A Comprehensive Introduction to DIFFERENTIAL GEOMETRY

VOLUME THREE
Third Edition

MICHAEL SPIVAK

PUBLISH OR PERISH, INC.

Houston, Texas 1999

Publish or Perish, Inc.
www.mathpop.com

Volume 1 ISBN 0-914098-70-5
Volume 2 ISBN 0-914098-71-3
Volume 3 ISBN 0-914098-72-1
Volume 4 ISBN 0-914098-73-X
Volume 5 ISBN 0-914098-74-8

Printed in the United States of America

PREFACE

These final three volumes are regarded as constituting a single volume, with Chapters 1 to 6 in Volume III, Chapters 7 to 9 in Volume IV, and Chapters 10 to 13 in Volume V. After finishing this multi-volume, I felt somewhat like a man who has tried to cleanse the Augean stables with a Johnny-Mop. Leafing through Mathematical Reviews for the past thirty years, and gazing at the dignified tomes which represent the glories of the classical era, one quickly senses that Differential Geometry is a field of overwhelming extent, beyond the comprehension of any mortal. I suppose such lucubrations ought to buoy up one's spirit with admiration for human achievement, but I must confess that they usually lead me instead to a state of brooding melancholy.

Although the strident word "comprehensive" still stands emblazoned in the title, the Bibliography, in Volume V, will begin to give some idea how much as has necessarily been left out. There are also mini-bibliographies in Volumes III and IV for the works explicitly cited there. Problems have been restricted practically to the absolute minimum, basically facts left to the reader as exercises.

As a glance at the table of contents will show, Volume III is essentially a course in classical surface theory, the only difference being that Chapter 1 prepares the ground for applying the intrinsic geometry of Riemannian manifolds, which was discussed in Volume II. Although much space is devoted to classical material, much of the generalized material in Chapter 7 would be almost incomprehensible without the prior treatment of surface theory. The only exception is the second half of Chapter 2 (from page 75 on), which can (and probably should) be omitted completely without loss of continuity. For those who don't care for the motivational twiddle-twaddle, an introduction to "modern" differential geometry can be extracted from Chapters 1, 7 (parts D and E), 8, and 13, with an assist from Chapter 5, and the first halves of Chapters 9 and 12.

References like Theorem 6-3 or 7-2, when quoted in Volumes III or IV, say, refer to Chapter 6 of Volume III, and Chapter 7 of Volume IV, respectively. References to results of Volumes I and II, or page numbers from any other volume, carry an additional Roman numeral, e.g., Theorem I.6-3, or pg. IV.167.

As acknowledged in the first edition, I am grateful to the Sonderforschungsbereich Theoretische Mathematik in Bonn, the University of California at Berkeley, and the following individuals (as well as many readers and others whom I may have inadvertently omitted): R. Bassein, R. Bishop, R. Böhme, D. Bourghelia, E. Calabi, B. Cenkl, S. S. Chern, M. do Carmo, P. Eberlein, J. Ehrbacher, W. Fulton, R. Gardner, P. Gilkey, R. Greene, R. Gulliver, R. Hartshorne, S. Hildebrandt, M. Hirsch, F. Hirzebruch, G. Hochschild, H. Jacobowitz, J. Kazdan, W. Klingenberg, S. Kobayashi, R. Kulkarni, B. Lawson, R. Maltz, P. Melvin, J. Milnor, T. Milnor, J. Moore, C. Morrey, J. Nash, L. Nirenberg, K. Nomizu, B. O'Neill, R. Osserman, T. Ōtsuki, R. Palais, H. Pittie, J. Polking, E. Portnoy, M. Protter, R. Reilly, R. Risch, H. Royden, E. Ruh, P. Ryan, R. Sacksteder, S. Sasaki, D. Schaeffer, J. Sjögren, C. Snygg, E. Spanier, J. Stallings, E. Thomas, F. Warner, J. Weiner, A. Weinstein, J. White, J. Wolf, H. Wu, A. Vasquez, D. Zagier, W. Ziller.

TABLE OF CONTENTS

Although the chapters are not divided into sections, the listing for each chapter gives some indication which topics are treated, and on what pages.

A

Comprehensive Introduction

to

DIFFERENTIAL GEOMETRY

VOLUME THREE

CHAPTER 1

THE FUNDAMENTAL EQUATIONS FOR HYPERSURFACES

In this chapter we are going to begin by considering a very general situation. Let $i\colon M^n \to N^m$ be an immersion of an n-dimensional manifold M into an m-dimensional manifold N; it is customary to refer to N as the "ambient space" and to define $m - n$ to be the **codimension** of M in N. We will be interested in the case where N has a Riemannian metric $\langle\ ,\ \rangle$, so that M can be given the induced Riemannian metric $i^*\langle\ ,\ \rangle$. (This setup is often described a little differently: we can begin with two Riemannian manifolds $(N, \langle\ ,\ \rangle)$ and $(M, \langle\!\langle\ ,\ \rangle\!\rangle)$, and consider isometric immersions of M in N, that is, immersions $i\colon M \to N$ which satisfy $i^*\langle\ ,\ \rangle = \langle\!\langle\ ,\ \rangle\!\rangle$.)

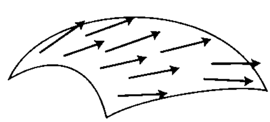
For every $p \in M$, we can consider the tangent space M_p as a subspace of the tangent space $N_{i(p)}$, by identifying M_p with $i_*M_p \subset N_{i(p)}$. Since all the results of this chapter are going to be local ones, it will simplify our notation considerably to assume that M is actually an imbedded submanifold of N, with $i\colon M \to N$ the inclusion map. We can then regard M_p as a subspace $M_p \subset N_p$. In the vector space N_p, with the inner product $\langle\ ,\ \rangle_p$, the subspace M_p has an orthogonal complement ${M_p}^\perp \subset N_p$, and we can use the decomposition $N_p = M_p \oplus {M_p}^\perp$ to define two projections

$$
\begin{aligned}
&\mathsf{T}\colon N_p \to M_p &&\text{(the tangential projection)}\\
&\perp\colon N_p \to {M_p}^\perp &&\text{(the normal, or perpendicular, projection)}
\end{aligned}
$$

with

$$X = \mathsf{T}X + \perp X \qquad \text{for all } X \in N_p.$$

Now let $X_p \in M_p$ be any vector, and let Y be a vector field on M which is everywhere "tangent to M", meaning that $Y_q \in M_q$ for $q \in M$. Then



$\nabla_{X_p} Y \in M_p$ is defined, where ∇ denotes the covariant differentiation in M which is determined by the induced Riemannian metric $i^*\langle\ ,\ \rangle$. If ∇' denotes the covariant differentiation determined by $\langle\ ,\ \rangle$ in the ambient space N, then $\nabla'_{X_p} Y$ is also well-defined; in fact, the value of $\nabla'_{X_p} Y$ depends only on the values of Y along some curve c with $c'(0) = X_p$. The relation between these two covariant differentiations is as nice as one could hope:

1. THEOREM. Let $i\colon M \to N$ be an immersion, where N has the Riemannian metric $\langle\ ,\ \rangle$, and let ∇ and ∇' be the covariant derivatives for $(M, i^*\langle\ ,\ \rangle)$ and $(N, \langle\ ,\ \rangle)$, respectively. If $X_p \in M_p$, and Y is a vector field on M which is everywhere tangent to M, then

$$\nabla_{X_p} Y = \top(\nabla'_{X_p} Y).$$

PROOF. Let X, Y, Z be vector fields on N. Since ∇' is compatible with $\langle\ ,\ \rangle$, we have (Corollary II.6-7)

$$\begin{aligned} X\langle Y, Z\rangle &= \langle \nabla'_X Y, Z\rangle + \langle Y, \nabla'_X Z\rangle \\ Y\langle Z, X\rangle &= \langle \nabla'_Y Z, X\rangle + \langle Z, \nabla'_Y X\rangle \\ -Z\langle X, Y\rangle &= -\langle \nabla'_Z X, Y\rangle - \langle X, \nabla'_Z Y\rangle. \end{aligned}$$

We also have $\nabla'_X Y - \nabla'_Y X = [X, Y]$, etc., and in particular, $\nabla'_X Y + \nabla'_Y X = 2\nabla'_X Y - [X, Y]$. Adding the above three equations, we thus obtain

$$(*)\quad \begin{aligned} &X\langle Y, Z\rangle + Y\langle Z, X\rangle - Z\langle X, Y\rangle \\ &\qquad = 2\langle \nabla'_X Y, Z\rangle - \langle [X, Y], Z\rangle + \langle [X, Z], Y\rangle + \langle [Y, Z], X\rangle. \end{aligned}$$

This equation shows that $\langle \nabla'_X Y, Z\rangle$ is completely determined by $\langle\ ,\ \rangle$ (and is essentially equivalent to our proof of Lemma II.6-8).

Now consider three vector fields X, Y, Z on N which are tangent to M at all points of M, so that there are vector fields $\bar{X}, \bar{Y}, \bar{Z}$ with $i_*\bar{X}(p) = X(p)$, etc. On M we have the same equation $(*)$ for the vector fields $\bar{X}, \bar{Y}, \bar{Z}$, but with $\nabla'_X Y$ replaced by $\nabla_{\bar{X}} \bar{Y}$. For a bracket term like $[\bar{X}, \bar{Y}]$ we have $[\bar{X}, \bar{Y}](p) = [X, Y](p)$ for $p \in M$, by Proposition I.6-3. We thus see that

$$\langle \nabla'_{X_p} Y, Z_p\rangle = \langle \nabla_{X_p} Y, Z_p\rangle \qquad \text{for } all\ Z_p \in M_p.$$

This is equivalent to the desired result. ❖

2. COROLLARY. If c is a curve in M and Y is a vector field along c which is tangent to M along c, then

$$\frac{DY}{dt} = \mathsf{T}\left(\frac{D'Y}{dt}\right).$$

Consequently, Y is parallel along c, in the sense of parallel that pertains to M, if and only if $D'Y/dt$ is always perpendicular to M. In particular, if Y is parallel along c in the sense that pertains to N, then it is also parallel along c in the sense that pertains to M.

PROOF. There is a unique operation $V \mapsto DV/dt$, from C^∞ vector fields V in M along c to C^∞ vector fields in M along c, with the properties in Proposition II.6-2. From the Theorem it is clear that $V \mapsto \mathsf{T}(D'V/dt)$ has these properties. ❖

Merely by combining this information with our previous formulation of other concepts in the ∇ setup, we can immediately deduce further results.

3. COROLLARY. A curve c in M is a geodesic if and only if $D'/dt(dc/dt)$ is everywhere perpendicular to M. In particular, if a geodesic c of N lies entirely in M, then c is also a geodesic in M.

PROOF. The curve c is a geodesic if and only if dc/dt is parallel along c. (The second part also follows from the fact that geodesics are precisely the critical points for the energy function.) ❖

4. COROLLARY. Let M be isometrically immersed in $\mathbb{R}^m$ (with its usual Riemannian metric). A curve c in M is a geodesic if and only if $c''(t)_{c(t)}$ is perpendicular to $M_{c(t)}$ for all t. In particular, a straight line in M is always a geodesic.

PROOF. This is a special case of Corollary 3, for if $\mathbb{R}^m$ has its usual metric, then ∇' is just the directional derivative, so $D'/dt(dc/dt) = c''(t)_{c(t)}$. (In Chapter II.3B we obtained the result (for $m = 3$) by a completely different method.) ❖

Remark: Classically, the tangential component $\mathsf{T}(D'/dt(dc/dt)) = \mathsf{T}(c''(t))$ was called the **geodesic curvature vector** of c. We will meet it again in Chapter 4.

Having considered the tangential component $\top(\nabla'_{X_p}Y)$, it is only fair that we next consider the normal component $\perp(\nabla'_{X_p}Y)$. Notice that

$$\perp(\nabla'_{X_p}fY) = \perp\big(X_p(f)\cdot Y_p + f(p)\cdot \nabla'_{X_p}Y\big) = f(p)\cdot \perp(\nabla'_{X_p}Y).$$

It follows from our general principal (Theorem I.4-2) that there is a well-defined tensor field s, with $s\colon M_p \times M_p \to M_p{}^{\perp}$ for each $p \in M$, such that

$$s(X_p, Y_p) = \perp(\nabla'_{X_p}Y)$$

for any vector field Y extending Y_p.

5. THEOREM. The tensor s is symmetric.

PROOF. Let X and Y be any extensions of $X_p, Y_p \in M_p$ to all of N which are tangent to M at all points of M. Then

$$\begin{aligned} \perp(\nabla'_{X_p}Y) - \perp(\nabla'_{Y_p}X) &= \perp(\nabla'_{X_p}Y - \nabla'_{Y_p}X) \\ &= \perp(\nabla'_X Y(p) - \nabla'_Y X(p)) \\ &= \perp([X,Y](p)) = 0, \end{aligned}$$

since $[X, Y]$ is also tangent to M at all points of M (Proposition I.6-3). ❖

By combining Theorems 1 and 5 we can now rewrite the decomposition

$$\nabla'_{X_p}Y = \perp(\nabla'_{X_p}Y) + \perp(\nabla'_{X_p}Y)$$

in the following form:

> The Gauss Formulas:
>
> $$\nabla'_{X_p}Y = \nabla_{X_p}Y + s(X_p, Y_p),$$
>
> where $X_p \in M_p$, and Y is a vector field tangent along M.

Although we seem to be dealing with a single formula here, we will obtain a set of formulas when we choose a coordinate system $x^1, \ldots, x^n$ on M and let $X_p = \partial/\partial x^i|_p$ and $Y = \partial/\partial x^j$. Consequently, we will adhere to this classical terminology; it should be compared to the (likewise classical) nomenclature of the next result.

6. THEOREM. Let M be isometrically immersed in N, and let R and R' denote the curvature tensors of M and N, respectively. Then for all $X_p, Y_p, Z_p, W_p \in M_p$ we have

Gauss' Equation (Gauss' Theorema Egregium):

$$\langle R'(X_p, Y_p)Z_p, W_p\rangle = \langle R(X_p, Y_p)Z_p, W_p\rangle + \langle s(X_p, Z_p), s(Y_p, W_p)\rangle - \langle s(Y_p, Z_p), s(X_p, W_p)\rangle$$

PROOF. Extend X_p, Y_p, Z_p, W_p to vector fields X, Y, Z, W which are tangent along M. Then the Gauss formulas yield

$$(1) \qquad \begin{aligned} \nabla'_X(\nabla'_Y Z) &= \nabla'_X(\nabla_Y Z) + \nabla'_X(s(Y, Z)) \\ &= \nabla_X(\nabla_Y Z) + s(X, \nabla_Y Z) + \nabla'_X(s(Y, Z)), \end{aligned}$$

and similarly

$$(1') \qquad \nabla'_Y(\nabla'_X Z) = \nabla_Y(\nabla_X Z) + s(Y, \nabla_X Z) + \nabla'_Y(s(X, Z)),$$

as well as

$$(2) \qquad \nabla'_{[X,Y]}Z = \nabla_{[X,Y]}Z + s([X, Y], Z).$$

Substituting (1), (1′), (2) into the formula $R'(X, Y)Z = \nabla'_X\nabla'_Y Z - \nabla'_Y\nabla'_X Z - \nabla'_{[X,Y]}Z$, and noting that W is orthogonal to any term $s(\ ,\ )$, we obtain

$$(3) \quad \langle R'(X, Y)Z, W\rangle = \langle R(X, Y)Z, W\rangle + \langle\nabla'_X(s(Y, Z)) - \nabla'_Y(s(X, Z)), W\rangle.$$

On the other hand, since $\langle s(Y, Z), W\rangle = 0$ we have

$$(4) \quad \begin{aligned} 0 = X(\langle s(Y, Z), W\rangle) &= \langle\nabla'_X s(Y, Z), W\rangle + \langle s(Y, Z), \nabla'_X W\rangle \\ &= \langle\nabla'_X s(Y, Z), W\rangle + \langle s(Y, Z), \nabla_X W + s(X, W)\rangle \\ &= \langle\nabla'_X s(Y, Z), W\rangle + \langle s(Y, Z), s(X, W)\rangle, \end{aligned}$$

since $\nabla_X W$ is orthogonal to $s(Y, Z)$. The desired result is now obtained by substituting (4), and the similar result with X and Y interchanged, into (3). ❖

Recall that if $P \subset M_p$ is a 2-dimensional subspace of M_p, we define the **sectional curvature** $K(P)$ as $\langle R(X, Y)Y, X\rangle$ for orthonormal $X, Y \in P$. We will let $K'(P)$ denote the corresponding sectional curvature in N.

7. COROLLARY (SYNGE'S INEQUALITY). Let M be isometrically immersed in N, and let $\gamma\colon [a,b] \to M$ be a curve in M which is a geodesic in N (and hence also a geodesic in M, by Corollary 3). Then for all 2-dimensional $P \subset M_{\gamma(t)}$ with $\gamma'(t) \in P$ we have

$$K(P) \leq K'(P).$$

In particular, if M is a *surface*, then for all $p = \gamma(t)$ we have

$$K(M_p) \leq K'(M_p).$$

Moreover, in this case equality holds for all $p = \gamma(t)$ if and only if $M_{\gamma(t)}$ is parallel along γ, in the sense that pertains to N.

PROOF. Assume γ is parameterized by arclength. Let $X_p = \gamma'(t)$ and let $Y_p \in P$ be a unit vector perpendicular to X_p. Applying Gauss' equation with $Z_p = Y_p$ and $W_p = X_p$, we obtain

$$K'(P) = K(P) + \langle s(X_p, Y_p), s(X_p, Y_p)\rangle - \langle s(Y_p, Y_p), s(X_p, X_p)\rangle.$$

If we let X be the vector field $X(t) = \gamma'(t)$ along γ, then X is parallel along γ, so we have $0 = \nabla'_X X$. This implies that

$$0 = \perp(\nabla'_X X)(p) = s(X_p, X_p),$$

which gives the desired inequality.

In the case of a surface, we choose $X(t) = \gamma'(t)$ once again, and we let $Y(t)$ be a unit vector in $M_{\gamma(t)}$ which is perpendicular to $X(t)$. Now Gauss' equation gives

$$K'(M_p) = K(M_p) + \langle s(X_p, Y_p), s(X_p, Y_p)\rangle - \langle s(Y_p, Y_p), s(X_p, X_p)\rangle.$$

Once again we have $s(X_p, X_p) = 0$, so equality holds for all p if and only if $s(X_p, Y_p) = 0$ for all p. Moreover, on the *surface* M, the vector field X is parallel along γ, while Y is a unit vector field which makes a constant angle with X along γ. It follows that Y is also parallel along γ, *in the sense that pertains to* M. Therefore

$$0 = \nabla_X Y = \top(\nabla'_X Y).$$

Since

$$s(X_p, Y_p) = \perp(\nabla'_X Y(p)),$$

this shows that $s(X_p, Y_p) = 0$ for all p if and only if $\nabla'_X Y(p) = 0$ for all p; the latter condition means that $M_{\gamma(t)}$ is parallel along γ. ❖

Remark: We can state the slightly more precise result for M a surface: $K(M_p) = K'(M_p)$ at a particular point p if and only if $\nabla'_X Y(p) = 0$.

As another application of Theorem 6, we give a new proof of an old result: If $W \subset N_p$ is 2-dimensional, and $\mathcal{O} \subset W$ is a sufficiently small neighborhood of 0, then $K(W)$ is the Gaussian curvature at p of the surface $M = \exp(\mathcal{O})$. Clearly we just have to show that $s(X_p, Y_p) = 0$ for $X_p, Y_p \in M_p$, so we just need to show that $s(X_p, X_p) = 0$ for all $X_p \in M_p$. But there is a vector field X tangent to M with $X = c'$ along the geodesic $c(t) = \exp(tX_p)$ of N. Then $\nabla'_X X(p) = 0$, so $s(X_p, X_p) = 0$.

The proof of Corollary 7 has probably already explained why Theorem 6 is referred to in the singular, as "Gauss' equation": when M is 2-dimensional and x^1, x^2 is a coordinate system on M, essentially the only interesting case of Theorem 6 occurs for $X_p = W_p = \partial/\partial x^1|_p$ and $Y_p = Z_p = \partial/\partial x^2|_p$, so that we really are dealing with a single equation. This equation actually occurs in Gauss' paper, as we shall soon see, when we specialize our results somewhat.

For the remainder of this chapter we consider the more specific situation where M is a **hypersurface** in N, that is, a submanifold of codimension 1; we will return only much later to the more general situation. In the case of hypersurfaces we can locally choose a **unit normal field** for M: on a neighborhood U of a point $p \in M$ we can choose a vector field ν such that $\langle \nu, \nu \rangle = 1$ and $\nu(q) \in M_q{}^{\perp}$ for all $q \in U$; in fact, there are only two possible choices for ν. Since ν is a vector field of N, defined along M, the symbol $\nabla'_{X_p}\nu$ makes sense for $X_p \in M_p$.

8. THEOREM. Let M be a hypersurface in N, and let ν be a unit normal field on a neighborhood of p in M.

(a) For all $X_p \in M_p$ we have

$$\nabla'_{X_p}\nu \in M_p.$$

(b) If Y is a vector field tangent along M, then we have

The Weingarten Equations:

$$\langle \nabla'_{X_p}\nu, Y_p \rangle = -\langle \nu, \nabla'_{X_p} Y \rangle = -\langle \nu, s(X_p, Y_p) \rangle.$$

(c) Consequently,

$$\langle \nabla'_{X_p}\nu, Y_p \rangle = \langle X_p, \nabla'_{Y_p}\nu \rangle.$$

PROOF. (a) Since $\langle \nu, \nu \rangle = 1$ along M, we have

$$0 = X_p(\langle \nu, \nu \rangle) = 2\langle \nabla'_{X_p} \nu, \nu \rangle,$$

which means that $\nabla'_{X_p} \nu \in M_p$, since ${M_p}^{\perp}$ is 1-dimensional.

(b) Since $\langle \nu, Y \rangle = 0$ along M, we have

$$0 = X_p(\langle \nu, Y \rangle) = \langle \nabla'_{X_p} \nu, Y \rangle + \langle \nu, \nabla'_{X_p} Y \rangle,$$

which implies the first equality in the Weingarten equations. The second equality comes from the definition $s(X_p, Y_p) = \perp(\nabla'_{X_p} Y)$, and the fact that $\perp(\nabla'_{X_p} Y)$ is a multiple of ν.

(c) follows from (b) and symmetry of s. ❖

The reader may recall that the "Weingarten equations" have already appeared in Volume II, pg. 124. The relationship between those equations and the ones in Theorem 8, as well as the reason for choosing the notation $s(X_p, Y_p)$, may come out in the following special case of Theorem 8.

9. COROLLARY. Let M^n be a hypersurface in $\mathbb{R}^{n+1}$ and let ν be a unit normal field on a neighborhood of p in M. Then for all $X_p, Y_p \in M_p$ we have

$$s(X_p, Y_p) = \mathrm{II}(X_p, Y_p) \cdot \nu(p),$$

where $\mathrm{II}(X_p, Y_p)$ is the second fundamental form of M defined for the choice ν of unit normal field, namely

$$\mathrm{II}(X_p, Y_p) = -\langle d\nu(X_p), Y_p \rangle.$$

(Here $d\nu(X_p)$ is interpreted as follows [pg. II.121ff.]: Since we can think of ν as a map $\nu\colon M \to S^{n-1} \subset \mathbb{R}^{n+1}$, we have the vector-valued differential form $d\nu\colon M_p \to \mathbb{R}^{n+1}$, and $d\nu(X_p) \in \mathbb{R}^{n+1}$ is to be moved back to a parallel vector in M_p; equivalently, $d\nu(X_p)$ denotes $\nu_*(X_p) \in {S^{n-1}}_{\nu(p)}$ moved back to a parallel vector in M_p.)

PROOF. Since $\nabla'_{X_p} \nu$ is now simply the directional derivative of ν, we have

$$\nabla'_{X_p} \nu = [X_p(\nu)]_p = [d\nu(X_p)]_p = d\nu(X_p),$$

in the notation we have just adopted. So the Theorem says that

$$\begin{aligned} \langle \nu, s(X_p, Y_p) \rangle &= -\langle d\nu(X_p), Y_p \rangle \\ &= \mathrm{II}(X_p, Y_p), \end{aligned}$$

which is equivalent to the desired result. ❖

The reader should now be able to see that the Weingarten equations of Theorem 8 reduce to equations (a)–(c) on pg. II.124 for a surface in $\mathbb{R}^3$. More precisely, the equation $\langle \nabla'_{X_p} \nu, Y_p \rangle = -\langle \nu, s(X_p, Y_p) \rangle$ establishes the relationship between s and II, and the equation $\langle \nu, \nabla'_{X_p} Y \rangle = \langle \nu, s(X_p, Y_p) \rangle$ then reduces to equations (a)–(c). One further point is worth checking: our present proof that s is symmetric is essentially equivalent to our second proof, in Volume II, that II is symmetric.

10. COROLLARY. Let M be a surface immersed in $\mathbb{R}^3$, and let $X_p, Y_p \in M_p$. Then

$$\langle R(X_p, Y_p)Y_p, X_p \rangle = \mathrm{II}(X_p, X_p) \cdot \mathrm{II}(Y_p, Y_p) - [\mathrm{II}(X_p, Y_p)]^2.$$

In particular, if (x, y) is a coordinate system on M, and we introduce the classical notation

$$\begin{aligned} \langle \ , \ \rangle = \mathrm{I} &= E\, dx \otimes dx + F\, dx \otimes dy + F\, dy \otimes dx + G\, dy \otimes dy \\ \mathrm{II} &= l\, dx \otimes dx + m\, dx \otimes dy + m\, dy \otimes dx + n\, dy \otimes dy, \end{aligned}$$

then

$$\frac{R_{1212}}{EG - F^2}(p) = \frac{ln - m^2}{EG - F^2}(p) = K(p),$$

where $K(p)$ is the Gaussian curvature of M at p.

PROOF. The first equation follows from Theorem 8, Corollary 9, and the fact that $R' = 0$ for $\mathbb{R}^3$. For the second equation we recall the formulas on pp. II.190 and 129. ❖

As we have already noted in the proof of Proposition II.4-7, when we expand R_{1212} using formula (***) on pg. II.188, the second equation in Corollary 10 is exactly equivalent to Gauss' equation for K. The reader probably suspects that our more general Gauss equations can be used to obtain generalizations of the Theorema Egregium to higher dimensions. However, we will defer all such considerations until after we have studied surfaces in more detail. For the present we wish to consider only one more result, which depends on a definition motivated by Corollary 9. If $M \subset N$ is a hypersurface, we produce a symmetric tensor II on M by *defining*

$$s(X_p, Y_p) = \mathrm{II}(X_p, Y_p) \cdot \nu(p);$$

naturally, the sign of II depends on the choice of the local unit normal field ν. Some authors call s the second fundamental form of $M \subset N$, while others reserve that name for II. The tensor II merely gives the length of s up to sign; but since it is real-valued, rather than $M_p{}^{\perp}$ valued, the symbol $\nabla_{Z_p}\mathrm{II}$ makes sense. Indeed, we have defined $\nabla_{Z_p}\mathcal{T}$ for any tensor field $\mathcal{T}$ (see Volume II, pp. 229ff.).

11. THEOREM. Let M be a hypersurface in N, and let ν be a unit normal field on a neighborhood of p in M, with corresponding II. Then for all $X_p, Y_p, Z_p \in M_p$, we have

> The Codazzi-Mainardi Equations:
>
> $$\langle R'(X_p, Y_p)Z_p, \nu(p)\rangle = (\nabla_{X_p}\mathrm{II})(Y_p, Z_p) - (\nabla_{Y_p}\mathrm{II})(X_p, Z_p).$$

Remark: This formula gives us the normal component of $R'(X_p, Y_p)Z_p$, while Gauss' equation essentially gives us the tangential component.

PROOF. We begin with the equations derived in the proof of Theorem 6:

$$(1)\qquad \nabla'_X(\nabla'_Y Z) = \nabla_X(\nabla_Y Z) + s(X, \nabla_Y Z) + \nabla'_X(s(Y, Z))$$

$$(1')\qquad \nabla'_Y(\nabla'_X Z) = \nabla_Y(\nabla_X Z) + s(Y, \nabla_X Z) + \nabla'_Y(s(X, Z))$$

$$(2)\qquad \begin{aligned} \nabla'_{[X,Y]}Z &= \nabla_{[X,Y]}Z + s([X, Y], Z) \\ &= \nabla_{[X,Y]}Z + s(\nabla_X Y, Z) - s(\nabla_Y X, Z). \end{aligned}$$

From these we see that the normal component of $R'(X, Y)Z$ is given by

$$(3)\qquad \begin{aligned} &\text{normal component of } R'(X, Y)Z = \\ &\quad [\perp\nabla'_X(s(Y, Z)) - s(\nabla_X Y, Z) - s(Y, \nabla_X Z)] \\ &\qquad - [\perp\nabla'_Y(s(X, Z)) - s(\nabla_Y X, Z) - s(X, \nabla_Y Z)]. \end{aligned}$$

On the other hand, since

$$(4)\qquad s(Y, Z) = \mathrm{II}(Y, Z)\cdot\nu,$$

we have

$$\nabla'_X(s(Y, Z)) = X(\mathrm{II}(Y, Z))\cdot\nu + \mathrm{II}(Y, Z)\cdot\nabla'_X\nu,$$

and consequently

$$(5)\qquad \langle\nabla'_X(s(Y, Z)), \nu\rangle = X(\mathrm{II}(Y, Z)),$$

since $\nabla'_X \nu$ is tangent to M. Using (3), (5), and the definition (4) again, we obtain

$$\begin{aligned}\langle R'(X,Y)Z,\nu\rangle = {} & [X(\mathrm{II}(Y,Z)) - \mathrm{II}(\nabla_X Y, Z) - \mathrm{II}(Y, \nabla_X Z)] \\ & - [Y(\mathrm{II}(X,Z)) - \mathrm{II}(\nabla_Y X, Z) - \mathrm{II}(X, \nabla_Y Z)].\end{aligned}$$

The result now follows from* Corollary II.6-5. ❖

It will be useful to examine the form which our fundamental equations take when the ambient space N has constant curvature K_0. Then by Lemma II.7-18 the curvature tensor R' of N satisfies

(1) $$\langle R'(X,Y)Z,W\rangle = K_0[\langle X,W\rangle \cdot \langle Y,Z\rangle - \langle X,Z\rangle \cdot \langle Y,W\rangle]$$

$$\Downarrow$$

(2) $$R'(X,Y)Z = K_0[\langle Y,Z\rangle X - \langle X,Z\rangle Y].$$

12. COROLLARY. Let N have constant curvature K_0. Then for M isometrically immersed in N we have

Gauss' Equation:

$$\begin{aligned}\langle R(X_p,Y_p)Z_p,W_p\rangle &+ \langle s(X_p,Z_p), s(Y_p,W_p)\rangle - \langle s(Y_p,Z_p), s(X_p,W_p)\rangle \\ &= K_0[\langle X_p,W_p\rangle \cdot \langle Y_p,Z_p\rangle - \langle X_p,Z_p\rangle \cdot \langle Y_p,W_p\rangle].\end{aligned}$$

And if M is a hypersurface we have

The Codazzi-Mainardi Equations:

$$(\nabla_{X_p}\mathrm{II})(Y_p,Z_p) = (\nabla_{Y_p}\mathrm{II})(X_p,Z_p).$$

PROOF. The first result follows from Theorem 6 and equation (1). The second follows from Theorem 11 and equation (2), which shows that $R'(X_p,Y_p)Z_p$ is tangent to M. ❖

*This Corollary holds also for tensors of type $\binom{k}{0}$; see also Problem 1.

We have carried our analysis of submanifolds as far as we presently wish to go. However, it is also important that we indicate how things work out when we use moving frames. Before doing this, we will first examine the classical tensor analysis treatment of submanifolds. This is included mainly for the sake of completeness, and because you may be unfortunate enough to encounter it again in a classical work which you need to consult. If you are inclined to skip this part, I cannot in good conscience caution you against such a course of action, except to say that reading it must be good for you, because you certainly won't like it.

We will simplify things slightly by beginning with hypersurfaces from the outset. We consider a coordinate system $y^1, \dots, y^{n+1}$ on N, with

$$\langle\ ,\ \rangle = \sum_{\alpha,\beta=1}^{n+1} g'_{\alpha\beta}\, dy^\alpha \otimes dy^\beta,$$

and let $x^1, \dots, x^n$ be a coordinate system on a hypersurface M, so that

$$\langle\ ,\ \rangle = \sum_{i,j=1}^{n} g_{ij}\, dx^i \otimes dx^j \qquad \text{on } M,$$

for certain functions g_{ij}. We adopt the convention that the indices i, j, etc., range from 1 to n, while α, β, etc., range from 1 to $n+1$, even in summation signs, so that $\sum_i$, for example, denotes $\sum_{i=1}^n$. It is easy to see that

$$\sum_{\alpha,\beta} g'_{\alpha\beta} \frac{\partial y^\alpha}{\partial x^i} \frac{\partial y^\beta}{\partial x^j} = g_{ij} \qquad \text{on } M.$$

It will be convenient to let y^α also denote the restriction of y^α to M. Then we can use the symbol $y^\alpha{}_{;i} = \partial y^\alpha / \partial x^i$, introduced on pg. II.211, for the components of dy^α on M, and we can write

(1) $$\sum_{\alpha,\beta} g'_{\alpha\beta} y^\alpha{}_{;i} y^\beta{}_{;j} = g_{ij} \qquad \text{on } M.$$

If $\nu = \sum_\alpha \nu^\alpha \cdot \partial/\partial y^\alpha$ is a unit normal field, then we also have

(2) $$\sum_{\alpha,\beta} g'_{\alpha\beta} y^\alpha{}_{;i} \nu^\beta = 0,$$

(3) $$\sum_{\alpha,\beta} g'_{\alpha\beta} \nu^\alpha \nu^\beta = 1.$$

We now wish to take the covariant derivative of (1) on M. Notice that $y^\alpha{}_{;i}\, y^\beta{}_{;j}$ is the (i, j) component of the tensor $dy^\alpha \otimes dy^\beta$ on M; on the other hand, each $g'_{\alpha\beta}$ is just a function on M. Writing

$$\frac{\partial g'_{\alpha\beta}}{\partial x^k} \qquad \text{as} \qquad \sum_\gamma \frac{\partial g'_{\alpha\beta}}{\partial y^\gamma}\frac{\partial y^\gamma}{\partial x^k} = \sum_\gamma \frac{\partial g'_{\alpha\beta}}{\partial y^\gamma} y^\gamma{}_{;k},$$

and using Proposition II.5-2, we obtain from (1)

$$\sum_{\alpha,\beta,\gamma} \frac{\partial g'_{\alpha\beta}}{\partial y^\gamma} y^\alpha{}_{;i}\, y^\beta{}_{;j}\, y^\gamma{}_{;k} + \sum_{\alpha,\beta} g'_{\alpha\beta}(y^\alpha{}_{;ik}\, y^\beta{}_{;j} + y^\beta{}_{;jk}\, y^\alpha{}_{;i}) = g_{ij;k}$$
$$= 0, \qquad \text{by Ricci's Lemma (Proposition II.5-3).}$$

If we write this equation with i and k interchanged, the term

$$\sum_{\alpha,\beta,\gamma} \frac{\partial g'_{\alpha\beta}}{\partial y^\gamma} y^\alpha{}_{;k}\, y^\beta{}_{;j}\, y^\gamma{}_{;i} \qquad \text{can be replaced by} \qquad \sum_{\alpha,\beta,\gamma} \frac{\partial g'_{\gamma\beta}}{\partial y^\alpha} y^\alpha{}_{;i}\, y^\beta{}_{;j}\, y^\gamma{}_{;k}.$$

A similar replacement can be made when we rewrite the original equation with j and k interchanged. Adding the two equations so obtained, and subtracting the original, we get

$$\sum_{\alpha,\beta} g'_{\alpha\beta}\, y^\alpha{}_{;k}\, y^\beta{}_{;ij} + \sum_{\alpha,\beta,\gamma} [\alpha\beta, \gamma]'\, y^\alpha{}_{;i}\, y^\beta{}_{;j}\, y^\gamma{}_{;k} = 0,$$

where $[\ ,\ ]'$ indicates the Christoffel symbols for the y coordinate system. This can also be written as

$$(4) \qquad \sum_{\alpha,\beta} g'_{\alpha\beta}\, y^\beta{}_{;k}\left(y^\alpha{}_{;ij} + \sum_{\rho,\sigma} \Gamma'^\alpha_{\rho\sigma}\, y^\rho{}_{;i}\, y^\sigma{}_{;j}\right) = 0,$$

which shows that the expression in parentheses is the α component of a vector perpendicular to M. As a matter of fact, a calculation (Problem 2) shows that it is the coefficient of $\partial/\partial y^\alpha$ in the expression for $\nabla'_{\partial/\partial x^j}\partial/\partial x^i - \nabla_{\partial/\partial x^j}\partial/\partial x^i$. Consequently, (4) is equivalent to Theorem 1, and despite the ugliness of the equations involved, its derivation is clearly closely related to that of Theorem 1.

Since $\sum_\alpha \nu^\alpha \cdot \partial/\partial y^\alpha$ is a unit normal field, and $M_p{}^\perp$ has dimension 1, equation (4) implies that

$$(5) \qquad y^\alpha{}_{;ij} + \sum_{\rho,\sigma} \Gamma'^\alpha_{\rho\sigma}\, y^\rho{}_{;i}\, y^\sigma{}_{;j} = \mathrm{II}_{ij}\, \nu^\alpha$$

for certain functions II_{ij}; multiplying by $\sum_\beta g'_{\alpha\beta}\nu^\beta$ and using (3), this can be written

(6) $$\mathrm{II}_{ij} = \sum_{\alpha,\beta} g'_{\alpha\beta}\nu^\beta y^\alpha{}_{;ij} + \sum_{\rho,\sigma,\beta} [\rho\sigma,\beta]' y^\rho{}_{;i} y^\sigma{}_{;j}\nu^\beta.$$

This shows that the II_{ij} satisfy the transformation rule for a covariant tensor of order 2 on M, since the $y^\alpha{}_{;ij}$ and $y^\rho{}_{;i}y^\sigma{}_{;j}$ do, and since the other terms don't involve the coordinate system x but only the coordinate system y on N (at the same time we see that the whole right side doesn't even depend on y). It is also clear that $\mathrm{II}_{ij} = \mathrm{II}_{ji}$. Equation (5) is equivalent to the Gauss formulas.

We next take the covariant derivative of equation (2) on M (now both $g'_{\alpha\beta}$ and ν^β are functions on M). We obtain

$$\sum_{\alpha,\beta} g'_{\alpha\beta}\left(y^\alpha{}_{;ij}\nu^\beta + y^\alpha{}_{;i}\nu^\beta{}_{;j}\right) = -\sum_{\alpha,\beta,\sigma} y^\alpha{}_{;i} y^\sigma{}_{;j}\nu^\beta \frac{\partial g'_{\alpha\beta}}{\partial y^\sigma}$$
$$= -\sum_{\alpha,\beta,\sigma} y^\alpha{}_{;i} y^\sigma{}_{;j}\nu^\beta([\alpha\sigma,\beta]' + [\beta\sigma,\alpha]').$$

Then (6) gives

$$\mathrm{II}_{ij} = -\sum_{\alpha,\beta} g'_{\alpha\beta} y^\alpha{}_{;i}\nu^\beta{}_{;j} - \sum_{\rho,\sigma,\beta} [\beta\sigma,\rho]' y^\rho{}_{;i} y^\sigma{}_{;j}\nu^\beta,$$

or

$$-\mathrm{II}_{ij} = \sum_{\alpha,\beta} g'_{\alpha\beta} y^\alpha{}_{;i}\left(\nu^\beta{}_{;j} + \sum_{\rho,\sigma} \Gamma'^{\beta}_{\rho\sigma} y^\rho{}_{;j}\nu^\sigma\right),$$

which can also be written as

(7) $$-\mathrm{II}_{ij} = \sum_{\alpha,\beta} g'_{\alpha\beta} y^\alpha{}_{;i}\left(\sum_\rho \nu^\beta{}_{:\rho} y^\rho{}_{;j}\right),$$

where $\nu^\beta{}_{:\rho}$ now denotes the covariant derivative, *in* N, *of the vector field with components* ν^β, so that

$$\nu^\beta{}_{:\rho} = \frac{\partial \nu^\beta}{\partial y^\rho} + \sum_\sigma \Gamma'^{\beta}_{\rho\sigma}\nu^\sigma.$$

Equation (7) is clearly equivalent to one part of the Weingarten equations, namely $-\langle \nu, s(X_p, Y_p)\rangle = \langle \nabla'_{X_p}\nu, Y_p\rangle$. If we treat (3) in a similar manner, we end up with

(8) $$\sum_{\alpha,\beta} g'_{\alpha\beta}\nu^\alpha\left(\sum_\rho \nu^\beta{}_{:\rho} y^\rho{}_{;j}\right) = 0,$$

which is equivalent to the fact that $\nabla'_{X_p}\nu \in M_p$.

Since (8) shows that we can write

$$(9)\qquad \sum_{\rho} \nu^{\beta}{}_{;\rho}\, y^{\rho}{}_{;j} = \sum_{k} A^k_j\, y^{\beta}{}_{;k} \qquad \text{for some functions } A^k_j \text{ on } M,$$

equation (7) gives

$$\begin{aligned} -\mathrm{II}_{ij} &= \sum_{\alpha,\beta,k} g'_{\alpha\beta}\, y^{\alpha}{}_{;i}\, y^{\beta}{}_{;k}\, A^k_j \\ &= \sum_k g_{ik} A^k_j \qquad \text{by (1)}, \end{aligned}$$

so

$$A^m_j = \sum_l -\mathrm{II}_{lj}\, g^{lm},$$

and hence from (9)

$$\sum_{\rho} \nu^{\beta}{}_{;\rho}\, y^{\rho}{}_{;j} = -\sum_{l,m} \mathrm{II}_{lj}\, g^{lm}\, y^{\beta}{}_{;m},$$

or equivalently

$$(10)\qquad \nu^{\beta}{}_{;j} + \sum_{\rho,\sigma} \Gamma'^{\beta}_{\rho\sigma}\, y^{\rho}{}_{;j}\, \nu^{\sigma} = -\sum_{l,m} \mathrm{II}_{lj}\, g^{lm}\, y^{\beta}{}_{;m}.$$

We return to equation (5), equivalent to Gauss' formulas. We can apply Ricci's identity (Proposition II. 5-4)

$$\lambda_{i;jk} - \lambda_{i;kj} = \sum_m \lambda_m R^m{}_{ijk} = \sum_{m,h} \lambda_m g^{mh} R_{hijk}$$

to $\lambda_i = y^{\alpha}{}_{;i}$, obtaining

$$\sum_{m,h} y^{\alpha}{}_{;m}\, g^{mh} R_{hijk} = y^{\alpha}{}_{;ijk} - y^{\alpha}{}_{;ikj}.$$

Computing the $y^{\alpha}{}_{;ijk}$ from (5), and using (5) and (10) in the result, we obtain finally

$$\sum_{m,h} y^{\alpha}{}_{;m}\, g^{mh}[R_{hijk} - (\mathrm{II}_{hj}\mathrm{II}_{ik} - \mathrm{II}_{hk}\mathrm{II}_{ij})] - \nu^{\alpha}(\mathrm{II}_{ij;k} - \mathrm{II}_{ik;j})$$
$$- \sum_{\rho,\sigma,\lambda} R'^{\alpha}{}_{\rho\sigma\lambda}\, y^{\rho}{}_{;i}\, y^{\sigma}{}_{;j}\, y^{\lambda}{}_{;k} = 0.$$

Multiplying by $\sum_\alpha g'_{\alpha\beta} y^\beta{}_{;l}$ or by $\sum_\alpha g'_{\alpha\beta}\nu^\beta$, we get

(11) $$R_{ijkl} = (\mathrm{II}_{ik}\mathrm{II}_{jl} - \mathrm{II}_{il}\mathrm{II}_{jk}) + \sum_{\alpha,\beta,\gamma,\delta} R'_{\alpha\beta\gamma\delta}\, y^\alpha{}_{;i}\, y^\beta{}_{;j}\, y^\gamma{}_{;k}\, y^\delta{}_{;l}$$

(12) $$\mathrm{II}_{ij;k} - \mathrm{II}_{ik;j} = \sum_{\alpha,\beta,\gamma,\delta} R'_{\alpha\beta\gamma\delta}\, y^\alpha{}_{;i}\, y^\gamma{}_{;j}\, y^\delta{}_{;k}\, \nu^\beta.$$

Equations (11) and (12) are equivalent to Gauss' Equation and the Codazzi-Mainardi equations, respectively. Whew!

When we turn to the method of moving frames, we find ourselves in a situation completely different from the mass of calculations in which we have just been mired. Although the moving frame method will not have the geometric appeal of the ∇ theory, it is far superior computationally. Not only are all the equations short and simple, but all the results follow naturally, almost without thought, from the structural equations. Indeed, everything happens so quickly that the real problem is recognizing a result when it appears.

Consider first an orthonormal moving frame $X_1, \ldots, X_n$ on an open subset of M. Recall that the **dual 1-forms** θ^i are defined by $\theta^i(X_j) = \delta^i_j$, and that there are unique 1-forms ω^i_j, the **connection forms**, satisfying the two equations

(1) $$\omega^i_j = -\omega^j_i$$

(2) $$d\theta^i = -\sum_{k=1}^n \omega^i_k \wedge \theta^k \qquad \text{(The first structural equation).}$$

The **curvature forms** Ω^i_j are then defined by

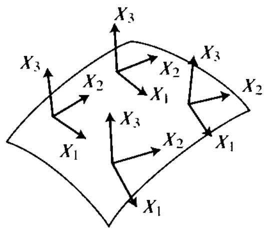
(3) $$d\omega^i_j = -\sum_{k=1}^n \omega^i_k \wedge \omega^k_j + \Omega^i_j \qquad \text{(The second structural equation).}$$

The relationship between ω^i_j, Ω^i_j and ∇, R is given by

(4) $$\nabla_{X_k} X_j = \sum_{i=1}^n \omega^i_j(X_k) X_i \qquad \text{or} \qquad \langle \nabla_X X_j, X_i\rangle = \omega^i_j(X)$$

(5) $$R(X_k, X_l)X_j = \sum_{i=1}^n \Omega^i_j(X_k, X_l) X_i \quad \text{or} \quad \langle R(X,Y)X_j, X_i\rangle = \Omega^i_j(X,Y).$$

Now let us consider an orthonormal moving frame $X_1, \ldots, X_m$, defined on an open subset of N, with the property that $X_1, \ldots, X_n$ are tangent to M at points of M, and consequently $X_{n+1}, \ldots, X_m$ are normal to M at points of M; such an orthonormal moving frame is said to be **adapted to** M. An adapted



orthonormal moving frame gives us an orthonormal moving frame $X_1, \ldots, X_n$ along M, with corresponding forms θ^i, ω^i_j, Ω^i_j $(i, j \leq n)$. We also want to consider the corresponding forms for the entire moving frame $X_1, \ldots, X_m$ on N; these will be denoted by ϕ^α, ψ^α_β, Ψ^α_β. We adopt the convention that i, j, etc., always range from 1 to n, while α, β, etc., always range from 1 to m, even in summation signs, so that $\sum_i$, for example, means $\sum_{i=1}^n$; it will also be convenient to use r, s, etc., for numbers that range from $n+1$ to m.

