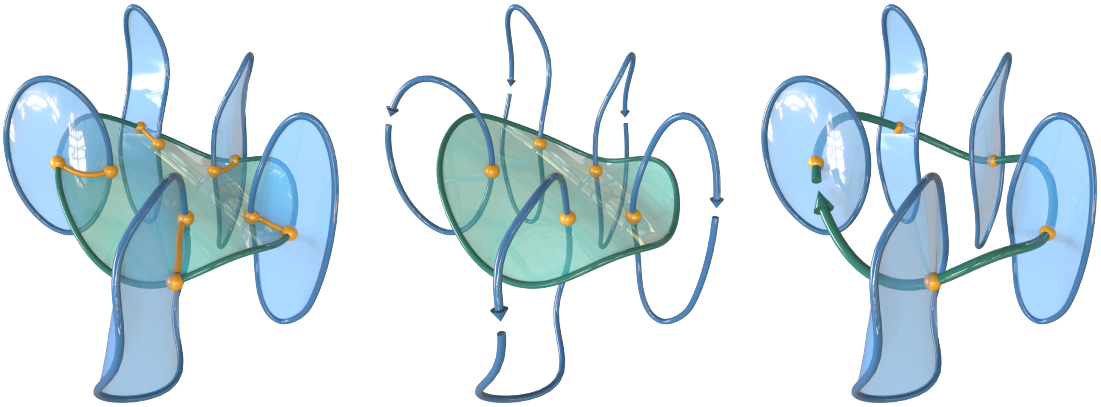
Exterior Calculus in Graphics

Course notes for a SIGGRAPH 2023 course

Stephanie Wang, Mohammad Sina Nabizadeh, and Albert Chern



Version 1.00: June 6, 2023
Latest version available at
https://stephaniewang.page/ExteriorCalculusInGraphics/

SIGGRAPH '23 Courses, August 06-10, 2023, Los Angeles, CA, USA
ACM 979-8-4007-0145-0/23/08.
10.1145/3587423.3595525

Abstract

The demand for a more advanced multivariable calculus has rapidly increased in computer graphics research, such as physical simulation, geometry synthesis, and differentiable rendering. Researchers in computer graphics often have to turn to references outside of graphics research to study identities such as the Reynolds Transport Theorem or the geometric relationship between stress and strain tensors. This course presents a comprehensive introduction to exterior calculus, which covers many of these advanced topics in a geometrically intuitive manner. The course targets anyone who knows undergraduate-level multivariable calculus and linear algebra and assumes no more prerequisites. Contrary to the existing references, which only serve the pure math or engineering communities, we use timely and relevant graphics examples to illustrate the theory of exterior calculus. We also provide accessible explanations to several advanced topics, including continuum mechanics, fluid dynamics, and geometric optimizations. The course is organized into two main sections: a lecture on the core exterior calculus notions and identities with short examples of graphics applications, and a series of mini-lectures on graphics topics using exterior calculus.

Preface

Disclaimer

These notes are still being developed and they contain mistakes. Please visit the course project page at https://stephaniewang.page/ExteriorCalculusInGraphics/ for the newest version. We encourage the readers to examine the statements and proofs in these notes and determine the correctness themselves. If you come across any errors / omissions, please email the authors while providing the page number and notes version (available on the title page).

Prerequisites

This is a comprehensive beginner-level course to introduce *exterior calculus* to students, researchers, educators, and developers. The course prerequisites include only undergraduate-level linear algebra and multivariable calculus.

Course Rationale

The course aims to extend the mathematical tool set for the audiences to gain short cuts and geometric intuitions through a variety of important advancements in modern computer graphics. The scope where exterior calculus applies is not limited to geometry processing which has been taught in computer graphics [Grinspun et al.(2006), Crane et al.(2013)]. In particular, we cover examples including differentiable rendering, optimization techniques, continuum mechanics, optimal transports, and shape synthesis. This course serves to fulfill the following goals:

I. *Fill in the missing episodes of the mathematical language for graphics.* Multivariable calculus and linear algebra lie at the center of continuous mathematics,

which is the pillar of computer graphics that aims at synthesizing our continuous world. Furthermore, differential calculus has become increasingly important in the recent developments of major areas of computer graphics such as rendering (*e.g.* differentiable rendering) and simulations (*e.g.* constitutive models). However, the computation of differentials can appear less than intuitive using the traditional tools from multivariable calculus. For instance, how does one differentiate a surface integral $\iint_\Sigma f \, dA$ regarding perturbations in the integration domain Σ? How does one differentiate a Dirac-δ function relative to its placement? These questions naturally arise from the sensitivity analysis of occlusion geometry in rendering and the differentiation of discontinuity and collision in simulations. While the standard multivariable calculus curriculum covers *partial derivatives* using coordinates, *double and triple integrals*, *gradient operators*, and a few variations of *Stokes Theorems*, there remains a gap between elementary vector calculus and cutting-edge graphics research. This course provides a bridge to cross this gap by building on top of freshman calculus and linear algebra and explaining various aspects of *exterior calculus* using examples in graphics.

II. *Provide a reference for exterior calculus.* Most existing texts on differential forms and exterior calculus are written for the community of differential geometers. While a few texts do provide engineering-related expositions [Flanders(1963), Frankel(2011)], most coverage of continuum mechanics does not include physical simulations or rendering references. More and more graphics papers (*e.g.* [Wang and Chern(2021)] in geometry optimization, [Sellán et al.(2021), Yang et al.(2021)] in physics simulations, [Nicolet et al.(2021), Zhao et al.(2020a)] for inverse rendering) use the language of geometric calculus; there has been an increasing demand for a comprehensive reference for exterior calculus in the graphics community. This course serves to fulfil the demand by publishing an accompanying course notes that document a list of exterior calculus notions and identities.

III. *Illustrate continuous, geometric intuition to differential forms.* Past appearance of exterior calculus has been limited to courses covering discrete differential geometry and geometry processing [Grinspun et al.(2006), Crane et al.(2013), Crane(2018)]. These material focuses on a discrete and computational aspect of exterior calculus. In particular, differential forms are described as data assigned to points, edges, faces, *etc.* Expanding these discrete ideas to continuous ones is not immediately obvious. This course introduces geometric interpretations for the continuous exterior calculus that best illustrate the structure dependency. With these interpretations in mind, one gains more accurate intuitions to the important concepts in simulations and optimization methods.

Course Contents

The course is organized into two main sections. The first section consists of core technical exposition of exterior calculus with short applications in graphics along the

way. The second section is a few application-focused topics formulated with exterior calculus.

Part I: Fundamentals

The course starts with the linear algebra of *vectors* and *covectors*. Vectors are *arrows* that represent direction, whereas covectors are *contour planes* that represent slopes. These two concepts are distinct. The significant application for vectors and covectors is the optimization method of *gradient descent*. How does one translate the information of slope of an objective function to an update direction in an iterative optimization? It turns out that the method is not unique. Leveraging the distinction between vectors and covectors can lead to efficient optimization schemes (see Example 3.9.)

The course moves on to *differential forms* and introduces three pictures of differential forms:

(i) *Alternating multilinear form.* A direct generalization of covectors from contour planes to lines and points. This picture is the linear algebra aspect of differential forms.

(ii) *Fields to be integrated along geometries.* Differential forms are to be paired with test geometries. For example, *flux* and *stress* is to be evaluated on a piece of surface. Calculus operators are adjoint operators of geometric operations across the form-geometry pairing. This picture has been integrated in the computation tools of Discrete Exterior Calculus.

(iii) *Codimensional geometry.* A point cloud represents a probability measure. Similarly, a curve cloud, surface cloud, *etc.* are differential forms. Calculus operators directly translate to geometric operations on these codimensional geometries. This picture is the Geometric Measure Theory aspect of differential forms.

Exterior calculus operations including the wedge product, interior product, exterior derivative, and pullback are associated to geometric operations of intersection, extrusion, boundary, and deformation, respectively. The course demonstrates that many vector identities are examples of self-evident topological statements. The differentiation of these calculus operations are also covered. The course also covers Lie derivatives and their applications in continuum mechanics.

Part II: Graphics Topics in Exterior Calculus

While short applications are mentioned along the previous part, the second part of the course focuses on a few applied topics.

(i) *Chapter 3: Geometric optimizations.* Optimization problems involving geometry is a crucial bit of computer graphics research. We discuss different geometry representations and how they affect the downstream computations of geometric properties like boundary and topological changes. We also provide a *geometric picture* of the theory of optimizations, where readers will find out about the important of choosing the correct metric for their *gradient descent method*.

(ii) *Chapter 4: Continuum Mechanics.* Hyperelasticity models are the fundamental physical models for soft body simulations. These physical models usually involve many matrix fields representing the stress and strain. The course decodes these matrices and describe them using the appropriate differential forms. The formulation makes constitutive models more transparent.

(iii) *Chapter 5: Fluid dynamics.* The inviscid incompressible fluid flow is governed by the Euler equation. By applying exterior calculus, we elucidate an alternative formulation for the Euler equation that uses the velocity covector field as the primary variable. Using this reformulation, one can obtain deeper insights in the flow of fluids and come up with better algorithm for their simulation.

Speaker Biographies

Stephanie Wang – https://stephaniewang.page/

Dr. Stephanie Wang is a postdoctoral researcher in Computer Science and Engineering at UCSD. She received her PhD in Mathematics from UCLA in 2020 with a dissertation on simulating various materials using the Material Point Method. Her research interests lie in in the intersection of mathematical analysis, scientific computing, and computer graphics. She works on problems arising from physical simulations and geometry processing by analyzing the mathematical models and improving the numerical methods. In her spare time, she enjoys walking for many hours in the mountains.

Mohammad Sina Nabizadeh – https://sinabiz.github.io/

Mohammad Sina Nabizadeh is a PhD student in Computer Science and Engineering at UCSD. There, he is advised by Professors Albert Chern and Ravi Ramamoorthi. In 2020, he received his bachelors in Computer Engineering at the University of Waterloo, Canada, where he was advised by Professor Christopher Batty at the Computational Motion Group. His interests lie in many different fields of computer graphics research such as geometric fluid dynamics, geometry processing, and rendering.

Albert Chern – https://cseweb.ucsd.edu/~alchern/

Prof. Albert Chern is an assistant professor in Computer Science and Engineering at UCSD. He received his PhD in Applied and Computational Mathematics at Caltech in 2017. He worked as a postdoctoral researcher in Mathematics at TU Berlin from 2017 to 2020 and began his assistant professorship at UCSD in 2020. Prof. Chern's research interests lie in the interplay among differential geometry, numerical partial differential equations, and their applications in geometry processing and physics simulations in computer graphics.

Acknowledgements

This work was funded by NSF CAREER Award 2239062 and UC San Diego Center for Visual Computing. Additional support was provided by the WiGRAPH Rising Stars Program, SideFX Software, and Activision Blizzard.

Contents

I Fundamentals

1. Vectors and Covectors

This chapter is a review of Linear Algebra, but with emphasis on drawing distinction between vectors and covectors. We state a few theorems from linear algebra along the exposition. Most linear algebra textbooks would state the theorems for matrices, and hence inherently basis dependent. We reprise the theorems using the language of vectors and covectors. We skip the proofs of most of these theorems. Readers can easily reconstruct the proofs by following the matrix counterpart from any matrix-based linear algebra textbook.

1.1 Dual Pairing

The **dual pairing** is a recurring operation throughout linear algebra. Linear combinations of vectors $a_1\vec{v}_1 + a_2\vec{v}_2 + a_3\vec{v}_3$ are pairing between a list of coefficients and a list of vectors. Matrix multiplications pair row vectors and column vectors:

$$\begin{bmatrix} — & \mathbf{r}_1^\intercal & — \\ — & \mathbf{r}_2^\intercal & — \end{bmatrix} \begin{bmatrix} | & | & | \\ \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 \\ | & | & | \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1^\intercal\mathbf{c}_1 & \mathbf{r}_1^\intercal\mathbf{c}_2 & \mathbf{r}_1^\intercal\mathbf{c}_3 \\ \mathbf{r}_2^\intercal\mathbf{c}_1 & \mathbf{r}_2^\intercal\mathbf{c}_2 & \mathbf{r}_2^\intercal\mathbf{c}_3 \end{bmatrix}. \tag{1.1}$$

In a linear system of equations, for example

$$\begin{bmatrix} 1\,\mathrm{kg} & 1\,\mathrm{kg} \\ 2\,\mathrm{kg} & 3\,\mathrm{kg} \end{bmatrix} \begin{bmatrix} x\,\$/\mathrm{kg} \\ y\,\$/\mathrm{kg} \end{bmatrix} = \begin{bmatrix} \$\,3 \\ \$\,8 \end{bmatrix}, \tag{1.2}$$

we see pairings between arrays quantities (rows with unit of kilograms) and an array of prices (x and y) that produce scalars in costs (right-hand side). From these examples, we observe that the pairings are between *different types* of objects. In particular the pairings are *not* inner product. It does not make sense to talk about "cosine of angles" between two vectors of different units!

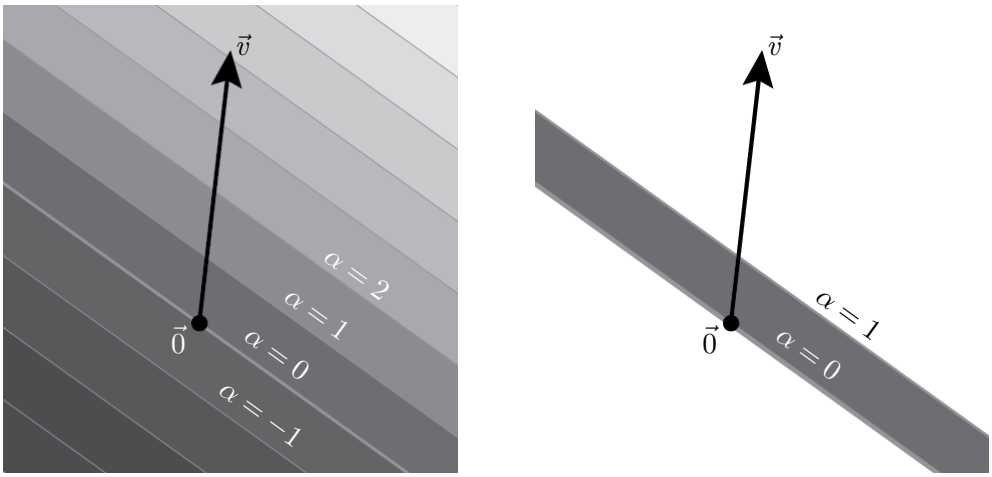


Figure 1.1 Covector-vector pairing (left) and a simplified representation of a covector using only a single slab (right). A covector α is a linear scalar function that can be visualized as a series of equidistant codimension-one planes (hyperplanes) representing its level sets. A vector is represented by an arrow based at the origin $\vec{0}$. The dual pairing of covector α and vector $\vec{v}$ is computed by counting the number of piercings of the arrow ($\vec{v}$) through the hyperplanes (α).

To distinguish the two types of objects in the linear pairing, we call one of the object **vector** and the other one **covector**. A covector is to be paired with a vector.

Definition 1.1 A covector is a scalar linear function of vectors. The evaluation, also known as the dual pairing, of a covector α and a vector $\vec{v}$ is denoted by

$$\alpha(\vec{v}) = \langle \alpha | \vec{v} \rangle = \langle \vec{v} | \alpha \rangle. \tag{1.3}$$

Definition 1.2 — Dual space. Let V be a real vector space. The space of all covectors on V is called the **dual space** $V^* = \{\alpha \colon V \xrightarrow{\text{linear}} \mathbb{R}\}$.

Theorem 1.1 For a finite dimensional space V, we have $V^{**} \cong V$.[a] In infinite dimensional cases, we may have $V^{**} \supseteq V$.

[a] This isomorphism is canonical, given by $\text{ev} \colon V \to V^{**}, \text{ev}(\vec{v}) \coloneqq (\alpha \mapsto \langle \alpha | \vec{v} \rangle)$. In particular, this is a stronger isomorphism than just having the same dimension. For example, when V is a finite dimensional space, one can say that $V \cong V^*$ because their dimensions are the same. But there is no canonical map between V and V^*.

Throughout the note, we will postulate that $V^{**} = V$.

1.1.1 Geometric picture of vectors and covectors

The standard geometric picture of a vector $\vec{v} \in V$ is an arrow based at the origin $\vec{0} \in V$. The geometric postulates for a vector space V is that it is an affine space with a distinguished point called the origin $\vec{0}$. An affine space is a space equipped with the notion of parallelism, or a projective space with a distinguished hyperplane

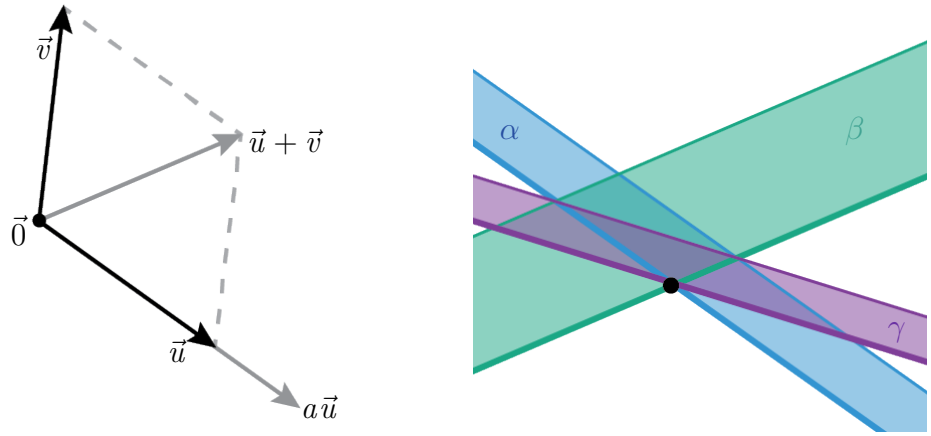


Figure 1.2 Addition of vectors (left) and covectors (right).

at infinity, or equivalently a Euclidean space without metric. The vector addition $\vec{u}+\vec{v}$ is constructed by drawing a parallelogram formed by $\vec{u}$, $\vec{v}$ and connecting the origin to its opposite vertex (see Figure 1.2, left). The scaling of a vector $a\vec{v}$ can be constructed using similar triangles and diagonal bisection in parallelograms.

The geometric picture of a covector $\alpha \in V^*$ is a foliation of equidistant parallel hyperplanes in V, each of which represents a level set $\{\alpha = k\} \subset V$ of the scalar linear function α. The dual paring $\langle \alpha | \vec{v} \rangle$ measures the number of planes of α pierced through by the arrow $\vec{v}$. See Figure 1.1, left.

Since we know how to geometrically construct equidistant parallel hyperplanes (again via the parallel postulate) we only need to draw two special level sets: $\{\alpha = 0\}$ and $\{\alpha = 1\}$ for illustrating α. Note that $\{\alpha = 0\} = \ker \alpha$ passes through the origin. Hence each covector is a slab with one side leaning against the origin. See Figure 1.1, right.

The covector addition $\alpha + \beta$ is constructed by first intersecting the two slabs, which forms a parallelogram prism, followed by building a new slab whose far side from the origin passes through the diagonal of the parallelogram that is not incident to the origin (see Figure 1.2). The scaling of a covector $a\alpha$ is defined by scaling the thickness of the slab by $1/a$. In particular, the narrower the slab is, the more dense the foliation of level sets of α is, and the larger the slope the linear function α has.

1.1.2 Annihilator

It is important to note that the dual pairing is different from an inner product. The dual pairing pairs a vector and a covector, whereas the inner product measures how aligned two vectors are. We do not need the inner product structure (a metric) to perform dual pairing. In fact, many inner product related notions we have seen in a first course in linear algebra are independent of metric. Here, we give an example: the notion of "orthogonal complement" of a subspace will be metric-independent if we treat it as a subspace in the dual space.

Definition 1.3 — Annihilator. Let $U \subset V$ be a linear subspace of the vector space V. The **annihilator** $U^\circ \subset V^*$ of U is a linear subspace in the dual space defined by

$$U^\circ = \{\alpha \in V^* \,|\, \langle \alpha | \vec{u} \rangle = 0 \text{ for all } \vec{u} \in U\} \tag{1.4}$$

Theorem 1.2 $U^{\circ\circ} = U$.

1.1.3 Adjoint

A linear map $A\colon U \xrightarrow{\text{linear}} V$ from a vector space U to another vector space V induces a dual (adjoint) map from the dual target space V^* back to the dual domain U^*.

Definition 1.4 The **adjoint** of a linear map $A\colon U \xrightarrow{\text{linear}} V$ is the linear map $A^*\colon V^* \xrightarrow{\text{linear}} U^*$ defined by

$$\langle A^*\lambda | \vec{u} \rangle = \langle \lambda | A\vec{u} \rangle \quad \text{for all } \vec{u} \in U \text{ and } \lambda \in V^*. \tag{1.5}$$

Theorem 1.3 $A^{**} = A$.

In case U, V are equipped with bases that allow a matrix representation $\mathbf{A} = (A_{ij})$ for A, the adjoint A^* is simply the transpose $\mathbf{A}^\intercal$.

Definition 1.5 The **four fundamental subspaces** of a linear map $A\colon U \xrightarrow{\text{linear}} V$ are

- kernel: $\ker(A) \coloneqq \{\vec{u} \in U \,|\, A\vec{u} = 0\} \subset U$,
- image: $\operatorname{im}(A) \coloneqq \{A\vec{u} \,|\, \vec{u} \in U\} \subset V$,
- cokernel: $\ker(A^*) \coloneqq \{\lambda \in V^* \,|\, A^*\lambda = 0\} \subset V^*$,
- coimage $\operatorname{im}(A^*) \coloneqq \{A^*\lambda \,|\, \lambda \in V^*\} \subset U^*$.

Theorem 1.4 — Theorem of four fundamental subspaces.

$$\ker(A)^\circ = \operatorname{im}(A^*), \quad \operatorname{im}(A)^\circ = \ker(A^*) \tag{1.6}$$

1.2 Endomorphisms and Bilinear Forms

In the previous section, we distinguish the difference between vectors and covectors. In this section, we continue to draw the distinction between them, and separate linear transformations of type $V \xrightarrow{\text{linear}} V$ and $V \xrightarrow{\text{linear}} V^*$. While both types can be represented by a square matrix, their geometric picture are vastly different.

A linear map $A\colon V \xrightarrow{\text{linear}} V$ that maps a vector space to itself is called an **endomorphism**. For example, the identity map is an endomorphism. An eigenvalue

problem $A\vec{v} = \lambda\vec{v}$ is formulated for endomorphisms A. Powers A^k, exponentials e^A and so on are defined only for endomorphisms.

A linear map $B\colon V \xrightarrow{\text{linear}} V^*$ that maps a vector space to its dual space is called a **bilinear form**. For each $\vec{v} \in V$, the resulting value $B(\vec{v}) \in V^*$ is a covector. Recall that a covector is a scalar-valued linear function that can take another vector and return a scalar. That is, a bilinear map is to be evaluated like $B(\vec{v})(\vec{w})$ bi-linearly dependent on the two input vectors $\vec{v}, \vec{w} \in V$. In other words, a bilinear map can also be viewed as $B\colon V \times V \xrightarrow{\text{bilinear}} \mathbb{R}$.

Note that the adjoint of a bilinear form $B\colon V \xrightarrow{\text{linear}} V^*$ would have the same type $B^*\colon V \xrightarrow{\text{linear}} V^*$ (using $V^{**} = V$). Only for bilinear forms we can compare B with B^* as they are objects of the same type.

A bilinear form B is said to be **symmetric** or **self-adjoint** if $B = B^*$. Equivalently, a bilinear form is symmetric if $B(\vec{u})(\vec{v}) = B(\vec{v})(\vec{u})$ for all $\vec{u}, \vec{v} \in V$. A bilinear form is said to be **skew-symmetric** if $B = -B^*$; *i.e.* $B(\vec{u})(\vec{v}) = -B(\vec{v})(\vec{u})$.

1.2.1 Quadratic form

A symmetric bilinear form is equivalent to a quadratic form. A quadratic form $q\colon V \xrightarrow{\text{quadratic}} \mathbb{R}$ is a function such that $q(a\vec{v}) = a^2 q(\vec{v})$ for all scalar $a \in \mathbb{R}$. A bilinear form $B\colon V \xrightarrow{\text{linear}} V^*$ gives rise to a quadratic form by $q(\vec{v}) \coloneqq B(\vec{v})(\vec{v})$. Conversely, a quadratic form $q\colon V \xrightarrow{\text{quadratic}} \mathbb{R}$ gives rise to a symmetric bilinear form $B\colon V \xrightarrow{\text{linear}} V^*$ using the formula for expanding a square

$$q(\vec{u} + \vec{v}) = q(\vec{u}) + q(\vec{v}) + 2B(\vec{u})(\vec{v}). \tag{1.7}$$

The symmetric bilinear form $B(\vec{u})(\vec{v}) = \frac{1}{2}(q(\vec{u} + \vec{v}) - q(\vec{u}) - q(\vec{v}))$ is called the **polarization** of q. Therefore, a symmetric bilinear form is fundamentally the same as a quadratic form.

1.3 Vector–Covector Conversion 1: Dual Basis

Vectors and covectors are different. The former is an arrow based at the origin, and the latter is a slab bounded by a pair of parallel hyperplanes, one of which passes through the origin. Without additional structure, there is no method converting a vector to a covector.

One scenario where we can convert vectors to covectors is when we have a full set of basis vectors. Suppose V is an n-dimensional vector space. Suppose $\vec{v}_1, \dots, \vec{v}_n$ are n linearly independent vectors. Then they form a basis. Geometrically, they span a non-degenerate parallelepiped. This parallelepiped has faces giving rise to n pairs of parallel hyperplanes, one of which passes through the origin. As such, there emerges n covectors. This construction is canonical and unique. The resulting covectors is called the **dual basis** for the dual space V^*. See Figure 1.3, left.

Conversely, given n linearly independent covectors, as a basis for V^*, we can geometrically visualize them as n slabs which intersect into parallelepiped with a corner being the origin. By reading off the edge vectors of the parallelepiped based at the origin, we get the find the dual basis for V.

The following is the algebraic definition of the above geometric construction.

Theorem 1.5 — Dual basis. Let $\vec{v}_1, \dots, \vec{v}_n \in V$ be a basis for a vector space V. Then there exists a unique set of covectors $\alpha_1, \dots, \alpha_n \in V^*$ forming a basis for V^* such that

$$\langle \alpha_i | \vec{v}_j \rangle = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases} \tag{1.8}$$

If $V = \mathbb{R}^n$ and $\vec{v}_1, \dots, \vec{v}_n$ are column vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$, then the relationship between $\mathbf{v}_1, \dots, \mathbf{v}_n$ and the dual basis as row vectors $\alpha_1, \dots, \alpha_n$ is a matrix inversion

$$\begin{bmatrix} — & \alpha_1 & — \\ & \vdots & \\ — & \alpha_n & — \end{bmatrix} = \begin{bmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{bmatrix}^{-1}. \tag{1.9}$$

Note that we need a full set of basis vectors to produce a full set of basis covectors. We cannot turn a single vector to a single covector without any additional structure.

■ **Example 1.1 — Diagonalization.** $\mathbb{R}^n = \mathbb{R} \oplus \mathbb{R} \oplus \cdots \oplus \mathbb{R}$ is a special vector space called Cartesian space. Each element is an n-tuple of numbers. A map $\mathbf{D} \colon \mathbb{R}^n \to \mathbb{R}^n$ is called a **diagonal matrix** if the map is just a channel-wise scaling

$$\mathbf{D} \colon (r_1, \dots, r_n) \mapsto (\lambda_1 r_1, \dots, \lambda_n r_n). \tag{1.10}$$

A basis $\vec{v}_1, \dots, \vec{v}_n$ for a vector space V induces a map $X \colon \mathbb{R}^n \xrightarrow{\text{linear}} V$ by $X((r_1, \dots, r_n)) \coloneqq \sum_{i=1}^n r_i \vec{v}_i$. Note that the inverse map $X^{-1} \colon V \xrightarrow{\text{linear}} \mathbb{R}^n$ is given by $X^{-1}(\vec{u}) = (\langle \alpha_1 | \vec{u} \rangle, \dots \langle \alpha_n | \vec{u} \rangle)$ where $\alpha_1, \dots, \alpha_n \in V^*$ is the dual basis.

A diagonalization of an endomorphism $A \in \mathrm{End}(V)$ is to express the endomorphism by

$$A = X \circ \mathbf{D} \circ X^{-1}, \tag{1.11}$$

which is achieved by solving the eigenvalue problem.

Definition 1.6 The trivial Cartesian metric on $\mathbb{R}^n$ and $\mathbb{R}^{n*} = \mathbb{R}^n$ is the identity map $\mathbf{I}$.

If V is equipped with a metric, then a basis $X \colon \mathbb{R}^n \xrightarrow{\text{linear}} V$ is orthonormal if and only if $X \colon (\mathbb{R}^n, \mathbf{I}) \to (V, \flat)$ is an **isometry** (a map that preserves the metric).

Theorem 1.6 If $X \colon (\mathbb{R}^n, \mathbf{I}) \xrightarrow{\text{isometry}} (V, \flat)$ is an orthonormal basis, then

$$X^{-1} = X^* \circ \flat \tag{1.12}$$

■

■ **Example 1.2 — Polar decomposition.** Let U, V be two spaces equipped with

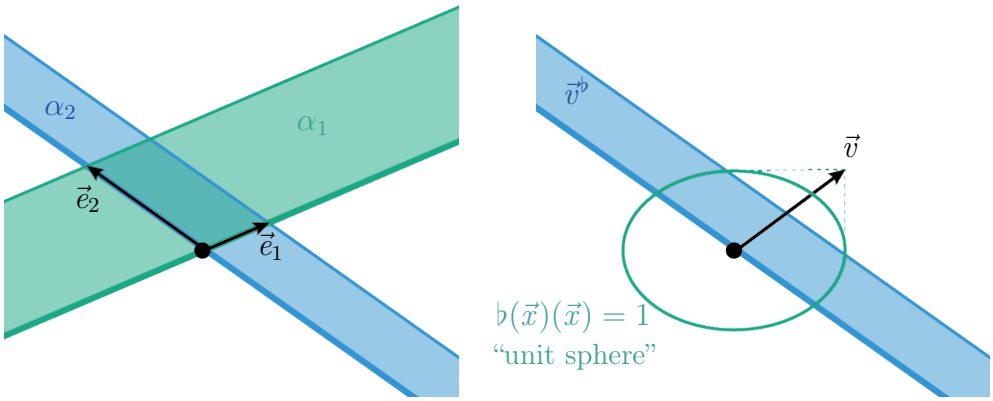


Figure 1.3 Converting covectors to vectors and vice versa using either the dual basis (left) or a metric (right).

metric $\flat_U, \flat_V$. The **polar decomposition** of a linear map $A\colon U \to V$ is

$$A = RY, \qquad \text{where } R\colon U \xrightarrow{\text{isometry}} V \text{ and } Y \in \mathrm{End}(U) \tag{1.13}$$

such that $\flat_U \circ Y$ is a symmetric positive definite bilinear form. ∎

1.4 Vector–Covector Conversion 2: Metric

A **metric** or an **inner product** is an additional structure for a vector space. The most common vector space equipped with a metric is the Euclidean space $(\mathbb{R}^n, \langle\cdot,\cdot\rangle)$ where the Euclidean inner product $\langle\vec{u},\vec{v}\rangle$ measures the length and the angle between vectors $\vec{u}$ and $\vec{v}$.

Definition 1.7 A **metric** is an injective symmetric bilinear form $\flat\colon V \xrightarrow{\text{linear}} V^*$ (injectivity means that $\ker(\flat) = \{0\}$ or that $\flat$ is invertible). We usually denote $\flat(\vec{u})(\vec{v}) = \langle\vec{u},\vec{v}\rangle$ and $\flat(\vec{u}) = \vec{u}^\flat$. The inverse $\flat^{-1}$ of the bilinear form $\flat$ is denoted by $\sharp = \flat^{-1}\colon V^* \xrightarrow{\text{linear}} V$.

The symbols $\flat, \sharp$ are called **flat** and **sharp** respectively, together they are called **musical isomorphisms**.

Recall that a symmetric bilinear form is equivalent to a quadratic form. In the case of a metric, we denote the quadratic form by $|\cdot|^2$:

$$|\vec{v}|^2 = \flat(\vec{v})(\vec{v}) = \langle\vec{v}^\flat|\vec{v}\rangle = \langle\vec{v},\vec{v}\rangle. \tag{1.14}$$

Conversely, we can derive the musical isomorphism $\flat$ from the quadratic form $|\cdot|^2$ by using the polarization identity if V is a $\mathbb{R}$-vector space:

$$\flat(\vec{u})(\vec{v}) = \langle\vec{u},\vec{v}\rangle = \frac{1}{2}\left(|\vec{u}+\vec{v}|^2 - |\vec{u}|^2 - |\vec{v}|^2\right). \tag{1.15}$$

A metric $|\cdot|^2$ on V induces a metric on V^* given by

$$|\alpha|^2_{V^*} \coloneqq \langle\alpha|\alpha^\sharp\rangle. \tag{1.16}$$

Geometrically, we can visualize a metric structure $|\cdot|^2$ on V by the "unit sphere"

$$Q \coloneqq \left\{ \vec{x} \in V \,\middle|\, |\vec{x}|^2 = 1 \right\} \subset V, \tag{1.17}$$

which is a non-degenerate quadratic surface in the affine space centered at the origin. In addition to the parallelism structure in the vector space, we have a "compass" allowing us to define distances and angles. The relationship between the arrow $\vec{v} \in V$ and the slab $\vec{v}^\flat \in V^*$ is that the tip $\vec{v}$ of the arrow and the $\{\vec{v}^\flat = 1\}$ hyperplane of the slab are in **polarity** with respect to the quadratic surface Q. See Figure 1.3, right.

Definition 1.8 In projective geometry, the **polar hyperplane** h to a point x with respect to a quadratic hypersurface Q is the hyperplane that passes through all the points on Q at which the tangent plane to Q contains x. If x is inside Q so that there is no tangent plane of Q containing x, then h is defined by the collection of points (outside of Q) whose polar hyperplane contains x. The point x is called the **pole** of the plane h with respect to Q if h is the **polar** of x.

If $\alpha = \vec{v}^\flat$, then the parallel hyperplanes of α are orthogonal to $\vec{v}$ with respect to the metric $\flat$.

Musical isomorphisms generally give different vector–covector conversions from the dual basis construction. In fact:

Theorem 1.7 Let $\vec{v}_1, \ldots, \vec{v}_n \in V$ be a basis for a vector space V and let $\alpha_1, \ldots, \alpha_n$ be its dual basis. Let $\flat$ be a metric on V. Then $\alpha_i = \vec{v}_i^\flat$ if and only if $\vec{v}_1, \ldots, \vec{v}_n$ is an **orthonormal basis** with respect to the metric $\flat$.

Definition 1.9 A metric is called **positive definite** if $|\vec{v}|^2 > 0$ for all $\vec{v} \neq 0$.

For most discussion we do not need to assume that a metric is positive definite.

■ **Example 1.3 — Eigenvalue problems for symmetric matrices.** Eigenvalue problems are posed for endomorphisms and symmetry are adjective only for bilinear forms. This makes us wonder what is going on with the famous theorem stating that "if a matrix is symmetric, then it has real eigenvalues and the eigenvectors are orthogonal." The answer is that there is a hidden metric that is often omitted. Here is the restored version of the statement.

Let B, C be two bilinear forms of type $V \xrightarrow{\text{linear}} V^*$. Then it is sensible to pose the eigenvalue problem as

$$B\vec{v} = \lambda C\vec{v}. \tag{1.18}$$

Theorem 1.8 If $\flat$ is a positive definite metric on V, and $B\colon V \xrightarrow{\text{linear}} V^*$ is a symmetric bilinear form. Then there exists an orthonormal basis $\vec{v}_1, \ldots, \vec{v}_n$ with respect to $\flat$ and real numbers $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ such that

$$B\vec{v}_i = \lambda_i \flat \vec{v}_i. \tag{1.19}$$

$\vec{v}_i$ and $\vec{v}_j$ are perpendicular under both $\flat$ and B: $\flat(\vec{v}_i)(\vec{v}_j) = B(\vec{v}_i)(\vec{v}_j) = 0$ for $i \neq j$.

Theorem 1.9 The result of Theorem 1.8 holds (except that the eigenvalues $\lambda_1, \ldots, \lambda_n$ may be complex) for the more general **normal bilinear forms**. A bilinear form $B\colon V \xrightarrow{\text{linear}} V^*$ is called normal if

$$B^* \circ \sharp \circ B = B \circ \sharp \circ B^*. \tag{1.20}$$

■

■ **Example 1.4 — Rotation matrices.** A matrix $\mathbf{R}$ is called a special orthogonal matrix or a rotation matrix if $\mathbf{R}^\intercal \mathbf{R} = \mathbf{I}$. Rotation is supposed to be an endomorphism, as it maps a vector to a rotated vector in the same space. But then the adjoint $\mathbf{R}^\intercal$ would be an endomorphism on the dual space V^*. It does not make sense to compose $\mathbf{R}^\intercal$ with $\mathbf{R}$. What is missing here is some metric dependency.

Definition 1.10 Let V be a vector space with metric $\flat\colon V \xrightarrow{\text{linear}} V^*$. An endomorphism $R \in \mathrm{End}(V)$ is said to be **orthogonal** or **unitary** if the composition of the following cyclic sequence of linear maps is the identity endomorphism

$$\begin{array}{ccc} V & \xrightarrow{R} & V \\ {\scriptstyle\sharp}\uparrow & & \downarrow{\scriptstyle\flat} \\ V^* & \xleftarrow[R^*]{} & V^* \end{array} \tag{1.21}$$

If $\det(R) = 1$, then R is called a special unitary transform. Another way to describe unitarity is that the pullback metric $R^* \circ \flat \circ R$ by R is the same as the metric $\flat = R^* \circ \flat \circ R$. In other words, it is a **linear isometry**.

Definition 1.11 In general, for two spaces U, V both equipped with metric $\flat_U, \flat_V$, a linear map $A\colon U \xrightarrow{\text{linear}} V$ is an **isometry** if

$$A^* \circ \flat_V \circ A = \flat_U. \tag{1.22}$$

Equivalently,

$$\langle A\vec{x}, A\vec{y}\rangle_V = \langle \vec{x}, \vec{y}\rangle_U \quad \text{for all } \vec{x}, \vec{y} \in U. \tag{1.23}$$

■

■ **Example 1.5 — Singular value decomposition.** Let U, V be two spaces equipped with metric $\flat_U, \flat_V$. The **singular value decomposition (SVD)** of a linear map $A\colon U \to V$ is to find an orthonormal basis $E_U\colon \mathbb{R}^m \xrightarrow{\text{linear}} U$ and an orthonormal basis $E_V\colon \mathbb{R}^n \xrightarrow{\text{linear}} V$ such that

$$A = E_V \circ \begin{bmatrix} \sigma_1 & & \\ & \sigma_2 & \\ & & \ddots \end{bmatrix} \circ \underbrace{E_U^{-1}}_{E_U^* \circ \flat_U} \tag{1.24}$$

where $\sigma_1 \geq \sigma_2 \geq \cdots$ and the diagonal matrix may be truncated to a rectangular matrix if $m \neq n$. ■

1.5 Vector Fields and Covector Fields

A typical picture of a vector field is many arrows scattered over a domain, each having a different based point. This collection of arrows represents a flow where each arrow gives away the flow direction at the location of its base point. Note that the arrows composing the vector field are based at different based points, which is in contrast to vectors in a vector space which are all based at the same origin. Therefore, to describe vector fields sensibly, we must consider a collection of vector spaces V_p indexed by the base point p in the domain. A vector field $\vec{v}$ is an assignment of a vector $\vec{v}_p \in V_p$ for every point p in the domain.

With the above picture in mind, let us introduce the standard mathematical notations for vector fields.

Let $M \subset \mathbb{R}^n$ be a region in an n-dimensional space representing a domain. For each $p \in M$, define a vector space T_pM given by a copy of $\mathbb{R}^n$. We call T_pM the **tangent space** to the domain M at p. The elements of T_pM are called **tangent vectors** at p.

The elements of T_pM are associated with *velocities* at which a particle at p can travel. For each smooth parameterized curve $\gamma\colon (-\epsilon, \epsilon) \to M$ that passes through p at $t = 0$, that is $\gamma(0) = p$, the instantaneous velocity $\frac{d\gamma}{dt}$ at $t = 0$ is an element of T_pM; that is, $\dot{\gamma}(0) \in T_pM$. Conversely, for each tangent vector $\vec{v}_p \in T_pM$, there exists (non-uniquely) a parameterized curve $\gamma\colon (-\epsilon, \epsilon) \to M$ such that $\gamma(0) = p$ and $\dot{\gamma}(0) = \vec{v}_p$.

The collection $TM = \{(p, \vec{v}_p) \mid p \in M, \vec{v}_p \in T_pM\} = \bigsqcup_{p\in M} T_pM$ of all tangent vectors at various base points p is called the **tangent bundle** over M. There is a projection operator $\pi\colon TM \to M$, $\pi\colon \vec{v}_p \mapsto p$. That is, π takes in a tangent vector $\vec{v}_p \in TM$ and returns the information $p \in M$ of which base point $\vec{v}_p$ has. The tangent space T_pM at p is the same as the preimage $\pi^{-1}(\{p\})$ of the point p via the projection. The preimage $\pi^{-1}(\{p\})$ of a point via the projection of a bundle is often called a **fiber** of the bundle.

A vector field is an assignment of an element in each fiber of the tangent bundle. Such an object is called a **section** of the bundle.

Definition 1.12 A **section** $\vec{v}$ of a bundle $\pi\colon TM \to M$ is a map $\vec{v}\colon M \to TM$ such that $\pi \circ \vec{v} = \mathrm{id}_M$. That is, $\pi(\vec{v}(p)) = p$ for all $p \in M$.

The formal definition given above is equivalent to saying that at any given point $p \in M$, the map $\vec{v}(p)$ is a vector in the tangent space T_pM of the point p, and hence $\pi(\vec{v}(p)) = p$.

The space of all sections of the tangent bundle (*i.e.* the space of all vector fields defined over M) is denoted by $\Gamma(TM)$. So, a vector field would be instanced as $\vec{v} \in \Gamma(TM)$.

A covector field $\alpha \in \Gamma(T^*M)$ is a section of the **cotangent bundle** T^*M over M. The cotangent bundle is defined such that each of its fiber T_p^*M is the dual space of the tangent space T_pM at the same base point.

1.6 Differential of a Function

One of the most important applications of distinguishing vectors and covectors is to understand the meaning of taking derivatives of a function.

Let $M \subset \mathbb{R}^n$ be a region in an n-dimensional space representing a domain, on which we will consider generic scalar-valued non-linear functions $g\colon M \to \mathbb{R}$.

