|\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}| = \sqrt{\det(g)}$, this proves (7.5). \square

Lemma 7.3 (pointwise identity). *Suppose that Γ is a smooth surface with smooth boundary $\partial\Gamma$, and assume \mathbf{X} is given by (5.9) and \mathbf{V} is a smooth perturbation. Then,*

$$\kappa \cdot \dot{\kappa} = -\kappa \cdot \Delta_\Gamma \mathbf{V} - 2\kappa(\nabla_\Gamma \nu) : (\nabla_\Gamma \mathbf{V}) \quad (7.8)$$

is true pointwise on Γ .

Proof. As usual, assume $\tilde{\mathbf{V}} = \mathbf{V} \circ \mathbf{X}$ has support in a reference domain U_i :

$$\begin{aligned}\kappa \cdot \dot{\kappa} &= -2\kappa^2 \{(\nabla_\Gamma \text{id}_\Gamma \circ \mathbf{X}) : (\nabla_\Gamma \mathbf{V} \circ \mathbf{X})\} \\ &\quad + \kappa \cdot \frac{\{\partial_{s_1} \mathbf{W} \times \partial_{s_2} \mathbf{X} + \partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{W}\}}{|\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}|} \\ &\quad + \kappa \cdot \frac{\{(\partial_{s_1} \boldsymbol{\nu} \times \partial_{s_2} \tilde{\mathbf{V}}) + (\partial_{s_1} \tilde{\mathbf{V}} \times \partial_{s_2} \boldsymbol{\nu})\}}{|\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}|},\end{aligned}\tag{7.9}$$

where we have used (7.2) and

$$\mathbf{W} := \frac{(\partial_{s_1} \tilde{\mathbf{V}} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \tilde{\mathbf{V}})}{|\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}|}.$$

Next, apply (4.27) to \mathbf{V} :

$$\begin{aligned}(\Delta_\Gamma \mathbf{V}) \circ \mathbf{X} &= \frac{1}{|\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}|} \left\{ -\partial_{s_1} \mathbf{W} \times \partial_{s_2} \mathbf{X} - \partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{W} \right. \\ &\quad + \partial_{s_1} \left[\frac{(\partial_{s_1} \tilde{\mathbf{V}} \cdot \partial_{s_2} \mathbf{X}) - (\partial_{s_2} \tilde{\mathbf{V}} \cdot \partial_{s_1} \mathbf{X})}{|\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}|} \right] \partial_{s_2} \mathbf{X} \\ &\quad - \partial_{s_2} \left[\frac{(\partial_{s_1} \tilde{\mathbf{V}} \cdot \partial_{s_2} \mathbf{X}) - (\partial_{s_2} \tilde{\mathbf{V}} \cdot \partial_{s_1} \mathbf{X})}{|\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}|} \right] \partial_{s_1} \mathbf{X} \\ &\quad \left. + (\partial_{s_1} \boldsymbol{\nu} \times \partial_{s_2} \tilde{\mathbf{V}}) + (\partial_{s_1} \tilde{\mathbf{V}} \times \partial_{s_2} \boldsymbol{\nu}) \right\},\end{aligned}$$

which implies that

$$\begin{aligned}\kappa \cdot \{(\partial_{s_1} \mathbf{W} \times \partial_{s_2} \mathbf{X}) + (\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{W})\} &= -\kappa \cdot [(\Delta_\Gamma \mathbf{V}) \circ \mathbf{X}] |\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}| \\ &\quad + \kappa \cdot \{(\partial_{s_1} \boldsymbol{\nu} \times \partial_{s_2} \tilde{\mathbf{V}}) + (\partial_{s_1} \tilde{\mathbf{V}} \times \partial_{s_2} \boldsymbol{\nu})\},\end{aligned}\tag{7.10}$$

because $\kappa = \kappa \boldsymbol{\nu}$ and $\partial_{s_i} \mathbf{X} \cdot \boldsymbol{\nu} = 0$. Thus, combining

$$(\nabla_\Gamma \text{id}_\Gamma \circ \mathbf{X}) : (\nabla_\Gamma \mathbf{V} \circ \mathbf{X}) = (\nabla_\Gamma \cdot \mathbf{V}) \circ \mathbf{X}$$

and (7.10) with (7.9) gives

$$\kappa \cdot \dot{\kappa} = -2\kappa^2 (\nabla_\Gamma \cdot \mathbf{V}) \circ \mathbf{X} - \kappa \cdot [(\Delta_\Gamma \mathbf{V}) \circ \mathbf{X}] + 2\kappa \cdot \frac{\{(\partial_{s_1} \boldsymbol{\nu} \times \partial_{s_2} \tilde{\mathbf{V}}) + (\partial_{s_1} \tilde{\mathbf{V}} \times \partial_{s_2} \boldsymbol{\nu})\}}{|\partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X}|}.$$

Now use (7.5) (and map back to Γ) to get the assertion (7.8). \square

Lemma 7.4 (integral identity). Suppose that Γ is a regular surface with smooth boundary $\partial\Gamma$. Assume $\beta : \Gamma \rightarrow \mathbb{R}$ is a smooth function. Then,

$$\begin{aligned} \int_{\Gamma} \beta \kappa \cdot \dot{\kappa} &= \int_{\Gamma} \kappa (\nabla_{\Gamma} \beta \cdot \nabla_{\Gamma} \phi) + \int_{\Gamma} \beta \nabla_{\Gamma} \kappa \cdot \nabla_{\Gamma} \phi \\ &\quad - \int_{\Gamma} \beta \kappa \phi |\nabla_{\Gamma} \nu|^2 - \int_{\partial\Gamma} \beta \kappa [(\tau \times \nu) \cdot \nabla_{\Gamma} \phi], \end{aligned} \quad (7.11)$$

where Γ is parameterized by \mathbf{X} (see (5.9)) and is perturbed by $\mathbf{V} = \phi \nu$, where $\phi : \Gamma \rightarrow \mathbb{R}$ is a smooth function (perturbation).

Proof. By (7.8), we have

$$\int_{\Gamma} \beta \kappa \cdot \dot{\kappa} = - \int_{\Gamma} \beta \kappa \cdot \Delta_{\Gamma} \mathbf{V} - 2 \int_{\Gamma} \beta \kappa (\nabla_{\Gamma} \nu) : (\nabla_{\Gamma} \mathbf{V}),$$

and then we integrate by parts to get (see (4.37))

$$\int_{\Gamma} \beta \kappa \cdot \dot{\kappa} = \underbrace{\int_{\Gamma} (\nabla_{\Gamma}(\beta \kappa)) : (\nabla_{\Gamma} \mathbf{V})}_{=I_1} - 2 \underbrace{\int_{\Gamma} \beta \kappa (\nabla_{\Gamma} \nu) : (\nabla_{\Gamma} \mathbf{V})}_{=I_2} - \underbrace{\int_{\partial\Gamma} \beta \kappa \cdot \{[(\tau \times \nu) \cdot \nabla_{\Gamma}] \mathbf{V}\}}_{=I_3}.$$

Next, we continue to manipulate the individual terms:

$$\begin{aligned} I_1 &= \int_{\Gamma} (\kappa \otimes \nabla_{\Gamma} \beta + \beta \nabla_{\Gamma}(\kappa \nu)) : (\nabla_{\Gamma}(\phi \nu)) \\ &= \int_{\Gamma} (\kappa \nu \otimes \nabla_{\Gamma} \beta + \beta \nu \otimes \nabla_{\Gamma} \kappa + \beta \kappa \nabla_{\Gamma} \nu) : (\nu \otimes \nabla_{\Gamma} \phi + \phi \nabla_{\Gamma} \nu) \\ &= \int_{\Gamma} \kappa (\nu \otimes \nabla_{\Gamma} \beta) : (\nu \otimes \nabla_{\Gamma} \phi) + \beta (\nu \otimes \nabla_{\Gamma} \kappa) : (\nu \otimes \nabla_{\Gamma} \phi) + \beta \kappa (\nabla_{\Gamma} \nu) : (\nu \otimes \nabla_{\Gamma} \phi) \\ &\quad + \int_{\Gamma} \kappa \phi (\nu \otimes \nabla_{\Gamma} \beta) : (\nabla_{\Gamma} \nu) + \beta \phi (\nu \otimes \nabla_{\Gamma} \kappa) : (\nabla_{\Gamma} \nu) + \beta \kappa \phi (\nabla_{\Gamma} \nu) : (\nabla_{\Gamma} \nu), \\ &= \int_{\Gamma} \kappa \nabla_{\Gamma} \beta \cdot \nabla_{\Gamma} \phi + \beta \nabla_{\Gamma} \kappa \cdot \nabla_{\Gamma} \phi + \beta \kappa \phi (\nabla_{\Gamma} \nu) : (\nabla_{\Gamma} \nu), \end{aligned}$$

where we have used (A.23), $\nabla_{\Gamma} \nu = (\nabla_{\Gamma} \nu)^T$, and the fact that $\nu \cdot \nabla_{\Gamma}(\cdot) = 0$. For I_2 , we have

$$I_2 = \int_{\Gamma} \beta \kappa (\nabla_{\Gamma} \nu) : (\nu \otimes \nabla_{\Gamma} \phi + \phi \nabla_{\Gamma} \nu) = \int_{\Gamma} \beta \kappa \phi (\nabla_{\Gamma} \nu) : (\nabla_{\Gamma} \nu),$$

and I_3 gives

$$\begin{aligned} I_3 &= \int_{\partial\Gamma} \beta \kappa \cdot \{[(\tau \times \nu) \cdot \nabla_{\Gamma}] (\phi \nu)\} \\ &= \int_{\partial\Gamma} \beta \kappa \cdot \{\nu [(\tau \times \nu) \cdot \nabla_{\Gamma} \phi] + \phi [(\tau \times \nu) \cdot \nabla_{\Gamma}] \nu\} \\ &= \int_{\partial\Gamma} \beta \kappa [(\tau \times \nu) \cdot \nabla_{\Gamma} \phi] + \beta \kappa \phi \underbrace{[(\tau \times \nu) \cdot (\nabla_{\Gamma} \nu)] \nu}_{=0}. \end{aligned}$$

Combining everything, we get (7.11). \square

7.2.2 ■ Normal Perturbations

Theorem 7.5 (normal perturbations of the bending energy). Suppose that Γ is a regular surface with smooth boundary $\partial\Gamma$. Assume $\beta : \Gamma \rightarrow \mathbb{R}_+$ is a strictly positive bending coefficient and define $\beta_\epsilon : \Gamma \rightarrow \mathbb{R}_+$ (on the perturbed surface Γ_ϵ) by $\beta_\epsilon = \beta(\mathbf{X} \circ \mathbf{X}_\epsilon^{-1})$. Then,

$$\begin{aligned}\delta J_b(\Gamma; \phi) &= \int_{\Gamma} \kappa (\nabla_{\Gamma} \beta \cdot \nabla_{\Gamma} \phi) + \int_{\Gamma} \beta \nabla_{\Gamma} \kappa \cdot \nabla_{\Gamma} \phi - \int_{\Gamma} \beta \kappa \phi |\nabla_{\Gamma} \nu|^2 + \frac{1}{2} \int_{\Gamma} \beta \kappa^3 \phi \\ &\quad - \int_{\partial\Gamma} \beta \kappa [(\boldsymbol{\tau} \times \boldsymbol{\nu}) \cdot \nabla_{\Gamma} \phi],\end{aligned}\tag{7.12}$$

where Γ is parameterized by \mathbf{X} (see (5.9)) and is perturbed by $\mathbf{V} = \phi \boldsymbol{\nu}$, where $\phi : \Gamma \rightarrow \mathbb{R}$ is a smooth function.