Now the forms ϕ^α, ψ^α_β, Ψ^α_β can be restricted to TM (that is, to tangent vectors of M). Clearly

$$\phi^i = \theta^i \qquad \text{on } TM \quad i \leq n$$
$$\phi^r = 0 \qquad \text{on } TM \quad r > n.$$

To obtain some information about the forms ψ^α_β on TM, we look at the first structural equation,

$$d\phi^\alpha = -\sum_\gamma \psi^\alpha_\gamma \wedge \phi^\gamma.$$

Restricting to TM we obtain

(a) $$d\theta^i = -\sum_k \psi^i_k \wedge \theta^k \qquad \text{on } TM \quad i \leq n$$

(b) $$0 = \sum_k \theta^k \wedge \psi^r_k \qquad \text{on } TM \quad r > n.$$

Recall that ω^i_j were the *unique* forms satisfying (1), (2). Since $\psi^\alpha_\beta = -\psi^\beta_\alpha$, equation (a) therefore shows that

(c) $$\psi^i_j = \omega^i_j \qquad \text{on } TM \quad i, j \leq n.$$

This equation already contains some information! In fact, equation (4) shows that

$$\langle \nabla_X X_j, X_i \rangle = \omega_j^i(X) = \psi_j^i(X) = \langle \nabla'_X X_j, X_i \rangle \qquad \text{for } X \in TM;$$

this is exactly equivalent to Theorem 1, for it shows that $\nabla'_X X_j(p)$ has the same inner product with every element of M_p as $\nabla_X X_j(p) \in M_p$.

Next, we look at the forms $\psi_k^r = -\psi_r^k$ on TM, for $k \le n < r$. Equation (b) tells us that they satisfy the hypothesis of the following Lemma.

13. LEMMA (CARTAN'S LEMMA). If $\lambda^1, \dots, \lambda^n$ are (C^∞) linearly independent 1-forms on a manifold M (of dimension $n' \ge n$), and $\mu_1, \dots, \mu_n$ are (C^∞) 1-forms on M satisfying

$$(*) \qquad \sum_{i=1}^{n} \lambda^i \wedge \mu_i = 0,$$

then there are unique (C^∞) functions f_{ij} on M such that

$$\mu_i = \sum_{j=1}^{n} f_{ij} \lambda^j;$$

moreover,

$$f_{ij} = f_{ji}.$$

Remark: This result (or at least the corresponding result for vector spaces) has already been given in Problem I.7-11.

PROOF. In a neighborhood of any point we can choose (C^∞) 1-forms $\lambda^{n+1}, \dots, \lambda^{n'}$ so that $\lambda^1, \dots, \lambda^{n'}$ are everywhere independent. Then there are (C^∞) functions f_{ij} ($i \le n$, $j \le n'$) with

$$\mu_i = \sum_{j=1}^{n'} f_{ij} \lambda^j.$$

Equation (*) implies that

$$0 = \sum_{i=1}^{n} \sum_{j=1}^{n'} f_{ij}\, \lambda^i \wedge \lambda^j = \sum_{1 \le i < j \le n} (f_{ij} - f_{ji})\, \lambda^i \wedge \lambda^j + \sum_{i=1}^{n} \sum_{j>n} f_{ij}\, \lambda^i \wedge \lambda^j.$$

Since the $\lambda^i \wedge \lambda^j$ for $i < j$ are linearly independent, we have $f_{ij} - f_{ji} = 0$ for $i, j \le n$ and $f_{ij} = 0$ for $j > n$. ❖

Applying Cartan's Lemma to the ψ^r_k, we conclude that there are unique functions s^r_{ij} on M satisfying

(d) $$\psi^r_j = -\psi^j_r = \sum_i s^r_{ij}\theta^i \qquad \text{on } TM \quad r > n,$$

$$s^r_{ij} = s^r_{ji}.$$

Equation (4) now shows that

$$\langle \nabla'_{X_k} X_j, X_r \rangle = \psi^r_j(X_k) = s^r_{kj} \qquad r > n,$$

and hence

$$\langle \nabla'_{X_j} X_k, X_r \rangle = \langle \nabla'_{X_k} X_j, X_r \rangle \qquad r > n.$$

This is basically Theorem 5, asserting the symmetry of s, which in our present notation can be defined by setting

(e) $$s(X_j, X_k) = \sum_r \Psi^r_j(X_k) \cdot X_r = \sum_r s^r_{jk} X_r,$$

and extending s by linearity. It should be noted that this definition of s involves a choice of a moving frame; when one is developing everything from the moving frame approach, a little calculation (Problem 3) must be supplied to show that the definition of s is really independent of the choice. Equations (c) and (d) together are equivalent to the Gauss formulas.

Now let us look at the second structural equation

$$d\psi^\alpha_\beta = -\sum_\gamma \psi^\alpha_\gamma \wedge \psi^\gamma_\beta + \Psi^\alpha_\beta.$$

Restricting to TM we obtain, for $\alpha, \beta = i, j \leq n$,

(f) $$d\omega^i_j = -\sum_k \omega^i_k \wedge \omega^k_j + \sum_r \psi^r_i \wedge \psi^r_j + \Psi^i_j \qquad \text{on } TM \quad i, j \leq n.$$

Comparing with (3) we obtain

(g) $$\Psi^i_j = \Omega^i_j - \sum_r \psi^r_i \wedge \psi^r_j \quad \text{on } TM.$$

Then equation (5) gives

$$\langle R'(X,Y)X_j, X_i \rangle = \Psi^i_j(X,Y) = \langle R(X,Y)X_j, X_i \rangle - \sum_r (\psi^r_i(X)\psi^r_j(Y) - \psi^r_i(Y)\psi^r_j(X)).$$

Since we have, for example,

$$\sum_r \psi_i^r(X)\psi_j^r(Y) = \sum_r \langle s(X_i, X), X_r\rangle \cdot \langle s(X_j, Y), X_r\rangle$$
$$= \langle s(X_i, X), s(X_j, Y)\rangle,$$

it follows that (g) is exactly equivalent to Theorem 6 (Gauss' Equation).

If we instead choose $\alpha = r > n$, and $\beta = j \le n$, we obtain

(h) $$d\psi_j^r = -\sum_i \psi_i^r \wedge \omega_j^i - \sum_s \psi_s^r \wedge \psi_j^s + \Psi_j^r \qquad \text{on } TM.$$

As before, we now restrict ourselves to the case $m = n + 1$; then X_{n+1} is a unit normal field on M. Notice that the equation $\psi_{n+1}^j = -\psi_j^{n+1}$ gives

$$\langle \nabla'_{X_k} X_{n+1}, X_j\rangle = \psi_{n+1}^j(X_k) = -\psi_j^{n+1}(X_k) = -\langle X_{n+1}, s(X_j, X_k)\rangle,$$

which are the Weingarten equations. Equation (h) takes the form

$$d\psi_j^{n+1} = -\sum_i \psi_i^{n+1} \wedge \omega_j^i + \Psi_j^{n+1}.$$

A little work (Problem 4) shows that this is equivalent to the Codazzi-Mainardi equations.

SUMMARY

$\phi^i = \theta^i$ on TM

$\phi^r = 0$ on TM

Consequences of the first structural equation:

- $\psi_j^i = \omega_j^i$ on TM .... (Theorem 1) — The Gauss formulas
- $\psi_j^r = \sum_i s_{ij}^r \theta^i$ on TM .... (Theorem 6) — The Gauss formulas
- $s_{ij}^r = s_{ji}^r$

Consequences of the second structural equation:

- $\Psi_j^i = \Omega_j^i - \sum_r \psi_i^r \wedge \psi_j^r$ on TM .. Gauss' Equation
- For $m = n + 1$:
  $\Psi_j^{n+1} = d\psi_j^{n+1} + \sum_i \psi_i^{n+1} \wedge \omega_j^i$ ...... on TM — Codazzi-Mainardi Equations

Although the derivation of the fundamental equations was so much easier in terms of moving frames than in terms of tensors, the resultant equations have the same disadvantage as the moving frame treatment of connections itself—our equations are not "invariant", they are merely a set of equations which hold for each choice of adapted orthonormal moving frame. Moreover, it is very hard to get any geometric feel for the equations—the tensor form of the equations seem much more geometric, especially Gauss' equation. As one might guess, an invariant description of the moving frame equations can be obtained by considering an appropriate principal bundle—the "bundle of adapted orthonormal frames". In Chapter 7 we will actually consider this construction in detail, even for submanifolds of higher codimension, but we will do this mainly for the sake of completeness, since the results which we will derive from this construction will also be obtained in other ways. On the whole, it seems uneconomical to construct the elaborate machinery of a principal bundle just to have a gadget on which we can give an invariant formulation of the fundamental equations for submanifolds, especially since the invariant equations are even more abstract and ungeometric. It is much easier, and more satisfying, to state these equations in terms of tensors down on the submanifold itself. On the other hand, when it comes to *using* these equations to prove theorems about submanifolds, the equations in terms of moving frames will almost always prove to be superior.

ADDENDUM

AUTO-PARALLEL AND TOTALLY GEODESIC SUBMANIFOLDS

The material of this section, besides being of interest in its own right, will play an important role on several occasions later on. We will not require any tools not already developed within the chapter, even though we will be dealing with submanifolds of arbitrary codimension, and even with manifolds whose connection does not come from a Riemannian metric.

Let N be a manifold with a connection ∇', and let M be a submanifold of N. We say that M is **auto-parallel** if the parallel translation in N along a curve c in M always takes vectors tangent to M into vectors tangent to M. For example, a straight line or a plane in $\mathbb{R}^3$ is auto-parallel.

14. PROPOSITION. A submanifold M of (N, ∇') is auto-parallel if and only if $\nabla'_X Y$ is tangent to M whenever X and Y are.

PROOF. We know from Proposition II.6-3 that

$$\nabla'_{X_p} Y = \lim_{h \to 0} \frac{1}{h}({\tau_h}^{-1} Y_{c(h)} - Y_p),$$

where c is a curve with $c'(0) = X_p$, and τ_h is parallel translation along c from $c(0)$ to $c(h)$. This makes it immediately clear that if M is auto-parallel, then $\nabla'_{X_p} Y$ is tangent to M if X and Y are.

Conversely, suppose $\nabla'_{X_p} Y$ is tangent to M whenever X and Y are. Let c be a curve in M and let V be a parallel vector field along c. Choose a coordinate system $x^1, \dots, x^n, x^{n+1}, \dots, x^m$ for N such that $x^r = 0$ on M for all $r > n$. If $V_t \in N_{c(t)}$ is given by

$$V_t = \sum_{\alpha=1}^{m} v^\alpha(t) \cdot \left.\frac{\partial}{\partial x^\alpha}\right|_c (t),$$

then we have (pg. II.233)

$$0 = \frac{dv^\gamma(t)}{dt} + \sum_{\alpha,\beta} \frac{dc^\alpha(t)}{dt} \Gamma^\gamma_{\alpha\beta}(c(t)) v^\beta(t). \tag{1}$$

Now $c^r(t) = 0$ for $r > n$, since c lies in M. Moreover, if $i, j \leq n$, then the vector

$$\nabla'_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} = \sum_{\gamma} \Gamma^{\gamma}_{ij} \frac{\partial}{\partial x^{\gamma}}$$

is tangent to M by hypothesis, so we must have $\Gamma^r_{ij} = 0$ for $r > n$. So for $\gamma = s > n$, equation (1) becomes

$$\frac{dv^s(t)}{dt} = -\sum_{r=n+1}^{m} \sum_{i=1}^{n} \frac{dc^i(t)}{dt} \Gamma^s_{ir}(c(t)) v^r(t).$$

This set of $m - n$ equations for the $m - n$ functions v^s has a unique solution for a given initial condition. The solution with all $v^s(0) = 0$ is clearly just $v^s(t) = 0$ for all $s > n$. In other words, if V_0 is tangent to M, then so are all V_t. ❖

15. COROLLARY. If M is an auto-parallel submanifold of (N, ∇'), then

$$R'(X_p, Y_p)Z_p \in M_p \qquad \text{for all } X_p, Y_p, Z_p \in M_p.$$

PROOF. Use the definition

$$R'(X, Y)Z = \nabla'_X \nabla'_Y Z - \nabla'_Y \nabla'_X Z - \nabla'_{[X,Y]} Z$$

(and Proposition I.6-3). ❖

In the particular case where the connection ∇' on N is the unique symmetric connection compatible with a Riemannian metric $\langle\ ,\ \rangle$ on N, we can characterize auto-parallel submanifolds $M \subset N$ in a different way.

16. PROPOSITION. If $(N, \langle\ ,\ \rangle)$ is a Riemannian manifold, then a submanifold $M \subset N$ is auto-parallel if and only if the second fundamental form s of M is zero.

PROOF. By definition, s is zero if and only if $\nabla'_X Y$ is tangent to M whenever X and Y are. So the result follows from Proposition 14. ❖

Notice that whenever M is an auto-parallel submanifold of (N, ∇'), we can define a connection ∇ on M by letting $\nabla_X Y = \nabla'_X Y$ for X and Y tangent to M. This connection ∇ on M is called the **induced connection** on M. (In the

Riemannian case, ∇ clearly coincides with the connection M has as a Riemannian submanifold.) If c is a curve in M and V is a vector field along c which is everywhere tangent to M, then the covariant derivative DV/dt along c which is determined by ∇ is exactly the same as the covariant derivative $D'V/dt$ along c which is determined by ∇': for the proof we just apply Proposition II.6-2, which essentially defines DV/dt. In particular, if V is parallel along c with respect to the connection ∇ in M, then it is also parallel along c with respect to the connection ∇' in N.

Auto-parallel submanifolds can also be characterized in yet another way. A submanifold M of (N, ∇') is called **geodesic at** p if every geodesic γ with $\gamma(0) = p$ and $\gamma'(0) \in M_p$ remains in M on some interval $(-\varepsilon, \varepsilon)$. It is called **totally geodesic** if it is geodesic at every point; it is easy to see that $M \subset N$ is totally geodesic if and only if every geodesic in M is also a geodesic in N.

17. PROPOSITION. Let M be a submanifold of a manifold (N, ∇').

(1) If M is auto-parallel, then M is totally geodesic.

(2) If M is totally geodesic, and ∇' is symmetric, then M is auto-parallel.

PROOF. (1) Let c be a geodesic of N with $c'(0) \in M_p$. Let $\bar{c}$ be the geodesic in M, with respect to the induced connection ∇, satisfying $\bar{c}'(0) = c'(0)$. To prove that an interval of c lies in M, it certainly suffices to prove that $\bar{c} = c$ along some interval containing 0. Now by the definition of a geodesic, $d\bar{c}/dt$ is parallel along c with respect to ∇. As we have already noted, this implies that $d\bar{c}/dt$ is parallel along c with respect to ∇'. Thus $\bar{c}$ is a geodesic in N. Since $\bar{c}'(0) = c'(0)$, the geodesics $\bar{c}$ and c must coincide on their common domain.

(2) In a neighborhood of a point $p \in M$ we can choose a coordinate system $x^1, \dots, x^n, x^{n+1}, \dots, x^m$ for N such that $x^r = 0$ on M for $r > n$. Let c be a geodesic with

$$c(0) = p \quad \text{and} \quad c'(0) = \sum_{i=1}^{n} a^i \left.\frac{\partial}{\partial x^i}\right|_p \in M_p.$$

Then by hypothesis we have $c(t) \in M$ for sufficiently small t. Now c satisfies (pg. II.246)

$$\frac{d^2 c^\gamma}{dt^2} + \sum_{\alpha,\beta} \Gamma^\gamma_{\alpha\beta}(c(t)) \frac{dc^\alpha}{dt}\frac{dc^\beta}{dt} = 0.$$

For $\gamma = s > n$ we have

$$\sum_{i,j=1}^{n} \Gamma_{ij}^{s}(c(t))\frac{dc^i}{dt}\frac{dc^j}{dt} = 0 \qquad \text{for small } t.$$

Letting $t = 0$, we obtain

$$\sum_{i,j=1}^{n} \Gamma_{ij}^{s}(p)a^i a^j = 0.$$

Choosing $a^i = 1$, all other $a^j = 0$, we get

$$\text{(a)} \qquad 0 = \Gamma_{ii}^{s}(p).$$

Choosing $a^i = a^j = 1$, all other $a^k = 0$, we get

$$\text{(b)} \qquad 0 = \Gamma_{ii}^{s}(p) + \Gamma_{ij}^{s}(p) + \Gamma_{ji}^{s}(p) + \Gamma_{jj}^{s}(p) = \Gamma_{ij}^{s}(p) + \Gamma_{ji}^{s}(p).$$

Using symmetry of the Γ's, we find that $\Gamma_{ij}^{s}(p) = 0$ for all i, j. This is true for all $p \in M$, so we find that $\nabla'_X Y$ is tangent to M if X and Y are. The result then follows from Proposition 14. ❖

Every n-dimensional plane $P \subset \mathbb{R}^m$ is clearly totally geodesic. (Conversely, if $M \subset \mathbb{R}^m$ is a totally geodesic submanifold, and p is a point of M, then M must clearly contain a neighborhood of its tangent space $M_p \subset \mathbb{R}^m$; so if M is a connected n-dimensional totally geodesic submanifold of $\mathbb{R}^m$, then M must be part of an n-dimensional plane $P \subset \mathbb{R}^m$.) It is just as clear that if we give S^m its standard Riemannian metric, then any n-sphere $S^n \subset S^m$ is totally geodesic. Now let us consider the Riemannian manifold $(N, \langle\ ,\ \rangle)$ mentioned on pg. II. 301, with constant curvature -1: the manifold N is

$$N = \left\{ a \in \mathbb{R}^m : \sum_{\alpha=1}^{m} (a^\alpha)^2 < 4 \right\},$$

and the components $g_{\alpha\beta}$ of $\langle\ ,\ \rangle$ with respect to the usual coordinate system $x^1, \dots, x^m$ are given by

$$g_{\alpha\beta} = \frac{\delta_{\alpha\beta}}{\left[1 - \frac{1}{4}\sum_{\alpha=1}^{m}(x^\alpha)^2\right]^2}.$$

Let $M \subset N$ be

$$M = \{a \in N : a^{n+1} = \cdots = a^m = 0\}.$$

The formulas for the Γ's on pg. II.299 show that $\Gamma_{ij}^r = 0$ on M whenever $i, j \le n$ and $r > n$, which means that $\nabla'_X Y$ is tangent to M whenever X and Y are. So M is a totally geodesic submanifold of N, by Propositions 14 and 16. Since the metric $\langle\ ,\ \rangle$ is radially symmetric around 0, it is clear that we can find a totally geodesic submanifold M of N with M_0 being any n-dimensional subspace of N_0. The same is true at any other point $p \in N$, because the fact that N has constant curvature implies that p has a neighborhood isometric to a neighborhood of 0 (Corollary II.7-13); in fact, since N is simply-connected and complete, there is actually an isometry of N onto itself taking any point p to 0 (Problem 5).*

The possibility of finding so many totally geodesic submanifolds is very exceptional:

18. THEOREM. Let $(N, \langle\ ,\ \rangle)$ be a connected Riemannian manifold of dimension $m \ge 3$. Suppose that for all $p \in N$ and all 2-dimensional subspaces $P \subset N_p$ there is a totally geodesic submanifold M of N with $p \in M$ and $M_p = P$. Then N has constant curvature.

PROOF. Each submanifold M is auto-parallel, by Proposition 17, so Corollary 15 shows that

$$\langle R'(X_p, Y_p)Z_p, W_p\rangle = 0$$

for $X_p, Y_p, Z_p \in M_p$ and $W_p \in {M_p}^{\perp}$. Since M_p can be any 2-dimensional subspace $P \subset M_p$, we see that

$$(1) \qquad \langle R'(X_p, Y_p)X_p, W_p\rangle = 0 \qquad \text{for orthonormal } X_p, Y_p, W_p \in N_p.$$

Applying (1) to $X_p, \bar{Y}_p, \bar{W}_p$ with

$$\begin{aligned} \bar{Y}_p &= (\cos\alpha)Y_p + (\sin\alpha)W_p \\ \bar{W}_p &= (-\sin\alpha)Y_p + (\cos\alpha)W_p, \end{aligned}$$

we obtain

$$\begin{aligned} 0 &= \sin\alpha\cos\alpha[\langle R'(X_p, W_p)X_p, W_p\rangle - \langle R'(X_p, Y_p)X_p, Y_p\rangle] \\ &\qquad + \cos^2\alpha\langle R'(X_p, Y_p)X_p, W_p\rangle - \sin^2\alpha\langle R'(X_p, W_p)X_p, Y_p\rangle \\ &= \sin\alpha\cos\alpha[\langle R'(X_p, W_p)X_p, W_p\rangle - \langle R'(X_p, Y_p)X_p, Y_p\rangle] \qquad \text{by (1).} \end{aligned}$$

*More detailed information about the manifold N will be found in Chapter 7, Part A.

Thus $\langle R'(X_p, W_p)X_p, W_p \rangle = \langle R'(X_p, Y_p)X_p, Y_p \rangle$ for all orthonormal X_p, Y_p, W_p, which implies that all sectional curvatures at p are equal. Since this is true for all p, Schur's Theorem (II.7-19) shows that M has constant curvature. ❖

It seems rather clear that if one takes a Riemannian manifold $(N, \langle\ ,\ \rangle)$ "at random", then it will not have any totally geodesic submanifolds of dimension > 1. But I must admit that I don't know of any specific example of such a manifold.

PROBLEMS

1. (a) In Corollary II.6-5, each $A(p)$ is regarded as a map $M_p \times \cdots \times M_p \to M_p$. If we instead regard each $A(p)$ as a map $M_p \times \cdots \times M_p \times M_p{}^* \to \mathbb{R}$, show that

$$(\nabla_{X_p} A)(Y_1(p), \ldots, Y_k(p), \omega(p)) = \nabla_{X_p}(A(Y_1, \ldots, Y_k, \omega))$$
$$- \sum_{i=1}^{k} A(Y_1(p), \ldots, \nabla_{X_p} Y_i, \ldots, Y_k(p), \omega(p)) + A(Y_1(p), \ldots, Y_k(p), \nabla_{X_p}\omega).$$

(b) If A is a tensor field of type $\binom{k}{l}$, where each $A(p)$ is regarded as a map $M_p \times \cdots \times M_p \times M_p{}^* \times \cdots \times M_p{}^* \to \mathbb{R}$, show that

$$(\nabla_{X_p} A)(Y_1(p), \ldots, \omega_l(p)) = \nabla_{X_p}(A(Y_1, \ldots, \omega_l))$$
$$- \sum_{i=1}^{k} A(\ldots, \nabla_{X_p} Y_i, \ldots, \omega_l(p))$$
$$+ \sum_{i=1}^{l} A(Y_1(p), \ldots, \nabla_{X_p}\omega_i, \ldots).$$

Consider in particular the cases $l = 0$ and $k = 0$.
(c) If we instead regard each $A(p)$ as a map from $M_p \times \cdots \times M_p$ to the set of maps $M_p{}^* \times \cdots \times M_p{}^* \to \mathbb{R}$, then

$$(\nabla_{X_p} A)(Y_1(p), \ldots, Y_k(p)) = \nabla_{X_p}(A(Y_1, \ldots, Y_k))$$
$$- \sum_{i=1}^{k} A(Y_1(p), \ldots, \nabla_{X_p} Y_i, \ldots, Y_k(p)).$$

2. Consider the situation on page 12. Writing

$$\frac{\partial}{\partial x^i} = \sum_{\rho} \frac{\partial y^\rho}{\partial x^i} \frac{\partial}{\partial y^\rho} = \sum_{\rho} y^\rho{}_{;i} \frac{\partial}{\partial y^\rho},$$

and similarly for $\partial/\partial x^j$, show that

$$\nabla'_{\partial/\partial x^j} \partial/\partial x^i = \sum_{\alpha} \left(\sum_{\rho,\sigma} \Gamma'^{\alpha}_{\rho\sigma} y^\rho{}_{;i} y^\rho{}_{;j} \right) \frac{\partial}{\partial y^\alpha} + \sum_{\alpha} \left(\sum_{\rho} y^\rho{}_{;j} \frac{\partial y^\alpha{}_{;i}}{\partial y^\rho} \right) \frac{\partial}{\partial y^\alpha}.$$

Using

$$y^\alpha{}_{;ij} = \frac{\partial y^\alpha{}_{;i}}{\partial x^j} - \sum_{k} y^\alpha{}_{;k} \Gamma^k_{ji} = \sum_{\rho} y^\rho{}_{;j} \frac{\partial y^\alpha{}_{;i}}{\partial y^\rho} - \sum_{k} y^\alpha{}_{;k} \Gamma^k_{ji},$$

verify the assertion near the bottom of page 13.

3. (The calculations in this Problem are similar to those on pages 79–82 and 97–100 and might be postponed until then, or they may be regarded as a rehearsal for the latter.) Let $\mathbf{X} = X_1, \dots, X_m$ and $\mathbf{X}' = X'_1, \dots, X'_m$ be two adopted orthonormal moving frames on $M^n \subset N$. Let s^r_{ij} and s'^r_{ij} be the unique functions with

$$\psi^r_j = \sum_i s^r_{ij}\theta^i, \qquad \psi'^r_j = \sum_i s'^r_{ij}\theta'^i.$$

Say that $\mathbf{X}' = \mathbf{X} \cdot a$ for an orthogonal matrix of functions a, so that we have (pp. II.280, 282)

$$\theta' = a^{-1} \cdot \theta \qquad \text{and} \qquad \psi' = a^{-1}da + a^{-1}\psi a.$$

The matrix a must satisfy $a^r_j = 0 = a^j_r$, since $\mathbf{X}$ and $\mathbf{X}'$ are both adopted to M. Conclude that

$$\psi'^r_j = \sum_{i,t} (a^{-1})^r_t \psi^t_i a^i_j,$$

and thus that

$$\sum_i s'^r_{ij}(a^{-1})^i_h = \sum_{i,t} (a^{-1})^r_t s^t_{ih} a^i_j \implies \sum_r s'^r_{kj} a^u_r = \sum_{i,h} s^u_{ih} a^i_j a^h_k.$$

Hence show that the definition

$$s(X'_j, X'_k) = \sum_r s'^r_{jk} X'_r$$

is compatible with the definition

$$s(X_i, X_h) = \sum_u s^u_{ih} X_u.$$

4. Apply the equation for ψ^{n+1}_j on page 20 to (X_k, X_l). Noting that

$$d\psi^{n+1}_j(X_k, X_l) = X_k(\psi^{n+1}_j(X_l)) - X_l(\psi^{n+1}_j(X_k)) - \psi^{n+1}_j([X_k, X_l])$$
$$[X_k, X_l] = \nabla'_{X_k} X_l - \nabla'_{X_l} X_k = \sum_i \omega^i_l(X_k) X_i - \omega^i_k(X_l) X_i,$$

deduce the last equation in the proof of Theorem 11.

5. Let N and $\bar{N}$ be two n-dimensional Riemannian manifolds of constant curvature K_0. Let $X_1, \ldots, X_n$ be an orthonormal basis of N_p, and $\bar{X}_1, \ldots, \bar{X}_n$ be an orthonormal basis of $\bar{N}_{\bar{p}}$. Let $c\colon [0,1] \to N$ be a curve with $c(0) = p$. By Corollary II.7-13, there is an isometry f from a neighborhood U_0 of p to a neighborhood of $\bar{p}$, with $f_*(X_i) = \bar{X}_i$. A **continuation** of f along c is a family $\{f_t\}$ of isometries $f_t\colon U_t \to \bar{N}$ where U_t is a neighborhood of $c(t)$, with $f_0 = f$, satisfying the following condition: for each t there is $\delta > 0$ so that $|t - t'| < \delta \Rightarrow f_t = f_{t'}$ on $U_t \cap U_{t'}$.

(a) If $\{f_t\}$ and $\{g_t\}$ are two continuations of f, then each $f_t = g_t$ in some neighborhood of $c(t)$.
(b) If N is connected, then there is at most one isometry $\phi\colon N \to \bar{N}$ with $\phi_* X_i = \bar{X}_i$.
(c) Let $\bar{N}$ be complete, and let $K = \{q \in \bar{N} : d(\bar{p}, q) \leq \text{ length of } c\}$. Then K is compact (see pg. I.343). Conclude that there is $\delta > 0$ such that for any orthonormal $Y_1, \ldots, Y_n \in N_{c(t)}$ $(0 \leq t \leq 1)$ and orthonormal $\bar{Y}_1, \ldots, \bar{Y}_n \in \bar{N}_q$ $(q \in K)$, there is an isometry taking Y_i to $\bar{Y}_i$ whose domain contains all $c(t')$ for $|t - t'| < \delta$. Then show that a continuation always exists.
(d) If N is simply-connected, show that for two paths $c, \gamma\colon [0,1] \to M$ with $c(0) = \gamma(0) = p$ and $c(1) = \gamma(1) = q$, the continuations $\{f_t\}$, $\{g_t\}$ of f along c and γ must satisfy $f_1(q) = g_1(q)$. Conclude that for $\bar{N}$ complete and N simply-connected, there is a (unique) isometry $\phi\colon N \to \bar{N}$ with $\phi_* X_i = \bar{X}_i$.
(e) Any two simply-connected complete manifolds of constant curvature K_0 are isometric, and there is an isometry taking any orthonormal basis of such a manifold to any other orthonormal basis.

CHAPTER 2

ELEMENTS OF THE THEORY OF SURFACES IN $\mathbb{R}^3$

This chapter will parallel as closely as possible the first chapter of Volume II. Our interest will now be directed away from the intrinsic geometry of surfaces, and toward those properties which describe the particular ways they are immersed in $\mathbb{R}^3$.

We recall that in our study of curves, we defined certain quantities, the curvature κ and the torsion τ, which describe the local appearance of a curve in $\mathbb{R}^3$. The curvature was first defined in an extremely geometric way, by taking limits of circles passing through three points of the curve. But it could be defined quite simply as $\kappa = |\mathbf{t}'|$, and exactly this approach was used to define τ. We then showed that these quantities actually describe the curve completely, up to Euclidean motions. We also investigated certain global properties connected with positive curvature. Finally, we showed how our investigations could be reformulated in terms of Lie groups, with Theorem I.10-18 playing a leading role, and then went on to investigate properties of curves invariant under a different group of motions of Euclidean space.

Our investigation of surfaces will proceed in just this order; however it will differ from the study of curves in one important respect. For curves we found that the definition and study of κ and τ [or of the affine curvature $\varkappa$] was greatly simplified by considering curves parameterized only by arclength [or affine arclength]. But for surfaces there is no natural choice of a parameterization; to a certain extent this is responsible for the considerably greater complications one encounters in surface theory.

In Chapter II.3B we considered a submanifold $M \subset \mathbb{R}^3$, with $i: M \to \mathbb{R}^3$ the inclusion map, and we defined the **first fundamental form** I by $\mathrm{I} = i^*\langle\ ,\ \rangle$, where $\langle\ ,\ \rangle$ is the usual Riemannian metric on $\mathbb{R}^3$. In terms of a coordinate system $\chi = (x, y)$ on M, we wrote the tensor I on M as

$$\mathrm{I} = E\, dx \otimes dx + F\, dx \otimes dy + F\, dy \otimes dx + G\, dy \otimes dy$$

for certain functions E, F, G on M; and we noted that if the inverse of χ is

$f: U \to \mathbb{R}^3$ (for $U \subset \mathbb{R}^2$ open), then

$$
\begin{aligned}
E(f(s,t)) &= \left\langle \frac{\partial f}{\partial s}(s,t), \frac{\partial f}{\partial s}(s,t) \right\rangle = \langle f_1(s,t), f_1(s,t) \rangle \\
F(f(s,t)) &= \langle f_1(s,t), f_2(s,t) \rangle \\
G(f(s,t)) &= \langle f_2(s,t), f_2(s,t) \rangle.
\end{aligned}
$$

This means that $E = \langle f_1, f_1 \rangle \circ f^{-1}$, etc., which sometimes makes the functions E, F, G rather awkward to work with; consequently, we will often find it convenient to change our view slightly, and define everything explicitly in terms of a given immersion.

If $f: M \to \mathbb{R}^3$ is an immersion, we define the **first fundamental form I_f of f** to be the tensor $f^*\langle\ ,\ \rangle$ on M. In particular, when $f: U \to \mathbb{R}^3$ (for $U \subset \mathbb{R}^2$ open) we have a form I_f on U, with

$$
\mathrm{I}_f(s,t)(v,w) = \langle f_* v, f_* w \rangle \qquad \text{for} \quad v, w \in \mathbb{R}^2{}_{(s,t)}.
$$

We can then define functions E, F, G directly on U by

$$
\begin{aligned}
E &= \left\langle \frac{\partial f}{\partial s}, \frac{\partial f}{\partial s} \right\rangle = \langle f_1, f_1 \rangle \\
F &= \langle f_1, f_2 \rangle \\
G &= \langle f_2, f_2 \rangle.
\end{aligned}
$$

These functions are nothing but the components of $\mathrm{I}_f = f^*\langle\ ,\ \rangle$ with respect to the standard coordinate system (s,t) on $\mathbb{R}^2$ [and they have essentially already been introduced on pg. II.128]. Since $f^*\langle\ ,\ \rangle$ is positive definite, we have $EG - F^2 > 0$ (pg. I.308); moreover (see Problem I.9-5), we have

$$
|f_1 \times f_2| = \sqrt{EG - F^2}.
$$

It will often be much more convenient to use the subscript notation, which was classically used for the higher dimensional cases,

$$
g_{ij} = \langle f_i, f_j \rangle, \qquad \text{so that} \qquad g_{11} = E, \quad g_{12} = g_{21} = F, \quad g_{22} = G.
$$

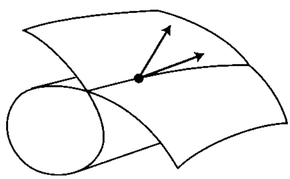
We also introduce the functions g^{ij} which satisfy

$$
\sum_k g_{ik} g^{kj} = \delta_i^j.
$$

The functions E, F, G are analogous to the single function $t \mapsto |c'(t)|$, defined for a curve c. We usually reparameterized our curves so that this function was equal to 1, but for surfaces, where no convenient reparameterization is available, the functions E, F, G always play a vital role, analogous to the arclength function of a curve.

We next seek an analogue of the functions κ and τ of a curve c. The definition of κ and τ depended very much on the possibility of parameterizing c by arclength, so that $\mathbf{t}(s) = c'(s)$ has length 1, and consequently $\mathbf{t}'(s) = \kappa(s)\mathbf{n}(s)$ for some unit vector $\mathbf{n}(s)$ perpendicular to the curve. In the case of a surface $M \subset \mathbb{R}^3$, we do not have a special parameterization to work with, but we already have an analogue of $\mathbf{n}$, namely a unit normal field ν, which can at least be defined in a neighborhood of each point. We recall that a choice of ν is equivalent to a choice of an orientation for M, for we can let $(X_1, X_2) \in M_p$ be positively oriented if and only if $(X_1, X_2, \nu(p))$ is positively oriented in $\mathbb{R}^3$.

When we are not dealing with a submanifold, but with an immersion $f: M \to \mathbb{R}^3$, the normal field should be considered as a "vector field along f", since we may have points $p, q \in M$ with $f(p) = f(q)$, but with different normals at this



point. We will denote this vector field along f by

$$q \mapsto N(q)_{f(q)} \in \mathbb{R}^3{}_{f(q)}.$$

Thus N is a function $N: M \to S^2 \subset \mathbb{R}^3$ We will always adhere to the convention of using ν when we are specifically considering imbedded submanifolds $M \subset \mathbb{R}^3$, and using N when we are considering immersions (and imbeddings) $f: M \to \mathbb{R}^3$; when necessary, we will write N_f to indicate the dependence of N on f. If $W \subset M$ is an open set on which f is an imbedding, then a unit normal field ν on $M = f(W) \subset \mathbb{R}^3$ is determined by the condition that $N = \nu \circ f$ on W (naturally we have to regard ν as a map $\nu: M \to S^2 \subset \mathbb{R}^3$ in order to write this). In terms of ν we have already defined the **second fundamental form II on** M by

$$\begin{aligned} \mathrm{II}(p)(v_p, w_p) &= \langle -d\nu(v_p), w_p \rangle \\ &= \langle -\nu_*(v_p), w_p \rangle \qquad \text{for} \quad v_p, w_p \in M_p. \end{aligned}$$

We can define the **second fundamental form II_f of f** to be the tensor on M defined by

$$\begin{aligned}\mathrm{II}_f(q)(v_q, w_q) &= \langle -dN(v_q), f_*(w_q)\rangle \\ &= \langle -N_*(v_q), f_*(w_q)\rangle \qquad \text{for} \quad v_q, w_q \in M_q.\end{aligned}$$

Equivalently, we have $\mathrm{II}_f = f^*\mathrm{II}$.

In particular, let us consider an immersion $f\colon U \to \mathbb{R}^3$, for $U \subset \mathbb{R}^2$ open. We can specifically choose N to be

$$N = \frac{f_1 \times f_2}{|f_1 \times f_2|} = \frac{f_1 \times f_2}{\sqrt{EG - F^2}}$$ (the negative square root gives the other possible choice for N).

Then

$$\begin{aligned}\mathrm{II}_f(s,t)(v, w) &= \langle -dN(v), f_*(w)\rangle \\ &= \langle -N_*(v), f_*(w)\rangle \qquad v, w \in \mathbb{R}^2{}_{(s,t)}.\end{aligned}$$

We will define functions l, m, n directly on U by

$$\begin{aligned}l &= \left\langle -\frac{\partial N}{\partial s}, \frac{\partial f}{\partial s}\right\rangle = \langle -N_1, f_1\rangle = \langle N, f_{11}\rangle \\ m &= \langle -N_1, f_2\rangle = \langle N, f_{12}\rangle \\ n &= \langle -N_2, f_2\rangle = \langle N, f_{22}\rangle.\end{aligned}$$

These functions l, m, n are simply the components of II_f with respect to the standard coordinate system (s, t) on $\mathbb{R}^2$ (compare with the proof of Theorem 1 on pg. II.123). Once again, it will often be more convenient to use subscript notation:

$$l_{ij} = \langle -N_i, f_j\rangle = \langle N, f_{ij}\rangle.$$

Before we go any further, we should clear up one problem. Let us suppose that $f\colon U \to \mathbb{R}^3$ is actually an imbedding, and let $M = f(U) \subset \mathbb{R}^3$. In geometric considerations, it is usually the linear transformation $-d\nu\colon M_p \to M_p$ which interests us, rather than the second fundamental form II itself. Now for $p = f(s,t)$, the map $-d\nu\colon M_p \to M_p$ is related to the matrix $(l_{ij}(s,t))$ in the following way:

$$l_{ij}(s,t) = \langle -d\nu(f_i(s,t)_p), f_j(s,t)_p\rangle.$$

Unfortunately, this does *not* mean that $(l_{ij}(s,t))$ is the matrix of $-d\nu$ with respect to the basis $f_1(s,t)_p, f_2(s,t)_p$, because these vectors are not necessarily orthonormal. To avoid going out of our minds now, and especially in the next chapter, we make note of the following:

0. FACT. Let $v_1, \ldots, v_n$ be a basis for the vector space V, with the inner product $\langle\ ,\ \rangle$. Suppose that A is the matrix of a linear transformation $T\colon V \to V$ with respect to $v_1, \ldots, v_n$, while B and C are the matrices $B = (\langle Tv_i, v_j\rangle)$ and $C = (\langle v_i, v_j\rangle)$. Then

$$A = (BC^{-1})^{\mathbf{t}},$$

where $^{\mathbf{t}}$ denotes the transpose.

PROOF. We have $Tv_i = \sum_j A_{ji} v_j$, and consequently

$$B_{ik} = \langle Tv_i, v_k\rangle = \sum_j A_{ji}\langle v_j, v_k\rangle = (A^{\mathbf{t}} \cdot C)_{ik},$$

so $B = A^{\mathbf{t}} \cdot C$. ❖

We apply this observation with

$$A = \begin{pmatrix} \text{matrix of } -d\nu\colon M_p \to M_p \\ \text{with respect to } (f_1)_p, (f_2)_p \end{pmatrix}$$

and

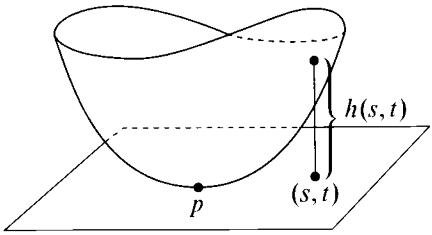
$$B = (l_{ij}), \qquad C = (g_{ij}),$$

where f_i, l_{ij}, g_{ij} are all evaluated at (s,t), and $p = f(s,t)$. Since B and C are symmetric, we have $(BC^{-1})^{\mathbf{t}} = C^{-1}B$; so we find that

$$\text{(I)}\quad \begin{pmatrix} \text{matrix of } -d\nu\colon M_p \to M_p \\ \text{with respect to } (f_1)_p, (f_2)_p \end{pmatrix} = (g_{ij})^{-1}(l_{ij})$$
$$= \frac{1}{EG - F^2} \cdot \begin{pmatrix} G & -F \\ -F & E \end{pmatrix}\begin{pmatrix} l & m \\ m & n \end{pmatrix}$$
$$[f_i, g_{ij}, l_{ij} \text{ evaluated at } (s,t);\ p = f(s,t)].$$

The functions l, m, n certainly seem to be good candidates for the desired analogues of the functions κ and τ. As a first test of their appropriateness, we will investigate how well these functions describe f up to second order in a neighborhood of a point p.

What we are interested in is the shape of $M = \text{image } f$, not the particular parameterization f itself; so we will essentially fix the parameterization by describing our surface M in terms of the distance of a point from the tangent plane at p. For convenience we will assume that $p = 0 \in \mathbb{R}^3$ and that the tangent plane at p is the (x, y)-plane. Then our surface M is the graph of a



function $h\colon \mathbb{R}^2 \to \mathbb{R}$ with $h_1(0,0) = h_2(0,0) = 0$. Applying Taylor's formula for functions of two variables, we have

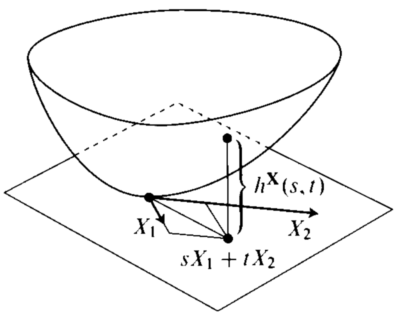
$$h(s,t) = \tfrac{1}{2}\big(h_{11}(0,0)\cdot s^2 + 2h_{12}(0,0)\cdot st + h_{22}(0,0)\cdot t^2\big) + R(s,t),$$

where $R(s,t)/|(s,t)|^2 = R(s,t)/(s^2+t^2) \to 0$ as $(s,t) \to 0$. We therefore say that the quadratic surface

$$P = \left\{\big(s,t,\tfrac{1}{2}\big(h_{11}(0,0)\cdot s^2 + 2h_{12}(0,0)\cdot st + h_{22}(0,0)\cdot t^2\big)\big)\right\}$$

"approximates M up to order 2 at 0".

Now we want to make an elementary observation which is crucial for avoiding confusion. Suppose that $\mathbf{X} = (X_1, X_2)$ is any basis for $\mathbb{R}^2$, say with $X_1 = (a_{11}, a_{21})$ and $X_2 = (a_{12}, a_{22})$. Our surface M can also be described as the graph of a function in terms of the "$X_1, X_2, (0,0,1)$ coordinate system". In



other words, we can consider the function $h^{\mathbf{X}}$ with $h^{\mathbf{X}}(s,t) =$ the height above the (x,y)-plane of the point of M lying above $sX_1 + tX_2$. This means that

$$h^{\mathbf{X}}(s,t) = h(sX_1 + tX_2) = h(a_{11}s + a_{12}t, a_{21}s + a_{22}t).$$

If we momentarily denote (s, t) by (s^1, s^2), then we can write more conveniently

$$h^{\mathbf{X}}(s^1, s^2) = h\Big(\sum_{i=1}^{2} a_{1i}s^i, \sum_{i=1}^{2} a_{2i}s^i\Big),$$

and for the partial derivatives of $h^{\mathbf{X}}$ we easily compute that

$$h^{\mathbf{X}}{}_{\alpha}(s^1, s^2) = \sum_{j=1}^{2} a_{j\alpha}h_j\Big(\sum_{i=1}^{2} a_{1i}s^i, \sum_{i=1}^{2} a_{2i}s^i\Big)$$

$$h^{\mathbf{X}}{}_{\alpha\beta}(s^1, s^2) = \sum_{j,k=1}^{2} a_{j\alpha}a_{k\beta}h_{jk}\Big(\sum_{i=1}^{2} a_{1i}s^i, \sum_{i=1}^{2} a_{2i}s^i\Big),$$

so that in particular

$$(*) \qquad \begin{cases} h^{\mathbf{X}}{}_{\alpha}(0,0) = 0 \\ h^{\mathbf{X}}{}_{\alpha\beta}(0,0) = \displaystyle\sum_{j,k=1}^{2} a_{j\alpha}a_{k\beta}h_{jk}(0,0). \end{cases}$$

Now the quadratic surface $Q \subset \mathbb{R}^3 = \mathbb{R}^2 \times \mathbb{R}$ defined by

$$Q = \big\{\big(sX_1 + tX_2, \tfrac{1}{2}\big(h^{\mathbf{X}}{}_{11}(0,0) \cdot s^2 + 2h^{\mathbf{X}}{}_{12}(0,0) \cdot st + h^{\mathbf{X}}{}_{22}(0,0) \cdot t^2\big)\big)\big\}$$

can equally well be said to approximate M up to order 2 at 0. But we claim that the surfaces P and Q are *exactly the same.* The best way to express this claim is as follows. For each basis $\mathbf{X} = (X_1, X_2)$, let us define a function $\Phi^{\mathbf{X}}\colon \mathbb{R}^2 \to \mathbb{R}$ by

$$\Phi^{\mathbf{X}}(sX_1 + tX_2) = \tfrac{1}{2}\big(h^{\mathbf{X}}{}_{11}(0,0) \cdot s^2 + 2h^{\mathbf{X}}{}_{12}(0,0) \cdot st + h^{\mathbf{X}}{}_{22}(0,0) \cdot t^2\big).$$

Then the functions $\Phi^{\mathbf{X}}$ are all the *same* function $\Phi\colon \mathbb{R}^2 \to \mathbb{R}$, and we can therefore describe both P and Q simply as the graph of Φ. To check that all $\Phi^{\mathbf{X}}$ are the same, let us use Φ for the function we obtain when $\mathbf{X}$ is the standard basis $(1,0), (0,1)$. Then

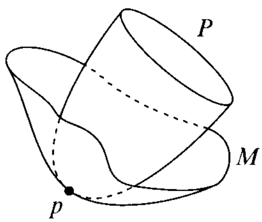
$$\begin{aligned} \Phi(s^1X_1 + s^2X_2) &= \Phi\Big(\sum_{i=1}^{2} a_{1i}s^i, \sum_{i=1}^{2} a_{2i}s^i\Big) \\ &= \tfrac{1}{2}\sum_{j,k=1}^{2} h_{jk}(0,0) \cdot \Big(\sum_{i=1}^{2} a_{ji}s^i\Big)\Big(\sum_{i=1}^{2} a_{ki}s^i\Big) \end{aligned}$$

$$= \frac{1}{2} \sum_{\alpha,\beta=1}^{2} \left(\sum_{j,k=1}^{2} h_{jk}(0,0) a_{j\alpha} a_{k\beta} \right) s^\alpha s^\beta$$

$$= \frac{1}{2} \sum_{\alpha,\beta=1}^{2} h^{\mathbf{X}}{}_{\alpha\beta}(0,0) s^\alpha s^\beta \qquad \text{by } (*)$$

$$= \Phi^{\mathbf{X}}(s^1 X_1 + s^2 X_2),$$

which is what we wanted to prove. It is also easy to see that if we describe M in terms of the "$X_1, X_2, (0,0,-1)$ coordinate system", then Φ becomes $-\Phi$, while the resulting second order approximating surface is unchanged. Thus, for every point p of a surface M in $\mathbb{R}^3$, there is a well-defined quadratic surface P which approximates M up to order 2 at p.