Definition 1.13 The differential of a function $g\colon M \to \mathbb{R}$ at a point $p \in M$ in a domain M is a covector $dg|_p$ based at p. This covector takes in a vector $\vec{v}_p$ based at p (representing a small displacement about p) and returns the rate of change $\langle dg_p|\vec{v}_p\rangle$ of g in the direction of $\vec{v}_p$. Concretely, for each $\vec{v}_p \in T_pM$ consider a parameterized curve $\gamma\colon (-\epsilon, \epsilon) \to M$ with $\gamma(0) = p$ and $\dot{\gamma}(0) = \vec{v}_p$, and define

$$\left\langle dg|_p \middle| \vec{v}_p \right\rangle \coloneqq \left.\frac{d}{dt}\right|_{t=0} g(\gamma(t)). \tag{1.25}$$

The differential dg of a scalar function $g\colon M \to \mathbb{R}$ is a covector field $dg \in \Gamma(T^*M)$ describing the "slope" of g at every point by the best fitting linear function on every tangent space.

1.6.1 Conversion 1: to partial derivatives

A **coordinate system** on M is a set of scalar functions $x_1\colon M \to \mathbb{R}, \ldots, x_n\colon M \to \mathbb{R}$ such that the covectors $dx_1, \ldots, dx_n$ form a basis for every cotangent space T_p^*M.

Let $\vec{e}_1, \ldots, \vec{e}_n \in \Gamma(TM)$ be the dual basis of $dx_1, \ldots, dx_n$.

Definition 1.14 — Partial derivatives. The partial derivative of g along the k-th direction in a coordinate system $(x_1, \ldots, x_n)$ is defined by

$$\frac{\partial g}{\partial x_k} \coloneqq \langle dg|\vec{e}_k\rangle. \tag{1.26}$$

Equivalently, $\frac{\partial g}{\partial x_1}, \ldots, \frac{\partial g}{\partial x_n}$ are the coefficients when expressing the covector dg in

terms of the covector basis from the coordinate system

$$dg = \frac{\partial g}{\partial x_1} dx_1 + \cdots \frac{\partial g}{\partial x_n} dx_n. \tag{1.27}$$

R The partial derivative $\frac{\partial g}{\partial x_1}$ depends not only on the coordinate function x_1 but it depends on the entire coordinate system $x_1, \ldots, x_n$. This is because the construction relies on taking the dual basis. Intuitively, the partial derivative depends on which other variables are fixed during the variation.

1.6.2 Conversion 2: to gradient vector

The **gradient vector** of a function g is the vector whose direction is the steepest ascending direction and whose magnitude is the slope along that direction.

Definition 1.15 — Gradient. Let M be a domain equipped with metric $\flat$ on every tangent space. Then the **gradient** $\operatorname{grad} g$ of a function $g\colon M \to \mathbb{R}$ is defined by

$$\operatorname{grad} g \coloneqq (dg)^\sharp. \tag{1.28}$$

One may use the polarity geometric picture of the musical isomorphism about the unit sphere to see why the gradient vector points in the steepest ascending direction.

Note that the gradient of a function is independent of the coordinate. It does however depend on a choice of metric.

R In many expositions in optimization or machine learning, the gradient is defined by the array of partial derivatives

$$\operatorname{grad} g \stackrel{?}{=} \left(\frac{\partial g}{\partial x_1}, \ldots, \frac{\partial g}{\partial x_n} \right). \tag{1.29}$$

This is not true unless the coordinate system is orthonormal and therefore it is misleading. Note that partial derivatives arise from the conversion of dual basis, and the gradient arises from the conversion using a metric.

1.6.3 Pushforward of vectors

The idea of the differential of a function $dg_p\colon T_pM \xrightarrow{\text{linear}} \mathbb{R}$ of a scalar function $g\colon M \to \mathbb{R}$ can be extended for general maps.

Let $\phi\colon M \to N$ be a general nonlinear map from a space M to another space N. Then the differential $d\phi$ of the map ϕ is a linear map between the corresponding tangent spaces

$$d\phi_p\colon T_pM \xrightarrow{\text{linear}} T_{\phi(p)}N. \tag{1.30}$$

Such a linear map transforms a rate of change $\vec{v}_p \in T_pM$ at the input $p \in M$ to a rate of change of $(d\phi_p)(\vec{v}) \in T_{\phi(p)}N$ at the output $\phi(p) \in N$ (Figure 1.4). That is, $d\phi_p(\vec{v}_p)$ describes how sensitive the value of ϕ is depending on variation in the $\vec{v}$ direction.

Definition 1.16 — Pushforward. Let $\phi\colon M \to N$ be a general nonlinear map. The linear map $\phi_* = d\phi\colon TM \to TN$, with $(\phi_*)_p\colon T_pM \xrightarrow{\text{linear}} T_{\phi(p)}N$, is called the **pushforward** map of ϕ. It is defined such that for every scalar function $g\colon N \to \mathbb{R}$, the composited scalar function $f \coloneqq (g \circ \phi)\colon M \to \mathbb{R}$ has the differential given by the compositions of linear maps

$$df = dg \circ \phi_*. \tag{1.31}$$

Theorem 1.10 — Jacobian matrix. Suppose we put a coordinate system $x_1, \dots, x_m\colon M \to \mathbb{R}$ on M and a coordinate system $y_1, \dots, y_n\colon N \to \mathbb{R}$ on N. Let $\vec{a}_1, \dots, \vec{a}_m$ be the coordinate vectors as the dual basis of the covector basis $dx_1, \dots, dx_m$. Similarly, let $\vec{e}_1, \dots, \vec{e}_n$ be the dual basis of $dy_1, \dots, dy_n$. Now, call $\phi_i \coloneqq y_i \circ \phi\colon M \to \mathbb{R}$ as the resulting coordinate value of the mapping ϕ. For each vector $\vec{v}$ on M, we can write it under the basis as $\vec{v} = \sum_{i=1}^m v_i \vec{a}_i$. The resulting vector after being applied by pushforward is $\phi_*\vec{v} = \sum_{j=1}^n w_j \vec{e}_j$. Then,

$$\begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{\partial \phi_1}{\partial x_1} & \cdots & \frac{\partial \phi_1}{\partial x_m} \\ \vdots & & \vdots \\ \frac{\partial \phi_n}{\partial x_1} & \cdots & \frac{\partial \phi_n}{\partial x_m} \end{bmatrix}}_{\mathbf{F}} \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} \tag{1.32}$$

where the partial derivatives are defined in the way of Section 1.6.1. The matrix $\mathbf{F} \in \mathbb{R}^{n \times m}$ is called the **Jacobian matrix** or the **deformation gradient** of ϕ under the coordinates $(x_1, \dots, x_m)$ and $(y_1, \dots, y_n)$.

Theorem 1.11 The sensitivity analysis for a composition $\psi \circ \phi\colon M \to L$ of maps $\phi\colon M \to N$ and $\psi\colon N \to L$ is simply the composition of their pushforwards:

$$(\psi \circ \phi)_* = \psi_* \phi_*, \quad \text{or} \quad d(\psi \circ \phi)_p = d\psi_{\phi(p)}\, d\phi_p. \tag{1.33}$$

This rule is called the **chain rule**.

Evaluating the differential by composing the chain of differentials of subcomponents is the basis for the **forward mode auto-differentiation** programs.

1.6.4 Pullback of covectors

Definition 1.17 — Pullback of covectors. The adjoint

$$\phi^*\colon T^*_{\phi(p)}N \xrightarrow{\text{linear}} T^*_pM \tag{1.34}$$

of the pushforward $\phi_*\colon T_pM \xrightarrow{\text{linear}} T_{\phi(p)}N$ is called the **pullback** operator (for covectors) via the mapping $\phi\colon M \to N$. For each covector (field) α on N, $\phi^*\alpha$ becomes a covector (field) on M and is called the pullback covector through ϕ.

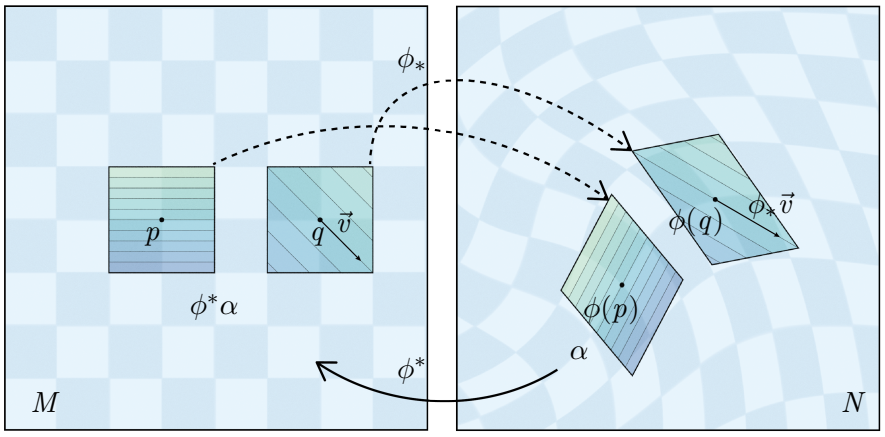


Figure 1.4 Given a map $\phi\colon M \to N$, a point $p \in M$ is mapped to $\phi(p) \in N$, a vector $\vec{v} \in T_qM$ is sent to $\phi_*\vec{v} \in T_{\phi(q)}N$ using the pushforward operator, and a covector $\alpha \in T^*_{\phi(p)}N$ is sent to a covector $\phi^*\alpha \in T^*_pM$ using the pullback operator.

By definition of adjoint,

$$\langle \phi^*\alpha | \vec{v} \rangle = \langle \alpha | \phi_* \vec{v} \rangle \quad \text{for all } \alpha \in T^*_{\phi(p)}N \text{ and } \vec{v} \in T_pM. \tag{1.35}$$

See Figure 1.4.

When written in coordinates, the pullback operator is the transpose of the Jacobian matrix.

The word *pullback* can also be used for composition of function. Suppose $g\colon N \to \mathbb{R}$ is a function on N and suppose $\phi\colon M \to N$. Then $f\colon M \to \mathbb{R}$ defined by $f = g \circ \phi =: \phi^* g$ is called the **pullback** of g via the mapping ϕ.

In summary, the overloaded concept of pullback are given as follows. When it is acted on functions and on covector fields: for $\phi\colon M \to N$,

$$\underset{\text{func}}{\phi^*} : (N \to \mathbb{R}) \xrightarrow{\text{linear}} (M \to \mathbb{R}), \qquad \underset{\text{func}}{\phi^*}\, g = g \circ \phi \tag{1.36}$$

$$\underset{\text{covec}}{\phi^*} : \Gamma(T^*N) \xrightarrow{\text{linear}} \Gamma(T^*M), \qquad (\underset{\text{covec}}{\phi^*}\, \alpha)|_p = (d\phi|_p)^*(\alpha|_{\phi(p)}). \tag{1.37}$$

Now, the chain rule can be summarized as:

Theorem 1.12 — Pullback and differential commute. The pullbacks ϕ^* for functions g and covectors dg satisfy

$$d(\underset{\text{func}}{\phi^*}\, g) = \underset{\text{covec}}{\phi^*}\, (dg) \tag{1.38}$$

Theorem 1.13 Let $\phi\colon M \to N$ and $\psi\colon N \to L$. Then

$$(\psi \circ \phi)^* = \phi^* \circ \psi^*. \tag{1.39}$$

■ **Example 1.6 — Back-propagation.** Suppose we have the sensitivity $\phi_* = d\phi$, $\psi_* = d\psi$ of maps $\phi\colon M \to N$ and $\psi\colon N \to L$. Suppose $g\colon L \to \mathbb{R}$ is some "cost function." Through ϕ and ψ we can pull the function back and formulate a cost function f on the parameter space M defined by $f = g \circ \psi \circ \phi$.

How do we compute the differential df_p at a particular set of parameters $p \in M$?

In a forward-mode differentiation method, one would first say that df_p is, under a coordinate system $(x_1, \dots, x_m)$, given by $df = \frac{\partial f}{\partial x_1} dx_1 + \cdots + \frac{\partial f}{\partial x_m} dx_m$ where $\frac{\partial f}{\partial x_i} = df(\vec{a}_i)$ where $\vec{a}_i$ is the coordinate vector. To find each of these partial derivatives, we expand $df = (dg)(d\psi)(d\phi)$, and evaluate $df(\vec{a}_i) = (dg)(d\psi)(d\phi)(\vec{a}_i)$ for each i. These pushforward operations have to be performed for m numbers of time.

An alternative approach is the so-called **reversed-mode differentiation** (a.k.a. back-propagation)

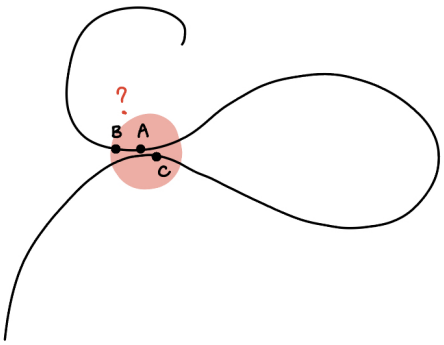
$$df_p = (\phi_p^*)(\psi_{\phi(p)}^*)(dg_{\psi(\phi(p))}). \tag{1.40}$$

In this case, we only need to pullback one covector (since g is scalar valued), rather than pushing forward m vectors like in the forward mode differentiation. ■

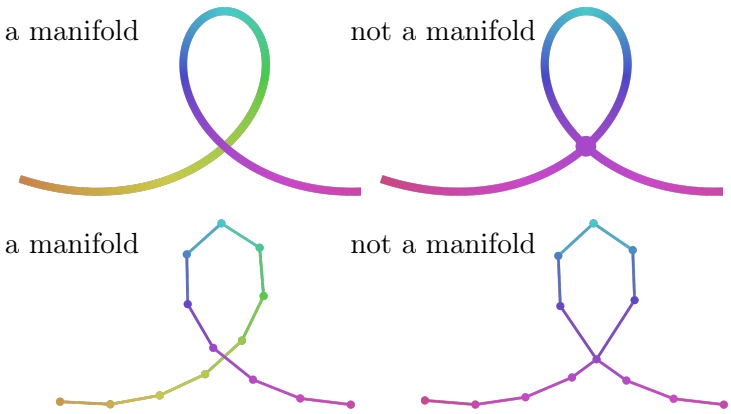
1.7 Manifolds

The domain M discussed in Section 1.5 and onward were assumed to be a subset of $\mathbb{R}^n$. In general, the above above discussion also works seamlessly when M is a **manifold**. What is a manifold? At high level it is a generalization of curves and surfaces, which are particularly interested in computer graphics. Its precise definition in mathematics is somewhat subtle.

Consider the following curve that almost touches itself somewhere. When per-



forming computations on this curve, we care about the relation between point A and point B a lot more than the relation between point A and point C, despite point C being closer to point A in the "ambient" space, which is the whiteboard where the curve was drawn on. This leads to us defining a customary "neighboring relation" over a geometry instead of using the absolute Euclidean distance. Such neighboring relation is formally referred to as *topology*. Mathematicians define *topological spaces* in the broadest sense, and later define *manifolds* as topological spaces that can afford differentiation and integration, similar to a Euclidean space.



In essence, a manifold is a space that can be locally parameterized by a Cartesian space. Through pullback and pushforward, calculus that is built upon differentials of functions automatically carries over from Cartesian space to manifolds. So, in the remainder of the course note, we can either think of a domain as a manifold or a subset of $\mathbb{R}^n$; the language is the same.

2. Exterior Calculus

In this chapter, we assume that the reader has a firm understanding of multivariable calculus, as well as the linear algebra that involves vectors and covectors (Chapter 1). However, you might wonder, why then, do we need to explore **exterior calculus** if multivariable calculus already provides a comprehensive mathematical framework? Historically, exterior calculus has emerged as a tool primarily because it provides powerful notations to elucidate the principles of differential and integral calculus on *manifolds*, abstract entities that can represent general domains and curved geometries. Under the framework, one carefully distinguishes coordinate-dependent multivariable quantities, and directly works with their geometrically meaningful, coordinate-free representations.

It turns out that exterior calculus is a powerful language not limited to the abstract manifold study. It proves to be useful in the broader application of multivariable calculus, even for problems defined on flat domains. Examples include solid mechanics, fluid dynamics, and optimization problems. The coordinate-free formalism exposes the underlying structures far better than piles of indices pointing to various coordinate components. One may write down equations that relate geometric measurements, say in a continuum, at a higher level that is closer to our physical and geometric intuition. Consequently, this approach results in a clearer, more straightforward pathway for mathematical reasoning for the phenomena and algebraic identities within these differential equations.

2.1 From Multivariable Calculus to Exterior Calculus

When it comes to multivariable integration in the introductory calculus, there are two important topics: change of variables (change of coordinates) formula, and integral theorems namely the Green, Gauss, Stokes theorems. One of the motivations of exterior calculus is to work in a framework that the change of variables are automatic,

and all the Stokes-like theorems are unified. Once we have such a framework, we have a more unified way for talking about, for example, derivatives of changes of variables (*e.g.* derivatives with respect to change of integration domain), which are crucial in applications such as continuum mechanics and differentiable rendering.

Here is a quickly recap of these topics in multivariable calculus. We follow one of the very first written textbooks on these topics, Maxwell's 1873 "a Treatise on Electricity and Magnetism, Vol. 1, Preliminary." At the dawn of vector calculus, Maxwell has made several remarks indicating that the natural structure to unify multivariable calculus is to use the skew-symmetric structure of differential forms. What exterior calculus is built upon is the linear algebra of these skew-symmetric forms.

2.1.1 Line and surface integral

An important operation is the integration of the component of a vector field projected along a line or a curve. This is called the **circulation** of a vector field along a curve (not necessarily closed).

Let $\boldsymbol{\gamma} = (x_1, x_2, x_3)\colon [0, S] \to \mathbb{R}^3$ be a curve, and $\mathbf{v} = (v_1, v_2, v_3)\colon \mathbb{R}^3 \to \mathbb{R}^3$ a vector field. Then the circulation of the curve is given by

$$\begin{aligned} C &= \int_0^S \langle \mathbf{v}, \boldsymbol{\gamma}' \rangle \, ds \quad \text{which can be expressed as} \\ &= \int_0^S v_1 \frac{dx_1}{ds} \, ds + v_2 \frac{dx_2}{ds} \, ds + v_3 \frac{dx_3}{ds} \, ds \quad \text{or simply} \\ &= \int_0^S \langle \mathbf{v}, d\boldsymbol{\gamma} \rangle = \int_{\gamma([0,S])} v_1 \, dx_1 + v_2 \, dx_2 + v_3 \, dx_3 \end{aligned}$$

Here $v_1 \, dx_1 + v_2 \, dx_2 + v_3 \, dx_3$ is called a **1-form**. This quantity C with integral $\int_{\boldsymbol{\gamma}([0,S])}$ generally depends on the entire path $\boldsymbol{\gamma}([0, S])$, as opposed to an integral $\int_{\mathbf{a}}^{\mathbf{b}}$ depending only on the two end points $\mathbf{a} = \boldsymbol{\gamma}(0), \mathbf{b} = \boldsymbol{\gamma}(S)$. However, when within a certain region the one-form takes the form of

$$v_1 \, dx_1 + v_2 \, dx_2 + v_3 \, dx_3 = d\Psi \quad \text{for some potential function } \Psi$$

that is, is an exact differential within that region, the value of C becomes

$$C = \Psi(\mathbf{b}) - \Psi(\mathbf{a})$$

and is the same for any two paths between $\mathbf{a}$ and $\mathbf{b}$, provided the path can be changed into the other by continuous motion without passing out of this region.[1]

[1]The necessary condition for the 1-form $v_1 \, dx_1 + v_2 \, dx_2 + v_3 \, dx_3$ to be an exact differential of a potential is

$$\operatorname{curl} \mathbf{v} = \left(\frac{\partial v_3}{\partial x_2} - \frac{\partial v_2}{\partial x_3}, \frac{\partial v_1}{\partial x_3} - \frac{\partial v_3}{\partial x_1}, \frac{\partial v_2}{\partial x_1} - \frac{\partial v_1}{\partial x_2} \right) = (0, 0, 0).$$

However, there are cases in which the condition $\operatorname{curl} \mathbf{v} = 0$ is fulfilled throughout a domain, and yet the line integral C from $\mathbf{a}$ to $\mathbf{b}$ may be different for two curves. This may be the case if the domain is multiply connected, *i.e.* there exists closed loop that cannot deform and shrink to a

Another important operation that involves a surface integral is the **flux** of a vector field through a surface. Let Σ be a surface, dA the surface area element,[2] and $\mathbf{n}$ the unit normal to the surface drawn towards the positive side of the surface. The flux of $\mathbf{v}$ over the surface is given and expressed by

$$\begin{aligned}\iint_\Sigma \langle \mathbf{v}, \mathbf{n}\rangle \, dA &= \iint_\Sigma v_1 n_1 \, dA + \iint_\Sigma v_2 n_2 \, dA + \iint_\Sigma v_3 n_3 \, dA \\ &= \iint_\Sigma v_1 \, dx_2 \, dx_3 + \iint_\Sigma v_2 \, dx_3 \, dx_1 + \iint_\Sigma v_3 \, dx_1 \, dx_2.\end{aligned}$$

The quantity $v_1 \, dx_2 \, dx_3 + v_2 \, dx_3 \, dx_1 + v_3 \, dx_1 \, dx_2$ is called a **2-form**.

Here $dx_1 \, dx_2$ is the area element projected to the x_1x_2-plane. The flux of a vector field through a surface is the sum of its contribution from all three projected components, each of which is the area integral of the vector component normal to the respective plane. It is important to note that these area integrals are signed integral inherited from the orientation of the space. The normal of the $dx_2 \, dx_3$-plane is x_1, the normal of the $dx_3 \, dx_1$-plane is x_2, *etc.* In particular, when writing flux integral in this form, the ordering of differentials matter $\iint v_1 \, dx_2 \, dx_3 = -\iint v_1 \, dx_3 \, dx_2$.

Similarly, in this context, a volumetric integral $\iiint h \, dx_1 \, dx_2 \, dx_3$ should have the ordering of the differential $dx_1 \, dx_2 \, dx_3$ positively oriented.

More generally, it turns out that the right structure for integration that evaluate circulations, fluxes, *etc.*, is infinitesimally a **skew symmetric** product of differentials $dx_{i_1} \cdots dx_{i_k}$. Later, in order not to be confused with the unsigned integrals, we shall use the notation $dx_{i_1} \wedge \cdots \wedge dx_{i_k}$ to emphasize the skew symmetry nature. We will spend 2.3 to study the linear algebra of these forms.

2.1.2 Changing variables for multivariable integrals

Suppose D is a domain of in the Cartesian space $\mathbb{R}^n$ with coordinates labeled $x_1, \ldots, x_n$. Let $F\colon D \to \mathbb{R}^n$ be a map into another Cartesian space with coordinates labeled $y_1, \ldots, y_n$. This map parametrizes the image $F(D)$ (as a signed multiset) with $y_i = F_i(x_1, \ldots, x_n)$ for $i = 1, \ldots, n$. Let $h\colon F(D) \to \mathbb{R}$ be a scalar function. Then the integral

$$\int_{F(D)} h(y_1, \ldots, y_n) \, dy_1 \cdots dy_n$$

can be written in terms of an integral over $(x_1, \ldots, x_n) \in D$ as

$$\int_{F(D)} h(y_1, \ldots, y_n) \, dy_1 \cdots dy_n = \int_D (h \circ F)(x_1, \ldots, x_n) \, J \, dx_1 \cdots dx_n,$$

point (nontrivial **cycles**), and if the two paths from $\mathbf{a}$ and $\mathbf{b}$ follow the opposite segments of such nontrivial loops. In this case, the one path cannot be transformed into the other. Two cycles are said to be equivalent if they together form the boundary of a surface embedded in the space. In a multiply connected domain, every cycle (up to equivalence) is a multiple, or a linear combination, of finitely many basis cycles.

If $\operatorname{curl} \mathbf{v} = 0$ and the circulations along those basis cycles vanish, then $v_1 \, dx + v_2 \, dx_2 + v_3 \, dx_3 = d\Psi$.

[2] Here dA does not mean the derivative d of a function A.

where J is the Jacobian determinant:

$$J = \det \begin{bmatrix} \partial F_1/\partial x_1 & \cdots & \partial F_1/\partial x_n \\ \vdots & \ddots & \vdots \\ \partial F_n/\partial x_1 & \cdots & \partial F_n/\partial x_n \end{bmatrix} = \det \Big(dF(e_1), \ldots, dF(e_n) \Big).$$

Here $e_1, \ldots, e_n$ is the canonical basis for the Cartesian space.

In exterior calculus, we will simply write

$$\int_{F(D)} \omega = \int_D F^* \omega,$$

i.e., the integral of ω over the image $F(D)$ of a map F is the same as the integral of the **pullback** differential $F^*\omega$ via the map over the domain D. The notion of pullback in exterior calculus will be the abstraction of the routine procedure of changing variables for the integrand.

2.1.3 Green, Gauss and Kelvin–Stokes theorems

Another important topic of multivariable calculus includes theorems about integrals of derivatives, which are higher dimensional versions of the fundamental theorem of calculus.

- The surface integral of the flux through a closed surface may be expressed as an integral over the enclosed volume.
- The circulation taken around a closed curve may be expressed in terms of a surface integral taken over a surface subtended by the curve.

Let $\mathbf{v} \colon \mathbb{R}^2 \to \mathbb{R}^2$ be a vector field in the 2-dimensional Cartesian space. Let $D \subset \mathbb{R}^2$ be a region, and let ∂D be its boundary curve. Then by integrating the derivatives in each variable and noting the integration orientation to assign an appropriate sign, we obtain **Green's Theorem**

$$\iint_D \left(\frac{\partial v_1}{\partial x_1} + \frac{\partial v_2}{\partial x_2} \right) dx_1 \, dx_2 = \oint_{\partial D} v_1 \, dx_2 - v_2 \, dx_1.$$

In general, in n-dimension, $\sum_{i=1}^n \partial v_i/\partial x_i$ is called the **divergence** of $\mathbf{v}$. Its integral over an n-dimensional region equals to the total (outward) flux through the boundary surface. Take 3D for example. Suppose $\mathbf{u} \colon \mathbb{R}^3 \to \mathbb{R}^3$ is a vector field in 3D (with coordinates labeled with y), $D \subset \mathbb{R}^3$ is a 3D domain, and ∂D is its boundary surface with normal vectors denoted by $\mathbf{n} \colon \partial D \to \mathbb{R}^3$. Then we have **Gauss' Divergence Theorem**

$$\begin{aligned} \iiint_D \sum_{i=1}^3 \frac{\partial u_i}{\partial y_i} \, dy_1 \, dy_2 \, dy_3 &= \oiint_{\partial D} u_1 \, dy_2 \, dy_3 + u_2 \, dy_3 \, dy_1 + u_3 \, dy_1 \, dy_2 \\ &= \oiint_{\partial D} \langle \mathbf{v}, \mathbf{n} \rangle \, dA. \end{aligned}$$

Let us come back to Green's Theorem. Suppose : $\mathbf{D} \subset \mathbb{R}^2 \to \mathbb{R}^3$ is a parametrized surface, whose restriction to the boundary $\mathbf{f}\colon \partial D \to \mathbb{R}^3$ is a parametrized space curve. Let $\vec{u}\colon \mathbb{R}^3 \to \mathbb{R}^3$ be a vector field in 3D. Then the **circulation** of $\mathbf{u}$ along $\mathbf{f}(\partial D)$ is

$$\begin{aligned}
\oint_{\mathbf{f}(\partial D)} \langle \mathbf{u}, d\mathbf{f} \rangle &= \oint_{\mathbf{f}(\partial D)} \sum_{i=1}^{3} u_i \, dy_i = \oint_{\partial D} \sum_{i=1}^{3} \sum_{j=1}^{2} u_i \frac{\partial f_i}{\partial x_j} \, dx_j \\
&\overset{\text{Green's Thm.}}{=} \sum_{i=1}^{3} \iint_D \left(\frac{\partial}{\partial x_2} \left(u_i \frac{\partial f_i}{\partial x_1} \right) - \frac{\partial}{\partial x_1} \left(u_i \frac{\partial f_i}{\partial x_2} \right) \right) dx_1 \, dx_2 \\
&= \sum_{i=1}^{3} \sum_{j=1}^{3} \iint_D \left(\frac{\partial u_i}{\partial y_j} \frac{\partial f_i}{\partial x_1} \frac{\partial f_j}{\partial x_2} - \frac{\partial u_i}{\partial y_j} \frac{\partial f_i}{\partial x_2} \frac{\partial f_j}{\partial x_1} \right) dx_1 \, dx_2 \\
&= \sum_{i=1}^{3} \sum_{j=1}^{3} \iint_D \left(\frac{\partial u_i}{\partial y_j} - \frac{\partial u_j}{\partial y_i} \right) \frac{\partial f_i}{\partial x_1} \frac{\partial f_j}{\partial x_2} \, dx_1 \, dx_2 \\
&= \iint_D \det \left(\operatorname{curl} \mathbf{u}, d\mathbf{f}(\mathbf{e}_1), d\mathbf{f}(\mathbf{e}_2) \right) \, dx_1 \, dx_2 = \iint_{\mathbf{f}(D)} \langle \operatorname{curl} \mathbf{u}, \mathbf{n} \rangle \, dA,
\end{aligned}$$

which is the total flux of $\operatorname{curl} \mathbf{u} = (\partial u_2/\partial y_3 - \partial u_3/\partial y_2, \partial u_3/\partial y_1 - \partial u_1/\partial y_3, \partial u_1/\partial y_2 - \partial u_2/\partial y_1)$ through the surface $\mathbf{f}(D)$. This is the **Kelvin–Stokes Theorem**.

In exterior calculus, vector calculus operators such as grad (∇), curl ($\nabla\times$), div ($\nabla\cdot$) all coalesce into a single derivative d. Similarly, important theorems such as the Fundamental Theorem of Calculus, or Gauss' and Green's Theorem all become instances of one single general statement, the so-called Stokes' theorem

$$\int_M d\omega = \int_{\partial M} \omega,$$

i.e., the integral of the differential of ω over the domain M is the same of ω over the boundary of the domain.

2.2 An Overview of the Language

The main objects that exterior calculus deals with are **differential forms**. Functions, or scalar fields, take a given point of the domain, and return a value. Extending this idea, differential forms are to be evaluated (integrated) over given curves, surfaces, or volumes, *etc.*

Specifically, let M be a 3-dimensional domain. On M, a k-form is a differential form that is to be integrated over a k-dimensional (oriented) submanifold:

- A 3-form is to be integrated over a volumetric region. For example, the mass density ρ of physical matter should be regarded as a 3-form, denoted by $\rho \in \Omega^3(M)$. Given a region $U \subset M$, the total mass within U is denoted by $\int_U \rho$.
- A 2-form is to be integrated over an oriented surface. For example, a flux ω of a flow should be regarded as a 2-form, denoted by $\omega \in \Omega^2(M)$. It describes each total flux over a given surface Σ, denoted by $\int_\Sigma \omega$.

- A 1-form is to be line integrated over a path. For example, a force field η should be regarded as a 1-form, denoted by $\eta \in \Omega^1(M)$. The total work done by the force along a given path Γ is denoted by $\int_\Gamma \eta$.
- A 0-form is a function, or a scalar field. For example, the temperature u of a material is a 0-form, denoted by $u \in \Omega^0(M)$. It is to be evaluated at each point.

We will encounter only 5 algebraic and differential operators for differential forms:

- **Pullback operator.** If we have a differential form $\alpha \in \Omega^k(W)$, and a map $f\colon M \to W$, then we have a pullback differential form $f^*\alpha \in \Omega^k(M)$. It has a natural definition that its evaluation over a k-dimensional surface $\Sigma \subset M$ is given by $\int_\Sigma (f^*\alpha) = \int_{f(\Sigma)} \alpha$.
- **Exterior derivative.** Denoted by d, the exterior derivative is a differential operator that sends a k-form α to a $(k+1)$-form $d\alpha$. The evaluation of $d\alpha$ over a $(k+1)$-dimensional surface Σ is designed so that $\int_\Sigma d\alpha = \int_{\partial\Sigma} \alpha$.
- **Wedge product.** It is an algebraic operator that produces higher-degree forms. If $\alpha \in \Omega^k(M)$ and $\beta \in \Omega^\ell(M)$ then their wedge product $\alpha \wedge \beta \in \Omega^{k+\ell}(M)$.
- **Interior product.** Provided a given direction, the interior product as an algebraic operator lowers the degree of a form. Suppose X is a tangent vector of M, and α is a k-form, then the interior product $i_X\alpha$ is a $(k-1)$-form.
- **Hodge star.** Denoted by $\star$, it is an algebraic operator that turns a k-form α into an $(n-k)$-form $\star\alpha$, where $n = \dim(M)$. Among these 5 operators, this Hodge star operator is the only one that requires a Riemannian metric for M.

Macroscopic and Microscopic View

A differential 1-form η is a quantity that is waiting to be integrated over a curve γ. This is the "macroscopic" viewpoint for differential forms. Let us take a short parametrized curve γ, which is so short that γ is well-described by a point $x = \gamma(0)$, a tangent vector $\mathbf{v} = \gamma'(0)$ at x as the velocity for the curve, and a small ε as the length of the interval of the parameter for the curve. Abusing the notation slightly, we write $\gamma = \varepsilon\mathbf{v}$. Then with continuity in η, we expect $\int_{\varepsilon\mathbf{v}} \eta \sim O(\varepsilon)$. The limit

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_{\varepsilon\mathbf{v}} \eta =: \eta(\mathbf{v})$$

is a quantity depending linearly on the velocity vector $\mathbf{v}$. We say $\eta(\mathbf{v})$ is the evaluation of the 1-form on a tangent vector $\mathbf{v}$. This is the algebraic, or the "microscopic," viewpoint of differential forms.

Conversely, given a 1-form η as a scalar-valued linear function on tangent vectors, we can understand its "macroscopic" counterpart as follows. The integral of η over a curve γ can be obtained by first partitioning the parameter interval of γ and then taking the sum of microscopic evaluation

$$\int_\gamma \eta \coloneqq \lim_{\substack{\text{partion}\\ \text{refines}}} \sum_i \eta\big(\gamma'(t_i)\big)(t_{i+1} - t_i).$$

Similarly, a differential 2-form ω in a "macroscopic" view is a quantity to be evaluated over an oriented surface. Its "microscopic" version is a function linearly evaluating on an infinitesimal oriented parallelogram spanned by two vectors v_1, v_2. More precisely, $\omega(\mathbf{v}_1, \mathbf{v}_2)$ is bilinear (linear in each component if fixing the other) and skew symmetric $\omega(\mathbf{v}_1, \mathbf{v}_2) = -\omega(\mathbf{v}_2, \mathbf{v}_1)$. To assemble a microscopic 2-form into its integral over a surface, partition the surface into infinitesimal parallelograms and sum the values of the bilinear forms.

As mentioned in Section 2.1, it turns out that the right structure for integration that evaluate circulations, fluxes, *etc.*, is infinitesimally a **skew symmetric multilinear form**. In Section 2.3 we study the linear algebra of these forms.

Starting from Section 2.4 we will expand on related concepts about vector fields. This will lead up to an important operation called *Lie derivative* in Section 2.6. This operation is essential in many applications such as continuum mechanics that is often overlooked in multivariable calculus. Finally, we introduce Hodge star in Section 2.7.

2.3 Differential Forms

Algebraically, differential forms are skew symmetric multilinear form. With such an object defined over a domain, it becomes a mathematical object that can be integrated over a submanifold (Section 2.3.1). In Section 2.3.2, we introduce how differential forms are made from covectors by the *theory of extension*. Using the same extension technique, we extend the notion of "the differential of a function" to general differential forms (Section 2.3.3). Finally, in Section 2.3.4, we give a geometric interpretation of differential forms and the operations among them.

Definition 2.1 Let M be a manifold. A **k-form** ω is a skew symmetric k-linear form field; that is, for each $p \in M$, we have a function linear in each of its k arguments

$$\omega_p[\![\cdot, \dots, \cdot]\!] \colon \underbrace{T_pM \times \cdots \times T_pM}_{k} \xrightarrow{\text{multilinear}} \mathbb{R} \tag{2.1}$$

satisfying

$$\omega_p[\![X_1, \dots, X_i, \dots, X_j, \dots, X_k]\!] = -\omega_p[\![X_1, \dots, X_j, \dots, X_i, \dots, X_k]\!]. \tag{2.2}$$

Such a skew-symmetric k-linear form at p is collectively

$$\omega_p \in \bigwedge^k T_p^*M. \tag{2.3}$$

The space of k-forms is denoted by $\Omega^k(M) = \Gamma(\bigwedge^k T^*M)$.[a]

[a] The motivation for the notation "$\wedge^k T^*M$" will become clear in Section 2.3.2.

Space $\Omega^0(M)$ of 0-forms are just space of functions $\Omega^0(M) = \{f \colon M \to \mathbb{R}\}$. Space $\Omega^1(M)$ of 1-forms are the space of covector fields.

For each p in an n-dimensional manifold M, the dimension of the vector space $\bigwedge^k T_p^*M$ is $\binom{n}{k}$. Every k-form would be 0 if $k > n$.

■ **Example 2.1** In the 3D Cartesian space, each vector written in coordinates $\mathbf{a} = (a_1, a_2, a_3)^\intercal$ can be converted into a 1-form $(\mathbf{a})_{1\text{-form}}$ or a 2-form $(\mathbf{a})_{2\text{-form}}$, whose evaluation on a vector $\mathbf{u}$ or a pair of vectors $\mathbf{u}, \mathbf{v}$ are defined by

$$(\mathbf{a})_{1\text{-form}}[\![\mathbf{u}]\!] = a_1 u_1 + a_2 u_2 + a_3 u_3 \tag{2.4}$$

$$(\mathbf{a})_{2\text{-form}}[\![\mathbf{u}, \mathbf{v}]\!] = \det \begin{bmatrix} a_1 & u_1 & v_1 \\ a_2 & u_2 & v_2 \\ a_3 & u_3 & v_3 \end{bmatrix}. \tag{2.5}$$

A scalar function a, which is a 0-form $a = (a)_{0\text{-form}}$, can be converted into a 3-form $(a)_{3\text{-form}}$

$$(a)_{3\text{-form}}[\![\mathbf{u}, \mathbf{v}, \mathbf{w}]\!] = a \det \begin{bmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{bmatrix}. \tag{2.6}$$

■

2.3.1 Integration and pullback

Differential forms can be treated as objects to-be-integrated along a k-dimensional surface. Suppose $S\colon \mathbb{D}^k \to M$ is a parametric k-dimensional surface defined over a k-dimensional parameter space $\mathbb{D}^k$ with values in M. Recall that $dS = S_*$ is its pushforward map. Let $\theta_1, \dots, \theta_k$, $\theta_i\colon \mathbb{D}^k \to \mathbb{R}$, be an arbitrary coordinate system. That is, we have $d\theta_1, \dots, d\theta_k$ as basis covector field, whose dual basis $\vec{e}_1, \dots, \vec{e}_k \in \Gamma(T\mathbb{D}^k)$ is a coordinate vector field.

Definition 2.2 Let $\omega \in \Omega^k(M)$ be a k-form on M. The integral $\int_S \omega$ of ω over a surface S described above is defined by

$$\int_S \omega := \left(\iint \cdots \int\right)_{\mathbb{D}^k} \omega_{S(\theta_1,\dots,\theta_k)}[\![S_*\vec{e}_1, \dots, S_*\vec{e}_k]\!]\, d\theta_1 \cdots d\theta_k. \tag{2.7}$$

This definition is independent of reparametrization of S.

■ **Example 2.2** In the 3D Cartesian space, for a scalar function a and a vector field $\mathbf{a}$ we have

$$\int_p a = a(p), \tag{2.8}$$

$$\int_C (\mathbf{a})_{1\text{-form}} = \int_C \mathbf{a} \cdot d\mathbf{l}, \tag{2.9}$$

$$\int_S (\mathbf{a})_{2\text{-form}} = \iint_S \mathbf{a} \cdot \mathbf{n}\, dA, \tag{2.10}$$

$$\int_V (a)_{3\text{-form}} = \iiint_V a\, dV. \tag{2.11}$$

■

Definition 2.3 For each smooth map $\phi\colon M \to N$, we define the **pullback operator**

$$\phi^*\colon \Omega^k(N) \xrightarrow{\text{linear}} \Omega^k(M) \tag{2.12}$$

that for each $\omega \in \Omega^k(N)$,

$$(\phi^*\omega)_p[\![X_1, \dots, X_k]\!] \coloneqq \omega_{\phi(p)}[\![\phi_* X_1, \dots, \phi_* X_k]\!]. \tag{2.13}$$

Theorem 2.1 — Pullback as change of integration variables.

$$\int_{\phi(S)} \omega = \int_S \phi^*\omega. \tag{2.14}$$

■ **Example 2.3** Let $\boldsymbol{\phi}\colon \mathbb{R}^3 \to \mathbb{R}^3$ be a smooth map between the 3D Cartesian space. Let $\mathbf{F} = \nabla\boldsymbol{\phi}$, denoted as a matrix $F_{ij} = \frac{\partial \boldsymbol{\phi}_i}{\partial x_j}$ called the **deformation gradient**. Let $J = \det(\mathbf{F})$. Then

$$\underset{\text{0-form}}{\boldsymbol{\phi}^*}\ (a)_{\text{0-form}} = (a \circ \boldsymbol{\phi})_{\text{0-form}}, \tag{2.15}$$

$$\underset{\text{1-form}}{\boldsymbol{\phi}^*}\ (\mathbf{a})_{\text{1-form}} = (\mathbf{F}^\intercal \mathbf{a} \circ \boldsymbol{\phi})_{\text{1-form}}, \tag{2.16}$$

$$\underset{\text{2-form}}{\boldsymbol{\phi}^*}\ (\mathbf{a})_{\text{2-form}} = (J\mathbf{F}^{-1}\mathbf{a} \circ \boldsymbol{\phi})_{\text{2-form}}, \tag{2.17}$$

$$\underset{\text{3-form}}{\boldsymbol{\phi}^*}\ (a)_{\text{3-form}} = (Ja \circ \boldsymbol{\phi})_{\text{3-form}}. \tag{2.18}$$

■

2.3.2 Wedge product and interior product

In Definition 2.1, we described what a differential form is. At each point, a differential form is a skew symmetric k-linear form. *If we have such an object*, we can talk about its integration and the associated changes of variables (pullback) (Section 2.3.1). But *how do we obtain these skew symmetric k-linear form in the first place*? Some instances of differential forms are concretely given, but only under a 3D Cartesian coordinate system (Example 2.1). Is there a way to systematically construct skew symmetric k-linear forms?

Well, we do have a starting point. *Covectors are 1-forms.* They are canonically defined by taking the dual space of a vector space. By the **theory of extension** introduced by Grassmann (1862), we can extend the elementary covectors into k forms through a new multiplication called **wedge product**. We can also extend the elementary vector–covector dual pairing to pairing k-forms and k vectors, making k-forms truly objects satisfying Definition 2.1. The extension of dual pairing to k-forms is called the **interior product**.