Proof. First note that, for all \mathbf{x} in Γ ,

$$\dot{\beta}(\mathbf{x}) = \frac{d}{d\epsilon} \beta_\epsilon(\mathbf{X}_\epsilon \circ \mathbf{X}^{-1}(\mathbf{x})) \Big|_{\epsilon=0^+} = \frac{d}{d\epsilon} \beta(\mathbf{x}) \Big|_{\epsilon=0^+} = 0.$$

Hence, starting from (7.1) and applying (5.49), we have

$$\begin{aligned}\delta J_b(\Gamma; \mathbf{V}) &= \frac{1}{2} \left(\int_{\Gamma} \dot{\beta} \kappa^2 + \int_{\Gamma} \beta \dot{\kappa}^2 + \int_{\Gamma} \beta \kappa^2 (\nabla_{\Gamma} \cdot \mathbf{V}) \right) \\ &= \int_{\Gamma} \beta \kappa \cdot \dot{\kappa} + \frac{1}{2} \int_{\Gamma} \beta \kappa^2 \nabla_{\Gamma} \cdot (\phi \boldsymbol{\nu}) \\ &= \int_{\Gamma} \kappa (\nabla_{\Gamma} \beta \cdot \nabla_{\Gamma} \phi) + \int_{\Gamma} \beta \nabla_{\Gamma} \kappa \cdot \nabla_{\Gamma} \phi - \int_{\Gamma} \beta \kappa \phi |\nabla_{\Gamma} \nu|^2 + \frac{1}{2} \int_{\Gamma} \beta \kappa^2 \phi \nabla_{\Gamma} \cdot \boldsymbol{\nu} \\ &\quad - \int_{\partial\Gamma} \beta \kappa [(\boldsymbol{\tau} \times \boldsymbol{\nu}) \cdot \nabla_{\Gamma} \phi],\end{aligned}\tag{7.13}$$

where we have used (7.11) and $\boldsymbol{\nu} \cdot \nabla_{\Gamma} \phi = 0$. We then obtain (7.12) by using (4.48): $\nabla_{\Gamma} \cdot \boldsymbol{\nu} = \kappa$. \square

Corollary 2. Assume the hypothesis of Theorem 7.5. Moreover, assume β is constant and Γ is a closed surface (see section 2.3.4). If there exists Γ such that $\delta J_b(\Gamma; \phi) = 0$ for all smooth ϕ , then

$$-\Delta_{\Gamma} \kappa - \frac{1}{2} \kappa^3 + 2\kappa \kappa_G = 0 \quad \text{on } \Gamma.\tag{7.14}$$

Proof. By hypothesis, (7.12) becomes

$$\delta J_b(\Gamma; \phi) = \beta \left(\int_{\Gamma} \nabla_{\Gamma} \kappa \cdot \nabla_{\Gamma} \phi - \int_{\Gamma} \kappa \phi |\nabla_{\Gamma} \nu|^2 + \frac{1}{2} \int_{\Gamma} \kappa^3 \phi \right),$$

and applying integration by parts yields

$$\begin{aligned}\delta J_b(\Gamma; \phi) &= \beta \int_{\Gamma} \left(-\Delta_{\Gamma} \kappa - \kappa |\nabla_{\Gamma} \nu|^2 + \frac{1}{2} \kappa^3 \right) \phi \\ &= \beta \int_{\Gamma} \left(-\Delta_{\Gamma} \kappa - \frac{1}{2} \kappa^3 + 2\kappa \kappa_G \right) \phi,\end{aligned}$$

where we have used $|\nabla_{\Gamma} \nu|^2 = \kappa^2 - 2\kappa_G$ (see (4.45)). The assertion directly follows. \square

Remark 28. The equation in (7.14) is a partial differential equation (PDE) in the variable \mathbf{X} . So the solution is the surface Γ . It is also nonlinear because of the product of the curvatures, and because the PDE is defined on Γ which is a priori unknown.

7.2.3 • General Perturbations

We now consider general vector perturbations of the Willmore functional, which can be useful for developing numerical algorithms to approximate solutions of (7.14). First, we need some basic lemmas.

Lemma 7.6 (integral identity). Assume the hypothesis of Theorem 7.5, except that \mathbf{V} is an arbitrary vector perturbation. Then,

$$\begin{aligned} \int_{\Gamma} \beta \nabla_{\Gamma} \phi \cdot \nabla_{\Gamma} \kappa &= \int_{\Gamma} \beta \nabla_{\Gamma} \kappa : \nabla_{\Gamma} \mathbf{V} - \int_{\Gamma} \beta [(\nabla_{\Gamma} \text{id}_{\Gamma} + \nabla_{\Gamma} \text{id}_{\Gamma}^T) \nabla_{\Gamma} \mathbf{V}] : \nabla_{\Gamma} \kappa \\ &\quad - \int_{\Gamma} \beta \kappa \Delta_{\Gamma} \boldsymbol{\nu} \cdot \mathbf{V} - \int_{\Gamma} \kappa \nabla_{\Gamma} \beta \cdot (\nabla_{\Gamma} \boldsymbol{\nu}) \mathbf{V} + \int_{\partial\Gamma} \beta \kappa (\boldsymbol{\tau} \times \boldsymbol{\nu}) \cdot (\nabla_{\Gamma} \boldsymbol{\nu}) \mathbf{V}, \end{aligned} \quad (7.15)$$

where \mathbf{V} satisfies $\mathbf{V} \cdot \boldsymbol{\nu} = \phi$.

Proof. Note that $\mathbf{V} = \phi \boldsymbol{\nu} + (\nabla_{\Gamma} \text{id}_{\Gamma}) \mathbf{V}$ and

$$\begin{aligned} \nabla_{\Gamma} \kappa &= \nabla_{\Gamma} (\kappa \cdot \boldsymbol{\nu}) = \boldsymbol{\nu}^T (\nabla_{\Gamma} \kappa) + \kappa^T (\nabla_{\Gamma} \boldsymbol{\nu}) \\ &= \boldsymbol{\nu}^T (\nabla_{\Gamma} \kappa) + (\kappa \cdot \nabla_{\Gamma}) \boldsymbol{\nu} = \boldsymbol{\nu}^T (\nabla_{\Gamma} \kappa), \end{aligned} \quad (7.16)$$

where we have used the fact that $[\nabla_{\Gamma} \boldsymbol{\nu}] = [\nabla_{\Gamma} \boldsymbol{\nu}]^T$ and $\kappa = \kappa \boldsymbol{\nu}$. Next, we derive a simple identity:

$$\begin{aligned} \nabla_{\Gamma} \phi \cdot \nabla_{\Gamma} \kappa &= \nabla_{\Gamma} (\mathbf{V} \cdot \boldsymbol{\nu}) \cdot [\boldsymbol{\nu}^T (\nabla_{\Gamma} \kappa)] \\ &= [\boldsymbol{\nu}^T (\nabla_{\Gamma} \mathbf{V}) + \mathbf{V}^T (\nabla_{\Gamma} \boldsymbol{\nu})] \cdot [\boldsymbol{\nu}^T (\nabla_{\Gamma} \kappa)] \\ &= (\nabla_{\Gamma} \kappa) : [(\boldsymbol{\nu} \otimes \boldsymbol{\nu}) \nabla_{\Gamma} \mathbf{V}] + [\boldsymbol{\nu}^T (\nabla_{\Gamma} \kappa)] \cdot (\nabla_{\Gamma} \boldsymbol{\nu}) \mathbf{V} \\ &= [(\boldsymbol{\nu} \otimes \boldsymbol{\nu}) \nabla_{\Gamma} \mathbf{V}] : (\nabla_{\Gamma} \kappa) + \nabla_{\Gamma} \kappa \cdot (\nabla_{\Gamma} \boldsymbol{\nu}) \mathbf{V}, \end{aligned} \quad (7.17)$$

where we have used (A.23) and (7.16). Now, integration by parts (see (4.38)) gives

$$\begin{aligned} \int_{\Gamma} \beta \nabla_{\Gamma} \kappa \cdot (\nabla_{\Gamma} \boldsymbol{\nu}) \mathbf{V} &= \int_{\Gamma} \nabla_{\Gamma} \kappa \cdot [\beta (\nabla_{\Gamma} \boldsymbol{\nu}) \mathbf{V}] \\ &= - \int_{\Gamma} \kappa \nabla_{\Gamma} \cdot [\beta (\nabla_{\Gamma} \boldsymbol{\nu}) \mathbf{V}] + \int_{\partial\Gamma} \beta \kappa (\boldsymbol{\tau} \times \boldsymbol{\nu}) \cdot (\nabla_{\Gamma} \boldsymbol{\nu}) \mathbf{V} \\ &= - \int_{\Gamma} \kappa \nabla_{\Gamma} \beta \cdot (\nabla_{\Gamma} \boldsymbol{\nu}) \mathbf{V} - \int_{\Gamma} \beta \kappa (\nabla_{\Gamma} \boldsymbol{\nu}) : (\nabla_{\Gamma} \mathbf{V}) - \int_{\Gamma} \beta \kappa \Delta_{\Gamma} \boldsymbol{\nu} \cdot \mathbf{V} \\ &\quad + \int_{\partial\Gamma} \beta \kappa (\boldsymbol{\tau} \times \boldsymbol{\nu}) \cdot (\nabla_{\Gamma} \boldsymbol{\nu}) \mathbf{V}, \end{aligned}$$

and using $\kappa \nabla_\Gamma \nu = \nabla_\Gamma \kappa - \nu \otimes \nabla_\Gamma \kappa$ gives

$$\begin{aligned} \int_\Gamma \beta \nabla_\Gamma \kappa \cdot (\nabla_\Gamma \nu) \mathbf{V} &= - \int_\Gamma \kappa \nabla_\Gamma \beta \cdot (\nabla_\Gamma \nu) \mathbf{V} - \int_\Gamma \beta (\nabla_\Gamma \kappa) : (\nabla_\Gamma \mathbf{V}) \\ &\quad + \int_\Gamma \beta (\nu \otimes \nabla_\Gamma \kappa) : (\nabla_\Gamma \mathbf{V}) - \int_\Gamma \beta \kappa \Delta_\Gamma \nu \cdot \mathbf{V} + \int_{\partial\Gamma} \beta \kappa (\tau \times \nu) \cdot (\nabla_\Gamma \nu) \mathbf{V} \\ &= - \int_\Gamma \kappa \nabla_\Gamma \beta \cdot (\nabla_\Gamma \nu) \mathbf{V} - \int_\Gamma \beta \nabla_\Gamma \kappa : \nabla_\Gamma \mathbf{V} + \int_\Gamma \beta [(\nu \otimes \nu) \nabla_\Gamma \mathbf{V}] : \nabla_\Gamma \kappa \\ &\quad - \int_\Gamma \beta \kappa (\Delta_\Gamma \nu) \cdot \mathbf{V} + \int_{\partial\Gamma} \beta \kappa (\tau \times \nu) \cdot (\nabla_\Gamma \nu) \mathbf{V}, \end{aligned}$$

where we have used $\nabla_\Gamma \kappa = \nu^T (\nabla_\Gamma \kappa)$ and (A.23). Now multiply (7.17) by β , integrate, and combine with the previous result:

$$\begin{aligned} \int_\Gamma \beta \nabla_\Gamma \phi \cdot \nabla_\Gamma \kappa &= \int_\Gamma \beta [(\nu \otimes \nu) \nabla_\Gamma \mathbf{V}] : \nabla_\Gamma \kappa + \int_\Gamma \beta \nabla_\Gamma \kappa \cdot (\nabla_\Gamma \nu) \mathbf{V}, \\ &= 2 \int_\Gamma \beta [(\nu \otimes \nu) \nabla_\Gamma \mathbf{V}] : \nabla_\Gamma \kappa - \int_\Gamma \kappa \nabla_\Gamma \beta \cdot (\nabla_\Gamma \nu) \mathbf{V} - 2 \int_\Gamma \beta \nabla_\Gamma \mathbf{V} : \nabla_\Gamma \kappa \\ &\quad - \int_\Gamma \beta \kappa \Delta_\Gamma \nu \cdot \mathbf{V} + \int_\Gamma \beta \nabla_\Gamma \kappa : \nabla_\Gamma \mathbf{V} + \int_{\partial\Gamma} \beta \kappa (\tau \times \nu) \cdot (\nabla_\Gamma \nu) \mathbf{V} \\ &= - \int_\Gamma \beta [2(\mathbf{I} - \nu \otimes \nu) \nabla_\Gamma \mathbf{V}] : \nabla_\Gamma \kappa - \int_\Gamma \kappa \nabla_\Gamma \beta \cdot (\nabla_\Gamma \nu) \mathbf{V} \\ &\quad - \int_\Gamma \beta \kappa \Delta_\Gamma \nu \cdot \mathbf{V} + \int_\Gamma \beta \nabla_\Gamma \kappa : \nabla_\Gamma \mathbf{V} + \int_{\partial\Gamma} \beta \kappa (\tau \times \nu) \cdot (\nabla_\Gamma \nu) \mathbf{V}. \end{aligned}$$

Finally, noting that $\nabla_\Gamma \text{id}_\Gamma + \nabla_\Gamma \text{id}_\Gamma^T = 2(\mathbf{I} - \nu \otimes \nu)$, we obtain (7.15). \square

Remark 29. Note that $\nabla_\Gamma \text{id}_\Gamma + \nabla_\Gamma \text{id}_\Gamma^T = 2\nabla_\Gamma \text{id}_\Gamma$. Depending on the numerical approximation used, it may be that $(\nabla_\Gamma \text{id}_\Gamma)_h$ is not symmetric (here, h denotes a numerical discretization). Thus, the symmetrized form may be more useful for preserving that symmetry.