Let us for simplicity stick to the case where $p = 0 \in \mathbb{R}^3$, the tangent plane at p is the (x,y)-plane, and M is the graph of $h\colon \mathbb{R}^2 \to \mathbb{R}$ (in the standard coordinate system). The surface M is the image of the immersion

$$f(s,t) = (s,t,h(s,t))$$

for which we have

$$\begin{aligned} N &= N(0,0) = (0,0,1) \\ l &= l(0,0) = \langle (0,0,1), (0,0,h_{11}(0,0)) \rangle \\ &= h_{11}(0,0) \\ m &= m(0,0) = h_{12}(0,0) \\ n &= n(0,0) = h_{22}(0,0). \end{aligned}$$

Thus our approximating quadratic surface P at 0 is described explicitly as the graph of

$$\alpha(s,t) = \tfrac{1}{2}(ls^2 + 2mst + nt^2) = \tfrac{1}{2}\left\langle (s,t),\ (s,t) \cdot \begin{pmatrix} l & m \\ m & n \end{pmatrix} \right\rangle.$$

To see just what this graph looks like, we choose two *orthonormal* eigenvectors $X_1, X_2 \in \mathbb{R}^2$ for the symmetric matrix $\begin{pmatrix} l & m \\ m & n \end{pmatrix}$, with corresponding eigenvalues k_1, k_2. Then

$$\begin{aligned} \alpha(sX_1 + tX_2) &= \tfrac{1}{2}\left\langle sX_1 + tX_2,\ sX_1 + tX_2 \cdot \begin{pmatrix} l & m \\ m & n \end{pmatrix}\right\rangle \\ &= \tfrac{1}{2}\langle sX_1 + tX_2,\ sk_1X_1 + tk_2X_2\rangle \\ &= \tfrac{1}{2}(k_1s^2 + k_2t^2). \end{aligned}$$

In other words, after a rotation of our axes so that they point along X_1, X_2, the graph of α becomes the graph of $\tilde{\alpha}\colon \mathbb{R}^2 \to \mathbb{R}$ defined by

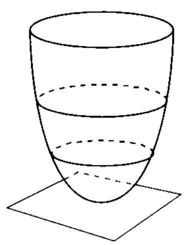
$$\tilde{\alpha}(s,t) = \tfrac{1}{2}(k_1s^2 + k_2t^2).$$

The shape of this graph depends on the sign of $k_1k_2 = \det\begin{pmatrix} l & m \\ m & n \end{pmatrix} = ln - m^2$, and leads us to classify the points p into four types.

1. *Elliptic point*: $ln - m^2 > 0$. Then k_1, k_2 have the same sign. If $k_1, k_2 > 0$, then the graph of

$$\tilde{\alpha}(s,t) = \frac{s^2}{\left(\sqrt{2/k_1}\right)^2} + \frac{t^2}{\left(\sqrt{2/k_2}\right)^2}$$

is an *elliptic paraboloid*; planes parallel to the (x, y)-plane intersect the graph of $\tilde{\alpha}$ in similar ellipses, while planes parallel to the other coordinate planes intersect the graph in parabolas.



In our new coordinate system, the original surface is the graph of $\tilde{h}$, where

$$\tilde{h}(s,t) = \tilde{\alpha}(s,t) + \tilde{R}(s,t),$$

with $\tilde{R}(s,t)/(s^2+t^2) \to 0$. There is clearly a constant $A > 0$ such that $\tilde{\alpha}(s,t) > A(s^2 + t^2)$, so

$$0 = \lim_{(s,t)\to 0} -\frac{\tilde{R}(s,t)}{s^2+t^2} = \lim_{(s,t)\to 0}\left(\frac{\tilde{\alpha}(s,t) - \tilde{h}(s,t)}{s^2+t^2}\right) \geq A - \lim_{(s,t)\to 0}\frac{\tilde{h}(s,t)}{s^2+t^2},$$

and hence

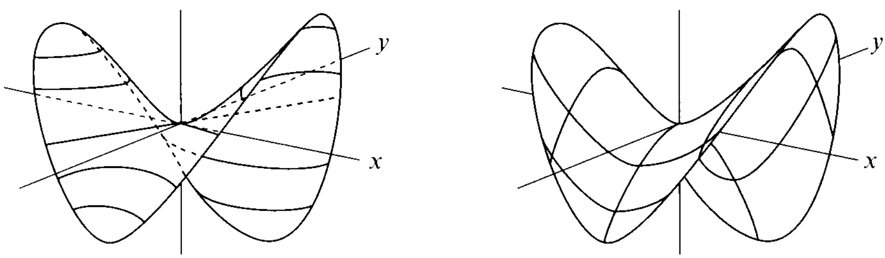
$$\frac{\tilde{h}(s,t)}{s^2+t^2} \geq A/2 > 0 \qquad \text{for sufficiently small } (s,t).$$

Therefore $\tilde{h}(s,t) > 0$ for small (s,t). Thus points of our surface which are near p lie on the same side of the tangent plane at p as $\nu(p)$. If $k_1, k_2 < 0$, then the graph of $\tilde{\alpha}$ is an elliptic paraboloid pointing in the other direction, and points of our surface which are near p lie on the other side of the tangent plane at p.

2. *Hyperbolic point*: $ln - m^2 < 0$. Then k_1, k_2 have opposite signs, say $k_1 > 0 > k_2$. The graph of

$$\tilde{\alpha}(s,t) = \frac{s^2}{\left(\sqrt{2/k_1}\right)^2} - \frac{t^2}{\left(\sqrt{-2/k_2}\right)^2}$$

is a *hyperbolic paraboloid*; planes parallel to the (x,y)-plane intersect the graph of $\tilde{\alpha}$ in similar hyperbolas [except that the (x,y)-plane itself intersects the graph in two straight lines through $(0,0)$], while planes parallel to the other coordinate planes intersect the graph in parabolas. It is easy to see that there are points

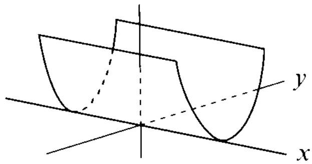


of the original surface arbitrarily close to p lying on both sides of the tangent plane at p.

3. *Parabolic point*: $ln - m^2 = 0$, *but not all of* l, m, n *are* 0. Then exactly one eigenvalue is 0; say $k_1 = 0$ but $k_2 \neq 0$. The graph of

$$\tilde{\alpha}(s,t) = \tfrac{1}{2}k_2 t^2$$

is a *parabolic cylinder.* If $k_2 > 0$, then our original surface must contain points close to p on the same side of the tangent plane as $\nu(p)$. But there can also be



points arbitrarily close to p on the other side of the tangent plane. For example, our surface might be the graph of

$$h(s,t) = s^3 + t^2.$$

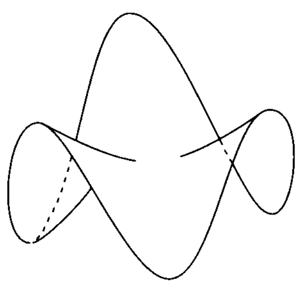
4. *Planar point*: $l = m = n = 0$. The graph of $\tilde{\alpha}$ is the (x, y)-plane.

In the planar case, nothing at all can be said about which side of the tangent plane our surface lies on. For example, our surface might be the graph of any one of the following functions:

$h(s,t) = s^4$	graph lies above the (x, y)-plane
$h(s,t) = -s^4$	graph lies below the (x, y)-plane
$h(s,t) = s^3$	graph lies above and below the (x, y)-plane.

A more interesting example of a planar point is provided by the "monkey saddle", the graph of

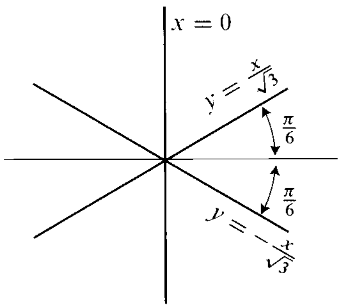
$$h(s,t) = s^3 - 3st^2 = \text{real part of } (s + it)^3,$$

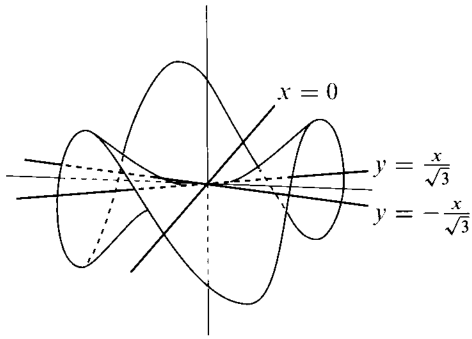


with a planar point at $0 \in \mathbb{R}^3$. I was always very confused by the name of this surface, because I thought it was supposed to be a saddle that you put on a monkey. Actually it's a saddle that a monkey *uses* (to ride a bicycle, say)—there are two depressions for its legs, and an extra one for its tail. The monkey saddle intersects the (x, y)-plane in the set

$$\{(x, y) : x^3 - 3xy^2 = 0\}$$

which consists of 3 straight lines all making equal angles with each other.





Notice that our classification of points on a surface as elliptic, hyperbolic, parabolic, or planar, does not at all depend on the special parameterization which we introduced; for equation (I) on page 35 shows that

$$\det(-d\nu\colon M_p \to M_p) = \frac{\det(l_{ij})}{\det(g_{ij})} = \frac{ln - m^2}{EG - F^2},$$

which means that the sign of $ln - m^2$ is always the same as the sign of $\det(-d\nu)$. When we do introduce our special parameterization, we have $E = G = 1$ and $F = 0$ at $(0, 0)$, so equation (I) then gives

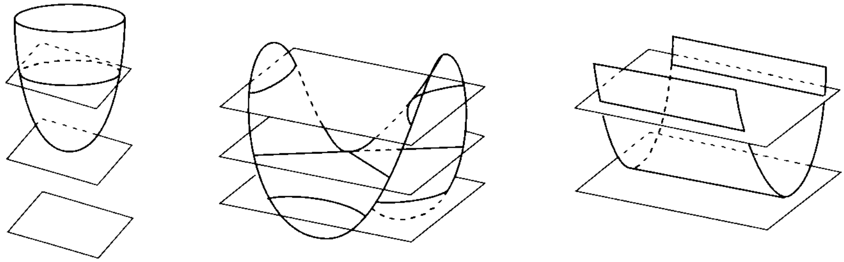
$$(l_{ij}) = \text{ matrix of } -d\nu\colon M_p \to M_p.$$

Consequently, the numbers k_1, k_2 which we have found can also be described invariantly as the eigenvalues of $-d\nu$; the orthonormal vectors X_1, X_2 (in $\mathbb{R}^2$, which we have identified with M_p) are just the eigenvectors of $-d\nu$.

The quadratic surface P which approximates M up to order 2 at a point $p \in M$ is called the **osculating paraboloid** at p; when $p = 0 \in \mathbb{R}^2$, the tangent plane at p is the (x, y)-plane, and X_1, X_2 point along the x- and y-axes, it is the graph of

$$\alpha(s, t) = \tilde{\alpha}(s, t) = \tfrac{1}{2}(k_1 s^2 + k_2 t^2).$$

For space curves we obtained an analogous osculating curve (pg. II.31), and we used this osculating curve to examine the original curve more closely by projecting it on the coordinate planes. That procedure wouldn't make much sense here, but there is something else we can do. Suppose we first intersect the osculating paraboloid with the two planes parallel to the (x, y)-plane and at distance d from it, and then project the intersection onto the (x, y)-plane.



We obtain the set

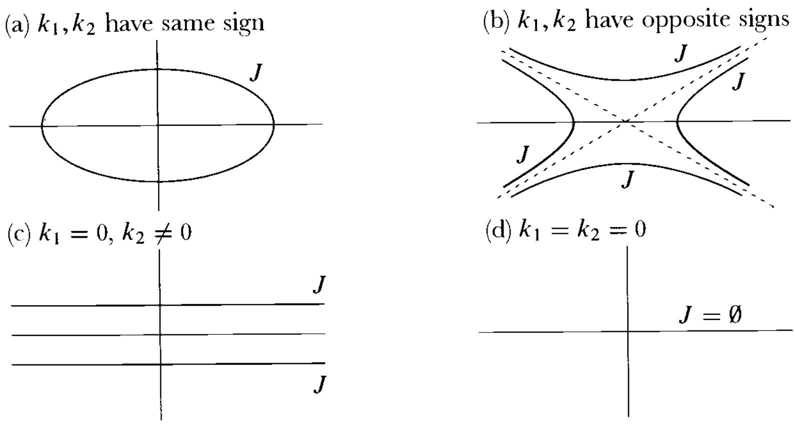
$$J_d = \{(x, y) : k_1x^2 + k_2y^2 = \pm 2d\},$$

which is either an ellipse, a pair of hyperbolas, a pair of parallel lines, or nothing. Clearly, the sets

$$\frac{J_d}{\sqrt{2d}} = \left\{\left(\frac{x}{\sqrt{2d}}, \frac{y}{\sqrt{2d}}\right) : (x, y) \in J_d\right\}$$

are all the same, namely

$$\frac{J_d}{\sqrt{2d}} = J = \{(x, y) : k_1x^2 + k_2y^2 = \pm 1\}.$$



Suppose now that we repeat this procedure, except that we intersect the two planes with our original surface instead of with its osculating paraboloid. We would expect the limiting set to be the same, since the surface agrees with its osculating paraboloid up to order 2. Actually, one has to be a little careful in formulating this result, and the corresponding proof is somewhat long, but not very interesting.

1. PROPOSITION. Let $p \in M$ be a point of an imbedded surface $M \subset \mathbb{R}^3$, and let X_1, X_2 be orthonormal eigenvectors of $-d\nu\colon M_p \to M_p$, with corresponding eigenvalues k_1, k_2. Let $I \subset M_p$ be the set

$$I = \{X \in M_p : k_1\langle X, X_1\rangle^2 + k_2\langle X, X_2\rangle^2 = \pm 1\},$$

so that I is congruent to

(a)	the ellipse $k_1x^2 + k_2y^2 = \pm 1$	p an elliptic point
(b)	the hyperbolas $k_1x^2 + k_2y^2 = \pm 1$	p a hyperbolic point
(c)	the parallel lines $k_2y^2 = \pm 1$	p a parabolic point with $k_2 \neq 0$
(d)	$\emptyset$	p a planar point.

For $d > 0$, let I_d be the projection on M_p of the intersection of M with the two planes parallel to M_p and at distance d from it. Then

$$\lim_{d\to 0} \frac{I_d}{\sqrt{2d}} = I,$$

where this limit has the following meaning: For every $\varepsilon > 0$, and every compact set $C \subset M_p$, there is some $\delta > 0$ such that if $0 < d < \delta$, then

(i) every point of $I \cap C$ is within ε of some point of $I_d/\sqrt{2d}$

(ii) every point of $\left(I_d/\sqrt{2d}\right) \cap C$ is within ε of some point of I.

Remark: In case (d), this just means that $I_d/\sqrt{2d}$ eventually lies outside of any compact set. In case (a), we clearly do not have to use the compact set C in condition (i), since I itself is compact. And in condition (ii) the compact set C is needed only to exclude points of I_d coming from extraneous points of M which are not near p.

PROOF. We assume that $p = 0 \in \mathbb{R}^3$, and that M_p is the (x, y)-plane, with the eigenvectors X_1, X_2 of $-d\nu(p)$ pointing along the x- and y-axes. Then M is locally the graph of a function $h\colon U \to \mathbb{R}$ (for $U \subset \mathbb{R}^2$ open), and we can assume that

$$I_d = \{(s,t) \in U : h(s,t) = \pm d\}$$

[there is no need to consider the points (s,t) outside U, since for sufficiently small d, the corresponding points $\left(s/\sqrt{2d}, t/\sqrt{2d}\right)$ lie outside of any given compact set $C \subset M_p$]. We thus have

$$I_d = \left\{(s,t) \in U : \frac{k_1s^2}{2} + \frac{k_2t^2}{2} + R(s,t) = \pm d\right\},$$

where

(1) $$\frac{R(s,t)}{s^2+t^2} \to 0 \qquad \text{as } (s,t) \to 0.$$

Hence

$$\frac{I_d}{\sqrt{2d}} = \left\{ (\sigma, \tau) : \begin{array}{c} (\sqrt{2d}\sigma, \sqrt{2d}\tau) \in U \text{ and} \\ k_1\sigma^2 + k_2\tau^2 + \dfrac{R(\sqrt{2d}\sigma, \sqrt{2d}\tau)}{d} = \pm 1 \end{array} \right\}.$$

Setting $s = \sqrt{2d}\sigma$ and $t = \sqrt{2d}\tau$ in (1), we see that

(2) $$\frac{R(\sqrt{2d}\sigma, \sqrt{2d}\tau)}{d(\sigma^2+\tau^2)} \to 0 \qquad \text{as} \qquad (\sqrt{2d}\sigma, \sqrt{2d}\tau) \to 0.$$

Now suppose we are given $\varepsilon > 0$, and a compact set $C \subset M_p$, which we might as well assume is of the form

$$C = \left\{(\sigma, \tau) : \sqrt{\sigma^2+\tau^2} \le A\right\}.$$

Choose $\varepsilon_0 > 0$ so that

(3) $$|\alpha| < \varepsilon_0 \implies |1 - \sqrt{1 \pm \alpha}| < \varepsilon/A.$$

Let $(\sigma, \tau) \in C \cap I$, so that

(4) $$k_1\sigma^2 + k_2\tau^2 = \pm 1 \quad \text{and} \quad \sqrt{\sigma^2+\tau^2} \le A.$$

Consider the function

(5) $$d \mapsto \frac{R(\sqrt{2d}\sigma, \sqrt{2d}\tau)}{d} \qquad 0 < d \le 1.$$

This is continuous in d, and approaches 0 as $d \to 0^+$, by (2). So there is $\delta > 0$ such that

(6) $$0 < d < \delta \implies \left| \frac{R(\sqrt{2d}\sigma, \sqrt{2d}\tau)}{d} \right| < \varepsilon_0.$$

Let

$$\alpha = \frac{R(\sqrt{2d}\sigma, \sqrt{2d}\tau)}{d},$$

and consider the point

$$(\sigma\sqrt{1 \mp \alpha}, \tau\sqrt{1 \mp \alpha}) = \begin{cases} (\sigma\sqrt{1-\alpha}, \tau\sqrt{1-\alpha}) & \text{if } +1 \text{ holds in equation (4)} \\ (\sigma\sqrt{1+\alpha}, \tau\sqrt{1+\alpha}) & \text{if } -1 \text{ holds in equation (4).} \end{cases}$$

We have

$$k_1(\sigma\sqrt{1-\alpha})^2 + k_2(\tau\sqrt{1-\alpha})^2 = (1-\alpha)[k_1\sigma^2 + k_2\tau^2] = 1-\alpha$$
$$k_1(\sigma\sqrt{1+\alpha})^2 + k_2(\tau\sqrt{1+\alpha})^2 = (1+\alpha)[k_1\sigma^2 + k_2\tau^2] = -1-\alpha,$$

so $(\sigma\sqrt{1 \mp \alpha}, \tau\sqrt{1 \mp \alpha})$ is in $I_d/\sqrt{2d}$. Its distance from (σ, τ) is

$$\left|1 - \sqrt{1 \mp \alpha}\right|\sqrt{\sigma^2+\tau^2} < \frac{\varepsilon}{A}\sqrt{\sigma^2+\tau^2} \qquad \text{by (3), since } |\alpha| < \varepsilon_0 \quad \text{by (6)}$$
$$< \varepsilon \qquad \text{by (4).}$$

We have thus found a $\delta > 0$ so that the given point $(\sigma, \tau) \in C \cap I$ has distance less than ε from a point of $I_d/\sqrt{2d}$, for all $d < \delta$. To conclude that one δ can be found which works for all $(\sigma, \tau) \in C \cap I$ we need the fact that the function (5) approaches 0 uniformly in (σ, τ); this follows from compactness of C.

We will now prove (ii). Given $\varepsilon > 0$ and $A > 0$, pick $\varepsilon_0 > 0$ so that

$$(7) \qquad |\alpha| < \varepsilon_0 \implies \left|1 - \frac{1}{\sqrt{1 \pm \alpha}}\right| < \frac{\varepsilon}{A}.$$

Then pick $\delta_0 > 0$ so that

$$s^2 + t^2 < \delta_0 \implies \frac{R(s,t)}{s^2+t^2} < \frac{\varepsilon_0}{2A^2}.$$

Setting $s = \sqrt{2d}\sigma$ and $t = \sqrt{2d}\tau$, we see that

$$(8) \qquad \sigma^2 + \tau^2 < \frac{\delta_0}{2d} \implies \frac{R(\sqrt{2d}\sigma, \sqrt{2d}\tau)}{d} < \frac{\varepsilon_0}{A^2}(\sigma^2+\tau^2).$$

Let $\delta = \delta_0/2A^2$. Then

$$0 < d < \delta \implies \frac{\delta_0}{2d} > A^2.$$

So if $0 < d < \delta$, then

$$\text{(9)} \qquad \textit{either} \quad \sigma^2 + \tau^2 > A^2 \quad \textit{or} \quad \sigma^2 + \tau^2 < A^2 < \frac{\delta_0}{2d}.$$

In the first case, the point $(\sigma, \tau) \in I_d/\sqrt{2d}$ is at distance $> A$ from the origin. In the second case, we have

$$\begin{aligned} \frac{R(\sqrt{2d}\sigma, \sqrt{2d}\tau)}{d} &< \frac{\varepsilon_0}{A^2}(\sigma^2 + \tau^2) && \text{by (8)} \\ &< \varepsilon_0 && \text{by (9),} \end{aligned}$$

so the point $(\sigma, \tau) \in I_d/\sqrt{2d}$ satisfies the equation

$$\text{(10)} \qquad \begin{aligned} k_1\sigma^2 + k_2\tau^2 &= \pm 1 - \frac{R(\sqrt{2d}\sigma, \sqrt{2d}\tau)}{d} \\ &= \pm 1 - \alpha, \qquad \text{where } 0 \leq |\alpha| < \varepsilon_0. \end{aligned}$$

Now the point

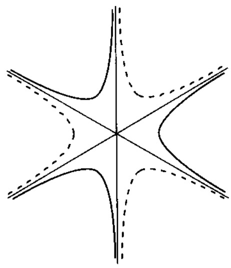
$$\left(\frac{\sigma}{\sqrt{1 \mp \alpha}}, \frac{\tau}{\sqrt{1 \mp \alpha}} \right)$$

is in I, and its distance from (σ, τ) is

$$\begin{aligned} \left| 1 - \frac{1}{\sqrt{1 \mp \alpha}} \right| \sqrt{\sigma^2 + \tau^2} &< \frac{\varepsilon}{A}\sqrt{\sigma^2 + \tau^2} && \text{by (7), since } |\alpha| < \varepsilon_0 \text{ by (10)} \\ &< \varepsilon && \text{by (9).} \end{aligned}$$

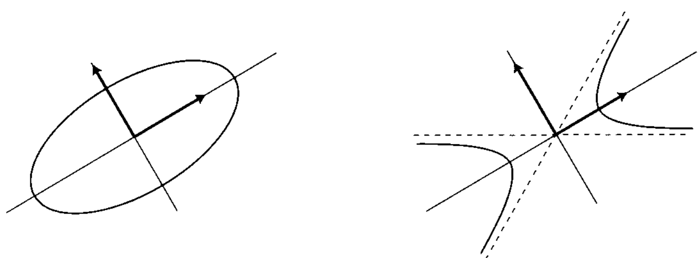
This completes the proof. ❖

The limiting set $I \subset M_p$ of Proposition 1 is called the **Dupin indicatrix** at p. In the case of a planar point, we can obtain a more meaningful indicatrix by considering $\lim_{d \to 0} I_d/\sqrt[3]{d}$—the adjustment factor $1/\sqrt[3]{d}$ is just what we need in order that the projection on M_p of the intersections of parallel planes with the osculating *cubic* will be the same. The figure below shows the resulting indicatrix for the monkey saddle; the continuous lines come from intersections with planes



on one side of the tangent plane, and the dashed lines from intersections with planes on the other side. Similarly, if all derivatives of $h: U \to \mathbb{R}$ up to order $k-1$ are 0 at $(0,0)$, then we can look at the generalized indicatrix $\lim_{d\to 0} I_d/\sqrt[k]{d}$.

Certain geometric terminology concerning conic sections has been taken over, via the Dupin indicatrix, to surfaces. Given a conic section in the plane, the directions which we have chosen as the x- and y- axes are called its **principal axes**. Consequently, the unit vectors $X_1, X_2 \in M_p$ (that is, the unit eigenvectors for $-d\nu: M_p \to M_p$) are called the **principal vectors**. They are really defined



only up to sign, so it is often more convenient to speak of the **principal directions**; moreover, if $k_1 = k_2$, then all unit vectors are to be considered to be principal. The eigenvalues k_1 and k_2 are called the **principal curvatures** at p. We have already met these vectors and curvatures in Volume II, and we recall that if $X = (\cos\theta)X_1 + (\sin\theta)X_2$ is any other unit vector, then

$$\text{(II)}\quad \begin{aligned}\langle -d\nu(X), X\rangle &= \langle k_1(\cos\theta)X_1 + k_2(\sin\theta)X_2,\ (\cos\theta)X_1 + (\sin\theta)X_2\rangle \\ &= k_1\cos^2\theta + k_2\sin^2\theta,\end{aligned}$$

which shows that k_1, k_2 are the minimum and maximum of $\langle -d\nu(X), X\rangle$ for unit vectors $X \in M_p$. Recall also that we defined

$$K(p) = \text{Gaussian curvature at } p = k_1 \cdot k_2$$
$$H(p) = \text{mean curvature at } p = \tfrac{1}{2}(k_1 + k_2).$$

A surface M is called **flat at** p if $K(p) = 0$. So p is a flat point if and only if p is either parabolic or planar.

For hyperbolas, there are two other important lines, the *asymptotes* (the dashed lines in the previous figure). The unit vectors which point along these lines in the Dupin indicatrix are called the **asymptotic directions**. If the hyperbola has the equation

$$k_1x^2 + k_2y^2 = \pm 1 \qquad (k_1, k_2 \text{ of different signs}),$$

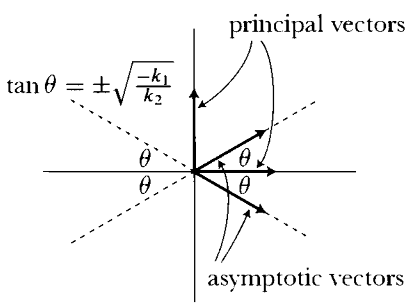
then the equation of the asymptotic lines $y = mx$ is found by noting that for large x the point (x, mx) is almost on the hyperbola, so

$$k_1x^2 + k_2m^2x^2 \quad \text{is close to } \pm 1 \implies k_1 + k_2m^2 \quad \text{is close to } 0,$$

and hence $m = \pm\sqrt{-k_1/k_2}$. On the other hand, if we consider a unit vector $X = (\cos\theta)X_1 + (\sin\theta)X_2$, then formula (II) gives

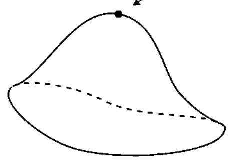
$$\langle -d\nu(X), X\rangle = k_1\cos^2\theta + k_2\sin^2\theta,$$

and this clearly equals 0 precisely when $\tan\theta = \pm\sqrt{-k_1/k_2}$. We therefore see that the vector $X \in M_p$ points along an asymptotic direction if and only if $\langle d\nu(X), X\rangle = 0$, and hence $\mathrm{II}(X, X) = 0$.



Asymptotic directions do not exist at elliptic points, while there are two distinct asymptotic directions at hyperbolic points, and these directions are bisected by the principal directions. At a parabolic point there is only *one* asymptotic direction—the principal direction with principal curvature 0. At planar points all directions are both principal and asymptotic. It is also clear that the asymptotic directions are perpendicular precisely when $k_1 = -k_2$, or $H = 0$.

Finally, there is one more important term, which describes a point where the Dupin indicatrix is actually a circle, so that the principal curvatures are equal, and all directions are principal directions. A point where all directions are principal is called an **umbilic** or **navel point**. This rather gross anatomical metaphor



is meant to suggest that the surface is very round at the point, like a sphere, on which all points are umbilics. Notice, however, that our definition also makes planar points umbilics, which turns out to be a convenient arrangement. At an umbilic, the map $-d\nu$ is just multiplication by some number k; equation (I) therefore shows that

$$l_{ij} = k g_{ij} \qquad \text{at an umbilic.}$$

At this stage it seems reasonable to begin asking to what extent the functions l, m, n describe f globally. The simplest question we can ask concerns surfaces f all of whose points are planar. Just as a curve with everywhere 0 curvature is a straight line, so we would expect a surface with $l = m = n = 0$ everywhere to be a plane. This is easy to prove. For, $l_{ij} = 0$ means $\langle -N_i, f_j \rangle = 0$; since N_i is a linear combination of f_1, f_2, this means that $-N_i = 0$, so N is constant. Therefore $\langle f, N \rangle_i = \langle f_i, N \rangle + \langle f, N_i \rangle = 0 + 0$, and hence $\langle f, N \rangle = b$, where N is a constant vector and b is a constant number. This is just the equation of a plane.

It would next seem reasonable to prove an analogue of the fact that a circle is the only plane curve with constant κ. Here the situation is a little different, however: we cannot expect to characterize a sphere totally in terms of l, m, n, because we have not picked out a preferred parameterization; the functions E, F, G must also play a role. In fact, the simplest criterion to consider is that all points of f be umbilics. This means that

$$l = kE, \quad m = kF, \quad n = kG,$$

for a certain function k. Of course, in the case of a sphere, k is constant, but we do not even have to assume that. Although the following analysis is quite easy, it is worth recording as a theorem, which also includes the result about surfaces with all points planar.

2. THEOREM. If $M \subset \mathbb{R}^3$ is a connected surface such that every point is an umbilic, then M is part of a plane or a sphere.

PROOF. Choose an immersion $f\colon U \to M$. By assumption, we have $\langle -N_i, f_j \rangle = \langle k f_i, f_j \rangle$. Since N_i is a linear combination of f_1, f_2, we thus have

$$N_i = -k f_i. \tag{1}$$

Consequently,

$$N_{ij} = -k_j f_i - k f_{ij}.$$

Since $N_{ij} = N_{ji}$, we obtain

$$-k_j f_i - k f_{ij} = -k_i f_j - k f_{ij},$$

and hence $k_i f_j = k_j f_i$. Setting $i = 1$, $j = 2$, and using linear independence of f_1, f_2, we obtain $k_i = 0$, so k is constant. Thus equation (1) gives

$$N = -kf + v_0 \quad \text{for some } v_0 \in \mathbb{R}^3. \tag{2}$$

If $k = 0$, then as we already showed above, f lies in a plane. If $k \neq 0$, we have

$$f - \frac{v_0}{k} = \frac{-N}{k} \implies \left| f - \frac{v_0}{k} \right|^2 = \frac{1}{k^2},$$

so f lies in a sphere of radius $1/|k|$. Simple supplementary considerations then allow one to deduce the stated result. ❖

We now want to carry the analogy with curves still further, and see whether every immersion $f\colon U \to \mathbb{R}^3$ is described completely by the corresponding g_{ij} and l_{ij}. Of course, we only expect g_{ij} and l_{ij} to determine f up to proper Euclidean motions (translations followed by rotations [elements of SO(3)]), since g_{ij} and l_{ij} are already "invariant under proper Euclidean motions"— if A is a proper Euclidean motion, then the g_{ij} and l_{ij} for $A \circ f$ are the same as those for f. In the theory of curves we showed that κ and τ formed a complete set of invariants for a curve up to translations and rotations, by showing that they were a complete set of invariants up to rotation for the function

$s \mapsto (\mathbf{t}(s), \mathbf{n}(s), \mathbf{b}(s))$; this was accomplished by using the Serret-Frenet formulas, which are differential equations for $(\mathbf{t}, \mathbf{n}, \mathbf{b})$, involving only κ and τ. In the case of surfaces, we have the three vectors (f_1, f_2, N), and so we want first to express the derivatives of each of these vectors as linear combinations of these same three vectors.

We begin by considering the f_{ik}, which we want to write as

$$f_{ik} = \sum_{h=1}^{2} A_{ik}^h f_h + B_{ik} N. \tag{1}$$

First we will worry about finding the A_{ik}^h, which amounts to finding the $\langle f_{ik}, f_j \rangle$. To do this, we should obviously begin with the definition $\langle f_i, f_j \rangle = g_{ij}$ and differentiate, to get

$$\langle f_{ik}, f_j \rangle + \langle f_i, f_{jk} \rangle = g_{ij,k} \qquad \left[\text{here } g_{ij,1} = D_1 g_{ij} = \frac{\partial g_{ij}}{\partial s}, \text{ etc.}\right].$$

Now comes the familiar old switcheroo: We also have

$$\begin{aligned} \langle f_{ji}, f_k \rangle + \langle f_j, f_{ki} \rangle &= g_{jk,i} \\ \langle f_{kj}, f_i \rangle + \langle f_k, f_{ij} \rangle &= g_{ki,j}; \end{aligned}$$

adding the first two of these three equations and subtracting the third, we get

$$\begin{aligned} \langle f_{ik}, f_j \rangle &= \frac{1}{2}(g_{ij,k} + g_{jk,i} - g_{ik,j}) \\ &= [ik, j], \end{aligned}$$

where $[ik, j]$ is the Christoffel symbol for the metric $\mathrm{I}_f = f^*\langle\ ,\ \rangle$ on U with respect to the standard coordinate system (s, t) on $\mathbb{R}^2$. Plugging back into (1) we have

$$[ik, j] = \langle f_{ik}, f_j \rangle = \sum_{h=1}^{2} A_{ik}^h g_{hj},$$

and, of course, we can solve explicitly for A_{ik}^h, using the g^{ij}:

$$A_{ik}^{\rho} = \sum_{j=1}^{2} g^{\rho j}[ik, j] = \Gamma_{ik}^{\rho}.$$

There is no problem finding the B_{ik}, since we have already introduced a name for them:

$$B_{ik} = \langle f_{ik}, N \rangle = -\langle N_i, f_k \rangle = l_{ik}.$$

We have thus found that

$$(*) \qquad f_{ik} = \sum_{h=1}^{2} \Gamma_{ik}^{h} f_h + l_{ik} N \qquad \text{The Gauss Formulas.}$$

The reader can easily check that these equations are indeed precisely the Gauss Formulas on page 4. Of course, in our present derivation, it is unnecessary to use the ∇ operator on the image of f, or even to have it defined; the only ∇ operator one needs to know about is the one for $\mathbb{R}^3$, and the Γ's just appear as weird combinations of the g_{ij} and their derivatives. (It is hard to see why these formulas should be named after Gauss, for he never explicitly solves for the f_{ik}. The closest results he writes down are certain formulas [for m, m', m'', n, n', n'', on pg. II.91] equivalent to the equations $\langle f_{ik}, f_j \rangle = [ik, j]$.)

We next want to express N_i in terms of f_1, f_2, N, as

$$N_i = \sum_{h=1}^{2} C_i^h f_h + 0 \cdot N.$$

Once again, there is no problem here, since we have already introduced a name for the relevant inner products. We have

$$l_{ij} = \langle -N_i, f_j \rangle = -\sum_{h=1}^{2} C_i^h g_{hj},$$

and consequently

$$C_i^\rho = -\sum_{j=1}^{2} g^{\rho j} l_{ij}.$$

Introducing new symbols l_i^h, we can therefore write

$$(**) \quad N_i = -\sum_{h=1}^{2} \left(\sum_{j=1}^{2} g^{hj} l_{ij} \right) f_h = -\sum_{h=1}^{2} l_i^h f_h \qquad \text{The Weingarten Equations.}$$

Of course, equations (**) amount to little more than the definition of l_{ij} and l_i^h, but in the classical literature it is always precisely these equations which are called the Weingarten equations.

The Gauss and Weingarten equations constitute an exact analogue of the Serret-Frenet formulas for a curve—the derivatives of f_1, f_2, N have been expressed in terms of these same vectors, and only the g_{ij} and l_{ij} enter. We thus seem to be in a good position to produce an immersion f with given g_{ij} and l_{ij}: we should first solve the Gauss and Weingarten equations for f_1, f_2, N, and then solve for f. However, these equations are *partial* differential equations (just 15 in all, for the 9 component functions f_i^j, n^j), and we know that these equations have solutions only if certain compatibility conditions are satisfied. These conditions are given explicitly on pg. I.187, but there is no need to turn back to them: the required conditions are obtained simply by setting mixed partial derivatives equal, and substituting the original equation into the results so obtained. We will now derive these conditions explicitly.

We begin by using (∗) to compute

$$f_{ikj} = \sum_{h=1}^{2} \Gamma_{ik,j}^{h} f_h + \sum_{h=1}^{2} \Gamma_{ik}^{h} f_{hj} + l_{ik,j} N + l_{ik} N_j$$

$$[\text{here } \Gamma_{ik,1}^{h} = \partial \Gamma_{ik}^{h}/\partial s, \text{ etc.}]$$

$$= \sum_{\rho=1}^{2} \Gamma_{ik,j}^{\rho} f_\rho + \sum_{h=1}^{2} \Gamma_{ik}^{h} \left(\sum_{\rho=1}^{2} \Gamma_{hj}^{\rho} f_\rho + l_{hj} N \right)$$

$$+ l_{ik,j} N - l_{ik} \left(\sum_{\rho=1}^{2} l_j^\rho f_\rho \right), \qquad \text{using } (*) \text{ and } (**).$$

Setting $f_{ikj} = f_{ijk}$, and using linear independence of f_1, f_2, N, we have

$$\text{(A)} \qquad \Gamma_{ik,j}^{\rho} - \Gamma_{ij,k}^{\rho} + \sum_{h=1}^{2} (\Gamma_{ik}^{h} \Gamma_{hj}^{\rho} - \Gamma_{ij}^{h} \Gamma_{hk}^{\rho}) = l_{ik} l_j^\rho - l_{ij} l_k^\rho$$

$$\text{(B)} \qquad l_{ik,j} - l_{ij.k} + \sum_{h=1}^{2} \Gamma_{ik}^{h} l_{hj} - \sum_{h=1}^{2} \Gamma_{ij}^{h} l_{hk} = 0.$$

In this mess, some things should be looking familiar. Indeed, comparing with pg. II.188, we see that equation (A) says that

$$R^\rho{}_{kji} = l_{ik} l_j^\rho - l_{ij} l_k^\rho,$$

which is equivalent to

$$\text{(A}') \qquad \begin{aligned} R_{hkji} &= \sum_{\rho=1}^{2} g_{h\rho} R^{\rho}{}_{kji} = \sum_{\rho=1}^{2} g_{h\rho}(l_j^{\rho} l_{ik} - l_k^{\rho} l_{ij}) \\ &= \sum_{\rho=1}^{2} g_{h\rho}\left(\sum_{\sigma=1}^{2} g^{\rho\sigma} l_{\sigma j} l_{ik} - \sum_{\sigma=1}^{2} g^{\rho\sigma} l_{\sigma k} l_{ij}\right) \\ &\qquad\qquad \text{by the definition of } l_j^{\rho} \text{ in } (**) \\ &= l_{hj} l_{ik} - l_{hk} l_{ij}. \end{aligned}$$

A special case is

$$R_{1212} = l_{11} l_{22} - l_{12} l_{12} = ln - m^2 \qquad \begin{array}{l}\text{Gauss' Equation} \\ \text{(Gauss' Theorema Egregium).}\end{array}$$

This really is equivalent to Gauss' Theorema Egregium, for it says (cf. pg. II.190) that

$$\left\langle R\left(\frac{\partial f}{\partial s}, \frac{\partial f}{\partial t}\right)\frac{\partial f}{\partial t}, \frac{\partial f}{\partial s}\right\rangle = ln - m^2,$$

and hence that the intrinsically defined Gaussian curvature K is given by

$$K = \frac{\left\langle R\left(\frac{\partial f}{\partial s}, \frac{\partial f}{\partial t}\right)\frac{\partial f}{\partial t}, \frac{\partial f}{\partial s}\right\rangle}{\left\langle\frac{\partial f}{\partial s}, \frac{\partial f}{\partial s}\right\rangle\left\langle\frac{\partial f}{\partial t}, \frac{\partial f}{\partial t}\right\rangle - \left\langle\frac{\partial f}{\partial s}, \frac{\partial f}{\partial t}\right\rangle^2} = \frac{ln - m^2}{EG - F^2},$$

the final expression being the Gaussian curvature as originally (extrinsically) defined for a surface in $\mathbb{R}^3$. In Volume II we gave a simpler looking proof of this result, but the present proof is philosophically more satisfying, since it relies only on standard techniques for dealing with a system of partial differential equations. Notice that our proof of Theorem 1-6 was basically the same, since it used the fact that $R(X, Y)Z$ measures the difference of $\nabla_X \nabla_Y Z$ and $\nabla_Y \nabla_X Z$.

It is easy to see that all other cases of (A′) are equivalent to this particular one, or are trivial, because of the identities

$$R_{ijkl} = -R_{jikl} \quad \text{and} \quad R_{klij} = R_{ijkl},$$

which always hold (pp. II.194ff.), and the fact that the right side of (A′) has the same symmetry properties.

Now let us take a look at (B). If $j = k$ it says nothing. Moreover, the equation for $j = 2$, $k = 1$ is equivalent to the one for $j = 1$, $k = 2$. So we take the latter pair for j and k, and let $i = 1$ or 2, obtaining

$$(\mathrm{B}')\quad \begin{cases} l_{12,1} - l_{11,2} + \sum_{h=1}^{2} \Gamma_{12}^{h} l_{h1} - \sum_{h=1}^{2} \Gamma_{11}^{h} l_{h2} = 0 \\ l_{22,1} - l_{21,2} + \sum_{h=1}^{2} \Gamma_{22}^{h} l_{h1} - \sum_{h=1}^{2} \Gamma_{21}^{h} l_{h2} = 0 \end{cases}$$

The Codazzi-Mainardi Equations.

It is easy to see (Problem 1) that these equations can be derived from the ones given in Corollary 1-12. [Note also that equations (A′) and (B′) are precisely what the classical tensor analysis equations (11) and (12) on page 16 become in this case.]

There is still one more set of equations which we must consider, obtained by setting $N_{ij} = N_{ji}$. However (Problem 2), it turns out that these reduce to the Codazzi-Mainardi equations. We have thus found altogether three conditions which must be satisfied, and our general theory (Theorem I.6-1) tells us that these are the only conditions we need. We are all ready for a theorem.

3. FUNDAMENTAL THEOREM OF SURFACE THEORY (BONNET; 1867). Let $U \subset \mathbb{R}^2$ be a convex open set containing $(0,0)$.

(1) Let $f, \bar{f}\colon U \to \mathbb{R}^3$ be two immersions, and define

$$g_{ij} = \langle f_i, f_j \rangle \qquad \bar{g}_{ij} = \langle \bar{f}_i, \bar{f}_j \rangle$$

$$N = \frac{f_1 \times f_2}{\sqrt{g_{11}g_{22} - {g_{12}}^2}} \qquad \bar{N} = \frac{\bar{f}_1 \times \bar{f}_2}{\sqrt{\bar{g}_{11}\bar{g}_{22} - {\bar{g}_{12}}^2}}$$

$$l_{ij} = \langle -N_i, f_j \rangle = \langle N, f_{ij} \rangle \qquad \bar{l}_{ij} = \langle -\bar{N}_i, \bar{f}_j \rangle = \langle \bar{N}, \bar{f}_{ij} \rangle.$$

Suppose that $g_{ij} = \bar{g}_{ij}$ and $l_{ij} = \bar{l}_{ij}$ on U. Then there is a proper Euclidean motion A such that $\bar{f} = A \circ f$.