In fact, using the same technique of extension, we can extend operators other than the vector–covector dual pairing. For example, at the level between functions and covector fields, we have the notion of taking differentials (the differential of a function is a covector field). We can extend such a differentiation notion, and obtain

exterior derivative (Section 2.3.3).

The following Definitions 2.4, 2.5 and 2.6 collectively serve as an alternative axiomatic definition for differential forms that can replace Definition 2.1.

Definition 2.4 — Wedge product. Let V be a vector space, and let V^* be its dual space. Define

$$\bigwedge^\bullet V^* = \mathbb{R} \oplus V^* \oplus \left(\bigwedge^2 V^*\right) \oplus \left(\bigwedge^3 V^*\right) \oplus \cdots \tag{2.19}$$

as the free algebra[a] generated by a new multiplication called **wedge product**

$$\wedge\colon \bigwedge^k V^* \times \bigwedge^\ell V^* \xrightarrow{\text{bilinear}} \bigwedge^{k+\ell} V^* \tag{2.20}$$

that satisfies the following two rules:

- Associativity: $(\alpha \wedge \beta) \wedge \gamma = \alpha \wedge (\beta \wedge \gamma)$; (2.21)
- Anti-commutativity: $\alpha \wedge \alpha = 0$ for 1-forms $\alpha \in V^*$. (2.22)

Here, $\bigwedge^k V^*$ in (2.19) is the collection of all linear combinations of degree-k wedge products of covectors.

[a] The space of all polynomial expressions but without commutativity on multiplication.

Definition 2.5 — Interior product. For each vector $X \in V$ in a vector space V, define a linear map

$$i_X\colon \wedge^k V^* \xrightarrow{\text{linear}} \wedge^{k-1} V^* \tag{2.23}$$

that satisfies

- Base definition: $i_X \alpha = \langle \alpha | X \rangle$ for all 1-forms $\alpha \in V^*$; (2.24)
- Nilpotent: $i_X i_X = 0$; (2.25)
- Graded Leibniz rule: $i_X(\alpha \wedge \beta) = (i_X \alpha) \wedge \beta + (-1)^{\deg(\alpha)} \alpha \wedge (i_X \beta)$. (2.26)

Definition 2.6 — Evaluation of k-forms. For $\omega \in \wedge^k V^*$ and $X_1, \ldots, X_k \in V$, define

$$\omega[\![X_1, \ldots, X_k]\!] \coloneqq i_{X_k} \cdots i_{X_2} i_{X_1} \omega. \tag{2.27}$$

In particular,

$$(i_X \omega)[\![Y_1, \ldots, Y_{k-1}]\!] = \omega[\![X, Y_1, \ldots, Y_{k-1}]\!]. \tag{2.28}$$

The wedge product and interior product generalize to the pointwise wedge product and interior product for the space of k-form fields $\Omega^k(M) \coloneqq \Gamma(\wedge^k T^*M)$

$$\wedge\colon \Omega^k(M) \times \Omega^\ell(M) \xrightarrow{\text{bilinear}} \Omega^{k+\ell}(M), \quad (\alpha \wedge \beta)_p \coloneqq \alpha_p \wedge \beta_p \tag{2.29}$$

$$i_X\colon \Omega^k(M) \xrightarrow{\text{linear}} \Omega^{k-1}(M), \quad (i_X \omega)_p \coloneqq i_{X_p} \omega_p, \tag{2.30}$$

where $X \in \Gamma(TM)$ is a vector field.

Theorem 2.2 — Graded anti-commutativity.

$$\alpha \wedge \beta = (-1)^{k\ell} \beta \wedge \alpha \tag{2.31}$$

when α is a k-form and β is an ℓ-form.

Proof. First, we have

$$\alpha \wedge \beta = -\beta \wedge \alpha \quad \text{when } k = \ell = 1 \tag{2.32}$$

as a consequence of (2.22):

$$0 = (\alpha + \beta) \wedge (\alpha + \beta) = \underbrace{\alpha \wedge \alpha}_{0} + \alpha \wedge \beta + \beta \wedge \alpha + \underbrace{\beta \wedge \beta}_{0}. \tag{2.33}$$

For general k-forms and ℓ-forms, write α, β as linear combinations of $\wedge$ of 1-forms, and reorder $\alpha \wedge \beta$ into $\beta \wedge \alpha$ by applying the swaps (2.32) $k\ell$ number of times. □

Theorem 2.3 For $\alpha_1, \dots, \alpha_k \in \Omega^1(M)$,

$$(\alpha_1 \wedge \cdots \wedge \alpha_k)[\![X_1, \dots, X_k]\!] = \det \begin{bmatrix} \langle \alpha_1 | X_1 \rangle & \cdots & \langle \alpha_1 | X_k \rangle \\ \vdots & \ddots & \vdots \\ \langle \alpha_k | X_1 \rangle & \cdots & \langle \alpha_k | X_k \rangle \end{bmatrix}. \tag{2.34}$$

Proof. Using the definition (2.27) and repeatedly apply the Leibniz rule (2.26) and (2.24), we obtain an expression $\sum_{\sigma \in S_k} (-1)^{|\sigma|} \prod_{i=1}^{k} \langle \alpha_i | X_{\sigma(i)} \rangle$, which is Leibniz's formula for the determinant of the matrix $\langle \alpha i | X_j \rangle$. □

From Theorem 2.3, it is apparent that the k-forms generated by k forms are precisely the ones described in Definition 2.1.

■ **Example 2.4** In the 3D Cartesian space $\mathbb{R}^3$ with covector basis $dx, dy, dz \in \Omega^1(\mathbb{R}^3)$, the conversion (Example 2.1) from Cartesian vectors/scalars to forms are explicitly given by

$$(a)_{0\text{-form}} = a, \tag{2.35}$$
$$(\mathbf{a})_{1\text{-form}} = a_1\, dx + a_2\, dy + a_3\, dz, \tag{2.36}$$
$$(\mathbf{a})_{2\text{-form}} = a_1\, dy \wedge dz + a_2\, dz \wedge dx + a_3\, dx \wedge dy \tag{2.37}$$
$$= i_{(\mathbf{a})_{\text{vec}}}(dx \wedge dy \wedge dz), \tag{2.38}$$
$$(a)_{3\text{-form}} = a\, dx \wedge dy \wedge dz. \tag{2.39}$$

One can check that this is consistent with Example 2.1 using Theorem 2.3. ■

■ **Example 2.5** In the 3D Cartesian space,

$$(a)_{0\text{-form}} \wedge (b)_{0\text{-form}} = (ab)_{0\text{-form}}, \tag{2.40}$$
$$(a)_{0\text{-form}} \wedge (\mathbf{b})_{1\text{-form}} = (a\mathbf{b})_{1\text{-form}}, \tag{2.41}$$
$$(a)_{0\text{-form}} \wedge (\mathbf{b})_{2\text{-form}} = (a\mathbf{b})_{2\text{-form}}, \tag{2.42}$$
$$(a)_{0\text{-form}} \wedge (b)_{3\text{-form}} = (ab)_{3\text{-form}}, \tag{2.43}$$
$$(\mathbf{a})_{1\text{-form}} \wedge (\mathbf{b})_{1\text{-form}} = (\mathbf{a} \times \mathbf{b})_{2\text{-form}}, \tag{2.44}$$
$$(\mathbf{a})_{1\text{-form}} \wedge (\mathbf{b})_{2\text{-form}} = (\mathbf{a} \cdot \mathbf{b})_{3\text{-form}}. \tag{2.45}$$

■

■ **Example 2.6** In the 3D Cartesian space,

$$i_{\mathbf{v}}(\mathbf{a})_{1\text{-form}} = (\mathbf{a} \cdot \mathbf{v})_{0\text{-form}}, \tag{2.46}$$
$$i_{\mathbf{v}}(\mathbf{a})_{2\text{-form}} = (\mathbf{a} \times \mathbf{v})_{1\text{-form}}, \tag{2.47}$$
$$i_{\mathbf{v}}(a)_{3\text{-form}} = (a\mathbf{v})_{2\text{-form}}. \tag{2.48}$$

■

■ **Example 2.7 — Vector algebra identities.** The following 3D vector identities are special cases of the Leibniz rule (2.26) for the interior product written out in 3D vector algebra.

Theorem 2.4 — Vector triple product. $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}$.

Proof. $(\mathbf{a} \times (\mathbf{b} \times \mathbf{c}))_{1\text{-form}} \overset{(2.47)}{=} -i_{\mathbf{a}}(\mathbf{b} \times \mathbf{c})_{2\text{-form}} \overset{(2.44)}{=} -i_{\mathbf{a}}((\mathbf{b})_{1\text{-form}} \wedge (\mathbf{c})_{1\text{-form}}) \overset{(2.26)}{=} -i_{\mathbf{a}}(\mathbf{b})_{1\text{-form}}(\mathbf{c})_{1\text{-form}} + (\mathbf{b})_{1\text{-form}} i_{\mathbf{a}}(\mathbf{c})_{1\text{-form}} \overset{(2.46)}{=} (-(\mathbf{a} \cdot \mathbf{b})\mathbf{c} + (\mathbf{a} \cdot \mathbf{c})\mathbf{b})_{1\text{-form}}$. □

Theorem 2.5 — Cauchy–Binet identity. $(\mathbf{a}\times\mathbf{b})\cdot(\mathbf{c}\times\mathbf{d}) = (\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d}) - (\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{c})$.

Proof. $[(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d})]_{3\text{-form}} \overset{(2.45)}{=} (\mathbf{a} \times \mathbf{b})_{1\text{-form}} \wedge (\mathbf{c} \times \mathbf{d})_{2\text{-form}} \overset{\substack{(2.47)\\(2.44)}}{=} -i_{\mathbf{a}}(\mathbf{b})_{2\text{-form}} \wedge (\mathbf{c})_{1\text{-form}} \wedge (\mathbf{d})_{1\text{-form}} \overset{(2.26)}{=} (\mathbf{b})_{2\text{-form}} \wedge i_{\mathbf{a}}(\mathbf{c})_{1\text{-form}} \wedge (\mathbf{d})_{1\text{-form}} - (\mathbf{b})_{2\text{-form}} \wedge (\mathbf{c})_{1\text{-form}} \wedge i_{\mathbf{a}}(\mathbf{d})_{1\text{-form}} \overset{\substack{(2.46)\\(2.45)}}{=} (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d})_{3\text{-form}} - (\mathbf{a} \cdot \mathbf{d})(\mathbf{b} \cdot \mathbf{c})_{3\text{-form}}$. □

Theorem 2.6 — Vector quadruple product. $(\mathbf{a} \times \mathbf{b}) \times (\mathbf{c} \times \mathbf{d}) = \det(\mathbf{a}, \mathbf{c}, \mathbf{d})\mathbf{b} - \det(\mathbf{b}, \mathbf{c}, \mathbf{d})\mathbf{a}$.

Proof. $[(\mathbf{a} \times \mathbf{b}) \times (\mathbf{c} \times \mathbf{d})]_{2\text{-form}} \overset{(2.44)}{=} (\mathbf{a} \times \mathbf{b})_{1\text{-form}} \wedge (\mathbf{c} \times \mathbf{d})_{1\text{-form}} \overset{(2.47)}{=} i_{\mathbf{a}}(\mathbf{b})_{2\text{-form}} \wedge i_{\mathbf{c}}(\mathbf{d})_{2\text{-form}} \overset{(2.26)}{=} i_{\mathbf{a}}[(\mathbf{b})_{2\text{-form}} \wedge i_{\mathbf{c}}(\mathbf{d})_{2\text{-form}}] - (\mathbf{b})_{2\text{-form}} i_{\mathbf{a}} i_{\mathbf{c}}(\mathbf{d})_{2\text{-form}} \overset{\substack{(2.47),(2.45)\\(2.6)}}{=} i_{\mathbf{a}}(\mathbf{b} \cdot (\mathbf{d} \times \mathbf{c}))_{3\text{-form}} - (\mathbf{b})_{2\text{-form}} \det(\mathbf{d}, \mathbf{c}, \mathbf{a}) \overset{(2.48)}{=} (\mathbf{a})_{2\text{-form}} \det(\mathbf{b}, \mathbf{d}, \mathbf{c}) - (\mathbf{b})_{2\text{-form}} \det(\mathbf{d}, \mathbf{c}, \mathbf{a})$. □

■

Interaction with pullbacks

The pullback operator Definition 2.3 preserves the wedge product and interior product. This is because Definition 2.3 is consistent to the pullback operator for covectors (Definition 1.17 and Definition 1.4), and the rest of the extension construction is canonical (Definition 2.4–2.6).

Theorem 2.7 — Pullback distributes over wedge products.

$$\phi^*(\alpha \wedge \beta) = (\phi^*\alpha) \wedge (\phi^*\beta). \tag{2.49}$$

Theorem 2.8 — Passing pullback through interior product.

$$\phi^*(i_{\phi_*(X)}\omega) = i_X(\phi^*\omega). \tag{2.50}$$

■ **Example 2.8 — Matrix vector identities.** We can write Theorem 2.7 and Theorem 2.8 in the 3D Cartesian space using the correspondence given by Example 2.3.

Theorem 2.9 Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$, $\mathbf{F} \in \mathbb{R}^{3\times 3}$, and $J = \det(\mathbf{F})$. Then

$$J\mathbf{F}^{-1}(\mathbf{u} \times \mathbf{v}) = (\mathbf{F}^\intercal \mathbf{u}) \times (\mathbf{F}^\intercal \mathbf{v}). \tag{2.51}$$
$$\mathbf{F}^\intercal(\mathbf{u} \times (\mathbf{F}\mathbf{v})) = (J\mathbf{F}^{-1}\mathbf{u}) \times \mathbf{v}. \tag{2.52}$$

Proof. (2.51) is (2.49) when both α and β are 1-forms. (2.52) is (2.50) when ω is a 2-form. □

■

Integral picture for interior product

Differential k-forms can be evaluated by inserting k vectors (Definition 2.6). Differential k-forms cam also be integrated over k-dimensional surface (Section 2.3.1). In the former algebraic picture, the interior product is the "insertion" operation (2.28). For the latter integral picture, it turns out that the interior product is dual to the "extrusion" operation, stated as follows.

Suppose $S\colon \mathbb{D}^k \to M$ is a $(k-1)$-dimensional surface. Let $X \in \Gamma(TM)$ be a vector field. Consider extruding the $(k-1)$-dimensional surface S into a k-dimensional surface $\operatorname{ext}^\epsilon_X S$ along X. Concretely, it is the solution to the initial value problem of flowing along X:

$$(\operatorname{ext}^\epsilon_X S)\colon [0,\epsilon] \times \mathbb{D}^k \to M, \tag{2.53}$$
$$\frac{\partial(\operatorname{ext}^\epsilon_X S)}{\partial t}(t,\theta_1,\dots,\theta_{k-1}) = X|_{(\operatorname{ext}^\epsilon_X S)(t,\theta_1,\dots,\theta_{k-1})} \tag{2.54}$$
$$(\operatorname{ext}^\epsilon_X S)(0,\theta_1,\dots,\theta_{k-1}) = S(\theta_1,\dots,\theta_{k-1}). \tag{2.55}$$

This geometric operation of extrusion is illustrated in Figure 2.1. The following theorem states that the interior product using a vector field X measures the rate of change of measurement when the integration domain is an extrusion along the vector field.

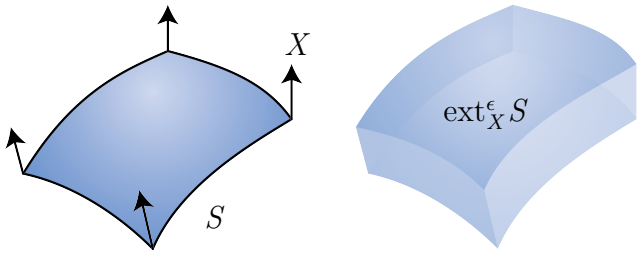


Figure 2.1 The extrusion of a $(k-1)$-dimensional surface S along a vector field X (left) yields a k-dimensional surface (right).

Theorem 2.10 — Extrusion and interior product.

$$\left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \int_{\text{ext}_X^\epsilon S} \omega = \int_S i_X \omega. \tag{2.56}$$

2.3.3 Exterior derivative

Taking the differential of a function produces a covector field (Definition 1.6). Using a similar extension rule as in Definition 2.5, we obtain:

Definition 2.7 — Exterior derivative. Define $d\colon \Omega^k(M) \xrightarrow{\text{linear}} \Omega^{k+1}(M)$ satisfying

- Base case: df is the differential of f for any 0-form f; (2.57)
- Nilpotent: $d \circ d = 0$; (2.58)
- Graded Leibniz rule: $d(\alpha \wedge \beta) = (d\alpha) \wedge \beta + (-1)^{\deg(\alpha)} \alpha \wedge (d\beta)$. (2.59)

■ **Example 2.9** In the 3D Cartesian space,

$$d(f)_{\text{0-form}} = (\nabla f)_{\text{1-form}}, \tag{2.60}$$
$$d(\mathbf{v})_{\text{1-form}} = (\nabla \times \mathbf{v})_{\text{2-form}}, \tag{2.61}$$
$$d(\mathbf{v})_{\text{2-form}} = (\nabla \cdot \mathbf{v})_{\text{3-form}}. \tag{2.62}$$

■

Next, we state two important theorems about the exterior derivative. The first is that exterior derivative and pullback commute. This is a consequence of the same theorem for the base case (Theorem 1.12), and that its extension to the general exterior derivatives (Definition 2.7) is defined using its interaction with $\wedge$, which is conserved under pullback (Theorem 2.7).

Theorem 2.11 — Exterior derivative and pullback commute.

$$\phi^*(d\alpha) = d(\phi^*\alpha). \tag{2.63}$$

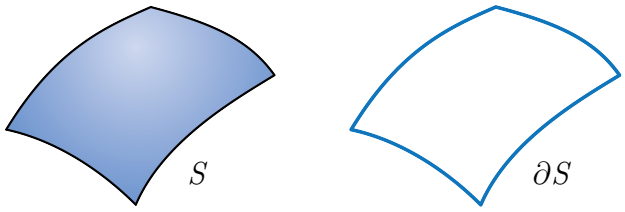


Figure 2.2 The boundary of a k-dimensional surface S (left) is a $(k-1)$-dimensional surface (right).

The second important theorem is the Stokes theorem. Similar to how the interior product is dual to extrusion (Theorem 2.10), the exterior derivative is dual to the boundary operation on the integration domain (Figure 2.2).

Theorem 2.12 — Stokes Theorem.

$$\int_S d\alpha = \int_{\partial S} \alpha. \tag{2.64}$$

Proof. Let the k-dimensional surface S be parameterized by $\phi\colon \Omega \subset \mathbb{R}^k \to M$. By Theorem 2.1 and Theorem 2.11, the Stokes theorem reduces to the divergence theorem on the k-dimensional region $\Omega \subset \mathbb{R}^k$. □

■ **Example 2.10 — Nilpotence (2.58) in 3D.**

$$(\nabla \times)(\nabla) = 0, \tag{2.65}$$
$$(\nabla \cdot)(\nabla \times) = \mathbf{0}. \tag{2.66}$$

■

■ **Example 2.11 — Leibniz rule (2.59) in 3D.** Let $f\colon \mathbb{R}^3 \to \mathbb{R}$, $\mathbf{u}, \mathbf{v}\colon \mathbb{R}^3 \to \mathbb{R}^3$ be function and vector fields. The Leibniz rule (2.59) becomes the following vector calculus identities. When it is applied to a 0-form and a 1-form, we get

Theorem 2.13 $\nabla \times (f\mathbf{u}) = (\nabla f) \times \mathbf{u} + f(\nabla \times \mathbf{u})$.

For 0-form and 2-form:

Theorem 2.14 $\nabla \cdot (f\mathbf{u}) = (\nabla f) \cdot \mathbf{u} + f(\nabla \cdot \mathbf{u})$.

For 1-form and 1-form:

Theorem 2.15 $\nabla \cdot (\mathbf{u} \times \mathbf{v}) = (\nabla \times \mathbf{u}) \cdot \mathbf{v} - \mathbf{u} \cdot (\nabla \times \mathbf{v})$.

■

■ **Example 2.12 — Theorem 2.11 in 3D.** In the 3D Cartesian space, Theorem 2.11 becomes the following vector identities. Let $\boldsymbol{\phi}\colon \mathbb{R}^3 \to \mathbb{R}^3$, $\mathbf{F} = \nabla\boldsymbol{\phi}$, and $J =$

$\det(\mathbf{F})$. Recall Example 2.3 for the pullback formulas in 3D. Theorem 2.11 applied to 0-forms yields

Theorem 2.16 — Chain rule revisited. Let $f\colon \mathbb{R}^3 \to \mathbb{R}$. Then,

$$\mathbf{F}^\intercal(\nabla f \circ \boldsymbol{\phi}) = \nabla(f \circ \boldsymbol{\phi}) \tag{2.67}$$

Theorem 2.11 applied to 1-forms yields

Theorem 2.17 — Change of variables for curl. Let $\mathbf{u}\colon \mathbb{R}^3 \to \mathbb{R}^3$. Then,

$$J\mathbf{F}^{-1}((\nabla \times \mathbf{u}) \circ \boldsymbol{\phi}) = \nabla \times (\mathbf{F}^\intercal \mathbf{u} \circ \boldsymbol{\phi}). \tag{2.68}$$

Theorem 2.11 applied to 2-forms yields

Theorem 2.18 — Piola identity. Let $\mathbf{u}\colon \mathbb{R}^3 \to \mathbb{R}^3$. Then,

$$J((\nabla \cdot \mathbf{u}) \circ \boldsymbol{\phi}) = \nabla \cdot (J\mathbf{F}^{-1}\mathbf{u} \circ \boldsymbol{\phi}). \tag{2.69}$$

This example really shows the power of Theorem 2.11. The simple concept of commutativity Theorem 2.11 becomes rather opaque vector identities especially for the Piola identity (Theorem 2.18) and its counterpart for curl (Theorem 2.17) which is in fact too complicated to be discovered with a name in vector calculus.

The Piola identity also implies the following identity [Evans(1998), Sec 8.1]. Specialize $\mathbf{u}$ into a constant coordinate vector, and note that the cofactor matrix of $\mathbf{F}$ is given by $\operatorname{cof}(\mathbf{F}) = J\mathbf{F}^{-\intercal}$.

Theorem 2.19 The cofactor matrix of the Jacobian matrix of any map has divergence-free rows.

■

■ **Example 2.13 — Theorem 2.12 in 3D.** In the 3D Cartesian space, Theorem 2.12 becomes the following theorems. Theorem 2.12 applied to 0-forms becomes

Theorem 2.20 — Fundamental theorem of calculus. Let $f\colon \mathbb{R}^3 \to \mathbb{R}$, and C be a curve connecting $\mathbf{a} \in \mathbb{R}^3$ to $\mathbf{b} \in \mathbb{R}^3$. Then,

$$\int_C \nabla f \cdot d\mathbf{l} = f(\mathbf{b}) - f(\mathbf{a}). \tag{2.70}$$

Theorem 2.12 applied to 1-forms becomes

Theorem 2.21 — Kelvin–Stokes curl theorem. Let $\mathbf{u}\colon \mathbb{R}^3 \to \mathbb{R}$, and S be a surface in $\mathbb{R}^3$ with normal vector $\mathbf{n}$. Then,

$$\iint_S (\nabla \times \mathbf{u}) \cdot \mathbf{n} dA = \oint_{\partial S} \mathbf{u} \cdot d\mathbf{l}. \tag{2.71}$$

Theorem 2.12 applied to 2-forms becomes

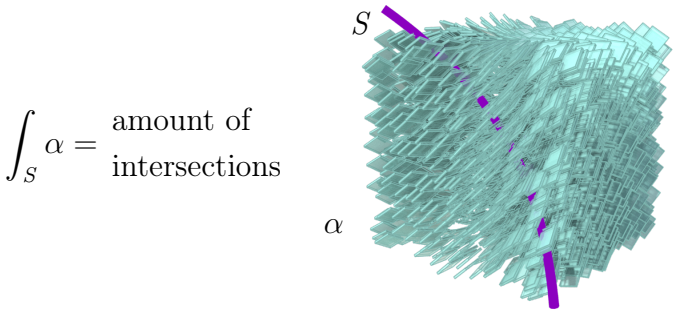


Figure 2.3 The integration of a k-form α along a k-dimensional surface S measures the total amount of intersections between S and the codimension-k geometries representing α.

Theorem 2.22 — Gauss' divergence theorem. Let $\mathbf{u}\colon \mathbb{R}^3 \to \mathbb{R}$, and V be a volumetric region in $\mathbb{R}^3$ with boundary normal $\mathbf{n}$. Then,

$$\iiint_V (\nabla \cdot \mathbf{u}) dV = \oiint_{\partial V} \mathbf{u} \cdot \mathbf{n} dA. \tag{2.72}$$

■

2.3.4 Differential forms as codimensional geometries

In this last part of Section 2.3, we provide visual representations for differential forms, wedge products, interior products, and exterior derivatives. These geometric interpretations are given in the appendix of [Yin et al.(2023)].

Geometrically, a vector field X is an assignment of an infinitesimal "arrow" $X_p \in T_pM$ at every point $p \in M$, whose directions and magnitudes depict some "instantaneous flow velocity" within the domain M. A k-form $\alpha \in \Omega^k(M)$, on the other hand, is a distribution of $(n-k)$-dimensional (codimension-k) oriented planes over M. See Figure 2.4. For example, in 3D, a 3-form is illustrated as a point cloud, a 2-form is a line segment cloud, a 1-form is a plane field, and a 0-form is a superposition of sublevel sets of the corresponding scalar function.

Integration is counting signed intersection

The orientations and densities of the codimension-k plane cloud are given so that the integration $\int_S \alpha$ over a test k-surface S is the total signed intersection between S and the codimensional-k plane cloud (Figure 2.3).[3]

Pullback is taking preimage

The pullback $\phi^*\alpha$ of a k-form $\alpha \in \Omega^k(M)$ has its codimension cloud given by the preimage of the codimension cloud of α through ϕ.

[3]For $n = 2, 3$, every k-form at each tangent space admits a distinguished $(n-k)$-subspace that represents the orientation of the plane field. For $n > 3$ and $1 < k < n-1$, the codimension-k plane fields are generally no longer described by a distinguished oriented subspace but a superposition of many. This is related to the *decomposability* of forms (Definition 4.3).

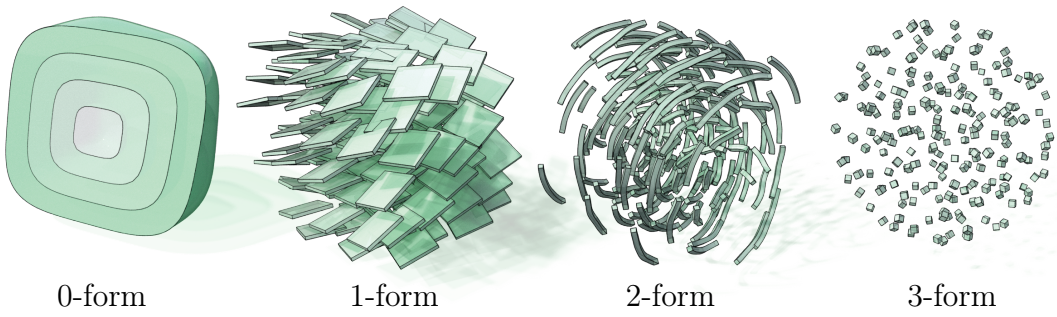


Figure 2.4 Differential k-forms can be represented by clouds of codimension-k geometries.

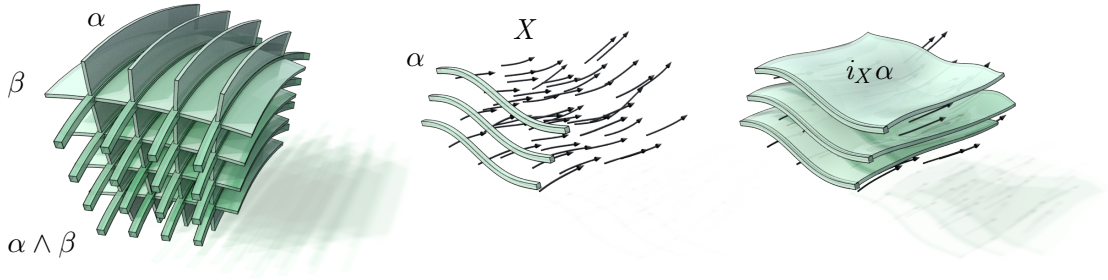


Figure 2.5 Wedge products are intersections (left) and interior products are extrusions (right).

Wedge product is taking intersection

The wedge product $\alpha \wedge \beta$ of a k-form $\alpha \in \Omega^k(M)$ and an ℓ-form $\beta \in \Omega^\ell(M)$ is a $(k+\ell)$-form whose codimension cloud is given by the intersection of the codimension clouds of α and β (Figure 2.5, left).

Interior products are extrusions

The interior product $i_X\alpha$ of a k-form $\alpha \in \Omega^k$ with a vector field $X \in \Gamma(TM)$ is a $(k-1)$-form whose codimension cloud is given by the extrusion of the codimension cloud of α along X (Figure 2.5, right.)

Exterior derivatives are boundary operators

The exterior derivative operator $d\colon \Omega^k(M) \to \Omega^{k+1}(M)$ takes the boundary of the codimensional geometric representation. For each $\alpha \in \Omega^k(M)$, its exterior derivative $d\alpha \in \Omega^{k+1}(M)$ has the $(n-k-1)$ dimensional cloud elements given by the boundaries of the $(n-k)$-dimensional cloud elements of α (Figure 2.6). The Stokes Theorem $\int_S d\alpha = \oint_{\partial S} \alpha$ can be interpreted as the invariant of intersection number when swapping the boundary operation. A k-form $\alpha \in \Omega^k(M)$ is **closed** if $d\alpha = 0$. For a closed form, the boundaries of each instance of the codimensional cloud cancel out with the neighboring boundaries. As a consequence, the codimensional cloud of a closed form stitches together into global pieces of $(n-k)$-dimensional surfaces foliating the space (Figure 2.7).

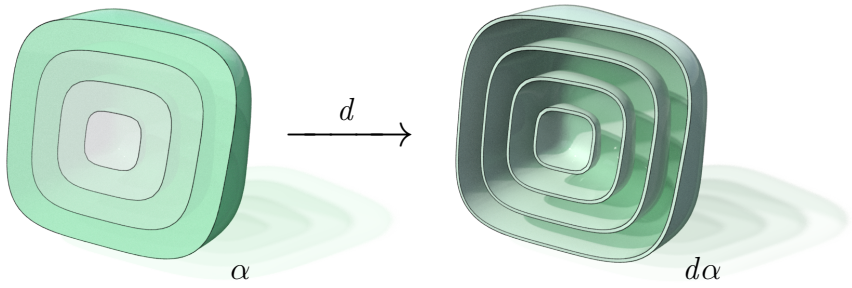


Figure 2.6 Taking exterior derivative is taking boundary.

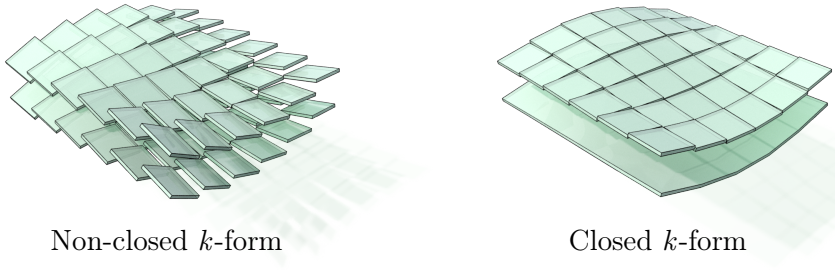


Figure 2.7 A k-form is closed when the boundaries of its codimension-k geometries cancel out.

2.4 Vector Fields as Derivations

Now, we will take a pause on studying the operations on differential forms, and expand on the idea of directional derivatives of functions. Directional derivatives are the starting point of multivariable calculus. It turns out that the collection of directional derivative operations form a space that is naturally endowed with an algebra that encodes important information about the direction fields themselves. This algebra is called a *Lie algebra*. Such an algebra is used in mechanics and robotics for distinguishing constrained systems (holonomic constraints or non-holonomic constraints).

2.4.1 Algebra and derivations

Definition 2.8 — Algebra. A vector space $\mathcal{A}$ is called an **algebra** if it is equipped with a bilinear function $\times : \mathcal{A} \times \mathcal{A} \xrightarrow{\text{bilinear}} \mathcal{A}$.

■ **Example 2.14** The space $\mathbb{R}^{n\times n}$ of n-by-n matrices is an algebra, with matrix-matrix multiplication as the bilinear function. ■

■ **Example 2.15** The space $\Omega^0(M)$ of functions is an algebra, with pointwise multiplication as the bilinear function. ■

Note that the multiplication structure (bilinear function) for an algebra does not need to be associative:

■ **Example 2.16** $(\mathbb{R}^3, \times)$ with cross product $\times$ is an algebra. Note that $\mathbf{u}\times(\mathbf{v}\times\mathbf{w}) \neq (\mathbf{u}\times\mathbf{v})\times\mathbf{w}$ in general. ■

Definition 2.9 — Derivations. Let $(\mathcal{A}, \times)$ be an algebra. A linear map $X : \mathcal{A} \xrightarrow{\text{linear}} \mathcal{A}$ is called a **derivation** if

$$X(a \times b) = (Xa) \times b + a \times (Xb) \quad a, b \in \mathcal{A} \tag{2.73}$$

The space of all derivations over an algebra is denoted by $\mathrm{Der}(\mathcal{A})$.

Definition 2.10 — View vector fields as derivations. Each vector field $X \in \Gamma(TM)$ is also viewed as an operator on functions $X : \Omega^0(M) \to \Omega^0(M)$ defined by

$$Xf \coloneqq df[\![X]\!]. \tag{2.74}$$

Note that this functional operator is a derivation on functions

$$X(fg) = (Xf)g + f(Xg). \tag{2.75}$$

In fact, a converse statement is true: *every derivation on function is a vector field*:

$$\Gamma(TM) = \mathrm{Der}(\Omega^0(M)). \tag{2.76}$$

This equivalence between vector fields and the algebraic definition of derivations on functions motivates many differential geometry textbooks to adopt (2.76) as the *definition* for vector fields.

The takeaway for this abstraction is that we can think of each vector field as the directional derivative operator (2.74) that obeys the derivation property (2.75).

2.4.2 Lie algebra for vector fields

Definition 2.11 — Lie algebra. An algebra $(\mathcal{A}, [\cdot,\cdot])$ is called a **Lie algebra** if the bilinear function $[\cdot,\cdot] : \mathcal{A}\times\mathcal{A} \xrightarrow{\text{bilinear}} \mathcal{A}$ satisfies skew symmetry

$$[a, b] = -[b, a] \tag{2.77}$$

and the *Jacobi identity*

$$[a, [b, c]] + [b, [c, a]] + [c, [a, b]] = 0. \tag{2.78}$$

Here are two elementary examples for Lie algebra.

■ **Example 2.17** $(\mathbb{R}^3, \times)$ with the 3D cross product $\times$ is a Lie algebra. ■

■ **Example 2.18** $(\mathbb{R}^{n\times n}, [\cdot,\cdot])$ is a Lie algebra, where the bracket operation is the matrix commutator $[A, B] = AB - BA$. ■

Theorem 2.23 The space of vector fields $\Gamma(TM)$ is a Lie algebra, with Lie bracket $[X, Y]$ of vector fields X, Y defined by

$$[X, Y]f \coloneqq X(Yf) - Y(Xf) \tag{2.79}$$

when it is operated on functions.

Proof. We only need to verify two properties. First, $[X, Y] \coloneqq X \circ Y - Y \circ X$ is still a derivation on functions, and hence $[\cdot, \cdot]$ is a valid algebra operation. Second, $[X, Y]$ satisfies the Jacobi identity. Each of these properties is easy to check straight from definitions. □

R Space of tangent vector fields, or equivalently the space of derivations on functions, is naturally equipped with a Lie algebra structure.

2.4.3 Practical calculation of Lie bracket

Suppose M is a subset of $\mathbb{R}^n$ with coordinate system $x^1, \dots, x^n$. Let X, Y be two vector fields. As derivations, we represent them as directional differential operators

$$X = \sum_{i=1}^{n} X^i \frac{\partial}{\partial x^i}, \quad Y = \sum_{i=i}^{n} Y^i \frac{\partial}{\partial x^i} \tag{2.80}$$

where each components X^i, Y^i of the vector fields are functions defined over M. Let $Z = [X, Y]$. Our goal is to determine its component Z^i as in $Z = \sum_{i=1}^{n} Z^i \frac{\partial}{\partial x^i}$ in terms of the components of X^i, Y^i. Expanding the definition of Lie bracket,

$$Zf = X(Yf) - Y(Xf) = \sum_{i,j} X^i \frac{\partial}{\partial x^i}\left(Y^j \frac{\partial f}{\partial x^j}\right) - Y^i \frac{\partial}{\partial x^i}\left(X^j \frac{\partial f}{\partial x^j}\right) \tag{2.81}$$

$$= \sum_{i,j} \left(X^i \frac{\partial Y^j}{\partial x^i} - Y^i \frac{\partial X^j}{\partial x^i}\right) \frac{\partial f}{\partial x^j} + \underbrace{\left(X^i Y^j - Y^i X^j\right) \frac{\partial^2 f}{\partial x^i \partial x^j}}_{=0}. \tag{2.82}$$

Therefore,

$$Z^j = X^i \frac{\partial Y^j}{\partial x^i} - Y^i \frac{\partial X^j}{\partial x^i}. \tag{2.83}$$

If "∇" denotes taking componentwise differential, we can write

$$Z = [X, Y] = \nabla_X Y - \nabla_Y X. \tag{2.84}$$

Note that each of the coordinate-componentwise differential $\nabla_X Y$ or $\nabla_Y X$ *depends on the choice of coordinate*. But the difference $\nabla_X Y - \nabla_Y X$ becomes $[X, Y]$ which is coordinate independent.

In practice, to compute the Lie bracket of two vector fields, one can choose *any* coordinate system (which doesn't need to be orthonormal) and take the mutual componentwise derivative (2.84).

2.4.4 Integrability

Why do we care about the Lie algebra structure for vector fields? One of the most important application of the Lie structure for vector fields is the **integrability to coordinate system**.

Theorem 2.24 Suppose $x_1, \dots, x_n \in \Omega^0(M)$ be a coordinate system, which gives rise to covector basis field $dx_1, \dots, dx_n \in \Omega^1(M)$. Let $\vec{e}_1, \dots, \vec{e}_n \in \Gamma(TM)$ be the dual basis from $dx_1, \dots, dx_n \in \Omega^1(M)$, called coordinate vectors. Then $[\vec{e}_i, \vec{e}_j] = 0$.

Conversely, if a basis vector field $X_1, \dots, X_n \in \Gamma(TM)$ satisfies $[X_i, X_j] = 0$ for each pair of $i, j = 1, \dots, n$. Then they are locally the coordinate vectors of some coordinate system.

Definition 2.12 — Distribution. A **distribution** $\mathcal{A}$ is an assignment of a linear subspace $\mathcal{A}_p \subset T_pM$ of the tangent space at each point $p \in M$. We say a vector field $X \in \Gamma(TM)$ belongs to the distribution, denoted by $X \in \Gamma(\mathcal{A})$, if $X_p \in \mathcal{A}_p$ for all $p \in M$.

A distribution can be seen as a plane field defined over the space M. Like Figure 2.7, there may be two cases: a general plane field may not stitch together with its neighbor, or it may stitch together and form a foliation of submanifolds. In the former case, we say the distribution is non-integrable. In the latter case, we say that the distribution is integrable.

Definition 2.13 A distribution $\mathcal{A} \subset TM$ is said to be **integrable** if the plane field associated to $\mathcal{A}$ is locally the tangent plane of a family of submanifolds in M.

Theorem 2.25 — Frobenius theorem of integrability. A distribution $\mathcal{A} \subset TM$ is integrable if and only if $\Gamma(\mathcal{A}) \subset \Gamma(TM)$ is not only a vector subspace but also a Lie subalgebra; that is, the Lie bracket of vector fields in the distribution stays in the distribution: $[X, Y] \in \Gamma(\mathcal{A})$ for all $X, Y \in \Gamma(\mathcal{A})$.

This notion of integrability plays a crucial role in characterizing robotics and mechanical constructs. Such systems are parameterized through general coordinates like joint angles. Often, these parameters are subject to limitations concerning their infinitesimal mobility.

The range of feasible infinitesimal movements constitutes a distribution over the tangent bundle of the parameter space. If this distribution proves to be integrable, the constraint is referred to as a **holonomic constraint**. Otherwise, if the distribution is non-integrable, it is termed a **non-holonomic constraint**.

A system with holonomic constraints can be simplified to a lower-dimensional system, which is the submanifold integrated from the distribution. On the other hand, a system with non-holonomic constraints allows for certain maneuvers within the realm of feasible movements, yet these can lead to directions transversal to the distribution. Essentially, non-holonomic constraints are velocity restrictions that do not sufficiently confine the positions to lower-dimensional spaces.

■ **Example 2.19 — Rolling ball.** A rolling ball moving on a surface without slipping presents an example of a nonholonomically constrained system. In this case, the constraint arises from the no-slip condition, which dictates that the point of the ball in contact with the surface has zero velocity relative to the surface. Under this constraint, the 3D orientation of the ball have only 2 degrees of freedom among the 3-dimensional space of all 3D rotations. However, moving in an infinitesimal cycle using the 2 degrees of freedom can yield a net motion into the 3rd degree of freedom.