Lemma 7.7 (integral identity). Assume the hypothesis of Theorem 7.5, except that \mathbf{V} is an arbitrary vector perturbation. Then,

$$-\int_\Gamma \beta \kappa \Delta_\Gamma \nu \cdot \mathbf{V} = \int_\Gamma \beta \kappa |\nabla_\Gamma \nu|^2 \phi - \frac{1}{2} \int_\Gamma \beta \nabla_\Gamma (\kappa^2) \cdot \mathbf{V}, \quad (7.18)$$

where \mathbf{V} satisfies $\mathbf{V} \cdot \nu = \phi$.

Proof. From (4.49), we know that $\Delta_\Gamma \nu = -|\nabla_\Gamma \nu|^2 \nu + (\nabla_\Gamma \kappa)^T$, so then

$$\begin{aligned} \beta \kappa (\Delta_\Gamma \nu) \cdot \mathbf{V} &= -\beta \kappa |\nabla_\Gamma \nu|^2 \phi + \beta \kappa (\nabla_\Gamma \kappa) \cdot (\nabla_\Gamma \text{id}_\Gamma) \mathbf{V} \\ &= -\beta \kappa |\nabla_\Gamma \nu|^2 \phi + \frac{1}{2} \beta \nabla_\Gamma (\kappa^2) \cdot \mathbf{V}. \end{aligned}$$

Integrating this relation delivers the assertion. \square

Finally, we arrive at the main theorem.

Theorem 7.8 (general perturbations of the bending energy). *Assume the hypothesis of Theorem 7.5, except that \mathbf{V} is an arbitrary vector perturbation. Then,*

$$\begin{aligned}\delta J_b(\Gamma; \mathbf{V}) = & \int_{\Gamma} \kappa (\nabla_{\Gamma} \beta) \cdot (\nabla_{\Gamma} \mathbf{V})^T \boldsymbol{\nu} + \int_{\Gamma} \beta \nabla_{\Gamma} \kappa : \nabla_{\Gamma} \mathbf{V} \\ & - \int_{\Gamma} \beta [(\nabla_{\Gamma} \text{id}_{\Gamma} + \nabla_{\Gamma} \text{id}_{\Gamma}^T) \nabla_{\Gamma} \mathbf{V}] : \nabla_{\Gamma} \kappa \\ & + \frac{1}{2} \left(\int_{\Gamma} \beta (\nabla_{\Gamma} \cdot \kappa) (\nabla_{\Gamma} \cdot \mathbf{V}) + \int_{\Gamma} \kappa^2 (\nabla_{\Gamma} \beta) \cdot \mathbf{V} \right) \\ & - \int_{\partial\Gamma} \beta \kappa (\boldsymbol{\tau} \times \boldsymbol{\nu}) \cdot (\nabla_{\Gamma} \mathbf{V})^T \boldsymbol{\nu} - \frac{1}{2} \int_{\partial\Gamma} \beta \kappa^2 (\boldsymbol{\tau} \times \boldsymbol{\nu}) \cdot \mathbf{V}\end{aligned}\tag{7.19}$$

for all smooth $\mathbf{V}: \Gamma \rightarrow \mathbb{R}^3$.

Proof. Let us decompose \mathbf{V} as $\mathbf{V} = \phi \boldsymbol{\nu} + [\nabla_{\Gamma} \text{id}_{\Gamma}] \mathbf{V}$. Starting with Theorem 7.5,

$$\begin{aligned}\delta J_b(\Gamma; \phi) = & \int_{\Gamma} \kappa (\nabla_{\Gamma} \beta \cdot \nabla_{\Gamma} \phi) + \int_{\Gamma} \beta \nabla_{\Gamma} \kappa \cdot \nabla_{\Gamma} \phi - \int_{\Gamma} \beta \kappa \phi |\nabla_{\Gamma} \boldsymbol{\nu}|^2 + \frac{1}{2} \int_{\Gamma} \beta \kappa^3 \phi \\ & - \int_{\partial\Gamma} \beta \kappa [(\boldsymbol{\tau} \times \boldsymbol{\nu}) \cdot \nabla_{\Gamma} \phi] \\ = & \int_{\Gamma} \kappa (\nabla_{\Gamma} \beta \cdot \nabla_{\Gamma} \phi) - \int_{\Gamma} \beta \kappa \Delta_{\Gamma} \boldsymbol{\nu} \cdot \mathbf{V} - \int_{\Gamma} \beta \kappa \phi |\nabla_{\Gamma} \boldsymbol{\nu}|^2 + \frac{1}{2} \int_{\Gamma} \beta \kappa^3 \phi \\ & + \int_{\Gamma} \beta \nabla_{\Gamma} \kappa : \nabla_{\Gamma} \mathbf{V} - \int_{\Gamma} \beta [(\nabla_{\Gamma} \text{id}_{\Gamma} + \nabla_{\Gamma} \text{id}_{\Gamma}^T) \nabla_{\Gamma} \mathbf{V}] : \nabla_{\Gamma} \kappa \\ & - \int_{\Gamma} \kappa \nabla_{\Gamma} \beta \cdot (\nabla_{\Gamma} \boldsymbol{\nu}) \mathbf{V} + \int_{\partial\Gamma} \beta \kappa (\boldsymbol{\tau} \times \boldsymbol{\nu}) \cdot (\nabla_{\Gamma} \boldsymbol{\nu}) \mathbf{V} \\ & - \int_{\partial\Gamma} \beta \kappa [(\boldsymbol{\tau} \times \boldsymbol{\nu}) \cdot \nabla_{\Gamma} \phi],\end{aligned}$$

where we have used (7.15). Next, we use (7.18):

$$\begin{aligned}\delta J_b(\Gamma; \phi) = & \int_{\Gamma} \kappa (\nabla_{\Gamma} \beta \cdot \nabla_{\Gamma} \phi) - \frac{1}{2} \int_{\Gamma} \beta \nabla_{\Gamma} (\kappa^2) \cdot \mathbf{V} + \frac{1}{2} \int_{\Gamma} \beta \kappa^3 \phi \\ & + \int_{\Gamma} \beta \nabla_{\Gamma} \kappa : \nabla_{\Gamma} \mathbf{V} - \int_{\Gamma} \beta [(\nabla_{\Gamma} \text{id}_{\Gamma} + \nabla_{\Gamma} \text{id}_{\Gamma}^T) \nabla_{\Gamma} \mathbf{V}] : \nabla_{\Gamma} \kappa \\ & - \int_{\Gamma} \kappa \nabla_{\Gamma} \beta \cdot (\nabla_{\Gamma} \boldsymbol{\nu}) \mathbf{V} + \int_{\partial\Gamma} \beta \kappa (\boldsymbol{\tau} \times \boldsymbol{\nu}) \cdot (\nabla_{\Gamma} \boldsymbol{\nu}) \mathbf{V} \\ & - \int_{\partial\Gamma} \beta \kappa [(\boldsymbol{\tau} \times \boldsymbol{\nu}) \cdot \nabla_{\Gamma} \phi].\end{aligned}\tag{7.20}$$

With the identities

$$\begin{aligned}\nabla_{\Gamma} \cdot (\beta \kappa^2 \mathbf{V}) &= \beta \kappa^2 (\nabla_{\Gamma} \cdot \mathbf{V}) + \kappa^2 (\nabla_{\Gamma} \beta) \cdot \mathbf{V} + \beta \nabla_{\Gamma} (\kappa^2) \cdot \mathbf{V}, \\ \nabla_{\Gamma} \cdot \kappa &= \kappa^2,\end{aligned}$$

we get by integration by parts:

$$\begin{aligned}\int_{\Gamma} \beta \kappa^3 \phi &= \int_{\Gamma} \kappa \cdot (\beta \kappa^2 \mathbf{V}) = \int_{\Gamma} \beta (\nabla_{\Gamma} \cdot \kappa) (\nabla_{\Gamma} \cdot \mathbf{V}) + \int_{\Gamma} \kappa^2 (\nabla_{\Gamma} \beta) \cdot \mathbf{V} \\ &\quad + \int_{\Gamma} \beta \nabla_{\Gamma} (\kappa^2) \cdot \mathbf{V} - \int_{\partial\Gamma} (\boldsymbol{\tau} \times \boldsymbol{\nu}) \cdot (\beta \kappa^2 \mathbf{V}).\end{aligned}$$

Hence, (7.20) becomes

$$\begin{aligned}\delta J_b(\Gamma; \mathbf{V}) &= \int_{\Gamma} \kappa \nabla_{\Gamma} \beta \cdot \nabla_{\Gamma} \phi + \frac{1}{2} \int_{\Gamma} \beta (\nabla_{\Gamma} \cdot \kappa) (\nabla_{\Gamma} \cdot \mathbf{V}) + \frac{1}{2} \int_{\Gamma} \kappa^2 (\nabla_{\Gamma} \beta) \cdot \mathbf{V} \\ &\quad + \int_{\Gamma} \beta \nabla_{\Gamma} \kappa : \nabla_{\Gamma} \mathbf{V} - \int_{\Gamma} \beta [(\nabla_{\Gamma} \text{id}_{\Gamma} + \nabla_{\Gamma} \text{id}_{\Gamma}^T) \nabla_{\Gamma} \mathbf{V}] : \nabla_{\Gamma} \kappa \\ &\quad - \int_{\Gamma} \kappa \nabla_{\Gamma} \beta \cdot (\nabla_{\Gamma} \boldsymbol{\nu}) \mathbf{V} + \int_{\partial\Gamma} \beta \kappa (\boldsymbol{\tau} \times \boldsymbol{\nu}) \cdot (\nabla_{\Gamma} \boldsymbol{\nu}) \mathbf{V} \\ &\quad - \int_{\partial\Gamma} \beta \kappa [(\boldsymbol{\tau} \times \boldsymbol{\nu}) \cdot \nabla_{\Gamma} \phi] - \frac{1}{2} \int_{\partial\Gamma} (\boldsymbol{\tau} \times \boldsymbol{\nu}) \cdot (\beta \kappa^2 \mathbf{V}).\end{aligned}\tag{7.21}$$

Next, plug $\nabla_{\Gamma} \phi = \nabla_{\Gamma} (\mathbf{V} \cdot \boldsymbol{\nu}) = \boldsymbol{\nu}^T (\nabla_{\Gamma} \mathbf{V}) + \mathbf{V}^T (\nabla_{\Gamma} \boldsymbol{\nu})$ into (7.21) and use the symmetry of $\nabla_{\Gamma} \boldsymbol{\nu}$ to obtain the assertion. \square

7.3 ■ Gradient Flow

We apply the gradient flow framework (section 6.3) to derive a method for computing solutions of (7.14). The formula (7.19) allows for arbitrary vector-valued perturbations, which are useful for developing finite element discretizations of the problem.