(2) Let g_{ij} and l_{ij} $(i, j = 1, 2)$ be functions on U which satisfy

(i) $g_{ij} = g_{ji}$ and $l_{ij} = l_{ji}$, and (g_{ij}) is positive definite on U, so that we can define corresponding g^{ij} and Γ_{ij}^{k}

(ii) Gauss' Equation:

$$l_{11}l_{22} - (l_{12})^2 = R_{1212}$$
$$= \sum_{\rho=1}^{2} g_{1\rho}\left(\Gamma^{\rho}_{22,1} - \Gamma^{\rho}_{21,2} + \sum_{h=1}^{2}(\Gamma^{h}_{22}\Gamma^{\rho}_{h1} - \Gamma^{h}_{21}\Gamma^{\rho}_{h2})\right)$$

(iii) The Codazzi-Mainardi Equations:

$$l_{12,1} - l_{11,2} + \sum_{h=1}^{2}\Gamma^{h}_{12}l_{h1} - \sum_{h=1}^{2}\Gamma^{h}_{11}l_{h2} = 0$$
$$l_{22,1} - l_{21,2} + \sum_{h=1}^{2}\Gamma^{h}_{22}l_{h1} - \sum_{h=1}^{2}\Gamma^{h}_{21}l_{h2} = 0.$$

Then there is an immersion $f\colon U \to \mathbb{R}^3$ such that

$$g_{ij} = \langle f_i, f_j\rangle$$
$$l_{ij} = \langle -N_i, f_j\rangle = \langle N, f_{ji}\rangle, \quad \text{for } N = \frac{f_1 \times f_2}{\sqrt{g_{11}g_{22} - g_{12}{}^2}}.$$

PROOF. Let us adopt the more systematic notation

$$\mathbf{v}_1 = f_1, \quad \mathbf{v}_2 = f_2, \quad \mathbf{v}_3 = N$$
$$\bar{\mathbf{v}}_1 = \bar{f}_1, \quad \bar{\mathbf{v}}_2 = \bar{f}_2, \quad \bar{\mathbf{v}}_3 = \bar{N}.$$

To prove (1), we first choose a rotation $B \in \mathrm{SO}(3)$ such that

$$B(\mathbf{v}_\alpha(0,0)) = \bar{\mathbf{v}}_\alpha(0,0) \qquad \alpha = 1,2,3.$$

This is possible because $g_{ij}(0) = \bar{g}_{ij}(0)$ for $i,j = 1,2$, and because the two triples of vectors $(\mathbf{v}_1(0,0), \mathbf{v}_2(0,0), \mathbf{v}_3(0,0))$ and $(\bar{\mathbf{v}}_1(0,0), \bar{\mathbf{v}}_2(0,0), \bar{\mathbf{v}}_3(0,0))$ are both positively oriented, with the third vector perpendicular to the first two. If we let $\tilde{f} = B \circ f$, then it is easy to see that

$$\tilde{g}_{ij} = g_{ij} = \bar{g}_{ij}$$
$$\tilde{\mathbf{v}}_3 = B \circ \mathbf{v}_3$$
$$\tilde{l}_{ij} = l_{ij} = \bar{l}_{ij}.$$

We claim that the maps

$$(\bar{\mathbf{v}}_1, \bar{\mathbf{v}}_2, \bar{\mathbf{v}}_3),\ (\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \tilde{\mathbf{v}}_3)\colon U \to \mathbb{R}^3,$$

which we know are equal at $(0,0)$, are actually equal everywhere. To prove this, we recall that the Gauss formulas and the Weingarten equations give

$$(***)\qquad \begin{cases} \bar{\mathbf{v}}_{i,k}(s,t) = \displaystyle\sum_{h=1}^{2} \bar{\Gamma}_{ik}^{h}(s,t)\bar{\mathbf{v}}_h(s,t) + \bar{l}_{ik}(s,t)\bar{\mathbf{v}}_3(s,t) & i = 1,2 \\ \bar{\mathbf{v}}_{3,k}(s,t) = -\displaystyle\sum_{h=1}^{2}\left(\sum_{j=1}^{2} \bar{g}^{hj}(s,t)\bar{l}_{kj}(s,t)\right)\bar{\mathbf{v}}_h(s,t) \end{cases}$$

for the $\bar{\mathbf{v}}_\alpha$, while for the $\tilde{\mathbf{v}}_\alpha$ we obtain the corresponding equations with $\tilde{\Gamma}_{ik}^{h}$, $\tilde{l}_{ik}$ and $\tilde{g}^{hj}$. But $\tilde{l}_{ik} = \bar{l}_{ik}$, and since $\tilde{g}_{ij} = \bar{g}_{ij}$ we also have $\tilde{g}^{hj} = \bar{g}^{hj}$ and $\tilde{\Gamma}_{ik}^{h} = \bar{\Gamma}_{ik}^{h}$. So the two maps $(\bar{\mathbf{v}}_1, \bar{\mathbf{v}}_2, \bar{\mathbf{v}}_3)$ and $(\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \tilde{\mathbf{v}}_3)$ satisfy the *same* equations (***) and have the same values at $(0,0)$. Therefore they must be equal on U (Theorem I.6-1). But this means that $\bar{f}$ and $\tilde{f} = B \circ f$ have the same partial derivatives, and therefore differ by a constant vector. Consequently, there is a translation $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ with $\bar{f} = T \circ \tilde{f} = (T \circ B) \circ f$.

To prove (2), we use Theorem I.6-1 to conclude that equation (***), written in terms of the given g_{ij} and l_{ij}, has a solution $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\colon U \to \mathbb{R}^3$ with any desired initial conditions; we have already seen that the required conditions in Theorem I.6-1 amount precisely to Gauss' equation and the Codazzi-Mainardi equations. Moreover, the functions $\mathbf{v}_\alpha$ can be defined on all of U because the equations (***) are linear (compare pg. I.165). Since (g_{ij}) is positive definite at $(0,0)$, there is a solution for which the the following conditions are satisfied at $(s,t) = (0,0)$:

(a) $\langle \mathbf{v}_i(s,t), \mathbf{v}_j(s,t)\rangle = g_{ij}(s,t)$ $\qquad i,j = 1,2$

(b) $\langle \mathbf{v}_i(s,t), \mathbf{v}_3(s,t)\rangle = 0$ $\qquad i = 1,2$

(c) $|\mathbf{v}_3(s,t)| = 1$

(d) $(\mathbf{v}_1(s,t), \mathbf{v}_2(s,t), \mathbf{v}_3(s,t))$ is positively oriented.

We will show that conditions (a)–(d) actually hold at all points of U.

Our equation (***) for $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ gives the equations

$$\text{(A)}\quad \langle \mathbf{v}_i, \mathbf{v}_j\rangle_k = \langle \mathbf{v}_{i,k}, \mathbf{v}_j\rangle + \langle \mathbf{v}_i, \mathbf{v}_{jk}\rangle$$
$$= \sum_{h=1}^{2} \Gamma_{ik}^{h}\langle \mathbf{v}_h, \mathbf{v}_j\rangle + \sum_{h=1}^{2} \Gamma_{jk}^{h}\langle \mathbf{v}_h, \mathbf{v}_i\rangle + l_{ik}\langle \mathbf{v}_3, \mathbf{v}_i\rangle + l_{jk}\langle \mathbf{v}_3, \mathbf{v}_j\rangle$$

for $i, j = 1, 2$, as well as

$$\text{(B)} \qquad \langle \mathbf{v}_i, \mathbf{v}_3 \rangle_k = \langle \mathbf{v}_{i,k}, \mathbf{v}_3 \rangle + \langle \mathbf{v}_i, \mathbf{v}_{3,k} \rangle = l_{ik} - \sum_{h=1}^{2} \left(\sum_{j=1}^{2} g^{hj} l_{kj} \right) \langle \mathbf{v}_i, \mathbf{v}_h \rangle$$

and

$$\text{(C)} \qquad \langle \mathbf{v}_3, \mathbf{v}_3 \rangle_k = 2 \langle \mathbf{v}_{3,k}, \mathbf{v}_3 \rangle = 0.$$

[Equations (A)–(C) all hold for $k = 1, 2$.]

But we also have

$$g_{ij,k} = [ik, j] + [jk, i] \qquad \text{(by pg. I.331)}$$
$$= \sum_{h=1}^{2} \Gamma_{ik}^{h} g_{hj} + \sum_{h=1}^{2} \Gamma_{jk}^{h} g_{hi}.$$

This shows that the set of equations (A)–(C) are satisfied both by

the set of functions: $\langle \mathbf{v}_i, \mathbf{v}_j \rangle$ $(j = 1, 2)$, $\langle \mathbf{v}_3, \mathbf{v}_1 \rangle$, $\langle \mathbf{v}_3, \mathbf{v}_2 \rangle$, $\langle \mathbf{v}_3, \mathbf{v}_3 \rangle$

and by

the set of functions: g_{ij} $(j = 1, 2)$, 0, 0, 1.

Moreover, we chose the $\mathbf{v}_i$ so that these two collections of functions have the same value at $(0, 0)$. It follows that they have the same values on all of U. In other words, equations (a)–(c) hold on all of U. Moreover, (a) and (b) [and non-singularity of (g_{ij})] imply that $(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ are always linearly independent. So condition (d) at $(0, 0)$ implies condition (d) everywhere.

We now claim that there is a function $f\colon U \to \mathbb{R}^3$ satisfying $f_i = \mathbf{v}_i$. In order to prove this, we just have to show that $\mathbf{v}_{i,j} = \mathbf{v}_{j,i}$. But this follows from (***), by symmetry of the Γ_{ik}^{h} and l_{ik}. We now have $\langle f_i, f_j \rangle = g_{ij}$ by (a). Moreover, (b)–(d) then show that $\mathbf{v}_3 = n$. Consequently,

$$\langle f_{ij}, n \rangle = \langle \mathbf{v}_{i,j}, \mathbf{v}_3 \rangle = l_{ij}$$

by (***), together with (b) and (c). ❖

Theorem 3 is exactly the sort of result we would want if we were primarily interested in immersions $f\colon U \to \mathbb{R}^3$. But what we really want to study are submanifolds of $\mathbb{R}^3$, without relying on a particular choice of a parameterization. For example, let us consider two surfaces $M, \bar{M} \subset \mathbb{R}^3$ and a diffeomorphism $\phi\colon M \to \bar{M}$. We would like to have conditions which insure that ϕ is the restriction to M of some proper Euclidean motion. If we arbitrarily choose some immersion $f\colon U \to M \subset \mathbb{R}^3$, and let $\bar{f}\colon U \to \mathbb{R}^3$ be $\bar{f} = \phi \circ f$, then Theorem 2 tells us that such a proper Euclidean motion exists if the g_{ij} and l_{ij} for f equal the $\bar{g}_{ij}$ and $\bar{l}_{ij}$ for $\bar{f}$. Now the individual functions g_{ij} and l_{ij} for f do not have an "invariant meaning": given a submanifold $M \subset \mathbb{R}^3$, we cannot, for example, find functions γ_{ij} on M so that every $f\colon U \to M$ has its g_{ij}'s given simply by $g_{ij} = \gamma_{ij} \circ f$. Fortunately, however, the tensors

$$g_{11}\, ds \otimes ds + g_{12}\, ds \otimes dt + g_{21}\, dt \otimes ds + g_{22}\, dt \otimes dt$$
$$l_{11}\, ds \otimes ds + l_{12}\, ds \otimes dt + l_{21}\, dt \otimes ds + l_{22}\, dt \otimes dt$$

do have an invariant meaning: they are just $f^*\mathrm{I}$ and $f^*\mathrm{II}$. So we can formulate the first part of Theorem 3 for submanifolds:

4. COROLLARY. Let $M, \bar{M} \subset \mathbb{R}^3$ be two connected oriented surfaces imbedded in $\mathbb{R}^3$, let $\nu\colon M \to S^2 \subset \mathbb{R}^3$ and $\bar{\nu}\colon \bar{M} \to S^2 \subset \mathbb{R}^3$ be the unit normal vector fields determined by the orientations, and let I, II and $\bar{\mathrm{I}}$, $\bar{\mathrm{II}}$ be the first and second fundamental forms for M and $\bar{M}$ (the forms II and $\bar{\mathrm{II}}$ being defined with respect to ν and $\bar{\nu}$, respectively). Let $\phi\colon M \to \bar{M}$ be an orientation preserving diffeomorphism which preserves the first and second fundamental forms,

$$\phi^*\bar{\mathrm{I}} = \mathrm{I} \qquad \text{(i.e., } \phi \text{ is an isometry)}$$
$$\phi^*\bar{\mathrm{II}} = \mathrm{II}.$$

Then there is a proper Euclidean motion A such that $\phi = A|M$ and $A_*\nu = \bar{\nu}$.

PROOF. Let $f\colon U \to M \subset \mathbb{R}^3$ be an orientation preserving immersion, and let $\bar{f} = \phi \circ f\colon U \to \bar{M} \subset \mathbb{R}^3$; the immersion $\bar{f}$ is also orientation preserving, since ϕ is. The $\bar{g}_{ij}$ for $\bar{f}$ are the coefficients, with respect to the standard coordinate system (s, t), of

$$\bar{f}^*\bar{\mathrm{I}} = (\phi \circ f)^*\bar{\mathrm{I}} = f^*\phi^*\bar{\mathrm{I}} = f^*\mathrm{I}.$$

Consequently, $\bar{g}_{ij} = g_{ij}$. Similarly, since $\bar{f}$ is orientation preserving, the $\bar{l}_{ij}$ are the coefficients of $\bar{f}^*\bar{\mathrm{II}} = f^*\mathrm{II}$; since f is also orientation preserving, we find that $\bar{l}_{ij} = l_{ij}$. By Theorem 3, there is some proper Euclidean motion A such that $\phi = A$ on $f(U)$. If we choose immersions $\{f_\alpha\colon U_\alpha \to M\}$ whose images cover M, it is easy to see that the corresponding A_α must all be the same proper Euclidean motion A. ❖

We also want to formulate the existence part of Theorem 3 for manifolds, rather than immersions. So we consider an oriented surface M with a Riemannian metric $\langle\!\langle\ ,\ \rangle\!\rangle$ [corresponding to the g_{ij}] and a symmetric tensor S covariant of order 2 [corresponding to the l_{ij}]. In the previous chapter we have already seen how to give an invariant version of Gauss' equation and the Codazzi-Mainardi equations. This allows us to state

5. COROLLARY. Let $(M, \langle\!\langle\ ,\ \rangle\!\rangle)$ be an oriented Riemannian 2-manifold, with covariant derivative ∇ and curvature tensor R, and let S be a symmetric tensor on M, covariant of order 2. Suppose that S satisfies

(1) Gauss' Equation:

$$\langle\!\langle R(X,Y)Y,X\rangle\!\rangle = S(X,X)S(Y,Y) - [S(X,Y)]^2$$

(2) The Codazzi-Mainardi Equations:

$$(\nabla_X S)(Y,Z) = (\nabla_Y S)(X,Z).$$

Then for any $p \in M$ there is a neighborhood U of p and an immersion $f\colon U \to \mathbb{R}^3$ such that

$$\langle\!\langle\ ,\ \rangle\!\rangle = f^*\langle\ ,\ \rangle$$
$$S = f^*\mathrm{II},$$

where $\langle\ ,\ \rangle$ is the usual Riemannian metric on $\mathbb{R}^3$ and II is the second fundamental form on $f(U)$ defined in terms of the unit normal field ν which is determined by the orientation that $f(U)$ gets from the orientation on $U \subset M$.

PROOF. Left to the reader. ❖

Unlike Corollary 4, where a global result comes almost automatically, in Corollary 5 we cannot generally choose U to be all of M. As an example, we take the torus $S^1 \times S^1$ with a flat metric $\langle\!\langle\ ,\ \rangle\!\rangle$ [pg. II.179] and let $S = 0$. The Gauss and Codazzi-Mainardi equations are trivially satisfied. But the only connected submanifolds of $\mathbb{R}^3$ with $\mathrm{II} = 0$ everywhere are subsets of a plane (Theorem 2), so we certainly cannot find an immersion $f\colon S^1 \times S^1 \to \mathbb{R}^3$ with $f^*\mathrm{II} = S = 0$ everywhere. On the other hand (Problem 3), we *can* take U to be all of M in Corollary 5 when M is simply-connected.

Now that we have adequately documented the importance of the second fundamental form in surface theory, we will take this opportunity to slip in something new. The reader has perhaps already surmised with subconscious dread that there is a third fundamental form, and even yet higher numbered monsters, but these bogey men turn out to be very nicely behaved creatures which are in no way to be feared. For a submanifold $M \subset \mathbb{R}^3$, with unit normal $\nu\colon M \to S^2 \subset \mathbb{R}^3$, we define the **third fundamental form** III of M by

$$\begin{aligned}\mathrm{III}(p)(v_p, w_p) &= \langle -d\nu(v_p), -d\nu(w_p)\rangle\\ &= \langle d\nu(v_p), d\nu(w_p)\rangle \qquad v_p, w_p \in M_p.\end{aligned}$$

Similarly, if $f\colon U \to \mathbb{R}^3$ is an immersion (for $U \subset \mathbb{R}^2$ open), we define III_f by

$$\begin{aligned}\mathrm{III}_f(s,t)(v,w) &= \langle dN_f(v), dN_f(w)\rangle\\ &= \langle (N_f)_*(v), (N_f)_*(w)\rangle \qquad v, w \in \mathbb{R}^2{}_{(s,t)}.\end{aligned}$$

This is equivalent to defining $\mathrm{III}_f = f^*\mathrm{III}$, where III is the third fundamental form for image f. Remembering that $d\nu\colon M_p \to M_p$ is self-adjoint (Theorem 1-8 or Theorem II.3-1), we see that

$$\mathrm{III}(p)(v_p, w_p) = \langle (d\nu)^2(v_p), w_p\rangle.$$

This suggests defining

$$\mathrm{IV}(p)(v_p, w_p) = \langle (d\nu)^3(v_p), w_p\rangle,$$

and so forth. There is no notational way to write down the general definition, since no one has ever addressed the burning question of how we should indicate the n^{th} Roman numeral (come to think of it, no one even knows how to write down arbitrarily large Roman numerals). But that doesn't matter very much, especially as all these forms are expressible in terms of I and II anyway:

6. PROPOSITION. For a surface $M \subset \mathbb{R}^3$ we have

$$\mathrm{III} - 2H\cdot\mathrm{II} + K\cdot\mathrm{I} = 0.$$

(Similarly, for an immersion $f\colon U \to \mathbb{R}^3$, we have

$$\mathrm{III}_f - 2H\cdot\mathrm{II}_f + K\cdot\mathrm{I}_f = 0.$$

where $H(s,t)$ is the mean curvature of image f at $f(s,t)$, etc.)

PROOF. Remember the Cayley-Hamilton Theorem! The map $-d\nu\colon M_p \to M_p$ satisfies its characteristic polynomial $\chi(\lambda)$, which is given by

$$\chi(\lambda) = \lambda^2 - [\mathrm{trace}(-d\nu)]\lambda + \det(-d\nu) = \lambda^2 - 2H\lambda + K.$$

Consequently,

$$(-d\nu)^2 - 2H(-d\nu) + K \cdot \text{identity} = 0 \quad \text{on } M_p.$$

Applying this equation to v_p, and taking the inner product with w_p, we obtain the desired result. ❖

It is clear that we can also express IV in terms of III and II, etc. Proposition 6 does not necessarily mean that III is not worth considering, for it is still a useful tool for expressing certain quantities. Suppose, for example, that we have an immersion $f: U \to \mathbb{R}^3$, with normal map N_f ($= \nu \circ f$ for the unit normal map ν on image f). Since the normal map N_f plays such a vital role in describing the geometry of f, it is not at all unreasonable to ask what the first and second fundamental forms of N_f look like. Notice that in this instance we certainly want to explicitly consider the forms I_{N_f} and II_{N_f} for the *map* N_f: the *image* of N_f is just part of S^2, so *its* first and second fundamental forms aren't very interesting.

7. PROPOSITION. Let $f: U \to \mathbb{R}^3$ be an immersion with normal map $N_f = \nu \circ f$. Then the third fundamental form of f is both the first fundamental form of N_f and the negative of the second fundamental form of N_f:

$$\mathrm{III}_f = \mathrm{I}_{N_f} = -\mathrm{II}_{N_f}.$$

PROOF. Our original definition,

$$\mathrm{III}_f(s,t)(v,w) = \langle (N_f)_*(v), (N_f)_*(w)\rangle,$$

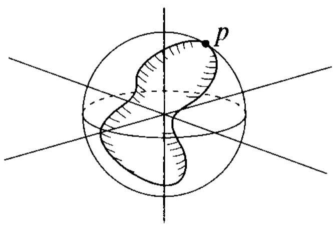
shows that $\mathrm{III}_f = \mathrm{I}_{N_f}$. Since the unit normal vector at any point $p \in S^2$ is just p itself, it is also clear that the normal map of N_f is just N_f itself, so we have $N_{N_f} = N_f$. Thus

$$\begin{aligned}\mathrm{II}_{N_f}(s,t)(v,w) &= \langle -(N_{N_f})_*(v), (N_f)_*(w)\rangle \\ &= \langle -(N_f)_*(v), (N_f)_*(w)\rangle. \text{ ❖}\end{aligned}$$

This result will come in handy at one point in Chapter 9, but we have no more to say about III at present. Following the route set forth in the first chapter of Volume II, we will now briefly look at some global properties of surfaces which are related to positive curvature. At the very outset we note one respect in which the situation for surfaces in $\mathbb{R}^3$ is different from that of curves in $\mathbb{R}^2$: Although we are able to define a signed curvature κ for a curve c in $\mathbb{R}^2$, the sign of κ depends on the "orientation" of c, and is reversed when we traverse c in the opposite direction; but for a surface $M \subset \mathbb{R}^3$, the Gaussian curvature K, which may also be positive or negative, does not depend on the orientation of M. We begin with a simple, but sometimes useful, observation.

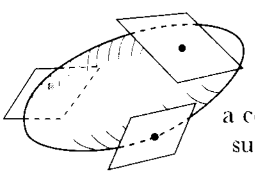
8. PROPOSITION. If M is a compact surface immersed in $\mathbb{R}^3$, then there is at least one point $p \in M$ where $K(p) > 0$.

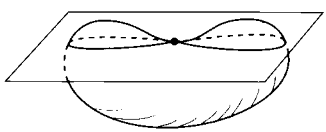
PROOF. The trick is to choose a point $p \in M$ whose distance from 0 is a maximum. Then M is even more curved at p than the sphere of radius $|p|$—in fact, each principal curvature is $> 1/|p|$, by Proposition II.3-0, and the corresponding result for curves in $\mathbb{R}^2$. Details are left as an exercise. ❖



This result gives us another way of seeing that the flat torus cannot be immersed in $\mathbb{R}^3$ (no matter what tensor S we choose on it).

We now want to consider surfaces M with $K(p) > 0$ for *all* $p \in M$. We naturally hope to relate this condition to convexity, so a brief discussion of that concept is in order. We define an imbedded surface $M \subset \mathbb{R}^3$ to be **convex** if it lies on one side of each of its tangent planes. As in the case of curves, we would first like to relate this definition to the more common one.

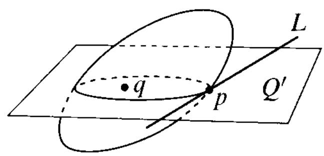

a convex surface


a non-convex surface

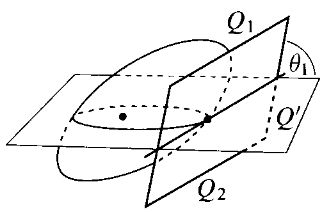
Any subset A of $\mathbb{R}^3$ is called **convex** if the line segment $\overline{pq}$ from p to q is contained in A whenever $p, q \in A$. Suppose A is convex and p is a point in the boundary of A. A plane P containing p is called a **support plane** of A if A lies completely in one of the closed half-spaces into which P divides $\mathbb{R}^3$.

9. PROPOSITION. If A is convex, and p is in the boundary of A, then there is at least one support plane P containing p.

PROOF. If A has no interior points, it lies in a plane, and the Theorem follows easily from the corresponding result, Proposition II.1-3, for subsets of $\mathbb{R}^2$. If A has an interior point q, let Q be a plane containing q and p, and let L be a support line for $A \cap Q$ through p. This line L divides Q into two closed half-planes; let Q' be the one such that $Q' - L$ contains no points of $Q \cap A$.



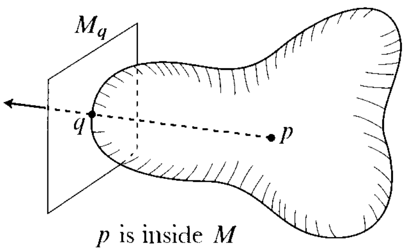
Now we will consider the various closed half-planes having L as their edge. Choose one side of Q' and consider angles θ such that the half-plane with L as its edge which makes an angle of θ with Q' on this side does not intersect A except along L (it may be that $\theta = 0$ is the only possibility). Let θ_1 be the least upper bound of all such θ, and let Q_1 be the half-plane with L as edge which makes an angle of θ_1. Let Q_2 be the corresponding half-plane on the other side of Q'.



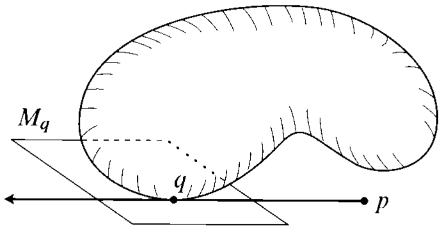
To prove the theorem, it clearly suffices to prove that the angle between Q_1 and Q_2 is $\geq \pi$. We note that there are points of A on planes arbitrarily close to Q_1 and Q_2. If the angle between Q_1 and Q_2 were $< \pi$, then we could consider a suitable pair of such points, together with points of A in a whole neighborhood of q, and find that A must contain some point of L in its interior, which is impossible. ❖

We now want to show that a compact connected 2-dimensional submanifold M of $\mathbb{R}^3$ is convex if and only if the set A consisting of all points on M or inside M is a convex subset of $\mathbb{R}^3$. Once again, we are assuming Corollary I.11-15, and the following easy consequence:

> Suppose $M \subset \mathbb{R}^3$ is a compact connected surface, and l is a ray from p which intersects M at just one point $q \neq p$. Suppose, moreover, that l does *not* lie along the tangent plane of M at q. Then p is inside M.



p is inside M



p is not inside M

10. PROPOSITION. Let $M \subset \mathbb{R}^3$ be a compact connected surface, and let A be the set of all points on or inside M. Then M is convex (that is, M lies on one side of each of its tangent planes) if and only if A is convex.

PROOF. Exactly like the proof of Proposition II.1-4. ❖

We are finally ready to relate convexity and curvature. If you have found yourself nodding drowsily at the rather obvious generalizations of old material which occupied the last few pages, it is time to wake up now, because our result for surfaces is *not* just an analogue of the result for curves.

11. THEOREM (HADAMARD). (1) If M is a convex surface in $\mathbb{R}^3$, then $K(p) \geq 0$ for all $p \in M$.

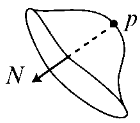
(2) Let M be a compact connected 2-manifold, and $f\colon M \to \mathbb{R}^3$ an immersion with $K(p) > 0$ for all $p \in M$. Then

(i) The manifold M is orientable, and the normal map $N\colon M \to S^2 \subset \mathbb{R}^3$ is a diffeomorphism,

(ii) The map $f\colon M \to \mathbb{R}^3$ is an imbedding, and $f(M)$ is convex.

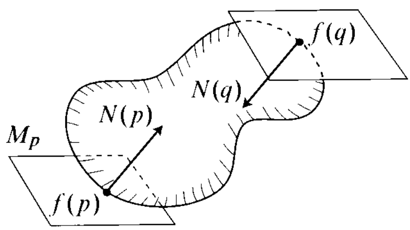
PROOF. The first part is immediate, for we have already seen that if $K(p) < 0$, then M lies on both sides of M_p.

To prove (2), we first recall that since $K(p) > 0$, points of M near p all lie on one side of M_p. We choose N so that it always points on this side. Since this

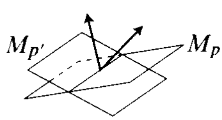


gives us a continuous choice of N, it also gives us an orientation of M. To show that $N: M \to S^2$ is a diffeomorphism, we note first that N_* is always one-one, since $K = \det N_*$. So $N(M) \subset S^2$ is open. It is also closed, since M is compact; so $N(M) = S^2$. Now we need to use a few properties of covering spaces. The fact that $N: M \to S^2$ is onto and locally one-one does not immediately imply that N is a covering space map; however it is an easy exercise (Problem 4) to show that this follows from the fact that M is compact. But S^2 is simply-connected, and therefore has no non-trivial covering spaces. So $N: M \to S^2$ is a diffeomorphism, and we have proved (i).

To prove (ii), we consider a point $p \in M$ and the tangent space $M_p \subset \mathbb{R}^3{}_{f(p)}$. At least some points of $f(M)$ lie on the same side of M_p as $N(p)$. Let $f(q)$ be



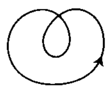
a point on this side which is furthest from M_p. Then clearly $N(q) = -N(p)$. Suppose that $f(p) = f(p')$ for some other $p' \in M$. Since $N: M \to \mathbb{R}^3$ is one-one, $N(p')$ cannot be either $N(p)$ or $-N(p) = N(q)$. So the tangent plane $M_{p'}$ must cross the tangent plane M_p. It is then easy to see that $f(M)$ must



contain points on both sides of M_p. Let $f(r)$ be a point of M furthest from M_p on the *other* side from q. Then $N(r)$ must equal $N(p)$, contradicting the fact that N is one-one. Thus we have shown that f is an embedding.

To show that $f(M)$ is convex, we use a similar argument. Given $p \in M$, we just have to show that all of $f(M)$ lies on the same side of $M_p \subset \mathbb{R}^3{}_{f(p)}$ as $N(p)$ does. If there were points on the opposite side, and $f(q)$ were a point on the opposite side which is furthest from M_p, then $N(q)$ would have to equal $N(p)$, again a contradiction. ❖

This result naturally invites comparison with Theorem II.1-8, which states that a simple closed curve c in $\mathbb{R}^2$ is convex if and only if it satisfies $\kappa \geq 0$ or $\kappa \leq 0$ (depending on the direction in which c is traversed). The proof of part (1) of Theorem 11 is much simpler than the proof of the corresponding part of Theorem II.1-8. This is because the sign of the Gaussian curvature K has a local geometric meaning, while the sign of κ has none; the only meaningful assertion about κ is the *global* statement that it is ≥ 0 or ≤ 0 everywhere. In part (2) of Theorem 11 we have the significant circumstance that we do not have to assume that M is imbedded—this comes out as part of the conclusion. The analogous assertion is *false* in the case of curves: the figure below shows an immersed, but not imbedded, curve with $\kappa \geq 0$ everywhere. In one respect



our result does not improve on Theorem II.1-8, for in order to prove part (2), we needed to assume the strict inequality $K(p) > 0$. Actually the result holds even when we assume only that $K \geq 0$, but the proof in this case is much more difficult (for further discussion, and references, see pp. IV. 82–83).

Our approach to surface theory has so far been very classical, but we are now ready to jazz it up a bit. First we want to examine the moving frame approach again, and write out explicitly all the equations (which in the case of surfaces in $\mathbb{R}^3$ boil down to almost nothing). In addition to their importance in the remainder of this chapter, some of these formulas will be crucial in Chapter 6 (and the equations in the general case will be even more crucial in Chapter 7).

If X_1, X_2 is a positively oriented orthonormal moving frame on an oriented surface $M \subset \mathbb{R}^3$, and we let $X_3 = \nu$, then X_1, X_2, X_3 is a positively oriented adapted orthonormal moving frame on M. There are just the following forms to consider:

θ^1, θ^2	the dual 1-forms
$\omega_1^2 = -\omega_2^1$	the connection form
ψ_1^3, ψ_2^3.	

We want to relate these forms to the tensors and functions on M already considered. Notice first that I is given by

$$\mathrm{I} = \theta^1 \otimes \theta^1 + \theta^2 \otimes \theta^2.$$

We also have

$$dA = \theta^1 \wedge \theta^2,$$

where dA is the volume element determined by the metric I on M and the given orientation. Since we have (see page 19ff.)

$$\psi_1^3 = s_{11}^3\theta^1 + s_{21}^3\theta^2 = \mathrm{II}(X_1, X_1)\theta^1 + \mathrm{II}(X_2, X_1)\theta^2$$
$$\psi_2^3 = s_{12}^3\theta^1 + s_{22}^3\theta^2 = \mathrm{II}(X_1, X_2)\theta^1 + \mathrm{II}(X_2, X_2)\theta^2,$$

we can write

$$\mathrm{II} = \psi_1^3 \otimes \theta^1 + \psi_2^3 \otimes \theta^2.$$

It is also easy to see that the Gaussian and mean curvatures

$$K = \mathrm{II}(X_1, X_1) \cdot \mathrm{II}(X_2, X_2) - [\mathrm{II}(X_1, X_2)]^2$$
$$H = \tfrac{1}{2}\{\mathrm{II}(X_1, X_1) + \mathrm{II}(X_2, X_2)\}$$

are given by

$$\psi_1^3 \wedge \psi_2^3 = K\theta^1 \wedge \theta^2$$
$$\psi_1^3 \wedge \theta^2 - \psi_2^3 \wedge \theta^1 = 2H\theta^1 \wedge \theta^2.$$

On the other hand, we have a much more important expression for K, in terms of the connection form ω_1^2. We note first that equation (3) on page 16 now reduces to

$$d\omega_1^2 = \Omega_1^2.$$

Then Gauss' equation (page 20) becomes simply

$$0 = d\omega_1^2 - \psi_2^3 \wedge \psi_1^3 = d\omega_1^2 + K\theta^1 \wedge \theta^2,$$

so that

$$d\omega_1^2 = -K\theta^1 \wedge \theta^2.$$

Since this equation is equivalent to Gauss' equation, it must somehow demonstrate the Theorema Egregium, and it surely does, since the form ω_1^2 does not depend on the imbedding (it is the unique form with $d\theta^1 = \omega_1^2 \wedge \theta^2$ and $d\theta^2 = -\omega_1^2 \wedge \theta^1$). Some elementary treatments of surface theory proceed to use this equation to *define* the Gaussian curvature of an arbitrary 2-dimensional Riemannian manifold—it is only necessary to check that K does not depend on the choice of the orthonormal moving frame; this is a special case of the moving frame definition of the curvature tensor given in Chapter II.7. Finally, we mention that the Codazzi-Mainardi equations (page 20) now become

$$d\psi_1^3 = \omega_1^2 \wedge \psi_2^3$$
$$d\psi_2^3 = -\omega_1^2 \wedge \psi_1^3.$$

An introduction to surface theory carried out purely in terms of this moving frame and structural equation approach can be very frustrating. Instead of dealing with geometrically tangible things like dN and II, one has only the 1-forms ω_1^2, ψ_1^3, ψ_2^3 to play with, and the simplicity of the Gauss and Codazzi-Mainardi equations as given above seems vitiated by their lack of intuitive geometric content. But this simplicity is a great advantage in proving theorems, and can be attributed, in large measure, to the fact that they express integrability conditions so neatly in terms of d. For example, they allow us to give a proof of the fundamental theorem of surface theory which uses the differential form version of the Frobenius integrability theorem (Proposition I.7-14), instead of mucking around with the classical integrability conditions; we will present this proof, in a more general situation, in Chapter 7. The truly overwhelming advantage of the moving frame approach becomes apparent when one is seriously investigating questions about the shape of surfaces in space; any information one can hope to get has to come out of the three simple equations

$$d\omega_1^2 = -\psi_1^3 \wedge \psi_2^3, \quad d\psi_1^3 = \omega_1^2 \wedge \psi_2^3, \quad d\psi_2^3 = -\omega_1^2 \wedge \psi_1^3.$$

Usually one just picks a frame suited to the problem and reads off the information from these equations. As a very simple example, we consider an all-umbilic surface $M \subset \mathbb{R}^3$. In this situation any adapted orthonormal moving frame X_1, X_2, X_3 on M is suitable (the hypothesis that p is an umbilic essentially says that all orthonormal frames at p are indistinguishable from one another), and we have

$$\psi_i^3 = \lambda\theta^i \qquad i = 1, 2$$

for some function λ on M. Thus

$$d\lambda \wedge \theta^1 + \lambda d\theta^1 = d\psi_1^3 = \omega_1^2 \wedge \psi_2^3 = \lambda\omega_1^2 \wedge \theta^2$$
$$d\lambda \wedge \theta^2 + \lambda d\theta^2 = -\lambda\omega_1^2 \wedge \theta^1,$$

while

$$d\theta^1 = \omega_1^2 \wedge \theta^2$$
$$d\theta^2 = -\omega_1^2 \wedge \theta^1.$$

So we find that

$$d\lambda \wedge \theta^i = 0 \qquad i = 1, 2.$$

But this implies that the 1-form $d\lambda$ is 0, so we find, once again, that λ is constant. More interesting examples will occur in later chapters.

The moving frame approach was brought in at this point not to launch an extended investigation into the geometry of surfaces—that occurs in later chapters—but with a completely different goal in mind. We want to show that the Gauss and Codazzi-Mainardi equations for surfaces in $\mathbb{R}^3$ are, from the proper point of view, nothing more than the "equations of structure" of the Lie group SO(3), and that the Fundamental Theorem of Surface Theory reduces to Theorems I.10-17 and I.10-18 about Lie groups. After doing this, we will then proceed to bring another group into the picture by examining properties of surfaces in $\mathbb{R}^3$ which are invariant under the group of maps $A\colon \mathbb{R}^3 \to \mathbb{R}^3$ of the form $A = T \circ B$ for T a translation and $B \in \mathrm{SL}(3)$ = group of 3×3 matrices with det = 1.

We begin with some preliminaries about notation. Nowadays, an "affine" map $A\colon \mathbb{R}^n \to \mathbb{R}^n$ is usually defined to be one of the form $A = T \circ B$ for T a translation and $B \in \mathrm{GL}(n, \mathbb{R})$. Thus the proper Euclidean motions, $A = T \circ B$ for $B \in \mathrm{SO}(n)$, might be described as "special orthogonal affine" maps, while maps $A = T \circ B$ for $B \in \mathrm{SL}(n)$ might best be described as "special linear affine" maps. We will employ this terminology regularly for maps $A\colon \mathbb{R}^n \to \mathbb{R}^n$, but we will also find it convenient to abbreviate the phrase "special linear affine" to "special affine" when we speak of such concepts as "special affine curvature". Thus when we speak of the "special affine geometry of surfaces" in $\mathbb{R}^3$, we mean properties of surfaces invariant under special linear affine maps. On those occasions when we want to consider properties of surfaces invariant under all affine maps, we will emphasize this fact by speaking of "general affine" invariants.

We will also include a brief review of the relevant facts about Lie groups (pp. II.36–37), since we are going to make a slight change of notation. If G is a Lie group with Lie algebra $\mathfrak{g}$, then we define the natural $\mathfrak{g}$-valued 1-form* $\boldsymbol{\omega}$ on G by $\boldsymbol{\omega}(a)(\widetilde{X}(a)) = X$, where $\widetilde{X}$ is the left invariant vector field with $\widetilde{X}(e) = X$. If $X_1, \ldots, X_n$ is a basis of $\mathfrak{g}$, then we have $\boldsymbol{\omega} = \sum_{i=1}^n \boldsymbol{\omega}^i X_i$ for ordinary ($\mathbb{R}$-valued) left invariant 1-forms $\boldsymbol{\omega}^i$, and these forms $\boldsymbol{\omega}^i$ are a basis for the left invariant 1-forms. These forms are important because two maps

*We are now using $\boldsymbol{\omega}$ to distinguish this form on G from the forms ω_1^2 for a moving frame on a surface. This wasn't important in Volume II, where we considered only curves.

$f, g: M \to G$ differ by a left translation [$f = L_a \circ g$ for some $a \in G$] if and only if $f^*(\omega^i) = g^*(\omega^i)$ for all i (Theorem I.10-18). When G is a subgroup of $\mathrm{GL}(n, \mathbb{R})$, with $P: G \to \mathrm{GL}(n, \mathbb{R}) \subset \mathbb{R}^{n^2}$ the inclusion map, then we can form $P^{-1} \cdot dP$, where P^{-1} denotes (somewhat confusingly) the map $A \mapsto A^{-1}$, and the differential dP of P can be considered either as an $\mathbb{R}^{n^2}$-valued 1-form on G, or as the matrix of 1-forms $dP = (dx^{ij})$, where dx^{ij} denotes the differential of $x^{ij}|G$. Then (pp. II.36–37) $P^{-1} \cdot dP$ is the natural $\mathfrak{g}$-valued 1-form $\boldsymbol{\omega}$ on G. Among the entries of this matrix will be a basis for the left invariant 1-forms on G (the entries are generally not linearly independent, since the forms dx^{ij} are not linearly independent on G). So if $f: M \to \mathrm{GL}(n, \mathbb{R})$ is a C^∞ map, and we want to look at the forms $f^*(\omega^i)$ for a basis $\{\omega^i\}$ of left invariant 1-forms on G, it suffices to look at the entries of the matrix

$$f^*(P^{-1} \cdot dP) = f^{-1} \cdot df.$$

To study properties of immersions $f: \mathbb{R}^2 \to \mathbb{R}^3$ which are invariant under special orthogonal affine maps of $\mathbb{R}^3$, we need to define an associated map $\alpha_f: \mathbb{R}^2 \to \mathrm{SO}(3)$. This can be done in the following way. Let $X_3 = N$ be the normal map of f, and let X_1, X_2 be the result of applying the Gram-Schmidt orthonormalization process to the vectors f_1, f_2. We then have an adapted orthonormal moving frame X_1, X_2, X_3 on $\mathbb{R}^2$, and if we also consider X_i as a column vector, then $\alpha_f = (X_1, X_2, X_3): \mathbb{R}^2 \to \mathrm{SO}(3)$ is the desired map. Notice that we can reconstruct f_1, f_2 from X_1, X_2 when we are also given the functions g_{ij}.

To find $\alpha_f{}^*(\omega^i)$, where the ω^i are a basis for the left invariant 1-forms on $\mathrm{SO}(3)$, we look at the entries of the matrix of 1-forms

$$a = \alpha_f{}^{-1} \cdot d\alpha_f, \quad \text{with} \quad d\alpha_f = \alpha_f \cdot a.$$

This equation means that

$$(dX_1, dX_2, dX_3) = (X_1, X_2, X_3) \begin{pmatrix} 0 & -a_{21} & -a_{31} \\ a_{21} & 0 & -a_{32} \\ a_{31} & a_{32} & 0 \end{pmatrix},$$

where the a_{ij} are 1-forms, and the X_i and dX_i are considered as column vectors; the latter equation stands for

$$dX_1 = a_{21} X_2 + a_{31} X_3, \quad \text{etc.}$$

Thus, if ∇' denotes the covariant differentiation in $\mathbb{R}^3$, we have

$$\nabla'_X X_1 = dX_1(X) = a_{21}(X) X_2 + a_{31}(X) X_3, \quad \text{etc.}$$

But

$$\nabla'_X X_1 = \omega_1^2(X) \cdot X_2 + \psi_1^3(X) \cdot X_3, \quad \text{etc.}$$

(where $\omega_1^2, \psi_1^3, \psi_2^3$ really denote f^* of the corresponding forms on image f). Thus we see that

> *The forms* ω_1^2, ψ_1^3, ψ_2^3 *are precisely* $\alpha_f{}^*(\boldsymbol{\omega}^i)$ *for* $\boldsymbol{\omega}^i$ *a basis of the left invariant 1-forms on* SO(3).

Theorem I.10-18 then tells us that for two immersions $f, \bar{f}\colon \mathbb{R}^2 \to \mathbb{R}^3$, the maps $\alpha_f, \alpha_{\bar{f}}\colon \mathbb{R}^2 \to \mathrm{SO}(3)$ differ by an element of SO(3) if and only if

$$\bar{\omega}_1^2 = \omega_1^2, \quad \bar{\psi}_1^3 = \psi_1^3, \quad \bar{\psi}_2^3 = \psi_2^3;$$

here the forms $\omega_1^2, \psi_1^3, \psi_2^3$ are formed for the moving frame $X_1, X_2, X_3 = N$, where X_1, X_2 are obtained by applying the Gram-Schmidt orthonormalization process to f_1, f_2, while the forms $\bar{\omega}_1^2, \bar{\psi}_1^3, \bar{\psi}_2^3$ are formed for the moving frame $\bar{X}_1, \bar{X}_2, \bar{X}_3 = \bar{N}$, where $\bar{X}_1, \bar{X}_2$ are obtained by applying the Gram-Schmidt orthonormalization process to $\bar{f}_1, \bar{f}_2$.

From this fact we can easily derive the first part of the Fundamental Theorem of Surface Theory. For if $\bar{g}_{ij} = g_{ij}$, then $\bar{X}_1, \bar{X}_2$ are the same linear combination of $\bar{f}_1, \bar{f}_2$ as X_1, X_2 are of f_1, f_2. Consequently, if we are also given that $\bar{l}_{ij} = l_{ij} \implies \bar{\mathrm{II}}(\bar{f}_i, \bar{f}_j) = \mathrm{II}(f_i, f_j)$, then we conclude that $\bar{\mathrm{II}}(\bar{X}_i, \bar{X}_j) = \mathrm{II}(X_i, X_j)$. The formulas on page 69 then show that $\bar{\psi}_1^3 = \psi_1^3$ and $\bar{\psi}_2^3 = \psi_2^3$, while the equation $\bar{\omega}_1^2 = \omega_1^2$ follows from $\bar{g}_{ij} = g_{ij}$. Thus $\alpha_f, \alpha_{\bar{f}}\colon \mathbb{R}^2 \to \mathrm{SO}(3)$ differ by an element of SO(3); this implies that (f_1, f_2, N), $(\bar{f}_1, \bar{f}_2, \bar{N})$ differ by an element of SO(3), and hence that $f, \bar{f}$ differ by a special orthogonal affine map of $\mathbb{R}^3$.

In the case of curves, the equations of structure of $\mathrm{SO}(n)$ or $\mathrm{SL}(n, \mathbb{R})$ could not give any interesting information, since there are no non-zero 2-forms on $\mathbb{R}$. But they do give information for surfaces. To figure out the equations of structure of SO(3), we proceed as follows. Since the Lie algebra

$$\mathfrak{o}(3) = \{\text{tangent space of SO(3) at } I\}$$

is just the set of skew-symmetric 3×3 matrices, the matrices

$$X_1 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad X_2 = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \qquad X_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$$

are a basis for $\mathfrak{o}(3)$. The bracket operation in $\mathfrak{o}(3)$ is (pp. I.378–379)

$$[M, N] = MN - NM.$$

In particular, we compute that

$$[X_1, X_2] = X_3, \qquad [X_1, X_3] = -X_2, \qquad [X_2, X_3] = X_1,$$

so that if we write

$$[X_i, X_j] = \sum_{k=1}^{3} C_{ij}^k X_k,$$

then the "constants of structure" C_{ij}^k are

$$\begin{array}{lll} C_{12}^1 = 0 & C_{12}^2 = 0 & C_{12}^3 = 1 \\ C_{13}^1 = 0 & C_{13}^2 = -1 & C_{13}^3 = 0 \\ C_{23}^1 = 1 & C_{23}^2 = 0 & C_{23}^3 = 0. \end{array}$$

Now let $\boldsymbol{\omega}^i$ be the left invariant 1-forms on $\mathfrak{o}(3)$ with $\boldsymbol{\omega}^1(I), \boldsymbol{\omega}^2(I), \boldsymbol{\omega}^3(I)$ dual to X_1, X_2, X_3. The equation on pg. I.396 (which is equivalent to the "equations of structure" on pg. I.404) then gives

$$\begin{aligned} d\boldsymbol{\omega}^1 &= -\boldsymbol{\omega}^2 \wedge \boldsymbol{\omega}^3 \\ d\boldsymbol{\omega}^2 &= \boldsymbol{\omega}^1 \wedge \boldsymbol{\omega}^3 \\ d\boldsymbol{\omega}^3 &= -\boldsymbol{\omega}^1 \wedge \boldsymbol{\omega}^2. \end{aligned}$$

Now we have seen that if $f\colon \mathbb{R}^2 \to \mathbb{R}^3$ is an immersion, then the forms ω_1^2, ψ_1^3, ψ_2^3 for f are given by

$$\begin{aligned} \omega_1^2 &= \alpha_f{}^*(\boldsymbol{\omega}^1) \\ \psi_1^3 &= \alpha_f{}^*(\boldsymbol{\omega}^2) \\ \psi_2^3 &= \alpha_f{}^*(\boldsymbol{\omega}^3). \end{aligned}$$

Therefore

$$\begin{aligned} d\omega_1^2 &= \alpha_f{}^*(d\boldsymbol{\omega}^1) = -\alpha_f{}^*(\boldsymbol{\omega}^2 \wedge \boldsymbol{\omega}^3) = -\psi_1^3 \wedge \psi_2^3 \\ d\psi_1^3 &= \alpha_f{}^*(d\boldsymbol{\omega}^2) = -\alpha_f{}^*(\boldsymbol{\omega}^1 \wedge \boldsymbol{\omega}^3) = \omega_1^2 \wedge \psi_2^3 \\ d\psi_2^3 &= \alpha_f{}^*(d\boldsymbol{\omega}^3) = -\alpha_f{}^*(\boldsymbol{\omega}^1 \wedge \boldsymbol{\omega}^2) = -\omega_1^2 \wedge \psi_1^3. \end{aligned}$$

As promised, these are precisely the Gauss Equation and Codazzi-Mainardi Equations, in the form given on pages 69–70. The reader can now easily see that the second part of the Fundamental Theorem of Surface Theory follows immediately from Theorem I.10-17.