Let $M = \mathrm{SO}(3)$ be the space of 3D rotations. Consider a basis of vector fields

$$X_{\mathbf{R}} = [\mathbf{e}_1\times]\mathbf{R}, \quad Y_{\mathbf{R}} = [\mathbf{e}_2\times]\mathbf{R}, \quad Z_{\mathbf{R}} = [\mathbf{e}_3\times]\mathbf{R} \quad \in T_{\mathbf{R}}\,\mathrm{SO}(3) \tag{2.85}$$

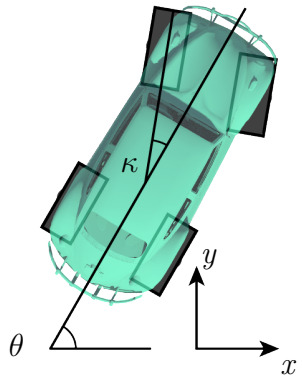
at each $\mathbf{R} \in \mathrm{SO}(3) \subset \mathbb{R}^{3\times 3}$. Here, $[\begin{bmatrix} a \\ b \\ c \end{bmatrix}\times] = \begin{bmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{bmatrix}$ is the matrix representation of cross product. The rolling-without-slipping condition means that the allowed motion is in the span of X, Y. Note that, using the matrix components as our coordinate system, the directional derivative of the coordinate is $\nabla_X \mathbf{R} = X_{\mathbf{R}}$, $\nabla_Y \mathbf{R} = Y_{\mathbf{R}}$. In particular, $\nabla_X(Y) = [\mathbf{e}_2\times]\nabla_X\mathbf{R} = [\mathbf{e}_2\times][\mathbf{e}_1\times]\mathbf{R}$. By a similar computation, we get

$$[X, Y]_{\mathbf{R}} = \nabla_X Y_{\mathbf{R}} - \nabla_Y X_{\mathbf{R}} = [\mathbf{e}_1\times][\mathbf{e}_2\times]\mathbf{R} - [\mathbf{e}_2\times][\mathbf{e}_1\times]\mathbf{R} \tag{2.86}$$

$$= [\mathbf{e}_1 \times \mathbf{e}_2]\mathbf{R} = [\mathbf{e}_3\times]\mathbf{R} = Z_{\mathbf{R}}. \tag{2.87}$$

In particular, Lie bracket between vector fields within the distribution $\mathrm{span}(X, Y)$ can leave the distribution. Therefore, we have a non-holonomically constrained system. ■

■ **Example 2.20 — Parallel parking.** The configuration space of a four-wheel car on a plane is $M = \{(x, y, \theta, \kappa) : x, y \in \mathbb{R}, \theta, \kappa \in [0, 2\pi)\}$, where (x, y) denotes the location of the car on the plane, θ denotes the angle between the central axis of the car and the x-axis, and κ denotes the angle at which the front wheels are turned from the car's central axis.



The car can move by hitting its acceleration and going forward in its current direction θ and changing its direction by its current steering wheel angle κ, or the

car can move by changing the angle of its steering wheel. These two actions each corresponds to the following two generators:

$$X_{\text{forward}} = \begin{bmatrix} \cos\theta \\ \sin\theta \\ \kappa \\ 0 \end{bmatrix} \quad \text{and} \quad X_{\text{turn}} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}. \tag{2.88}$$

To see the effect of applying the following sequence of actions: move forward, steer to the left, move backward, steer to the right, we can look at the Lie bracket of these two generators:

$$\begin{aligned} [X_{\text{forward}}, X_{\text{turn}}] &= \nabla_{X_{\text{forward}}} X_{\text{turn}} - \nabla_{X_{\text{turn}}} X_{\text{forward}} \\ &= \begin{bmatrix} 0 & 0 & -\sin\theta & 0 \\ 0 & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \cos\theta \\ \sin\theta \\ \kappa \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}. \end{aligned} \tag{2.89}$$

This means that the action sequence generates a spinning motion. We denote its generator by

$$X_{\text{spin}} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}. \tag{2.90}$$

We can further analyze the effect of the following sequence of actions: spin to the left, move forward, spin to the right, move backward by calculating the Lie bracket

$$\begin{aligned} [X_{\text{spin}}, X_{\text{forward}}] &= \nabla_{X_{\text{spin}}} X_{\text{forward}} - \nabla_{X_{\text{forward}}} X_{\text{spin}} \\ &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \cos\theta \\ \sin\theta \\ \kappa \\ 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 & -\sin\theta & 0 \\ 0 & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -\sin\theta \\ \cos\theta \\ 0 \\ 0 \end{bmatrix}. \end{aligned} \tag{2.91}$$

This means that the action sequence generates a sideway motion where the car moves towards $-\sin\theta, \cos\theta$ while its front points towards $(\cos\theta, \sin\theta$. We can denote this generator as X_{sideway}. We've thus far obtained four generators of motion:

$$X_{\text{forward}} = \begin{bmatrix} \cos\theta \\ \sin\theta \\ \kappa \\ 0 \end{bmatrix}, X_{\text{turn}} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, X_{\text{spin}} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \text{and } X_{\text{sideway}} = \begin{bmatrix} -\sin\theta \\ \cos\theta \\ 0 \\ 0 \end{bmatrix}. \tag{2.92}$$

What we are seeing is that by introducing Lie bracket, the space of all possible motion made from sequencing moving forward and turning the steering wheel is actually enough to cover the entire tangent space T_pM of the configuration manifold M, meaning the car is capable of traveling over the entire plane even though it only has two motions. ∎

2.5 Covariant Derivatives and Vector-Valued Forms

As we briefly remarked at (2.84), the componentwise differential "$\nabla_X Y$" of a vector field Y along the X direction is coordinate dependent. What is a coordinate-independent notion of $\nabla_X Y$?

So far, we have only been looking at scalar-valued k-forms. (Componentwise differentials in (2.84) is treating vectors as multi-component scalars.) Another frequently occurring type of object is *vector-valued k-forms*. Recall a k-form is a

section of the bundle $\bigwedge^k T^*M$. One may take a pointwise tensor product with the tangent bundle $(\bigwedge^k T^*M) \otimes TM$, whose sections are both a k-form and a vector field. These are vector-valued k-forms. For notation convenience we call

$$\Gamma\left((\bigwedge^k T^*M) \otimes TM\right) = \Omega^k(M; TM). \tag{2.93}$$

A vector field $Y \in \Gamma(TM)$ can also be seen as a vector-valued 0-form $Y \in \Omega^0(M; TM)$. One may ask what is its directional derivative along a direction $X \in T_pM$ at a point $p \in M$.

The natural definition of the directional derivative of a vector would depend on a metric.

Definition 2.14 Let $(M, \flat)$ be a Riemannian manifold. A differential operator that takes a vector field to a vector-valued 1-form

$$\nabla\colon \Omega^0(M; TM) \to \Omega^1(M; TM) \tag{2.94}$$

is a **Levi-Civita connection** or a **covariant derivative** if

- Compatible with differential of scalar: $\nabla(fY) = (df)\,Y + f\nabla Y$ for $f \in \Omega^0(M)$ and $Y \in \Omega^0(M; TM)$; (2.95)
- Compatible with metric: $d\langle X, Y\rangle = \langle \nabla X, Y\rangle + \langle X, \nabla Y\rangle$; (2.96)
- Compatible with Lie bracket (torsion free): $\nabla_X Y - \nabla_Y X = [X, Y]$. (2.97)

Theorem 2.26 — Fundamental theorem of Riemannian geometry. The Levi-Civita connection exists and is unique.

So, on a manifold with metric, we can just take differential of a vector field like taking differential of a standard 0-form. The only difference in notation is that we use ∇ instead of d.

Definition 2.15 — Exterior covariant derivative. Using the covariant derivative ∇ we can apply exterior derivative to vector-coefficient k-forms (*i.e.* vector-valued k-forms). The operator

$$d^\nabla\colon \Omega^k(M; TM) \to \Omega^{k+1}(M; TM) \tag{2.98}$$

is characterized by that

- It is the same as ∇ when $k = 0$. (2.99)
- Graded Leibniz rule: $d^\nabla(\alpha \wedge A) = (d\alpha) \wedge A + (-1)^{\deg(\alpha)}\alpha \wedge d^\nabla A$ for scalar-valued form α and vector-valued form A. (2.100)

R Despite the similarity between d^∇ and d, there are some differences. For example,

$$d^\nabla \circ d^\nabla \tag{2.101}$$

is generally nonzero. In fact, $d^\nabla \circ d^\nabla = R\wedge$ where R is a matrix-valued 2-form called **Riemann curvature tensor**.

Definition 2.16 The identity is a vector valued 1-form

$$I \in \Omega^1(M; TM), \quad I[\![X]\!] \coloneqq X. \tag{2.102}$$

The identity 1-form I is useful for expressing the flat of a vector

$$X^\flat = \langle X, I \rangle. \tag{2.103}$$

One checks that $X^\flat[\![Y]\!] = \langle X, I[\![Y]\!] \rangle = \langle X, Y \rangle$.

Theorem 2.27 The torsion free condition (2.97) for a derivative operator ∇ is equivalent to

$$d^\nabla I = 0. \tag{2.104}$$

Proof. The exterior (covariant) derivative d on a 1-form α has an explicit formula of $(d\alpha)[\![X, Y]\!] = d(\alpha[\![Y]\!])[\![X]\!] - d(\alpha[\![X]\!])[\![Y]\!] - \alpha[\![[X, Y]]\!]$, which we will give a proof later in Theorem 2.36. Now, letting α be I, and X, Y be arbitrary vector fields, we get

$$\begin{aligned} (d^\nabla I)[\![X, Y]\!] &= \nabla(I[\![Y]\!])[\![X]\!] - \nabla(I[\![X]\!])[\![Y]\!] - I[\![[X, Y]]\!] && (2.105)\\ &= \nabla_X Y - \nabla_Y X - [X, Y]. && (2.106) \end{aligned}$$

This expression vanishes if and only if the torsion free condition (2.97) holds. □

2.6 Lie Derivative

Lie derivative is an extremely fundamental derivative in multivariable calculus, which is unfortunately not mentioned in most of the course on multivariable calculus.

Lie derivative is defined by *the rate of change of a time-dependent pullback operator*. Recall that the pullback operator is the just the unification of general changes of coordinates/variables. The question about what the rate of change is of a time-dependent change of coordinates/variables naturally arises. The unifying concept for this rate of change is the Lie derivative, which is as important as how changes of variables are important in calculus.

The use of Lie derivative is independently introduced, with different names, in applied research fields when rates of change of changes of variables are unavoidable. For example, its special cases include Reynolds transport theorem that is used in differentiable rendering [Zhao et al.(2020b)]. Lie derivative also unifies convection and stretching terms in continuum mechanics, which is the study of the differential relations of a deformable body.

2.6.1 General definition

Let $A \in \langle \text{type} \rangle_M$ be a field defined over M of a particular type $\langle \text{type} \rangle$. This space $\langle \text{type} \rangle_M$ can be functions (0-forms) $\Omega^0(M)$, k-forms $\Omega^k(M)$, vector fields $\Gamma(TM)$, or any other types that has a pullback operator

$$\underset{\langle \text{type} \rangle}{\varphi^*} : \langle \text{type} \rangle_M \xrightarrow{\text{linear}} \langle \text{type} \rangle_M \tag{2.107}$$

associated to any map $\varphi \colon M \to M$. The Lie derivative is the differential operator that measures the rate of change of the pullback $\underset{\langle \text{type} \rangle}{\varphi_t^*} A$ of A when the time-dependent map φ_t as a flow map has a known velocity field.

Definition 2.17 — Flow map generated by vector field. For each vector field $X \in \Gamma(TM)$ on M, the **flow map** $\varphi_t \colon M \to M$, $t \in (-\epsilon, \epsilon)$, generated by X is the solution to

$$\frac{\partial \varphi}{\partial t} = X \circ \varphi \quad \text{and} \quad \varphi_0 = \mathrm{id}_M. \tag{2.108}$$

Definition 2.18 — Lie derivative. Let $X \in \Gamma(TM)$ be a vector field on M and let $\varphi_t \colon M \to M$ be the flow map generated by X (Definition 2.17). Define the **Lie derivative** for $\langle \text{type} \rangle$

$$\underset{\langle \text{type} \rangle}{\mathscr{L}_X} : \langle \text{type} \rangle_M \xrightarrow{\text{linear}} \langle \text{type} \rangle_M \tag{2.109}$$

by

$$\underset{\langle \text{type} \rangle}{\mathscr{L}_X} A \coloneqq \frac{\partial}{\partial t}\bigg|_{t=0} \left(\underset{\langle \text{type} \rangle}{\varphi_t^*} A \right). \tag{2.110}$$

This notion of Lie derivative can be used in situations beyond taking derivative of the infinitesimally flow map (2.108) generated around the identity map. Whenever there is a time dependent map, the rate of change can of the pullback can be written in terms of the Lie derivative.

Theorem 2.28 — Rate of change of pullback. Let M and W be two manifolds. Let $\phi_t \colon M \to W$ be a time-dependent map, and suppose there is $\vec{u} \in \Gamma(TW)$ so that

$$\frac{\partial \phi_t}{\partial t} = \vec{u} \circ \phi_t. \tag{2.111}$$

Let $A \in \langle \text{type} \rangle_W$ be some field defined on W. Then

$$\frac{\partial}{\partial t} \left(\underset{\langle \text{type} \rangle}{\phi_t^*} A \right) = \underset{\langle \text{type} \rangle}{\phi_t^*} \left(\underset{\langle \text{type} \rangle}{\mathscr{L}_{\vec{u}}} A \right). \tag{2.112}$$

Corollary 2.29 — Passing the time derivative through a pullback. Continue the setup of Theorem 2.28. Suppose $A_t \in \langle \mathsf{type} \rangle_W$ is a time-dependent field defined on W. Then

$$\frac{\partial}{\partial t}\left(\underset{\langle \mathsf{type}\rangle}{\phi_t^*}\, A_t\right) = \underset{\langle \mathsf{type}\rangle}{\phi_t^*}\left(\frac{\partial}{\partial t}A_t + \underset{\langle \mathsf{type}\rangle}{\mathscr{L}_{\vec{u}}}\, A_t\right). \tag{2.113}$$

We call the operator $(\frac{\partial}{\partial t} + \mathscr{L}_{\vec{u}})$ the **Lie material derivative**.

Definition 2.19 We say that a time-dependent field $A_t \in \langle \mathsf{type} \rangle_W$ is **Lie-advected** by $\vec{u}$ if

$$\frac{\partial}{\partial t}A_t + \underset{\langle \mathsf{type}\rangle}{\mathscr{L}_{\vec{u}}}\, A_t = 0; \tag{2.114}$$

equivalently, for $\frac{\partial \phi_t}{\partial t} = \vec{u} \circ \phi_t$,

$$\underset{\langle \mathsf{type}\rangle}{\phi_t^*}\, A_t = \underset{\langle \mathsf{type}\rangle}{\phi_0^*}\, A_0 \quad \text{is time independent.} \tag{2.115}$$

2.6.2 Lie derivative for functions

The Lie derivative for functions is just the directional derivative.

Theorem 2.30 — Lie derivative for functions. For $f \in \Omega^0(M)$

$$\underset{0\text{-form}}{\mathscr{L}_X}\, f = df[\![X]\!] = Xf. \tag{2.116}$$

Proof. Let φ_t be the flow map generated by $X \in \Gamma(TM)$ (Definition 2.17). Then $\underset{0\text{-form}}{\mathscr{L}_X}\, f = \frac{\partial}{\partial t}|_{t=0}(\underset{0\text{-form}}{\varphi_t^*}\, f) = \frac{\partial}{\partial t}|_{t=0}(f \circ \varphi_t) = df|_{\varphi_t}[\![\frac{\partial \varphi_t}{\partial t}]\!]|_{t=0} = df[\![X]\!]$. □

2.6.3 Lie derivative for vector fields

The Lie derivative for vector fields is the Lie bracket (Theorem 2.23).

Theorem 2.31 — Lie derivative for vector fields. For $X, Y \in \Gamma(TM)$

$$\underset{\text{vec}}{\mathscr{L}_X}\, Y = [X, Y]. \tag{2.117}$$

To show Theorem 2.31, first note that the pullback of a vector field is given by

$$\underset{\text{vec}}{\varphi_t^*}\, Y = (d\varphi_t)^{-1}\, Y. \tag{2.118}$$

This pullback of vectors is natural, as we have

$$\underset{0\text{-form}}{\varphi_t^*}\, (Yf) = (\underset{\text{vec}}{\varphi_t^*}\, Y)(\underset{0\text{-form}}{\varphi_t^*}\, f). \tag{2.119}$$

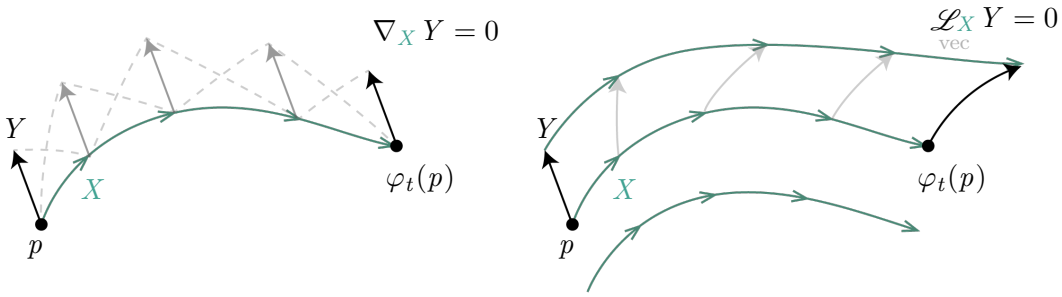


Figure 2.8 The difference between the covariant derivative $\nabla_X Y$ and the Lie derivative $\underset{\text{vec}}{\mathscr{L}_X} Y = [X, Y]$. Vanishing covariant derivative implies that Y is parallel transported by the path generated by X (left). The dashed lines are geodesics constituting infinitesimal parallelograms for the transportation (known as *Schild's ladder*), indicating the metric dependency of ∇_X. On the other hand, vanishing Lie derivative implies that Y is the pushed forward by the flow map generated by X (right), which is metric independent but requires the neighborhood information of X instead of only along a path.

Proof of Theorem 2.31. Let φ_t be the flow map generated by $X \in \Gamma(TM)$, and let f be an arbitrary function. Taking the time derivative of (2.119) yields

$$\underset{\text{0-form}}{\mathscr{L}_X}\,(Yf) = (\underset{\text{vec}}{\mathscr{L}_X}\, Y)f + Y(\underset{\text{0-form}}{\mathscr{L}_X}\, f) \tag{2.120}$$

$$\Longrightarrow \quad X(Yf) = (\underset{\text{vec}}{\mathscr{L}_X}\, Y)f + Y(Xf) \tag{2.121}$$

$$\Longrightarrow \quad (\underset{\text{vec}}{\mathscr{L}_X}\, Y)f = X(Yf) - Y(Xf) = [X, Y]f. \tag{2.122}$$

□

Comparison between covariant derivative $\nabla_X Y$ and Lie derivative $\underset{\text{vec}}{\mathscr{L}_X} Y$

We have learned two derivatives of a vector field $Y \in \Gamma(TM)$ along another vector field $X \in \Gamma(TM)$. One is the covariant derivative $\nabla_X Y$ (Definition 2.14), and the other is the Lie derivative $\underset{\text{vec}}{\mathscr{L}_X} Y$ (Theorem 2.31). Figure 2.8 shows their differences. Consider the vector fields Y that satisfies $\nabla_X Y = 0$, and respectively $\underset{\text{vec}}{\mathscr{L}_X} Y = 0$.

The equation $\nabla_X Y = 0$ implies that Y is *parallel transported* along X, and $\nabla_X Y$ measures how much Y deviates from being parallel when we take a step in the X direction. This type of X-directional derivative of Y only depends on the direction of X at a single point. But the result $\nabla_X Y$ depends on a metric as is required in Definition 2.14. In fact, since $\nabla_X Y$ is a directional derivative, we have

$$\nabla_X Y = \underset{\text{0-form}}{\mathscr{L}_X^{\nabla}}\, Y \tag{2.123}$$

where Y is viewed as a vector-valued 0-form, and the Lie derivative is built by taking the rate of change of the 0-form-typed pullback.

The equation $\underset{\text{vec}}{\mathscr{L}_X} Y = 0$, on the other hand, implies that Y is pushed forward by the flow generated by X. The quantity $\underset{\text{vec}}{\mathscr{L}_X} Y$ measures how much Y deviates from

being purely pushed forward by the flow. The idea of pushforward by a flow, as well as the Lie bracket, is independent of any metric. However, this type of "directional derivative" of Y along a flow X requires the information of X in a neighborhood, not only at one point. It is the neighborhood behavior of X that establishes an infinitesimal pushforward (deformation gradient or Jacobian matrix) of the flow.

2.6.4 Lie derivative for differential forms

Finally, we give the formula for the Lie derivative for differential forms.

Theorem 2.32 — Cartan's magic formula. For $\alpha \in \Omega^k(M)$

$$\underset{k\text{-form}}{\mathscr{L}_X}\ \alpha = i_X\, d\alpha + d i_X \alpha. \tag{2.124}$$

Lie derivatives for differential forms are particularly important, as they allow us to take derivative on the integration domain:

Theorem 2.33 — Derivative of an integral. Let X be the vector field so that $\dot{\phi} = X \circ \phi$. Then,

$$\frac{d}{dt}\int_{\phi_t(S)} \alpha_t = \int_{\phi_t(S)} \left(\frac{\partial}{\partial t}\alpha_t + \underset{k\text{-form}}{\mathscr{L}_X}\ \alpha_t\right) \tag{2.125}$$

Proof of Theorem 2.33. $\frac{\partial}{\partial t}\int_{\phi_t(S)} \alpha_t = \frac{\partial}{\partial t}\int_S \phi_t^* \alpha_t = \int_S \frac{\partial}{\partial t}(\phi_t^* \alpha_t) = \int_S \phi_t^*(\frac{\partial}{\partial t}\alpha + \mathscr{L}_X\, \alpha)$ $= \int_{\phi_t(S)}(\frac{\partial}{\partial t}\alpha + \mathscr{L}_X\, \alpha)$. □

Now we give a simple but abstract proof for Cartan's magic formula (2.124). This proof still deserves geometric intuition, which we will provide afterward.

Proof of Theorem 2.32. The operator $\underset{k\text{-form}}{\mathscr{L}_X}$ must satisfy the following 3 properties:

- Base definition: $\underset{k\text{-form}}{\mathscr{L}_X} = \underset{0\text{-form}}{\mathscr{L}_X}$ when $k = 0$;
- Derivation: $\underset{\text{form}}{\mathscr{L}_X}(\alpha \wedge \beta) = (\underset{\text{form}}{\mathscr{L}_X}\alpha) \wedge \beta + \alpha \wedge (\underset{\text{form}}{\mathscr{L}_X}\beta)$;
- Commutativity with d: $\underset{\text{form}}{\mathscr{L}_X}\, d\alpha = d\underset{\text{form}}{\mathscr{L}_X}\alpha$.

The "derivation" and "commutativity with d" properties follow directly from the analogous properties Theorem 2.7 and Theorem 2.11 for the pullback operator. Now, notice that the operators on differential forms that satisfy these 3 properties must be *unique*. This is because any differential form ω can be written as a linear combination (with 0-form as coefficients) of wedge products of the d of some coordinate functions $\omega = \sum_{i_1<\ldots<i_k} \omega_{i_1\ldots i_k}\, dx^{i_1} \wedge \cdots \wedge dx^{i_k}$, and the above 3 properties form a sufficient set of rules to evaluate the Lie derivative of this expression. To complete the proof, check that (2.124) satisfy all the 3 properties. □

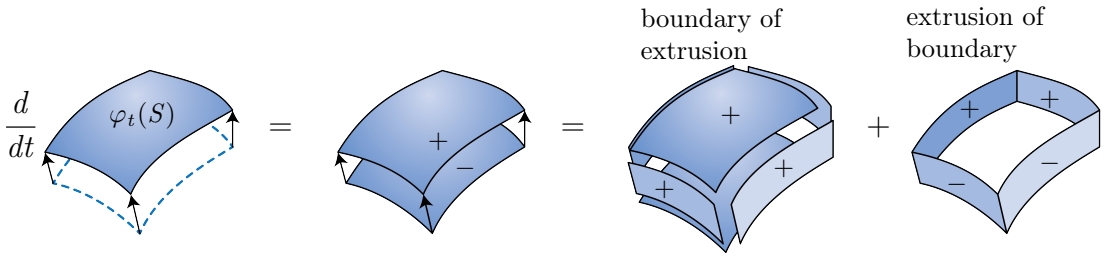


Figure 2.9 An intuition behind Cartan's magic formula. The Lie derivative of a differential form is dual to the rate of change of a moving integration domain. This rate of change of a domain can be written as the boundary of the extrusion plus the extrusion of the boundary.

Geometric intuition behind Cartan's magic formula

The geometric intuition behind (2.124) is best appreciated when viewed under an integral sign. To understand this, consider (2.125), but with α being independent of t, and $\phi_t = \varphi_t$ representing the flow map generated by X around the identity map as given by Definition 2.17:

$$\int_S \mathscr{L}_X \, \alpha = \frac{d}{dt} \int_{\varphi_t(S)} \alpha. \tag{2.126}$$

In essence, the Lie derivative, through the integral pairing between an integration domain and a differential form, acts as the dual operator to the rate of change of the integration domain.

As illustrated in Figure 2.9, this rate of change of the integration domain can be expressed as the sum of the boundary of the extrusion along the flow and the extrusion of the boundary. Recall from Theorem 2.10 and Theorem 2.12 that the operations of extrusion and taking boundary are dual to the interior product and the exterior derivative, respectively. Thus, we arrive at Cartan's magic formula.

2.6.5 Useful identities

We have explored different types of Lie derivatives in Sections 2.6.2–2.6.4. Here, we collect a few identities by combining these various types of Lie derivatives.

Theorem 2.34 Let $X, Y_1, \ldots, Y_k$ be vector fields and α be a k-form. Then

$$\begin{aligned}\mathscr{L}_X(\alpha[\![Y_1, \ldots, Y_k]\!]) = (\mathscr{L}_X \, \alpha)[\![Y_1, \ldots, Y_k]\!] + \alpha[\![[X, Y_1], \ldots, Y_k]\!] + \\ \cdots + \alpha[\![Y_1, \ldots, [X, Y_k]]\!]\end{aligned} \tag{2.127}$$

Proof. This follows from taking the t-derivative of

$$\underset{\text{0-form}}{\varphi_t^*} \; (\alpha[\![Y_1, \ldots, Y_k]\!]) = (\underset{k\text{-form}}{\varphi_t^*}) [\![\underset{\text{vec}}{\varphi_t^*} Y_1, \ldots, \underset{\text{vec}}{\varphi_t^*} Y_k]\!]. \tag{2.128}$$

□

Similar to Theorem 2.34, we have:

Theorem 2.35 Let X, Y be vector fields and α be a k-form. Then we have

$$\mathscr{L}_X(i_Y\alpha) = i_{[X,Y]}\alpha + i_Y(\mathscr{L}_X\alpha). \tag{2.129}$$

The following identity is an explicit formula for the evaluation of the d of a 1-form.

Theorem 2.36 For $\alpha \in \Omega^1(M)$, we have

$$(d\alpha)[\![X, Y]\!] = X\langle\alpha|Y\rangle - Y\langle\alpha|X\rangle - \langle\alpha|[X, Y]\rangle. \tag{2.130}$$

Proof. $(d\alpha)[\![X, Y]\!] = i_Y i_X d\alpha \overset{(2.124)}{=} i_Y(\mathscr{L}_X\alpha) - \underbrace{i_Y d i_X \alpha}_{Y\langle\alpha|X\rangle} \overset{(2.129)}{=} \mathscr{L}_X(i_Y\alpha) - i_{[X,Y]}\alpha - Y\langle\alpha|X\rangle = X\langle\alpha|Y\rangle - \langle\alpha|[X, Y]\rangle - Y\langle\alpha|X\rangle.$ □

2.6.6 Lie derivatives in the 3D Cartesian space

■ **Example 2.21** In the 3D Cartesian space, the Lie derivatives for differential forms and vector fields correspond to the following expressions in vector calculus.

Theorem 2.37 — Lie derivatives in vector calculus.

$$\underset{\text{0-form}}{\mathscr{L}_{\mathbf{u}}}\,(a)_{\text{0-form}} = (\mathbf{u}\cdot\nabla a)_{\text{0-form}}, \tag{2.131}$$

$$\underset{\text{1-form}}{\mathscr{L}_{\mathbf{u}}}\,(\mathbf{a})_{\text{1-form}} = (\mathbf{u}\cdot\nabla\mathbf{a} + (\nabla\mathbf{u})^\intercal\mathbf{a})_{\text{1-form}}, \tag{2.132}$$

$$\underset{\text{2-form}}{\mathscr{L}_{\mathbf{u}}}\,(\mathbf{a})_{\text{2-form}} = (\mathbf{u}\cdot\nabla\mathbf{a} - \mathbf{a}\cdot\nabla\mathbf{u} + (\nabla\cdot\mathbf{u})\mathbf{a})_{\text{2-form}}, \tag{2.133}$$

$$\underset{\text{3-form}}{\mathscr{L}_{\mathbf{u}}}\,(a)_{\text{3-form}} = (\mathbf{u}\cdot\nabla a + (\nabla\cdot\mathbf{u})a)_{\text{3-form}}. \tag{2.134}$$

$$\underset{\text{vec}}{\mathscr{L}_{\mathbf{u}}}(\mathbf{a})_{\text{vec}} = (\mathbf{u}\cdot\nabla\mathbf{a} - \mathbf{a}\cdot\nabla\mathbf{u})_{\text{vec}}. \tag{2.135}$$

The traditional **material derivative** along a vector field $\mathbf{u}$ defined in most existing textbook is given by

$$\tfrac{D}{Dt} \coloneqq \tfrac{\partial}{\partial t} + \mathbf{u}\cdot\nabla. \tag{2.136}$$

The Lie material derivative, which measures the real rate of change of various forms, deviates from the material derivative by additional terms:

$$(\tfrac{\partial}{\partial t} + \underset{\text{0-form}}{\mathscr{L}_{\mathbf{u}}})(a)_{\text{0-form}} = (\tfrac{D}{Dt}a)_{\text{0-form}}, \tag{2.137}$$

$$(\tfrac{\partial}{\partial t} + \underset{\text{1-form}}{\mathscr{L}_{\mathbf{u}}})(\mathbf{a})_{\text{1-form}} = (\tfrac{D}{Dt}\mathbf{a} + (\nabla\mathbf{u})^\intercal\mathbf{a})_{\text{1-form}}, \tag{2.138}$$

$$(\tfrac{\partial}{\partial t} + \underset{\text{2-form}}{\mathscr{L}_{\mathbf{u}}})(\mathbf{a})_{\text{2-form}} = (\tfrac{D}{Dt}\mathbf{a} - \mathbf{a}\cdot\nabla\mathbf{u} + (\nabla\cdot\mathbf{u})\mathbf{a})_{\text{2-form}}, \tag{2.139}$$

$$(\tfrac{\partial}{\partial t} + \underset{\text{3-form}}{\mathscr{L}_{\mathbf{u}}})(a)_{\text{3-form}} = (\tfrac{D}{Dt}a + (\nabla\cdot\mathbf{u})a)_{\text{3-form}}, \tag{2.140}$$

$$(\tfrac{\partial}{\partial t} + \underset{\text{vec}}{\mathscr{L}_{\mathbf{u}}})(\mathbf{a})_{\text{vec}} = (\tfrac{D}{Dt}\mathbf{a} - \mathbf{a}\cdot\nabla\mathbf{u})_{\text{vec}}. \tag{2.141}$$

Now, let us show Theorem 2.37.

Proof of (2.131). Eq. (2.131) follows directly from Theorem 2.30. □

Proof of (2.132). Using (2.103), we write $(\mathbf{a})_{\text{1-form}} = \vec{a}^\flat = \langle \vec{a}, I \rangle$, where $\vec{a} = (\mathbf{a})_{\text{vec}}$ and I is the identity vector-valued 1-form (Definition 2.16). For clarity we also write $\vec{u} = (\mathbf{u})_{\text{vec}}$. Now, $\mathscr{L}_{\mathbf{u}}(\mathbf{a})_{\text{1-form}} = \mathscr{L}_{\vec{u}}\langle \vec{a}, I\rangle \overset{(2.124)}{=} di_{\vec{u}}\langle \vec{a}, I\rangle + i_{\vec{u}}d\langle \vec{a}, I\rangle \overset{(2.96)}{=}$ $d\langle \vec{a}, \vec{u}\rangle + i_{\vec{u}}\langle \nabla \vec{a} \wedge I\rangle + i_{\vec{u}}\langle \vec{a}, d^\nabla I\rangle \overset{\text{Thm. 2.27}}{\overset{(2.96),(2.26)}{=}} \langle \nabla \vec{a}, \vec{u}\rangle + \langle \vec{a}, \nabla \vec{u}\rangle + \langle \nabla_{\vec{u}}\vec{a}, I\rangle - \langle \nabla \vec{a}, \vec{u}\rangle =$ $\langle \vec{a}, \nabla \vec{u}\rangle + \langle \nabla_{\vec{u}}\vec{a}, I\rangle = ((\nabla \mathbf{u})^\intercal \mathbf{a} + \nabla_{\mathbf{u}}\mathbf{a})_{\text{1-form}}$. □

Proof of (2.135). Eq. (2.135) follows from Theorem 2.31 and (2.97). □

Proof of (2.134). Let $\mu = (1)_{\text{3-form}}$ be the 3D volume form. Then $i_{\mathbf{u}}\mu = (\mathbf{u})_{\text{2-form}}$ and $\mathscr{L}_{\mathbf{u}}\,\mu \overset{(2.124)}{=} di_{\mathbf{u}}\mu = d(\mathbf{u})_{\text{2-form}} = (\nabla \cdot \mathbf{u})_{\text{3-form}}$. Now, $\mathscr{L}_{\mathbf{u}}(a)_{\text{3-form}} = \mathscr{L}_{\mathbf{u}}((a)_{\text{0-form}}\mu)$ $= \mathscr{L}_{\mathbf{u}}(a)_{\text{0-form}}\mu + (a)_{\text{0-form}}\,\mathscr{L}_{\mathbf{u}}\,\mu = (\nabla_{\mathbf{u}}a + (\nabla \cdot \mathbf{u})a)_{\text{3-form}}$. □

Proof of (2.133). Writing $(\mathbf{a})_{\text{2-form}} = i_{\vec{a}}\mu$ where $\vec{a} = (\mathbf{a})_{\text{vec}}$ and μ is the 3D volume form, we have $\mathscr{L}_{\vec{u}}(\mathbf{a})_{\text{2-form}} = \mathscr{L}_{\vec{u}}(i_{\vec{a}}\mu) \overset{(2.129)}{=} i_{[\vec{u},\vec{a}]}\mu + i_{\vec{a}}(\mathscr{L}_{\vec{u}}\,\mu) \overset{(2.134)}{=} i_{[\vec{u},\vec{a}]}\mu + (\nabla \cdot$ $\mathbf{u})i_{\vec{a}}\mu \overset{(2.135)}{=} (\nabla_{\mathbf{u}}\mathbf{a} - \nabla_{\mathbf{a}}\mathbf{u} + (\nabla \cdot \mathbf{u})\mathbf{a})_{\text{2-form}}$. □

In fact, during our proof of (2.133), we only used the fact that a 2-form is the contraction of a vector and a 3-form, each of which has an established Lie derivatives (2.134) and (2.135). What we may also explore is to expand the left-hand side of (2.133) using Cartan's formula (2.124):

$$\mathscr{L}_{\mathbf{u}}(\mathbf{a})_{\text{2-form}} = di_{\mathbf{u}}(\mathbf{a})_{\text{2-form}} + i_{\mathbf{u}}d(\mathbf{a})_{\text{2-form}} \tag{2.142}$$

$$= (\nabla \times (\mathbf{a} \times \mathbf{u}) + (\nabla \cdot \mathbf{a})\mathbf{u})_{\text{2-form}}. \tag{2.143}$$

Combining this result with (2.133) yields the following vector calculus identity

Theorem 2.38 — Curl of cross product.

$$\nabla \times (\mathbf{a} \times \mathbf{u}) = \mathbf{u} \cdot \nabla \mathbf{a} - \mathbf{a} \cdot \nabla \mathbf{u} + (\nabla \cdot \mathbf{u})\mathbf{a} - (\nabla \cdot \mathbf{a})\mathbf{u}. \tag{2.144}$$

Similarly, combining Cartan's magic formula for 1-forms with (2.132) gives us the following vector calculus identity

Theorem 2.39 — Cross product of curl.

$$(\nabla \times \mathbf{a}) \times \mathbf{u} = \mathbf{u} \cdot \nabla \mathbf{a} + \nabla(\mathbf{a} \cdot \mathbf{u}) - (\nabla \mathbf{u})^\intercal \mathbf{a}. \tag{2.145}$$

■

■ **Example 2.22 — Reynolds' transport theorems.** Consider (2.125) with $S = \Omega \subset \mathbb{R}^n$ being an n-dimensional region, and $\alpha = (f\mu)$ being an n-form, where f is some function and μ is the volume form in $\mathbb{R}^n$. Let $\Omega_t = \phi_t(\Omega)$ denote a moving domain, and let $\mathbf{u}$ be the velocity of the moving boundary; more precisely,

$\mathbf{u}$ is the velocity field written in coordinates so that $\dot{\phi}_t = (\mathbf{u})_{\text{vec}} \circ \phi_t$. Then, by Cartan's magic formula and Stokes' theorem, $\int_\Omega \underbrace{\mathscr{L}_{\vec{u}}\,(f\mu)}_{n\text{-form}} = \int_\Omega di_{\vec{u}}(f\mu) = \oint_{\partial\Omega} f(\mathbf{u})_{(n-1)\text{-form}}$. Hence (2.125) can be written as the following vector calculus identity:

Theorem 2.40 — Reynolds' transport theorem. Let Ω_t be a domain moving with a velocity $\mathbf{u}$. Then

$$\frac{d}{dt}\int_{\Omega_t} f\,dV = \int_{\Omega_t} \frac{\partial f}{\partial t} dV + \oint_{\partial\Omega_t} f\mathbf{u}\cdot\mathbf{n}dA. \tag{2.146}$$

The special case of the Reynolds' transport theorem in $n = 1$ is:

Theorem 2.41 — Leibniz's integral rule. In 1D calculus,

$$\frac{d}{dt}\int_{a(t)}^{b(t)} f(t,x)dx = \int_{a(t)}^{b(t)} \frac{\partial f(t,x)}{\partial t}dx + b'(t)f(t,b(t)) - a'(t)f(t,a(t)). \tag{2.147}$$

Specializing (2.125) to the case of a surface in the 3D Cartesian space yields

Theorem 2.42 Let S_t be a surface in 3D moving at a velocity $\mathbf{u}$. Let $\mathbf{f}$ be a (time-dependent) vector field defined on the 3D space. Then,

$$\frac{d}{dt}\iint_{S_t} \mathbf{f}\cdot\mathbf{n}dA = \iint_{S_t}\left(\frac{\partial \mathbf{f}}{\partial t} + (\nabla\cdot\mathbf{f})\mathbf{u}\right)\cdot\mathbf{n}dA + \oint_{\partial S_t}(\mathbf{f}\times\mathbf{u})\cdot d\mathbf{l}. \tag{2.148}$$

An alternative formula (using (2.133)) is

$$\frac{d}{dt}\iint_{S_t} \mathbf{f}\cdot\mathbf{n}dA = \iint_{S_t}\left(\frac{\partial \mathbf{f}}{\partial t} + \mathbf{u}\cdot\nabla\mathbf{f} - \mathbf{f}\cdot\nabla\mathbf{u} + (\nabla\cdot\mathbf{u})\mathbf{f}\right)\cdot\mathbf{n}dA \tag{2.149}$$

Specializing (2.125) to the case of a curve in the 3D Cartesian space yields

Theorem 2.43 Let C_t be a curve in 3D moving at a velocity $\mathbf{u}$. Let $\mathbf{f}$ be a (time-dependent) vector field defined on the 3D space. Then,

$$\frac{d}{dt}\int_{C_t} \mathbf{f}\cdot d\mathbf{l} = \int_{C_t}\left(\frac{\partial \mathbf{f}}{\partial t} + (\nabla\times\mathbf{f})\times\mathbf{u}\right)\cdot d\mathbf{l} + (\mathbf{f}\cdot\mathbf{u})|_{\partial C_t}. \tag{2.150}$$

Alternatively, using (2.132), we also have

$$\frac{d}{dt}\int_{C_t} \mathbf{f}\cdot d\mathbf{l} = \int_{C_t}\left(\frac{\partial \mathbf{f}}{\partial t} + \mathbf{u}\cdot\nabla\mathbf{f} + (\nabla\mathbf{u})^\intercal\mathbf{f}\right)\cdot d\mathbf{l}. \tag{2.151}$$

■

2.7 Pairings, Inner Products and Hodge Stars

In this last section of the chapter, we introduce the **Hodge star**. The previous operators we have introduced ($\wedge$, i_X, d, Lie brackets, *etc.*) are built with canonical

construction without depending on metric. In particular, they all have simple interplay with the pullback operator. The Hodge star that we are studying here is *metric dependent*. In particular, it does not have simple interplay with pullback or other operators.

The Hodge star is defined with the notion of an L^2 inner product. Let us first clarify dual pairing and inner product.

Theorem 2.44 — Metric-independent pairing. The dual space of k-forms is isomorphic to $(n-k)$-forms.

$$\big(\Omega^k(M)\big)^* \cong \Omega^{n-k}(M). \tag{2.152}$$

This is done through the following dual pairing

$$\begin{aligned} \Omega^k(M) \times \Omega^{n-k}(M) &\xrightarrow{\text{bilinear}} \mathbb{R} \\ (\alpha, \beta) &\mapsto \int_M \alpha \wedge \beta. \end{aligned} \tag{2.153}$$

■ **Example 2.23** In the 3D Cartesian space, $\int_M (\mathbf{a})_{\text{1-form}} \wedge (\mathbf{b})_{\text{2-form}} = \iiint_M \mathbf{a} \cdot \mathbf{b}\, dV$.
■

Definition 2.20 — Inner product between forms. Suppose $\flat$ is an inner product structure on a vector space V. This inner product structure defines a natural **volume form**

$$\mu \in \wedge^n V^*, \quad \mu[\![X_1, \dots, X_n]\!] = 1 \text{ if } X_1, \dots, X_n \text{ are positively orthonormal.} \tag{2.154}$$

The metric also defines a natural inner product on k-forms: for $\alpha, \beta \in \wedge^k V^*$

$$\langle \alpha, \beta \rangle \coloneqq \sum_{1 \le i_1 < \cdots < i_k \le n} \alpha[\![X_{i_1}, \dots, X_{i_k}]\!]\, \beta[\![X_{i_1}, \dots, X_{i_k}]\!] \tag{2.155}$$

using any orthonormal basis $X_1, \dots, X_n$.