7.3.1 ■ Constraints

For simplicity, let Γ be a surface without boundary that encloses a volume Ω , i.e., $\Gamma = \partial\Omega$ and $\partial\Gamma = \emptyset$. In addition, take $\beta(\mathbf{x}) = 1$, and define a Lagrangian shape functional

$$\mathcal{L}(\Gamma, p, \lambda) = J_b(\Gamma) - p \left(\int_{\Omega} 1 d\mathbf{x} - C_V \right) - \lambda \left(\int_{\Gamma} 1 dS(\mathbf{x}) - C_A \right),\tag{7.22}$$

where we have constrained the volume $|\Omega| = C_V$ and the surface area $|\Gamma| = C_A$. The volume constraint is necessary because we assume that the biomembrane encloses an incompressible fluid; the surface area constraint is due to the material properties of the biomembrane which, in applications, is a lipid bilayer [89, 90, 113, 114].

Just as in (6.11), a constrained minimizer of (7.22) must satisfy the following first order optimality conditions:

$$\begin{aligned}\delta_{\Gamma} \mathcal{L}(\Gamma, p, \lambda; \mathbf{V}) &= \delta J_b(\Gamma; \mathbf{V}) - p \int_{\Gamma} \mathbf{V} \cdot \boldsymbol{\nu} - \lambda \int_{\Gamma} \kappa \boldsymbol{\nu} \cdot \mathbf{V} = 0, \\ \partial_p \mathcal{L}(\Gamma, p, \lambda) &= \int_{\Omega} 1 d\mathbf{x} - C_V = 0, \\ \partial_{\lambda} \mathcal{L}(\Gamma, p, \lambda) &= \int_{\Gamma} 1 dS(\mathbf{x}) - C_A = 0,\end{aligned}\tag{7.23}$$

where \mathbf{V} is an arbitrary smooth perturbation of Γ .

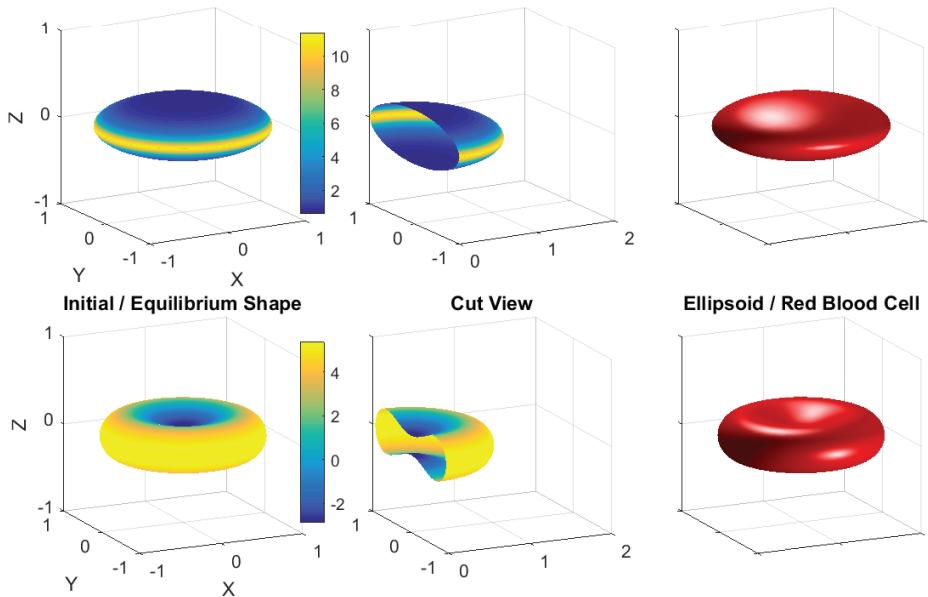


Figure 7.1. An equilibrium shape of the Willmore functional (7.1). Starting from a flattened ellipsoid (top row), the gradient flow (7.24) deforms the shape toward a disk with a thickened outer ring and inner “dimples” with prescribed volume and surface area (bottom row). The summed curvature κ is plotted in color (first two columns). Note that κ is negative in the dimples because the surface is concave there.

7.3.2 ■ Flow Toward a Minimizer

A gradient flow for the bending energy with constraints can be derived by the approach used in section 6.4.2 (see (6.35)). Hence, for each $t \geq 0$, find $\mathbf{V}(t)$ in $\mathbb{V}(\Gamma(t))$ and $p(t)$, $\lambda(t)$ in \mathbb{R} , such that

$$\begin{aligned} \int_{\Gamma(t)} \mathbf{V}(t) \cdot \mathbf{Y} &= -\delta \mathcal{L}(\Gamma(t), p, \lambda; \mathbf{Y}) = -\delta J_b(\Gamma(t); \mathbf{Y}) + p(t) \int_{\Gamma(t)} \mathbf{Y} \cdot \boldsymbol{\nu} + \lambda(t) \int_{\Gamma(t)} \kappa \boldsymbol{\nu} \cdot \mathbf{Y}, \\ \int_{\Omega} 1 d\mathbf{x} &= C_V, \quad \int_{\Gamma} 1 dS(\mathbf{x}) = C_A, \end{aligned} \tag{7.24}$$

for all \mathbf{Y} in $\mathbb{V}(\Gamma(t))$. The space $\mathbb{V}(\Gamma)$ is given in (6.23). The multipliers $p(t)$ and $\lambda(t)$ control the volume enclosed by $\Gamma(t)$ and surface area of $\Gamma(t)$ at each time t , respectively.

Figure 7.1 shows the results of a numerical simulation using the method in [10]. The initial guess for Γ in the gradient flow scheme is an ellipsoid with aspect ratio $1 : 1 : 0.3125$; the summed curvature κ of the ellipsoid falls in the range $[0.6262, 11.3846]$. The final (equilibrium) shape is a biconcave disk with aspect ratio $0.93141 : 0.93141 : 0.25505$; the summed curvature of the biconcave disk falls in the range $[-2.8754, 5.3607]$. One can see that κ is generally smaller for the biconcave disk. The volume and surface area are constrained to be $C_V = 1.1$, $C_A = 7.1$. The equilibrium surface is reminiscent of the shapes of red blood cells [55].

7.3.3 ■ Summary

The purpose of this chapter is to give an introduction to Willmore flow and show an application of it. The interested reader is directed toward the previously cited references. For more information on the numerical simulation of these flows, see for instance, [6, 10, 30, 33, 34]. We conclude this chapter with some exercises.

- Prove Lemma 7.3 using a simpler approach. Hint: use Lemmas 5.5 and 5.6.
- Show that $\mathbf{V}(t)$ in (7.24) satisfies

$$\mathbf{V} \cdot \boldsymbol{\nu} = \Delta_{\Gamma} \kappa + \frac{1}{2} \kappa^3 - 2\kappa \kappa_G + p + \lambda \kappa.$$

Hint: choose $\mathbf{Y} = \phi \boldsymbol{\nu}$ and use Theorem 7.5 and Corollary 2.

- If Γ has a nonempty boundary $\partial\Gamma$, then the shape perturbation of J_b has boundary terms $\int_{\partial\Gamma} \dots$; see (7.19). Let \mathbf{X} be a parameterization of Γ . What boundary conditions must \mathbf{V} (and/or \mathbf{X}) satisfy in order for the boundary terms in (7.19) to vanish? For example, is setting $\mathbf{V} = \mathbf{0}$ on $\partial\Gamma$ enough?

Appendix A

Vectors and Matrices

This appendix gives some basic facts for manipulating vectors and matrices.

A.1 ▪ Vector Operations

A.1.1 ▪ Notation

Let $\mathbf{a} \cdot \mathbf{b}$ be the dot product of the vectors $\mathbf{a} = (a_1, \dots, a_n)^T$ and $\mathbf{b} = (b_1, \dots, b_n)^T$ in \mathbb{R}^n , i.e.,

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^n a_i b_i.$$

The Euclidean norm of a vector \mathbf{a} in \mathbb{R}^n is defined by $|\mathbf{a}| = \sqrt{\mathbf{a} \cdot \mathbf{a}}$ for any $n \geq 1$. The zero vector is denoted by $\mathbf{0}$.

A.1.2 ▪ Basic Identities

Let $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ be vectors in \mathbb{R}^3 . Denote by \times the cross product operation; see section A.2.2. Then the following basic vector identities hold:

$$\mathbf{a} \times \mathbf{a} = \mathbf{0}, \tag{A.1}$$

$$\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}, \tag{A.2}$$

$$(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) = (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}) - (\mathbf{a} \cdot \mathbf{d})(\mathbf{b} \cdot \mathbf{c}), \tag{A.3}$$

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}(\mathbf{a} \cdot \mathbf{c}) - \mathbf{c}(\mathbf{a} \cdot \mathbf{b}), \tag{A.4}$$

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}. \tag{A.5}$$

A.2 ▪ Matrix Operations

A.2.1 ▪ Notation and Definitions

We denote a $k \times n$ matrix as a rectangular array of numbers \mathbf{M} in $\mathbb{R}^{k \times n}$, i.e.,

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{k1} & m_{k2} & \cdots & m_{kn} \end{bmatrix} \Leftrightarrow \mathbf{M} = [m_{ij}], \quad 1 \leq i \leq k, 1 \leq j \leq n. \tag{A.6}$$

We also use the notation $(\mathbf{M})_{ij} \equiv m_{ij}$ to denote the (i, j) element of the matrix \mathbf{M} ; similarly, for a vector \mathbf{a} , we write $(\mathbf{a})_i \equiv a_i$ to denote the i th element of \mathbf{a} .

For any column vector \mathbf{x} in \mathbb{R}^n , define the product \mathbf{Mx} by

$$\mathbf{y} = \mathbf{Mx}, \quad \Leftrightarrow \quad y_i = \sum_{j=1}^n m_{ij} x_j, \quad 1 \leq i \leq k, \quad (\text{A.7})$$

where \mathbf{y} is in \mathbb{R}^k . The matrix transpose \mathbf{M}^T in $\mathbb{R}^{n \times k}$ is given by

$$\mathbf{M}^T = [m_{ij}^T] := [m_{ji}], \quad \text{where } 1 \leq i \leq n, 1 \leq j \leq k. \quad (\text{A.8})$$

Thus, if \mathbf{a} and \mathbf{b} are column vectors in \mathbb{R}^n , then we can view \mathbf{a}^T as a $1 \times n$ matrix, and $\mathbf{a} \cdot \mathbf{b} = \mathbf{a}^T \mathbf{b}$. If \mathbf{A} is in $\mathbb{R}^{k \times n}$ and \mathbf{B} is in $\mathbb{R}^{n \times r}$, then the product \mathbf{AB} is defined by

$$\mathbf{C} = \mathbf{AB}, \quad \Leftrightarrow \quad c_{ij} = \sum_{q=1}^n a_{iq} b_{qj}, \quad 1 \leq i \leq k, 1 \leq j \leq r, \quad (\text{A.9})$$

where \mathbf{C} is in $\mathbb{R}^{k \times r}$.

If \mathbf{M} is in $\mathbb{R}^{n \times n}$, then we call \mathbf{M} a square matrix. The identity matrix \mathbf{I} in $\mathbb{R}^{n \times n}$ satisfies $(\mathbf{I})_{ii} = 1$ for $1 \leq i \leq n$ and $(\mathbf{I})_{ij} = 0$ for $i \neq j$; it also satisfies $\mathbf{Ix} = \mathbf{x}$ for any $n \times 1$ vector \mathbf{x} and $\mathbf{AI} = \mathbf{A}$, $\mathbf{IB} = \mathbf{B}$, where \mathbf{A} is in $\mathbb{R}^{k \times n}$ and \mathbf{B} is in $\mathbb{R}^{n \times r}$.

The trace of \mathbf{M} is defined by

$$\text{trace}(\mathbf{M}) = \sum_{i=1}^n m_{ii}. \quad (\text{A.10})$$

There is an analogous “dot product” for matrices, i.e., if \mathbf{A} and \mathbf{B} are $k \times n$ matrices, then

$$\mathbf{A} : \mathbf{B} = \sum_{i=1}^k \sum_{j=1}^n a_{ij} b_{ij} = \text{trace}(\mathbf{A}^T \mathbf{B}), \quad (\text{A.11})$$

which leads to the so-called Frobenius norm of a matrix:

$$|\mathbf{A}| = \sqrt{\mathbf{A} : \mathbf{A}}. \quad (\text{A.12})$$

For column vectors \mathbf{a} in \mathbb{R}^k and \mathbf{b} in \mathbb{R}^n , we have the tensor product of \mathbf{a} and \mathbf{b} defined by

$$\mathbf{C} := \mathbf{a} \otimes \mathbf{b} = \mathbf{ab}^T, \quad \Leftrightarrow \quad c_{ij} = a_i b_j, \quad 1 \leq i \leq k, 1 \leq j \leq n, \quad (\text{A.13})$$

where \mathbf{C} is a $k \times n$ matrix.