For the remainder of this chapter* we will be considering the special affine theory of surfaces $M \subset \mathbb{R}^3$. If we try to follow the approach used for ordinary surface theory, then instead of working with an adapted moving frame X_1, X_2, ν on M which is orthonormal, we want to work with an adapted moving frame X_1, X_2, X_3 on M with $\det(X_1, X_2, X_3) = 1$. If $f\colon U \to M$ is an immersion (for $U \subset \mathbb{R}^2$ open), then you might think that we should use the moving frame

$$(af_1, af_2, aN), \qquad \text{where} \qquad a = \frac{1}{\sqrt[3]{\det(f_1, f_2, N)}}.$$

But this moving frame has no significance for special affine geometry, for it is not a "special affine invariant": if $A\colon \mathbb{R}^3 \to \mathbb{R}^3$ is special linear affine, then the normal $N_{A\circ f}$ for $A \circ f$ is not necessarily the image $A_*(N_f)$ of the normal N_f for f. As a matter of fact, not only the length, but even the direction of $A_*(N_f)$ will be wrong; the whole concept of "orthogonality" has no meaning in special affine geometry. Our first problem, therefore, is to pick out a "special affine normal" for M which is a special affine invariant. This is going to take quite a bit of doing.

In ordinary surface theory, the normal $\nu(p)$ is defined in terms of the tangent plane M_p of M, which is the first order surface which approximates M up to order 1 at p. Since special affine geometry always seems to involve higher order approximations to our given geometric object, we might expect to find a reasonable candidate for the special affine normal by looking at the second order approximation to our surface. As before, let us assume that $p = 0 \in \mathbb{R}^3$ and that the tangent plane at p is the (x, y)-plane, so that M is the graph of a function $h\colon \mathbb{R}^2 \to \mathbb{R}$ with $h(0,0) = h_1(0,0) = h_2(0,0) = 0$. The quadratic surface approximating M up to order 2 at 0 is

$$P = \left\{\left(s, t, \tfrac{1}{2}\left(h_{11}(0,0)\cdot s^2 + 2h_{12}(0,0)\cdot st + h_{22}(0,0)\cdot t^2\right)\right)\right\}.$$

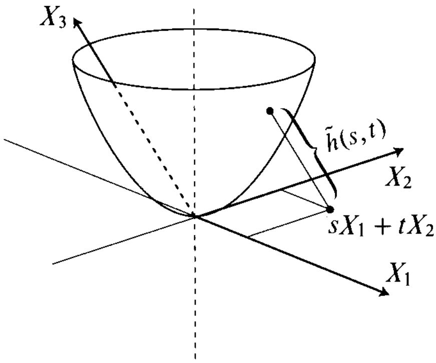
We have already seen that this surface does not depend on the particular choice of a basis in $\mathbb{R}^2$: if $\mathbf{X} = X_1, X_2$ is any basis of $\mathbb{R}^2$ and $h^{\mathbf{X}}(s,t)$ is the third coordinate of the point of M lying above $sX_1 + tX_2$, then the surface

$$Q = \left\{\left(sX_1 + tX_2, \tfrac{1}{2}\left({h^{\mathbf{X}}}_{11}(0,0)\cdot s^2 + 2{h^{\mathbf{X}}}_{12}(0,0)\cdot st + {h^{\mathbf{X}}}_{22}(0,0)\cdot t^2\right)\right)\right\}$$

is exactly the same as P. We also noted that we still have $Q = P$ when we change the direction of the z-axis; it is just as easy to see that $Q = P$ even when we change the *unit* on the z-axis, so that we are describing M in terms of the "$X_1, X_2, (0, 0, c)$ coordinate system".

*This material will not be needed later, except in Problem 4-16. Problems for this chapter are on page 134.

In all this arbitrariness, however, one essential prejudice of ordinary geometry remains: we have always picked the third axis perpendicular to the plane of the first two. Suppose now that we choose X_1, X_2 in $\mathbb{R}^2$ and a linearly independent vector X_3 which does not necessarily point along the z-axis. We can



still describe M as a graph in terms of the "(X_1, X_2, X_3) coordinate system": we let $\tilde{h}(s,t)$ be the X_3 component of the point of M with $s = X_1$ component and $t = X_2$ component. Now we look at the surface

$$Q = \left\{sX_1 + tX_2 + \tfrac{1}{2}\left(\tilde{h}_{11}(0,0)\cdot s^2 + 2\tilde{h}_{12}(0,0)\cdot st + \tilde{h}_{22}(0,0)\cdot t^2\right)X_3\right\}.$$

This surface is *not* the same surface as P. In fact, consider the case where $X_3 = (0,0,1) + \lambda X_1 + \mu X_2$ for certain numbers λ, μ. To say that M is the graph of h in the $X_1, X_2, (0,0,1)$ system means that

$$\text{(1)} \qquad M = \{sX_1 + tX_2 + h(s,t)\cdot(0,0,1)\} \qquad [h_\alpha(0,0) = 0];$$

similarly, if M is the graph of $\tilde{h}$ in the X_1, X_2, X_3 system, then

$$\begin{aligned} \text{(2)} \qquad M &= \{sX_1 + tX_2 + \tilde{h}(s,t)X_3\} \\ &= \left\{[s + \lambda\tilde{h}(s,t)]X_1 + [t + \mu\tilde{h}(s,t)]X_2 + \tilde{h}(s,t)\cdot(0,0,1)\right\} \\ &\qquad\qquad [\tilde{h}_\alpha(0,0) = 0]. \end{aligned}$$

Comparing (1) and (2) we find that

$$\tilde{h}(s,t) = h(s + \lambda\tilde{h}(s,t),\ t + \mu\tilde{h}(s,t)).$$

From this we easily compute that $\tilde{h}_{\alpha\beta}(0,0) = h_{\alpha\beta}(0,0)$, so that the approximating *functions of* s *and* t

$$\tilde{h}_{11}(0,0)s^2 + 2\tilde{h}_{12}(0,0)st + \tilde{h}_{22}(0,0)t^2$$
$$h_{11}(0,0)s^2 + 2h_{12}(0,0)st + h_{22}(0,0)t^2$$

are the same; this means that the approximating *surfaces*

$$\begin{aligned} P &= \{(sX_1 + tX_2, \tfrac{1}{2}(h_{11}(0,0)s^2 + 2h_{12}(0,0)st + h_{22}(0,0)t^2))\} \\ Q &= \{sX_1 + tX_2 + \tfrac{1}{2}(\tilde{h}_{11}(0,0)s^2 + 2\tilde{h}_{12}(0,0)st + \tilde{h}_{22}(0,0)t^2)X_3\} \\ &= \{sX_1 + tX_2 + \tfrac{1}{2}(h_{11}(0,0)s^2 + 2h_{12}(0,0)st + h_{22}(0,0)t^2)X_3\} \end{aligned}$$

are definitely different. In fact, we clearly have

$$Q = A(P),$$

where $A\colon \mathbb{R}^3 \to \mathbb{R}^3$ is the affine map which keeps M_p fixed and takes $(0,0,1)$ to X_3. All we can say is that Q does not change when we multiply X_3 by a constant, just as P does not change when we multiply $(0,0,1)$ by a constant; we can merely speak of the osculating paraboloid corresponding to any given line through p which does not lie in M_p.

Thus we see that we do not get a special affine invariant osculating paraboloid simply by looking at M up to order 2. There are some things that we do get, however. Consider first a fixed basis X_1, X_2 for $M_p = (x,y)$-plane. We have just seen that the matrices

$$S = \begin{pmatrix} h_{11}(0,0) & h_{12}(0,0) \\ h_{21}(0,0) & h_{22}(0,0) \end{pmatrix} \qquad \text{and} \qquad S' = \begin{pmatrix} \tilde{h}_{11}(0,0) & \tilde{h}_{12}(0,0) \\ \tilde{h}_{21}(0,0) & \tilde{h}_{22}(0,0) \end{pmatrix},$$

defined in terms of the third axes $(0,0,1)$ and $X_3 = (0,0,1) + \lambda X_1 + \mu X_2$, respectively, are exactly the same; if we had picked X_3 to be the most general possible choice, $X_3 = (0,0,c) + \lambda X_1 + \mu X_2$, then S' would clearly be

$$S' = \frac{1}{c} \cdot S.$$

On the other hand, suppose we consider the coordinate system

$$a_{11}X_1 + a_{21}X_2, \quad a_{12}X_1 + a_{22}X_2, \quad (0,0,1).$$

Equation $(*)$ on page 37 shows that the matrix S' in this case is related to the matrix S by

$$S' = A^{\mathbf{t}}SA, \qquad A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

where $^{\mathbf{t}}$ denotes the transpose. In general, if we are given any two bases (X_1, X_2, X_3) and (X'_1, X'_2, X'_3) of $\mathbb{R}^3{}_p$, with X_1, X_2 and X'_1, X'_2 bases for M_p, and

$$X'_3 = cX_3 + \lambda X_1 + \mu X_2,$$

then the matrices S and S' are related by

(1) $$S' = \frac{1}{c} \cdot A^{\mathrm{t}} S A,$$

where $A = (a_{ij})$ is given by

(2) $$X'_i = \sum_{j=1}^{2} a_{ji} X_j .$$

From equation (1) we see that

$$\det S' \gtreqless 0 \iff \det S \gtreqless 0 \qquad \text{and} \qquad S' = 0 \iff S = 0.$$

Thus we can determine whether p is an elliptic, hyperbolic, parabolic, or planar, point of M by means of an arbitrary basis (X_1, X_2, X_3) of $\mathbb{R}^3{}_p$ with $X_1, X_2 \in M_p$. This clearly implies that the "type" of a point (elliptic, hyperbolic, parabolic, or planar) is a "general affine invariant": If $A\colon \mathbb{R}^3 \to \mathbb{R}^3$ is any affine map and $M \subset \mathbb{R}^3$ is a surface, then $A(p)$ is the same type of point on $A(M)$ as p is on M.

Consider, for the time being, an elliptic point $p \in M$. As we observed in the proof of Theorem 11, there is a natural orientation for M_p, the one that makes an ordered basis (X_1, X_2) of M_p positively oriented whenever (X_1, X_2, X_3) is positively oriented in $\mathbb{R}^3{}_p$ for any $X_3 \in \mathbb{R}^3{}_p$ which points "inward" (that is, in the direction of the osculating paraboloid). Now any basis $X_1, X_2, X_3 \in \mathbb{R}^3{}_p$ with $X_1, X_2 \in M_p$ determines a matrix $S = (s_{ij})$, and we can use S to define an inner product on M_p by

$$\langle X_i, X_j \rangle = s_{ij};$$

this inner product is positive definite if and only if X_3 is inward pointing. If X_1, X_2, X_3 happens to be orthonormal, then the inner product is just the second fundamental form II of ordinary surface theory. Now consider another basis X'_1, X'_2, X'_3, with X'_3 inward pointing. This determines a matrix S', and hence another positive definite inner product $\langle\ ,\ \rangle'$ on M_p by

$$\langle X'_i, X'_j \rangle' = s'_{ij}.$$

Equation (1) shows that the inner products $\langle\ ,\ \rangle$ and $\langle\ ,\ \rangle'$ are closely related. Indeed,

$$\begin{aligned}\langle X'_i, X'_j\rangle' = s'_{ij} &= (\text{constant})\cdot\sum_{k,l} a_{ki}s_{kl}a_{lj}\\ &= (\text{constant})\cdot\sum_{k,l} a_{ki}a_{lj}\langle X_k, X_l\rangle\\ &= (\text{constant})\cdot\left\langle\sum_k a_{ki}X_k, \sum_l a_{lj}X_l\right\rangle\\ &= (\text{constant})\cdot\langle X'_i, X'_j\rangle \qquad \text{by (2),}\end{aligned}$$

so that we have

$$\langle\ ,\ \rangle' = (\text{constant})\cdot\langle\ ,\ \rangle.$$

By restricting our attention to inward pointing X_3, we thus obtain a class of positive definite inner products on M_p, any one being a (positive) constant multiple of any other. We can express this by saying that we have defined a "conformal structure" on M_p (compare pg. II.296). This conformal structure is invariant under all affine maps, provided only that they are orientation preserving: If $A\colon \mathbb{R}^3 \to \mathbb{R}^3$ is any orientation preserving affine map and $p \in M \subset \mathbb{R}^3$ is an elliptic point, then the class of inner products defined on M_p is precisely A^* of the class of inner products defined on the tangent space $A(M)_{A(p)}$ of $A(M)$ at $A(p)$.

A conformal structure on M_p does not allow us to pick out orthonormal bases, but it does make sense to consider bases X_1, X_2 of M_p with

$$\langle X_i, X_j\rangle = (\text{constant} > 0)\cdot\delta_{ij},$$

since this condition does not depend on the choice of the inner product $\langle\ ,\ \rangle$ in our conformal structure. Such bases may be called "quasi-orthonormal". They can clearly be characterized quite simply as follows: for any inward pointing vector $X_3 \in \mathbb{R}^3{}_p$, the corresponding osculating paraboloid P is given by

$$P = \{sX_1 + tX_2 + (\text{constant} > 0)\cdot(s^2 + t^2)\cdot X_3\}.$$

These considerations can be made in a less geometric way, but with the calculations going through more smoothly, by using moving frames. For an adapted moving frame X_1, X_2, X_3 on a surface $M \subset \mathbb{R}^3$ we will still use the dual and connection forms ϕ^α and ψ^α_β for moving frames in $\mathbb{R}^3$, and we again let θ^1, θ^2

be the dual forms determined by the moving frame X_1, X_2 on M. As in ordinary surface theory, we have

$$\begin{aligned} \phi^i &= \theta^i \quad \text{on } TM \qquad i = 1,2 \\ \phi^3 &= 0 \quad \text{on } TM; \end{aligned}$$

moreover the first structural equation,

$$d\phi^3 = -\sum_{\gamma=1}^{3} \psi_\gamma^3 \wedge \phi^\gamma,$$

still implies that

$$0 = \sum_{k=1}^{2} \theta^k \wedge \psi_k^3 \quad \text{on } TM,$$

so that by Cartan's Lemma there is a matrix $S = (s_{ij})$ with

(a)
$$\begin{aligned} \psi_j^3 &= \sum_{i=1}^{2} s_{ij}\theta^i \qquad \text{on } TM \\ s_{ij} &= s_{ji}. \end{aligned}$$

We will soon be able to compare this matrix S with the one defined previously. First we want to consider another adapted moving frame X'_1, X'_2, X'_3 on M. The matrix a with $X'_\alpha = \sum_\beta a_{\beta\alpha} X_\beta$ must be of the form

$$a = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{pmatrix},$$

and we easily find that

(b)
$$(a^{-1})_{3i} = 0 \qquad i = 1,2; \qquad (a^{-1})_{33} = \frac{1}{a_{33}}.$$

The dual forms ϕ'^α for the X'_α are given by (see pg. II. 282)

(c)
$$\begin{aligned} \phi' = a^{-1}\phi \implies \phi &= a \cdot \phi' \\ \implies \theta^i &= \sum_{j=1}^{2} a_{ij}\theta'^j, \end{aligned}$$

while the connection forms ψ'^{α}_{β} are related to the ψ^{α}_{β} by (pg. II.280)

(d) $$\psi' = a^{-1}\,da + a^{-1}\psi a.$$

In particular,

$$\begin{aligned}\psi'^3_i &= (a^{-1}\,da)_{3i} + (a^{-1}\psi a)_{3i}\\ &= \sum_{\alpha=1}^{3}(a^{-1})_{3\alpha}\,da_{\alpha i} + \sum_{\alpha,\beta=1}^{3}(a^{-1})_{3\alpha}\psi^{\alpha}_{\beta}a_{\beta i}\\ &= 0 + \sum_{j=1}^{2}\frac{1}{a_{33}}\psi^3_j a_{ji} \qquad \text{by (b).}\end{aligned}$$

Hence

$$\begin{aligned}\psi'^3_i &= \frac{1}{a_{33}}\sum_{j,k} a_{ji}s_{jk}\theta^k\\ &= \frac{1}{a_{33}}\sum_{j,k,l} a_{ji}s_{jk}a_{kl}\theta'^l.\end{aligned}$$

So if we also write

$$\psi'^3_j = \sum_{i=1}^{2} s'_{ij}\theta'^i, \qquad s'_{ij} = s'_{ji},$$

then the matrix S' is related to the matrix S by

(e) $$S' = \frac{1}{a_{33}}A^{\mathrm{t}}SA,$$

where A is the 2×2 matrix $A = (a_{ij})$.

As a first consequence of this equation we see, what is not *a priori* clear, that

The matrix $S(p)$ depends only on the vectors $X_1(p), X_2(p), X_3(p)$.

Taking X_3 to be a parallel vector field in $\mathbb{R}^3$, we easily see that $S(p)$ is, in fact, the same as the matrix S on page 77, for the basis $X_1(p), X_2(p), X_3(p)$ of $\mathbb{R}^3{}_p$. Then equation (e) is just equation (I) on page 78. As before, we then see from equation (e) that if p is an elliptic point, then the inner product

$$\sum_{i,j} s_{ij}(p)\cdot\theta^i(p)\otimes\theta^j(p)$$

on M_p is well-defined up to constant multiple, and hence, by considering only the case where $X_3(p)$ is inward pointing, we again obtain the (general orientation preserving affine invariant) conformal structure on M_p. Clearly the basis X_1, X_2 of M_p is quasi-orthonormal [that is, $\langle X_i(p), X_j(p)\rangle =$ (constant > 0) $\cdot\, \delta_{ij}$] if and only if

$$S(p) = (\text{constant } > 0)\cdot\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \qquad \text{or} \qquad \psi_i^3(p) = (\text{constant} > 0)\cdot\theta^i(p).$$

Now from among our class of inner products on M_p we can distinguish a particular one $\langle\ ,\ \rangle_p$, which will be a special affine invariant. We do this by defining $X_1, X_2 \in M_p$ to be **orthonormal with respect to** $\langle\ ,\ \rangle_p$ if and only if

$$(*) \qquad S(p) = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \qquad \begin{array}{l}\text{for all inward pointing } X_3 \in \mathbb{R}^3{}_p \text{ with}\\ \det(X_1, X_2, X_3) = \pm 1\end{array}$$

(the sign depending on whether or not (X_1, X_2) is positively oriented). To check that this is well-defined, note first that if we also have

$$\det(X_1, X_2, X'_3) = \pm 1$$

for an inward pointing X'_3, then clearly $a_{33}(p) = 1$, so equation (e) shows that $S'(p) = S(p)$; consequently, condition $(*)$ does not depend on the choice of X_3. Moreover, for fixed $X_3 \in \mathbb{R}^3{}_p$, and different $X'_1, X'_2 \in M_p$, equation (e) shows that $S' = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$ if and only if $A^{\mathbf{t}}A = I$, which means that X'_1, X'_2 is related to X_1, X_2 by the orthogonal transformation A; hence the inner product which makes X_1, X_2 orthonormal also makes X'_1, X'_2 orthonormal. It should be clear, from the very definition, that $\langle\ ,\ \rangle_p$ is a special affine invariant: If $A\colon \mathbb{R}^3 \to \mathbb{R}^3$ is special linear affine, and $p \in M$ is an elliptic point, then the inner product $\langle\ ,\ \rangle_p$ on M_p is A^* of the inner product $\langle\ ,\ \rangle_{A(p)}$ on the tangent space $A(M)_{A(p)}$ of $A(M)$ at $A(p)$. Clearly a basis $X_1, X_2 \in M_p$ is orthonormal with respect to $\langle\ ,\ \rangle_p$ if and only if

$$\psi_i^3(p) = \theta^i(p) \qquad \text{or} \qquad P = \left\{sX_1 + tX_2 + \tfrac{1}{2}(s^2 + t^2)X_3\right\}$$

for every inward pointing $X_3 \in \mathbb{R}^3{}_p$ with $\det(X_1, X_2, X_3) = \pm 1$.

On a surface M with all points elliptic, we now have an inner product $\langle\ ,\ \rangle_p$ defined on each M_p, and thus we have a Riemannian metric $\langle\ ,\ \rangle$ on M. This Riemannian metric will also be denoted by I, and called the **special affine first fundamental form** of M. If it seems strange that we can find a special affine invariant metric on elliptically curved surfaces in $\mathbb{R}^3$ even though there is clearly no special affine invariant metric on $\mathbb{R}^3$ itself, it might help to observe that we have essentially the same situation for 1-dimensional manifolds M in $\mathbb{R}^3$, for we can define the unit tangent vectors of M to be those of the form $c'(\sigma)$, where $c\colon [0,1] \to M$ is a curve parameterized by special affine arclength. Once again the manifold M cannot be too flat (and in fact the requisite condition is more stringent, since it involves third derivatives). Naturally, if $f\colon M \to \mathbb{R}^3$ is an immersion with all points of the image elliptic, then we define the **special affine first fundamental form** $\mathcal{I}_f$ of f to be the tensor $\mathcal{I}_f = f^*(\langle\!\langle\ ,\ \rangle\!\rangle)$ on M, where $\langle\!\langle\ ,\ \rangle\!\rangle$ is the special affine first fundamental form on $f(M)$. Notice that this is not completely analogous to the definition in ordinary surface theory, where we can simply define $\mathrm{I}_f = f^*\langle\ ,\ \rangle$ for $\langle\ ,\ \rangle$ the usual Riemannian metric on $\mathbb{R}^3$; it is much closer to the definition of II_f. In fact, we have already noted that $\langle\!\langle\ ,\ \rangle\!\rangle_p$ is a multiple of $\mathrm{II}(p)$; just which multiple will soon be determined.

When $f\colon U \to \mathbb{R}^3$ for $U \subset \mathbb{R}^2$ open, and all points of $f(U)$ are elliptic, we define the functions $g_{ij}\colon U \to \mathbb{R}^3$ to be the components of $\mathcal{I}_f$ with respect to the standard coordinate system (s,t) on $\mathbb{R}^2$, so that

$$\mathcal{I}_f = g_{11}\, ds \otimes ds + g_{12}\, ds \otimes dt + g_{21}\, dt \otimes ds + g_{22}\, dt \otimes dt.$$

We would naturally like to be able to compute the g_{ij} in terms of f. We first take the case where f is simply

$$f(s,t) = (s,t,h(s,t)), \qquad h(0,0) = h_1(0,0) = h_2(0,0) = 0,$$

with $p = f(0,0) = 0 \in \mathbb{R}^3$. If $X_1, X_2, X_3 \in \mathbb{R}^3{}_0$ is the standard basis, then the corresponding osculating paraboloid P is the graph of

$$(s,t) \mapsto \tfrac{1}{2}(\alpha s^2 + 2\beta st + \gamma t^2) \qquad \text{for} \qquad \begin{cases} \alpha = h_{11}(0,0) \\ \beta = h_{12}(0,0) \\ \gamma = h_{22}(0,0). \end{cases}$$

If p is an elliptic point, then $\alpha\gamma - \beta^2 > 0$. For the sake of concreteness, suppose also that P lies above the (x,y)-plane, so that X_3 is inward pointing. There is another basis $X'_j = \sum_{i=1}^{2} a_{ij} X_i$ of $\mathbb{R}^2$ such that P is the graph of

$$(s,t) \mapsto \tfrac{1}{2}(s^2 + t^2)$$

in the X'_1, X'_2, X_3 system. Equation (1) on page 78 (or equation (e) on page 81) shows that the matrix $A = (a_{ij})$ satisfies

$$(1) \qquad A^{\mathrm{t}} \cdot \frac{1}{2}\begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix} \cdot A = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \implies \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix} = (A^{\mathrm{t}})^{-1} A^{-1}$$
$$\implies \det A = (\alpha\gamma - \beta^2)^{-1/2}.$$

Now P is also the graph of $(s,t) \mapsto \frac{1}{2}(s^2+t^2)$ in the $(\lambda X'_1, \lambda X'_2, \lambda^2 X_3)$ coordinate system. In order to have

$$\begin{aligned} 1 &= \det(\lambda X'_1, \lambda X'_2, \lambda^2 X_3) = \lambda^4 \det(X'_1, X'_2, X_3) \\ &= \lambda^4 \cdot \det A = \lambda^4 (\alpha\gamma - \beta^2)^{-1/2}, \end{aligned}$$

we must take

$$\lambda = \sqrt[8]{\alpha\gamma - \beta^2}.$$

So the vectors

$$X''_1 = \left(\sqrt[8]{\alpha\gamma - \beta^2}\right) X'_1, \qquad X''_2 = \left(\sqrt[8]{\alpha\gamma - \beta^2}\right) X'_2$$

are orthonormal with respect to $\langle\ ,\ \rangle_0$. To figure out the numbers $g_{ij}(0,0)$, we note that if $B = (b_{ij})$ is the inverse of the matrix A, then

$$X_i = \sum_{j=1}^{2} b_{ji} X'_j = \frac{1}{\sqrt[8]{\alpha\gamma - \beta^2}} \sum_{j=1}^{2} b_{ji} X''_i,$$

and consequently

$$\begin{aligned} g_{ij}(0,0) = \langle X_i, X_j \rangle_0 &= \frac{1}{\sqrt[4]{\alpha\gamma - \beta^2}} \left\langle \sum_{k=1}^{2} b_{ki} X''_k, \sum_{k=1}^{2} b_{kj} X''_k \right\rangle_0 \\ &= \frac{1}{\sqrt[4]{\alpha\gamma - \beta^2}} \sum_{k=1}^{2} b_{ki} b_{kj} \\ &= \frac{1}{\sqrt[4]{\alpha\gamma - \beta^2}} (B^{\mathrm{t}} B)_{ij} = \frac{1}{\sqrt[4]{\alpha\gamma - \beta^2}} \left((A^{\mathrm{t}})^{-1} A^{-1}\right)_{ij}. \end{aligned}$$

Using (1), we see that

(2) $$g_{ij}(0,0) = \frac{h_{ij}(0,0)}{\sqrt[4]{\det(h_{ij}(0,0))}}.$$

This can also be written

(3) $$g_{ij}(0,0) = \frac{d_{ij}}{\sqrt[4]{\det(d_{ij})}}(0,0),$$
$$\text{where } d_{ij} = \det(f_1, f_2, f_{ij}).$$

These calculations were all carried out for the case where f is of the form $f(s,t) = (s,t,h(s,t))$, with $h(0,0) = h_1(0,0) = h_2(0,0) = 0$. We could try to deal with the general map $f\colon U \to \mathbb{R}^3$ by reducing it to this case. For example, if $f_1(0,0)$ and $f_2(0,0)$ both lie in the (x,y)-plane, we could first determine a function h by the condition

$$\{(f^1(s,t), f^2(s,t), f^3(s,t))\} = f(U) = \{(s,t,h(s,t))\}$$
$$\Longrightarrow h(f^1(s,t), f^2(s,t)) = f^3(s,t),$$

use this equation to relate the $h_{ij}(0,0)$ to the $f_i(0,0)$ and $f_{ij}(0,0)$, and then use (2) to find $g_{ij}(0,0)$ in terms of these numbers (we would still have to take care of the general case when $f_1(0,0)$ and $f_2(0,0)$ do not lie in $\mathbb{R}^2$). A much simpler course of action is to guess from (3) that the answer should be

$$\mathrm{I}_f = \frac{d_{11}\,ds\otimes ds + d_{12}\,ds\otimes dt + d_{21}\,dt\otimes ds + d_{22}\,dt\otimes dt}{\sqrt[4]{\det(d_{ij})}}$$
$$= \frac{1}{\sqrt[4]{\det(d_{ij})}}\sum_{i,j=1}^{2} d_{ij}\,ds^i\otimes ds^j, \quad \text{using } (s^1,s^2) \text{ for } (s,t).$$

Now this guess cannot be precisely correct, for if we define $\tilde f(s^1,s^2) = f(s^2,s^1)$, then we have $\tilde d_{ij}(s^1,s^2) = \det(\tilde f_1, \tilde f_2, \tilde f_{12})(s^1,s^2) = \det(f_2, f_1, f_{12})(s^2,s^1) = -d_{ij}(s^2,s^1)$, so our formula changes sign. The problem, of course, is that $\tilde f$ is orientation reversing if f is orientation preserving. The right guess is that the above formula holds whenever $f\colon U \to M$ is orientation preserving. To prove this, we note first that the right side is clearly a special affine invariant, since it involves only determinants d_{ij}; consequently, there is no loss of generality in assuming that the tangent plane at the point in question is the (x,y)-plane. We still based our calculations on a very special parameterization, so we want to check that the right side is "invariant under orientation preserving change of parameter": if $f\colon U \to \mathbb{R}^3$ is any immersion (for $U \subset \mathbb{R}^2$ open), and $p = (p^1,p^2)\colon V \to U$ is an orientation preserving diffeomorphism (for $V \subset \mathbb{R}^2$ open), then we want to check that

$$(4)\quad p^*\left(\frac{1}{\sqrt[4]{\det(d_{ij})}}\sum_{i,j=1}^{2} d_{ij}\,ds^i\otimes ds^j\right) = \frac{1}{\sqrt[4]{\det(\tilde d_{ij})}}\sum_{i,j=1}^{2} \tilde d_{ij}\,ds^i\otimes ds^j,$$

where the $\tilde d_{ij}$ are the d_{ij} for $\tilde f = f\circ p$.

Now

$$(5)\quad p^*\left(\sum_{i,j=1}^{2} d_{ij}\, ds^i \otimes ds^j\right) = \sum_{i,j=1}^{2} (d_{ij} \circ p)\left(\sum_{\rho=1}^{2} p^i{}_\rho\, ds^\rho\right) \otimes \left(\sum_{\sigma=1}^{2} p^j{}_\sigma\, ds^\sigma\right)$$
$$= \sum_{i,j=1}^{2} \sum_{\rho,\sigma=1}^{2} p^i{}_\rho p^j{}_\sigma (d_{ij} \circ p)\, ds^\rho \otimes ds^\sigma$$
$$= \sum_{i,j=1}^{2} \left(\sum_{\rho,\sigma=1}^{2} p^\rho{}_i p^\sigma{}_j (d_{\rho\sigma} \circ p)\right) ds^i \otimes ds^j .$$

On the other hand, since

$$\tilde{f}_i = D_i(f \circ p) = \sum_{\rho=1}^{2} p^\rho{}_i (f_\rho \circ p)$$
$$\tilde{f}_{ij} = \sum_{\rho=1}^{2} p^\rho{}_{ij} (f_\rho \circ p) + \sum_{\rho,\sigma=1}^{2} p^\rho{}_i p^\sigma{}_j (f_{\rho\sigma} \circ p),$$

we have

$$(6)\quad \tilde{d}_{ij} = \det(\tilde{f}_1, \tilde{f}_2, \tilde{f}_{ij})$$
$$= \det\left(\sum_{\mu=1}^{2} (f_\mu \circ p) \cdot p^\mu{}_1,\ \sum_{\nu=1}^{2} (f_\nu \circ p) \cdot p^\nu{}_2,\ \sum_{\rho,\sigma=1}^{2} p^\rho{}_i p^\sigma{}_j (f_{\rho\sigma} \circ p) + \sum_{\rho=1}^{2} p^\rho{}_{ij} (f_\rho \circ p)\right)$$
$$= \det\left(\sum_{\mu=1}^{2} (f_\mu \circ p) \cdot p^\mu{}_1,\ \sum_{\nu=1}^{2} (f_\nu \circ p) \cdot p^\nu{}_2,\ \sum_{\rho,\sigma=1}^{2} p^\rho{}_i p^\sigma{}_j (f_{\rho\sigma} \circ p)\right)$$
$$= \sum_{\mu,\nu=1}^{2} p^\mu{}_1 p^\nu{}_2 \sum_{\rho,\sigma=1} p^\rho{}_i p^\sigma{}_j \det(f_\mu \circ p, f_\nu \circ p, f_{\rho\sigma} \circ p)$$
$$= (\det p') \cdot \left[\sum_{\rho,\sigma=1}^{2} p^\rho{}_i p^\sigma{}_j (d_{\rho\sigma} \circ p)\right],$$

and consequently

$$(7)\qquad \sum_{i,j=1}^{2} \tilde{d}_{ij}\, ds^i \otimes ds^j = (\det p') \sum_{i,j=1}^{2} \left(\sum_{\rho,\sigma=1}^{2} p^\rho{}_i p^\sigma{}_j (d_{\rho\sigma} \circ p)\right) ds^i \otimes ds^j .$$

If it weren't for the factor (det p'), the tensors in (5) and (7) would already be the same. From (6) we easily see that

$$(8)\qquad \det(\tilde{d}_{ij}) = (\det p')^4 \cdot \det(d_{ij}\circ p)$$
$$\Longrightarrow \sqrt[4]{\det(\tilde{d}_{ij})} = (\det p') \cdot \sqrt[4]{\det(d_{ij}\circ p)} \qquad \text{for } \det p' > 0.$$

Together with (7), this gives exactly the equation (4) which we want. We have thus shown that for orientation preserving $f\colon U \to M$ we always have

$$\mathrm{I}_f = \sum_{i,j=1}^{2} g_{ij}\, ds^i \otimes ds^j$$
$$\text{for}\quad g_{ij} = \frac{d_{ij}}{\sqrt[4]{\det(d_{ij})}}, \qquad d_{ij} = \det(f_1, f_2, f_{ij}).$$

This formula allows us to compare the special affine first fundamental form I_f with the ordinary *second* fundamental form II_f, whose coefficients l_{ij} are

$$l_{ij} = \langle n, f_{ij}\rangle = \left\langle \frac{f_1\times f_2}{|f_1\times f_2|}, f_{ij}\right\rangle$$
$$= \frac{\det(f_1, f_2, f_{ij})}{|f_1\times f_2|} = \frac{d_{ij}}{\det(g_{ij})}.$$

In the classical literature, the tensor I_f is introduced a little differently. One simply notes that $\sum_{i,j} d_{ij}\, ds^i \otimes ds^j$ is a nice tensor to consider, because it is a special affine invariant. Then one asks whether it is also an invariant under change of parameter. After deriving equation (7) one sees that it isn't, but upon noticing (8), one realizes that dividing by $\sqrt[4]{\det(d_{ij})}$ will give a tensor that *is* invariant under orientation preserving change of parameter, yet still a special affine invariant.

Now consider a hyperbolic point $p \in M$, and a basis $X_1, X_2, X_3 \in \mathbb{R}^3{}_p$ with $X_1, X_2 \in M_p$. This again determines a matrix $S = (s_{ij})$, and we can still define an inner product on M_p by

$$\langle X_i, X_j\rangle = s_{ij},$$

but now the inner product is merely non-degenerate, and neither positive definite nor negative definite. Any other such inner product, defined for a different basis, is a constant multiple of this one. If we want these constant multiples all to be positive, then we will have to have a way of selecting a permissible direction for the vectors X_3. So we have to choose, arbitrarily, an orientation for M_p, and then *define* X_3 to "point inward" if and only if (X_1, X_2, X_3) is positively oriented in $\mathbb{R}^3{}_p$ whenever (X_1, X_2) is positively oriented in M_p. By considering only inward pointing X_3 we obtain a class of non-degenerate inner products on M_p, any one being a positive constant multiple of any other. If $A\colon \mathbb{R}^3 \to \mathbb{R}^3$ is any affine map and we give the tangent space $A(M)_{A(p)}$ the orientation which makes $A_*\colon M_p \to A(M)_{A(p)}$ orientation preserving, then the class of inner products on M_p is precisely A^* of the class of inner products on $A(M)_{A(p)}$.

In the present set-up it makes sense to consider ordered bases X_1, X_2 of M_p with

$$\begin{aligned} \langle X_1, X_1\rangle &= -\langle X_2, X_2\rangle = \text{constant} > 0 \\ \langle X_1, X_2\rangle &= 0. \end{aligned}$$

These ordered bases will again be called "quasi-orthonormal". For any such ordered basis, and any inward pointing $X_3 \in \mathbb{R}^3{}_p$, the corresponding osculating paraboloid P is given by

$$P = \{sX_1 + tX_2 + (\text{constant} > 0)\cdot(s^2 - t^2)X_3\}.$$

In terms of moving frames we have

$$S(p) = (\text{constant} > 0)\cdot\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \text{or} \quad \begin{cases} \psi_1^3(p) = (\text{constant} > 0)\cdot\theta^1(p) \\ \psi_2^3(p) = -(\text{constant} > 0)\cdot\theta^2(p). \end{cases}$$

From among our class of inner products on M_p we can again distinguish a particular one $\langle\ ,\ \rangle_p$. We define an ordered basis X_1, X_2 to be "orthonormal",

$$\begin{aligned} \langle X_1, X_1\rangle_p &= -\langle X_2, X_2\rangle_p = 1 \\ \langle X_1, X_2\rangle_p &= 0, \end{aligned}$$

if and only if $S(p) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ for all inward pointing $X_3 \in \mathbb{R}^3{}_p$ satisfying $\det(X_1, X_2, X_3) = 1$. The verification that this is well-defined is similar to the case of an elliptic point. If $A\colon \mathbb{R}^3 \to \mathbb{R}^3$ is a map of the form $A = T\circ B$, where T is a translation and $B\colon \mathbb{R}^3 \to \mathbb{R}^3$ is a linear map of determinant ± 1, and we give the tangent space $A(M)_{A(p)}$ the orientation which makes $A_*\colon M_p \to A(M)_{A(p)}$ orientation preserving, then $\langle\ ,\ \rangle_p$ is A^* of the inner product $\langle\ ,\ \rangle_{A(p)}$ on $A(M)_{A(p)}$. Clearly an ordered basis (X_1, X_2) of M_p is orthonormal if and only if

$$\left.\begin{aligned} \psi_1^3(p) &= \theta^1(p) \\ \psi_2^3(p) &= -\theta^2(p) \end{aligned}\right\} \quad \text{or} \quad P = \{sX_1 + tX_2 + \tfrac{1}{2}(s^2 - t^2)X_3\}$$

for all inward pointing $X_3 \in \mathbb{R}^3{}_p$ with $\det(X_1, X_2, X_3) = 1$.

On an oriented surface M with all points hyperbolic, we now have an inner product $\langle\ ,\ \rangle_p$ defined on each M_p, and hence an "indefinite Riemannian metric" $\langle\ ,\ \rangle$ on M. Once again we also denote $\langle\ ,\ \rangle$ by I, and call it the **special affine first fundamental form** of M. If $f\colon M \to \mathbb{R}^3$ is an immersion of an oriented surface with all points of the image hyperbolic, then we define the **special affine first fundamental form** I_f of f to be the tensor $\mathrm{I}_f = f^*(\langle\ ,\ \rangle)$ on M, where $\langle\ ,\ \rangle$ is the special affine first fundamental form on $f(M)$, when $f(M)$ is given the orientation that makes f orientation preserving.

When $f\colon U \to \mathbb{R}^2$, for $U \subset \mathbb{R}^2$ open (with the usual orientation), and all points of $f(U)$ are hyperbolic, we again define functions $g_{ij}\colon U \to \mathbb{R}^3$ by

$$\mathrm{I}_f = g_{11}\, ds \otimes ds + g_{12}\, ds \otimes dt + g_{21}\, dt \otimes ds + g_{22}\, dt \otimes dt.$$

Again take the case where $f(s,t) = (s,t,h(s,t))$, with $p = f(0,0) = 0 \in \mathbb{R}^3$ and $M_p = (x,y)$-plane. If $X'_j = \sum_{i=1}^{2} a_{ij} X_i$ is a new basis of $\mathbb{R}^2$ such that P is the graph of

$$(s,t) \mapsto \tfrac{1}{2}(s^2 - t^2)$$

in the X'_1, X'_2, X_3 system, then we have

$$\text{(1')}\qquad A^{\mathrm{t}} \cdot \frac{1}{2}\begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix} \cdot A = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$\Longrightarrow \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix} = (A^{\mathrm{t}})^{-1}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} A^{-1}$$

$$\Longrightarrow \det A = (\beta^2 - \alpha\gamma)^{-1/2}.$$

As before, we see that the vectors

$$X''_1 = \left(\sqrt[8]{\beta^2 - \alpha\gamma}\right) X'_1, \qquad X''_2 = \left(\sqrt[8]{\beta^2 - \alpha\gamma}\right) X'_2$$

satisfy

$$\langle X''_1, X''_1\rangle_0 = -\langle X''_2, X''_2\rangle_0 = 1$$
$$\langle X''_1, X''_2\rangle_0 = 0.$$

So, introducing the matrix B as before, we have

$$\begin{aligned} g(0,0) = \langle X_i, X_j \rangle_0 &= \frac{1}{\sqrt[4]{\beta^2 - \alpha\gamma}} \left\langle \sum_{k=1}^{2} b_{ki} X''_k, \sum_{k=1}^{2} b_{kj} X''_k \right\rangle_0 \\ &= \frac{1}{\sqrt[4]{\beta^2 - \alpha\gamma}} \cdot (b_{1i}b_{1j} - b_{2i}b_{2j}) \\ &= \frac{1}{\sqrt[4]{\beta^2 - \alpha\gamma}} \left[B^{\mathrm{t}} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} B \right]_{ij} \\ &= \frac{1}{\sqrt[4]{\beta^2 - \alpha\gamma}} \left[(A^{\mathrm{t}})^{-1} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} A^{-1} \right]_{ij}. \end{aligned}$$

Thus (1′) gives

$$g_{ij}(0,0) = \frac{d_{ij}}{\sqrt[4]{-\det(d_{ij})}}(0,0).$$

The same calculations as before show that this formula holds for any $f\colon U \to \mathbb{R}^3$. We can refer to the elliptic and hyperbolic cases jointly by means of equation

$$\text{(I)} \qquad \mathrm{I}_f = \sum_{i,j=1}^{2} g_{ij}\, ds^i \otimes ds^j$$

$$\text{for} \quad g_{ij} = \frac{d_{ij}}{\sqrt[4]{|\det(d_{ij})|}}, \qquad d_{ij} = \det(f_1, f_2, f_{ij}).$$

If M consists entirely of elliptic points, then the map $f\colon U \to M$ must be orientation preserving when M is given its natural orientation, and if M consists entirely of hyperbolic points, then M must be given the orientation which makes f orientation preserving. (Henceforth we will not bother to mention the subsidiary conditions on orientation which must be added to all our considerations.)

Finally, what do we do at points $p \in M$ which are flat (parabolic or planar)? The answer is, we don't do anything. We do not define $\langle\ ,\ \rangle$ and *we cannot expect to.* To see that this must be so, just consider the surface $\mathbb{R}^2 \subset \mathbb{R}^3$, consisting entirely of flat points. If we could define a metric $\langle\ ,\ \rangle$ on $\mathbb{R}^2$ which was a special affine invariant, then we would have to have $A^*\langle\ ,\ \rangle = \langle\ ,\ \rangle$ for every special linear affine map A from $\mathbb{R}^2$ to $\mathbb{R}^2$, since such an A can always be extended to a special linear affine map $A'\colon \mathbb{R}^3 \to \mathbb{R}^3$. But of course, there is no metric on $\mathbb{R}^2$ which is invariant under all special linear affine maps. In affine geometry we will simply always assume, without explicitly mentioning it again, that all surfaces have no flat points. Thus connected surfaces will consist either entirely of elliptic points, or entirely of hyperbolic points.

After all this work, we are hardly any closer to the problem of picking out a "special affine normal". In order to do this we will have to consider *third* order approximations to M. We will still assume that $p = 0 \in \mathbb{R}^3$ and that M_p is the (x, y)-plane. Choose some vector $X_3 \in \mathbb{R}^3$ which is not in $\mathbb{R}^2 = (x, y)$-plane. For every basis $\mathbf{X} = (X_1, X_2)$ of $\mathbb{R}^2$, we can then consider the function $h^{\mathbf{X}}$ which describes M in the X_1, X_2, X_3 coordinate system, and we can look at the quadratic and cubic polynomials

$$\tfrac{1}{2}\big(h^{\mathbf{X}}{}_{11}(0,0)\cdot s^2 + 2h^{\mathbf{X}}{}_{12}(0,0)\cdot st + h^{\mathbf{X}}{}_{22}(0,0)\cdot t^2\big)$$
$$\tfrac{1}{6}\big(h^{\mathbf{X}}{}_{111}(0,0)\cdot s^3 + 3h^{\mathbf{X}}{}_{112}(0,0)\cdot s^2t + 3h^{\mathbf{X}}{}_{122}(0,0)\cdot st^2 + h^{\mathbf{X}}{}_{222}(0,0)\cdot t^3\big)$$

which appear in the Taylor series for $h^{\mathbf{X}}$. As on page 37, we can also define functions $\Phi^{\mathbf{X}}, \Psi^{\mathbf{X}}\colon \mathbb{R}^2 \to \mathbb{R}$ by

$$\begin{aligned}\Phi^{\mathbf{X}}(sX_1 + tX_2) &= \tfrac{1}{2}\big(h^{\mathbf{X}}{}_{11}(0,0)\cdot s^2 + 2h^{\mathbf{X}}{}_{12}(0,0)\cdot st + h^{\mathbf{X}}{}_{22}(0,0)\cdot t^2\big)\\ \Psi^{\mathbf{X}}(sX_1 + tX_2) &= \tfrac{1}{6}\big(h^{\mathbf{X}}{}_{111}(0,0)\cdot s^3 + 3h^{\mathbf{X}}{}_{112}(0,0)\cdot s^2t\\ &\qquad + 3h^{\mathbf{X}}{}_{122}(0,0)\cdot st^2 + h^{\mathbf{X}}{}_{222}(0,0)\cdot t^3\big).\end{aligned}$$

All the $\Phi^{\mathbf{X}}$ are really the same function $\Phi\colon \mathbb{R}^2 \to \mathbb{R}$, and all the $\Psi^{\mathbf{X}}$ are really the same function $\Psi\colon \mathbb{R}^2 \to \mathbb{R}$. We have already checked this for the $\Phi^{\mathbf{X}}$'s, using the relation

(1) $$h^{\mathbf{X}}{}_{\alpha\beta}(0,0) = \sum_{j,k=1}^{2} a_{j\alpha}a_{k\beta}h_{jk}(0,0);$$

the check for the $\Psi^{\mathbf{X}}$'s is similar, using the easily derived relation

(2) $$h^{\mathbf{X}}{}_{\alpha\beta\gamma}(0,0) = \sum_{j,k,l=1}^{2} a_{j\alpha}a_{k\beta}a_{l\gamma}h_{jkl}(0,0).$$

Remember that these functions Φ and Ψ *do* depend on the original choice of X_3. We would now like to ask if there is a particular choice for the direction of X_3 which will make the functions Φ and Ψ be related to each other in some especially nice way; so the real problem here is to formulate a definite question, by deciding on a suitable criterion for declaring that Φ and Ψ are nicely related.