Definition 2.21 — Hodge star. Suppose $\flat$ is an inner product structure on a vector space V. Let μ be the associated volume form. Define

$$\star_k \colon \wedge^k V^* \xrightarrow{\text{linear}} \wedge^{n-k} V^* \tag{2.156}$$

such that

$$\alpha \wedge \star\beta = \langle \alpha, \beta \rangle \mu. \tag{2.157}$$

Definition 2.22 — L^2 inner product. Suppose $\flat$ is an inner product structure on

T_pM for each $p \in M$. Define the L^2 **inner product** for $\Omega^k(M)$ as

$$\langle\!\langle \alpha, \beta \rangle\!\rangle := \int_M \langle \alpha, \beta \rangle \, \mu = \int_M \alpha \wedge \star\beta. \tag{2.158}$$

■ **Example 2.24** In the 3D Cartesian space

$$\star\,(a)_{\text{0-form}} = (a)_{\text{3-form}}, \tag{2.159}$$
$$\star\,(\mathbf{a})_{\text{1-form}} = (\mathbf{a})_{\text{2-form}}, \tag{2.160}$$
$$\star\,(\mathbf{a})_{\text{2-form}} = (\mathbf{a})_{\text{1-form}}, \tag{2.161}$$
$$\star\,(a)_{\text{3-form}} = (a)_{\text{0-form}}. \tag{2.162}$$

■

Theorem 2.45 The Hodge star gives a duality relation between the wedge product and the interior product

$$i_X \star_k \alpha = (-1)^k \star_{k+1} (X^\flat \wedge \alpha). \tag{2.163}$$

That is,

$$\begin{array}{ccccccccccc}
\wedge^0 V^* & \xleftarrow{i_{\mathbf{v}}} & \wedge^1 V^* & \xleftarrow{i_{\mathbf{v}}} & \wedge^2 V^* & \xleftarrow{i_{\mathbf{v}}} & \wedge^3 V^* & \longleftarrow\cdots & \wedge^{n-1} V^* & \xleftarrow{i_{\mathbf{v}}} & \wedge^n V^* \\
\downarrow{\scriptstyle \star_0} & \uparrow{\scriptstyle \mathbf{v}^\flat\wedge} & \downarrow{\scriptstyle \star_1} & \uparrow{\scriptstyle \mathbf{v}^\flat\wedge} & \downarrow{\scriptstyle \star_2} & \uparrow{\scriptstyle \mathbf{v}^\flat\wedge} & \downarrow{\scriptstyle \star_3} & & \downarrow{\scriptstyle \star_{n-1}} & \uparrow{\scriptstyle \mathbf{v}^\flat\wedge} & \downarrow{\scriptstyle \star_n} \\
\wedge^n V^* & \xrightarrow[i_{\mathbf{v}}]{} & \wedge^{n-1} V^* & \xrightarrow[-i_{\mathbf{v}}]{} & \wedge^{n-2} V^* & \xrightarrow[i_{\mathbf{v}}]{} & \wedge^{n-3} V^* & \longrightarrow\cdots & \wedge^1 V^* & \xrightarrow[(-1)^{n-1} i_{\mathbf{v}}]{} & \wedge^0 V^*
\end{array} \tag{2.164}$$

We conclude our exterior calculus chapter here. There are more objects one can build with the Hodge star that we have not covered, such as the codifferential $\delta = (-1)^k \star^{-1} d\star$ and Hodge Laplacian $\Delta = d\delta + \delta d$.

Technically, exterior calculus with metric (*i.e.* with Hodge star mixed in) makes it the so-called **geometric calculus**. In particular, one can replace the wedge product with the *geometric product* formed by the geometric algebra (Clifford algebra) of the given metric. One obtains the Dirac operator $d + \delta$ as the natural generalization of d to geometric algebra. In this sense, just as exterior algebra is the geometric algebra with 0 metric, the exterior calculus is the metric-independent geometric calculus. That is why it reveals the important structures about the differential topological aspect of multivariable calculus such as pullback, Lie derivative.

II Graphics Topics in Exterior Calculus

3. Geometric Optimizations

In this chapter, we take a new, geometric look at the study of optimizations. This includes optimization problems where the optimal solution in search is itself a geometry (Section 3.1) and the geometric picture of the computation of optimization problems (Section 3.2 and Section 3.3).

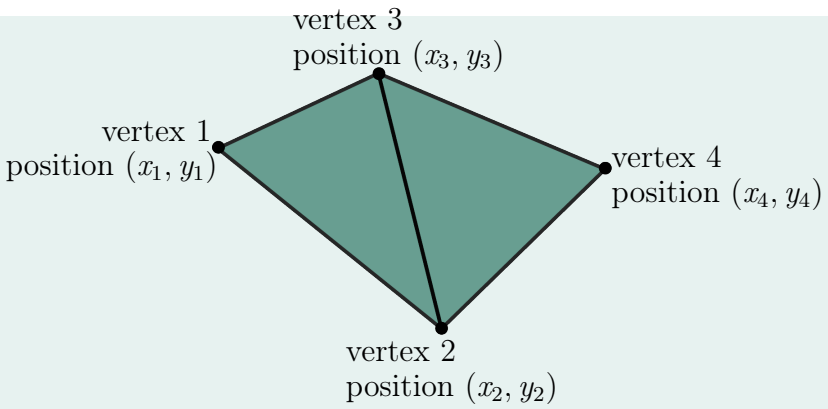
3.1 Geometry representations

In graphics, geometry can refer to a point distribution (*a point cloud*), a curve, a surface, or a volumetric object. They each have co-dimension 3, 2, 1, and 0, and they correspond to the *domains* on which 0-form, 1-form, 2-form, 3-forms are integrated.

Notation 3.1 Throughout this chapter, we use M to denote the **material space** and W to denote the **world space**. M is the *abstract* space of the manifold, like the index of particles of a point cloud, or the combinatorial structure of a mesh. W is the world space where the manifold is *embedded* and in most case, just the 2- or 3-dimensional Euclidean space. We always denotes the dimension of M as m and the dimension of W as n.

3.1.1 Lagrangian representations: manifold embedding

Definition 3.1 An **embedding** is an injective smooth map $f: M \to W$ that preserves the topological and the manifold structure. In the context of computer graphics, this corresponds *the positions of the vertices* of a mesh / polycurve, while the manifold M itself corresponds to the *combinatorial structure* of the mesh / polycurve.



Manifold M

Triangle id	Vertices
1	$(1, 2, 3)$
2	$(3, 2, 4)$

Embedding $f: M \to W$

Vertex id	Position
1	(x_1, y_1)
2	(x_2, y_2)
3	(x_3, y_3)
4	(x_4, y_4)

■ **Example 3.1 — Function graph.** Given a function graph $z = g(x, y)$, we can view it as an embedding

$$\begin{aligned} f: M &\to W \\ (x, y) &\mapsto (x, y, f(x, y)). \end{aligned}$$

Here the material space M is the 2-dimensional parameter space of x and y, and W is the 3-dimensional world space of x, y, and z. ■

Many classical geometric properties like *curvature* and *second fundamental form* are defined using the embedding of a manifold. We give a few examples of them below.

Definition 3.2 — Differential geometry of curves by embedding function. Let M be an 1-dimensional manifold and $f: M \to W$ is an embedding of M into $W = \mathbb{R}^3$. Let $p \in M$ be a point. Locally (in a neighborhood around p), we can parametrize M as an interval (a, b) of the real line, and use the Cartesian coordinates x, y, z for $\mathbb{R}^3$ and write the embedding $f: (a, b) \to \mathbb{R}^3$ as $f(t) = (x(t), y(t), z(t))$ for $t \in (a, b)$. The **unit tangent vector** is given as

$$\mathbf{T}(t) = \frac{(x'(t), y'(t), z'(t))}{|(x'(t), y'(t), z'(t))|_{\ell^2}}, \tag{3.1}$$

where $|(x, y, z)|_{\ell^2} = \sqrt{x^2 + y^2 + z^2}$ denotes the Euclidean norm. An intuitive observation tells us that the faster the tangent direction changes, the *curvier* the curve is. Notice that the curviness property only depends on the tangent *direction*, not the derivative of the embedding function. We hence define the **curvature** as

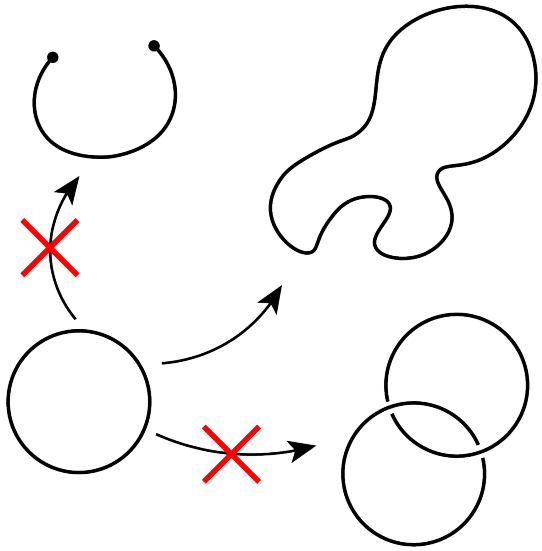


Figure 3.1 An abstract 1-dimensional manifold M (bottom left) that resembles a circle can be embedded to any wiggly shape (top right) but we can never embed it to something that has an open endpoints (top left) or something knotted (bottom right).

the rate of change of the unit tangent:

$$\kappa(t) = \left|\dot{\mathbf{T}}\right|_{\ell^2}. \tag{3.2}$$

R It is crucial that the readers see that the above definition of unit tangent and curvature are both *embedding dependent*! Without an embedding function $f\colon M \to W$, the manifold still has its own (albeit abstract) tangent space $T_pM \cong \mathbb{R}$. An embedding $f\colon M \to W$ can convert the abstract tangent to an $\mathbb{R}^3$ vector by the push-forward map

$$\begin{aligned} df\colon T_pM \cong \mathbb{R} &\xrightarrow{\text{linear}} T_{f(p)}W \cong \mathbb{R}^3 \\ 1 &\longmapsto \begin{bmatrix} x'(t) \\ y'(t) \\ z'(t) \end{bmatrix}. \end{aligned} \tag{3.3}$$

R Another crucial insight is that the embedding function $f\colon M \to W$ has to be *smooth*. The smoothness requirement means that the embedding must preserve the manifold's **topology**, including *topological features* like genus, knottedness, or open endpoints. See Figure 3.1 and Figure 3.2. This is why when a simulation artist uses meshed surfaces to model fracturing material (*e.g.* tearing a piece of paper), their simulation pipeline cannot simply modify the vertex positions but it sometimes have to modify the combinatorial structure (often called *remeshing*) when the surface go through *topological changes*!

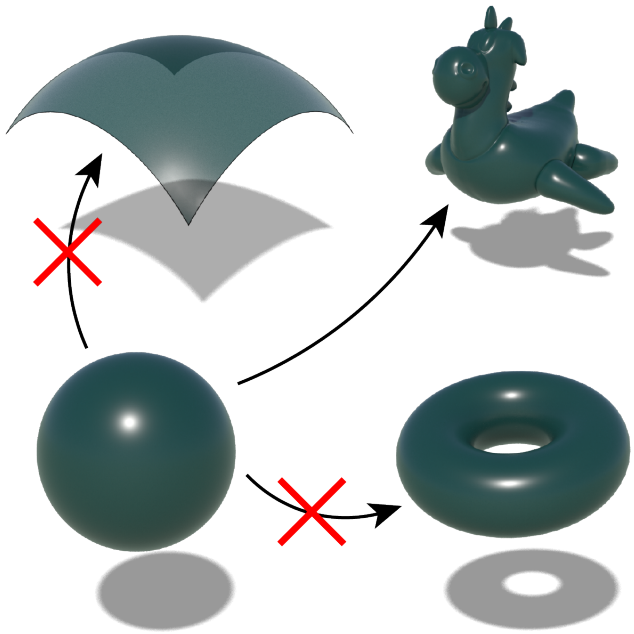


Figure 3.2 An abstract 2-dimensional manifold M (bottom left) that resembles a sphere can be embedded to any wiggly shape (top right) but we can never embed it to something that has boundary (top left) or something with a nonzero genus (bottom right).

3.1.2 Integration pairing

Now that we've learned about the distinction between an abstract manifold M and their embedding $f\colon M \to W$ into the world space, we can dive into the relation between an embedded manifold $M \overset{f}{\hookrightarrow} W$ and differential forms in the world space $\Omega^m(W)$.

Definition 3.3 We define the **integration pairing** between an embedded k-dimensional manifold $M \overset{f}{\hookrightarrow} W$ [a] and a k-form $\alpha \in \Omega^k(W)$ as the k-dimensional integral:

$$\langle M|\alpha\rangle \coloneqq \int_M \alpha. \tag{3.4}$$

The integration pairing corresponds to function evaluation ($k = 0$), oriented line integral ($k = 1$), oriented surface integral ($k = 2$), and the usual volumetric integral ($k = 3$).

[a] It is possible to consider a broader range of integration pairings: as long as the differential forms are smooth, we can integrate them not only on smoothly embedded manifolds, but also on rectifiable sets, which are patches of smoothly embedded manifolds joined together in a merely continuous sense.

Theorem 3.1 The integration pairing is bilinear, meaning that we can take linear

combination in both the k-dimensional embedded manifolds and the k-forms: for $a_1, a_2 \in \mathbb{R}$ two real numbers, $M_1, M_2 \subset W$ two embedded k-dimensional manifolds, $\alpha_1, \alpha_2 \in \Omega^k(W)$ two k-forms,

$$\langle a_1 M_1 + a_2 M_2 | \alpha \rangle = a_1 \int_{M_1} \alpha + a_2 \int_{M_2} \alpha = a_1 \langle M_1 | \alpha \rangle + a_2 \langle M_2 | \alpha \rangle, \tag{3.5}$$

$$\langle M | a_1 \alpha_1 + a_2 \alpha_2 \rangle = a_1 \int_{M} \alpha_1 + a_2 \int_{M} \alpha_2 = a_1 \langle M | \alpha_1 \rangle + a_2 \langle M | \alpha_2 \rangle. \tag{3.6}$$

Theorem 3.2 By Stokes' Theorem (Theorem 2.12), the exterior derivative d of differential forms and the operation of taking boundary ∂ of the integration domain are **adjoint** of each other under the integration pairing:

$$\langle S | d\alpha \rangle = \int_S d\alpha \overset{\text{Stokes}}{=} \int_{\partial S} \alpha = \langle \partial S | \alpha \rangle. \tag{3.7}$$

3.1.3 Exterior calculus on combinatorial surfaces and curves (DEC)

Definition 3.4 A **tessellated (meshed)** surface is a tuple (V, E, F) of vertices V, edges E, and faces F. A vertex has dimension 0, so we called it a *0-cell*. We call an edge a *1-cell* and a face a *2-cell*. Together, we call (V, E, F) a *cell complex*. The **boundary operator** ∂ is defined by the following:

- Since, for example, a face can have multiple edges that form its boundary, the boundary operator needs to map a face to a combination of multiple edges.
- We denote the set of all weighted sum of faces as C_F, weighted sum of edges as C_E, and weighted sum of vertices as C_V. The orientation of an edge / face corresponds to the sign of the weights $e_{ij} = -e_{ji}, f_{ijk} = f_{jki} = f_{kij} = -f_{jik} = -f_{ikj} = -f_{kji}$. For reference, we referred to a weighted sum of cells as **a chain**.
- We then define the boundary operator as the linear maps ∂_1, ∂_2

$$C_V \xleftarrow{\partial_1} C_E \xleftarrow{\partial_2} C_F \tag{3.8}$$

  such that if e_{ij}, e_{jk}, e_{ki} are the three edges forming the boundary of a triangular face f_{ijk}, then $\partial_2(f_{ijk}) = e_{ij} + e_{jk} + e_{ki}$, and if v_i, v_j are the two endpoints forming the boundary of an edge e_{ij}, then $\partial_1(e_{ij}) = -v_i + v_j$. Readers can check that $\partial_1 \circ \partial_2 = 0$ as the boundary of a face is a loop, which has no endpoints.

Definition 3.5 With the above example in mind, we can define a generalized m-dimensional **cell complex** as $M = (M_0, \cdots, M_m)$ together with a boundary

operator and the chain complex:

$$C_0 \xleftarrow{\partial_1} C_1 \xleftarrow{\partial_2} \cdots \xleftarrow{\partial_m} C_m. \tag{3.9}$$

This definition should cover poly-lines (1-dimensional cell complex), tesselated (meshed) surfaces (2-dimensional cell complex), and volumetric meshes (*e.g.* tetrahedral mesh) (3-dimensional cell complex).

Since each C_k is a finite dimensional vector space, its dual space $C^k \coloneqq (C_k)^*$ is also finite dimensional, too. Just like the dual of vectors is called a covector (linear maps from vectors to real numbers), the dual of chains is called a **cochain** (linear maps from chains to real numbers). In Theorem 3.2 we saw that the adjoint of boundary operator ∂ is the exterior derivative d, thus gives rise to the **cochain complex**, similar to (3.9):

$$C^0 \xrightarrow{\partial_1^* = d_0} C^1 \xrightarrow{\partial_2^* = d_1} \cdots \xrightarrow{\partial_m^* = d_{m-1}} C^m. \tag{3.10}$$

The k-cells $M_k = \{c_1, \cdots, c_N\}$ automatically form a basis of the space of weighted sums of them, *a.k.a.* k-chains C_k. Each cell $c \in M_k$ corresponds to a linear functional on k-cochains by the integration pairing:

$$c \colon \alpha \in \Omega^k(M) \mapsto \int_c \alpha. \tag{3.11}$$

We can therefore view the space of cochains C^k as differential k-forms up to minor difference that cannot be distinguished by integration on k-cells. More precisely, suppose $\alpha, \beta \in \Omega^k(M)$ are two differential k-forms such that their integral values are the same on all k-cells,

$$\int_c \alpha = \int_c \beta, \text{ for all } c \in C_k. \tag{3.12}$$

For the purpose of any computation on the cell complex, α and β are virtually indistinguishable. As a result, they can be represented as the same N real numbers:

$$\begin{bmatrix} \int_{c_1} \beta \\ \int_{c_2} \beta \\ \vdots \\ \int_{c_N} \beta \end{bmatrix} = \begin{bmatrix} \int_{c_1} \alpha \\ \int_{c_2} \alpha \\ \vdots \\ \int_{c_N} \alpha \end{bmatrix} \in \mathbb{R}^N. \tag{3.13}$$

In other words, α and β belongs to the same equivalence class, and the space of cochains C^k is exactly the collection of these equivalence classes. We can further derive an entire theory about the exterior calculus of these "differential forms up to minor differences that cannot be distinguished by the cell complex" (equivalence classes). Another SIGGRAPH course [Crane(2018)] already covered a wide range of exterior calculus operations on discrete surfaces and curves and we omit the discussions here.

3.1.4 Eulerian representations: geometric measure theory

Eulerian representations, contrary to Lagrangian ones, seeks to represent geometries contained in the world space W as functionals of differential forms.

Recall that a Dirac-δ *measure* ρ_p of a point $p \in W$ is an n-form (sometimes also called a *measure*) where for any (volumetric) subset $A \subset W$,

$$\int_A \rho_p = \begin{cases} 1, & \text{if } p \in A \\ 0, & \text{if } p \notin A. \end{cases} \tag{3.14}$$

The Dirac-δ n-form ρ_p induces a linear functional T via dual pairing:

$$\begin{aligned} T\colon \Omega^0(W) &\xrightarrow{\text{linear}} \mathbb{R} \\ f &\longmapsto \int_W f\rho_p. \end{aligned} \tag{3.15}$$

Readers can verify that this functional T is linear in its input $f \in \Omega^0(W)$ by verifying that $T(af + bg) = aT(f) + bT(g)$ for $a, b \in \mathbb{R}$ and $f, g \in \Omega^0(W)$.

We can approximate the Dirac-δ n-form by ρ_ϵ defined as

$$\int_A \rho_\epsilon = \frac{\text{Vol}(A \cap B_p(\epsilon))}{\text{Vol}(B_p(\epsilon))}. \tag{3.16}$$

Here $B_p(\epsilon)$ denotes the neighborhood ball centered at point p with radius $\epsilon > 0$. The induced linear functional on 0-forms is:

$$\begin{aligned} T_\epsilon\colon \Omega^0(W) &\xrightarrow{\text{linear}} \mathbb{R} \\ f &\longmapsto \frac{\int_{B_p(\epsilon)} f\mu}{\int_{B_p(\epsilon)} \mu}. \end{aligned} \tag{3.17}$$

For continuous functions f, the above corresponds to the *average* of f on the ϵ-ball $B_p(\epsilon)$ and it converges to the function value $f(p)$ as $\epsilon \to 0$.[1]

The intuition behind Dirac-δ n-form is that each point $p \in W$ corresponds to a linear functional T_p that maps a 0-form $f \in \Omega^0(W)$ to its evaluation $f(p)$ at the point p. A point p is of dimension 0. For geometries of other dimension, we can similarly define a linear function on differential forms.

Definition 3.6 — Currents / Dirac-δ forms. One way to to represent a k-dimensional geometry $S \subset W$ is by the linear functional T_S that maps differential k-forms $\Omega^k(W)$ to their integral on S:

$$\begin{aligned} T_S\colon \Omega^k(W) &\xrightarrow{\text{linear}} \mathbb{R} \\ \alpha &\xrightarrow{\text{linear}} \langle S|\alpha\rangle = \int_S \alpha. \end{aligned} \tag{3.18}$$

We call the functional T_S the **k-current associated with S**. In lieu of the dual pairing between $(n-k)$-forms and k-forms:

$$\alpha \in \Omega^k(W), \beta \in \Omega^{n-k}(W) \mapsto \int_W \alpha \wedge \beta, \tag{3.19}$$

[1] This is true for more general, integrable functions for almost all $p \in W$. Interested readers can read about Lebesgue differentiation theorem.

we can now view k-dimensional geometries as $(n-k)$-forms!

$$\{k\text{-dimensional geomeries}\} \cong \big(\Omega^k(W)\big)^* \cong \Omega^{n-k}(W). \tag{3.20}$$

We call the $(n-k)$-form $\delta_S \in \Omega^{n-k}(W)$ satisfying the following equation the **Dirac-δ form associated with S**:

$$\int_W \alpha \wedge \delta_S = \int_S \alpha. \tag{3.21}$$

We saw in Theorem 3.2 that the adjoint of the exterior derivative d is the boundary operator ∂. In the context of currents / Dirac-δ forms, we have a similar insight.

Theorem 3.3 Let $S \subset W$ be a k-dimensional geometry and $\delta_S \in \Omega^{n-k}(W)$ be the associated Dirac-δ form. If we take an exact k-form $d\alpha \in \Omega^k(W)$ for some smooth $\alpha \in \Omega^{k-1}(W)$ as the input of the dual pairing, then

$$\int_W d\alpha \wedge \delta_S = \int_S d\alpha \overset{\text{Stokes}}{=} \int_{\partial S} \alpha = \int_W \alpha \wedge \delta_{\partial S}. \tag{3.22}$$

We then apply Leibniz rule (2.59)

$$d\underbrace{(\alpha \wedge \delta_S)}_{(n-1)\text{-form}} = d\alpha \wedge \delta_S + \alpha \wedge (-1)^{(k-1)} d\delta_S, \tag{3.23}$$

Since $\int_W d(\alpha \wedge \delta_S) = \int_{\partial W} \alpha \wedge \delta_S = 0$[a], the left hand side of (3.22) turns into $\int_W \alpha \wedge (-1)^k d\delta_S$. Since the equation holds true for all $\alpha \in \Omega^{k-1}(W)$, we conclude that as an $(n-k+1)$-form,

$$d\delta_S = (-1)^k \delta_{\partial S}. \tag{3.24}$$

[a] In most cases, we can assume S is a bounded geometry, *i.e.* it doesn't intersect the boundary $\partial W$the world space W; therefore $\delta_S|_{\partial M} = 0$. We will need to use testing smooth forms $\alpha \in \Omega_c^{k-1}(M)$ that vanish at the boundary ∂W / has compact support if S is not as well-behaved.

R Recall that if we represented the geometry $S \subset W$ as a manifold with an embedding map, the process of taking the boundary ∂S would involve dealing with the combinatorial structure representing the abstract manifold. It is when we adopt the Eulerian view of geometry representations that we can use a simple *linear* operation d to take the boundary of a geometry.

■ **Example 3.2 — Surface area.** Let W be the 3-dimensional Euclidean space. Suppose $S \subset W$ is a 2-dimensional geometry (surface) with boundary ∂S. The surface area of S is the same as the surface integral of the unit normal vector field

$\mathbf{n}\colon S \to TW$

$$\text{Area}(S) = \int_S 1\, dA = \int_S \langle \mathbf{n}, \mathbf{n} \rangle, \tag{3.25}$$

where $\langle \cdot, \cdot \rangle$ denotes the Euclidean metric. At any given point $p \in S$, the unit normal vector $\mathbf{n}_p \in T_p W$ is the vector that yields the largest *flux*, *i.e.*

$$\langle \mathbf{n}_p, \mathbf{n}_p \rangle = \max_{\mathbf{v} \in T_p W \cong \mathbb{R}^3} \langle \mathbf{v}, \mathbf{n}_p \rangle. \tag{3.26}$$

Combining the above equations, we see that surface area is determined by the maximal flux amongst all unit vector fields

$$\text{Area}(S) = \max_{\mathbf{X}\colon W \to \mathbb{R}^3, |\mathbf{X}| \leq 1} \int_S \mathbf{X}^\flat. \tag{3.27}$$

Using the Hodge star and musical isomorphisms associated with Euclidean metric, we can associate any vector field $\mathbf{X}\colon W \to \mathbb{R}^3$ with a 2-form $\alpha \in \Omega^2(W)$ such that $\mathbf{X} = (\star\alpha)^\sharp$. The constraint that $\mathbf{X}$ is a unit vector field is equivalent to

$$\|\alpha\|_{\max} := \max_{p \in M} |\alpha_p| \leq 1. \tag{3.28}$$

We now write everything using the Dirac-δ 1-form δ_S representing S:

$$\begin{aligned} \text{Area}(S) &= \max_{\alpha \in \Omega^2(W), \|\alpha\|_{\max} \leq 1} \int_S \alpha && (3.29)\\ &= \max_{\alpha \in \Omega^2(W), \|\alpha\|_{\max} \leq 1} \int_W \alpha \wedge \delta_S && (3.30)\\ &=: \|\delta_S\|_{\text{mass}}. && (3.31) \end{aligned}$$

The last two lines and is the same as saying $\|\cdot\|_{\text{mass}}$ is the **dual norm** of $\|\cdot\|_{\max}$.

By combining the boundary operator from Theorem 3.3 and the above derivation, we can now solve the classical **Plateau problem** of finding surface S of minimal surface area subject to a fixed boundary curve $\Gamma \subset W$:

$$\begin{aligned} &\underset{S \subset W\text{: a surface}}{\text{minimize}} \quad \text{Area}(S) \\ &\text{subject to } \partial S = \Gamma. \end{aligned} \tag{3.32}$$

The equivalent problem written in terms of Dirac-δ forms look like

$$\begin{aligned} &\underset{\eta \in \Omega^1(W)\text{: a 1-form}}{\text{minimize}} \quad \|\eta\|_{\text{mass}} \\ &\text{subject to } d\eta = \delta_\Gamma. \end{aligned} \tag{3.33}$$

■

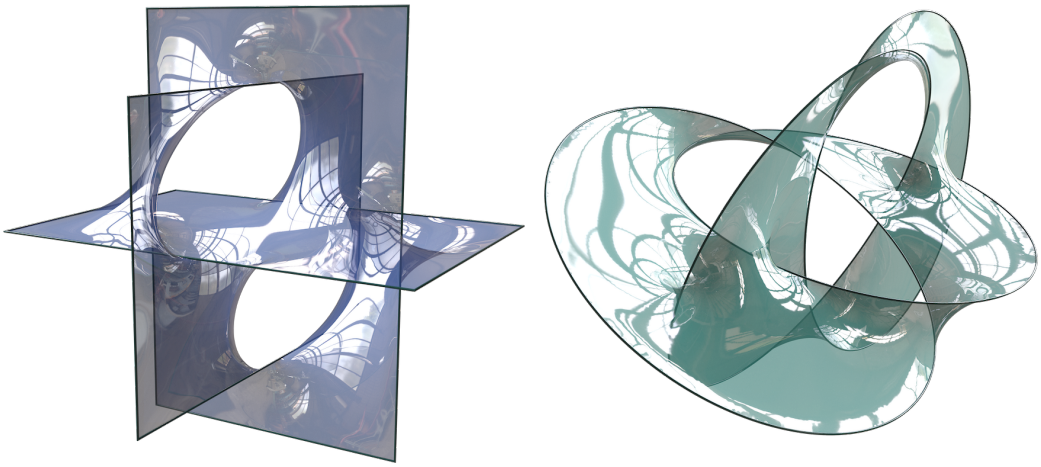


Figure 3.3 Minimal surfaces computed using Eulerian representation as described by Problem 3.33 [Wang and Chern(2021)]. Unlike Lagrangian representation that requires tricky combinatorial effort in computing surface boundary, problems formulated using Eulerian representations only require a linear equation when assigning the boundary!

3.2 A Geometric view of optimization problems

Optimization is a type of math problems commonly found in engineering and science. The most basic optimization problem was perhaps introduced in one's first Calculus class.

Problem 3.1 — Optimization on the real line. Given an **objective function** $f: \mathbb{R} \to \mathbb{R}$, find the **optimal input** $x^{\text{opt}} \in \mathbb{R}$ such that f attains the minimal value at x^{opt}, *i.e.*

$$f(x^{\text{opt}}) = \min_{x \in \mathbb{R}} f(x). \tag{3.34}$$

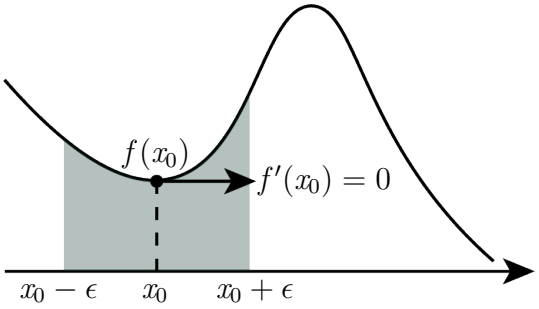
We learned in Calculus that for differentiable f, a necessary condition that an input x_0 is optimal is that the first derivative of f attains zero:

$$f'(x_0) = 0. \tag{3.35}$$

The above equation is referred to as the **first order optimality condition** for the optimization problem (3.34). If f is twice differentiable, then the first order optimality condition together with the inequality

$$f''(x_0) \geq 0 \tag{3.36}$$

ensures that x_0 is a **local minimum**, which means all nearby points $x \in (x_0 - \epsilon, x_0 + \epsilon)$ are less optimal than (or equally optimal to) x_0, *i.e.* $f(x) \leq f(x_0)$.



In this section, we explore the geometric picture of optimization problems: replacing real line with finite-dimensional manifolds or even function spaces, adding equality constraints, adding inequality constraints, the geometric meaning of Lagrange multipliers, and iterative numerical methods for solving optimization methods. We focus on deriving the necessary condition for an optimal solution, similar to (3.35), and leave the discussions of the sufficient condition for other textbooks, *e.g.* [Boyd and Vandenberghe(2004)].

3.2.1 Unconstrained problems on (finite-dimensional) manifolds

Problem 3.2 M is an n-dimensional manifold and $f\colon M \to \mathbb{R}$ is a function on M. An unconstrained optimization problem on the manifold M is about finding a point $p \in M$ such that $f(p)$ attains minimal value:

$$\underset{p \in M}{\text{minimize}}\, f(p). \tag{3.37}$$

The *locally optimal solution* $p \in M$ must satisfy that all its neighboring points are less optimal than (or equally optimal to) p. This property can be captured by the property that if $\gamma\colon [-\epsilon, \epsilon] \to M$ is a path in M passing through the point $p = \gamma(0)$, then the function $F = f \circ \gamma\colon [-\epsilon, \epsilon] \xrightarrow{\gamma} M \xrightarrow{f} \mathbb{R}$ takes a local minimum when $s = 0$ and $\gamma(s) = p$. Apply the first order optimality condition (3.35), we get

$$0 = F'(0) = \frac{d}{ds} f \circ \gamma(s)|_{s=0} = df_{\gamma(0)}[\![\dot{\gamma}(0)]\!] = df_p[\![v]\!], \tag{3.38}$$

where $v \in T_pM = \dot{\gamma}(0)$ is the tangent of the path γ at the point p. Since this has to hold true for all path γ at all tangent direction $v \in T_pM$, we conclude that the first order condition for a local optimal solution p is the same as the differential df of the objective function f is zero as a 1-form at p:

$$df_p = 0. \tag{3.39}$$

3.2.2 Unconstrained problems on functions and differential forms

We now take a look at unconstrained optimization problems on the space of functions and differential forms $\Omega^k(M)$. When $k = 0$, the study of optimization problems on the space of functions $\Omega^0(M)$ is often called **Calculus of Variations**. One nice property about differential forms is that they have a linear structure. To perturb an input $\alpha \in \Omega^k(M)$, one can simply pick another k-form $\mathring{\alpha} \in \Omega^k(M)$ and add it to α. In Figure 3.5, we visualize this concept for 0-forms *a.k.a.* functions.

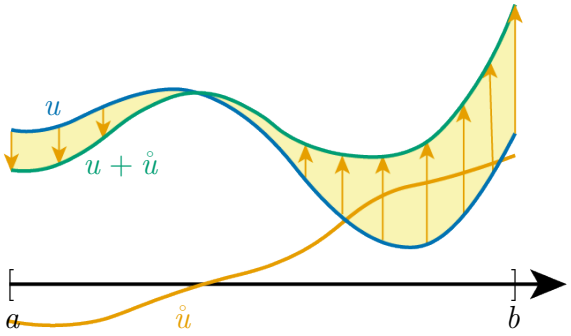


Figure 3.4 The function space $\mathcal{V} = \Omega^0([a, b])$ is a vector space. With any function $u\colon [a, b] \to \mathbb{R}$, we can always take another function $\mathring{u}\colon [a, b] \to \mathbb{R}$ and add it to u to get a modified function $u + \mathring{u}\colon [a, b] \to \mathbb{R}$. Effectively, the tangent space at u is the entire function space $T_u\mathcal{V} = \mathcal{V}$.

Problem 3.3 M is an n-dimensional manifold and $\mathcal{V} = \Omega^k(M)$ is the vector space of all k-forms defined on M. Let $\mathcal{E}\colon \mathcal{V} \to \mathbb{R}$ be a functional[2] defined on $\mathcal{V}$. We use the notation $\mathcal{E}$ because a common origin of optimization problems on function spaces is about minimizing an *energy functional*. An unconstrained optimization problem on k-forms $\mathcal{V} = \Omega^k(M)$ is about finding an optimal $\alpha \in \mathcal{V}$ such that $\mathcal{E}[\alpha]$ attains minimal value:

$$\underset{\alpha \in \mathcal{V}}{\text{minimize}}\, \mathcal{E}(\alpha). \tag{3.40}$$

The differential $d\mathcal{E}$ at an input $\alpha \in \mathcal{V}$ is a covector $d\mathcal{E}|_\alpha \in \mathcal{V}^*$. We can understand this infinite dimensional covector by pairing it with a tangent vector $\mathring{\alpha} \in T_\alpha\mathcal{V} = \mathcal{V}$:

$$d\mathcal{E}_\alpha[\![\mathring{\alpha}]\!] = \langle d\mathcal{E}_\alpha | \mathring{\alpha} \rangle \coloneqq \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \mathcal{E}(\alpha + \epsilon\mathring{\alpha}). \tag{3.41}$$

Each tangent vector $\mathring{\alpha} \in T_\alpha\mathcal{V} = \mathcal{V}$ acts on α by varying its value α_p to $\alpha_p + \mathring{\alpha}_p$ at any point $p \in M$. Hence we call the tangent vector $\mathring{\alpha}$ a **variation**. We use the notation of $\mathring{\alpha}$ to indicate that it is a variation in the object α. Likewise, we sometimes write

$$\mathring{\mathcal{E}} = \langle d\mathcal{E}_\alpha | \mathring{\alpha} \rangle \tag{3.42}$$

to denote the corresponding variation in $\mathcal{E}$.

The *locally optimal solution* $\alpha \in \mathcal{V}$ must satisfy that $\alpha + \epsilon\mathring{\alpha}$ is less optimal than (or equally optimal to) α for small $\epsilon > 0$. As a result, the function $F(s) = \mathcal{E}(\alpha + s\mathring{\alpha})$ must attain local minimum when $s = 0$ and $F(0) = \mathcal{E}(\alpha)$. By (3.35),

$$0 = F'(0) = \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \mathcal{E}(\alpha + \epsilon\mathring{\alpha}) = d\mathcal{E}_\alpha[\![\mathring{\alpha}]\!]. \tag{3.43}$$

Since this has to hold true for all variations $\mathring{\alpha} \in \mathcal{V}$, we conclude that the first order condition for a local optimal solution α is the same as the differential $d\mathcal{E}$ of the

[2] A *functional* is just another name for function with an emphasis that it may be a function over a space of functions.

energy functional $\mathcal{E}$ vanishes at α as a covector:

$$d\mathcal{E}|_\alpha = 0 \in \mathcal{V}^*. \tag{3.44}$$

■ **Example 3.3** Let $\mathcal{V} = \{u\colon [a,b] \to \mathbb{R}\} = \Omega^0([a,b])$. Define $\mathcal{E}\colon \mathcal{V} \to \mathbb{R}$ as $\mathcal{E}(u) = \int_a^b u^2(x)\,dx$. The differential $d\mathcal{E}|_u\colon \mathcal{V} \xrightarrow{\text{linear}} \mathbb{R}$ is

$$\begin{aligned} d\mathcal{E}_u[\![\mathring{u}]\!] &= \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \int_a^b (u(x) + \epsilon\mathring{u}(x))^2 dx && (3.45)\\ &= \left.\int_a^b 2(u(x) + \epsilon\mathring{u}(x))\mathring{u}(x)\,dx\right|_{\epsilon=0} && (3.46)\\ &= \int_a^b 2u(x)\mathring{u}(x)\,dx. && (3.47) \end{aligned}$$

■

■ **Example 3.4 — Dirichlet energy.** Let $\mathcal{V} = \Omega^0([a,b])$. Define $\mathcal{E}\colon \mathcal{V} \to \mathbb{R}$ as $\mathcal{E}(u) = \int_a^b \frac{1}{2}u'(x)^2\,dx$, which is often called the **Dirichlet energy**. We can use integration by parts to understand the differential $d\mathcal{E}|_u\colon \mathcal{V} \xrightarrow{\text{linear}} \mathbb{R}$,

$$\begin{aligned} d\mathcal{E}_u[\![\mathring{u}]\!] &= \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \int_a^b (u'(x) + \epsilon\mathring{u}'(x))^2 dx && (3.48)\\ &= \int_a^b u'(x)\mathring{u}'(x)\,dx && (3.49)\\ &\overset{\text{IbyP}}{=} u'(b)\mathring{u}(b) - u'(a)\mathring{u}(a) - \int_a^b u''(x)\mathring{u}(x)\,dx. && (3.50) \end{aligned}$$

The expression after the integration by parts appear more directly as an L^2 linear pairing (see Definition 2.22) with $\mathring{u}$ together with some finite dimensional dot product pairing at the boundary. ■

$d\mathcal{E}|_\alpha$ is the differential of $\mathcal{E}$ at α, which is a covector in $\mathcal{V}^* = \left(\Omega^k(M)\right)^* \cong \Omega^{n-k}(M)$ [3]. It is not to be confused with the gradient of the functional $\operatorname{grad}\mathcal{E}|_\alpha \in \mathcal{V} = \Omega^k(M)$. It is also not to be confused with the differential of the input variable $d\alpha \in \Omega^{k+1}(M)$. Let $\flat_\mathcal{V}\colon \mathcal{V} \to \mathcal{V}^*$, or "uncurried" as $\langle\!\langle \cdot,\cdot \rangle\!\rangle_\mathcal{V} = \flat_\mathcal{V}(\cdot)(\cdot)$, be some metric (inner product structure) defined on $\mathcal{V}$. Then we have a unique vector $\operatorname{grad}\mathcal{E}|_\alpha := d\mathcal{E}|_\alpha^{\sharp_\mathcal{V}}$ defined such that

$$\langle\!\langle \operatorname{grad}\mathcal{E}|_\alpha, \mathring{\alpha} \rangle\!\rangle = d\mathcal{E}|_\alpha[\![\mathring{\alpha}]\!] \quad \forall\, \mathring{\alpha} \in \mathcal{V}. \tag{3.51}$$

■ **Example 3.5** Let $\mathcal{V} = \{u\colon [a,b] \to \mathbb{R}\}$. Define $\mathcal{E}\colon \mathcal{V} \to \mathbb{R}$ as $\mathcal{E}(u) = \int_a^b u^2(x)\,dx$, whose variation is known as $d\mathcal{E}_u[\![\mathring{u}]\!] = \int_a^b 2u(x)\mathring{u}(x)\,dx$. If we define an inner

[3] See Theorem 2.44.

product structure by

$$\langle\!\langle u, v \rangle\!\rangle := \int_a^b u(x)v(x)m(x)\,dx \tag{3.52}$$

using some given positive function $m\colon [a,b] \to \mathbb{R}_{>0}$. Then the gradient of $\mathcal{E}$ is given by

$$\operatorname{grad} \mathcal{E}|_u(x) = \frac{2u(x)}{m(x)}. \tag{3.53}$$

■

The domain for $\mathcal{E}$ does not need to be a vector space. One typical extension is that $\mathcal{A}$ is an affine space parallel to $\mathcal{V}$. That is, $\mathcal{V}$ is the tangent space of $\mathcal{A}$ at any point $\alpha \in \mathcal{A}$. Then $d\mathcal{E}|_\alpha$ is a covector in $\mathcal{V}^*$, and $\operatorname{grad} \mathcal{E}|_\alpha$ is a vector in $\mathcal{V}$. The first order optimality condition for a minimization problem on an affine space $\mathcal{A}$ is the same: the differential $d\mathcal{E}|_\alpha$ is zero as a covector in $\mathcal{V}^*$.

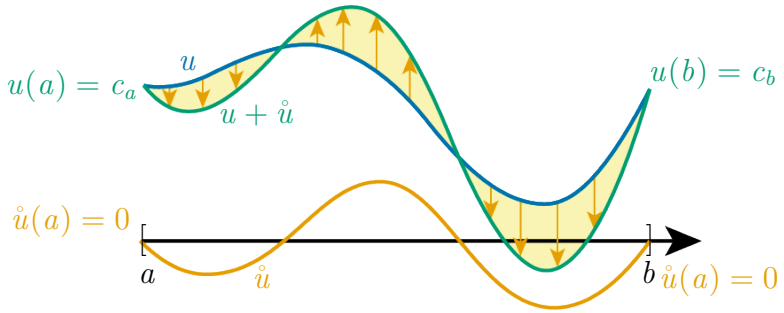


Figure 3.5 The function space with boundary conditions $\mathcal{A} = \{u \in \Omega^0([a,b]), u(a) = c_a, u(b) = c_b\}$ is an affine space. With any function $u \in \mathcal{A}$, we can always take another function that vanishes on the boundary $\mathring{u} \in \mathcal{V} = \{\mathring{u}\colon [a,b] \to \mathbb{R}, \mathring{u}(a) = \mathring{u}(b) = 0\}$ and add it to u to get a modified function $u + \mathring{u}\colon [a,b] \to \mathbb{R}$. Effectively, the tangent space at u is the space of functions with vanishing boundary values $T_u\mathcal{A} = \mathcal{V}$.