A.2.2 ■ Cross Product

The cross product \times operation (in \mathbb{R}^3) can be written as

$$[\mathbf{a}]_\times := \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix} \Rightarrow \mathbf{a} \times \mathbf{b} = [\mathbf{a}]_\times \mathbf{b} = [\mathbf{a}]_\times \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}. \quad (\text{A.14})$$

Note that any 3×3 skew symmetric matrix has the form $[\mathbf{d}]_\times$ for some vector \mathbf{d} in \mathbb{R}^3 . We also have these identities

$$\begin{aligned} [\mathbf{a}]_\times^T &= -[\mathbf{a}]_\times, \\ \mathbf{b}^T [\mathbf{a}]_\times &= -\mathbf{b}^T [\mathbf{a}]_\times^T = -([\mathbf{a}]_\times \mathbf{b})^T = -(\mathbf{a} \times \mathbf{b})^T = (\mathbf{b} \times \mathbf{a})^T, \\ \nabla(\mathbf{a} \times \mathbf{x}) &= [\mathbf{a}]_\times \nabla \mathbf{x} = [\mathbf{a}]_\times, \\ \mathbf{b} \cdot [\nabla(\mathbf{a} \times \mathbf{x})] &= [\mathbf{a}]_\times \mathbf{b} = \mathbf{a} \times \mathbf{b}. \end{aligned} \quad (\text{A.15})$$

A.2.3 ▪ Identities

Proposition 29 (properties of the cofactor matrix). Let \mathbf{M} be an $n \times n$ matrix (with entries m_{ij}) and let $\text{cof}\mathbf{M}$ be the matrix of cofactors of \mathbf{M} , whose (i, j) th entry is defined by

$$(\text{cof}\mathbf{M})_{ij} = (-1)^{i+j} d(\mathbf{M})_{ij}, \quad (\text{A.16})$$

where $d(\mathbf{M})_{ij}$ is the determinant of the $(n-1) \times (n-1)$ matrix obtained by deleting the i th row and j th column from \mathbf{M} . Then $\text{cof}\mathbf{M}$ satisfies the following properties:

1. $\text{cof}\mathbf{M} = \det(\mathbf{M})(\mathbf{M}^T)^{-1}$.
2. $(\text{cof}\mathbf{M})_{kj}$ is independent of $\{m_{k1}, m_{k2}, \dots, m_{kn}\}$ and $\{m_{1q}, m_{2q}, \dots, m_{nq}\}$.
3. $\partial \det(\mathbf{M}) / \partial m_{kj} = (\text{cof}\mathbf{M})_{kj}$.

Proof. Property 1 can be found in standard linear algebra texts. Property 2 follows from the definition of $d(\mathbf{M})_{ij}$. Property 3 follows by first rewriting property 1 as

$$\det(\mathbf{M})\mathbf{I} = \mathbf{M}^T \text{cof}\mathbf{M} \Leftrightarrow \det(\mathbf{M})\delta_{ij} = \sum_{l=1}^n m_{li}(\text{cof}\mathbf{M})_{lj}. \quad (\text{A.17})$$

Now differentiate both sides with respect to m_{kj} , where $1 \leq k, q \leq n$ (k, q fixed) and set i, j equal to q :

$$\begin{aligned} \frac{\partial \det(\mathbf{M})}{\partial m_{kj}} &= \sum_{l=1}^n \underbrace{\left(\frac{\partial}{\partial m_{kj}} m_{lj} \right)}_{=0 \text{ when } l \neq k} (\text{cof}\mathbf{M})_{lj} + m_{kj} \underbrace{\left(\frac{\partial}{\partial m_{kj}} (\text{cof}\mathbf{M})_{lj} \right)}_{=0 \text{ by (2)}} \\ &= (\text{cof}\mathbf{M})_{kj}. \quad \square \end{aligned}$$

Proposition 30 (matrix cross product identity). Let \mathbf{M} be a 3×3 matrix and let \mathbf{a}, \mathbf{b} be 3×1 vectors. Then we have

$$(\mathbf{M}\mathbf{a}) \times (\mathbf{M}\mathbf{b}) = \det(\mathbf{M})(\mathbf{M}^T)^{-1}(\mathbf{a} \times \mathbf{b}). \quad (\text{A.18})$$

Proof. Let m_{ij} be the elements of \mathbf{M} and recall Proposition 29:

$$\begin{aligned} \det(\mathbf{M})(\mathbf{M}^T)^{-1} &= \begin{bmatrix} (m_{33}m_{22} - m_{32}m_{23}) & -(m_{33}m_{21} - m_{31}m_{23}) & (m_{32}m_{21} - m_{31}m_{22}) \\ -(m_{33}m_{12} - m_{32}m_{13}) & (m_{33}m_{11} - m_{31}m_{13}) & -(m_{32}m_{11} - m_{31}m_{12}) \\ (m_{23}m_{12} - m_{22}m_{13}) & -(m_{23}m_{11} - m_{21}m_{13}) & (m_{22}m_{11} - m_{21}m_{12}) \end{bmatrix} \\ &= \text{cof}\mathbf{M} =: \mathbf{C} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix}. \end{aligned} \quad (\text{A.19})$$

Let \mathbf{e}_n be the n th unit basis vector in \mathbb{R}^3 for $1 \leq n \leq 3$. Using (A.5) and (A.15), we have

$$\mathbf{e}_n \cdot [(\mathbf{M}\mathbf{a}) \times (\mathbf{M}\mathbf{b})] = [\mathbf{e}_n \times (\mathbf{M}\mathbf{a})] \cdot (\mathbf{M}\mathbf{b}) = \mathbf{b}^T \underbrace{\mathbf{M}^T [\mathbf{e}_n]_{\times} \mathbf{M} \mathbf{a}}_{\text{skew symmetric}}. \quad (\text{A.20})$$

From (A.19), one can show that

$$(\mathbf{M}^T [\mathbf{e}_n]_{\times} \mathbf{M})_{ij} = \sum_{k=1}^3 \sum_{l=1}^3 m_{ki} ([\mathbf{e}_n]_{\times})_{kl} m_{lj} = ([\mathbf{C}^T \mathbf{e}_n]_{\times})_{ij}. \quad (\text{A.21})$$

Thus, we get

$$\begin{aligned} \mathbf{e}_n \cdot [(\mathbf{M}\mathbf{a}) \times (\mathbf{M}\mathbf{b})] &= \mathbf{b} \cdot [\mathbf{C}^T \mathbf{e}_n]_{\times} \mathbf{a} = ((\mathbf{C}^T \mathbf{e}_n) \times \mathbf{a}) \cdot \mathbf{b} \\ &= (\mathbf{C}^T \mathbf{e}_n) \cdot (\mathbf{a} \times \mathbf{b}) = \mathbf{e}_n \cdot \mathbf{C}(\mathbf{a} \times \mathbf{b}) \text{ for } 1 \leq n \leq 3, \end{aligned} \quad (\text{A.22})$$

which is the assertion. \square

A.3 • Vector and Matrix Identities

Proposition 31. Let $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ be column vectors in \mathbb{R}^n , and let $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ be $n \times n$ matrices. Then,

$$\begin{aligned} (\mathbf{a} \otimes \mathbf{b}) : (\mathbf{c} \otimes \mathbf{d}) &= (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}), \\ (\mathbf{AB}) : \mathbf{C} &= (\mathbf{A}^T \mathbf{C}) : \mathbf{B}, \\ (\mathbf{a}^T \mathbf{C}) \cdot (\mathbf{b}^T \mathbf{D}) &= \mathbf{D} : [(\mathbf{b} \otimes \mathbf{a}) \mathbf{C}]. \end{aligned} \quad (\text{A.23})$$

Proof. The first identity follows by (A.11). The second identity follows by

$$(\mathbf{AB}) : \mathbf{C} = \sum_i \sum_j \left(\sum_k a_{ik} b_{kj} \right) c_{ij} = \sum_k \sum_j \left(\sum_i a_{ik} c_{ij} \right) b_{kj} = (\mathbf{A}^T \mathbf{C}) : \mathbf{B}.$$

For the last identity, we have

$$\begin{aligned} (\mathbf{a}^T \mathbf{C}) \cdot (\mathbf{b}^T \mathbf{D}) &= \sum_j \left(\sum_k (a_k c_{kj}) \sum_i (b_i d_{ij}) \right) \\ &= \sum_i \sum_j d_{ij} \sum_k (b_i a_k c_{kj}) = \mathbf{D} : [(\mathbf{b} \otimes \mathbf{a}) \mathbf{C}]. \quad \square \end{aligned}$$

Appendix B

Derivatives and Integrals

This appendix gives some basic facts involving derivatives and integrals.

B.1 ▪ Differential Formulas

B.1.1 ▪ Basic Identities

Proposition 32 (derivative of an inverse matrix). Let $\mathbf{G}(t)$ be an invertible matrix for all t . Moreover, assume \mathbf{G} and \mathbf{G}^{-1} are continuously differentiable with respect to t . Then,

$$\frac{d}{dt}\mathbf{G}^{-1}(t) = -\mathbf{G}^{-1}(t)\left(\frac{d}{dt}\mathbf{G}(t)\right)\mathbf{G}^{-1}(t). \quad (\text{B.1})$$

Proof. It follows by

$$0 = \frac{d}{dt}(\mathbf{G}^{-1}(t)\mathbf{G}(t)) = \left(\frac{d}{dt}\mathbf{G}^{-1}(t)\right)\mathbf{G}(t) + \mathbf{G}^{-1}(t)\left(\frac{d}{dt}\mathbf{G}(t)\right). \quad \square$$

Lemma B.1 (derivative of a determinant). Let $\mathbf{G}(t)$ be a family of matrices that depend on the parameter t . Assume that $\det(\mathbf{G}(t)) \neq 0$ for all t . Then,

$$\frac{d}{dt}\det(\mathbf{G}(t)) = \det(\mathbf{G}(t)) \cdot \text{trace}\left[\left(\frac{d}{dt}\mathbf{G}(t)\right)\mathbf{G}^{-1}(t)\right]. \quad (\text{B.2})$$

See the standard references for a proof. Note that \mathbf{G}^{-1} is the matrix inverse.