We now find ourselves placed in a purely algebraic situation, which we can formulate as follows. A function $\Phi\colon \mathbb{R}^2 \to \mathbb{R}$ will be called **quadratic** if for a basis X_1, X_2 of $\mathbb{R}^2$ we have

$$\Phi(sX_1 + tX_2) = \Phi_{11}s^2 + 2\Phi_{12}st + \Phi_{22}t^2,$$

for some numbers Φ_{jk}. More generally, for an n-dimensional vector space V, a function $\Phi\colon V \to \mathbb{R}$ will be called **quadratic** if for a basis $X_1, \dots, X_n$ of V we have

$$\Phi\left(\sum_{i=1}^{n} s^i X_i\right) = \sum_{j,k=1}^{n} \Phi_{jk} s^j s^k$$

for certain numbers Φ_{jk}, which it will be convenient to assume are symmetric with respect to the indices j and k. It is easy to see that if Φ has this form for one basis, then it has this form for any other basis. Indeed, if $\bar{X}_1, \dots, \bar{X}_n$ is another basis, with $\bar{X}_j = \sum_i a_{ij} X_i$, then

$$\begin{aligned}\Phi\left(\sum_{i=1}^{n} s^i \bar{X}_i\right) &= \Phi\left(\sum_{i=1}^{n}\sum_{\rho=1}^{n} s^i a_{\rho i} X_\rho\right) = \Phi\left(\sum_{\rho=1}^{n}\left(\sum_{i=1}^{n} s^i a_{\rho i}\right) X_\rho\right) \\ &= \sum_{j,k=1}^{n} \Phi_{jk}\left(\sum_{i=1}^{n} s^i a_{ji}\right)\left(\sum_{i=1}^{n} s^i a_{ki}\right) \\ &= \sum_{\alpha,\beta=1}^{n}\left(\sum_{j,k=1}^{n} \Phi_{jk} a_{j\alpha} a_{k\beta}\right) s^\alpha s^\beta = \sum_{\alpha,\beta=1}^{n} \bar{\Phi}_{\alpha,\beta} s^\alpha s^\beta,\end{aligned}$$

where the $\bar{\Phi}_{\alpha\beta}$ are given by

(3) $$\bar{\Phi}_{\alpha\beta} = \sum_{j,k=1}^{n} \Phi_{jk} a_{j\alpha} a_{k\beta}.$$

Naturally, (3) is just the n-dimensional analogue of the equations (1) which gave us a well-defined Φ in the first place. Notice that in terms of the matrix $A = (a_{ij})$ and the matrices $[\Phi] = (\Phi_{jk})$ and $[\bar{\Phi}] = (\bar{\Phi}_{\alpha\beta})$ we can write (3) as

(4) $$[\bar{\Phi}] = A^{\mathrm{t}} \cdot [\Phi] \cdot A.$$

We will define a function $\Psi\colon V \to \mathbb{R}$ to be **cubic** if for a basis $X_1, \dots, X_n$ of V we have

$$\Psi\left(\sum_{i=1}^{n} s^i X_i\right) = \sum_{j,k,l=1}^{n} \Psi_{jkl} s^j s^k s^l,$$

for certain numbers Ψ_{jkl}, which we can assume are symmetric in the indices. For the basis $\bar{X}$ we easily find that

$$\Psi\left(\sum_{i=1}^{n} s^i \bar{X}_i\right) = \sum_{\alpha,\beta,\gamma=1}^{n} \bar{\Psi}_{\alpha\beta\gamma} s^\alpha s^\beta s^\gamma,$$

where

$$\bar{\Psi}_{\alpha\beta\gamma} = \sum_{j,k,l=1}^{n} \Psi_{jkl} a_{j\alpha} a_{k\beta} a_{l\gamma}; \tag{5}$$

this equation, of course, is just the analogue of (2). It is not hard to see that we could also define a quadratic function $\Phi\colon V \to \mathbb{R}$ to be one of the form $\Phi(X) = B(X, X)$ for a symmetric bilinear function $B\colon V \times V \to \mathbb{R}$. Similarly, a cubic function $\Psi\colon V \to \mathbb{R}$ is one of the form $\Psi(X) = T(X, X, X)$ for a symmetric trilinear function $T\colon V \times V \times V \to \mathbb{R}$.

Now we would like to find some quantity depending on Φ and Ψ, but *not* on the choice of basis, and hence not on the particular coefficients Φ_{jk} and Ψ_{jkl} for this basis. As a warm-up, let's first take the case of two *quadratic* forms Φ and Θ. We will assume that the first quadratic form Φ is non-degenerate, by which we mean that the corresponding symmetric bilinear form is non-degenerate. More concretely, this means that if we choose a basis $X_1, \ldots, X_n$ for V, then the matrix $[\Phi] = (\Phi_{jk})$ is non-singular, and therefore has an inverse matrix $[\Phi]^{-1} = (\Phi^{jk})$ with

$$\sum_{k=1}^{n} \Phi_{jk}\Phi^{kl} = \delta_j^l.$$

Now consider the number

$$\sum_{j,k=1}^{n} \Phi^{jk}\Theta_{jk}.$$

If $\bar{X}_j = \sum_i a_{ij} X_i$ is a new basis, and $B = (b_{ij})$ is the inverse of $A = (a_{ij})$, then (4) shows that the matrix $[\bar{\Phi}] = (\bar{\Phi}_{\alpha\beta})$ satisfies

$$[\bar{\Phi}] = A^{\mathbf{t}}[\Phi]A \implies [\bar{\Phi}]^{-1} = B[\Phi]^{-1}B^{\mathbf{t}}$$

and hence

$$\bar{\Phi}^{\alpha\beta} = \sum_{j,k=1}^{n} \Phi^{jk} b_{\alpha j} b_{\beta k}. \tag{6}$$

Applying (3) to Θ, we thus find that

$$\begin{aligned}\sum_{\alpha,\beta=1}^{n} \bar{\Phi}^{\alpha\beta}\bar{\Theta}_{\alpha\beta} &= \sum_{\alpha,\beta=1}^{n}\sum_{j,k=1}^{n} \Phi^{jk} b_{\alpha j} b_{\beta k} \sum_{l,m=1}^{n} \Theta_{lm} a_{l\alpha} a_{m\beta} \\ &= \sum_{j,k=1}^{n}\sum_{l,m=1}^{n} \Phi^{jk}\Theta_{lm}\delta_{lj}\delta_{mk} \\ &= \sum_{j,k=1}^{n} \Phi^{jk}\Theta_{jk}.\end{aligned}$$

We thus have a well-defined number (Φ,Θ), which for any basis $\{X_i\}$ is given by

$$(\Phi,\Theta) = \sum_{j,k=1}^{n} \Phi^{jk}\Theta_{jk}.$$

Naturally, there must also be some invariant definition of (Φ,Θ) lurking around. Indeed, the bilinear function corresponding to Φ is just what we usually call an inner product $\langle\ ,\ \rangle$ on V. This gives rise to inner products on just about every other vector space in sight which is related to V. In particular, we could define an inner product $\langle\!\langle\ ,\ \rangle\!\rangle$ on the set of all bilinear functions $C\colon V \times V \to \mathbb{R}$; the quadratic function Θ corresponds to such a C, and (Φ,Θ) is just $\langle\!\langle C, C\rangle\!\rangle$. But in this case I don't think it's worth all the linear algebra which this involves; it's easier to just do the calculation.

To get a little more feeling for what the number (Φ,Θ) is, take the case where the inner product corresponding to Φ is actually positive definite. Then there is a basis $X_1,\dots,X_n$ of V such that $\Phi_{ij} = \delta_{ij}$ and also $\Theta_{ij} = 0$ for $i \neq j$ (compare Proposition II.4-14). So (Φ,Θ) is just the sum of the "diagonal terms", $(\Phi,\Theta) = \sum_i \Theta_{ii}$. When $(\Phi,\Theta) = 0$, the forms Φ and Θ are said to be **apolar**, a term that comes from the old invariant theory. We can give a very concrete geometric meaning to apolarity when V is 2-dimensional. For our special choice of basis, we see that Θ is apolar to Φ if and only if the graph of $\Theta\colon V \to \mathbb{R}$ is a hyperbolic paraboloid,

$$\begin{aligned}\Theta(sX_1 + tX_2) &= \Theta_{11}s^2 + \Theta_{22}t^2 \\ &= \Theta_{11}s^2 - \Theta_{11}t^2,\end{aligned}$$

for which the set $\Theta^{-1}(0) \subset V$ is a pair of straight lines making angles of $\pi/4$ with the lines spanned by X_1 and X_2 (we measure angles in V by using the inner product on V corresponding to Φ, so that X_1 and X_2 are orthonormal).

We can also express the apolarity of Φ and Θ in a way that does not involve the special choice of basis: Φ and Θ are apolar if and only if $\Theta^{-1}(0)$ is the union of two straight lines which are perpendicular to each other in the inner product on V corresponding to Φ.

Now consider a quadratic function Φ, which we will still assume is non-degenerate, and a cubic function Ψ. Instead of constructing a number from Φ and Ψ, we will construct an element of V^*, namely

$$X_i \mapsto \sum_{j,k=1}^{n} \Phi^{jk}\Psi_{ijk}.$$

Suppose we have a new basis $\bar{X}_j = \sum_i a_{ij} X_i$, and that we let $B = (b_{ij})$ be the inverse matrix of $A = (a_{ij})$, as before. Consider the map

$$\bar{X}_\alpha \mapsto \sum_{\beta,\gamma=1}^{n} \bar{\Phi}^{\beta\gamma}\bar{\Psi}_{\alpha\beta\gamma}.$$

This map takes $X_i = \sum_\alpha b_{\alpha i}\bar{X}_\alpha$ to

$$\sum_{\alpha=1}^{n} b_{\alpha i} \sum_{\beta,\gamma=1}^{n} \bar{\Phi}^{\beta\gamma}\bar{\Psi}_{\alpha\beta\gamma} = \sum_{\alpha=1}^{n} b_{\alpha i} \sum_{\beta,\gamma=1}^{n} \sum_{j,k=1}^{n} \Phi^{jk} b_{\beta j} b_{\gamma k} \sum_{l,p,q=1}^{n} \Psi_{lpq} a_{l\alpha} a_{p\beta} a_{q\gamma}$$

by (5) and (6)

$$= \sum_{j,k=1}^{n} \Phi^{jk}\Psi_{ijk}.$$

Thus we have a well-defined map $(\Phi, \Psi)\colon V \to \mathbb{R}$, which for any basis $\{X_i\}$ is given by

$$(\Phi, \Psi)(X_i) = \sum_{j,k=1}^{n} \Phi^{jk}\Psi_{ijk};$$

industrious readers can supply their own invariant definition. As before, we say that Φ and Ψ are **apolar** if $(\Phi, \Psi) = 0$. Suppose that V is 2-dimensional, with a basis X_1, X_2. Since

$$\begin{pmatrix} \Phi^{11} & \Phi^{12} \\ \Phi^{21} & \Phi^{22} \end{pmatrix} = \frac{1}{\det[\Phi]} \begin{pmatrix} \Phi_{22} & -\Phi_{21} \\ -\Phi_{12} & \Phi_{11} \end{pmatrix},$$

and the Φ_{jk} and Ψ_{jkl} are symmetric in the indices, we see that Φ and Ψ are apolar if and only if they satisfy the **apolarity conditions**

$$(*) \qquad \begin{cases} \Phi_{22}\Psi_{111} - 2\Phi_{12}\Psi_{112} + \Phi_{11}\Psi_{122} = 0 \\ \Phi_{22}\Psi_{112} - 2\Phi_{12}\Psi_{122} + \Phi_{11}\Psi_{222} = 0. \end{cases}$$

Later on we will give a geometric interpretation of apolarity when V is 2-dimensional, but for the moment, we content ourselves with the observation that apolarity is clearly just about the simplest well-defined relationship that one can posit between a quadratic and a cubic function. It also happens to do the trick:

12. PROPOSITION. Let M be a surface in $\mathbb{R}^3$ and $p \in M$ a point which is elliptic or hyperbolic. For each tangent vector $X_3 \in \mathbb{R}^3{}_p$ which is not in M_p, define a quadratic function $\Phi\colon M_p \to \mathbb{R}$ and a cubic function $\Psi\colon M_p \to \mathbb{R}$ by looking at the second and third order terms in the Taylor series for the function which describes M in terms of the X_1, X_2, X_3 coordinate system, for any basis X_1, X_2 of M_p. Then there is a unique direction for X_3 which makes Φ and Ψ apolar.

PROOF. For simplicity we will assume that $p = 0 \in \mathbb{R}^3$ and that M_p is the (x, y)-plane. We can choose a basis X_1, X_2 for $\mathbb{R}^2$ so that if

$$(1) \qquad M = \{sX_1 + tX_2 + h(s,t) \cdot (0,0,1)\},$$

then h has the form

$$(2) \qquad h(s,t) = \frac{A}{2}(s^2 \pm t^2) + \frac{1}{6}(Bs^3 + 3Cs^2t + 3Dst^2 + Et^3) + R(s,t),$$

where $R(s,t)/|(s,t)|^3 \to 0$ as $(s,t) \to 0$.

Now consider the basis $X_1, X_2, X_3 = (0,0,1) + \lambda X_1 + \mu X_2$, for two constants λ, μ. Let k be the function describing M in the X_1, X_2, X_3 coordinate system, so that

$$(3) \qquad \begin{aligned} M &= \{sX_1 + tX_2 + k(s,t) \cdot [(0,0,1) + \lambda X_1 + \mu X_2]\} \\ &= \{[s + \lambda \cdot k(s,t)]X_1 + [t + \mu \cdot k(s,t)]X_2 + k(s,t) \cdot (0,0,1)\}. \end{aligned}$$

Comparing with (1), we see that

$$k(s,t) = h(s + \lambda \cdot k(s,t),\ t + \mu \cdot k(s,t)).$$

Using (2), and noting that we have $k(s,t)/|(s,t)| \to 0$, we find that

$$\begin{aligned} k(s.t) = \frac{A}{2}(s^2 \pm t^2) + \frac{1}{6}(Bs^3 + 3Cs^2t + 3Dst^2 + Et^3) \\ + Ak(s,t)(\lambda \cdot s \pm \mu \cdot t) + R'(s,t), \end{aligned}$$

where $R'(s,t)/|(s,t)|^3 \to 0$ as $(s,t) \to 0$. If we plug the whole right side of this equation into the term $k(s,t)$ on the right, we then find that

$$k(s,t) = \frac{A}{2}(s^2 \pm t^2) + \frac{1}{6}(Bs^3 + 3Cs^2t + 3Dst^2 + Et^3)$$
$$+ \frac{1}{2}(s^2 \pm t^2)(\lambda \cdot s \pm \mu \cdot t) + R''(s,t),$$

where $R''(s,t)/|(s,t)|^3 \to 0$. Therefore the Φ and Ψ for the basis X_1, X_2, X_3 are given by

$$\Phi(sX_1 + tX_2) = \frac{A}{2}(s^2 \pm t^2)$$
$$\Psi(sX_1 + tX_2) = \frac{1}{6}([B + 3\lambda]s^3 + 3[C \pm \mu]s^2t + 3[D \pm \lambda]st^2 + [E + 3\mu]t^3).$$

The apolarity conditions (∗) for this Φ and Ψ are

$$\pm(B + 3\lambda) + (D \pm \lambda) = 0$$
$$\pm(C \pm \mu) + (E + 3\mu) = 0.$$

There are clearly unique λ and μ for which these equations hold. It is also clear that if the apolarity conditions hold for some X_3, then they also hold for any multiple of X_3. So there is a unique direction for X_3 which will make Φ and Ψ apolar. ❖

For an elliptic or hyperbolic point $p \in M$, the unique direction through p which is given by Proposition 12 is called the **affine normal direction** at p. It is clearly a general affine invariant: If $A\colon \mathbb{R}^3 \to \mathbb{R}^3$ is any affine map, then the affine normal direction through the point $A(p) \in A(M)$ is the image under A of the affine normal direction through $p \in M$.

The affine normal direction can also be characterized in terms of moving frames. For simplicity we first assume that $p = 0 \in \mathbb{R}^3$ and that $M_p =$ (x,y)-plane. If the vector v points along the affine normal direction through p, then M is the graph, in the $(1,0,0), (0,1,0), v$ coordinate system, of a function h satisfying the apolarity conditions (∗):

$$\left.\begin{aligned} 0 &= h_{22}h_{111} - 2h_{12}h_{112} + h_{11}h_{122} = \frac{\partial}{\partial s}\det h_{ij} \\ 0 &= h_{22}h_{112} - 2h_{12}h_{122} + h_{11}h_{222} = \frac{\partial}{\partial t}\det h_{ij} \end{aligned}\right\} \quad \text{at } (0,0).$$

We can write these two equations together simply as

(1) $$0 = d\det(h_{ij}) \qquad \text{at } (0,0).$$

Locally M is the image of the map $f\colon \mathbb{R}^2 \to \mathbb{R}^3$ defined by

$$f(s,t) = (s,t,0) + h(s,t)\cdot v.$$

First take the particularly simple moving frame

$$\mathring{X}_1 = f_1 = (1,0,0) + h_1 v, \qquad \mathring{X}_2 = f_2 = (0,1,0) + h_2 v, \qquad \mathring{X}_3 = v$$

and let $\mathring{\psi}^\alpha_\beta$ be the corresponding connection forms. Then

$$0\mathring{X}_1 + 0\mathring{X}_2 + h_{ik}\mathring{X}_3 = \nabla'_{\mathring{X}_k}\mathring{X}_i = \sum_{\alpha=1}^{3}\mathring{\psi}^\alpha_i(\mathring{X}_k)\cdot\mathring{X}_\alpha \qquad i,k = 1,2$$

$$0 = \nabla'_{\mathring{X}_k}\mathring{X}_3 = \sum_{\alpha=1}^{3}\mathring{\psi}^\alpha_3(\mathring{X}_k)\cdot\mathring{X}_\alpha \qquad k = 1,2,$$

which implies, among other things, that

(2) $$\begin{aligned} \mathring{\psi}^j_i &= 0 \qquad \text{on } TM \qquad i,j = 1,2 \\ \mathring{\psi}^\alpha_3 &= 0 \qquad \text{on } TM \qquad \alpha = 1,2,3. \end{aligned}$$

Now consider an arbitrary adapted moving frame X_1, X_2, X_3 subject only to the condition that

$$X_3 = \mathring{X}_3 \qquad \text{at } p.$$

This condition implies that the matrix a on page 80 relating the two frames also satisfies

(3) $$\begin{aligned} a_{i3}(p) &= 0 \qquad i = 1,2; \qquad a_{33}(p) = 1 \\ \implies (a^{-1})_{i3}(p) &= 0 \qquad i = 1,2; \qquad (a^{-1})_{33}(p) = 1. \end{aligned}$$

From the general equation (d) on page 81 we have

$$\begin{aligned} \psi^j_i &= (a^{-1}\,da)^j_i + (a^{-1}\mathring{\psi}a)^j_i \\ &= \sum_{\alpha=1}^{3}(a^{-1})_{j\alpha}\,da_{\alpha i} + \sum_{\alpha,\beta=1}^{3}(a^{-1})_{j\alpha}\mathring{\psi}^\alpha_\beta\, a_{\beta i} \qquad i,j = 1,2 \end{aligned}$$

$$\begin{aligned} \psi^3_3 &= (a^{-1}\,da)^3_3 + (a^{-1}\mathring{\psi}a)^3_3 \\ &= \sum_{\alpha=1}^{3}(a^{-1})_{3\alpha}\,da_{\alpha 3} + \sum_{\alpha,\beta=1}^{3}(a^{-1})_{3\alpha}\mathring{\psi}^\alpha_\beta\, a_{\beta 3}. \end{aligned}$$

Taking into account equations (2) and (3), as well as equation (b) on page 80, we obtain

$$\text{(4)} \qquad \psi_3^3 = da_{33} \qquad \text{on } TM, \text{ at } p$$

and

$$\psi_i^j = \sum_{k=1}^{2} (a^{-1})_{jk}\, da_{ki} \qquad \text{on } TM, \text{ at } p \qquad i,j = 1,2;$$

since our matrix a has the form

$$\begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \text{at } p,$$

we can write the latter equation as

$$\text{(5)} \qquad \psi_i^j = \sum_{k=1}^{2} (A^{-1})_{jk}\, dA_{ki} \qquad \text{on } TM, \text{ at } p \qquad i,j = 1,2,$$

where A is the 2×2 matrix $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$. But we can compute da_{33} in terms of the matrix S for the moving frame X_1, X_2, X_3. For equation (e) on page 81 relates S to $\mathring{S} = (h_{ij})$ by

$$\begin{aligned} a_{33} \cdot S = A^{\mathrm{t}} \mathring{S} A &\implies (a_{33})^2 \cdot (\det S) = (\det A)^2 \cdot (\det \mathring{S}) \\ &\implies 2 \log |a_{33}| + \log |\det S| = 2 \log |\det A| + \log |\det \mathring{S}|. \end{aligned}$$

Since $a_{33}(p) = 1$, at p we have

$$\begin{aligned} 2\, da_{33} + d \log |\det S| &= 2d \log |\det A| + d \log |\det \mathring{S}| \\ &= 2d \log |\det A|, \qquad \text{by (1).} \end{aligned}$$

Hence (4) becomes

$$\text{(6)} \qquad \psi_3^3 = d \log |\det A| - \frac{1}{2} d \log |\det S| \qquad \text{on } TM, \text{ at } p.$$

Now note that

$$\begin{aligned} d \log |\det A| &= \frac{d \det A}{\det A} = \frac{d(A_{11} A_{22} - A_{12} A_{21})}{\det A} \\ &= \frac{(A_{22}\, dA_{11} - A_{12}\, dA_{21}) + (-A_{21}\, dA_{12} + A_{11}\, dA_{22})}{\det A}. \end{aligned}$$

Since

$$A^{-1} = \frac{\begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}}{\det A},$$

this becomes

$$\begin{aligned} d \log |\det A| &= \sum_{k=1}^{2} (A^{-1})_{1k}\, dA_{k1} + \sum_{k=1}^{2} (A^{-1})_{2k}\, dA_{k2} \\ &= \psi_1^1 + \psi_2^2 \qquad \text{on } TM, \text{ at } p \qquad \text{by (5).} \end{aligned}$$

Substituting into (6) we see that

An adapted moving frame X_1, X_2, X_3 *on* $M \subset \mathbb{R}^3$ *has* $X_3(p)$ *pointing in the affine normal direction at* p *if and only if*

$$\psi_3^3 = \psi_1^1 + \psi_2^2 - \tfrac{1}{2} d \log |\det S| \qquad \textit{on } TM, \textit{ at } p.$$

From this fact we see, what is by no means *a priori* clear, that the condition $\psi_3^3 = \psi_1^1 + \psi_2^2 - \frac{1}{2} d \log |\det S|$ on TM at p *depends only on the direction of* X_3 *at* p. It is also possible to check this by a direct, quite unpleasant, calculation. If we had somehow independently observed this fact, we could have used this equation to *define* the affine normal direction. Of course, it is hard to see how one would ever be led to such an "observation", but there is at least a way to simplify the equation. We have already observed that the condition

$$S(p) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ or } \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad \text{or} \qquad \begin{cases} \psi_1^3 = \theta^1 \\ \psi_2^3 = \pm\theta^2 \end{cases}$$

depends only on the value of the moving frame X_1, X_2, X_3 at p, so it is reasonable to restrict our attention to frames satisfying this condition at all points. It is clear that

An adapted moving frame X_1, X_2, X_3 *on* $M \subset \mathbb{R}^3$ *with* $S = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ *or* $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ *everywhere, has* $X_3(p)$ *pointing in the affine normal direction at* p *if and only if*

$$\psi_3^3 = \psi_1^1 + \psi_2^2 \qquad \textit{on } TM, \textit{ at } p.$$

The direct verification that for such frames the condition $\psi_3^3 = \psi_1^1 + \psi_2^2$ on TM at p depends only on the direction of X_3 at p is a little easier, and gives a nice criterion for the direction of the affine normal; of course, as a definition it still leaves a lot to be desired in terms of motivation.

Now that we have picked out an affine normal direction which is a general affine invariant, it is a simple matter to define a special affine normal vector which is a special affine invariant. We define the **special affine normal** ν_p of M at p to be the unique vector $\nu_p \in \mathbb{R}^3{}_p$ such that

(1) ν_p points along the affine normal direction at p

(2) $\det(X_1, X_2, \nu_p) = 1$ for every positively oriented basis X_1, X_2 of M_p which is orthonormal for the metric $\langle\ ,\ \rangle_p$.

If $f\colon M \to \mathbb{R}^3$ is an immersion, we let $\mathcal{N}$ be the vector field along f such that $\mathcal{N}(p)_{f(p)} \in \mathbb{R}^3{}_{f(p)}$ is the special affine normal to $f(M)$ at $f(p)$. We will not try to derive a formula for $\mathcal{N}$ right away, since it will come out very naturally later on.

Suppose now that we have an adapted moving frame X_1, X_2, X_3 on M such that $\det(X_1, X_2, X_3) = 1$. Then for all tangent vectors X to M we have

$$\begin{aligned} 0 &= X(\det(X_1, X_2, X_3)) \\ &= \det(\nabla'_X X_1, X_2, X_3) + \det(X_1, \nabla'_X X_2, X_3) + \det(X_1, X_2, \nabla'_X X_3) \\ &= \det\left(\sum_{\alpha=1}^{3} \psi_1^\alpha(X)\cdot X_\alpha, X_2, X_3\right) + \cdots \\ &= \psi_1^1(X) + \psi_2^2(X) + \psi_3^3(X). \end{aligned}$$

Thus we have

$$\psi_1^1 + \psi_2^2 + \psi_3^3 = 0 \qquad \text{on } TM.$$

Suppose, moreover, that (X_1, X_2) is positively oriented and orthonormal with respect to the metric $\langle\ ,\ \rangle$ on M; in other words, suppose that $S = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ or $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ everywhere. We have seen that in this case $X_3(p)$ points in the affine normal direction if and only if

$$\psi_3^3 = \psi_1^1 + \psi_2^2 \qquad \text{on } TM, \text{ at } p.$$

Since we also have $\psi_1^1 + \psi_2^2 + \psi_3^3 = 0$ everywhere on TM, we find that

An adapted moving frame X_1, X_2, X_3 *on* $M \subset \mathbb{R}^3$ *with* (X_1, X_2) *positively oriented,* $\det(X_1, X_2, X_3) = 1$, *and* $S = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ *or* $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ *everywhere, has* $X_3(p) = \nu_p$ *if and only if*

$$\psi_3^3 = 0 \qquad \textit{on } TM, \textit{ at } p.$$

Now the condition $\psi_3^3(p) = 0$ on M_p means that

$$\nabla'_{X_p} X_3 \in M_p \qquad \text{for all } X_p \in M_p.$$

So we see, in particular, that we always have

$$\nabla'_{X_p} \mathfrak{v} \in M_p \qquad \text{for } X_p \in M_p,$$

just as we always have $\nabla'_{X_p} \nu \in M_p$ in ordinary surface theory. (Verifying this fact directly from the definition of $\mathfrak{v}$ involves a rather hideous computation, and in general all the formulas and computations for affine surface theory are considerably more complicated than those in ordinary surface theory; that is why we have brought in moving frames, with their attendant computational simplifications, so early in the game.) We will denote the map $X_p \mapsto \nabla'_{X_p} \mathfrak{v}$ from M_p to M_p by $d\mathfrak{v}\colon M_p \to M_p$. The reason for this notation is the same as in ordinary surface theory: since we have a vector $\mathfrak{v}_p \in \mathbb{R}^3{}_p$ for each $p \in M$, we have a map $\mathfrak{v}\colon M \to \mathbb{R}^3$, and $\nabla'_{X_p} \mathfrak{v}$ is the same as $[d\mathfrak{v}(X_p)]_p$. In the present case, the map $\mathfrak{v}$ doesn't go into any special subset of $\mathbb{R}^3$, but $\mathfrak{v}(M) \subset \mathbb{R}^3$ will be some surface, at least near the points where $d\mathfrak{v}\colon M_p \to M_p$ is non-singular. The relation $\nabla'_{X_p} \mathfrak{v} \in M_p$ tells us that the tangent plane M_p is parallel to the tangent plane $\mathfrak{v}(M)_{\mathfrak{v}(p)}$. So we could also denote $d\mathfrak{v}\colon M_p \to M_p$ by $\mathfrak{v}_*\colon M_p \to M_p$, as in the case of ordinary surface theory.

Before proceeding with the development of special affine surface theory, we pause briefly to describe the procedure usually found in those papers and books which present the theory totally from the moving frame point of view, and ignore the question of general affine invariants, like the affine normal direction. One works, first of all, only with adapted frames X_1, X_2, X_3 satisfying $\det(X_1, X_2, X_3) = 1$, and hence $\psi_1^1 + \psi_2^2 + \psi_3^3 = 0$ on TM. Equation (e) is first derived, except that now a_{33} must $= 1$; it shows that S is determined only up to a transformation $S \mapsto A^{\mathbf{t}} S A$. The canonical forms for symmetric matrices under this equivalence relation are precisely $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$; so one "normalizes" the frame by requiring that S be everywhere of this form. Then one further normalizes the frame by requiring that $\psi_3^3(p) = 0$ for all p, noting that the condition $\psi_3^3(p) = 0$ now uniquely determines $X_3(p)$; indeed if $\{X_\alpha\}$ and $\{X'_\alpha\}$ are two adapted frames with $\det(X_1, X_2, X_3) = \det(X'_1, X'_2, X'_3) = 1$ and $S = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ or $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ everywhere, so that

$$\begin{aligned} \psi_1^3 &= \theta^1 \\ \psi_2^3 &= \pm\theta^2, \end{aligned}$$

then the matrix a relating the X'_α to the X_α must be of the form

$$a = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{pmatrix}$$

and equation (d) on page 81 yields

$$\begin{aligned} \psi'^3_3 &= \psi^3_1 a_{13} + \psi^3_2 a_{23} + \psi^3_3 \\ &= a_{13}\theta^1 \pm a_{23}\theta^2 + \psi^3_3, \end{aligned}$$

so that $\psi^3_3(p) = \psi'^3_3(p) = 0 \implies a_{13}(p) = a_{23}(p) = 0 \implies X'_3(p) = X_3(p)$. The uniquely determined X_3 is now dubbed the special affine normal, and any basis X_1, X_2 with $S = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ or $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ is called orthonormal, thereby defining the special affine first fundamental form. These normalizations are usually carried out with hardly a word of motivation, as if they are so natural that any idiot would immediately think of doing them—in reality, of course, the authors already knew what results they wanted, since they were simply reformulating a classical theory.

Now that we have finally determined the special affine normal v_p at $p \in M$, we can imitate a basic construction from ordinary surface theory. First we introduce the unique projections

$$\begin{aligned} &\top \colon \mathbb{R}^3{}_p \to M_p \\ &\perp \colon \mathbb{R}^3{}_p \to \mathbb{R} \cdot v_p = \text{ all multiples of } v_p \end{aligned}$$

such that

$$X = \top X + \perp X \qquad \text{for all } X \in M_p.$$

Notice that $\top \colon \mathbb{R}^3{}_p \to M_p$ is generally different from the tangential projection $\top \colon \mathbb{R}^3{}_p \to M$ of ordinary surface theory. Given vector fields X, Y tangent along M, we would like to find $\top\nabla'_X Y$ and $\perp\nabla'_X Y$, where $\nabla'_X Y$ is the ordinary covariant differentiation in $\mathbb{R}^3$. Now in ordinary surface theory, $\perp\nabla'_X Y$ involved the second fundamental form II. Since II is so closely related to the affine first fundamental form I, we might expect I to be involved in $\perp\nabla'_X Y$. As a matter of fact,

13. PROPOSITION. If X, Y are tangent along $M \subset \mathbb{R}^3$, then

$$\perp\nabla'_X Y = \langle X, Y\rangle \cdot v = \mathrm{I}(X, Y) \cdot v.$$

PROOF. Note first that, as in the case of $\perp\nabla'_X Y$, we have

$$\perp\nabla'_{X_p} fY = \perp(X_p(f)\cdot Y_p + f(p)\cdot\nabla'_{X_p}Y) = f(p)\cdot\perp\nabla'_{X_p}Y,$$

so $\perp\nabla'_{X_p}Y$ actually depends only on the value of Y at p. Now let (X_1, X_2) be a positively oriented moving frame on M which is orthonormal for $\langle\ ,\ \rangle$; this means (taking the elliptic case for simplicity) that for the moving frame $(X_1, X_2, X_3) = (X_1, X_2, \nu)$ we have

$$\psi_i^3 = \theta^i \qquad \text{on } TM.$$

But then

$$\perp\nabla'_{X_p}X_i = \psi_i^3(X)\cdot\nu = \theta^i(X)\cdot\nu = \langle X, X_i\rangle\cdot\nu. \ ❖$$

Now let's consider $\top\nabla'_X Y$. In ordinary surface theory this is $\nabla_X Y$, where ∇ is the connection on M determined by the metric $i^*\langle\ ,\ \rangle$. In affine surface theory, we can consider the connection ∇ on M determined by the metric $\langle\ ,\ \rangle$ (this exists even if $\langle\ ,\ \rangle$ is not positive definite; compare pg. II.342). Now one can ask whether $\top\nabla'_X Y = \nabla_X Y$ for vector fields X and Y tangent along M. This simply isn't true, and we therefore define a map $\mathfrak{s}\colon M_p \times M_p \to M_p$ by

$$\top\nabla'_{X_p}Y = \nabla_{X_p}Y + \mathfrak{s}(X_p, Y_p),$$

where Y is a vector field tangent along M which extends Y_p. The notation $\mathfrak{s}(X_p, Y_p)$ is justified, since

$$\begin{aligned}\top\nabla'_{X_p} fY - \nabla_{X_p} fY &= \top(X_p(f)\cdot Y_p + f(p)\cdot\nabla'_{X_p}Y)\\ &\qquad - [X_p(f)\cdot Y_p + f(p)\cdot\nabla_{X_p}Y]\\ &= f(p)\cdot[\top\nabla'_{X_p}Y - \nabla_{X_p}Y].\end{aligned}$$

14. PROPOSITION. The tensor $\mathfrak{s}$ is symmetric.

PROOF. If X and Y are extensions of $X_p, Y_p \in M_p$ to vector fields which are tangent to M at all points of M, then

$$\begin{aligned}\mathfrak{s}(X_p, Y_p) - \mathfrak{s}(Y_p, X_p) &= \top(\nabla'_{X_p}Y - \nabla'_{Y_p}X) - (\nabla_{X_p}Y - \nabla_{Y_p}X)\\ &= [X, Y](p) - [X, Y](p) = 0. \ ❖\end{aligned}$$

We can now write the decomposition

$$\nabla'_X Y = \top\nabla'_X Y + \bot\nabla'_X Y$$

in the following form:

The Special Affine Gauss Formulas:
$\nabla'_X Y = \nabla_X Y + \mathfrak{s}(X,Y) + \langle X, Y\rangle \nu$
for vector fields X, Y tangent along M.

In ordinary surface theory we often found that the second fundamental form $\mathrm{II}(X,Y) = \langle s(X,Y), \nu\rangle$ was easier to work with than the map $s\colon M_p \times M_p \to M_p{}^{\perp}$ itself. Since $\mathfrak{s}$ now goes into M_p, we have to adopt a slightly different strategy. We define the **special affine second fundamental form** $\mathbb{II}$ **of** M to be the *tri*-linear map

$$\mathbb{II}(X,Y,Z) = \langle \mathfrak{s}(X,Y), Z\rangle.$$

Naturally, when $f\colon M \to \mathbb{R}^3$ is an immersion, we define the **special affine second fundamental form** $\mathbb{II}_f$ **of** f to be $f^*\mathbb{II}$, where $\mathbb{II}$ is the special affine second fundamental form of $f(M)$. For $f\colon U \to \mathbb{R}^3$ (with $U \subset \mathbb{R}^2$ open), we write the form $\mathbb{II}_f$ as

$$\mathbb{II}_f = \sum_{i,j,k=1}^{2} \ell_{ijk}\, ds^i \otimes ds^j \otimes ds^k.$$

Since the components of $\mathbb{II}_f$ have three indices, there is no possibility of confusing the ℓ_{ijk} with the l_{ij} (if things hadn't worked out like this, I think I would have committed suicide at this point). We also write

$$\mathfrak{s}(f_i, f_j) = \sum_{k=1}^{2} \ell_{ij}^k f_k,$$

so that

$$\begin{aligned}\ell_{ijk} = \langle \mathfrak{s}(f_i, f_j), f_k\rangle &= \left\langle \sum_{\rho=1}^{2} \ell_{ij}^{\rho} f_\rho, f_k \right\rangle \\ &= \sum_{\rho=1}^{2} g_{k\rho}\ell_{ij}^{\rho};\end{aligned}$$

equivalently,

$$\ell_{ij}^k = \sum_{\rho=1}^{2} g^{k\rho}\ell_{ij\rho}.$$

Suppose that (X_1, X_2) is a positively oriented moving frame on M which is orthonormal for $\langle\ ,\ \rangle$, and let $X_3 = \nu$. For the moment consider the case where all points of M are elliptic. We set

$$c_{ijk} = \mathrm{I\!I}(X_i, X_j, X_k) = \langle s(X_i, X_j), X_k\rangle,$$

so that we have

$$s(X, Y) = \sum_{k=1}^{2} c_{ijk}\theta^i(X)\theta^j(Y)\cdot X_k$$

$$c_{ijk} = c_{jik}.$$

We also let ω_i^j be the connection forms for the frame X_1, X_2 which are determined *by the metric* $\langle\ ,\ \rangle$; thus ω_i^j are the unique 1-forms on M satisfying

$$d\theta^i = -\sum_{j=1}^{2}\omega_j^i \wedge \theta^j$$

$$\omega_j^i = -\omega_i^j.$$

Since

$$\Upsilon\nabla'_X X_i = \nabla_X X_i + s(X_i, X),$$

we have

$$\psi_i^k(X) = \omega_i^k(X) + \sum_{j=1}^{2} c_{ijk}\theta^j(X)$$

and hence

$$(1) \qquad \psi_i^k = \omega_i^k + \sum_{j=1}^{2} c_{ijk}\theta^j.$$

Since $\omega_i^k = -\omega_k^i$, we obtain

$$(2) \qquad \psi_i^k + \psi_k^i = \sum_{j=1}^{2}(c_{ijk} + c_{kji})\theta^j.$$

(As usual, these formulas are understood as formulas on TM). From this we immediately deduce

15. PROPOSITION. The tensor $\mathfrak{s}$ satisfies the "apolarity condition"

$$\operatorname{trace}\big(X \mapsto \mathfrak{s}(X,Y)\big) = 0.$$

In terms of a map $f: U \to \mathbb{R}^3$ we have

$$0 = \sum_{i,k=1}^{2} g^{ik} \ell_{ijk} = \sum_{i,k=1}^{2} g^{ik} \ell_{jik};$$

in other words, the quadratic form determined by $\langle\ ,\ \rangle_p$ is apolar to the cubic form determined by $\mathrm{I\!I}(p)$.

PROOF. For our moving frame (X_1, X_2, X_3) we have $\psi_1^1 + \psi_2^2 + \psi_3^3 = 0$, and also $\psi_3^3 = 0$, so $\psi_1^1 + \psi_2^2 = 0$. Using equation (2), this gives

$$0 = \sum_{j=1}^{2}\left(\sum_{i=1}^{2} c_{iji}\right)\theta^j \implies 0 = \sum_{i=1}^{2} c_{iji}$$
$$= \sum_{i=1}^{2} \langle \mathfrak{s}(X_i, X_j), X_i \rangle$$
$$= \operatorname{trace}\big(X \mapsto \mathfrak{s}(X, X_j)\big).$$

Similar, but slightly more involved, computations give the same result in the hyperbolic case. The second part of the Proposition is merely a restatement of the first, as we easily compute by using Fact 0. ❖

By bringing in the structural equations, we obtain one other important piece of information.

16. PROPOSITION. The tensor $\mathrm{I\!I}$ is symmetric in all three arguments; equivalently,

$$\langle \mathfrak{s}(X,Y), Z \rangle = \langle \mathfrak{s}(X,Z), Y \rangle.$$

In terms of a map $f: U \to \mathbb{R}^3$ we have

$$\ell_{ijk} = \ell_{ikj}.$$

PROOF. Again we consider only the elliptic case, and leave the hyperbolic case to the reader. From equation (2) on page 106 we obtain

$$(1) \qquad \sum_{k=1}^{2} \psi_i^k \wedge \theta^k + \sum_{k=1}^{2} \psi_k^i \wedge \theta^k = \sum_{j,k=1}^{2} (c_{ijk} + c_{kji})\theta^j \wedge \theta^k.$$

Now

$$\sum_{k=1}^{2} \psi_k^i \wedge \theta^k = -d\theta^i \tag{2}$$

by the first structural equation. But we also have $\psi_i^3 = \theta^i$, so we also get

$$d\theta^i = d\psi_i^3 = -\sum_{k=1}^{2} \psi_k^3 \wedge \psi_i^k \qquad (\text{since } \psi_3^3 = 0) \tag{3}$$

$$= -\sum_{k=1}^{2} \theta^k \wedge \psi_i^k = \sum_{k=1}^{2} \psi_i^k \wedge \theta^k.$$

From (1), (2), (3) we have

$$0 = \sum_{j,k=1}^{2} (c_{ijk} + c_{kji})\theta^j \wedge \theta^k,$$

so for each i, j, k we have

$$0 = c_{ijk} + c_{kji} - c_{ikj} - c_{jki}$$
$$= c_{ijk} - c_{ikj}. \quad ❖$$

From the apolarity conditions we can now easily obtain an explicit formula for the vector field $\mathcal{N}$ along a map $f\colon U \to \mathbb{R}^3$. We write the special affine Gauss formulas as

$$f_{ij} = \nabla_{f_i} f_j + s(f_i, f_j) + g_{ij}\mathcal{N}$$
$$= \sum_{k=1}^{2} \Gamma_{ij}^k f_k + \sum_{k=1}^{2} \ell_{ij}^k f_k + g_{ij}\mathcal{N},$$

where the Γ_{ij}^k are the Christoffel symbols for the metric $\langle\ ,\ \rangle$, and hence computable in terms of the g_{ij}. Now

$$\sum_{i,j=1}^{2} g^{ij} g_{ij} = 2,$$

so we have

$$\begin{aligned}\mathcal{N} &= \frac{1}{2}\sum_{i,j=1}^{2} g^{ij}\left(f_{ij} - \sum_{k=1}^{2}\Gamma_{ij}^{k} f_k - \sum_{k=1}^{2}\ell_{ij}^{k} f_k\right)\\ &= \frac{1}{2}\sum_{i,j=1}^{2} g^{ij}\left(f_{ij} - \sum_{k=1}^{2}\Gamma_{ij}^{k} f_k\right) - \frac{1}{2}\sum_{i,j=1}^{2}\sum_{k,\rho=1}^{2} g^{ij}g^{k\rho}\ell_{ij\rho}\\ &= \frac{1}{2}\sum_{i,j=1}^{2} g^{ij}\left(f_{ij} - \sum_{k=1}^{2}\Gamma_{ij}^{k} f_k\right),\end{aligned}$$

using Proposition 15 and the fact that $\ell_{ij\rho} = \ell_{\rho ij}$, by Proposition 16. This equation means that each component $\mathcal{N}^{\alpha}$ of $\mathcal{N}$ is given by

$$\text{(II)} \qquad \mathcal{N}^{\alpha} = \frac{1}{2}\sum_{i,j=1}^{2} g^{ij}\left({f^{\alpha}}_{ij} - \sum_{k=1}^{2}\Gamma_{ij}^{k} {f^{\alpha}}_{k}\right) \qquad \alpha = 1,2,3.$$

Notice that the partial derivatives ${f^{\alpha}}_i$ are the components of the vector field df^{α} on U. So the tensor $\nabla(df^{\alpha})$ of type $\binom{2}{0}$ has components (pg. II.210)

$${f^{\alpha}}_{i;j} = {f^{\alpha}}_{ij} - \sum_{k=1}^{2}\Gamma_{ij}^{k} {f^{\alpha}}_{k}.$$

[In Chapter 1 we wrote $f_{;i}$ instead of f_i for $\partial f/\partial x^i$, but when we are dealing with the standard coordinate system on $\mathbb{R}^2$ we will revert to the standard subscript notation for partial derivatives; we use ; rather than ; to emphasize that we are using the covariant derivative ∇.] We can therefore also write

$$\text{(II}'\text{)} \qquad \mathcal{N}^{\alpha} = \frac{1}{2}\sum_{i,j=1}^{2} g^{ij} {f^{\alpha}}_{i;j}.$$

Our formula *is* rather complicated, but it involves only quantities computable in terms of the g_{ij}'s, and hence ultimately in terms of f; in Addendum 1 to Chapter 7 we will have a little more to say about it.