■ **Example 3.6 — Dirichlet energy with boundary conditions.** Let $\mathcal{A} = \{u\colon [a,b] \to \mathbb{R}, u(a) = c_a, u(b) = c_b\}$ for some constants c_a, c_b. Note that $\mathcal{A}$ is an affine space. That is, the difference between any two elements in $\mathcal{A}$ form a vector space $\mathcal{V} = \{\mathring{u}\colon [a,b] \to \mathbb{R}, \mathring{u}(a) = \mathring{u}(b) = 0\}$ which is closed under additions and scalar multiplications. Let $\mathcal{E}\colon \mathcal{A} \to \mathbb{R}$ be the Dirichlet energy $\mathcal{E}(u) = \int_a^b \frac{1}{2}u'(x)^2\,dx$ for $u \in \mathcal{V}$. (3.50) states that

$$d\mathcal{E}_u[\![\mathring{u}]\!] = -\int_a^b u''(x)\mathring{u}(x)\,dx. \tag{3.54}$$

Consider the $m(x)$-weighted inner product $\langle\!\langle \mathring{u}, \mathring{v} \rangle\!\rangle = \int_a^b \mathring{u}(x)\mathring{v}(x)m(x)\,dx$ for

$\mathring{u}, \mathring{v} \in \mathcal{V}$ and write

$$d\mathcal{E}_u[\![\mathring{u}]\!] = \int_a^b \left(\frac{-u''(x)}{m(x)}\right) \mathring{u}(x) m(x)\, dx = \left\langle\!\!\left\langle \underbrace{\frac{-u''}{m}}_{\operatorname{grad} \mathcal{E}_u}, \mathring{u} \right\rangle\!\!\right\rangle. \tag{3.55}$$

The gradient $\operatorname{grad} \mathcal{E}_u \in \mathcal{V}$ as a function is

$$\operatorname{grad} \mathcal{E}_u(x) = -\frac{u''(x)}{m(x)}, \quad x \in [a, b]. \tag{3.56}$$

■

■ **Example 3.7** Let $L\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $L = L(u_1, u_2)$ be some expression. Let its differential be denoted by

$$dL = \frac{\partial L}{\partial u_1}\, du_1 + \frac{\partial L}{\partial u_2}\, du_2. \tag{3.57}$$

Now, consider $\mathcal{A} = \{y\colon [a, b] \to \mathbb{R} \,|\, y(a) = y_a, y(b) = y_b\}$ for some given y_a, y_b. Note that $\mathcal{A}$ is an affine space with tangent vector space $\mathcal{V} = \{\mathring{y}\colon [a, b] \to \mathbb{R} \,|\, \mathring{y}(a) = \mathring{y}(b) = 0\}$. Let $\mathcal{E}\colon \mathcal{A} \to \mathbb{R}$,

$$\mathcal{E}(y) \coloneqq \int_a^b L\left(y(x), y'(x)\right)\, dx. \tag{3.58}$$

Then its variation is

$$d\mathcal{E}|_y[\![\mathring{y}]\!] = \int_a^b \left(\frac{\partial L}{\partial u_1}(y(x), y'(x))\mathring{y}(x) + \frac{\partial L}{\partial u_2}(y(x), y'(x))\mathring{y}'(x)\right)\, dx \tag{3.59}$$

$$= \int_a^b \left(\frac{\partial L}{\partial u_1}(y(x), y'(x)) - \frac{d}{dx}\left(\frac{\partial L}{\partial u_2}(y(x), y'(x))\right)\right) \mathring{y}(x)\, dx. \tag{3.60}$$

A common shorthand notation is

$$d\mathcal{E}|_y[\![\mathring{y}]\!] = \int_a^b \left(\frac{\partial L}{\partial y} - \frac{d}{dx}\frac{\partial L}{\partial y'}\right) \mathring{y}\, dx. \tag{3.61}$$

■

In a more general case, the domain of $\mathcal{E}$ is an infinite dimensional manifold $\mathcal{M}$. Its variation is a covector field $d\mathcal{E} \in \Gamma(T^*\mathcal{M})$ and its gradient with respect some inner product structure is a vector field $\operatorname{grad} \mathcal{E} \in \Gamma(T\mathcal{M})$.

Problem 3.4 $\mathcal{M}$ is an infinite dimensional manifold and $\mathcal{E}\colon \mathcal{M} \to \mathbb{R}$ is a function defined on $\mathcal{M}$. The following defines an unconstrained optimization problem on $\mathcal{M}$:

$$\underset{y \in \mathcal{M}}{\text{minimize}}\ \mathcal{E}(y). \tag{3.62}$$

The necessary condition for a *locally optimal solution* $y \in \mathcal{M}$ is that variation of

the energy along any variation $\mathring{y} \in T_y\mathcal{M}$ must vanish:

$$d\mathcal{E}|_y[\![\mathring{y}]\!] = 0, \quad \text{for all } \mathring{y} \in T_y\mathcal{M}, \tag{3.63}$$

or simply

$$d\mathcal{E}|_y = 0 \in T_y^*\mathcal{M}. \tag{3.64}$$

Using any inner product, this vanishing variation condition becomes

$$\operatorname{grad}\mathcal{E}|_y = 0. \tag{3.65}$$

We call (3.65) the **Euler–Lagrange equation**. It is not a surprise that the optimality condition for problems on finite dimensional manifold (Problem 3.2) and problems on infinite dimensional manifold (Problem 3.4) look almost identical, but we usually use Euler-Lagrange equation to refer to the metric-informed equation (3.65) for infinite dimensional cases.

In the sections to follow, we always denote our optimization variable as $x \in \mathcal{M}$ where $\mathcal{M}$ can be a finite- or infinite-dimensional manifold.

3.2.3 Optimization with equality constraints

An optimization with equality constraints takes the following general form. Let $\mathcal{M}$ be a manifold, $\mathcal{F}\colon \mathcal{M} \to \mathbb{R}$ be the objective function, and let $\mathcal{G}\colon \mathcal{M} \to \mathcal{U}$ be a constraint function for some vector space $\mathcal{U}$. The optimization problem is stated as

$$\begin{aligned} &\underset{x\in\mathcal{M}}{\text{minimize}}\ \mathcal{F}(x) \\ &\text{subject to } \mathcal{G}(x) = \mathbf{0}_{\mathcal{U}}. \end{aligned} \tag{3.66}$$

Here $\mathbf{0}_{\mathcal{U}}$ simply denotes the zero element in the vector space $\mathcal{U}$. We assume both $\mathcal{F}$ and $\mathcal{G}$ are smooth. The necessary condition for the optimal solution x is

$$d\mathcal{F}|_x[\![\mathring{x}]\!] = 0 \quad \text{for all } \mathring{x} \in (d\mathcal{G}|_x)^\perp \tag{3.67}$$

where $(d\mathcal{G}|_x)^\perp$ denotes the annihilator subspace

$$(d\mathcal{G}|_x)^\perp = \left\{ \mathring{x} \;\middle|\; d\mathcal{G}|_x[\![\mathring{x}]\!] = 0 \right\}. \tag{3.68}$$

Note that the type of $d\mathcal{G}|_x$ is $d\mathcal{G}|_x\colon T_x\mathcal{M} \xrightarrow{\text{linear}} \mathcal{U}$. The condition implies that there exists a **Lagrange multiplier** $\lambda_0 \in \mathcal{U}^*$ such that $\mathcal{G}|_x = \mathbf{0}$ and

$$d\mathcal{F}|_x + \left\langle \lambda_0 \middle| d\mathcal{G}|_x \right\rangle_{\mathcal{U}^*\times\mathcal{U}} = 0 \in T_x\mathcal{M}. \tag{3.69}$$

3.2.4 Karush–Kuhn–Tucker conditions

Let us look at optimization problems with both equality constraints and inequality constraints. The energy function is still given as a function $\mathcal{F}\colon \mathcal{M} \to \mathbb{R}$ on a general manifold $\mathcal{M}$. The equality constraints are given through a function $\mathcal{G}\colon \mathcal{M} \to \mathcal{U}$ to a vector space $\mathcal{U}$. The inequality constraints are defined by a (proper) convex cone

$\mathcal{C} \subset \mathcal{V}$ in a vector space $\mathcal{V}$ and a function $\mathcal{H}\colon \mathcal{M} \to \mathcal{V}$. The optimization problem takes the following form:

$$\begin{cases} \underset{y\in\mathcal{M}}{\text{minimize}}\ \mathcal{F}(y) \quad \text{subject to} \\ \mathcal{G}(y) = \mathbf{0}_{\mathcal{U}} \\ \mathcal{H}(y) \in \mathcal{C}. \end{cases} \tag{3.70}$$

The necessary stationary conditions for an optimal solution $y_0 \in \mathcal{M}$ are the following **Karush–Kuhn–Tucker (KKT) conditions**. There exist Lagrange multiplies $\lambda_0 \in \mathcal{U}^*$ and

$$\mu_0 \in \mathcal{C}^\circ \coloneqq \left\{ \mu \in \mathcal{V}^* \,\middle|\, \langle \mu | v \rangle_{\mathcal{V}^*\times\mathcal{V}} \le 0 \text{ for all } v \in \mathcal{C} \right\} \subset \mathcal{V}^* \tag{3.71a}$$

($\mathcal{C}^\circ$ is called the **polar cone** of $\mathcal{C}$) such that $\mathcal{G}|_{y_0} = \mathbf{0}$, $\mathcal{H}|_{y_0} \in \mathcal{C}$,

$$d\mathcal{F}|_{y_0} + \left\langle \lambda_0 \middle| d\mathcal{G}|_{y_0} \right\rangle_{\mathcal{U}^*\times\mathcal{U}} + \left\langle \mu_0 \middle| d\mathcal{H}|_{y_0} \right\rangle_{\mathcal{V}^*\times\mathcal{V}} = 0 \in T_{y_0}\mathcal{M}, \tag{3.71b}$$

and

$$\left\langle \mu_0 \middle| \mathcal{H}(y_0) \right\rangle_{\mathcal{V}^*\times\mathcal{V}} = 0. \tag{3.71c}$$

The last condition (3.71c) is called the **complementary slackness condition**. It implies that if $\mathcal{H}(y_0)$ is the interior of the cone $\mathcal{C}$, then the Lagrange multiplier term is deactivated $\mu_0 = 0$. Only when $\mathcal{H}(y_0)$ lies on the boundary of the cone, we see an associated "normal force" μ_0 emerges to keep $\mathcal{H}(y_0)$ in the interior of the cone. Condition (3.71c) is the pullback of the corresponding complementary slackness condition about the convex cone on $\mathcal{V}$ back to $\mathcal{M}$ via the map $\mathcal{H}$. In (3.71b), the 2nd and 3rd terms can also be interpreted as the pullbacks of the covectors $\lambda_0 \in \mathcal{U}^*, \mu_0 \in \mathcal{V}^*$ via $\mathcal{G}$ and $\mathcal{H}$ respectively.

3.3 Computational methods for optimization problems

3.3.1 Direct and iterative methods for unconstrained problems

For an unconstrained problem on a (finite- or infinite-dimensional) manifold $\mathcal{M}$

$$\underset{x\in\mathcal{M}}{\text{minimize}}\, \mathcal{F}(x), \tag{3.72}$$

there are two ways to find the optimal solution. The first method is to solve the optimality condition directly

$$d\mathcal{F}|_x = 0 \in T_x\mathcal{M}. \tag{3.73}$$

The second method is to find a metric $\sharp\colon T^*M \xrightarrow{linear} TM$ and solve the **gradient flow**: starting with an initial guess $x_0 \in \mathcal{M}$, the path $x\colon [0,\infty) \to \mathcal{M}$ satisfying the following differential equation

$$\dot{x}(t) = -(d\mathcal{F}_{x(t)})^\sharp = -\operatorname{grad} \mathcal{F}_{x(t)} \tag{3.74}$$

will[4] converge to the stable point $x^{\text{opt}} \in \mathcal{M}$ with vanishing the gradient

$$(d\mathcal{F}|_{x^{\text{opt}}})^\sharp = \operatorname{grad} \mathcal{F}|_{x^{\text{opt}}} = 0. \tag{3.75}$$

Numerically, this is often done via a **forward Euler** scheme, where a sequence of intermediate states $x^{(k)}$ is defined iteratively

$$\begin{aligned} x^{(0)} &= x_0 \quad \text{(initial guess)} \\ x^{(k+1)} &= x^{(k)} - \Delta t^{(k)} \left(d\mathcal{F}|_{x^{(k)}}\right)^\sharp, \end{aligned} \tag{3.76}$$

where $\Delta t^{(k)} > 0$ denotes the k-th time step, which can be a fixed value or determined via line search. The iterative scheme (3.76) is often called the **Gradient Descent** method.

■ **Example 3.8 — An extended discussion on Dirichlet energy from Example 3.6.** Let $\mathcal{A} = \{u\colon [a,b] \to \mathbb{R}, u(a) = c_a, u(b) = c_b\}$ for some constants c_a, c_b. Consider the Dirichlet energy $\mathcal{E}(u) = \int_a^b \frac{1}{2} u'(x)^2 \, dx$. We saw in Example 3.6 that

$$d\mathcal{E}_u[\![\mathring{u}]\!] = \int_a^b u'(x)\mathring{u}'(x) \, dx = -\int_a^b u''(x)\mathring{u}(x) \, dx, \tag{3.77}$$

and the gradient $\operatorname{grad} \mathcal{E}_u$ associated with the $m(x)$-weighted inner product $\langle\!\langle \mathring{u}, \mathring{v} \rangle\!\rangle_{L^2} = \int_a^b \mathring{u}(x)\mathring{v}(x)m(x)dx$ is

$$\operatorname{grad} \mathcal{E}_u(x) = -\frac{u''(x)}{m(x)}. \tag{3.78}$$

Using the direct method, we can solve the optimality condition (also called Euler-Lagrange equation)

$$\begin{cases} u''(x) = 0 \\ u(a) = c_a \\ u(b) = c_b. \end{cases} \tag{3.79}$$

Alternatively, one can use the gradient flow method, and solve the differential equation for $u\colon [a,b] \times [0,\infty) \to \mathbb{R}$

$$\begin{cases} \frac{d}{dt} u(x,t) = -\operatorname{grad} \mathcal{E}_{u(\cdot,t)} = \frac{u''(x,t)}{m(x)} \\ u(x,0) = u_0(x), \end{cases} \tag{3.80}$$

where $u_0(x) \in \mathcal{A}$ is an arbitrary initial function.

Interestingly, (3.79) corresponds to solving a **Laplace problem** with specified boundary value, while (3.80) corresponds to solving the **heat equation** with an arbitrary initial value. Both of them gives a **harmonic function** with the specified boundary value! ■

[4] The convergence to a stable point often requires additional assumptions on the objective function $\mathcal{F}$ such as convexity, superlinearity, *etc.*. Thankfully most objective functions in computer graphics do possess these nice properties.

■ **Example 3.9 — Newton's method.** Let M be an m-dimensional manifold and $f\colon M \to \mathbb{R}$ be an objective function and we want to solve the following optimization problem:

$$\underset{x\in M}{\text{minimize}}\, f(x). \tag{3.81}$$

Suppose M has a coordinate system $(x_1, \cdots, x_m)\colon M \to \mathbb{R}^m$ that are not necessarily orthonormal. Given a point $x \in M$ and a tangent $\mathring{x} \in T_xM$, the differential can be written in terms of the coordinates:

$$\mathring{f} = df_x[\![\mathring{x}]\!] \tag{3.82}$$

$$= \frac{\partial f}{\partial x_1}(x)\mathring{x}_1 + \cdots + \frac{\partial f}{\partial x_m}(x)\mathring{x}_m, \text{ where } \mathring{x} = \begin{bmatrix} \mathring{x}_1 \\ \vdots \\ \mathring{x}_m \end{bmatrix} \tag{3.83}$$

We need a metric $\langle\cdot,\cdot\rangle$ to define the gradient flow. Write the metric as a matrix-valued function $\mu = \big[\mu_{ij}\big] : M \to \mathbb{R}^{m\times m}$ using the coordinates:

$$\left\langle \begin{bmatrix} \mathring{x}_1 \\ \vdots \\ \mathring{x}_m \end{bmatrix}, \begin{bmatrix} \mathring{y}_1 \\ \vdots \\ \mathring{y}_m \end{bmatrix} \right\rangle = \begin{bmatrix} \mathring{x}_1 & \cdots & \mathring{x}_m \end{bmatrix} \begin{bmatrix} \mu_{11}(x) & \cdots & \mu_{1m}(x) \\ \vdots & \ddots & \vdots \\ \mu_{m1}(x) & \cdots & \mu_{mm}(x) \end{bmatrix} \begin{bmatrix} \mathring{y}_1 \\ \vdots \\ \mathring{y}_m \end{bmatrix} \tag{3.84}$$

We can use the usual matrix algebra and compute

$$df_x[\![\mathring{x}]\!] = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) & \cdots & \frac{\partial f}{\partial x_m}(x) \end{bmatrix}^T \begin{bmatrix} \mathring{x}_1 \\ \vdots \\ \mathring{x}_m \end{bmatrix} \tag{3.85}$$

$$= \Bigg\langle \underbrace{\begin{bmatrix} \mu_{11}(x) & \cdots & \mu_{1m}(x) \\ \vdots & \ddots & \vdots \\ \mu_{m1}(x) & \cdots & \mu_{mm}(x) \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_m}(x) \end{bmatrix}}_{\operatorname{grad}_\mu f|_x}, \begin{bmatrix} \mathring{x}_1 \\ \vdots \\ \mathring{x}_m \end{bmatrix} \Bigg\rangle. \tag{3.86}$$

Since the original optimization Problem 3.81 doesn't concern to assign a metric to the manifold M, we can choose a metric ourselves such that the matrix $\mu(x)$ coincides with the **Hessian** of f:

$$\begin{bmatrix} \mu_{11}(x) & \cdots & \mu_{1m}(x) \\ \vdots & \ddots & \vdots \\ \mu_{m1}(x) & \cdots & \mu_{mm}(x) \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_m}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_m \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_m^2}(x) \end{bmatrix} =: H_f(x). \tag{3.87}$$

Optimization methods based on the gradient flow of this Hessian metric are called **Newton's method**:

$$\dot{x} = H_f(x)^{-1} \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_m}(x) \end{bmatrix}. \tag{3.88}$$

■

4. Continuum Mechanics

Continuum mechanics is the study of the statics and dynamics of deformable bodies. These deformable bodies can behave like elastic solid, Newtonian fluid, elastoplastic solid, viscoplasticity (non-Newtonian) fluid, ferrofluid, or plasma. Which behavior should a deformable body follow depends on a constitutive model. These models boil down to describing how a potential energy is defined as a function of the deformation of the body, and how a dissipation function depends on both the deformation and the rate of deformation of the body. The equations of motion are derived from the model by variational principle.

- **(Hyper)elasticity** The potential energy is only a function of the metric induced by the deformation from the world metric. Examples include elastic bodies such as jelly, rubber, *etc.* Examples also include most solids undergone small deformations or rigid motions.

- **Newtonian fluid** The potential energy is only a function of the induced change of volume by the deformation; the dissipation is only a function of the rate of deformation.

- **Elastoplasticity** An elasticity setup but the system carries extra variables that encode a reference metric for rest state, and the dissipation function depends on the rate of change of the rest metric. Examples include solids undergone irreversible plastic deformation, hardening, and fracturing.

- **Viscoplasticity** In addition to elastoplasticity setup, the fluid carries a convected tensor field describing the statistics of the direction microscopic anisotropic molecules. The dissipation also depends on the rate of change of the deformation. Examples are fluids consisting long polymers such as starch water mixture.

- **Ferrofluid** A Newtonian fluid that carries an additional convected field of magnetization, adding magnetic terms to the potential energy and the dissipation function. Examples are ferromagnetic liquids.

- **Magnetohydrodynamic fluids** Magnetohydrodynamic fluids are fluids with free electrons, *i.e.* they are electrically conductive fluids. In particular, such a continuum is a Newtonian fluid that carries trapped magnetic field due to the electric conductivity of the fluid. This adds electromagnetic term to the potential energy. Examples include liquid metals and ionized gas (plasmas).

The advantage of employing exterior calculus in formulating continuum mechanical equations are

- Strong type system for tensors: Instead of identifying tensors as matrices or arrays of various sizes, a stronger type system is able to distinguish the type of, for example, stress and strain, which have different geometric and physical meanings, even though their matrix representations are the same size.

- More fundamental and simpler understanding of the relationship between objects: The language sorts out that many objects in continuum mechanics are related by the canonical pullback operator, which is an important insight when we need to invoke, for example, the commutativity between pullback and exterior derivative.

- Exterior calculus includes the Lie derivative, which is fundamental for doing calculus on deformations. Without the vocabulary of Lie derivatives, differential equations that describe conservation laws are broken into many obscure terms.

- Theorems such as symmetry of stress tensors, conservation of circulation in inviscid fluid etc can be reasoned just by type checking.

In this chapter, we will only focus on hyperelasticity.

4.1 Deformation map

A universal postulate for a continuum mechanical system is that the state of the system is a map representing the shape of a deformable body.

Postulate 4.1 — Continuum mechanical system. We have the following setups for a continuum mechanical system.

- M: a manifold representing **material domain** or **Lagrangian coordinate**.
- W: a manifold representing **world coordinate** or **Eulerian coordinate**.
- The state of deformation is described by a (time-dependent) embedding $\phi = \phi(t) \colon M \to W$. We call it **deformation map** or **flow map**.
- M is equipped with a time-independent $\rho_M \in \Omega^n(M)$ representing mass.

- W is equipped with a time-independent metric $\flat_W$.

The material domain M is typically $\mathbb{R}^n$ but sometimes just an index set of particles. When $M = \mathbb{R}^3$ coordinate system is denoted by (X, Y, Z) or interchangeably (X^1, X^2, X^3). The world domain W is typically $\mathbb{R}^n$. When $W = \mathbb{R}^3$ coordinate system is denoted by (x, y, z) or interchangeably (x^1, x^2, x^3).

Using the above objects, we can define the kinetic energy of a moving object.

Definition 4.1 — Kinetic energy. The kinetic energy of a deformation map is given by

$$\mathcal{K}(\phi, \dot{\phi}) \coloneqq \int_M \frac{1}{2} |\dot{\phi}|^2_{\flat_W} \rho_M. \tag{4.1}$$

In a conservative Newtonian mechanical system, we can derive the equations of motion by the kinetic energy $\mathcal{K}(\phi, \dot{\phi})$ and the potential energy $\mathcal{U}(\phi)$ of the system, both of which are functions of our state variable ϕ (and its time derivative $\dot{\phi}$ for the kinetic energy). Explicitly, the equation of motion is given by

$$\frac{\nabla}{dt} \frac{\partial \mathcal{K}}{\partial \dot{\phi}} = -d\mathcal{U}. \tag{4.2}$$

This equation is the critical point of the **action**

$$\mathcal{S}(\phi) = \int_0^T \Big(\mathcal{K}(\phi, \dot{\phi}) - \mathcal{U}(\phi) \Big)\, dt. \tag{4.3}$$

over the space of paths $\phi \colon [0, T] \to (M \to W)$ on the space $(M \to W)$ of deformations with fixed initial and final positions.

To describe elasticity, we only need a minimal assumption on the form of the potential energy $\mathcal{U}$ (Section 4.4). The remaining derivation for the equation of motion is to evaluate the differential $d\mathcal{U}$ with respect to ϕ. This calculation will lead to the emergence of the notion of stress with delicate tensor types (Section 4.5).

Before carrying out this calculation in Section 4.5, we first give an overview of the classical exposition of elasticity in Section 4.2, followed by a preliminary on tensor algebra in Section 4.3.

4.2 Elasticity overview

In this section, we give a quick overview of the classical exposition of elasticity. This exposition is based on how the system was discovered and explained, instead of how it could have been derived by calculating $d\mathcal{U}$. In this section, we will introduce several many matrices in this section, some of which represent strain and others represent stress of various kinds. Later in Section 4.5, we will re-introduce these tensors using a principled argument, from which we discover a more delicate understanding of these tensors.

When studying elasticity, we first introduce the concept of stress. Phenomenologically, a solid body has internal force. In the early 19th century, Augustin-Louis

Cauchy showed that the internal force $\mathbf{T}$ acting on an infinitesimal internal surface, known as the **traction force**, is a linear function of the surface normal. So we can write $\mathbf{T} = \boldsymbol{\sigma}\mathbf{n}$ where $\mathbf{n}$ is the surface normal, and $\boldsymbol{\sigma}$ is the matrix representing this linear function. This matrix $\boldsymbol{\sigma}$ is called the **Cauchy stress**.

The net force acting on an infinitesimal volume is the given by the sum of traction force on the volume's boundary. Hence, the net force density is given by the divergence $\mathbf{f}_{\text{net}} = \nabla \cdot \boldsymbol{\sigma}$ of the Cauchy stress. When the solid is in static equilibrium, one can show that $\boldsymbol{\sigma}$ must be a symmetric matrix. However, in fact, $\boldsymbol{\sigma}$ is symmetric even under accelerated motion for a different reason (under the assumption of hyperelasticity).

Hyperelasticity is the model for large deformation. When it comes to this large deformation, one starts to introduce many more other **alternative stress measures**. Only writing a few major ones, we have the **Kirchhoff stress** $\boldsymbol{\tau}$, the **first Piola–Kirchhoff stress** $\mathbf{P}$, the **second Piola–Kirchhoff stress** $\mathbf{S}$, the **nominal stress** $\mathbf{N} = \mathbf{P}^\intercal$. Their conversions are given by

Equation for	$\boldsymbol{\sigma}$	$\boldsymbol{\tau}$	$\mathbf{P}$	$\mathbf{S}$
$\boldsymbol{\sigma} =$	$\boldsymbol{\sigma}$	$\frac{1}{J}\boldsymbol{\tau}$	$\frac{1}{J}\mathbf{P}\mathbf{F}^\intercal$	$J^{-1}\mathbf{F}\mathbf{S}\mathbf{F}^\intercal$
$\boldsymbol{\tau} =$	$J\boldsymbol{\sigma}$	$\boldsymbol{\tau}$	$\mathbf{P}\mathbf{F}^\intercal$	$\mathbf{F}\mathbf{S}\mathbf{F}^\intercal$
$\mathbf{P} =$	$J\boldsymbol{\sigma}\mathbf{F}^{-\intercal}$	$\boldsymbol{\tau}\mathbf{F}^{-\intercal}$	$\mathbf{P}$	$\mathbf{F}\mathbf{S}$
$\mathbf{S} =$	$J\mathbf{F}^{-1}\boldsymbol{\sigma}\mathbf{F}^{-\intercal}$	$\mathbf{F}^{-1}\boldsymbol{\tau}\mathbf{F}^{-\intercal}$	$\mathbf{F}^{-1}\mathbf{P}$	$\mathbf{S}$

where the **deformation gradient** $\mathbf{F} = \nabla\boldsymbol{\phi}$ is Jacobian matrix of the deformation map $\boldsymbol{\phi}$, and $J = \det(\mathbf{F})$. A stress–strain relation is given by some function $\mathbf{S} = \mathbf{S}(\mathbf{C})$ where $\mathbf{C} = \mathbf{F}^\intercal\mathbf{F}$ is called the Cauchy–Green tensor.

Where do all these stresses come from? Also, they are all represented by square matrices of the same size, which are inherently coordinate dependent and there is no information of what type they are (are they endomorphism or bilinear form, or some other type?) What are their geometric intuition?

In fact, we do see something familiar in this conversion table. Recall Example 2.3 that the pullback of a 2-form in 3D involves $J\mathbf{F}^{-1}$. In particular, the relationship between $\mathbf{P}$ and $\boldsymbol{\sigma}$ looks like a pullback for 2-forms. This conversion relation between $\mathbf{P}$ and $\boldsymbol{\sigma}$ hints that, perhaps, both of them are 2-form. If we think about it, we realize that *of course!* $\boldsymbol{\sigma}$ *is a force-(co)vector valued 2-form.* The Cauchy stress represents the traction force of a given infinitesimal surface. Instead of representing the traction force as $\boldsymbol{\sigma}\mathbf{n}$ using a normal vector, a more geometrically accurate description is that the force is given by $\sigma[\![X, Y]\!]$ on an infinitesimal parallelogram spanned by vectors X, Y. In that case, the 1st Piola–Kirchhoff stress P is merely the change of coordinate to the material coordinate

$$P = \underset{\text{2-form}}{\phi^*}\ \sigma. \tag{4.4}$$

From this example, we see that the theory of elasticity would be much more natural when written in exterior calculus.

What is more, all of the above stress tensors arise automatically and carry specific differential form types when we try to derive $d\mathcal{U}$, essentially making exterior calculus unavoidable. Following this first-principle derivation, we do not need any prerequisite

phenomenological knowledge about the existence of internal traction force, Cauchy stress, or any other alternative stresses. They are just the results of variational calculus. Using representations other than exterior calculus remove some information about the physical and geometric nature of these stress tensors.

4.3 Tensor Algebra

What are tensors? In geometry, vectors, covectors, endomorphisms, bilinear forms, differential forms, *etc.*, are all special cases of tensors. When these objects are written in basis or coordinate, they become arrays of numbers. For example, vectors, covectors and $(n-1)$-forms are represented as $\mathbb{R}^n$ under basis; endomorphisms and bilinear forms are both represented as square matrices $\mathbb{R}^{n\times n}$.

For convenience, many introductions of tensors classify tensors by their array size. Tensors that are represented by 1D array $\mathbb{R}^n$ of numbers are called order-1 tensors; tensors that are represented by a matrix $\mathbb{R}^{n\times n}$ of numbers are called order-2 tensors. More generally, order-k tensors are arrays whose entries can be querried by k indices. However, just as the distinctions between vectors and covectors are important, describing tensors by their array size does not give enough information of their geometric meaning.

Here we define tensors with the starting point of vector spaces and vector bundles as building blocks. The construction of tensors retain the information of the type of these vector spaces (bundle).

4.3.1 Tensor product

The notion of **tensor product** allows us to combine two vector spaces and produce a new vector space.

Definition 4.2 — Tensor product space. Let U, V be two vector spaces. Consider a new vector space, denoted by $U \otimes V$, and a non-degenerate bilinear function called tensor product

$$(\cdot)\otimes(\cdot)\colon U\times V \xrightarrow{\text{bilinear}} U\otimes V. \tag{4.5}$$

By the bilinearity we have rules such as

$$(c\vec{u})\otimes\vec{v} = \vec{u}\otimes(c\vec{v}) = c(\vec{u}\otimes\vec{v}) \quad \text{for all } \vec{u}\in U, \vec{v}\in V \text{ and } c\in\mathbb{R}. \tag{4.6}$$

and by the non-degeneracy $\vec{u}_1\otimes\vec{v}$ and $\vec{u}_2\otimes\vec{v}$ are linearly independent in $U\otimes V$ whenever $\vec{u}_1, \vec{u}_2$ are linearly independent in U.

Concretely, let $(\vec{e}_i \in U)_{i\in I}$ be a basis for U and let $(\vec{f}_j \in V)_{j\in J}$ be a basis for V. Then the vector space $U\otimes V$ is the linear combination of $\vec{h}_{ij} \coloneqq \vec{e}_i\otimes\vec{f}_j$ as basis vectors with indices running over all $i\in I, j\in J$. The tensor product of $\vec{u} = \sum_{i\in I} u^i\vec{e}_i \in U$ and $\vec{v} = \sum_{j\in J} v^j\vec{f}_j \in V$ is given by

$$\vec{u}\otimes\vec{v} \coloneqq \sum_{i\in I}\sum_{j\in J} u^i v^j \vec{h}_{ij}. \tag{4.7}$$

In particular,

$$\dim(U \otimes V) = \dim(U)\dim(V), \tag{4.8}$$

in contrast to $\dim(U \times V) = \dim(U) + \dim(V)$.

A general element $A \in U \otimes V$ takes the form

$$A = \sum_{i\in I}\sum_{j\in J} A^{ij}\vec{h}_{ij}. \tag{4.9}$$

Here $\mathbf{A} = (A^{ij})_{i\in I, j\in J}$ is the matrix representation of the tensor A.

Definition 4.3 — Decomposable tensor. An element A in $U \otimes V$ is said to be **decomposable** if there exist $\vec{u} \in U$ and $\vec{v} \in V$ so that $A = \vec{u} \otimes \vec{v}$. In a basis representation, decomposability means $A^{ij} = u^i v^j$, *i.e.* the matrix $\mathbf{A} = (A^{ij})_{ij}$ can be expressed as an outer product $\mathbf{A} = \mathbf{u}\mathbf{v}^\intercal$.

A generic element in $U \otimes V$ is not decomposable, unless either U or V is a one-dimensional vector space.

■ **Example 4.1** The space of endomorphisms $\mathrm{End}(U) \coloneqq \{A\colon U \xrightarrow{\text{linear}} U\}$ on a vector space U is the tensor product space $\mathrm{End}(U) = U^* \otimes U$. An endomorphism is a vector-valued covector, as it takes in a vector and spits out a vector. ■

■ **Example 4.2** The space of bilinear forms $\{B\colon U \xrightarrow{\text{linear}} U^*\} = \{B\colon U \times U \xrightarrow{\text{bilinear}} \mathbb{R}\}$ on a vector space U is the tensor product space $U^* \otimes U^*$. ■

Definition 4.4 — Tensor product bundle. Let E_1, E_2 be two vector bundles defined over a common manifold M. Then $E_1 \otimes E_2$ is a new vector bundle over M where each fiber $(E_1 \otimes E_2)_p$ is defined by the tensor product space $(E_1)_p \otimes (E_2)_p$ for each $p \in M$.

4.3.2 Tensor power

A special case of constructing tensor product spaces is taking tensor product between the same vector space. Let U be a vector space. Then we may take $U \otimes U =: \otimes^2 U$ and continue tensoring more copies of the same space $\otimes^3 U \coloneqq U \otimes U \otimes U$ and so on.

Definition 4.5 — Tensor power space. We recursively define $\otimes^k U$ by $\otimes^0 U \coloneqq \mathbb{R}, \otimes^1 U = U$, and a non-degenerate bilinear function

$$\otimes\colon \otimes^k U \times \otimes^\ell U \xrightarrow{\text{bilinear}} \otimes^{k+\ell} U \tag{4.10}$$

such that it is associative

$$(A \otimes B) \otimes C = A \otimes (B \otimes C) \quad \text{for all } A \in \otimes^k U, B \in \otimes^\ell U, C \in \otimes^m U. \tag{4.11}$$

■ **Example 4.3** Space of bilinear forms on U is the tensor space $\otimes^2 U^*$. ■

Under a basis for U and its dual basis for U^*, bilinear forms on U is represented by a square matrix with size n-by-n where $n = \dim(U)$. Just as the consideration of symmetric matrices and skew-symmetric matrices are important when studying bilinear forms, we should also discuss the symmetric and skew-symmetric tensor powers.

Using a similar formal definition we define the following special spaces of tensor power spaces.

Definition 4.6 — Symmetric tensor power space. We recursively define $\odot^k U$ by $\odot^0 U \coloneqq \mathbb{R}, \odot^1 U = U$, and a non-degenerate bilinear function

$$\odot\colon\ \odot^k U \times \odot^\ell U \xrightarrow{\text{bilinear}} \odot^{k+\ell} U \tag{4.12}$$

satisfying

- associativity:

$$(A \odot B) \odot C = A \odot (B \odot C) \quad \text{for all } A \in \odot^k U, B \in \odot^\ell U, C \in \odot^m U. \tag{4.13}$$

- symmetry:

$$\vec{u} \odot \vec{v} = \vec{v} \odot \vec{u} \quad \text{for all } \vec{u}, \vec{v} \in U. \tag{4.14}$$

The space $\odot^k U$ is included in $\otimes^k U$ by the inclusion map for decomposable symmetric tensors:

$$\begin{aligned} \odot^k U &\hookrightarrow \otimes^k U \\ \vec{u}_1 \odot \cdots \odot \vec{u}_k &\mapsto \sum_{\sigma \in S_k} \vec{u}_{\sigma(1)} \otimes \cdots \otimes \vec{u}_{\sigma(k)} \end{aligned} \tag{4.15}$$

where S_k is the permutation group on $\{1, \ldots, k\}$.

Definition 4.7 — Skew-symmetric tensor power space (Exterior algebra revisited). We recursively define $\wedge^k U$ by $\wedge^0 U \coloneqq \mathbb{R}, \wedge^1 U = U$, and a non-degenerate bilinear function

$$\wedge\colon\ \wedge^k U \times \wedge^\ell U \xrightarrow{\text{bilinear}} \wedge^{k+\ell} U \tag{4.16}$$

satisfying

- associativity:

$$(A \wedge B) \wedge C = A \wedge (B \wedge C) \quad \text{for all } A \in \wedge^k U, B \in \wedge^\ell U, C \in \wedge^m U. \tag{4.17}$$

- skew-symmetry:

$$\vec{u} \wedge \vec{u} = 0 \quad \text{for all } \vec{u} \in U. \tag{4.18}$$

The space $\wedge^k U$ is included in $\otimes^k U$ by the inclusion map for decomposable symmetric tensors:

$$\begin{aligned} \wedge^k U &\hookrightarrow \otimes^k U \\ \vec{u}_1 \wedge \cdots \wedge \vec{u}_k &\mapsto \sum_{\sigma \in S_k} (-1)^{|\sigma|} \vec{u}_{\sigma(1)} \otimes \cdots \otimes \vec{u}_{\sigma(k)} \end{aligned} \tag{4.19}$$

where S_k is the permutation group on $\{1, \ldots, k\}$, and $(-1)^{|\sigma|}$ is $+1$ (resp. -1) when σ is an even (resp. odd) permutation.

4.3.3 Pullback bundle

Suppose we have a deformation map $\phi\colon M \to W$. What is the type of its temporal derivative $\dot{\phi} = \frac{\partial}{\partial t}\phi$ and its spatial derivative $d\phi$? The object $\dot{\phi}$ is to be evaluated at a point in M, but its value is a vector in TW. The object $d\phi$ can take a vector at a point in M, but spits out a vector in TW. The subtlety here is that TW is not a bundle over M.

To characterize the type of objects such as $\dot{\phi}$ and $d\phi$, we need the concept of pullback bundle.

Definition 4.8 — Pullback bundle. Let E be a vector bundle over W and let $\phi\colon M \to W$ be a map. Then there induces a vector bundle E_ϕ over M where each fiber $(E_\phi)_p$, $p \in M$, is defined by

$$(E_\phi)_p \coloneqq E_{\phi(p)}. \tag{4.20}$$

For example $T_\phi W$ becomes a bundle over M; the fibers are $(T_\phi W)_x = T_{\phi(x)} W$.

Using pullback bundle, we can characterize the tensor types of derivatives of ϕ:

$$\dot{\phi} \in \Gamma(T_\phi W), \quad d\phi \in \Gamma(T^*M \otimes T_\phi W). \tag{4.21}$$

4.3.4 Type algebra

We will be using the following isomorphisms between tensor spaces.

Recall that the space of linear homomorphisms between two vector spaces U, V is denoted by

$$\mathrm{Hom}(U; V) \coloneqq \{A\colon U \xrightarrow{\text{linear}} V\}. \tag{4.22}$$

Theorem 4.1 $\mathrm{Hom}(U; V) \cong U^* \otimes V$.

Theorem 4.2 $(U \otimes V)^* \cong U^* \otimes V^*$. The dual pairing between $U^* \otimes V^*$ and $U \otimes V$ is given by

$$\langle \alpha \otimes \beta | \vec{u} \otimes \vec{v} \rangle \coloneqq \langle \alpha | \vec{u} \rangle \langle \beta | \vec{v} \rangle \tag{4.23}$$

for $\alpha \in U^*$, $\beta \in V^*$, $\vec{u} \in U$, and $\vec{v} \in V$.

Combining Theorem 4.1 and Theorem 4.2 gives:

Theorem 4.3 $\mathrm{Hom}(U; V)^* \cong \mathrm{Hom}(U^*; V^*)$. The dual pairing between a linear operator $B \in \mathrm{Hom}(U^*; V^*)$ and $A \in \mathrm{Hom}(U; V)$ is given by

$$\langle B | A \rangle = \mathrm{tr}(A^* B) = \mathrm{tr}(B^* A) = \mathrm{tr}(A B^*) = \mathrm{tr}(B A^*). \tag{4.24}$$

Theorem 4.4 Let E be a vector bundle over M. Then $\Gamma(E)^* \cong \Gamma(E^* \otimes \wedge^n T^* M)$. The dual pairing between a field $S \in \Gamma(E^* \otimes \wedge^n T^* M)$ and a field $X \in \Gamma(E)$ is given by

$$\langle\!\langle S | X \rangle\!\rangle \coloneqq \int_M \langle S | X \rangle. \tag{4.25}$$

Note that in (4.25) the pairing $\langle S | X \rangle$ is a pointwise dual pairing between fibers of E^* and fibers of E, leaving $\langle S | X \rangle \in \Gamma(\wedge^n T^* M) = \Omega^n(M)$ that can still be integrated over M. That is why the dual space of the space scalar functions are measures:

$$(\Omega^0(M))^* \cong \Omega^n(M). \tag{4.26}$$

The dual space of the space of vector fields is the space of covector-valued measures.

$$\Gamma(TM)^* \cong \Gamma(T^* M \otimes \wedge^n T^* M). \tag{4.27}$$

Theorem 4.5 Let U be an n-dimensional vector space. Then

$$U \otimes \wedge^n U^* \cong \wedge^{n-1} U^*. \tag{4.28}$$

The isomorphism is given by

$$\vec{u} \otimes \mu \mapsto i_{\vec{u}} \mu \quad \vec{u} \in U, \mu \in \wedge^n U^*. \tag{4.29}$$

Proof. One can check that (4.36) is non-degenerate, and that $U \otimes \wedge^n U^*$ and $\wedge^{n-1} U^*$ share the same dimension (note that $\dim(\wedge^n U^*) = 1$). One can also construct an explicit inverse map for (4.36). For each $\omega \in \wedge^{n-1} U^*$, pick a $\mu \in \wedge^n U^*$ and our goal is to reconstruct $\vec{u} \in U$ so that $i_{\vec{u}} \mu = \omega$. (Note that a different choice of μ will only lead to rescaled $\vec{u}$, and still resulting in the same $i_{\vec{u}} \mu$ and $\vec{u} \otimes \mu$ using (4.6).) The reconstruction goes as picking an arbitrary basis $\vec{e}_1, \ldots, \vec{e}_n$ for U and

$$\vec{u} = \sum_{k=1}^{n} u^k \vec{e}_k, \quad u^k = (-1)^k \frac{\omega[\![\vec{e}_1, \ldots, \widehat{\vec{e}_k}, \ldots, \vec{e}_n]\!]}{\mu[\![\vec{e}_1, \ldots, \vec{e}_n]\!]} \tag{4.30}$$

where $\widehat{(\cdot)}$ means skipping over the term. □

■ **Example 4.4 — Cramer's rule.** Let $A\colon U \xrightarrow{\text{linear}} U$ be a linear map. Given $\vec{b} \in U$, find $\vec{x} \in U$ such that $A\vec{x} = \vec{b}$.