Lemma B.2 (expansion of the determinant). Consider $\det \mathbf{G}(t)$, where $\mathbf{G}(t) = \mathbf{I} + t\mathbf{A} + \frac{1}{2}t^2\mathbf{B} + \mathbf{F}(t)$ and \mathbf{A}, \mathbf{B} are fixed matrices. Suppose $\|\mathbf{F}(t)\| = o(t^2)$. Then,

$$\det \mathbf{G}(t) = 1 + t \text{trace } \mathbf{A} + \frac{1}{2}t^2 [\text{trace } \mathbf{B} + (\text{trace } \mathbf{A})^2 - \text{trace}(\mathbf{A}^2)] + o(t^2). \quad (\text{B.3})$$

Proof. Let $g(t) = \det \mathbf{G}(t)$ and do a Taylor expansion. From (B.2), we get

$$\begin{aligned} g'(t) &= \det(\mathbf{G}(t)) \text{trace}\left[\left(\frac{d}{dt}\mathbf{G}(t)\right)\mathbf{G}^{-1}(t)\right], \\ &= g(t) \text{trace}\left[(\mathbf{A} + t\mathbf{B} + \mathbf{F}'(t))\mathbf{G}^{-1}(t)\right], \\ \Rightarrow g'(0) &= \text{trace } \mathbf{A} \end{aligned} \quad (\text{B.4})$$

and

$$\begin{aligned}
 g''(t) &= g'(t)\text{trace}\left[\left(\mathbf{A} + t\mathbf{B} + \mathbf{F}'(t)\right)\mathbf{G}^{-1}(t)\right] \\
 &\quad + g(t)\text{trace}\left[\left(\mathbf{B} + \mathbf{F}''(t)\right)\mathbf{G}^{-1}(t) + \left(\mathbf{A} + t\mathbf{B} + \mathbf{F}'(t)\right)(\mathbf{G}^{-1}(t))'\right] \\
 (\text{B.1}) \rightarrow &= g'(t)\text{trace}\left[\left(\mathbf{A} + t\mathbf{B} + \mathbf{F}'(t)\right)\mathbf{G}^{-1}(t)\right] + g(t)\text{trace}\left[\left(\mathbf{B} + \mathbf{F}''(t)\right)\mathbf{G}^{-1}(t)\right] \quad (\text{B.5}) \\
 &\quad - g(t)\text{trace}\left[\left(\mathbf{A} + t\mathbf{B} + \mathbf{F}'(t)\right)\mathbf{G}^{-1}(t)\mathbf{G}'(t)\mathbf{G}^{-1}(t)\right] \\
 \Rightarrow g''(0) &= (\text{trace } \mathbf{A})^2 + \text{trace } \mathbf{B} - \text{trace}(\mathbf{A}^2).
 \end{aligned}$$

So the second order Taylor expansion of $g(t)$ gives

$$\begin{aligned}
 g(t) &= g(0) + g'(0)t + \frac{1}{2}g''(0)t^2 + o(t^2) \quad (\text{B.6}) \\
 &= 1 + t \text{ trace } \mathbf{A} + \frac{1}{2}t^2 \left[\text{trace } \mathbf{B} + (\text{trace } \mathbf{A})^2 - \text{trace}(\mathbf{A}^2) \right] + o(t^2). \quad \square
 \end{aligned}$$

B.1.2 ■ Change of Variables for Spatial Gradients

Mapping a Domain

Let $\hat{\Omega}$ and Ω be open sets of points contained in \mathbb{R}^n ; we call them *domains*. Let $\Phi : \hat{\Omega} \rightarrow \Omega$ be a **homeomorphism** (see section 2.1.4) that maps points in the domain $\hat{\Omega}$ to the domain Ω . Thus, given \mathbf{a} in $\hat{\Omega}$, there is a distinct point \mathbf{y} in Ω such that $\mathbf{y} = \Phi(\mathbf{a})$ and $\Phi^{-1}(\mathbf{y}) = \mathbf{a}$. Therefore, $\Omega = \Phi(\hat{\Omega})$, i.e., Φ maps the entire domain $\hat{\Omega}$ onto Ω . Similarly, Φ^{-1} maps the entire domain Ω onto $\hat{\Omega}$. In addition, assume that Φ (and its inverse Φ^{-1}) are at least C^1 (differentiable).

The Formula

Let \hat{f} be a scalar function defined on $\hat{\Omega}$, i.e., $\hat{f} \equiv \hat{f}(\mathbf{a})$ for all \mathbf{a} in $\hat{\Omega}$. With this, define f on Ω by $f(\mathbf{y}) = \hat{f}(\Phi^{-1}(\mathbf{y})) = \hat{f}(\mathbf{a})$. Then, we have that gradients transform as

$$\begin{aligned}
 \partial_{y_j} f(\mathbf{y}) &= \partial_{y_j} \hat{f}(\Phi^{-1}(\mathbf{y})) = \sum_i \partial_{a_i} \hat{f}(\mathbf{a}) \Big|_{\mathbf{a}=\Phi^{-1}(\mathbf{y})} \frac{\partial(\Phi^{-1}(\mathbf{y}) \cdot \mathbf{e}_i)}{\partial_{y_j}} \\
 \Rightarrow (\nabla_y f)(\mathbf{y}) &= \underbrace{\nabla_a \hat{f}(\Phi^{-1}(\mathbf{y}))}_{\text{row vector}} \underbrace{(\nabla_y \Phi^{-1})(\mathbf{y})}_{\text{matrix}}
 \end{aligned} \quad (\text{B.7})$$

which implies that

$$\begin{aligned}
 (\nabla_y f) \circ \Phi(\mathbf{a}) &= \nabla_a \hat{f}(\mathbf{a}) (\nabla_y \Phi^{-1} \circ \Phi)(\mathbf{a}) \\
 \text{using (5.24)} \rightarrow &= \underbrace{\nabla_a \hat{f}(\mathbf{a})}_{\text{row vector}} \underbrace{[\nabla_a \Phi(\mathbf{a})]^{-1}}_{\text{matrix}}.
 \end{aligned} \quad (\text{B.8})$$

B.2 ■ Integral Formulas

B.2.1 ■ Change of Variables for Integrals over Bulk Domains

Refer to the notation in section B.1.2. We have the standard formula for applying a change of variables to a bulk integral (see any multivariable calculus book for this formula):

$$\int_{\Omega} f(\mathbf{y}) d\mathbf{y} = \int_{\hat{\Omega}} f(\Phi(\mathbf{a})) |\det(\nabla_a \Phi(\mathbf{a}))| d\mathbf{a} = \int_{\hat{\Omega}} \hat{f}(\mathbf{a}) |\det(\nabla_a \Phi(\mathbf{a}))| d\mathbf{a}. \quad (\text{B.9})$$

And when a gradient is present, we get

$$\int_{\Omega} (\nabla_y f)(y) dy = \int_{\hat{\Omega}} (\nabla_y f) \circ \Phi(a) |\det(\nabla_a \Phi(a))| da$$

(B.10)

using (B.8) $\rightarrow = \int_{\hat{\Omega}} \nabla_a \hat{f}(a) [\nabla_a \Phi(a)]^{-1} |\det(\nabla_a \Phi(a))| da.$

B.2.2 • Surface Integrals

Refer to Definition 1 in section 2.2.2. The surface integral of $g : \Gamma \rightarrow \mathbb{R}$ over the parameterized surface Γ is defined by

$$\int_{\Gamma} g \equiv \int_{\Gamma} g(x) dS(x) = \int_U g(\mathbf{X}(s_1, s_2)) \left| \partial_{s_1} \mathbf{X} \times \partial_{s_2} \mathbf{X} \right| ds_1 ds_2. \quad (B.11)$$

Bibliography

- [1] E. A. ABBOTT, *Flatland: A Romance of Many Dimensions*, Roberts Brothers, Boston, MA, 1885. (Cited on p. 28)
- [2] R. A. ADAMS AND J. J. F. FOURNIER, *Sobolev Spaces*, 2nd ed., Pure and Applied Mathematics Series 140, Elsevier, Oxford, UK, 2003. (Cited on p. 114)
- [3] G. ALLAIRE, F. JOUVE, AND A.-M. TOADER, *Structural optimization using sensitivity analysis and a level-set method*, Journal of Computational Physics, 194 (2004), pp. 363–393. (Cited on pp. 3, 88, 113, 115)
- [4] H. W. ALT, *Verzweigungspunkte von H -Flächen I*, Mathematische Zeitschrift, 127 (1972), pp. 333–362. (Cited on pp. 109, 111)
- [5] T. AUBIN, *Nonlinear Analysis on Manifolds. Monge–Ampère Equations*, Fundamental Principles of Mathematical Sciences 252, Springer Verlag, New York, NY, 1982. (Cited on p. 7)
- [6] J. W. BARRETT, H. GARCKE, AND R. NÜRNBERG, *Parametric approximation of Willmore flow and related geometric evolution equations*, SIAM Journal on Scientific Computing, 31 (2008), pp. 225–253. (Cited on pp. 117, 135)
- [7] J. W. BARRETT, H. GARCKE, AND R. NÜRNBERG, *A variational formulation of anisotropic geometric evolution equations in higher dimensions*, Numerische Mathematik, 109 (2008), pp. 1–44. (Cited on p. 117)
- [8] G. K. BATCHELOR, *An Introduction to Fluid Dynamics*, Cambridge University Press, Cambridge, UK, 1967. (Cited on pp. 3, 113)
- [9] T. BAUMGART, S. T. HESS, AND W. W. WEBB, *Imaging coexisting fluid domains in biomembrane models coupling curvature and line tension*, Nature, 425 (2003), pp. 821–824. (Cited on pp. 9, 123)
- [10] A. BONITO, R. H. NOCHETTO, AND M. S. PAULETTI, *Parametric FEM for geometric biomembranes*, Journal of Computational Physics, 229 (2010), pp. 3171–3188. (Cited on pp. 123, 134, 135)
- [11] D. BONN, J. EGGLERS, J. INDEKEU, J. MEUNIER, AND E. ROLLEY, *Wetting and spreading*, Reviews of Modern Physics, 81 (2009), pp. 739–805. (Cited on p. 109)
- [12] A. BUNSE-GERSTNER, *An algorithm for the symmetric generalized eigenvalue problem*, Linear Algebra and Its Applications, 58 (1984), pp. 43–68. (Cited on p. 50)
- [13] M. BURGER, *A framework for the construction of level set methods for shape optimization and reconstruction*, Interfaces and Free Boundaries, 5 (2002), pp. 301–329. (Cited on pp. 88, 113, 114, 115, 121)

- [14] W. BÜRGER AND E. KUWERT, *Area-minimizing disks with free boundary and prescribed enclosed volume*, Journal für die reine und angewandte Mathematik, 2008 (2008), pp. 1–27. (Cited on p. 111)
- [15] A. CHAMBOLLE, *An algorithm for mean curvature motion*, Interfaces and Free Boundaries, 6 (2004), p. 195–218. (Cited on p. 117)
- [16] T. CHAN AND J. SHEN, *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*, SIAM, Philadelphia, PA, 2005. (Cited on pp. 117, 121, 122)
- [17] T. F. CHAN AND L. A. VESE, *Active contours without edges*, IEEE Transactions on Image Processing, 10 (2001), pp. 266–277. (Cited on pp. 117, 118, 122)
- [18] T. H. COLDING AND W. P. MINICOZZI II, *A Course in Minimal Surfaces*, Graduate Studies in Mathematics 121, American Mathematical Society, Providence, RI, 2011. (Cited on pp. 105, 107)
- [19] R. COURANT AND N. DAVIDS, *Minimal surfaces spanning closed manifolds*, Proceedings of the National Academy of Sciences of the United States of America, 26 (1940), pp. 194–199. (Cited on p. 111)
- [20] P.-G. DE GENNES, *Wetting: Statics and dynamics*, Reviews of Modern Physics, 57 (1985), pp. 827–863. (Cited on pp. 9, 107, 109, 110, 111)
- [21] P.-G. DE GENNES, F. BROCHARD-WYART, AND D. QUÉRÉ, *Capillarity and Wetting Phenomena: Drops, Bubbles, Pearls, Waves*, Springer-Verlag, New York, NY, 2004. (Cited on pp. 9, 107, 109, 110, 111)
- [22] K. DECKELNICK, G. DZIUK, AND C. M. ELLIOTT, *Computation of geometric partial differential equations and mean curvature flow*, Acta Numerica, 14 (2005), pp. 139–232. (Cited on pp. 7, 117)
- [23] M. C. DELFOUR AND J.-P. ZOLÉSIO, *Shapes and Geometries: Analysis, Differential Calculus, and Optimization*, 2nd ed., Advances in Design and Control 22, SIAM, Philadelphia, PA, 2011. (Cited on pp. 3, 7, 28, 87, 90, 104, 115)
- [24] M. P. DO CARMO, *Differential Geometry of Curves and Surfaces*, Prentice-Hall, Englewood Cliffs, NJ, 1976. (Cited on pp. 7, 19, 22, 23, 27, 30, 40, 42, 56, 105, 109)
- [25] M. P. DO CARMO, *Riemannian Geometry*, Mathematics: Theory and Applications, Birkhäuser, Boston, MA, 1992. (Cited on pp. 7, 16, 56, 76)
- [26] G. DOĞAN, *A Variational Shape Optimization Framework for Image Segmentation*, Ph.D. thesis, University of Maryland, College Park, MD, 2006. (Cited on pp. 7, 28)
- [27] G. DOĞAN, P. MORIN, AND R. H. NOCHETTO, *A variational shape optimization approach for image segmentation with a Mumford–Shah functional*, SIAM Journal Scientific Computing, 30 (2008), pp. 3028–3049. (Cited on pp. 28, 113)
- [28] G. DOĞAN, P. MORIN, R. H. NOCHETTO, AND M. VERANI, *Discrete gradient flows for shape optimization and applications*, Computer Methods in Applied Mechanics and Engineering, 196 (2007), pp. 3898–3914. (Cited on pp. 88, 113, 114, 115, 117, 120, 121)
- [29] Q. DU, C. LIU, R. RYHAM, AND X. WANG, *A phase field formulation of the Willmore problem*, Nonlinearity, 18 (2005), pp. 1249–1267. (Cited on p. 123)
- [30] Q. DU, C. LIU, AND X. WANG, *A phase field approach in the numerical study of the elastic bending energy for vesicle membranes*, Journal of Computational Physics, 198 (2004), pp. 450–468. (Cited on pp. 123, 135)