With our present proof of Proposition 15, the fact that our fundamental forms I, II satisfy the same apolarity conditions which we used to *define* II works out as some sort of miracle. So it will perhaps be reassuring to see this fact demonstrated by a computation, which will also be useful later. We will assume that $p = 0 \in \mathbb{R}^3$, the tangent plane M_p is the (x, y)-plane, *and the z-axis is the affine* *normal direction at p*. Then M is the image of an immersion $f(s,t) = (s, t, h(s,t))$ with $h(0,0) = h_1(0,0) = h_2(0,0) = 0$, and h satisfies the apolarity conditions

$$(1) \qquad 0 = d(\det h_{ij}) \qquad \text{at } (0,0).$$

Now $\nabla'_{f_i} f_j = (0, 0, h_{ij})$, and the tangential projection $\mathbf{T}$ at 0 is just the ordinary projection on the (x, y)-plane, so $\mathbf{T}\nabla'_{f_i} f_j = 0$ at p. Consequently, $\mathfrak{s}(f_i, f_j) = -\nabla_{f_i} f_j$ at $(0,0)$, which gives

$$(2) \qquad \begin{aligned} \ell_{ijk} &= \langle \mathfrak{s}(f_i, f_j), f_k \rangle = -\langle \nabla_{f_i} f_j, f_k \rangle && \text{at } (0,0) \\ &= -\sum_{\rho=1}^{2} \Gamma_{ij}^{\rho} \langle f_\rho, f_k \rangle = -\sum_{\rho=1}^{2} g_{\rho k} \Gamma_{ij}^{\rho} && \text{at } (0,0) \\ &= -[ij, k] && \text{at } (0,0) \end{aligned}$$

where $[ij, k]$ are Christoffel symbols for I_f. To compute them, we note that

$$\begin{aligned} \frac{\partial g_{ij}}{\partial s^k} &= \frac{\partial}{\partial s^k}[d_{ij} \cdot (\det(d_{ij}))^{-1/4}] && \text{at } (0,0) \\ &= (\det(d_{ij}))^{-1/4} \frac{\partial d_{ij}}{\partial s^k} && \text{at } (0,0) \text{ by (1)} \\ &= (\det(d_{ij}))^{-1/4} \frac{\partial}{\partial s^k} \det(f_1, f_2, f_{ij}) && \text{at } (0,0) \\ &= (\det(d_{ij}))^{-1/4}[\det(f_{1k}, f_2, f_{ij}) + \det(f_1, f_{2k}, f_{ij}) \\ &\qquad + \det(f_1, f_2, f_{ijk})] && \text{at } (0,0) \\ &= (\det(d_{ij}(0,0)))^{-1/4}[0 + 0 + h_{ijk}(0,0)], \end{aligned}$$

since each f_{ij} has its first 2 components equal 0 at $(0,0)$. Thus

$$\begin{aligned} [ij, k](0,0) &= \frac{1}{2}\left(\frac{\partial g_{ik}}{\partial s^j} + \frac{\partial g_{jk}}{\partial s^i} - \frac{\partial g_{ij}}{\partial s^k}\right)(0,0) \\ &= \frac{(h_{ikj} + h_{jki} - h_{ijk})}{2 \cdot \sqrt[4]{\det(d_{ij})}}(0,0), \end{aligned}$$

and (2) becomes

$$(3) \qquad \ell_{ijk}(0,0) = -\frac{h_{ijk}(0,0)}{2 \cdot \sqrt[4]{\det(d_{ij}(0,0))}}.$$

This demonstrates Propositions 15 and 16 (since we can always make the z-axis the direction of the affine normal direction by a suitable special affine linear map, and the assertions in question are invariant under such maps).

In ordinary surface theory we used $\mathrm{I}(p)$ and $\mathrm{II}(p)$ to define two numerical invariants, the principal curvatures $k_1(p), k_2(p)$ [or equivalently $K(p)$ and $H(p)$]. These invariants arise quite naturally from an algebraic point of view, as the invariants of a 2×2 symmetric matrix S under the map $S \mapsto A^{-1}SA$, for 2×2 orthogonal matrices A. Geometrically, these invariants arise when we describe the osculating paraboloid of M (using the ordinary normal ν_p as the third axis)—for $p \in M \subset \mathbb{R}^3$ and $p' \in M' \subset \mathbb{R}^3$ we have $\{k_1(p), k_2(p)\} = \{k_1(p'), k_2(p')\}$ if and only if there is a special orthogonal affine map $A: \mathbb{R}^3 \to \mathbb{R}^3$ which takes the osculating paraboloid P at p onto the osculating paraboloid P' at p'. In special affine surface theory we now want to use $\mathfrak{I}(p)$ and $\mathfrak{II}(p)$ to determine numerical invariants. Algebraically, we are asking for invariants of a cubic form Ψ under the map $\Psi \mapsto \bar{\Psi}$, where

$$\bar{\Psi}_{\alpha\beta\gamma} = \sum_{j,k,l=1}^{2} \Psi_{jkl} a_{j\alpha} a_{k\beta} a_{l\gamma}$$

for 2×2 matrices A which preserve a certain quadratic form Φ. In the olden days when mighty invariant theory held sway over most of the domains of mathematics, this question was as natural to consider as the first, and one could probably preface the answer with the standard refrain "As every undergraduate knows ... ". Nowadays, of course, not even graduate students know, or care, what the invariants for cubic forms are. Instead of describing them algebraically, we pass immediately to the (equivalent) geometric problem. For each $p \in M$ we have the function $\Phi + \Psi: M_p \to \mathbb{R}$ obtained by looking at the second and third order terms in the Taylor series for the function which describes M in the X_1, X_2, ν_p coordinate system, for any basis X_1, X_2 of M_p (which one doesn't matter). We will call the graph S of $\Phi + \Psi$ in the X_1, X_2, ν_p coordinate system the **osculating cubic** at p, and we want to classify these osculating cubics up to special linear affine maps.

17. PROPOSITION. Let S be the osculating cubic at a point p of a surface $M \subset \mathbb{R}^3$. If p is an elliptic point, then there is a special linear affine map $A: \mathbb{R}^3 \to \mathbb{R}^3$ such that $A(S) \subset \mathbb{R}^3$ is the graph (in the ordinary coordinate system for $\mathbb{R}^3$) of the function

(a) $$(s,t) \mapsto \frac{1}{2}(s^2 + t^2) + \frac{C}{6}(s^3 - 3st^2), \qquad \text{for some } C.$$

If p is a hyperbolic point, then we can choose A so that $A(S)$ is the graph of one of the three functions

(bl) $$(s,t) \mapsto \frac{1}{2}(s^2 - t^2) + \frac{C}{6}(s^3 + 3st^2)$$

(b2) $$(s,t) \mapsto \frac{1}{2}(s^2 - t^2) + \frac{C}{6}(t^3 + 3ts^2)$$

(c) $$(s,t) \mapsto \frac{1}{2}(s^2 - t^2) + \frac{1}{6}(s+t)^3;$$

we can also choose A so that $A(S)$ is the graph of one of the two functions

(b′) $$(s,t) \mapsto st + \frac{D}{6}(s^3 + t^3)$$

(c′) $$(s,t) \mapsto st + \frac{1}{6}s^3.$$

Remark: Changing from (bl) to (b2) involves interchanging the first two axes of $\mathbb{R}^3$ while leaving the third axis fixed, so we need both forms unless we allow maps $A\colon \mathbb{R}^3 \to \mathbb{R}^3$ of the form $A = T \circ B$ with T a translation and $\det B = \pm 1$. Alternatively, we can make do with only one of these forms if we allow ourselves to change the orientation of M, and hence the direction of ν_p.

PROOF. We might as well assume that $p = 0 \in \mathbb{R}^3$, the tangent plane M_p is the (x,y)-plane, the affine normal ν_p is $(0,0,1)$, and the vectors $(1,0)$ and $(0,1)$ are an orthonormal basis of M_p; for it is clear that the image of M under a suitable special linear affine map will have this property. Then M is the image of an immersion $f(s,t) = (s,t,h(s,t))$ with $h(0,0) = h_1(0,0) = h_2(0,0) = 0$; since the z-axis is the affine normal direction, the apolarity conditions become

(∗) $$\begin{cases} h_{22}h_{111} - 2h_{12}h_{112} + h_{11}h_{122} = 0 \\ h_{22}h_{112} - 2h_{12}h_{122} + h_{11}h_{222} = 0 \end{cases} \quad \text{at } (0,0).$$

Suppose that p is an elliptic point. The fact that $\nu_p = (0,0,1)$ and the basis $(1,0,0)$, $(0,1,0)$ is orthonormal, means that $h_{ij}(0,0) = \delta_{ij}$. So (∗) becomes

$$h_{111} + h_{122} = 0 \qquad \text{and} \qquad h_{112} + h_{222} = 0 \quad \text{at } (0,0),$$

and S is the graph of a function of the form

$$(s,t) \mapsto \frac{1}{2}(s^2 + t^2) + \frac{1}{6}(as^3 + 3bs^2t - 3ast^2 - bt^3).$$

Now suppose we apply one more special linear (in fact, special orthogonal) transformation

(1) $$(\sigma, \tau, z) \mapsto (\sigma \cos\theta - \tau \sin\theta,\ \sigma \sin\theta + \tau \cos\theta,\ z).$$

The image of S under this map is the graph of

$$(s,t) \mapsto \frac{1}{2}([s\cos\theta - t\sin\theta]^2 + [s\sin\theta + t\cos\theta]^2) + \frac{1}{6}(a[s\cos\theta - t\sin\theta]^3 + \cdots),$$

which works out to be

(2) $$(s,t) \mapsto \frac{1}{2}(s^2+t^2) + \frac{1}{6}(a^*s^3 + 3b^*s^2t - 3a^*st^2 - b^*t^3),$$

where

(3) $$\begin{cases} a^* = a\cos^3\theta + 3b\cos^2\theta\sin\theta - 3a\cos\theta\sin^2\theta - b\sin^3\theta \\ b^* = b\cos^3\theta - 3a\cos^2\theta\sin\theta - 3b\cos\theta\sin^2\theta + a\sin^3\theta. \end{cases}$$

To obtain the desired form (a), we just have to choose θ so that $b^* = 0$, which can be done by choosing θ so that $\cot\theta$ is a solution of the equation

$$b(\cot\theta)^3 - 3a(\cot\theta)^2 - 3b(\cot\theta) + a = 0.$$

Now suppose that p is a hyperbolic point. Then $h_{11}(0,0) = -h_{22}(0,0) = 1$ and $h_{12}(0,0) = 0$. So $(*)$ becomes

$$h_{122} = h_{111} \qquad \text{and} \qquad h_{222} = h_{112} \quad \text{at } (0,0),$$

and S is the graph of a function of the form

(4) $$(s,t) \mapsto \frac{1}{2}(s^2-t^2) + \frac{1}{6}(as^3 + 3bs^2t + 3ast^2 + bt^3).$$

We might as well assume that $a, b \neq 0$, for otherwise we already have the form (b1) or (b2). We apply one more special linear transformation

(5) $$(\sigma, \tau, z) \mapsto (\sigma\cosh u + \tau\sinh u,\ \sigma\sinh u + \tau\cosh u,\ z).$$

The image of S under this map works out to be the graph of

$$(s,t) \mapsto \frac{1}{2}(s^2-t^2) + \frac{1}{6}(a^*s^3 + 3b^*s^2t + 3a^*st^2 + b^*t^3),$$

where

$$(6)\qquad \begin{cases} a^* = a\cosh^3 u + 3b\cosh^2 u \sinh u + 3a\cosh u \sinh^2 u + b\sinh^3 u \\ b^* = b\cosh^3 u + 3a\cosh^2 u \sinh u + 3b\cosh u \sinh^2 u + a\sinh^3 u. \end{cases}$$

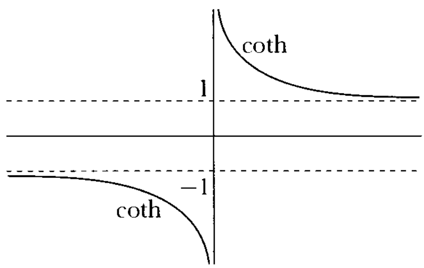
To obtain the form (b1), we have to find u so that $v = \coth u = (\cosh u)/(\sinh u)$ satisfies the equation

$$(7)\qquad bv^3 + 3av^2 + 3bv + a = 0.$$

Now coth does not take on all values, for the odd function

$$\coth u = \frac{\cosh u}{\sinh u} = \frac{e^u + e^{-u}}{e^u - e^{-u}} = \frac{e^{2u}+1}{e^{2u}-1}$$

is clearly > 1 for $u > 0$, and hence < -1 for $u < 0$. It is easy to see that coth actually takes on all values except those in $[-1, 1]$. So we can obtain the



form (b1) if equation (7) has a real root which is not in $[-1, 1]$. This certainly happens if the values of $bv^3 + 3av^2 + 3bv + a$ are either both positive or both negative at $v = 1$ and $v = -1$. So we can obtain (b1) if

$$(8)\qquad 4a + 4b \text{ and } 4a - 4b \text{ are both } > 0 \text{ or both } < 0.$$

But similarly, we can obtain (b2) if the values of $av^3 + 3bv^2 + 3av + b$ are both positive or both negative at $v = 1$ and $v = -1$, i.e., if

$$(9)\qquad 4a + 4b \text{ and } 4b - 4a \text{ are both } > 0 \text{ or both } < 0.$$

Clearly, either (8) or (9) must hold if $a \neq \pm b$. Thus (b1) or (b2) can be obtained when $a \neq \pm b$. We note right away that (b′) is obtained from (b1) by considering the special linear transformation

$$(10)\qquad (\sigma, \tau, z) \mapsto \left(\frac{\sigma+\tau}{\sqrt{2}}, \frac{-(\sigma-\tau)}{\sqrt{2}}, z\right),$$

the constant D being $-C\sqrt{2}$. Similarly, (b′) is obtained from (b2) by considering the special linear transformation

(11) $$(\sigma, \tau, z) \mapsto \left(\frac{\sigma - \tau}{\sqrt{2}}, \frac{\sigma + \tau}{\sqrt{2}}, z\right),$$

the constant D being $+C\sqrt{2}$.

When we have $a = \pm b$, then our original function (4) is

$$(s, t) \mapsto \frac{1}{2}(s^2 - t^2) + \frac{a}{6}(s \pm t)^3.$$

The transformation (11) or (10) will change this to

$$(s, t) \mapsto st + \frac{a'}{6}s^3.$$

If $a' = 0$, this is a special case of (b′). Otherwise, we can make $a' = 1$, and thus achieve (c′), by a transformation which changes s to $s/\sqrt[3]{a'}$, and t to $\sqrt[3]{a'}t$. Since we can achieve (c′), we can now achieve (c) by using the inverse of the transformation (11). ❖

Notice that in case (a), the cubic part of S is just (a multiple of) the monkey saddle. Implicit in the previous Proposition is a geometric interpretation of apolarity when $\Phi\colon V \to \mathbb{R}$ corresponds to a positive definite inner product on a 2-dimensional vector space V. The cubic form $\Psi\colon V \to \mathbb{R}$ is apolar to Φ when the set $\Psi^{-1}(0)$ consists of three lines that make angles of $\pi/3$ with each other (in the inner product on V corresponding to Φ). It is clear that if we apply a rotation around the z-axis through an angle of $\pi/3$ or $2\pi/3$, then $A(S)$ will go into itself, so in case (a) the affine linear map A is not unique. We can also change C to $-C$ in (a) by rotation through an angle of π, which changes s to $-s$ and t to $-t$; the same change occurs if we rotate through an angle of $\pi + \pi/3$ or $\pi + 2\pi/3$. If $C = 0$, then we can apply any rotation around the z-axis. However, from equations (3) it is easy to see that this is the only extent to which A and C are not unique.

In the hyperbolic case, if we can obtain (b1) or (b2) with one constant $C \neq 0$, then we can also achieve this form with $-C$ by applying the rotation $(\sigma, \tau, z) \mapsto (-\sigma, -\tau, z)$. On the other hand, it is easy to see that this is the only other constant we can obtain, and that there is precisely one A which will make $A(S)$ have each of these forms. When $C = 0$, then we can always compose A with any transformation of the form

$$(\sigma, \tau, z) \mapsto (\pm[\sigma \cosh u + \tau \sinh u],\ \pm[\sigma \sinh u + \tau \cosh u],\ z).$$

Bearing these remarks in mind, we see that Proposition 17 allows us to define a single new invariant: If p is a point of M for which the osculating cubic S has the form (a), (b1), or (b2) for some number C (unique up to sign), we define the **Pick invariant** J at p by

$$J = \frac{C^2}{2};$$

if the osculating cubic has the form (c), we define J to be 0.

It would be nice to have a straightforward invariant description of J in terms of the forms I, II, and it can be obtained as follows. Given an inner product $\langle\ ,\ \rangle$ on a vector space V, we can use it to define an inner product $\langle\!\langle\ ,\ \rangle\!\rangle$ on the vector space of all trilinear maps $\alpha\colon V \times V \times V \to \mathbb{R}$. Instead of describing this completely invariantly, let us consider a basis $v_1, \ldots, v_n$ of V and let $\gamma_{ij} = \langle v_i, v_j \rangle$. The matrix (γ_{ij}) is non-singular, and has an inverse matrix (γ^{ij}). Given two trilinear maps $\alpha, \beta\colon V \times V \times V \to \mathbb{R}$, let $\alpha_{ijk} = \alpha(v_i, v_j, v_k)$ and $\beta_{ijk} = \beta(v_i, v_j, v_k)$. Then

$$\langle\!\langle \alpha, \beta \rangle\!\rangle = \sum_{i,j,k=1}^{n} \sum_{\rho,\sigma,\tau=1}^{n} \alpha_{ijk}\beta_{\rho\sigma\tau}\gamma^{i\rho}\gamma^{j\sigma}\gamma^{k\tau}.$$

If $v_1, \ldots, v_n$ is orthonormal with respect to $\langle\ ,\ \rangle$, then we have simply

$$\langle\!\langle \alpha, \beta \rangle\!\rangle = \sum_{i,j,k=1}^{n} \alpha_{ijk}\beta_{ijk}.$$

The reader may easily check that this definition does not depend on the choice of basis, or else fashion an invariant definition.

18. PROPOSITION. For a point p of a surface $M \subset \mathbb{R}^3$, the Pick invariant $J(p)$ is

$$J(p) = \frac{1}{2}\langle\!\langle \mathrm{II}(p), \mathrm{II}(p) \rangle\!\rangle_p,$$

where $\langle\!\langle\ ,\ \rangle\!\rangle_p$ is the inner product on all trilinear maps on M_p determined by the inner product $\langle\ ,\ \rangle_p$ on M_p.

If $f\colon U \to \mathbb{R}^3$ is an immersion, then

$$J = \frac{1}{2} \sum_{i,j,k=1}^{2} \sum_{\rho,\sigma,\tau=1}^{2} \ell_{ijk}\ell_{\rho\sigma\tau}g^{i\rho}g^{j\sigma}g^{k\tau}.$$

PROOF. It suffices to consider the case where $M = \{(s, t, h(s, t))\}$, the point p is $0 \in \mathbb{R}^3$, and the osculating cubic S has one of the forms (a)–(c) in Proposition 17. In case (a) we thus have

$$\begin{aligned} h(0,0) &= h_i(0,0) = 0, \\ h_{ij}(0,0) &= \delta_{ij}, \\ h_{111}(0,0) &= C, \\ h_{211}(0,0) &= h_{121}(0,0) = h_{112}(0,0) = -C, \qquad \text{all other } h_{ijk}(0,0) = 0, \end{aligned}$$

with similar equations in cases (bl), (b2), (c). We can compute the $g_{ij}(0,0)$ from equation (I) on page 90 and the $\ell_{ijk}(0,0)$ from equation (3) on page 110. It is then a simple calculation to check that we do indeed have

$$\begin{aligned} J(p) &= \frac{1}{2} \sum_{i,j,k=1}^{2} \sum_{\rho,\sigma,\tau=1}^{2} \ell_{ijk} \ell_{\rho\sigma\tau} g^{i\rho} g^{j\sigma} g^{k\tau} \qquad \text{at } (0,0) \\ &= \frac{1}{2} \langle\!\langle \mathrm{I\!I}(p), \mathrm{I\!I}(p) \rangle\!\rangle_p. \end{aligned}$$

Since both sides of our formula have an invariant meaning, the formula must be true for any coordinate system. ❖

Notice also that for a moving frame X_1, X_2 as on page 106 we have

$$J = \frac{1}{2} \sum_{i,j,k} (c_{ijk})^2$$

in the elliptic case, with a similar formula in the hyperbolic case.

Now that we have obtained the invariant J, there is an obvious question staring us in the face: what are the surfaces with $J = 0$ everywhere? In ordinary surface theory we found that surfaces with $k_1 = k_2 = 0$ everywhere are planes. It seems natural to conjecture that the surfaces with $J = 0$ everywhere are just the surfaces which can be described by quadratic equations. Now this isn't true, and the reason is essentially because we have $J = 0$ in case (c) of Proposition 17. In cases (a), (bl), (b2), the vanishing of $J(p)$ implies the vanishing of $\mathrm{I\!I}(p)$ [as is also clear in case (a) from Proposition 18]. Leaving the complexities which arise because of case (c) to later Problems (4-14, 15), we restrict our attention to surfaces with $\mathrm{I\!I} = 0$ everywhere.

19. PROPOSITION. If $M \subset \mathbb{R}^3$ is a connected surface with all points elliptic or all points hyperbolic, and $\mathrm{II} = 0$ on M, then M is a quadratic surface, that is, $M = W^{-1}(0)$ for some $W: \mathbb{R}^3 \to \mathbb{R}$ of the form

$$W(x_1, x_2, x_3) = \sum_{i,j} a_{ij} x_i x_j + \sum_i b_i x_i + c.$$

(These surfaces are described in greater detail in the next chapter; at the moment it is only necessary to note that the affine linear image of a quadratic surface is also a quadratic surface.)

Conversely, every non-flat quadratic surface has $\mathrm{II} = 0$ everywhere.

PROOF. As usual, we consider only the elliptic case. Take a moving frame X_1, X_2, X_3 on M as on page 106. If $\mathrm{II} = 0$, then all $c_{ijk} = 0$, so equation (1) on page 106 shows that

$$\psi^i_j = \omega^i_j,$$

and in particular

$$\psi^i_j = -\psi^j_i.$$

This implies that

$$d\psi^i_j + \sum_{k=1}^{2} \psi^i_k \wedge \psi^k_j = -\left(d\psi^j_i + \sum_{k=1}^{2} \psi^j_k \wedge \psi^k_i\right).$$

But

$$(1) \qquad \begin{aligned} d\psi^i_j &= -\sum_{k=1}^{2} \psi^i_k \wedge \psi^k_j - \psi^i_3 \wedge \psi^3_j \\ &= -\sum_{k=1}^{2} \psi^i_k \wedge \psi^k_j - \psi^i_3 \wedge \theta^j, \end{aligned}$$

so we obtain

$$\psi^i_3 \wedge \theta^j = -(\psi^j_3 \wedge \theta^i).$$

Taking $i = j$, we see that ψ^i_3 is a multiple of θ^i, and taking $i \neq j$ we see that the multiples are the same for each i, so we have

$$(2) \qquad \psi^i_3 = \alpha \cdot \theta^i$$

for some function α. Taking d of equation (2) gives

$$-\sum_{k=1}^{2} \psi_k^i \wedge \psi_3^k = d\psi_3^i = d\alpha \wedge \theta^i + \alpha\, d\theta^i$$
$$\Downarrow$$
$$-\sum_{k=1}^{2} \psi_k^i \wedge \alpha \cdot \theta^k = d\alpha \wedge \theta^i - \alpha\left(\sum_{k=1}^{2} \psi_k^i \wedge \theta^k\right),$$

and hence

$$d\alpha \wedge \theta^i = 0.$$

Thus $d\alpha = 0$, and α is a constant. We consider two cases.

Case 1. $\alpha = 0$. Equation (2) shows that $\psi_3^i = 0$, so for tangent vectors X on M we have

$$\nabla'_X \nu = \nabla'_X X_3 = \sum_{i=1}^{2} \psi_3^i(X) \cdot X_i = 0,$$

i.e., ν is constant; we will use it as one of our axes. We also have, by (1),

$$d\psi_j^i = -\sum_{k=1}^{2} \psi_k^i \wedge \psi_j^k.$$

Since $\psi_j^i = \omega_j^i$, which are the connection forms for the metric $\langle\ ,\ \rangle$, this equation shows that $\langle\ ,\ \rangle$ is *flat*. So we could have chosen our orthonormal moving frame X_1, X_2 to be of the form $X_i = f_i$ for some isometry $f: U \to \mathbb{R}^3$, where $U \subset \mathbb{R}^2$ has the standard Riemannian metric; and with this choice we have $\psi_j^i = \omega_j^i = 0$. Now we have

$$f_{ij} = \nabla'_{f_j} f_i = \sum_{i=1}^{2} \psi_i^k(f_j) \cdot f_k + \psi_i^3(f_j) \cdot \nu$$
$$= 0 + \theta^i(f_j)\nu = \delta_{ij}\nu.$$

So f is of the form

$$f(s,t) = c + b_1 s + b_2 t + \frac{1}{2}(s^2 + t^2)\nu$$

for constants b_1, b_2, c.

Case 2. $\alpha \neq 0$. Now equation (2) shows that

$$\nabla'_X v = \sum_{i=1}^{2} \alpha\theta^i(X) \cdot X_i = \alpha X.$$

So for any $f \colon U \to M \subset \mathbb{R}^3$ we have

$$\mathcal{N}_i = \alpha f_i \implies \mathcal{N} = \alpha f + \beta \implies f = \frac{1}{\alpha}(\mathcal{N} - \beta)$$

for some constant vector β. Hence it suffices to show that v satisfies a quadratic equation.

Let $\mathbf{X}$ be the 3×3 matrix

$$\mathbf{X} = (X_1, X_2, X_3) = (X_1, X_2, v)$$

where the X_i are considered as column vectors. Then $d\mathbf{X}$ is given in terms of the 3×3 matrix $\psi = (\psi^\alpha_\beta)$ as

(1) $$d\mathbf{X} = \mathbf{X} \cdot \psi.$$

Since ψ is actually

$$\psi = \begin{pmatrix} \psi^1_1 & \psi^1_2 & \alpha\theta^1 \\ \psi^2_1 & \psi^2_2 & \alpha\theta^2 \\ \theta^1 & \theta^2 & 0 \end{pmatrix}, \qquad \psi^i_j = -\psi^j_i,$$

we easily see that if we set

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{\alpha} \end{pmatrix},$$

then

(2) $$\psi \cdot A + A \cdot \psi^{\mathbf{t}} = 0.$$

Thus we have

$$\begin{aligned} d(\mathbf{X} \cdot A \cdot \mathbf{X}^{\mathbf{t}}) &= d\mathbf{X} \cdot A \cdot \mathbf{X}^{\mathbf{t}} + \mathbf{X} \cdot A \cdot d\mathbf{X}^{\mathbf{t}} \\ &= \mathbf{X} \cdot \psi \cdot A \cdot \mathbf{X}^{\mathbf{t}} + \mathbf{X} \cdot A \cdot \psi^{\mathbf{t}} \cdot \mathbf{X}^{\mathbf{t}} \qquad \text{by (1)} \\ &= 0 \qquad \text{by (2).} \end{aligned}$$

We can assume that $\mathbf{X}$ is the identity matrix at some point, so we obtain

$$\begin{aligned}\mathbf{X}\cdot A\cdot \mathbf{X}^{\mathbf{t}} = A \implies (A^{-1}\cdot \mathbf{X}\cdot A)\cdot \mathbf{X}^{\mathbf{t}} &= \text{identity}\\ \implies \mathbf{X}^{\mathbf{t}}\cdot (A^{-1}\cdot \mathbf{X}\cdot A) &= \text{identity}\\ \implies \mathbf{X}^{\mathbf{t}}\cdot A^{-1}\cdot \mathbf{X} = A^{-1} &= \begin{pmatrix} 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & -\alpha\end{pmatrix}.\end{aligned}$$

The $(3,3)$ entry of this equation gives

$$\nu\cdot A^{-1}\cdot \nu^{\mathbf{t}} = -\alpha,$$

which is a quadratic equation for ν.

It is not hard to show, conversely, that all surfaces of the sort we have obtained have $\mathrm{I\!I} = 0$. These are the only non-flat quadrics (see Problem 3-6). ❖

We proceed in our study of special affine surface theory along the very same route followed in ordinary surface theory, by looking for the equations expressing the derivatives of a moving frame X_1, X_2, ν in terms of this frame, and by looking at the integrability conditions for these equations. We will always work only with the elliptic case, leaving the hyperbolic case to the reader. We already have the analogue of the Gauss formulas,

$$\nabla'_X Y = \nabla_X Y + \mathfrak{s}(X,Y) + \langle X,Y\rangle \nu,$$

or in terms of a map $f\colon U \to \mathbb{R}^3$,

$$\begin{aligned} f_{ij} &= \nabla_{f_i} f_j + \mathfrak{s}(f_i, f_j) + g_{ij}\mathcal{N}\\ &= \sum_{k=1}^{2}\Gamma_{ij}^k f_k + \sum_{i=1}^{2}\ell_{ij}^k f_k + g_{ij}\mathcal{N};\end{aligned}$$

in terms of a moving frame (X_1, X_2, ν) with X_1, X_2 orthonormal for $\langle\ ,\ \rangle$ we have

$$\psi_i^j = \omega_i^j + \sum_{k=1}^{2} c_{ijk}\theta^k, \qquad \psi_i^3 = \theta^i.$$

The strange way in which the special affine Gauss formulas interchange the roles of $\top\nabla'_X Y$ and $\perp\nabla'_X Y$, as compared to the ordinary Gauss formulas, will be reflected by funny twists throughout the development.

The next thing we seek is an analogue of the Weingarten equations. Now we know that

$$\nabla'_{X_p} \mathfrak{v} \in M_p,$$

so for a map $f \colon U \to \mathbb{R}^3$ we always have

$$\mathcal{N}_i = \sum_{j=1}^{2} \mathscr{b}_i^j f_j$$

for certain functions $\mathscr{b}_i^j$ with

$$\langle n_i, f_j \rangle = \sum_{\rho=1}^{2} \mathscr{b}_i^\rho \mathscr{g}_{\rho j} = \mathscr{b}_{ij}, \qquad \text{say;}$$

for our moving frame we have

$$\nabla'_X \mathfrak{v} = \nabla'_X X_3 = \sum_{i=1}^{2} \psi_3^i(X) \cdot X_i.$$

Unlike the situation in ordinary surface theory however, it is not at all clear how, or even whether, the $\mathscr{b}_i^j$ and $\mathscr{b}_{ij}$ are related to the ℓ_{ijk}. This is answered in a very unexpected way when we look at the integrability conditions for the special affine Gauss equations. These can be derived from the formulas for f_{ij}, exactly as in the first part of this chapter, or from the formulas for $\nabla'_X Y$, as in Chapter 1. But since the computations are quite involved, no matter how one does them, it will be easiest to use the moving frame version. We begin with the tangential part of these equations,

$$\text{(1)} \qquad \psi_i^j = \omega_i^j + \sum_{k=1}^{2} c_{ijk} \theta^k.$$

We take d of this equation, remembering that $\psi_i^3 = \theta^i$, and introducing the curvature forms Ω_i^j for the ω_i^j, to obtain

$$\text{(2)} \qquad -\sum_{\rho=1}^{2} \psi_\rho^j \wedge \psi_i^\rho - \psi_3^j \wedge \theta^i = -\sum_{\rho=1}^{2} \omega_\rho^j \wedge \omega_i^\rho + \Omega_i^j + \sum_{k=1}^{2} dc_{ijk} \wedge \theta^k + \sum_{k=1}^{2} c_{ijk} d\theta^k.$$

Substituting back from (1) we have

$$-\sum_\rho \left\{\omega_\rho^j + \sum_k c_{\rho jk}\theta^k\right\} \wedge \left\{\omega_i^\rho + \sum_k c_{i\rho k}\theta^k\right\} - \psi_3^j \wedge \theta^i$$
$$= -\sum_\rho \omega_\rho^j \wedge \omega_i^\rho + \Omega_i^j + \sum_k dc_{ijk} \wedge \theta^k - \sum_{k,\rho} c_{ijk}\omega_\rho^k \wedge \theta^\rho.$$

Using $\omega_i^\rho = -\omega_\rho^i$ and switching dummy indices in the last term, we get

(3) $$-\sum_{\rho,k,l} c_{\rho jk}c_{i\rho l}\theta^k \wedge \theta^l - \psi_3^j \wedge \theta^i$$
$$= \Omega_i^j + \sum_k \left[dc_{ijk} - \sum_\rho c_{\rho jk}\omega_i^\rho - \sum_\rho c_{i\rho k}\omega_j^\rho - \sum_\rho c_{ij\rho}\omega_k^\rho\right] \wedge \theta^k.$$

To interpret equation (3), we apply it to two vectors, X, Y. The first term becomes

(3a) $$\sum_{\rho,k,l} c_{jk\rho}c_{il\rho}\theta^k(Y)\theta^l(X) - \sum_{\rho,k,l} c_{jkl}c_{il\rho}\theta^k(X)\theta^l(Y)$$
$$= \langle \mathfrak{s}(X_i, X), \mathfrak{s}(X_j, Y)\rangle - \langle \mathfrak{s}(X_i, Y), \mathfrak{s}(X_j, X)\rangle.$$

The second term becomes

(3b) $$\psi_3^j(Y)\theta^i(X) - \psi_3^j(X)\theta^i(Y) = \langle d\nu(Y), X_j\rangle \cdot \langle X, X_i\rangle - \langle d\nu(X), X_j\rangle \cdot \langle Y, X_i\rangle,$$

while the first term on the right becomes simply

(3c) $$\langle \mathcal{R}(X, Y)X_i, X_j\rangle,$$

where $\mathcal{R}$ is the curvature tensor for $\langle\ ,\ \rangle$. To interpret the last term, we recall that the tensor

$$\mathrm{I\!I} = \sum c_{ijk}\theta^i \otimes \theta^j \otimes \theta^k$$

has a covariant derivative $\nabla\mathrm{I\!I}(X, Y, Z, W) = (\nabla_W \mathrm{I\!I})(X, Y, Z)$, which can be written

$$\nabla\mathrm{I\!I} = \sum_{i,j,k,l} c_{ijk;l}\theta^i \otimes \theta^j \otimes \theta^k \otimes \theta^l, \qquad \text{say.}$$

Using Corollary I.6-5 one easily checks (Problem 5) that

$$\sum_l c_{ijk;l}\theta^l = dc_{ijk} - \sum_\rho c_{\rho jk}\,\omega_i^\rho - \sum_\rho c_{i\rho k}\,\omega_j^\rho - \sum_\rho c_{ij\rho}\,\omega_k^\rho.$$

So the last term is

$$\sum_{k,l} c_{ijk;l}\theta^l \wedge \theta^k,$$

which when applied to X, Y gives

$$\text{(3d)} \qquad \sum_{k,l} c_{ijk;l}\theta^l(X)\theta^k(Y) - \sum_{k,l} c_{ijk;l}\theta^l(Y)\theta^l(X)$$
$$= (\nabla_X \mathrm{II})(X_i, X_j, Y) - (\nabla_Y \mathrm{II})(X_i, X_j, X).$$

Writing (3a) + (3b) = (3c) + (3d), but replacing X_i, X_j by arbitrary tangent vectors Z, W, we thus obtain

$$\text{(A)} \quad \langle \mathcal{R}(X,Y)Z, W\rangle = (\nabla_Y \mathrm{II})(W,Z,X) - (\nabla_X \mathrm{II})(W,Z,Y)$$
$$+ \langle \mathfrak{s}(W,Y), \mathfrak{s}(X,Z)\rangle - \langle \mathfrak{s}(W,X), \mathfrak{s}(Y,Z)\rangle$$
$$+ \langle dv(Y), W\rangle \cdot \langle X, Z\rangle - \langle dv(X), W\rangle \cdot \langle Y, Z\rangle.$$

In terms of a map $f\colon U \to \mathbb{R}^3$ this becomes

$$\mathcal{R}_{jik\mu} = \ell_{jik;\mu} - \ell_{ji\mu;k} + \sum_\rho (\ell_{ik}^\rho \ell_{j\rho\mu} - \ell_{i\mu}^\rho \ell_{j\rho k})$$
$$+ g_{ik} b_{\mu j} - g_{i\mu} b_{kj}.$$

Now what do these equations tell us? In ordinary surface theory we obtained the Theorema Egregium, telling us that the intrinsic Gaussian curvature K is equal to some expression involving the coefficients of II. But now we don't get anything of the sort, because we have the unknown expressions* dv (or b_{ij}). Instead, equation (A) allows us to solve for $\langle dv(Y), W\rangle$—we just have to choose a unit vector Z with $\langle Y, Z\rangle = 0$, and set $X = Z$.

*Note also that the terms involving $\mathfrak{s}$ in equation (A) are the negatives of the corresponding terms involving s in the ordinary Gauss equations.

An especially nice formula for $d\nu$ can be obtained with a little more work. Recall first of all that for a map $f: U \to \mathbb{R}^3$ we have the apolarity conditions

$$0 = \sum_{i,j} g^{ij}\ell_{\mu ij} \implies 0 = \sum_{i,j}(g^{ij}\ell_{\mu ij})_{;k}$$
$$= \sum_{i,j} g^{ij}\ell_{\mu ij;k}, \qquad \text{by Ricci's Lemma.}$$

For orthonormal X_1, X_2 this means that

$$\sum_i (\nabla_X \mathrm{I\!I})(X_i, X_i, Y) = 0.$$

[Naturally, this formula can also be derived directly from the apolarity condition $\sum_i \mathrm{I\!I}(X_i, X_i, Y) = 0$, but the coordinate treatment is much easier and quicker.] So if $\mathcal{K}$ is the intrinsic Gaussian curvature for the metric $\langle\ ,\ \rangle$, then equation (A) gives

$$\begin{aligned}
-2\mathcal{K} &= \sum_{i,j}\langle \mathcal{R}(X_i, X_j)X_i, X_j\rangle \\
&= \sum_j\left\{\sum_i(\nabla_{X_j}\mathrm{I\!I})(X_j, X_i, X_i)\right\} - \sum_i\left\{\sum_j(\nabla_{X_i}\mathrm{I\!I})(X_j, X_i, X_i)\right\} \\
&\quad + \sum_{i,j}\langle \mathfrak{s}(X_j, X_j), \mathfrak{s}(X_i, X_i)\rangle - \sum_{i,j}\langle \mathfrak{s}(X_i, X_j), \mathfrak{s}(X_i, X_j)\rangle \\
&\quad + \sum_{i,j}\langle d\nu(X_j), X_j\rangle \cdot \delta_{ii} - \sum_{i,j}\langle d\nu(X_i), X_j\rangle \cdot \delta_{ij} \\
&= 0 - 0 + \sum_{i,j,k} c_{jjk}c_{iik} - \sum_{i,j,k}(c_{ijk})^2 \\
&\quad + 2\sum_j\langle d\nu(X_j), X_j\rangle - \sum_i\langle d\nu(X_i), X_i\rangle \\
&= \sum_{j,k} c_{jjk}\left(\sum_i c_{iik}\right) - \sum_{i,j,k}(c_{ijk})^2 + \sum_j\langle d\nu(X_j), X_j\rangle.
\end{aligned}$$

Now $\sum_i c_{iik} = 0$ by the apolarity conditions, so we have (see page 117)

(1) $$-2\mathcal{K} = -2J + \sum_j\langle d\nu(X_j), X_j\rangle.$$

where J is the Pick invariant. But as another consequence of equation (A), and the symmetry properties of $\mathcal{R}$, we also have

$$
\begin{aligned}
(2) \qquad 0 &= \sum_i \langle \mathcal{R}(X_i, X)X_i, Y\rangle + \langle \mathcal{R}(X_i, X)Y, X_i\rangle \\
&= 0 - 2\sum_i (\nabla_{X_i} \mathrm{I\!I})(X, Y, X_i) \\
&\qquad + \sum_i \langle d\nu(X), Y\rangle \cdot \delta_{ii} - \sum_i \langle d\nu(X_i), Y\rangle \cdot \langle X, X_i\rangle \\
&\qquad + \sum_i \langle d\nu(X), X_i\rangle \cdot \langle X_i, Y\rangle - \sum_i \langle d\nu(X_i), X_i\rangle \cdot \langle X, Y\rangle \\
&= -2\sum_i (\nabla_{X_i} \mathrm{I\!I})(X, Y, X_i) + 2\langle d\nu(X), Y\rangle \\
&\qquad - \sum_i \langle d\nu(X_i), X_i\rangle \cdot \langle X, Y\rangle .
\end{aligned}
$$

To interpret the sum $\sum_i (\nabla_{X_i} \mathrm{I\!I})(X, Y, X_i)$, we again switch to coordinates for simplicity, noting to begin with that the sum is easily shown to be independent of the particular orthonormal X_1, X_2 chosen. Now this sum is just the value of

$$\sum_{i,j} g^{ij} \ell_{k\mu i;j} = \sum_j \ell^{j}_{k\mu;j}$$

when we happen to have $X_i = f_i$ $(i = 1, 2)$ and $f_k = X$, $f_\mu = Y$. Since $\sum_j \ell^{j}_{k\mu;j}$ are the components of a tensor, namely the tensor

$$\mathcal{S}(X, Y) = \operatorname{trace}\big(Z \mapsto (\nabla_Z \mathfrak{s})(X, Y)\big),$$

we must, in fact, always have

$$\sum_i (\nabla_{X_i} \mathrm{I\!I})(X, Y, X_i) = \mathcal{S}(X, Y).$$

Thus equation (2) can be written

$$(3) \qquad 2\langle d\nu(X), Y\rangle = \sum_i \langle d\nu(X_i), X_i\rangle \cdot \langle X, Y\rangle + 2\mathcal{S}(X, Y).$$

Substituting in from (1) we obtain

$$(4) \qquad \langle d\nu(X), Y\rangle = (J - \mathcal{K}) \cdot \langle X, Y\rangle + \mathcal{S}(X, Y),$$

the analogue in ordinary surface theory being

$$\langle d\nu(X), Y\rangle = -K \cdot \langle X, Y\rangle.$$

We can solve equation (4) for $d\mathfrak{v}(X)$ explicitly by introducing the tensor $\tilde{\mathcal{S}}$ of type $\binom{1}{1}$ defined by

$$\langle \tilde{\mathcal{S}}(X), Y\rangle = \mathcal{S}(X, Y).$$

Then we have

$$\text{(B)} \qquad d\mathfrak{v}(X) = (J - \mathcal{K}) \cdot X + \tilde{\mathcal{S}}(X).$$

In terms of a map $f\colon U \to \mathbb{R}^3$ we have

$$\begin{aligned} \mathscr{b}_{\mu k} &= \mathscr{g}_{\mu k} \cdot (J - \mathcal{K}) + \sum_j \ell^j_{\mu k;j} \\ &= \mathscr{g}_{\mu k} \cdot (J - \mathcal{K}) + \mathcal{L}_{\mu k}, \qquad \text{say;} \end{aligned}$$

in terms of our moving frame we have

$$\begin{aligned} \psi_3^j &= (J - \mathcal{K})\theta^j + \sum_{i,k} c_{ijk;k}\theta^i \\ &= (J - \mathcal{K})\theta^j + \sum_i C_{ij}\theta^i, \qquad \text{say.} \end{aligned}$$

Conversely, it is not hard to see that if we define $d\mathfrak{v}(X)$ [or the $\mathscr{b}_{\mu k}$, or the ψ_3^j] by these formulas, then equation (A) [or the coordinate equation right below it, or equation (3) on page 123] is satisfied.

As one immediate consequence of equation (4) we find that the special affine normal $\mathfrak{v}$ has another property in common with the ordinary normal ν:

20. PROPOSITION. The map $d\mathfrak{v}\colon M_p \to M_p$ is self-adjoint with respect to the inner product $\langle\ ,\ \rangle_p$ on M_p,

$$\langle d\mathfrak{v}(X_p), Y_p\rangle = \langle X_p, d\mathfrak{v}(Y_p)\rangle \qquad \text{for } X_p, Y_p \in M_p.$$

PROOF. We just need to show that $\mathcal{S}(X, Y) = \mathcal{S}(Y, X)$. This follows from the symmetry of $\mathscr{s}$, and the definition of $\mathcal{S}$ (on page 126). ❖

The eigenvectors of $-d\nu\colon M_p \to M_p$ are naturally called the **special affine principal directions** at p, and the corresponding eigenvalues are the **special affine principal curvatures** k_1 and k_2. The **special affine mean curvature** $\mathcal{H}$ and **special affine (extrinsic) curvature** $\mathcal{K}_{\text{ext}}$ are then defined by

$$\mathcal{H} = \tfrac{1}{2}(k_1 + k_2)$$
$$\mathcal{K}_{\text{ext}} = k_1 \cdot k_2.$$

If $X_1, X_2 \in M_p$ are orthonormal, then

$$\mathcal{H}(p) = -\frac{1}{2}\sum_{j=1}^{2}\langle d\nu(X_j), X_j\rangle.$$

So equation (1) on page 125 shows that we also have

$$\mathcal{H} = \mathcal{K} - J.$$

We have obtained these results by looking at the integrability conditions for the tangential part of the Gauss formulas. Now we will look at the integrability conditions for the $\perp$ component,

$$\psi_i^3 = \theta^i.$$

Exterior differentiation gives

$$-\sum_{k=1}^{2}\psi_k^3 \wedge \psi_i^k = d\theta^i = -\sum_{k=1}^{2}\psi_k^i \wedge \theta^k$$

or

$$-\sum_k \theta^k \wedge \psi_i^k = \sum_k \theta^k \wedge \psi_k^i \qquad \text{or} \qquad \sum_k (\psi_i^k + \psi_k^i) \wedge \theta^k = 0.$$

But these conditions are *automatic*, for we derived this equation in the proof of Proposition 16 (conversely, the equation follows immediately from equation (2) on page 106 and symmetry of the c_{ijk}).

On the other hand, we still have to look at the conditions which say that $\mathcal{N}_{ij} = \mathcal{N}_{ji}$. In ordinary surface theory they reduced to the Codazzi-Mainardi equations; now we will obtain new conditions. The moving frame version of the formula for $\mathcal{N}_i$ is that on page 127.