Pick an arbitrary basis $\vec{e}_1, \dots, \vec{e}_n$ for U and an arbitrary volume form $\mu \in \wedge^n U^*$. Note that $\det(A) = (\underset{n\text{-form}}{A^*\ \mu})/\mu$, which is independent of the choice of μ. We may assume $\mu[\![\vec{e}_1, \dots, \vec{e}_n]\!] = 1$. Now, $A\colon U \xrightarrow{\text{linear}} U$ induces a pullback on $(n-1)$-forms:

$$\underset{(n-1)\text{-form}}{A^*} : \wedge^{n-1} U^* \xrightarrow{\text{linear}} \wedge^{n-1} U^*. \tag{4.31}$$

Note that

$$\underset{(n-1)\text{-form}}{A^*}(i_{\vec{b}}\mu) = i_{A^{-1}\vec{b}}\underset{n\text{-form}}{A^*}\mu = \det(A) i_{\vec{x}}\mu. \tag{4.32}$$

Hence, using (4.30) we can reconstruct $\vec{x}$ as $\sum_{i=1}^n x^k \vec{e}_k$ with

$$x^k = \frac{1}{\det(A)}(-1)^k (A^* i_{\vec{b}}\mu)[\![\vec{e}_1, \dots, \widehat{\vec{e}_k}, \dots, \vec{e}_n]\!] \tag{4.33}$$

$$= \frac{1}{\det(A)}\mu[\![\vec{a}_1, \dots, \underbrace{\vec{b}}_{\substack{k\text{-th}\\ \text{slot}}}, \dots, \vec{a}_n]\!] \tag{4.34}$$

where $\vec{a}_i = A\vec{e}_i$. ■

The following is a more general version of Theorem 4.5.

Theorem 4.6 Let U be an n-dimensional vector space. Then

$$(\otimes^k U) \otimes \wedge^n U^* \cong \wedge^{n-k} U^*. \tag{4.35}$$

The isomorphism is given by

$$(\vec{u}_1 \otimes \cdots \otimes \vec{u}_k) \otimes \mu \mapsto i_{\vec{u}_k} \cdots i_{\vec{u}_1}\mu, \quad \vec{u}_1, \dots, \vec{u}_k \in U, \mu \in \wedge^n U^*. \tag{4.36}$$

Using Theorem 4.6 and 4.4 we can reason the dual space of the space of k-form fields:

$$\Omega^k(M)^* = \Gamma(\wedge^k T^*M)^* \cong \Gamma(\wedge^k TM \otimes \wedge^n T^*M) \cong \Gamma(\wedge^{n-k} T^*M) = \Omega^{n-k}(M). \tag{4.37}$$

Theorem 4.7 $\Omega^k(M)^* \cong \Omega^{n-k}(M)$. The dual pairing between $\alpha \in \Omega^{n-k}(M)$ and $\beta \in \Omega^k(M)$ is given by

$$\langle\!\langle \alpha | \beta \rangle\!\rangle = \int_M \alpha \wedge \beta. \tag{4.38}$$

See also Theorem 2.44.

4.4 Postulates of Elasticity

In addition to the general postulate Postulate 4.1 for continuum mechanics laid out in Section 4.1, (hyper)elasticity has the following additional assumption.

Postulate 4.2 — Hyperelasticity. A hyperelasticity system is a conservative continuum mechanical system with the potential energy $\mathcal{U}(\phi)$ taking the form of $\mathcal{U}(\phi) = \int_M \Psi(\underset{\odot^2 T^*W}{\phi^* \flat_W})$, where $\underset{\odot^2 T^*W}{\phi^* \flat_W}$ is the induced metric (a.k.a. pullback metric) from $\flat_W$ by the deformation map ϕ, and $\Psi\colon \odot^2(T^*M) \xrightarrow{\text{(nonlinear)}} \wedge^n T^*M$ is a base-point preserving map. This function Ψ is called the **strain energy density function**.

The equation of motion for a hyperelastic system is given by (4.2) (following the least action principle)

$$\rho_M \ddot{\phi} = -\frac{\partial \mathcal{U}}{\partial \phi} + f_{\text{ext}} \tag{4.39}$$

with possibly an external force $f_{\text{ext}} \in \Gamma(\wedge^n T^*M \otimes \wedge^n T^*_\phi W)$.

The induced metric is the measurement of world-space distances and angles, as a tensor defined on M. The elastic potential energy is the integral of some pointwise evaluation of local energy Ψ, and this local energy is only a function of the local induced metric.

We will derive $\frac{-\partial \mathcal{U}}{\partial \phi}$ in Section 4.5. We spend the remainder of this section to explain objects such as the pullback metric and the deformation gradient.

4.4.1 Deformation gradient

Let $\phi\colon M \to W$ be a deformation map. We define the **deformation gradient** as the world-vector-valued 1-form on M:

Definition 4.9 — Deformation gradient.

$$F := d\phi \in \Gamma(T^*M \otimes T_\phi W) \cong \Gamma(\text{Hom}(TM; T_\phi W)). \tag{4.40}$$

The deformation is the pushforward operator given from the flow map ϕ. Each displacement as a vector in the material is mapped to a displacement as a tangent vector in the world.

■ **Example 4.5** Under a Cartesian coordinate in $\mathbb{R}^3$, if we write

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \boldsymbol{\phi}(X, Y, Z) = \begin{bmatrix} \phi^1(X, Y, Z) \\ \phi^2(X, Y, Z) \\ \phi^3(X, Y, Z) \end{bmatrix} \tag{4.41}$$

then the deformation gradient has a matrix representation given by

$$\mathbf{F} = d\boldsymbol{\phi} = \begin{bmatrix} \frac{\partial \phi^1}{\partial X} & \frac{\partial \phi^1}{\partial Y} & \frac{\partial \phi^1}{\partial Z} \\ \frac{\partial \phi^2}{\partial X} & \frac{\partial \phi^2}{\partial Y} & \frac{\partial \phi^2}{\partial Z} \\ \frac{\partial \phi^3}{\partial X} & \frac{\partial \phi^3}{\partial Y} & \frac{\partial \phi^3}{\partial Z} \end{bmatrix} \tag{4.42}$$

■

4.4.2 Pullback metric

Recall that

$$\flat_W \in \Gamma(\odot^2 T^*W) \cong \Gamma(\mathrm{Hom}_{\mathrm{sym}}(TW; T^*W)) \tag{4.43}$$

denote the metric on the world space W. Now, with the deformation gradient F and its adjoint, we find a diagram

$$\begin{array}{ccc} TM & \xrightarrow{F} & TW \\ & & \downarrow{\scriptstyle \flat_W} \\ T^*M & \xleftarrow[F^*]{} & T^*W \end{array} \tag{4.44}$$

By composing these arrows, we obtain a symmetric bilinear form that can serve as a metric on M.

Definition 4.10 — Pullback metric. Define

$$\underset{\odot^2 T^*W}{\phi^*} : \odot^2 T^*W \to \odot^2 T^*M, \qquad \underset{\odot^2 T^*W}{\phi^*}\ \flat_W := F^* \flat_W F \tag{4.45}$$

where $F = d\phi \in \Gamma(\mathrm{Hom}(TM; T_\phi W))$ and $\flat_W \in \Gamma(\mathrm{Hom}_{\mathrm{sym}}(TW; T^*W))$ are treated as linear maps like in (4.44). When viewing $\phi^* \flat_W$ as a bilinear form that takes in two vectors, this definition is equivalent to

$$\langle \vec{a}, \vec{b} \rangle_{\phi^* \flat_W} = \left\langle d\phi[\![\vec{a}]\!], d\phi[\![\vec{b}]\!] \right\rangle_{\flat_W}, \quad \vec{a}, \vec{b} \in T_pM. \tag{4.46}$$

We call

$$C := F^* \flat_W F \in \Gamma(\odot^2 T^*M) \tag{4.47}$$

the **(right) Cauchy–Green tensor**.

$$\begin{array}{ccc} TM & \xrightarrow{F} & TW \\ & \overset{C}{\curvearrowleft} & \downarrow{\scriptstyle \flat_W} \\ T^*M & \xleftarrow[F^*]{} & T^*W \end{array} \tag{4.48}$$

The physical meaning of the pullback metric is best understood by (4.46). The pullback metric gives the following inner product structure: The inner product between two vectors $\vec{a}, \vec{b}$ in the material is evaluated by taking their world-space

inner product; the deformation gradient $F = d\phi$ first realizes these material vectors as world vectors before taking the world-space inner product.

Note that the inner product structure encodes the notion of lengths and angles. In particular, it encodes how much the material is stretched and sheared (after comparing it with a fixed reference material metric).

Definition 4.11 — Cauchy–Green constructor. Define

$$\mathcal{C}\colon \Gamma(T^*M \otimes T_\phi W) \to \Gamma(\odot^2 T^*M) \tag{4.49}$$
$$\mathcal{C}(F) \coloneqq F^* \flat_W F. \tag{4.50}$$

■ **Example 4.6** Continuing Example 4.5 in the Cartesian 3D space, where the metric tensor for $\flat_W$ is represented by the identity matrix, the right Cauchy–Green tensor is given by

$$\mathbf{C} = \mathbf{F}^\intercal \mathbf{F}. \tag{4.51}$$

■

4.4.3 Strain energy density function

One postulates that the elastic potential energy is only a function of the pullback metric C, which encodes all information of the local world-space measurements of distance and angle. This local potential energy is the strain energy density function of C. The strain energy density function is a base-point preserving map between vector bundles:

$$\Psi\colon \odot^2 T^*M \xrightarrow{\text{(nonlinear)}} \wedge^n T^*M. \tag{4.52}$$

It is the core material property. The value $\Psi(C)$ describes how much energy is stored (as an n-form) for a given pullback metric C. The function $\Psi_p\colon \odot^2 T_p^*M \to \wedge^n T_p^*M$ can depend on $p \in M$ for an inhomogeneous material.

Definition 4.12 — Derivative of Ψ. For each point $p \in M$, the differential $d(\Psi_p)$ of the function $\Psi_p\colon \odot^2 T_p^*M \to \wedge^n T_p^*M$ over the domain $\odot^2 T_p^*M$ is a $\wedge^n T_p^*M$-valued covector of $\odot^2 T_p^*M$; that is,

$$d(\Psi_p) \in (\odot^2 T_p^*M)^* \otimes \wedge^n T_p^*M \cong (\odot^2 T_pM) \otimes \wedge^n T_p^*M. \tag{4.53}$$

We denote this differential as

$$\frac{\partial \Psi}{\partial C} \in \Gamma\left((\odot^2 TM) \otimes \wedge^n T^*M\right), \quad \left.\frac{\partial \Psi}{\partial C}\right|_p \coloneqq d(\Psi_p). \tag{4.54}$$

The base-point-preserving function (as a function of C)

$$\frac{\partial \Psi}{\partial C}\colon \Gamma(\odot^2 T^*M) \to \Gamma\left((\odot^2 TM) \otimes \wedge^n T^*M\right) \tag{4.55}$$

is called a **stress–strain relation**.

For a general hyperelastic model, we do not need to make any further assumption. For whichever Ψ that is modeled, there is a derived $\frac{\partial \Psi}{\partial C}$. We will only need to know $\frac{\partial \Psi}{\partial C} \in \Gamma(\odot^2 TM \otimes \wedge^n T^*M)$ to derive $d\mathcal{U}$ (Section 4.5).

More discussion about special models and physical designs for Ψ will be given in Section 4.7.

4.5 Derivation of Elasticity

Our goal is to derive $\frac{\partial}{\partial \phi}\mathcal{U}(\phi)$ for (4.39) for hyperelastic systems (Postulate 4.2). The "computation tree" for evaluating $\mathcal{U}(\phi) = \int_M \Psi(C)$ is given by

$$\begin{array}{ccccccccc} C^\infty(M;W) & \xrightarrow{d} & \Gamma(T^*M \otimes T_\phi W) & \overset{\text{Def. 4.11}}{\xrightarrow{\mathcal{C}}} & \Gamma(\odot^2 T^*M) & \overset{\text{Sec. 4.4.3}}{\xrightarrow{\Psi}} & \Gamma(\wedge^n T^*M) & \xrightarrow{\int_M} & \mathbb{R} \\ \phi & & F = d\phi & & C = \mathcal{C}(F) & & \Psi(C) & & \mathcal{U}(\phi) \end{array} \tag{4.56}$$

To compute the variation $\frac{\partial \mathcal{U}}{\partial \phi}$ we apply the method of back propagation. The tangent spaces of (4.56) and the differentials are

$$\Gamma(T_\phi W) \xrightarrow{d^\nabla} \Gamma(T^*M \otimes T_\phi W) \xrightarrow{d\mathcal{C}|_F} \Gamma(\odot^2 T^*M) \xrightarrow{d\Psi|_C} \Gamma(\wedge^n T^*M) \xrightarrow{\int_M} \mathbb{R}. \tag{4.57}$$

Here, the variation of the first map d requires the Levi-Civita connection ∇^W for $\flat_W$ on W:

$$\frac{\partial}{\partial \epsilon}\Big|_{\epsilon=0} \phi_\epsilon = \mathring{\phi} \in \Gamma(T_\phi W) \quad \Longrightarrow \quad \frac{\partial}{\partial \epsilon}\Big|_{\epsilon=0} (\underset{\text{0-form}}{d\ \phi_\epsilon}) = \underset{\substack{TW\text{-valued} \\ \text{0-form}}}{d^\nabla \mathring{\phi}}. \tag{4.58}$$

Using the duality theorems in Section 4.3.4, we find the dual spaces of the sequence of vector spaces in (4.57) and connect them by the adjoint maps:

$$\begin{array}{ccccccccc} \Gamma(T_\phi W) & \xrightarrow{d^\nabla} & \Gamma(T^*M \otimes T_\phi W) & \xrightarrow{d\mathcal{C}|_F} & \Gamma(\odot^2 T^*M) & \xrightarrow{d\Psi|_C} & \Gamma(\wedge^n T^*M) & \xrightarrow{\int_M} & \mathbb{R} \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ \Gamma(\wedge^n T^*M & \xleftarrow{(d^\nabla)^*} & \Gamma(\wedge^{n-1} T^*M & \xleftarrow{(d\mathcal{C}_F)^*} & \Gamma(\odot^2 TM & \xleftarrow{(d\Psi_C)^*} & \Omega^0(M) & \xleftarrow{(\int_M)^*} & \mathbb{R}^* \\ \otimes T_\phi^* W) & & \otimes T_\phi^* W) & & \otimes \wedge^n T^*M) & & & & \end{array} \tag{4.59}$$

Take $1 \in \mathbb{R}^*$ as the seed and back-propagate it using these adjoint maps. The intermediate variables during the back-propagation are known as the **stress tensors** of various kinds.

The adjoint of integration distributes 1 into the constant function

$$\mathbb{1} = (\textstyle\int_M)^*(1) \in \Omega^0(M). \tag{4.60}$$

The next adjoint map $(d\Psi)^*$ yields

$$\frac{\partial \Psi}{\partial C} = (d\Psi|_C)^*(\mathbb{1}) \in \Gamma(\odot^2 TM \otimes \wedge^n T^*M), \tag{4.61}$$

which we assume can be looked up (Section 4.4.3) while further discussions about its evaluation can be found in Section 4.7.

Definition 4.13 — 2nd Piola–Kirchhoff stress tensor. The **2nd Piola–Kirchhoff stress tensor (PK2)** $S \in \Gamma(\odot^2 TM \otimes \wedge^n T^*M)$ is defined by

$$\frac{S}{2} := \frac{\partial \Psi}{\partial C} = (d\Psi|_C)^*(\mathbb{1}). \tag{4.62}$$

The next adjoint map $(d\mathcal{C})^*$ gives us:

Definition 4.14 — 1st Piola–Kirchhoff stress tensor.

$$P := (d\mathcal{C}|_F)^* \frac{\partial \Psi}{\partial C} = (d\mathcal{C}|_F)^*(d\Psi|_C)^*(\mathbb{1}) \in \Gamma(\wedge^{n-1} T^*M \otimes T^*_\phi W). \tag{4.63}$$

called the **1st Piola–Kirchhoff stress tensor (PK1)**.

Finally,

$$\frac{\partial \mathcal{U}}{\partial \phi} = (d^\nabla)^* P = (d^\nabla)^*(d\mathcal{C}|_F)^*(d\Psi|_C)^*(\mathbb{1}) \in \Gamma(\wedge^n T^*M \otimes T^*_\phi W). \tag{4.64}$$

To complete the calculation, we derive explicit formulas for the adjoints $(d^\nabla)^*$ and $(d\mathcal{C}|_F)^*$.

Theorem 4.8 Represent the domain $\Gamma(\odot^2 TM \otimes \wedge^n T^*M)$ and the target space $\Gamma(\wedge^{n-1} T^*M \otimes T^*_\phi W)$ of $(d\mathcal{C}|_F)^*$ as spaces of linear maps $\Gamma(\mathrm{Hom}_{\mathrm{sym}}(T^*M; TM) \otimes \wedge^n T^*M)$ and $\Gamma(\mathrm{Hom}(T^*M; T^*_\phi W) \otimes \wedge^n T^*M)$ respectively using Theorem 4.1 and Theorem 4.5. Then the linear map

$$d\mathcal{C}|_F^* : \Gamma(\mathrm{Hom}_{\mathrm{sym}}(T^*M; TM) \otimes \wedge^n T^*M) \xrightarrow{\text{linear}} \Gamma(\mathrm{Hom}(T^*M; T^*_\phi W) \otimes \wedge^n T^*M) \tag{4.65}$$

is explicitly given by

$$d\mathcal{C}|_F^*[\![H]\!] = 2\flat_W FH, \quad H \in \Gamma(\mathrm{Hom}_{\mathrm{sym}}(T^*M; TM) \otimes \wedge^n T^*M), \tag{4.66}$$

in terms of linear maps $F \in \Gamma(\mathrm{Hom}(TM; T_\phi W))$ and $\flat_W \in \Gamma(\mathrm{Hom}_{\mathrm{sym}}(T_\phi W; T^*_\phi W))$.

Proof. Differentiating (4.50) ($\mathcal{C}(F) = F^*\flat_W F$) with respect to F, we obtain

$$d\mathcal{C}_F[\![\mathring{F}]\!] = \mathring{F}^*\flat_W F + F^*\flat_W \mathring{F}, \quad \mathring{F} \in \Gamma(\mathrm{Hom}(TM; T_\phi W)). \tag{4.67}$$

Since this is a pointwise linear operation, we may focus on its pointwise adjoint as in

the following diagram (with $\Gamma(\cdots)$ and $\otimes \wedge^n T^*M$ dropped):

$$\begin{array}{ccc} \mathrm{Hom}(\overset{\mathring{F}}{TM};\, T_\phi W) & \xrightarrow{d\mathcal{C}|_F} & \mathrm{Hom}_{\mathrm{sym}}(TM;\, T^*M) \\ \vdots & & \vdots \\ \mathrm{Hom}(T^*M;\, T^*_\phi W) & \xleftarrow{d\mathcal{C}|_F^*} & \mathrm{Hom}_{\mathrm{sym}}(\underset{H}{T^*M};\, TM) \end{array} \qquad \overset{\mathring{F}^* \flat_W F + F^* \flat_W \mathring{F}}{} \tag{4.68}$$

Using the dual pairing between linear maps (Theorem 4.3) the adjoint $d\mathcal{C}^*_F$ must satisfy

$$\mathrm{tr}\left((d\mathcal{C}|^*_F [\![H]\!])^* \mathring{F}\right) \overset{\substack{\text{def'n of}\\ \text{adjoint}\\ \text{and (4.67)}}}{=} \mathrm{tr}(H^*(\mathring{F}^* \flat_W F + F^* \flat_W \mathring{F})) \tag{4.69}$$

$$= \mathrm{tr}(H^* \mathring{F}^* \flat_W F) + \mathrm{tr}(H^* F^* \flat_W \mathring{F}) \tag{4.70}$$

$$= 2\,\mathrm{tr}(H^* F^* \flat_W \mathring{F}) \tag{4.71}$$

where the last step uses invariance of trace under cyclic permutations ($\mathrm{tr}(ABCD) = \mathrm{tr}(BCDA)$) and adjoints $\mathrm{tr}(A) = \mathrm{tr}(A^*)$, and that $H^* = H$for $H \in \mathrm{Hom}_{\mathrm{sym}}(TM^*;\, TM)$. Therefore, $d\mathcal{C}|^*_F [\![H]\!] = 2\flat_W FH$. □

Theorem 4.9 The adjoint of $\underset{\substack{TW\text{-valued}\\ 0\text{-form}}}{d^\nabla}$ is $\underset{\substack{T^*W\text{-valued}\\ (n-1)\text{-form}}}{-d^\nabla}$ up to a boundary integral.

$$\begin{array}{ccc} \Gamma(T_\phi W) & \xrightarrow{\underset{0\text{-form}}{d^\nabla}} & \Gamma(T^*M \otimes T_\phi W) \\ \vdots & & \vdots \\ \Gamma(\wedge^n T^*M \otimes T^*_\phi W) & \xleftarrow[\substack{=\\ (n-1)\text{-form}}]{(d^\nabla)^*\quad -d^\nabla} & \Gamma(\wedge^{n-1} T^*M \otimes T^*_\phi W) \end{array} \tag{4.72}$$

Proof. For each $\mathring{\phi} \in \Gamma(T_\phi W)$ and $P \in \Gamma(\wedge^{n-1} T^*M \otimes T^*_\phi W)$, we have

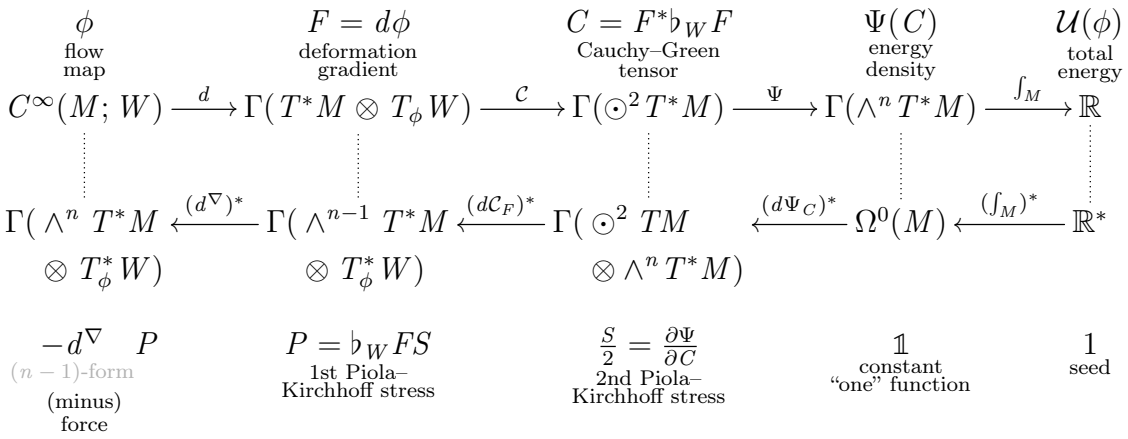
$$\underbrace{\int_M \underset{(n-1)\text{-form}}{d} \langle \mathring{\phi} | P \rangle}_{= \oint_{\partial M} \langle \mathring{\phi} | P \rangle} = \underbrace{\int_M \langle \underset{\substack{TW\text{-valued}\\ 0\text{-form}}}{d^\nabla} \mathring{\phi} \wedge P \rangle}_{\langle\!\langle d^\nabla \mathring{\phi} | P \rangle\!\rangle} + \underbrace{\int_M \langle \mathring{\phi} | \underset{\substack{T^*W\text{-valued}\\ (n-1)\text{-form}}}{d^\nabla} P \rangle}_{\langle\!\langle \mathring{\phi} | d^\nabla P \rangle\!\rangle} . \tag{4.73}$$

Therefore, up to a boundary term, the adjoint of $\underset{\substack{TW\text{-valued}\\ 0\text{-form}}}{d^\nabla}$ is $\underset{\substack{T^*W\text{-valued}\\ (n-1)\text{-form}}}{-d^\nabla}$. □

4.5.1 Summary of the derivation

The derivation for the differential of the potential energy $\mathcal{U}(\phi)$ follows the same logic as a reversed mode automatic differentiation. We first perform a forward evaluation, involving several intermediate spaces. Then we perform a back-propagation along their dual spaces. The intermediate variables are the Piola–Kirchhoff stress tensors.

The following diagram shows the forward nonlinear evaluations (upper row) and the linear back-propagation (lower row).



(4.74)

The equation of motion for the flow map is given by substituting the result of the derivative computation into (4.39)

$$\rho_M \ddot{\phi}^\nabla = \underset{(n-1)\text{-form}}{d^\nabla} P + f_{\text{ext}}. \tag{4.75}$$

■ **Example 4.7 — Summary in 3D.** We can translate the tensors in (4.74) into matrices under the 3D Cartesian coordinate, continuing Example 4.5 ($\mathbf{F} = \nabla\boldsymbol{\phi}$) and Example 4.6 ($\mathbf{C} = \mathbf{F}^\intercal\mathbf{F}$).

The 1st ($\mathbf{P}$) and 2nd ($\mathbf{S}$) Piola–Kirchhoff stress tensors are 3-by-3 matrices given by

$$\mathbf{S} = 2\frac{\partial \Psi}{\partial \mathbf{C}} \quad \left(S_{ij} = 2\frac{\partial \Psi(\mathbf{C})}{\partial C_{ij}}\right), \quad \mathbf{P} = \mathbf{F}\mathbf{S}. \tag{4.76}$$

The elastic force is given by taking the divergence of the 1st Piola–Kirchhoff stress tensor on the 2nd index:

$$\mathbf{f}_M = -\frac{\partial \mathcal{U}}{\partial \boldsymbol{\phi}} = \nabla \cdot \mathbf{P} \quad \left(f_i = \sum_j \frac{\partial}{\partial X^j} P_{ij}\right). \tag{4.77}$$

Here $\mathbf{f}_M$ is interpreted as force density per material volume. ■

4.6 Cauchy stress tensor

The 1st Piola–Kirchhoff stress tensor $P \in \Gamma(\wedge^{n-1} T^*M \otimes T^*_\phi W)$ is a world-covector-valued $(n-1)$-form. It is to be evaluated on a codimension-1 infinitesimal plane, and it returns the force as a covector in the world space.

> *The 1st Piola–Kirchhoff stress tensor assigns a world force covector on each codimension-1 plane in the material.*

By a change of coordinate of the same object, we have the **Cauchy stress tensor**.

The Cauchy stress tensor assigns a world force covector on each codimension-1 plane in the world.

Definition 4.15 — Cauchy stress tensor. Given a 1st Piola–Kirchhoff stress tensor $P \in \Gamma(\wedge^{n-1} T^*M \otimes T^*_\phi W)$, the Cauchy stress tensor is a covector-valued $(n-1)$-form

$$\sigma \in \Gamma(\wedge^{n-1} T^*W \otimes T^*W) \tag{4.78}$$

(on the image of ϕ) defined such that

$$P = \underset{(n-1)\text{-form}}{\phi^*} \sigma. \tag{4.79}$$

■ **Example 4.8** In a 3D Cartesian space, the relationship between the Cauchy stress tensor $\boldsymbol{\sigma}$ and the 1st Piola–Kirchhoff stress tensor $\mathbf{P}$ is given by

$$\mathbf{P} = J\boldsymbol{\sigma}\mathbf{F}^{-\intercal}, \quad J = \det(\mathbf{F}). \tag{4.80}$$

This is a direct application that in 3D

$$\phi^*(\mathbf{w})_{2\text{-form}} = (J\mathbf{F}^{-1}\mathbf{w})_{2\text{-form}} \circ \boldsymbol{\phi}. \tag{4.81}$$

The multiplication by $\mathbf{F}^{-1}$ in (4.80) is from the right with a transpose because it is the 2nd index of $\mathbf{P}$ and $\boldsymbol{\sigma}$ that corresponds to being a 2-form. ■

Theorem 4.10 Let $f \in \Gamma(\wedge^n T^*M \otimes T^*W)$ be the force covector measure defined on the world so that the total elastic force $d^\nabla\sigma$:

$$\underset{(n-1)\text{-form}}{d^\nabla} P = \underset{n\text{-form}}{\phi^*} f. \tag{4.82}$$

Then

$$\underset{(n-1)\text{-form}}{d^\nabla} \sigma = f. \tag{4.83}$$

Proof. Exterior derivative and pullback commutes:

$$\underset{n\text{-form}}{\phi^*} \underset{(n-1)\text{-form}}{d^\nabla} \sigma = \underset{(n-1)\text{-form}}{d^\nabla} \underset{(n-1)\text{-form}}{\phi^*} \sigma = \underset{(n-1)\text{-form}}{d^\nabla} P = \underset{n\text{-form}}{\phi^*} f. \tag{4.84}$$

□

■ **Example 4.9** In the 3D Cartesian setup, if $\mathbf{f}_W$ represents the force density per world-space volume and $\mathbf{f}_M$ represents the force density per material-space volume.

Then

$$\mathbf{f}_M = J\mathbf{f}_W \circ \boldsymbol{\phi}, \quad \nabla \cdot \boldsymbol{\sigma} = \mathbf{f}_W, \quad \nabla \cdot \mathbf{P} = \mathbf{f}_M. \tag{4.85}$$

■

Next, we show that the Cauchy stress tensor is symmetric. In contrast to physical arguments about angular momentum conservation (many versions of which apply only to W being Euclidean or under the assumption of equilibrium), we show the symmetry by type checking.

First, consider a version of the Cauchy stress so that its force-covector is converted into a vector using the inner product structure of the world:

$$\tilde{\sigma} \coloneqq \sharp_W \sigma \in \Gamma(\wedge^{n-1} T^*W \otimes TW) \tag{4.86}$$

$$\cong \Gamma(TW \otimes TW \otimes \wedge^n T^*W). \tag{4.87}$$

Note that in the 3D Cartesian representation $\tilde{\sigma}$ and σ share the same matrix representation. Now, in the type of (4.87) it is sensible to ask whether $\tilde{\sigma}$ belong to the symmetric sub-type

$$\tilde{\sigma} \overset{?}{\in} \Gamma(TW \odot TW \otimes \wedge^n T^*W). \tag{4.88}$$

Theorem 4.11 Under Postulate 4.2, we must have

$$\tilde{\sigma} \in \Gamma(TW \odot TW \otimes \wedge^n T^*W). \tag{4.89}$$

Proof. Regarding $\tilde{\sigma}$ as a homomorphism-valued measure

$$\tilde{\sigma} \in \Gamma(\mathrm{Hom}(T^*W; TW) \otimes \wedge^n T^*W) \tag{4.90}$$

we have

$$\underset{n\text{-form}}{\phi^*}\ \tilde{\sigma} = (\underset{(n-1)\text{-form}}{\phi^*}\ \tilde{\sigma})F^* = \sharp_W(\underset{(n-1)\text{-form}}{\phi^*}\ \sigma)F^* \tag{4.91}$$

$$= \sharp_W PF^* = \sharp_W \flat_W FSF^* = FSF^* \tag{4.92}$$

which is self-adjoint as S is self-adjoint. The pullback $\underset{n\text{-form}}{\phi^*}$ on the n-form part does not change the symmetry type of its homomorphism values. □

4.7 More on the strain energy density function

In Section 4.4.3 we described that the strain energy density function is a pointwise (base-point preserving) map from a Cauchy–Green tensor to an energy density

$$\Psi\colon\ \odot^2 T^*M \xrightarrow{\text{(nonlinear)}} \wedge^n T^*M. \tag{4.93}$$

How do we model Ψ? We need more context of geometry and physics.

Postulate 4.3 — Existence of rest metric. There is a time-independent **material metric** $\flat_M \in \Gamma(\odot^2 T^*M)$. Having $\flat_M$, one can compare the pullback metric $C = F^*\flat_W F$ with $\flat_M$ by their quotient

$$\hat{C} \coloneqq \sharp_M C = \flat_M^{-1} C \in \Gamma(\mathrm{Hom}(TM; TM)) \tag{4.94}$$

which is an endomorphism on TM (and by contrast C is a bilinear form):

$$\begin{array}{ccc} TM & \xrightarrow{F} & TW \\ & \overset{C}{\curvearrowleft} & \downarrow \flat_W \\ T^*M & \xleftarrow[F^*]{} & T^*W \end{array} \qquad \begin{array}{ccc} TM & \xrightarrow{F} & TW \\ \flat_M \downarrow & \overset{\hat{C}}{\curvearrowleft} & \downarrow \flat_W \\ T^*M & \xleftarrow[F^*]{} & T^*W \end{array} \tag{4.95}$$

The energy $\Psi(C)$ can be written as a scalar-valued function of $\hat{C}$

$$\Psi(C) = \hat{\Psi}(\hat{C})\rho_M, \quad \hat{\Psi}\colon \mathrm{End}(TM) \to \mathbb{R}, \tag{4.96}$$

and $\hat{\Psi}$ attains its minimum at $\hat{C} = \mathrm{id}_{TM}$. The value of $\hat{\Psi}$ is called the **internal energy** or **Helmholtz's free energy**.

The endomorphism $\hat{C}$ indicates how much itself deviates from the identity, as a measurement of relative stretching. When the internal energy is only a function of $\hat{C}$ and not on other "latent variables" such as entropy or temperature, then internal energy is equivalent to Helmholtz's free energy.

R In thermoelasticity, $\hat{\Psi}$ can depend on entropy (isentropic material) or on temperature (isothermal material). In the former case, $\hat{\Psi}$ is the internal energy, and in the latter case $\hat{\Psi}$ is Helmholtz's free energy.

In elastoplasticity, $\flat_M$ becomes variable as well. A changed rest metric $\flat_M$ models a permanent plastic deformation.

4.7.1 Strains

A strain is an expression of $\hat{C}$ that measures its deviation from identity.

- Green–St. Venant strain $E_{\mathrm{StV}} \coloneqq \frac{1}{2}(\hat{C} - I)$.
- Biot strain $E_{\mathrm{Biot}} \coloneqq \sqrt{\hat{C}} - I$.
- Hencky strain $E_{\mathrm{Hencky}} = \hat{H} \coloneqq \frac{1}{2}\ln \hat{C}$.
- Almansi strain $E_{\mathrm{Almansi}} \coloneqq \frac{1}{2}(I - \hat{C}^{-1})$.

All these strains agree with each other to first order when $\hat{C} \approx I$.

In many scenario for large deformations, it is natural to consider Hencky's logarithmic strain as our main measurement for the deviation from identity. By a

mere change of variable, one may also model the internal energy as a function of the Hencky strain:

$$\Psi(C) = \hat{\Psi}(\hat{C})\rho_M = \hat{\Psi}_{\log}(\hat{H})\rho_M, \quad 0 \in \operatorname*{argmin}_{\hat{H}\in \mathrm{End}(TM)} \hat{\Psi}_{\log}(\hat{H}) \tag{4.97}$$

■ **Example 4.10 — St. Venant–Kirchhoff model.** In the **St. Venant–Kirchhoff model**, one defines the internal energy as the following quadratic energy of some strain

$$\hat{\Psi}_{\mathrm{StVK}}(\hat{C}) = \frac{\hat{\lambda}}{2}\operatorname{tr}(E)^2 + \hat{\mu}\operatorname{tr}(E^2). \tag{4.98}$$

Here $\hat{\lambda}, \hat{\mu}$ are the (per unit mass) **Lamé constants**.

The corresponding stress–strain relation is

$$S = 2\frac{\partial \Psi(C)}{\partial C} = \sharp_M \frac{\partial \hat{\Psi}(\hat{C})}{\partial \hat{C}}\rho_M = \left(\hat{\lambda}\operatorname{tr}(E)\sharp_M + 2\hat{\mu}\sharp_M E\right)\rho_M \tag{4.99}$$

$$= \lambda \operatorname{tr}(E)\sharp_M + 2\mu\sharp_M E \tag{4.100}$$

where $\lambda = \hat{\lambda}\rho_M, \mu = \hat{\mu}\rho_M$ are Lamé constants as n-forms.

In Cartesian coordinate,

$$\mathbf{S} = \lambda \operatorname{tr}(\mathbf{E})\mathbf{I} + 2\mu\mathbf{E}. \tag{4.101}$$

■

4.7.2 Isotropic models

To design a more general $\hat{\Psi}(\hat{C})$ beyond a quadratic function on some strain, one takes advantage of another symmetry under the following isotropy assumption.

Postulate 4.4 — Isotropic material. We call a material **isotropic** if

$$\hat{\Psi}(\hat{C}) = \hat{\Psi}(R^{-1}\hat{C}R) \tag{4.102}$$

for all $R \in \mathrm{End}(TM)$ so that $R^*\flat_M R = \flat_M$ and $\det(R) = 1$ (*i.e.* for rotation transformations R with respect to $\flat_M$).

■ **Example 4.11** (4.98) is isotropic. ■

Since $\flat_M \hat{C} = C$ is self-adjoint, there exists a $\flat_M$-orthonormal basis, or equivalently a $\flat_M$-rotation transformation R, that diagonalizes $\hat{C}$. Therefore, for an isotropic material ($\flat_M$-rotation invariance), $\hat{\Psi}(\hat{C})$ depends only on the eigenvalues of $\hat{C}$. The eigenvalues of $\hat{C}$ are also the squares of the singular values of $F\colon TM \xrightarrow{\text{linear}} T_\phi W$ using $\flat_M$ and $\flat_W$ as the inner product structures for the singular value decomposition.

Theorem 4.12 A material is isotropic if and only if $\hat{\Psi}(\hat{C})$ is a function of the eigenvalues $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ of $\hat{C}$ modulo permutation.

A general function of n variables $\{\lambda_1, \dots, \lambda_n\}$ that is invariant under their permutation can be expressed as a function of the coefficients of the polynomial whose roots are $\{\lambda_1, \dots, \lambda_n\}$. Therefore, Theorem 4.12 is equivalent to the following Theorem 4.13.

Definition 4.16 — Invariants of an endomorphism. The coefficients $I_1, \dots, I_n$ of the characteristic polynomial of $\hat{C}$

$$\mathcal{P}_{\hat{C}}(z) = \det(zI - \hat{C}) = z^n - I_1 z^{n-1} + I_2 z^{n-2} - \cdots + (-1)^n I_n \tag{4.103}$$

are called the **principal invariants** of $\hat{C}$. The trace of the first n powers of $\hat{C}$

$$J_k \coloneqq \operatorname{tr}(\hat{C}^k), \quad k = 1, \dots, n \tag{4.104}$$

are called the **main invariants** of $\hat{C}$. When the Hencky strain $\hat{H} = \frac{1}{2}\ln(\hat{C})$ (4.97) is our main variable for material modeling, we also consider the principal invariants and the main invariants of $\hat{H}$. They are denoted by lower case letters:

$$\det(zI - \hat{H}) =: z^n - i_1 z^{n-1} + i_2 z^{n-2} - \cdots + (-1)^n i_n \tag{4.105}$$

$$j_k \coloneqq \operatorname{tr}(\hat{H}^k). \tag{4.106}$$

Theorem 4.13 — Principal invariants. A material is isotropic if and only if $\hat{\Psi}(\hat{C})$ can be written as

$$\hat{\Psi}(\hat{C}) = w(I_1, \dots, I_n) \tag{4.107}$$

where $I_1, \dots, I_n$ are the principal invariants of $\hat{C}$, and $w\colon \mathbb{R}^n \to \mathbb{R}$ is some function.

R Similarly, isotropic materials can be modeled with $\hat{\Psi}_{\log}(\hat{H}) = w(i_1, \dots, i_n)$.

The following example demonstrates that one can perform change of variables to convert any functions of $\operatorname{tr}(\hat{C}^k)$ (even when $k > n$) into an expression of the principal invariants.

■ **Example 4.12** In 3D, the 3 principal invariants of the Cauchy–Green endomorphism $\hat{C}$ are

$$I_1 = \lambda_1 + \lambda_2 + \lambda_3 = \operatorname{tr}(\hat{C}) \tag{4.108}$$

$$I_2 = \lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3 = \tfrac{1}{2}(\operatorname{tr}(\hat{C})^2 - \operatorname{tr}(\hat{C}^2)) \tag{4.109}$$

$$I_3 = \lambda_1\lambda_2\lambda_3 = \det(\hat{C}). \tag{4.110}$$

Conversely,

$$\operatorname{tr}(\hat{C}) = I_1 \tag{4.111}$$
$$\operatorname{tr}(\hat{C}^2) = I_1^2 - 2I_2 \tag{4.112}$$
$$\operatorname{tr}(\hat{C}^3) = I_1^3 - 3I_1 I_2 + 3I_3 \tag{4.113}$$

where the 3rd equation is obtained by taking trace on the **Cayley–Hamilton Theorem**

$$\hat{C}^3 - I_1 \hat{C}^2 + I_2 \hat{C} - I_3 I = 0. \tag{4.114}$$

Higher moments such as $\operatorname{tr}(\hat{C}^4)$ can be obtained by repeated substitutions using the Cayley–Hamilton equation: $\operatorname{tr}(\hat{C}^4) = \operatorname{tr}(\hat{C}(\hat{C}^3)) = \operatorname{tr}(\hat{C}(I_1\hat{C}^2 - I_2\hat{C} + I_3 I))$.

$$\operatorname{tr}(\hat{C}^4) = I_1^4 - 4I_1^2 I_2 + 4I_1 I_3 + 2I_2^2, \tag{4.115}$$
$$\operatorname{tr}(\hat{C}^5) = I_1^5 - 5I_1^3 I_2 + 5I_1^2 I_3 + 5I_1 I_2^2 - 5I_2 I_3. \tag{4.116}$$

In general,

$$\operatorname{tr}(\hat{C}^k) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ I_3 & -I_2 & I_1 \end{bmatrix}^k \begin{bmatrix} 1 \\ I_1 \\ I_1^2 - 2I_2 \end{bmatrix}. \tag{4.117}$$

Using power series, any expression $\operatorname{tr}(f(\hat{C}))$, for any analytic function f, can be written as an analytic function of I_1, I_2, I_3. ∎

A useful tip about taking the derivative of the determinant of a matrix with respect to a matrix (endomorphism) is

$$\frac{\partial \det(A)}{\partial A} = \operatorname{cof}(A) = \det(A) A^{*-1}. \tag{4.118}$$

Applying this identity to our characteristic polynomial we get

$$\frac{\partial \mathcal{P}_{\hat{C}}(z)}{\partial \hat{C}} = -\mathcal{P}_{\hat{C}}(z)(zI - \hat{C})^{*-1} \tag{4.119}$$

and

$$\frac{\partial I_k}{\partial \hat{C}} = \frac{1}{k!}\frac{d^k}{dz^k}\bigg|_{z=0} \frac{\partial \mathcal{P}_{\hat{C}}(z)}{\partial \hat{C}} = -\frac{1}{k!}\frac{d^k}{dz^k}\bigg|_{z=0} \left(\mathcal{P}_{\hat{C}}(z)(zI - \hat{C})^{*-1}\right). \tag{4.120}$$

This allows us to compute the 2nd Piola–Kirchhoff stress for a general function $w(I_1, \ldots, I_n)$ of the invariants:

$$\frac{\partial \Psi(C)}{\partial C} = \sharp_M \frac{\partial \hat{\Psi}(\hat{C})}{\partial \hat{C}} \rho_M, \quad \frac{\partial \hat{\Psi}(\hat{C})}{\partial \hat{C}} = \sum_{k=1}^{n} \frac{\partial I_k}{\partial \hat{C}} \frac{\partial w}{\partial I_k}. \tag{4.121}$$

■ **Example 4.13 — Neo-Hookean.**

$$\hat{\Psi}(\hat{C}) = w(I_1, \ldots, I_n) = \frac{\hat{\mu}}{2}(I_1 - n - \ln I_n) + \frac{\hat{\lambda}}{2}(\sqrt{I_n} - 1)^2. \tag{4.122}$$

■

The following are special model one can consider in the framework of hyperelasticity models. They are used in conformal mappings in geometry processing and fluid simulations. Although they are usually not discussed together with elastic solid simulation, we can describe them in our current discussion about the effects of stress–strain energy designs.