- [31] E. B. DUSSANY, *The moving contact line: The slip boundary condition*, Journal of Fluid Mechanics, 77 (1976), pp. 665–684. (Cited on p. 113)
- [32] K. ECKER AND G. HUISKEN, *Mean curvature evolution of entire graphs*, Annals of Mathematics, 130 (1989), pp. 453–471. (Cited on p. 115)
- [33] C. M. ELLIOTT AND B. STINNER, *Modeling and computation of two phase geometric biomembranes using surface finite elements*, Journal of Computational Physics, 229 (2010), pp. 6585–6612. (Cited on pp. 123, 135)
- [34] C. M. ELLIOTT, B. STINNER, AND C. VENKATARAMAN, *Modelling cell motility and chemotaxis with evolving surface finite elements*, Journal of the Royal Society Interface, 9 (2012), pp. 3027–3044. (Cited on p. 135)
- [35] S. ESEDOGLU, A. RÄTZ, AND M. RÖGER, *Colliding interfaces in old and new diffuse-interface approximations of Willmore-flow*, Communications in Mathematical Sciences, 12 (2014), pp. 125–147. (Cited on p. 123)
- [36] G. EVANS, J. BLACKLEDGE, AND P. YARDLEY, *Numerical Methods for Partial Differential Equations*, Springer-Verlag, London, UK, 2000. (Cited on p. 85)
- [37] L. C. EVANS, *Partial Differential Equations*, American Mathematical Society, Providence, RI, 1998. (Cited on pp. 76, 83, 84, 111, 114, 116)
- [38] L. C. EVANS AND J. SPRUCK, *Motion of level sets by mean curvature. I*, Journal of Differential Geometry, 33 (1991), pp. 635–681. (Cited on pp. 115, 116)
- [39] M. M. FALL, *Embedded disc-type surfaces with large constant mean curvature and free boundaries*, Communications in Contemporary Mathematics, 14 (2012), 1250037. (Cited on p. 111)
- [40] M. E. GAGE, *An isoperimetric inequality with applications to curve shortening*, Duke Mathematical Journal, 50 (1983), pp. 1225–1229. (Cited on p. 115)
- [41] M. GIAQUINTA AND S. HILDEBRANDT, *Calculus of Variations I: The Lagrangian Formalism*, Grundlehren der mathematischen Wissenschaften 310, Springer-Verlag, Berlin, Germany, 1996. (Cited on pp. 111, 116)
- [42] R. GLOWINSKI AND O. PIRONNEAU, *On the numerical computation of the minimum-drag profile in laminar flow*, Journal of Fluid Mechanics, 72 (1975), pp. 385–389. (Cited on p. 4)
- [43] M. A. GRAYSON, *The heat equation shrinks embedded plane curves to round points*, Journal of Differential Geometry, 26 (1987), pp. 285–314. (Cited on p. 115)
- [44] M. GROMOV, *Sign and geometric meaning of curvature*, Rendiconti del Seminario Matematico e Fisico di Milano, 61 (1991), pp. 9–123. (Cited on pp. 23, 92)
- [45] R. D. GULLIVER, *Regularity of minimizing surfaces of prescribed mean curvature*, Annals of Mathematics, 97 (1973), pp. 275–305. (Cited on pp. 109, 111)
- [46] M. D. GUNZBURGER, *Perspectives in Flow Control and Optimization*, SIAM, Philadelphia, PA, 2003. (Cited on p. 104)
- [47] M. D. GUNZBURGER AND H. KIM, *Existence of an optimal solution of a shape control problem for the stationary Navier–Stokes equations*, SIAM Journal on Control and Optimization, 36 (1998), pp. 895–909. (Cited on p. 115)
- [48] J. HASLINGER AND R. A. E. MÄKINEN, *Introduction to Shape Optimization: Theory, Approximation, and Computation*, Advances in Design and Control 7, SIAM, Philadelphia, PA, 2003. (Cited on pp. 3, 115)

- [49] F. B. HILDEBRAND, *Introduction to Numerical Analysis*, 2nd ed., Dover, Mineola, NY, 1987. (Cited on p. 84)
- [50] M. HINTERMÜLLER AND W. RING, *A second order shape optimization approach for image segmentation*, SIAM Journal on Applied Mathematics, 64 (2003), pp. 442–467. (Cited on pp. 113, 115)
- [51] M. HINZE, R. PINNAU, M. ULRICH, AND S. ULRICH, *Optimization with PDE Constraints*, Mathematical Modeling: Theory and Applications 23, Springer, New York, NY, 2009. (Cited on pp. 3, 104, 114, 121)
- [52] C. HUH AND L. E. SCRIVEN, *Hydrodynamic model of steady movement of a solid/liquid/fluid contact line*, Journal of Colloid and Interface Science, 35 (1971), pp. 85–101. (Cited on p. 113)
- [53] C. ISENBERG, *The Science of Soap Films and Soap Bubbles*, Dover, Mineola, NY, 1992. (Cited on pp. 105, 107)
- [54] K. ITO AND K. KUNISCH, *Lagrange Multiplier Approach to Variational Problems and Applications*, Advances in Design and Control 15, SIAM, Philadelphia, PA, 2008. (Cited on pp. 3, 104, 108, 114)
- [55] J. T. JENKINS, *Static equilibrium configurations of a model red blood cell*, Journal of Mathematical Biology, 4 (1977), pp. 149–169. (Cited on pp. 9, 123, 134)
- [56] J. F. JOANNY AND M. O. ROBBINS, *Motion of a contact line on a heterogeneous surface*, Journal of Chemical Physics, 92 (1990), pp. 3206–3212. (Cited on p. 113)
- [57] J. JOST, *Embedded minimal disks with a free boundary on a polyhedron in \mathbb{R}^3* , Mathematische Zeitschrift, 199 (1988), pp. 311–320. (Cited on p. 111)
- [58] J. JOST AND X. LI-JOST, *Calculus of Variations*, Cambridge University Press, Cambridge, UK, 1998. (Cited on pp. 111, 116)
- [59] E. E. KEAVNEY, S. W. WALKER, AND M. J. SHELLEY, *Optimization of chiral structures for microscale propulsion*, Nano Letters, 13 (2013), pp. 531–537. (Cited on pp. 3, 104)
- [60] E. KREYSZIG, *Differential Geometry*, Dover, Mineola, NY, 1991. (Cited on pp. 7, 22, 56)
- [61] S. LANG, *A First Course in Calculus*, 5th ed., Undergraduate Texts in Mathematics, Springer, New York, NY, 1986. (Cited on pp. 7, 9, 11)
- [62] S. LANG, *Linear Algebra*, 3rd ed., Undergraduate Texts in Mathematics, Springer, New York, NY, 1987. (Cited on pp. 7, 9, 16, 52)
- [63] S. LANG, *Real and Functional Analysis*, 3rd ed., Graduate Texts in Mathematics, Springer, New York, NY, 1993. (Cited on pp. 11, 114)
- [64] S. LANG, *Undergraduate Analysis*, 2nd ed., Undergraduate Texts in Mathematics, Springer, New York, NY, 1997. (Cited on pp. 9, 11, 21, 84)
- [65] P. D. LAX, *Functional Analysis*, Wiley-Interscience, New York, NY, 2002. (Cited on p. 114)
- [66] A. LAURAIN AND S. W. WALKER, *Droplet footprint control*, SIAM Journal on Control and Optimization, 53 (2015), pp. 771–799 (Cited on p. 109)
- [67] S. LUCKHAUS AND T. STURZENHECKER, *Implicit time discretization for the mean curvature flow equation*, Calculus of Variations and Partial Differential Equations, 3 (1995), pp. 253–271. (Cited on p. 115)

- [68] R. S. MARTIN AND J. H. WILKINSON, *Reduction of the symmetric eigenproblem $Ax = \lambda Bx$ and related problems to standard form*, in Linear Algebra, Handbook for Automatic Computation 2, F. L. Bauer et al., eds., Springer-Verlag, Berlin, Germany, 1971, pp. 303–314. (Cited on p. 50)
- [69] B. MOHAMMADI AND O. PIRONNEAU, *Applied Shape Optimization for Fluids*, Numerical Mathematics and Scientific Computation, Clarendon Press, Oxford, UK, 2001. (Cited on p. 4)
- [70] C. B. MOLER AND G. STEWART, *An algorithm for generalized matrix eigenvalue problems*, SIAM Journal on Numerical Analysis, 10 (1973), pp. 241–256. (Cited on p. 50)
- [71] J. D. MOORE AND T. SCHULTE, *Minimal disks and compact hypersurfaces in Euclidean space*, Proceedings of the American Mathematical Society, 94 (1985), pp. 321–328. (Cited on p. 111)
- [72] K. W. MORTON AND D. F. MAYER, *Numerical Solution of Partial Differential Equations*, Cambridge University Press, Cambridge, UK, 1994. (Cited on p. 85)
- [73] D. MUMFORD AND J. SHAH, *Optimal approximations by piecewise smooth functions and associated variational problems*, Communications on Pure and Applied Mathematics, 42 (1989), pp. 577–685. (Cited on pp. 117, 122)
- [74] C. G. NGAN AND E. B. DUSSANV, *The moving contact line with a 180° advancing contact angle*, Physics of Fluids, 27 (1984), pp. 2785–2787. (Cited on p. 113)
- [75] J. NOCEDAL AND S. J. WRIGHT, *Numerical Optimization*, 2nd ed., Springer Series in Operations Research, Springer, New York, NY, 2006. (Cited on pp. 3, 113)
- [76] J. OPREA, *The Mathematics of Soap Films: Explorations with Maple*, Student Mathematical Library 10, American Mathematical Society, Providence, RI, 2000. (Cited on pp. 105, 107)
- [77] S. OSHER AND R. FEDKIW, *Level Set Methods and Dynamic Implicit Surfaces*, Springer-Verlag, New York, NY, 2003. (Cited on p. 122)
- [78] J. F. PADDAY, ED., *Wetting, Spreading, and Adhesion*, Academic Press, London, UK, 1978. (Cited on p. 109)
- [79] R. L. PANTON, *Incompressible Flow*, 2nd ed., John Wiley and Sons, Inc., New York, NY, 1996. (Cited on p. 3)
- [80] A. PATERSON AND M. FERMIGIER, *Wetting of heterogeneous surfaces: Influence of defect interactions*, Physics of Fluids, 9 (1997), pp. 2210–2216. (Cited on p. 109)
- [81] G. PETERS AND J. H. WILKINSON, *$Ax = \lambda Bx$ and the generalized eigenproblem*, SIAM Journal on Numerical Analysis, 7 (1970), pp. 479–492. (Cited on p. 50)
- [82] O. PIRONNEAU, *On optimum profiles in Stokes flow*, Journal of Fluid Mechanics, 59 (1973), pp. 117–128. (Cited on p. 4)
- [83] O. PIRONNEAU, *On optimum design in fluid mechanics*, Journal of Fluid Mechanics, 64 (1974), pp. 97–110. (Cited on p. 4)
- [84] O. PIRONNEAU, *Optimal Shape Design for Elliptic Systems*, Springer Series in Computational Physics, Springer-Verlag, New York, NY, 1984. (Cited on pp. 4, 87, 104, 115)
- [85] A. QUARTERONI, R. SACCO, AND F. SALERI, *Numerical Mathematics*, Texts in Applied Mathematics 37, Springer, Berlin, Germany, 2000. (Cited on p. 84)
- [86] A. RALSTON AND P. RABINOWITZ, *A First Course in Numerical Analysis*, 2nd ed., Dover, Mineola, NY, 2001. (Cited on p. 84)