■ **Example 4.14 — Conformal energy.** In 2D ($n = 2$), an energy that only penalizes conformal distortion but not area distortion takes the form

$$\hat{\Psi}_{\log}(\hat{H}) = w(i_1, i_2) = \tilde{w}\left(i_1^2 - 2i_2\right) \tag{4.123}$$

In terms of the singular values $s_1 = \sqrt{\lambda_1}, s_2 = \sqrt{\lambda_2}$ of the deformation gradient F with respect to $\flat_M$ and $\flat_W$,

$$Q = i_1^2 - 2i_2 = \ln\left(\frac{s_1}{s_2}\right) \tag{4.124}$$

which is also known as the **dilatation** or the **quasiconformal error**. A **quasiconformal map** is when the maximal conformal distortion, for example $\sup_M(\frac{1+e^Q}{1-e^Q} = \coth\frac{-Q}{2})$ over the domain M, is bounded or minimized.

A deformation is **conformal** or **conformally constrained** if $Q = 0$ everywhere. When a deformation is conformal, F is locally a scaled isometry (scale rotation) $C = \phi^*\flat_W = e^{2u}\flat_M$ for some conformal factor e^{2u}. The conformal energy and the conformal constraint will induce a "conformal stress" that restores conformality, which are derived in [Soliman et al.(2021)]. Under the conformal constraint, the flow map ϕ is still free to deform conformally, and the energy is insensitive to those deformations. ■

■ **Example 4.15 — Volumetric energy.** In nD, an energy that only penalizes volume change but on any other mode of deformation takes the form

$$\hat{\Psi}(\hat{C}) = w(I_1, \ldots, I_n) = w(I_n). \tag{4.125}$$

Note that the nth principal invariant of $\hat{C}$ is $I_n = J^2$ where $J = (\phi^*\mu_W)/(\mu_M)$ where μ_W and μ_M are the volume forms with respect to the metrics $\flat_W$ and $\flat_M$ respectively. There are two notable consequences of this design of energy. First, as we will derive in Chapter 5, the Cauchy stress tensor will always be *a scalar times the identity*. This scalar is called the **pressure**. Second, *the energy is completely insensitive to any volume-preserving deformation*; that is, the material is unable to create any restoration force on other modes of deformation, such as volume-preserving stretching and shearing. When a material behaves like this, we

call it a **fluid**.

Inviscid barotropic compressible fluids are in fact exactly the same as a hyperelastic body with the assumption of (4.125). If the energy (4.125) is stiff, producing a **volumetrically constrained** system, then the flow map can only be a volume preserving one. Note that the material becomes rigid in its volume, but still free to shear and stretch. In that limit, we obtain an **incompressible fluid**. The governing equation of motion becomes the **incompressible Euler equation**.

Perhaps a philosophical take-away is: The idea of fluid and the concept of pressure can be derived from a minimalistic set of postulates, in contrast to phenomenologically reasoned. ∎

4.8 Conclusion

Exterior calculus naturally arises in continuum mechanics when we derive the equations of motion from first principle and keeping track of the types of objects. The *back-propagation* method that we employed for the derivation demands the distinction between vectors and their duals. The result is that we obtain stress tensors (1st Piola–Kirchhoff stress or cauchy stress) as covector-valued $(n-1)$-forms. This fits very well with the original physical intuition of a stress tensor: there is an assignment of force on every infinitesimal surface. The language of exterior calculus allows us to accurately describe this delicate tensor type and distinguish it from other matrices of the same array size. Moreover, changes of coordinates between the material space and world space use only one unified concept: pullback. Using the commutativity between pullbacks and exterior derivatives, we immediately obtain Theorem 4.10 without using any vector identity.

Overall, there is much less assumption and identities, and there is a more fine-grained understanding of formulas in elasticity and symmetry of stress tensor revealed by the exterior calculus language.

5. Fluid Dynamics

In this chapter, we study the motion of fluids. Similar to the philosophy of Chapter 4, we can characterize Newtonian fluids from a minimalistic set of postulates, and subsequently derive and study the equations of motion based on first principles. Our study will primarily concentrate on non-dissipative (inviscid) fluids. Hence, we only need to look at the kinetic and potential energy of the physical system.

The following postulates follows Example 4.15.

Postulate 5.1 — Inviscid barotropic fluid. An **inviscid barotropic fluid** is an isotropic hyperelastic body (Postulates 4.1–4.4) whose internal energy $\hat{\Psi}(\hat{C})$ is only a function of the volume change $\hat{\Psi}(\hat{C}) = w(\det(\hat{C}))$.

In the limit that the internal energy in response to a volume change is infinitely stiff, the deformable body becomes volumetrically rigid for every material element. In that case, a barotropic fluid becomes incompressible.

Postulate 5.2 — Inviscid incompressible fluid. An inviscid incompressible fluid is a volume-constrained ($\det(\hat{C})$ is time-independent) barotropic fluid (Postulate 5.1).

R "Barotropic" means the compressible fluid has its internal energy, and consequently the pressure, depend only on the change of fluid density (or the change of specific volume). A more general compressible fluid, such as an *isentropic fluid* governing gas dynamics, will have its internal energy depending on other hidden variables such as entropy.

R For fluids, Postulate 4.3 can be simplified to that the material space has a rest volume form μ_M, instead of a full-blown rest metric tensor $\flat_M$.

We first derive the equations of motion for these systems, known as **Euler equations**. These classical equations describe the motion of fluid's velocity as a *vector-valued 0-form*. Then, we move on to looking at the same equations in their *covector-based variants*, *i.e.* they describe the motion of fluid's velocity as a *1-form*. We can arrive at the covector-based Euler equation either by changes of variables at the PDE level, or from first principles. We show how this covector-based fluid equation provides visuals and computational efficiency advantages for fluid simulations.

5.1 Euler equations

We follow the same principles established in the previous chapter and study our dynamical system using its kinetic and potential energy, and its associated Euler-Lagrange equation to prove the equations of motion.

5.1.1 Vector-based Euler equations

Following Postulate 4.2, the kinetic energy for a flow map of a continuum is given by $\mathcal{K}(\dot{\phi}) = \int_M \frac{1}{2}|\dot{\phi}|^2_{\flat_W}\rho_M$, and together with the potential energy $\mathcal{U}(\phi) = \int_M \Psi(C)$, $C = \phi^{*\flat_W} = F^{*\flat_W}F$, we can find the equation of motion as in (4.2):

$$\frac{d}{dt}\frac{\partial \mathcal{K}}{\partial \dot{\phi}} = -\frac{\partial \mathcal{U}}{\partial \phi}. \tag{5.1}$$

Using (4.75), and assuming no external forces, the equation simplifies to:

$$\rho_M \overset{\nabla}{\ddot{\phi}} = \underset{(n-1)\text{-form}}{d^\nabla P} \tag{5.2}$$

whre $P = 2\flat_W F\frac{\partial \Psi}{\partial C}$. As demonstrated in Example 4.7, in 3D Cartesian space, the equation is further simplified to:

$$\rho_M \overset{\nabla}{\ddot{\phi}} = \nabla \cdot \mathbf{P}, \tag{5.3}$$

where $\mathbf{P} = \mathbf{FS}$, $\mathbf{S} = \frac{\partial \Psi}{\partial \mathbf{C}}$, $\mathbf{F} = \nabla\phi$, and $\mathbf{C} = \mathbf{F}^\intercal\mathbf{F}$.

Now let us substitute the assumptions of Postulate 5.1 (Example 4.15) for fluids. Specifically, the potential energy only depends on its top-dimensional principal invariant, or equivalently $\hat{\Psi}(\hat{C}) = w(I_3) = w(J^2)$ in 3D, where $J = \det(\mathbf{F})$. In the coordinate-free language, $J = \det(F) = \frac{\phi^*\mu_W}{\mu_M}$ where μ_W and μ_M are the volume forms on W and M respectively.

Definition 5.1 We denote the mass form in the world coordinate as $\rho \in \Omega^n(W)$, which given by the relation

$$\rho_M =: \underset{n\text{-form}}{\phi^* \rho}. \tag{5.4}$$

Define the material-space and world-space mass density 0-forms by

$$q_M \coloneqq \frac{\rho_M}{\mu_M} \in \Omega^0(M), \quad q \coloneqq \frac{\rho}{\mu_W} \in \Omega^0(W). \tag{5.5}$$

One can check that

$$\underset{\text{0-form}}{\phi^*}\ q = q \circ \phi = \frac{q_M}{J}. \tag{5.6}$$

Theorem 5.1 Under the assumption of Postulate 5.1, the Cauchy stress tensor takes the form of a scalar times identity

$$\sigma = -pI \otimes \mu_W, \quad p \in \Omega^0(W) \tag{5.7}$$

We call p the **pressure**.

Proof. Using (4.118), the derivative of $\hat{\Psi}(\hat{C}) = w(I_3)$ with respect to $\hat{C}$ is given by $\frac{\partial \hat{\Psi}}{\partial \hat{C}} = w'(I_3)\mathrm{cof}(\hat{C}) = w'(J^2)J^2\hat{C}^{*-1} = w'(J^2)J^2 \flat_M C^{-1}$. Then, by (4.121) we have the 2nd Piola stress tensor given by

$$\begin{aligned} S &= 2\frac{\partial \Psi(C)}{\partial C} = 2\sharp_M w'(J^2)J^2 \flat_M C^{-1}\rho_M && (5.8)\\ &= 2w'(J^2)J^2 C^{-1}\rho_M && (5.9)\\ &= 2w'(J^2)J^2 F^{-1}\sharp_W F^{*-1}\rho_M && (5.10) \end{aligned}$$

Transforming the 2nd Piola–Kirchhoff stress tensor to the 1st Piola–Kirchhoff and subsequently Cauchy stress tensor, we get

$$\begin{aligned} P &= \flat_W FS = 2w'(J^2)J^2 F^{*-1}\rho_M && (5.11)\\ P &= \underset{\text{2-form}}{\phi^*}\ \sigma \implies \sigma = 2w'(J^2)J(I \otimes \rho) && (5.12) \end{aligned}$$

We complete the proof by defining

$$p \coloneqq -2w'(J^2)Jq. \tag{5.13}$$

□

Definition 5.2 — Fluid velocity. We define $\vec{u} = \mathbf{u} \in \Gamma(TW)$ as the **fluid velocity**, defined by

$$\dot{\phi} = \mathbf{u} \circ \phi = \underset{\text{0-form}}{\phi^*}\ \mathbf{u} \tag{5.14}$$

We use the notation $\mathbf{u}$ and $\vec{u}$ interchangeably, both of which represent the vector type or the vector-valued 0-form type.

Also define

$$\eta \coloneqq \flat_W \mathbf{u} \in \Omega^1(W) \tag{5.15}$$

as the **fluid velocity covector**.

Theorem 5.2 — Barotropic Euler equation. Under the general equation of motion (5.2), the velocity field $\mathbf{u}$ satisfies

$$\frac{\partial q}{\partial t} + \operatorname{div}(q\mathbf{u}) = 0, \tag{5.16}$$

$$\frac{\partial \mathbf{u}}{\partial t} + \nabla_{\mathbf{u}}\mathbf{u} = -\frac{\operatorname{grad} p}{q}, \tag{5.17}$$

where p is a function of q derived from (5.13).

Proof. The continuity equation (5.16) follows from ρ_M being time-independent in terms of Lie derivative (2.134).

Noting that $P = \underset{\text{2-form}}{\phi^*}\ \sigma$ and $\sigma = -pI \otimes \mu_W$, we can write the equation of motion as:

$$\rho_M \overset{\nabla}{\ddot{\phi}} = d^\nabla P = d^\nabla \underset{\text{2-form}}{\phi^*}\ \sigma = \underset{\text{3-form}}{\phi^*}\ d^\nabla\sigma, \tag{5.18}$$

$$= \underset{\text{3-form}}{\phi^*}\ (-\operatorname{grad} p \otimes \mu_W) = -(\underset{\text{3-form}}{\phi^*}\ \mu_W)(\underset{\text{0-form}}{\phi^*}\ \operatorname{grad} p) \tag{5.19}$$

$$= -J\mu_M(\operatorname{grad} p \circ \phi). \tag{5.20}$$

Recall that $\dot{\phi} = \mathbf{u} \circ \phi = \underset{\text{0-form}}{\phi^*}\ \mathbf{u}$. From Corollary 2.29, we can interpret the second derivative of the flow map ϕ as the Lie material derivative of the velocity vector field as a 0-form:

$$\overset{\nabla}{\ddot{\phi}} = \frac{\nabla}{\partial t}\Big(\underset{\text{0-form}}{\phi^*}\ \mathbf{u}\Big) = \underset{\text{0-form}}{\phi^*}\Big(\frac{\partial \mathbf{u}}{\partial t} + \underset{\text{0-form}}{\mathscr{L}^\nabla_{\mathbf{u}}}\ \mathbf{u}\Big). \tag{5.21}$$

Further, note that $\frac{J\mu_M}{\rho_M} = \frac{J}{q_M} = \underset{\text{0-form}}{\phi^*}\ \frac{1}{q}$ using (5.6). This allows us to now write the Euler equations in the Eulerian coordinate:

$$\frac{\partial \mathbf{u}}{\partial t} + \nabla_{\mathbf{u}}\mathbf{u} = -\frac{\operatorname{grad} p}{q}. \tag{5.22}$$

□

R The formulation of the Euler equation through the above process describes the conservation law on linear momentum.

For the remainder of this chapter, we assume that $q = 1$ to reduce the system into an incompressible fluid system.

Corollary 5.3 — Incompressible Euler equation. Under Postulate 5.2, the velocity field $\mathbf{u}$ satisfies

$$(\operatorname{div}\mathbf{u})\mu_W = \underset{n\text{-form}}{\mathscr{L}_{\mathbf{u}}}\ \mu_W = 0, \tag{5.23}$$

$$\frac{\partial \mathbf{u}}{\partial t} + \nabla_{\mathbf{u}}\mathbf{u} = -\operatorname{grad} p, \tag{5.24}$$

5.1.2 Covector Euler equations as a constrained minimization

One may re-derive the equations of motion in terms of the velocity covector field. For this, we formulate the action minimization in terms of the Eulerian coordinate energy as follows:

$$\begin{aligned} &\underset{\mathbf{u},\psi}{\text{minimize}}\ S(\mathbf{u}) = \int_W \frac{1}{2}|\mathbf{u}|^2\mu_W \\ &\text{subject to } \dot{\psi} + \underset{0\text{-form}}{\mathscr{L}_{\mathbf{u}}}\ \psi = 0 \text{ and } \underset{n\text{-form}}{\mathscr{L}_{\mathbf{u}}}\ \mu = 0. \end{aligned} \tag{5.25}$$

Here, ψ is the inverse flow map given by $\psi \circ \phi = \underset{0\text{-form}}{\phi^*}\ \psi = \mathrm{id}$. The first constraint, referred to as **Lin's Constraint**, comes from the Lie material derivative for the 0-form backward flow map $\dot{\psi} + \underset{0\text{-form}}{\mathscr{L}_{\mathbf{u}}}\ \psi = 0$, and the second constraint enforces incompressibility for the fluid. Rewriting the action using Lagrange multipliers we get to:

$$S(\mathbf{u},\psi) = \int_W \frac{1}{2}|\mathbf{u}|^2\mu_W - \langle \eta_0 | \dot{\psi} + \mathscr{L}_{\mathbf{u}}\psi\rangle\mu_W - \langle\lambda|\,\mathscr{L}_{\mathbf{u}}\,\mu_W\rangle, \tag{5.26}$$

where $\lambda \in \Omega^0(W)$ and $\eta_0 \in \Gamma(T^*_\psi M)$ are the Lagrange multipliers for the incompressibility and Lin's constraint, respectively. By taking variations in terms of $\mathbf{u}$ and ψ we can find the following equations of motion:

$$\eta = \underset{1\text{-form}}{\psi^*}\ \eta_0 - d\lambda, \tag{5.27}$$

$$\frac{\partial \eta_0}{\partial t} = 0. \tag{5.28}$$

where $\eta = \mathbf{u}^\flat$. Using Theorem 2.28, we can take the time derivative of the above to the covector-based Euler equations:

$$\dot{\eta} + \underset{1\text{-form}}{\mathscr{L}_{\mathbf{u}}}\ \eta = -d\lambda. \tag{5.29}$$

R The covector formulation of the Euler equation describes a conservation law on the circulation of velocity 1-form.

Consistency between (5.29) and (5.24)

We can check that (5.29) is consistent with (5.24) as follows. As shown in Example 2.21, we can write the Lie material derivative of the velocity 1-form $\eta = \mathbf{u}^\flat$ for 3D Cartesian space:

$$\frac{\partial \eta}{\partial t} + \underset{\text{1-form}}{\mathscr{L}_{\mathbf{u}}}\, \eta = \left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} + (\nabla \mathbf{u})^\intercal \mathbf{u}\right)^\flat. \tag{5.30}$$

To get to the Euler equation for fluid dynamics, we can add a modified gradient pressure term $\lambda = p - \frac{1}{2} i_{\mathbf{u}} \eta = p - \frac{1}{2}|\mathbf{u}|^2$ to both sides.

$$\begin{aligned} \frac{\partial \eta}{\partial t} + \underset{\text{1-form}}{\mathscr{L}_{\mathbf{u}}}\, \eta + d\lambda &= \left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} + (\nabla \mathbf{u})^\intercal \mathbf{u} + \nabla \lambda\right)^\flat && (5.31) \\ &= \left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} + \nabla p\right)^\flat = 0. && (5.32) \end{aligned}$$

Corollary 5.4 — Vorticity equation. By taking d on (5.29), we obtain

$$\dot{\omega} + \underset{\text{2-form}}{\mathscr{L}_{\mathbf{u}}}\, \omega = 0, \tag{5.33}$$

where

$$\omega = d\eta \tag{5.34}$$

is called the **vorticity 2-form**.

5.2 Advantages of Covector Euler Equation

While the covector-based equation of motion is equivalent to the vector-based equation, it can provide alternative advantages that are desirable for the animation and simulation of fluid dynamics.

The common convention in fluid solvers is to freeze the velocity $\mathbf{v} \leftarrow \mathbf{u}$, and split the Euler equations in two steps of advection and projection. These steps, however, prove to induce additional numerical dissipation in the simulation which can hinder the computational performance and visual fidelity of the results. We can see that the covector-based formulation provides a formulation that can significantly improve the visual fidelity of the results while reducing numerical dissipation.

Given a frozen background velocity field $\mathbf{v}$, and splitting the Euler equations of motion (5.17) and (5.29), we get to the following transport equations for covector-based transport:

$$\dot{\eta} + \underset{\text{1-form}}{\mathscr{L}_{\mathbf{v}}}\, \eta = 0. \tag{5.35}$$

and for vector-based transport equation:

$$\dot{\mathbf{u}} + \mathbf{v} \cdot \nabla \mathbf{u} = 0. \tag{5.36}$$

Covector-based transport in vector calculus

From Example 2.21, we can transform (5.35) to a vector calculus equivalent of:

$$\dot{\mathbf{u}} + \mathbf{v} \cdot \nabla \mathbf{u} = -(\nabla \mathbf{v}) \cdot \mathbf{u}. \tag{5.37}$$

This form reveals how the covector-based Euler equation involves a stretching term that was not present in the vector-based transport (5.36).

This equation is also the basis for the *impulse methods* [Buttke(1993), Cortez(1995), Feng et al.(2022)]. In the literature, (5.35) is treated as a non-trivial dynamical system for each flowing particle $\frac{D\mathbf{u}}{Dt} = -(\nabla \mathbf{v}) \cdot \mathbf{u}$. The standard method for solving this PDE is to further split the transport equation which brings about additional instability due to the right-hand side stretching term.

Solving covector-based transport without inducing instability from the stretching term

To remedy the problems associated with solving (5.37), we can instead study (5.35). The solution to this PDE, as stated in Corollary 2.29, is given by:

$$\eta_t = e^{-\mathscr{L}_{\mathbf{v}}\, t} \eta_0, \tag{5.38}$$

$$= \psi_t^* \eta_0. \tag{5.39}$$

In 3D Cartisen domain (see Figure 5.1), this is simply

$$\mathbf{u}_t = (\nabla \psi_t)^\intercal \mathbf{u} \circ \psi_t. \tag{5.40}$$

Note the difference to the solution of vector-based Euler equation which is the component-wise Lie-material derivative of a vector-valued 0-form $\dot{\mathbf{u}} + \underset{\text{0-form}}{\mathscr{L}_{\mathbf{v}}\, \mathbf{u}} = 0$ given by:

$$\mathbf{u}_t = \mathbf{u}_0 \circ \psi_t. \tag{5.41}$$

As opposed to the **Impulse methods**, the advection through the pullback operator does not need the splitting of the stretching term, which allows for a more stable transport equation.

Conservation of Circulation

Kelvin's conservation of circulation [Thomson(1868), Frisch and Villone(2014)] states that the circulation $\oint_C \mathbf{u} \cdot d\mathbf{l}$ or equivalently $\oint_C \eta_t$, around every closed loop is conserved. Since the solution to the covector-based transport is simply the pullback of the 1-form field $\eta_t = \psi_t^* \eta_0$, this conservation law follows naturally according to Theorem 2.1. Concretely:

$$\oint_C \psi_t^* \eta_0 = \oint_{\psi(C)} \eta_0. \tag{5.42}$$

As opposed to the vector-based formulation of Euler equations which results in a conservation of linear momentum, the covector-based formulation leads to a conservation of circulation.

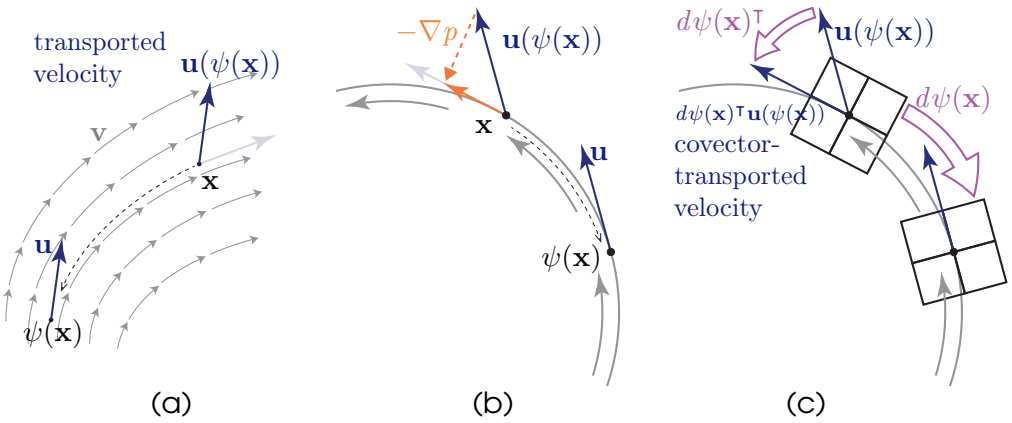


Figure 5.1 (a) In a standard fluid solver, the velocity $\mathbf{u}$ is transported by the flow $\mathbf{v}$ using an inverse flow map ψ. (b) This process can turn a rotation motion into a divergent one which is subsequently damped by the pressure projection. (c) Our advection method maintains the vorticity by a multiplication with the transposed Jacobian of the inverse flow map.

For simulating vortex-dominant phenomena, computations based on the conservation of circulation [Elcott et al.(2007), Nabizadeh et al.(2022)] are more advantageous. The difference between the two motion laws (5.17) and (5.29) becomes apparent in a flow dominated by vorticity. In (5.17), the pressure force is responsible for the concentration of vorticity. Without the pressure, the inertial motion (conservation of linear momentum) turns into a centrifugal force that makes vortices disintegrate. In (5.29), the Lagrangian pressure plays *no* role in the persistence of vorticity. The mechanism for the conservation of vorticity is entirely encoded in the left-hand side of the equation. Such a property makes an algorithm based on time-splitting into advection and pressure steps especially appealing.

Equivalence to a vortex method

The traditional advection-projection method introduces a splitting error that destroys vorticity [Zhang et al.(2015), Zehnder et al.(2018)]. This phenomenon arises solely from the advection step, since the projection step only modifies the velocity $\mathbf{u}$ with a pure gradient which leads to no change in its curl.

Theorem 5.5 — Vector-based transport as a vortex method. Under the classical advection equation $(\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla)\mathbf{u} = \mathbf{0}$, the vorticity $\mathbf{w} = \nabla\times\mathbf{u}$ evolves according to[a]

$$\frac{\partial}{\partial t}\mathbf{w} + \mathbf{v}\cdot\nabla\mathbf{w} - \mathbf{w}\cdot\nabla\mathbf{v} = \langle\nabla\mathbf{u}\times\nabla\mathbf{v}\rangle. \tag{5.43}$$

This *modified vorticity equation* deviates from the correct vorticity equation by a term $\langle\nabla\mathbf{u}\times\nabla\mathbf{v}\rangle$.

[a] In index notation $\langle\nabla\mathbf{u}\times\nabla\mathbf{v}\rangle_i = \epsilon_{ijk}\partial_j u^\ell \partial_k v^\ell$.

Proof. The vector-based transport equation $(\frac{\partial}{\partial t} + \mathbf{v}\cdot\nabla)\mathbf{u} = \mathbf{0}$, is equivalent to the following equation $\frac{\partial}{\partial t}\eta + \mathscr{L}_{\mathbf{v}}\,\eta = \langle\nabla\mathbf{v},\mathbf{u}\rangle$ using the velocity covector $\eta = \mathbf{u}^\flat$. Taking

d on both sides of the equation yields

$$\begin{aligned}&\tfrac{\partial}{\partial t}\omega + \mathscr{L}_{\mathbf{v}}\,\omega = d\langle \nabla\mathbf{v}, \mathbf{u}\rangle = \langle \underbrace{d^{\nabla}\nabla\mathbf{v}}_{=0}, \mathbf{u}\rangle - \langle \nabla\mathbf{v} \wedge \nabla\mathbf{u}\rangle \\ &= \langle \nabla\mathbf{u} \wedge \nabla\mathbf{v}\rangle,\end{aligned}$$

where the first identity is the Leibniz rule with metric $d\langle \mathbf{A} \wedge \mathbf{B}\rangle = \langle d^{\nabla}\mathbf{A} \wedge \mathbf{B}\rangle + (-1)^k\langle \mathbf{A} \wedge d^{\nabla}\mathbf{B}\rangle$, and the term $d^{\nabla}\nabla\mathbf{v}$ is zero since there is no curvature in flat space (*i.e.* $d^{\nabla}d^{\nabla} = 0$). Converting 2-forms to vector fields in 3D using $\omega = i_{\mathbf{w}}\mu$, we obtain

$$\tfrac{\partial}{\partial t}\mathbf{w} + \nabla_{\mathbf{v}}\mathbf{w} - \nabla_{\mathbf{w}}\mathbf{v} = \langle \nabla\mathbf{u} \times \nabla\mathbf{v}\rangle.$$

□

Theorem 5.6 — Covector-based transport as a vortex method. The evolution of $\mathbf{w} = \nabla \times \mathbf{u}$ that undergoes (5.37) is

$$\tfrac{\partial}{\partial t}\mathbf{w} + \mathbf{v} \cdot \nabla\mathbf{w} - \mathbf{w} \cdot \nabla\mathbf{v} = \mathbf{0} \tag{5.44}$$

which is the correct vorticity equation.

Proof. As shown in Corollary 2.29, Eq. (5.37) is equivalent to (5.35), i.e. $\frac{\partial}{\partial t}\eta + \mathscr{L}_{\mathbf{v}}\,\eta = 0$, given that $\eta = \mathbf{u}^{\flat}$. Applying d to (5.35), and using the commutativity between d and $\mathscr{L}_{\mathbf{v}}$, we obtain $\frac{\partial}{\partial t}\omega + \mathscr{L}_{\mathbf{v}}\,\omega = 0$ for the *vorticity 2-form* $\omega = d\eta$. In 3D, the relationship between the 2-form $\omega = d\mathbf{u}^{\flat}$ and the vector field $\mathbf{w} = \operatorname{curl}\mathbf{u}$ is given by $\omega = i_{\mathbf{w}}\mu$ where $\mu \in \Omega^3(M)$ is the volume form. Using this relationship we obtain

$$\begin{aligned}0 &= \tfrac{\partial}{\partial t}\omega + \mathscr{L}_{\mathbf{v}}\,\omega = \tfrac{\partial}{\partial t}(i_{\mathbf{w}}\mu) + \mathscr{L}_{\mathbf{v}}(i_{\mathbf{w}}\mu) \\ &= i_{\partial\mathbf{w}/\partial t}\mu + i_{[\mathbf{v},\mathbf{w}]}\mu + i_{\mathbf{w}}\underbrace{\mathscr{L}_{\mathbf{v}}\,\mu}_{=0} \quad (\mathscr{L}_{\mathbf{v}}\,\mu = 0 \text{ since } \operatorname{div}\mathbf{v} = 0) \\ &= i_{(\partial\mathbf{w}/\partial t + [\mathbf{v},\mathbf{w}])}\mu,\end{aligned}$$

where in the second line we used the identity of Lie derivative on contraction $\mathscr{L}_{\mathbf{v}}(i_{\mathbf{w}}\alpha) = i_{[\mathbf{v},\mathbf{w}]}\alpha + i_{\mathbf{w}}(\mathscr{L}_{\mathbf{v}}\,\alpha)$. Therefore, $\frac{\partial}{\partial t}\mathbf{w} + [\mathbf{v},\mathbf{w}] = \mathbf{0}$. Finally, we arrive at (5.44) by substituting $[\mathbf{v},\mathbf{w}] = \nabla_{\mathbf{v}}\mathbf{w} - \nabla_{\mathbf{w}}\mathbf{v}$ under Theorem 2.27. □

By advancing $\mathbf{u}$ via the covector transportation (5.35) or (5.37), we implicitly solve the vorticity equation (5.44), which is the modeling equation for vortex methods. The ability to solve (5.44) at the velocity level without using the vorticity variable is significant. Previous vortex methods which solve (5.44) have to include an expensive integration that converts vorticity back to velocity.

Commutativity between covector transportation and pressure projection

In a traditional fluid solver, the splitting error between traditional advection and the projection arises because two operations do not commute. Here we show that advection and projection *commute* in CF. This property of CF fundamentally removes the splitting error of these two operations.

Consider the equivalence classes of $\Omega^1(M)$ where $[\xi] = [\eta]$ whenever $\xi - \eta = d\varphi$ for some function φ. This is a natural abstraction for our discussion since two covector fields are equivalent if and only if they share the same pressure projection result.

The pressure projection can be understood as extracting the unique divergence-free representative in each equivalence class $[\eta] \in \Omega^1(M)/\operatorname{im}(d)$.

Now consider two covector fields $\xi_0, \eta_0 \in \Omega^1(M)$, divergence-free or not, and transport them by the covector advection equation (5.38) to obtain ξ_t, η_t respectively. Then

$$[\xi_0] = [\eta_0] \quad \text{if and only if} \quad [\xi_t] = [\eta_t]. \tag{5.45}$$

To see this assertion, use (5.38) to express $\xi_t - \eta_t$ as the pullback of $\xi_0 - \eta_0$ by the inverse flow map

$$\xi_t - \eta_t = \psi_t^*(\xi_0 - \eta_0) \tag{5.46}$$

and apply Theorem 2.11 so that the pullback of an exact differential is still exact. Therefore, whether one first projects then advects, advects then projects, or inserts a projection (or reflection [Zehnder et al.(2018)]) at the halfway point of the advection, one will obtain covector fields all in the same equivalence class and hence the same divergence-free representative.

Note that the same argument does not apply to the traditional advections. The transportation with a simple value look-up (5.41) generally turns an exact gradient vector field into a non-gradient field.

Extending CF with a long-time characteristic mapping

The analysis in Section 5.2 implies that we may delay the pressure projection for a longer time (rather than a time step) and just transport the velocity covector field η using a long-time flow map.

Let $\psi_t \colon W_t \to M_0$ be the inverse flow map (Eulerian-to-Lagrangian map) which is the Lagrangian marker carried by the history of the solution

$$\tfrac{\partial}{\partial t}\psi_t + \mathbf{u}_t \cdot \nabla \psi_t = \mathbf{0}, \quad \psi_0 = \mathrm{id}_M\,. \tag{5.47}$$

Then the velocity covector field at the current time t is the pressure projection of

$$\eta_t = \psi_t^* \eta_0. \tag{5.48}$$

In other words, if we maintain a Lagrangian marker ψ_t we obtain the fluid state through a single-step look up. This drastically reduces the amount of interpolation in the advection-projection iteration.

Such a long-time method of characteristic mapping (MCM) is proposed by [Tessendorf and Pelfrey(2011), Sato et al.(2018), Qu et al.(2019)]. We call the variant of the CF method based on (5.47) and (5.48) *CF+MCM*.

R CF+MCM is subtly different from the traditional MCM. The latter requires an accumulation of the pressure gradient over time whereas the former does not. In the traditional MCM [Qu et al.(2019)], one evaluates $\mathbf{u}_t$ by (5.24) integrating over time along particle trajectories:

$$\mathbf{u}_t(\mathbf{x}) = \mathbf{u}_0(\psi_t(\mathbf{x})) + \textstyle\int_0^t (\nabla p_\tau)(\phi_\tau(\psi_t(\mathbf{x})))\, d\tau. \tag{5.49}$$

In CF+MCM, the time integration of (5.29) yields

$$\begin{aligned}\eta_t &= \psi_t^* \eta_0 + \textstyle\int_0^t (\phi_\tau \circ \psi_t)^* d\lambda_\tau \, d\tau \\ &= \psi_t^* \eta_0 + d\left(\textstyle\int_0^t (\phi_\tau \circ \psi_t)^* \lambda_\tau \, d\tau\right).\end{aligned} \tag{5.50}$$

which is a comparable notation to (5.49) corresponds to

$$\begin{aligned}\mathbf{u}_t(\mathbf{x}) &= d\psi_t^\intercal \mathbf{u}_0(\psi_t(\mathbf{x})) \\ &\quad + \textstyle\int_0^t d\psi_t^\intercal \, d\phi_\tau^\intercal (\nabla(p_\tau - \frac{|\mathbf{u}_\tau|^2}{2}))(\phi_\tau(\psi_t(\mathbf{x}))) \, d\tau \\ &= d\psi_t^\intercal \mathbf{u}_0(\psi_t(\mathbf{x})) + \nabla\left(\textstyle\int_0^t (p_\tau - \frac{|\mathbf{u}_\tau|^2}{2})(\phi_\tau(\psi_t(\mathbf{x}))) \, d\tau\right).\end{aligned} \tag{5.51}$$

As omitted in (5.48), the second terms in (5.50) and (5.51) can be absorbed in a single pressure projection as they are exact differentials. This is possible since d (resp. ∇) in (5.50) (resp. (5.51)) can be pulled out of the integral by the commutativity property (*cf.* Theorem 2.11) between d and pullback operators. By contrast in (5.49), the ∇ in the pressure term cannot be pulled out of the time integral. The accumulated pressure $\int_0^t (\nabla p_\tau) \circ \phi_\tau \circ \psi_t \, d\tau$ is generally not an exact gradient. Therefore, a traditional MCM must carefully record the accumulated pressure. This procedure is entirely removed in CF+MCM.

5.3 Conclusion

One can write the equations of motion of fluid mechanics following the same process established in the previous chapter. While this gets us to a vector-based Euler equation following the conservation of linear momentum, with a modified procedure of minimizing action in the Lagrangian frame, one can write the same equations of motion in terms of velocity covector. These equation reveal multiple benefits when it comes to simulating incompressible ideal fluids, since the formulation naturally guarantees the preservation of circulation while involving a simple modification by multiplying the looked up velocity with the inverse transpose of the flow map's jacobian (i.e. $\mathbf{u} \leftarrow (\nabla\psi_t)^\intercal \mathbf{u} \circ \psi_t$ as opposed to $\mathbf{u} \leftarrow \mathbf{u} \circ \psi_t$). Additionally, this formulation completely removes the splitting error between the pressure projection and advection steps as it follows the commutativity between exterior derivative and the pullback operator. The method emulates a vorticity-based Euler equation, which is preferable in vortex-dominated flows, without incurring instability due to the stretching term as the pullback operator is exact solution to the Lie-material derivative for 1-form velocity field. This natural solution also allows for using long-term method of characteristics mapping since the pullback operator remains the exact solution regardless of the duration of the timesteps.

III

Appendix

Bibliography

[Boyd and Vandenberghe(2004)] Stephen P Boyd and Lieven Vandenberghe. 2004. *Convex optimization*. Cambridge university press.

[Buttke(1993)] Tomas F Buttke. 1993. Velicity methods: Lagrangian numerical methods which preserve the Hamiltonian structure of incompressible fluid flow. In *Vortex flows and related numerical methods*. Springer, 39–57.

[Cortez(1995)] Ricardo Cortez. 1995. Impulse-based methods for fluid flow. https://doi.org/10.2172/87798

[Crane(2018)] Keenan Crane. 2018. Discrete differential geometry: An applied introduction. *Notices of the AMS, Communication* (2018), 1153–1159.

[Crane et al.(2013)] Keenan Crane, Fernando De Goes, Mathieu Desbrun, and Peter Schröder. 2013. Digital geometry processing with discrete exterior calculus. In *ACM SIGGRAPH 2013 Courses*. 1–126.

[Elcott et al.(2007)] Sharif Elcott, Yiying Tong, Eva Kanso, Peter Schröder, and Mathieu Desbrun. 2007. Stable, circulation-preserving, simplicial fluids. *ACM Transactions on Graphics (TOG)* 26, 1 (2007), 4–es.

[Evans(1998)] Lawrence C Evans. 1998. *Partial differential equations*. American Mathematical Society.

[Feng et al.(2022)] Fan Feng, Jinyuan Liu, Shiying Xiong, Shuqi Yang, Yaorui Zhang, and Bo Zhu. 2022. Impulse Fluid Simulation. *IEEE Transactions on Visualization and Computer Graphics* (2022).

[Flanders(1963)] Harley Flanders. 1963. *Differential forms with applications to the physical sciences by harley flanders*. Elsevier.

[Frankel(2011)] Theodore Frankel. 2011. *The geometry of physics: an introduction*. Cambridge university press.

[Frisch and Villone(2014)] Uriel Frisch and Barbara Villone. 2014. Cauchy's almost forgotten Lagrangian formulation of the Euler equation for 3D incompressible flow. *The European Physical Journal H* 39, 3 (2014), 325–351.

[Grinspun et al.(2006)] Eitan Grinspun, Mathieu Desbrun, Konrad Polthier, Peter Schröder, and Ari Stern. 2006. Discrete differential geometry: an applied introduction. *ACM Siggraph Course* 7, 1 (2006).

[Nabizadeh et al.(2022)] Mohammad Sina Nabizadeh, Stephanie Wang, Ravi Ramamoorthi, and Albert Chern. 2022. Covector fluids. *ACM Transactions on Graphics (TOG)* 41, 4 (2022), 1–16.

[Nicolet et al.(2021)] Baptiste Nicolet, Alec Jacobson, and Wenzel Jakob. 2021. Large steps in inverse rendering of geometry. *ACM Transactions on Graphics (TOG)* 40, 6 (2021), 1–13.

[Qu et al.(2019)] Ziyin Qu, Xinxin Zhang, Ming Gao, Chenfanfu Jiang, and Baoquan Chen. 2019. Efficient and conservative fluids using bidirectional mapping. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 1–12.

[Sato et al.(2018)] Takahiro Sato, Christopher Batty, Takeo Igarashi, and Ryoichi Ando. 2018. Spatially adaptive long-term semi-Lagrangian method for accurate velocity advection. *Computational Visual Media* 4, 3 (2018), 223–230.

[Sellán et al.(2021)] Silvia Sellán, Jack Luong, Leticia Mattos Da Silva, Aravind Ramakrishnan, Yuchuan Yang, and Alec Jacobson. 2021. Breaking Good: Fracture Modes for Realtime Destruction. *arXiv preprint arXiv:2111.05249* (2021).

[Soliman et al.(2021)] Yousuf Soliman, Albert Chern, Olga Diamanti, Felix Knöppel, Ulrich Pinkall, and Peter Schröder. 2021. Constrained Willmore surfaces. *ACM Transactions on Graphics (TOG)* 40, 4 (2021), 1–17.

[Tessendorf and Pelfrey(2011)] Jerry Tessendorf and Brandon Pelfrey. 2011. The characteristic map for fast and efficient vfx fluid simulations. In *Computer Graphics International Workshop on VFX, Computer Animation, and Stereo Movies. Ottawa, Canada.*

[Thomson(1868)] William Thomson. 1868. On Vortex Motion. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh* 25, 1 (1868), 217–260.

[Wang and Chern(2021)] Stephanie Wang and Albert Chern. 2021. Computing minimal surfaces with differential forms. *ACM Transactions on Graphics (TOG)* 40, 4 (2021), 1–14.

[Yang et al.(2021)] Shuqi Yang, Shiying Xiong, Yaorui Zhang, Fan Feng, Jinyuan Liu, and Bo Zhu. 2021. Clebsch gauge fluid. *ACM Transactions on Graphics (TOG)* 40, 4 (2021), 1–11.

[Yin et al.(2023)] Hang Yin, Mohammad Sina Nabizadeh, Baichuan Wu, Stephanie Wang, and Albert Chern. 2023. Fluid cohomology. *ACM Transactions on Graphics (TOG)* 42, 4 (2023), 1–24.

[Zehnder et al.(2018)] Jonas Zehnder, Rahul Narain, and Bernhard Thomaszewski. 2018. An advection-reflection solver for detail-preserving fluid simulation. *ACM Transactions on Graphics (TOG)* 37, 4 (2018), 1–8.

[Zhang et al.(2015)] Xinxin Zhang, Robert Bridson, and Chen Greif. 2015. Restoring the missing vorticity in advection-projection fluid solvers. *ACM Transactions on Graphics (TOG)* 34, 4 (2015), 1–8.

[Zhao et al.(2020a)] Shuang Zhao, Wenzel Jakob, and Tzu-Mao Li. 2020a. Physics-based differentiable rendering: from theory to implementation. In *ACM SIGGRAPH 2020 Courses*. 1–30.

[Zhao et al.(2020b)] Shuang Zhao, Wenzel Jakob, and Tzu-Mao Li. 2020b. Physics-based differentiable rendering: from theory to implementation. In *ACM siggraph 2020 courses*. 1–30.

Index