- [87] A. ROS AND R. SOUAM, *On stability of capillary surfaces in a ball*, Pacific Journal of Mathematics, 178 (1997), pp. 345–361. (Cited on p. 111)
- [88] R. E. RUSU, *An algorithm for the elastic flow of surfaces*, Interfaces and Free Boundaries, 7 (2005), pp. 229–239. (Cited on p. 7)
- [89] S. A. SAFRAN, *Statistical Thermodynamics of Surfaces, Interfaces, and Membranes*, Frontiers in Physics 90, Westview Press, Berkeley, CA, 2003. (Cited on pp. 109, 111, 123, 133)
- [90] U. SIEFERT, *Fluid membranes in hydrodynamic flow fields: Formalism and an application to fluctuating quasispherical vesicles in shear flow*, The European Physical Journal B, 8 (1999), pp. 405–415. (Cited on pp. 9, 123, 133)
- [91] P. SMEREKA, *Semi-implicit level set methods for curvature and surface diffusion motion*, Journal of Scientific Computing, 19 (2003), pp. 439–456. (Cited on p. 117)
- [92] G. D. SMITH, *Numerical Solution of Partial Differential Equations*, 3rd ed., Oxford University Press, Oxford, UK, 1985. (Cited on p. 85)
- [93] J. SOKOŁOWSKI AND J.-P. ZOLÉSIO, *Introduction to Shape Optimization*, Springer Series in Computational Mathematics, Springer-Verlag, Berlin, Germany, 1992. (Cited on pp. 3, 7, 28, 87, 104)
- [94] K. STEFFEN AND H. C. WENTE, *The non-existence of branch points in solutions to certain classes of plateau type variational problems*, Mathematische Zeitschrift, 163 (1978), pp. 211–238. (Cited on p. 111)
- [95] G. W. STEWART, *Afternotes on Numerical Analysis*, SIAM, Philadelphia, PA, 1996. (Cited on p. 84)
- [96] J. STOER AND R. BULIRSCH, *Introduction to Numerical Analysis*, 3rd ed., Texts in Applied Mathematics 12, Springer, Berlin, Germany, 2002. (Cited on p. 84)
- [97] G. STRANG, *Linear Algebra and Its Applications*, 3rd ed., Brooks/Cole, New York, NY, 1988. (Cited on pp. 7, 16, 52, 80)
- [98] M. STRUWE, *The existence of surfaces of constant mean curvature with free boundaries*, Acta Mathematica, 160 (1988), pp. 19–64. (Cited on p. 111)
- [99] E. SÜLI AND D. MAYER, *An Introduction to Numerical Analysis*, Cambridge University Press, Cambridge, UK, 2003. (Cited on p. 84)
- [100] L. TARTAR, *An Introduction to Sobolev Spaces and Interpolation Spaces*, Lecture Notes of the Unione Matematica Italiana 3, Springer, Berlin, Germany, 2007. (Cited on p. 114)
- [101] R. TEMAM, *Navier-Stokes Equations: Theory and Numerical Analysis*, AMS Chelsea Publishing, Providence, RI, 2001. Reprint of the 1984 edition. (Cited on p. 3)
- [102] P. A. THOMPSON AND M. O. ROBBINS, *Simulations of contact-line motion: Slip and the dynamic contact angle*, Physical Review Letters, 63 (1989), pp. 766–769. (Cited on p. 113)
- [103] L. N. TREFETHEN AND D. BAU III, *Numerical Linear Algebra*, SIAM, Philadelphia, PA, 1997. (Cited on pp. 7, 16, 80, 84)
- [104] F. TRÖLTZSCH, *Optimal Control of Partial Differential Equations: Theory, Methods, and Applications*, Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2010. (Cited on p. 115)

- [105] S. W. WALKER, *A mixed formulation of a sharp interface model of Stokes flow with moving contact lines*, ESAIM: Mathematical Modelling and Numerical Analysis, 48 (2014), pp. 969–1009. (Cited on p. 113)
- [106] S. W. WALKER AND E. E. KEAVENY, *Analysis of shape optimization for magnetic microswimmers*, SIAM Journal on Control and Optimization, 51 (2013), pp. 3093–3126. (Cited on pp. 3, 104, 113, 115)
- [107] S. W. WALKER AND M. J. SHELLEY, *Shape optimization of peristaltic pumping*, Journal of Computational Physics, 229 (2010), pp. 1260–1291. (Cited on pp. 3, 104, 113, 114)
- [108] J. R. WEEKS, *The Shape of Space*, 2nd ed., Chapman & Hall/CRC Pure and Applied Mathematics 249, CRC Press, London, UK, 2001. (Cited on p. 28)
- [109] H. C. WENTE, *A general existence theorem for surfaces of constant mean curvature*, Mathematische Zeitschrift, 120 (1971), pp. 277–288. (Cited on pp. 109, 111)
- [110] T. J. WILLMORE, *Total mean curvature squared of surfaces*, in Geometry and Topology of Submanifolds, VIII (Brussels, 1995/Nordfjordeid, 1995), World Sci. Publ., River Edge, NJ, 1996, pp. 383–391. (Cited on pp. 7, 123)
- [111] E. WOLFRAM AND R. FAUST, *Liquid drops on a tilted plate, contact angle hysteresis and the Young contact angle*, in Wetting, Spreading and Adhesion, J. F. Padday, ed., Academic Press, London, UK, 1978, pp. 213–222. (Cited on p. 109)
- [112] E. C. ZACHMANOGLOU AND D. W. THOE, *Introduction to Partial Differential Equations with Applications*, Dover, Mineola, NY, 1987. (Cited on pp. 7, 83, 84, 116)
- [113] Z. OU-YANG, *Elastic theory of biomembranes*, Thin Solid Films, 393 (2001), pp. 19–23. (Cited on pp. 123, 133)
- [114] Z. OU-YANG AND W. HELFRICH, *Bending energy of vesicle membranes: General expressions for the first, second, and third variation of the shape energy and applications to spheres and cylinders*, Physical Reviews A, 39 (1989), pp. 5280–5288. (Cited on pp. 9, 123, 133)

Index

- area, 42
 cl , 92
 cof, 139
 κ_G , 52
 κ_1, κ_2 , 50
 k_n , 48
 κ , 52
 κ, κ , 52
 $d\mathbf{x}, dS, d\alpha, 8$
 dist, 23
 Ω_{all} , 89
 ∂ , 10
 Σ , 28, 33
 Ω, Γ, Σ , 8
 Γ , 87
 $\Omega, 82, 87$
 $\cdot, 137$
 $\cdot, 138$
 θ_{cl} , 109
 $\nu, \nu_s, \mathbf{b}, \mathbf{b}_s, \mathbf{b}_{s,g}$, 109
 f_{ext} , 92
 I_P , 34
 Π_P , 47
 $\text{id}_{\{\cdot\}}$, 8
 \mathbf{I} , 138
 I, \mathcal{D} , 117
 Ω_i, Ω_o , 118
 J , 16
 Δ , 83
 ϕ , 23
 Φ , 14
 \mathbf{X} , 14
 α , 14
 $D_{\{\cdot\}}$, 90
 $\nu, 2, 38$
 \mathbf{n} , 48
 $\mathbb{R}^{k \times n}$, 137
 \mathbb{R}, \mathbb{R}^n , 1
 J_b , 123
 $\mathcal{E}, \mathcal{J}, \mathcal{B}$, 88
 J_d , 4
- \mathcal{A} , 109
 \mathcal{L} , 108
 sgn (signum), 39
 ∇ , 8
 Δ , 67
 ∇_Γ , 62
 ∇_Σ , 64
 Δ_Γ , 67
 Δ_Σ , 73
 $\gamma, \gamma_s, \gamma_{s,g}$, 110
 τ , 48
 $T_x(\Gamma)$, 30
 Δt , 115
 trace, 138
 T , 7
 $T_b(\cdot)$, 92
 \mathbf{V} , 2, 90
 \mathbb{V} , 113
 $\mathbf{0}$, 137
- active contours, 119
 affine map, 13
 allowable coordinate transformation, 18
 arc-length, 33
 area functional, 105, 116
 atlas, 21
- ball, 10
 bending energy
 Euler–Lagrange equation, 129
 functional, 123
 general perturbations, 132
 gradient flow, 134
 normal perturbations, 129
 volume and area constraints,
 133
- bijective, 12
 boundary of a set, 10
 boundary, surface without, 22
- Chan–Vese functional, 118
- change of parameters, 27
 change of variables
 for gradient, 142
 for integrals, 142
 chart, 19
 closed surface, 23
 compact, 21
 definition, 11
 support, 11
 compactly contained, 11
 constant curvature surface, 108
 constant normal extension, 92
 continuous, 13
 coordinate curves, 29
 curvature
 Gauss, 52, 73
 mean, 52
 summed, 52, 71
 vector, 52, 71
 curvature (1-D curve), 48
 curvature tensor, 45
 curvature tensor transformation, 47
- derivative
 material, 90, 92
 of determinant, 141
 of inverse matrix, 141
 shape, 2, 91, 93
 determinant, expansion of the, 141
- diffeomorphism, 88
 differentiable, 59
 differentiable map, 88
 domain regularity, 87
 drag minimization, 3
 droplet equilibrium, 111
- Euclidean space, 9
 extension, 92
 extrinsic vs. intrinsic, 5

- first order optimality conditions, 108
fundamental form
 first, 34
 second, 44, 47
- gradient
 Euclidean, 8
 of inverse map, 93
 shape, 88
 surface, 62
- gradient flow velocity, 113
gradient flow velocity space, 114
gradient flow, time-discrete, 115
- hold-all domain, 89
homeomorphism, 13, 19
- image features, 118
image point, 12
image segmentation, 117
immersion, 16
injective, 12
integral notation, 8
integration by parts, 74
inverse image point, 12
isoperimetric inequality, 109
- Jacobian, 16, 18
- Lagrangian shape functional, 108
Laplace–Beltrami operator, 67
level set, 23
- mapping, 12
 inverse, 12
- material derivative
 definition, 90, 92
 normal vector, 96
 summed curvature squared, 123
 summed curvature vector, 97
 surface Jacobian, 96
- material point, 13
- matrices, 137
mean curvature flow, 116
metric tensor, 36
metric tensor transformation, 38
minimal surface, 106
- normal curvature, 48
normal vector, 38
notation, 7
- orientable surface, 40
orthogonal coordinates, 42
orthogonal transformation, 13
- parameterization, allowable, 16
parametric
 representation, 15
 surface, 17
- perturbation
 domain, 2, 89
 identity, 90
 shape, 2, 101
- prescribed mean curvature
 equation, 107
- principal curvatures and
 directions, 50
- red blood cell, 134
reference domain, 15
regular point, 17
regular surface, 19
Riemannian space, 36
rigid motion, 13
- sets (basics), 10
shape derivative
 definition, 91, 93
 normal vector, 99
 position vector, 98
 summed curvature, 100
- shape descent direction, 114
shape functional, 88, 101
 perturbation, 101
 perturbation structure
 theorem, 104
- shape operator, 51, 78
shape optimization, 3, 113
shape perturbation surface
 measure, 98
shape sensitivity, 101
soap film, 105
substrate surface tension, 109
support, 11
surface, 14
surface area, 42
surface divergence, 66
 of normal vector, 80
surface gradient, 62
 of normal vector, 78
surface heat equation, 83
surface Laplacian, 67
 of normal vector, 80
surface partial differential
 equations (PDEs), 82
surface tension, 107
surface types: elliptic,
 hyperbolic, parabolic,
 planar, 53
surface without boundary, 22
surjective, 12
- tangent plane, 17, 30
tangent plane/space, 28
tangent space projection, 65
tangent vector, 29
- tangential
 directional derivative, 61
 divergence, 66
 gradient, 62
 Laplacian, 67
 vector field, 61
- transformation, 12
tubular neighborhood, 92
- vectors, 137
volume constraint, 117
- Willmore functional, 123