(1) $$\psi_3^j = (J - \mathcal{K})\theta^j + \sum_i C_{ij}\theta^i.$$

Exterior differentiation gives

$$-\sum_{\rho} \psi_{\rho}^{j} \wedge \psi_{3}^{\rho} = d(J - \mathcal{K}) \wedge \theta^{j} + (J - \mathcal{K})d\theta^{j} + \sum_{i} dC_{ij} \wedge \theta^{i}$$
$$- \sum_{i} C_{ij} \wedge \left(\sum_{\rho} \omega_{\rho}^{i} \wedge \theta^{\rho}\right).$$

Substituting in for ψ_3^k from (1), noting that $-\sum_{\rho} \psi_{\rho}^{j} \wedge \theta^{\rho} = d\theta^{j}$, and switching dummy indices in the last term on the right, we have

$$-\sum_{i,\rho} C_{i\rho}\psi_{\rho}^{j} \wedge \theta^{i} = d(J - \mathcal{K}) \wedge \theta^{j} + \sum_{i} dC_{ij} \wedge \theta^{i} - \sum_{i,\rho} C_{\rho j} \wedge \omega_{i}^{\rho} \wedge \theta^{i}.$$

Now writing the ψ_{ρ}^{j} in terms of the ω_{ρ}^{j} we get

$$-\sum_{i,\rho} C_{i\rho}\omega_{\rho}^{j} \wedge \theta^{i} - \sum_{i,\rho,k} C_{i\rho}c_{j\rho k}\theta^{k} \wedge \theta^{i}$$
$$= d(J - \mathcal{K}) \wedge \theta^{j} + \sum_{i} dC_{ij} \wedge \theta^{i} - \sum_{i,\rho} C_{\rho j} \wedge \omega_{i}^{\rho} \wedge \theta^{i}.$$

Finally, since $\omega_{\rho}^{j} = -\omega_{j}^{\rho}$, we can write

(2) $$-\sum_{i,\rho,k} C_{i\rho}c_{j\rho k}\theta^{k} \wedge \theta^{i}$$
$$= d(J - \mathcal{K}) \wedge \theta^{j} + \sum_{i}\left[dC_{ij} - \sum_{\rho} C_{\rho j} \wedge \omega_{i}^{\rho} - \sum_{\rho} C_{i\rho}\omega_{j}^{\rho}\right] \wedge \theta^{i}.$$

To interpret this equation, we first apply it to (X, X_j). The left side gives

(2a) $$\sum_{i,\rho} C_{i\rho}c_{j\rho j}\theta^{i}(X) - \sum_{\rho,k} C_{j\rho}c_{j\rho k}\theta^{k}(X).$$

The first term on the right side gives

(2b) $$X(J - \mathcal{K}) - X_j(J - \mathcal{K}) \cdot \theta^{j}(X).$$

For the other term on the right side we note, as on page 123, that the tensor $\mathcal{S} = \sum_{i,j} C_{ij}\theta^{i} \otimes \theta^{j}$ has a covariant derivative

$$\nabla\mathcal{S} = \sum_{i,j,k} C_{ij;k}\theta^{i} \otimes \theta^{j} \otimes \theta^{k},$$

where (Problem 5)

$$\sum_k C_{ij;k}\theta^k = dC_{ij} - \sum_\rho C_{\rho j}\omega_i^\rho - \sum_\rho C_{i\rho}\omega_j^\rho.$$

So the second term on the right side of (2) is

$$\sum_{i,k} C_{ij;k}\theta^k \wedge \theta^i,$$

which when applied to (X, X_j) gives

(2c)
$$\sum_k C_{jj;k}\theta^k(X) - \sum_i C_{ij;j}\theta^i(X).$$

We now take the equations (2a) = (2b) + (2c) and add them for $j = 1, 2$ [the resultant equation is equivalent to the individual equations, for if a 2-form α satisfies $\sum_{j=1}^2 \alpha(X, X_j) = 0$, then also $\alpha(X_i, X_j) = 0$ for $i = 1, 2$], obtaining

(3)
$$\begin{aligned}\sum_{i,\rho} C_{i\rho}\Big(\sum_j c_{jj\rho}\Big)\cdot\theta^i(X) - \sum_k\Big(\sum_{j,\rho} C_{j\rho}c_{j\rho k}\Big)\theta^k(X)\\ = 2X(J - \mathcal{K}) - X(J - \mathcal{K})\\ + \sum_k\Big(\sum_j C_{jj;k}\Big)\theta^k(X) - \sum_i\Big(\sum_j C_{ij;j}\Big)\theta^i(X).\end{aligned}$$

Now we have

(3a)
$$\sum_{i,\rho} C_{i\rho}\Big(\sum_j c_{jj\rho}\Big)\cdot\theta^i(X) = 0$$

by the apolarity conditions $\sum_j c_{jj\rho} = 0$. In order to interpret the term involving $\sum_{j,\rho} C_{j\rho}c_{j\rho k}$ we introduce the tensor $\mathcal{S} * \mathrm{I\!I}$ of type $\binom{1}{0}$ defined by

$$\mathcal{S} * \mathrm{I\!I}(X) = \sum_{i,j=1}^{2} \mathcal{S}(X_i, X_j)\cdot \mathrm{I\!I}(X, X_i, X_j),$$

where X_1, X_2 is any orthonormal basis; it is easily checked that this definition is independent of the choice of such basis. [This is the simplest description of $\mathcal{S} * \mathrm{I\!I}$—a completely invariant definition requires an orgy of linear algebra.

Classically, $\mathcal{S} * \mathrm{II}$ is simply described in terms of its components, which are given by the 4-fold contraction

$$\sum_{i,\mu,\rho,\sigma} g^{\mu i} g^{\rho\sigma} \mathcal{L}_{\sigma i} \ell_{\mu\rho j},$$

where $\mathcal{L}_{\sigma i}$ are the components of $\mathcal{S}$.] Now we clearly have

$$\text{(3b)} \qquad -\sum_k \Big(\sum_{j,\rho} C_{j\rho} c_{j\rho k}\Big)\theta^k(X) = -\mathcal{S} * \mathrm{II}(X).$$

To deal with the term $\sum_j C_{jj;k}$ it is again simplest to switch to coordinates. We have the apolarity conditions

$$\begin{aligned}
& 0 = \sum_{i,j} g^{ij} \ell_{ij\rho} \\
\Longrightarrow\ & 0 = \sum_{i,j,\rho} g^{ij} g^{\mu\rho} \ell_{ij\rho} = \sum_{i,j} g^{ij} \ell^{\mu}_{ij} \\
\Longrightarrow\ & 0 = \sum_{i,j,\mu} (g^{ij} \ell^{\mu}_{ij})_{;\mu} = \sum_{i,j,\mu} g^{ij} (\ell^{\mu}_{ij;\mu}) \qquad \text{by Ricci's Lemma} \\
& = \sum_{i,j} g^{ij} \Big(\sum_{\mu} \ell^{\mu}_{ij;\mu}\Big) = \sum_{i,j} g^{ij} \mathcal{L}_{ij} \\
\Longrightarrow\ & 0 = \sum_{i,j} (g^{ij} \mathcal{L}_{ij})_{;k} = \sum_{i,j} g^{ij} \mathcal{L}_{ij;k}, \qquad \text{again by Ricci's Lemma.}
\end{aligned}$$

This tells us that

$$\text{(3c)} \qquad \sum_j C_{jj;k} = 0 \Longrightarrow \sum_k \Big(\sum_j C_{jj;k}\Big)\theta^k(X) = 0$$

[which can also be obtained, with some pain, directly from the apolarity conditions $\sum_j c_{jjk} = 0$]. Finally, we note that

$$\begin{aligned}
\text{(3d)} \qquad -\sum_i \Big(\sum_j C_{ij;j}\Big)\theta^i(X) &= -\sum_j (\nabla_{X_j} \mathcal{S})(X, X_j) \\
&= -\operatorname{trace}\big(Z \mapsto (\nabla_Z \tilde{\mathcal{S}})(X)\big),
\end{aligned}$$

by the very same argument that was used on page 126. Now the equation

$$(3a) + (3b) = X(J - \mathcal{K}) + (3c) + (3d)$$

yields

The Special Affine Codazzi-Mainardi Equations:
$X(J - \mathcal{K}) = \operatorname{trace}\big(Z \mapsto (\nabla_Z \tilde{\mathcal{S}})(X)\big) - \mathcal{S} * \mathrm{II}(X).$

In terms of a map $f: U \to \mathbb{R}^3$ we have

$$(J - \mathcal{K})_j = \sum_{\mu,i} g^{\mu i} \mathcal{L}_{\mu j;i} - \sum_{i,\mu,\rho,\sigma} g^{\mu i} g^{\rho\sigma} \mathcal{L}_{\sigma i} \ell_{\mu\rho j}.$$

Finally, we are ready to state

21. FUNDAMENTAL THEOREM OF SPECIAL AFFINE SURFACE THEORY (RADON; 1918).

(1) Let $M, \bar{M} \subset \mathbb{R}^3$ be two connected surfaces in $\mathbb{R}^3$, both consisting entirely of elliptic points or both consisting entirely of hyperbolic points; in the former case give both surfaces the usual orientation, and in the latter case, suppose that each surface is also oriented. Let $\nu: M \to \mathbb{R}^3$ and $\bar{\nu}: \bar{M} \to \mathbb{R}^3$ be the affine normal vector fields (determined by the orientations), and let I, II and $\bar{\mathrm{I}}, \bar{\mathrm{II}}$ be the first and second affine fundamental forms for M and $\bar{M}$, respectively. Let $\phi: M \to \bar{M}$ be an orientation preserving diffeomorphism such that

$$\phi^* \bar{\mathrm{I}} = \mathrm{I} \qquad \text{and} \qquad \phi^* \bar{\mathrm{II}} = \mathrm{II}.$$

Then there is a special linear affine motion $A: \mathbb{R}^3 \to \mathbb{R}^3$ such that $\phi = A|M$ and $A_* \nu = \bar{\nu}$.

(2) Let M be an oriented 2-manifold with a (not necessarily positive definite) metric $\langle\!\langle\ ,\ \rangle\!\rangle$ having covariant derivative ∇, curvature tensor $\mathcal{R}$, and curvature $\mathcal{K}$. Let $\mathcal{S}$ be a symmetric tensor on M of order 3. Define

$$J = \frac{1}{2} \langle\!\langle\!\langle \mathcal{S}, \mathcal{S} \rangle\!\rangle\!\rangle,$$

where $\langle\!\langle\!\langle\ ,\ \rangle\!\rangle\!\rangle$ is the inner product on tri-linear maps determined by the inner product $\langle\!\langle\ ,\ \rangle\!\rangle$, and define $s, \mathcal{S}, \tilde{\mathcal{S}}$ by

$$\begin{aligned}
\langle\!\langle s(X,Y), Z \rangle\!\rangle &= \mathcal{S}(X,Y,Z)\\
\mathcal{S}(X,Y) &= \operatorname{trace}\big(Z \mapsto (\nabla_Z s)(X,Y)\big)\\
\langle\!\langle \tilde{\mathcal{S}}(X), Y \rangle\!\rangle &= \mathcal{S}(X,Y).
\end{aligned}$$

Suppose that $\mathcal{S}$ satisfies

(1) The Apolarity Conditions:

$$\operatorname{trace}\big(X \mapsto \mathfrak{s}(X, Y)\big) = 0$$

(2) The Special Affine Codazzi-Mainardi Equations:

$$X(J - \mathcal{K}) = \operatorname{trace}\big(Z \mapsto (\nabla_Z \tilde{\mathcal{S}})(X)\big) - \mathcal{S} * \mathcal{S}(X).$$

Then for any point $p \in M$ there is a neighborhood U of p and an immersion $f\colon U \to \mathbb{R}^3$ such that

$$\langle\!\langle\ ,\ \rangle\!\rangle = f^*\mathrm{I}$$
$$\mathcal{S} = f^*\mathrm{II},$$

where I and II are the affine first and second fundamental forms on $f(U)$ determined by the orientation $f(U)$ gets from the orientation on $U \subset M$.

This can be proved in the same way that we proved Theorem 3, using the classical integrability theorem (I.6-1). Or the Frobenius form of the integrability conditions can be used (see the treatment for ordinary surface theory in Chapter 7). One can also reduce the theorem to Theorems I.10-17 and I.10-18; our integrability conditions reduce to the equations of structure of $SL(3, \mathbb{R})$.

PROBLEMS

1. Consider the Codazzi-Mainardi equations in Corollary 1-12, but write $(\nabla_X \mathrm{II})(Y,Z) = X(\mathrm{II}(Y,Z)) - \mathrm{II}(\nabla_X Y, Z) - \mathrm{II}(Y, \nabla_X Z)$, and similarly for $(\nabla_Y \mathrm{II})(X,Z)$. Choose $X = f_1$, $Y = f_2$, and then $Z = f_1$ or f_2, to obtain equations (B′) on page 56.

2. From equation (**) on page 53 show that

$$N_{ij} = -\sum_h \left(l^h_{i,j} + \sum_\rho l^\rho_i \Gamma^h_{\rho j} \right) f_h - \sum_{h,\rho} g^{h\rho} l_{\rho i} l_{hj} N.$$

Conclude that the equation $N_{ij} = N_{ji}$ is equivalent to

$$l^h_{i,j} + \sum_\rho l^\rho_i \Gamma^h_{\rho j} = l^h_{j,i} + \sum_\rho l^\rho_j \Gamma^h_{\rho i}.$$

Write $l^h_i = \sum_k g^{hk} l_{ki}$, and similarly for l^h_j, and expand. Multiply by $\sum_h g_{\tau h}$, and then use $\sum_h g_{\tau h} g^{hk}{}_{,j} = -\sum_h g_{\tau h,j} g^{hk}$. Show that the resulting equation is equivalent to the Codazzi-Mainardi equations, by making use of the identity

$$g_{ik,j} = [ij,k] + [jk,i].$$

3. Use the method of Problem 1-5 to prove that we can take U to be all of M in Corollary 5 when M is simply-connected.

4. (a) Find a continuous map $f\colon \mathbb{R} \to S^1$ which is onto and locally one-one, but not a covering map. *Hint*: Take part of the universal covering space of S^1.
(b) Let $f\colon X \to Y$ be a continuous map which is onto and locally a homeomorphism, and let X be compact. Then for every $y \in Y$, the set $f^{-1}(y) \subset X$ is finite, say $f^{-1}(y) = \{x_1, \ldots, x_k\}$. Choose disjoint open sets $U_i \ni x_i$ such that f is a homeomorphism on each U_i, and let $U = \bigcap_i f(U_i)$. Using compactness of X, show that there is a compact neighborhood $K \subset U$ of y such that $f^{-1}(K) \subset \bigcup_i U_i$. Conclude that f is a covering map.

5. Let $X_1, \ldots, X_n$ be a moving frame on a manifold with a connection ∇. Let

$$A = \sum a_{i_1 \ldots i_k} \theta^{i_1} \otimes \cdots \otimes \theta^{i_k}$$

be a tensor field of type $\binom{k}{0}$, and let

$$\nabla A = \sum a_{i_1 \ldots i_k; l} \theta^{i_1} \otimes \cdots \otimes \theta^{i_k} \otimes \theta^l.$$

Use Problem 1-1 to check that

$$\sum_l a_{i_1 \ldots i_k; l} \theta^l = da_{i_1 \ldots i_k} - \sum_\rho a_{\rho i_2 \ldots i_k} \omega^\rho_{i_1} - \cdots - \sum_\rho a_{i_1 \ldots i_{k-1} \rho} \omega^\rho_{i_k}.$$

CHAPTER 3

A COMPENDIUM OF SURFACES

In the following chapters it will often be quite useful to have a detailed knowledge of the classical surfaces. In this chapter we will list all the important ones systematically, together with many of their properties; other properties will be mentioned in later chapters, when we have the theorems necessary to derive them. A brief survey of the initial pages may prove quite discouraging, but the reader can be assured that after the basic formulas have been derived once and for all, the reading becomes a lot more pleasant—there are pretty pictures to draw, and interesting points to be made.

Usually we will represent a surface locally as the image of an immersion $f: U \to \mathbb{R}^3$. We collect here the formulas from the previous chapters, with a few additions which are used for actual calculations.

(A)

$$E = \langle f_1, f_1 \rangle \qquad F = \langle f_1, f_2 \rangle \qquad G = \langle f_2, f_2 \rangle$$

$$N = \frac{f_1 \times f_2}{|f_1 \times f_2|} = \frac{f_1 \times f_2}{\sqrt{EG - F^2}}$$

$$l = \langle -N_1, f_1 \rangle = \langle N, f_{11} \rangle = \frac{\det \begin{pmatrix} f_{11} \\ f_1 \\ f_2 \end{pmatrix}}{\sqrt{EG - F^2}} \qquad m = \langle -N_1, f_2 \rangle = \langle N, f_{12} \rangle = \frac{\det \begin{pmatrix} f_{12} \\ f_1 \\ f_2 \end{pmatrix}}{\sqrt{EG - F^2}} \qquad n = \langle -N_2, f_2 \rangle = \langle N, f_{22} \rangle = \frac{\det \begin{pmatrix} f_{22} \\ f_1 \\ f_2 \end{pmatrix}}{\sqrt{EG - F^2}}$$

(Here we have made a specific choice of N, which will influence the sign of various quantities to be computed later on.) We also recall that

$$\begin{pmatrix} \text{matrix of } -d\nu\colon M_p \to M_p \\ \text{with respect to } (f_1)_p, (f_2)_p \end{pmatrix} = (g_{ij})^{-1}(l_{ij}) = \frac{1}{EG - F^2} \cdot \begin{pmatrix} G & -F \\ -F & E \end{pmatrix} \begin{pmatrix} l & m \\ m & n \end{pmatrix}$$

$[f_i, g_{ij}, l_{ij}$ evaluated at (s,t); where $p = f(s,t)]$.

The principal curvatures k_1, k_2 are the eigenvalues of this matrix, and the Gaussian curvature K is $k_1 \cdot k_2$, while the mean curvature H is $(k_1 + k_2)/2$. So K and H are the determinant and half the trace, respectively, of this matrix. This gives us

(B)
$$K = \frac{ln - m^2}{EG - F^2}$$
$$H = \frac{En - 2Fm + Gl}{2(EG - F^2)}.$$

(The sign of H depends on the choice of N, but the sign of K does not.)

In equation (B), the left sides H and K must be evaluated at $p = f(s,t)$ when the right sides are evaluated at (s,t); similar conventions will be used in the remaining equations. We remind the reader that it is also possible, in principal at least, to compute K directly from E, F, G. Since k_1, k_2 are the roots of the equation $\lambda^2 - 2H\lambda + K = 0$, we have

(C)
$$k_1, k_2 = H \pm \sqrt{H^2 - K}.$$

(The signs of k_1, k_2 both change when N is changed.)

It is, as usual, rather more difficult to find the principal directions, that is, the *eigenvectors* of $-d\nu$. We leave it to the reader (Problem 1) to show that

(D)
$a_1 f_1 + a_2 f_2$ is a principal vector if and only if
$$\det \begin{pmatrix} {a_2}^2 & -a_1 a_2 & {a_1}^2 \\ E & F & G \\ l & m & n \end{pmatrix} = 0.$$

We have already pointed out that at an umbilic point we have the necessary and sufficient condition

(E)
$$l = kE, \quad m = kF, \quad n = kG$$ at an umbilic point, with $k_1 = k_2 = k$.

(The sign of k depends on N.)

This condition can also be derived from (D), since the determinant must be 0 for all choices of a_1 and a_2. It is also clear that

(F)
$a_1 f_1 + a_2 f_2$ is an asymptotic vector if and only if
$$l{a_1}^2 + 2ma_1a_2 + n{a_2}^2 = 0.$$

Finally, it is easy to see that

(G)

If $F = m = 0$ at a point, then f_1 and f_2 are principal vectors there, and

$$-d\nu(f_1) = \frac{l}{E} f_1, \qquad -d\nu(f_2) = \frac{n}{G} f_2.$$

When our surface is actually the graph of a function $h\colon \mathbb{R}^2 \to \mathbb{R}$, so that we can choose

$$f(s,t) = (s,t,h(s,t))$$
$$f_1 = (1,0,h_1(s,t))$$
$$f_2 = (0,1,h_2(s,t)) \qquad f_{ij} = (0,0,h_{ij}),$$

we obtain the following formulas:

(A′)

$$E = 1 + {h_1}^2 \qquad F = h_1 h_2 \qquad G = 1 + {h_2}^2$$

$$N = \frac{(-h_1, -h_2, 1)}{\sqrt{1 + {h_1}^2 + {h_2}^2}}$$

$$l = \frac{h_{11}}{\sqrt{1 + {h_1}^2 + {h_2}^2}} \qquad m = \frac{h_{12}}{\sqrt{1 + {h_1}^2 + {h_2}^2}} \qquad n = \frac{h_{22}}{\sqrt{1 + {h_1}^2 + {h_2}^2}}.$$

(B′)

$$K = \frac{h_{11}h_{22} - {h_{12}}^2}{[1 + {h_1}^2 + {h_2}^2]^2}$$

$$H = \frac{(1 + {h_1}^2)h_{22} + (1 + {h_2}^2)h_{11} - 2h_1h_2h_{12}}{2[1 + {h_1}^2 + {h_2}^2]^{3/2}}.$$

(C′)

$$k_1, k_2 = H \pm \sqrt{H^2 - K}.$$

(D′)

$a_1 f_1 + a_2 f_2 = (a_1, a_2, a_1 h_1 + a_2 h_2)$ is a principal vector if and only if

$$\det\begin{pmatrix} {a_2}^2 & -a_1a_2 & {a_1}^2 \\ 1 + {h_1}^2 & h_1h_2 & 1 + {h_2}^2 \\ h_{11} & h_{12} & h_{22} \end{pmatrix} = 0.$$

(E′)

$$h_{11} = k(1+h_1{}^2), \quad h_{12} = kh_1h_2, \quad h_{22} = k(1+h_2{}^2)$$
at an umbilic point, with $k_1 = k_2 = k\sqrt{1+h_1{}^2+h_2{}^2}$.

(F′)

$a_1f_1 + a_2f_2 = (a_1, a_2, a_1h_1 + a_2h_2)$ is an asymptotic vector if and only if
$$h_{11}a_1{}^2 + 2h_{12}a_1a_2 + h_{22}a_2{}^2 = 0.$$

It is also quite useful to be able to compute K and H for surfaces

$$M = \{p \in \mathbb{R}^2 : W(p) = 0\},$$

where $W\colon \mathbb{R}^3 \to \mathbb{R}$. Recall (pg. II.113) that we can choose

$$\nu = \frac{Z}{|Z|} \quad \text{for} \quad Z = (W_1, W_2, W_3).$$

If $X = (a_1, a_2, a_3)_p$, then

$$-d\nu(X) = -\nabla_X \frac{Z}{|Z|} = -\frac{1}{|Z|}\nabla_X Z - X\left(\frac{1}{|Z|}\right)Z$$
$$= -\frac{1}{|Z|}\left(\textstyle\sum_i a_iW_{1i}, \sum_i a_iW_{2i}, \sum_i a_iW_{3i}\right) - \underbrace{X\left(\frac{1}{|Z|}\right)Z}_{\text{normal to } M},$$

so

$$\langle -d\nu(X), X\rangle = -\frac{1}{|Z|}\textstyle\sum_{i,j} a_ia_jW_{ij} \qquad [W_{ij} \text{ evaluated at } p].$$

This means that the principal curvatures k_1, k_2 at p are $-1/|Z|$ times the maximum and minimum of

$$\textstyle\sum_{i,j} a_ia_jW_{ij} \quad \text{on} \quad S = \left\{(a_1,a_2,a_3) : \sum_i a_i{}^2 = 1 \text{ and } \sum_i a_iW_i = 0\right\}$$
$$[W_i,\ W_{ij} \text{ evaluated at } p].$$

Using Lagrangian multipliers (see Problem 3, if you have forgotten them), these extrema occur at $(a_1, a_2, a_3) \in S$ if and only if there are λ, μ such that

$$D_j\left(\textstyle\sum_i a_ia_kW_{ik}\right) = \lambda D_j\left(\sum_i a_i{}^2\right) + \mu D_j\left(\sum_i a_iW_i\right) \qquad \text{for all } j.$$

Using $W_{ik} = W_{ki}$, we find that

$$\begin{aligned} &\textstyle\sum_i a_i W_{ij} = \lambda a_j + \frac{\mu}{2} W_j \\ &\textstyle\sum_i a_i^2 = 1, \quad \sum_i a_i W_i = 0 \end{aligned} \qquad \text{for some } \lambda, \mu.$$

Since we then have $\sum_{i,j} a_i a_j W_{ij} = \lambda$, this shows that the desired maximum and minimum values of $\sum_{i,j} a_i a_j W_{ij}$ are precisely the numbers λ for which we have

$$\begin{cases} \sum_i a_i W_{ij} = \lambda a_j + \frac{\mu}{2} W_j \\ \sum_i a_i W_i = 0 \end{cases} \qquad \text{for some } (a_1, a_2, a_3) \neq 0, \text{ and some } \mu.$$

We can also write this as

$$(*) \quad \begin{cases} (W_{ij}) \cdot \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \lambda \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} + \frac{\mu}{2} \begin{pmatrix} W_1 \\ W_2 \\ W_3 \end{pmatrix} \\ \sum_i a_i W_i = 0 \end{cases} \qquad \begin{array}{l} \text{for some } (a_1, a_2, a_3) \neq 0, \\ \text{and some } \mu. \end{array}$$

Now, since

$$\left(\begin{array}{ccc|c} & (W_{ij}) - \lambda I & & \begin{matrix} W_1 \\ W_2 \\ W_3 \end{matrix} \\ \hline W_1 & W_2 & W_3 & 0 \end{array} \right) \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ t \end{pmatrix} = \left(\frac{(W_{ij}) \cdot \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} - \lambda \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} + t \begin{pmatrix} W_1 \\ W_2 \\ W_3 \end{pmatrix}}{a_1 W_1 + a_2 W_2 + a_3 W_3} \right),$$

we see that $(*)$ holds precisely when there exists $(a_1, a_2, a_3, t) \neq 0$ so that the right side of the above equation is the 0 column vector. Thus the desired λ's are those for which the left-hand 4×4 matrix has determinant 0:

(C″)

$$k_i = -\frac{1}{\sqrt{W_1^2 + W_2^2 + W_3^2}} \lambda_i,$$

where λ_i are the roots of the *quadratic* equation

$$(**) \qquad \det \left(\begin{array}{ccc|c} & (W_{ij}) - \lambda I & & \begin{matrix} W_1 \\ W_2 \\ W_3 \end{matrix} \\ \hline W_1 & W_2 & W_3 & 0 \end{array} \right) = 0.$$

Consequently,

(B″)

$$K = \frac{1}{W_1{}^2 + W_2{}^2 + W_3{}^2} \cdot \frac{\text{constant term in } (**)}{\text{coefficient of } \lambda^2 \text{ in } (**)}$$

$$H = -\frac{1}{2\sqrt{W_1{}^2 + W_2{}^2 + W_3{}^2}} \cdot \frac{\text{coefficient of } \lambda \text{ in } (**)}{\text{coefficient of } \lambda^2 \text{ in } (**)}.$$

(C″)

$$k_1, k_2 = H \pm \sqrt{H^2 - K}.$$

In particular, to anticipate the case of greatest interest for us, we find:

(***) If $(W_{ij}) = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$, then

$$K = \frac{1}{(W_1{}^2 + W_2{}^2 + W_3{}^2)^2}\left[W_1{}^2\lambda_2\lambda_3 + W_2{}^2\lambda_1\lambda_3 + W_3{}^2\lambda_1\lambda_2\right]$$

$$H = \frac{1}{2(W_1{}^2 + W_2{}^2 + W_3{}^2)^{3/2}}\left[W_1{}^2(\lambda_2+\lambda_3) + W_2{}^2(\lambda_1+\lambda_3) + W_3{}^2(\lambda_1+\lambda_2)\right].$$

[In all the formulas given so far, the sign of k_1, k_2 and of H depends on the choice of N as the normalized vector (W_1, W_2, W_3).]

Since X is a principal vector if and only if $\langle d\nu(X) \times X, \nu\rangle = 0$, we also see that

(D″)

$X = (a_1, a_2, a_3)_p$ is a principal vector if and only if

$$\det \begin{pmatrix} W_1 & \sum_i a_i W_{1i} & a_1 \\ W_2 & \sum_i a_i W_{2i} & a_2 \\ W_3 & \sum_i a_i W_{3i} & a_3 \end{pmatrix} = 0 \quad \text{and} \quad \sum_i a_i W_i = 0.$$

There does not seem to be any especially simple condition for an umbilic, but we do know that p is an umbilic if and only if the determinant in (D'') is 0 for *all* (a_1, a_2, a_3) with $\sum_i a_i W_i = 0$. Finally, we have

(F″) $X = (a_1, a_2, a_3)_p$ is an asymptotic vector if and only if $\sum_i a_i a_j W_{ij} = 0$ and $\sum_i a_i W_i = 0$.

We are now ready to begin our systematic list of surfaces. They come under five headings.

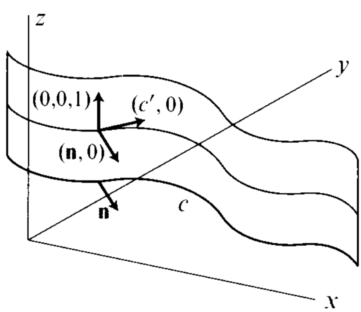
I. THE CLASSICAL FLAT SURFACES

1. *Plane*

For any plane, the normal map ν is constant, so $d\nu = 0$. Thus all points are umbilics with $k_1 = k_2 = 0$, and $K = H = 0$. All vectors are also asymptotic. The plane is actually a special case of

2. *Generalized Cylinder*

Here our surface is $M = \{(x, y, z) : (x, y) = c(s) \text{ for some } s\}$, where c is an immersed curve in $\mathbb{R}^2$, which we assume parameterized by arclength.



The normal map ν is always parallel to the (x, y) plane. One principal direction at any point is $(0, 0, 1)$, with $k_1 = 0$. The other principal direction at $(c(s), z)$ is $(c', 0) = (\mathbf{n}, 0)$, with $k_2 = \kappa(s)$, the curvature of c at s (here we choose the normal map ν to be $(\mathbf{n}, 0)$, where the normal $\mathbf{n}$ for c is picked as on pg. II.6; recall that $\mathbf{n}'(s) = -\kappa(s) \cdot c'(s)$, while k_2 is an eigenvalue of $-d\nu$). Hence $K = 0$ and $H = \frac{1}{2}\kappa(s)$.

The only asymptotic direction is $(0, 0, 1)$, unless $k(s) = 0$, in which case all directions are asymptotic.

3. *Generalized Cone*

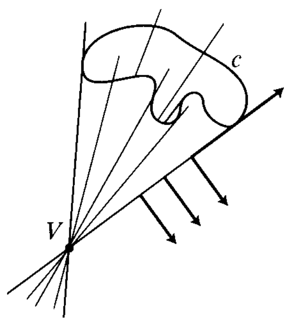
Our surface is parameterized by

$$f(s,t) = V + t[c(s) - V],$$

where $V \in \mathbb{R}^3$ is the *vertex*, and c is an immersed curve in $\mathbb{R}^3$. For f to be an immersion, the vectors

$$f_1 = tc' \quad \text{and} \quad f_2 = c - V$$

must be linearly independent, so we must have $t \neq 0$ (V cannot be in the surface) and $c'(s)$ linearly independent of $c(s) - V$.



Since the tangent space at $f(s,t)$ is spanned by $c'(s)$ and $c(s) - V$, the normal map is constant along the straight lines obtained by keeping s fixed. Consequently, the vectors $f_2(s,t) = c(s) - V$ are principal vectors, with $k_1 = 0$. Therefore $K = 0$. Once again, these vectors are also asymptotic, and there are no others, except at points where $H = k_2/2$ happens to be 0, in which case all vectors are asymptotic.

4. *Tangent Developable*

This surface consists of the tangents to a curve c in $\mathbb{R}^3$ (parameterized by arclength as usual). It can therefore be parameterized by

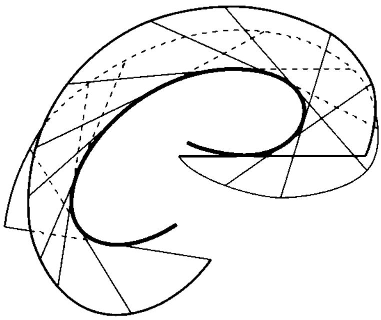
$$f(s,t) = c(s) + tc'(s).$$

We have

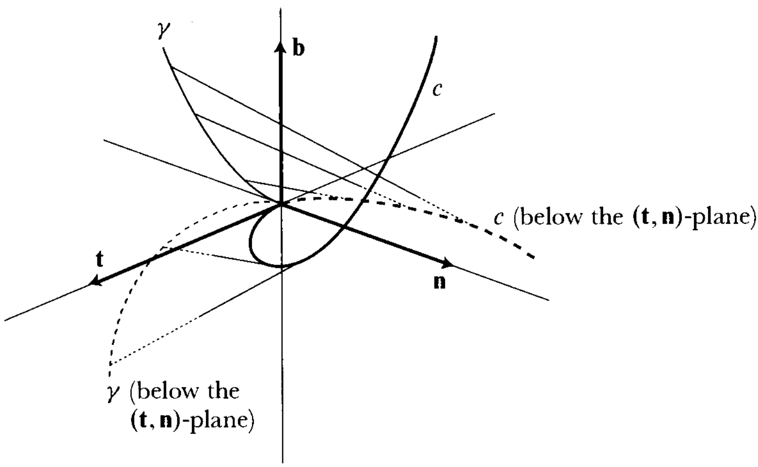
$$f_1 = c' + tc'' = c' + t\kappa\mathbf{n}, \qquad f_2 = c';$$

where $\mathbf{n}$ is the normal vector of c, and κ is the curvature

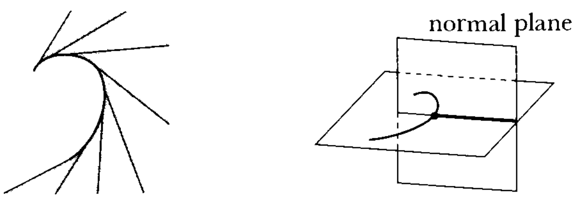
so f is regular if $t \neq 0$ and $\kappa \neq 0$. Thus this parameterization does not allow the curve itself to be part of the surface. In fact, the surface consists of two sheets which meet along the curve in a sharp edge (the **edge of regression** or



cuspidal edge); at any point of the curve, the normal plane intersects the surface in a curve γ with a cusp. This is shown below for our "standard curve" of pp. II.30, 31.



Actually, the assertion we have just made requires some careful interpretation, as can be seen by considering the case where c lies in a plane. The two sheets of the tangent developable are then the same portion of this plane, and their intersection with a normal plane is just a ray. In general, we can analyze the



intersection of the tangent developable and a normal plane to c as follows. For convenience, we consider the point $s = 0$, and assume $c(0) = 0$ and that $\mathbf{t}, \mathbf{n}, \mathbf{b}$ lie along the three coordinate axes. For small $s \neq 0$, the vector $c'(s)$, being close to $\mathbf{t} = c'(0)$, does not lie in the $(\mathbf{n}, \mathbf{b})$-plane; moreover, $c(s)$ also does not lie in the $(\mathbf{n}, \mathbf{b})$-plane. Now we can write (using the Serret-Frenet formulas)

$$\begin{aligned} c(s) &= c(0) + sc'(0) + \frac{s^2}{2}c''(0) + \frac{s^3}{6}c'''(0) + o(s^3) \\ &= (0,0,0) + s(1,0,0) + \frac{s^2}{2}(0,\kappa,0) + \frac{s^3}{6}(-\kappa^2,\kappa',\kappa\tau) + o(s^3); \end{aligned}$$

here $\kappa = \kappa(0)$ and $\tau = \tau(0)$, and $o(s^3)$ is a function with the property that $o(s^3)/s^3 \to 0$ as $s \to 0$. Similarly, we have

$$\begin{aligned} c'(s) &= c'(0) + sc''(0) + \frac{s^2}{2}c'''(0) + o(s^2) \\ &= (1,0,0) + s(0,\kappa,0) + \frac{s^2}{2}(-\kappa^2,\kappa',\kappa\tau) + o(s^2). \end{aligned}$$

Combining, we have

$$\begin{aligned} (*) \qquad f(s,t) &= c(s) + tc'(s) \\ &= (0,0,0) + s(1,0,0) + \frac{s^2}{2}(0,\kappa,0) + \frac{s^3}{6}(-\kappa^2,\kappa',\kappa\tau) + o(s^3) \\ &\quad + t\left[(1,0,0) + s(0,\kappa,0) + \frac{s^2}{2}(-\kappa^2,\kappa',\kappa\tau) + o(s^2)\right]. \end{aligned}$$

For each s there is $t(s)$ for which $f(s,t(s))$ lies in the $(\mathbf{n}, \mathbf{b})$-plane, which means that the first component of $f(s,t(s))$ equals zero:

$$s - \frac{\kappa^2}{6}s^3 + o(s^3) + t(s)\left[1 - \frac{\kappa^2}{2}s^2 + o(s^2)\right] = 0.$$

Dividing through, and giving just a little thought to the meaning of what we are doing, we find that

$$t(s) = -s - \frac{\kappa^2}{3}s^2 + o(s^3).$$

[Here, and below, $o(s^3)$ represents some *new* function with the property that $o(s^3)/s^3 \to 0$ as $0 \to 0$.] Substituting back into (*), we find that

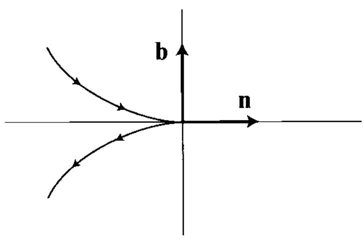
$$\begin{aligned} 2^{\text{nd}} \text{ component of } f(s,t(s)) &= \frac{\kappa}{2}s^2 + o(s^2) + t(s)\{\kappa s + o(s)\} \\ &= -\frac{\kappa}{2}s^2 + o(s^2), \\ 3^{\text{rd}} \text{ component of } f(s,t(s)) &= \frac{\kappa\tau}{6}s^3 + o(s^3) + t(s)\left\{\frac{\kappa\tau}{2}s^2 + o(s^2)\right\} \\ &= -\frac{\kappa\tau}{3}s^3 + o(s^3). \end{aligned}$$

Thus, in the $(\mathbf{n}, \mathbf{b})$-plane, or, in other words, in the (y, z)-plane, the intersection is described up to first order as the curve

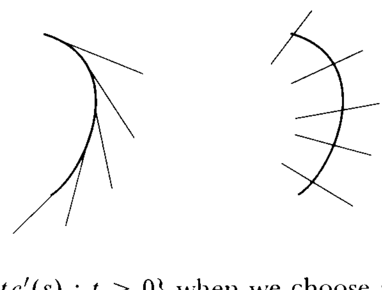
$$s \mapsto \left(-\frac{\kappa}{2}s^2, -\frac{\kappa\tau}{3}s^3\right) = (y(s), z(s));$$

the image is the graph of

$$z^2 = -\frac{8}{9}\frac{\tau^2}{\kappa}y^3.$$



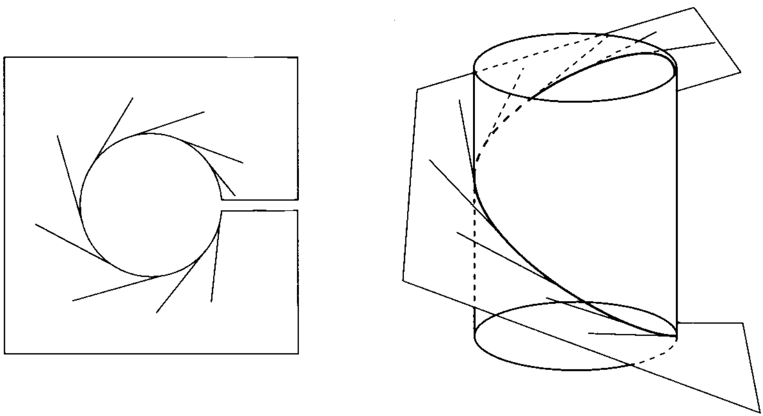
This analysis leads one to suspect that a *single* branch of the tangent developable can be extended so as to include the original curve, though, to be sure, a different parameterization is required. This is most obvious for a plane curve; one obtains an immediate extension if one uses lines perpendicular to the curve as t-parameter lines. Presumably in the general case we obtain a nice param-



eterization of $\{c(s) + tc'(s) : t \geq 0\}$ when we choose as one set of parameter lines the intersection of the surface with normal planes to the curve.

Since the tangent space at $f(s,t)$ is spanned by $c'(s)$ and $\mathbf{n}(s)$, it is the same along the straight lines obtained by keeping s fixed. So the normal map is constant along these lines, and the vectors $f_2(s,t) = c'(s)$ are principal vectors, with $k_1 = 0$. Once again, $K = 0$.

Since all the surfaces in our first category are flat, they are all locally isometric to the plane (that is the reason for the name "tangent developable"—a "development" of one surface on another is the very classical name for an isometry). The reader may easily construct isometries between the plane and generalized cylinders or cones. To map the tangent developable of c isometrically on the plane, we note that f_1, f_2, and consequently E, F, G, depend only on the curvature κ of c, *not on its torsion*. So if c_1 is a plane curve with the same curvature as c, then our original surface is isometric to the tangent developable of c_1, which is a subset of the plane. As an application of this fact, we note that the tangent developable of a helix (which has constant curvature) can be constructed from a piece of paper by cutting a circle out, and twisting the remaining portion around a cylinder.



All the surfaces in our first category are special cases of the surfaces in our second.

II. RULED SURFACES

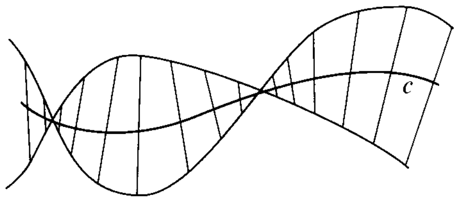
These are the surfaces which can be parameterized as

$$f(s,t) = c(s) + t\delta(s)$$

for two curves c, δ. Since

$$f_1 = c' + t\delta' \quad \text{and} \quad f_2 = \delta,$$

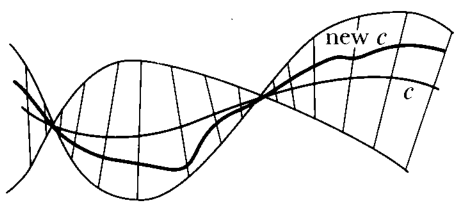
the map f is an immersion when δ and $c' + t\delta'$ are linearly independent. In particular, if δ and c' are linearly independent, this certainly happens for sufficiently small t. For fixed s, we obtain straight lines [or segments] through $c(s)$; these straight lines are called the *rulings* of the surface. The ruled surfaces which are not generalized cones or cylinders, or tangent developables, are sometimes called *scrolls*.



A calculation shows (Problem 4) that

$$K = \frac{-m^2}{EG - F^2} = \frac{-\langle c', \delta \times \delta' \rangle^2}{|(c' + t\delta') \times \delta|^2}.$$

A more reasonable formula for K is obtained when we choose our parameterization more carefully. Note first that we might as well choose δ to be a curve with $|\delta(s)| = 1$ (and consequently $\langle \delta(s), \delta'(s) \rangle = 0$); it is then called the **directrix** of the surface. Next note that if c is replaced by any curve which intersects each ruling only once, and the directrix is kept the same, then we obtain the same



surface. Let us assume that we always have $\delta'(s) \neq 0$ ("the directions of the rulings are always changing"). Then the ruling L_s through $c(s)$ is not parallel to the ruling $L_{s+\varepsilon}$ for small ε, so there is a unique point $P(\varepsilon)$ on L_s closest to $L_{s+\varepsilon}$. One can show (Problem 5) that as $\varepsilon \to 0$, the point $P(\varepsilon)$ approaches the point

$$\sigma(s) = c(s) - \frac{\langle c'(s), \delta'(s) \rangle}{\langle \delta'(s), \delta'(s) \rangle} \cdot \delta(s) \qquad \text{on } L_s.$$

We easily find from this that $\langle \sigma'(s), \delta'(s) \rangle = 0$. This is the only curve σ that can have this property: the point $\sigma(s)$ is simply the unique point on L_s where the tangent plane of the surface contains a vector perpendicular to $\delta'(s)$. The curve σ is called the **striction curve** of the surface, and clearly depends only on the surface, not on the original parameterization. One further alteration eliminates all trace of the original parameterization—we might as well change s so that it is the arclength of δ. We thus have the "standard parameterization"

$$f(s,t) = \sigma(s) + t\delta(s)$$
$$|\delta| = |\delta'| = 1, \quad \langle \sigma', \delta' \rangle = 0$$
$$[\delta, \sigma' \text{ linearly independent}].$$

[The only thing that might go wrong in all this is that $\sigma(s)$ might be a point where the original, and hence the new, f is not an immersion. As a matter of fact (Problem 5), for the tangent developable of c, the striction curve is just c itself.] It now turns out (Problem 4) that

$$K = \frac{-p^2(s)}{(p^2(s)+t^2)^2} \qquad \text{for} \quad p = \langle \sigma'(s), \delta(s) \times \delta'(s) \rangle.$$

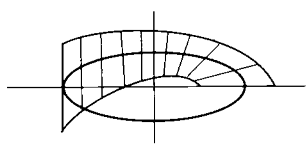
The function p is called the **distribution parameter**. Since p depends only on s, we see that $K \to 0$ as we go out along any ruling.

In addition to these general results, which we will put to use somewhat later, there are a few ruled surfaces of particular interest, two of which we will mention here, and two of which occur in our next category.

1. *Möbius Strip*

Our first example of a ruled surface is the "standard" Möbius strip, a slight modification of the one given on pg. I.10,

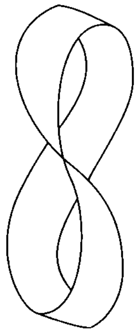
$$\begin{aligned} f(s,t) &= \left(\cos s + t\cos\frac{s}{2}\cos s,\ \sin s + t\cos\frac{s}{2}\sin s,\ t\sin\frac{s}{2}\right) \\ &= (\cos s, \sin s, 0) + t\left(\cos\frac{s}{2}\cos s,\ \cos\frac{s}{2}\sin s,\ \sin\frac{s}{2}\right) \qquad (|t| < \tfrac{1}{2}). \end{aligned}$$



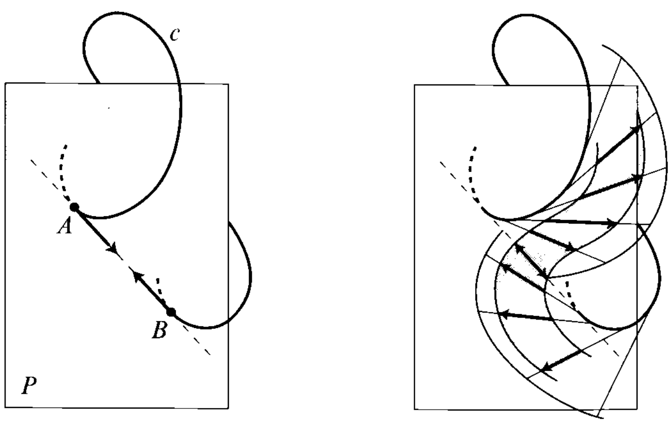
Computing directly from formulas (A) we find that

$$E = \frac{t^2}{4} + \left[1 + t\cos\left(\frac{s}{2}\right)\right]^2 \qquad F = 0 \qquad G = 1$$

(the values of F and G should be obvious!). Notice that this surface is *not* flat, so it is not the Möbius strip that one makes out of a strip of paper.



A C^∞ flat surface homeomorphic to the Möbius strip can be constructed as follows. We start with a curve c like the one shown in the first figure below.



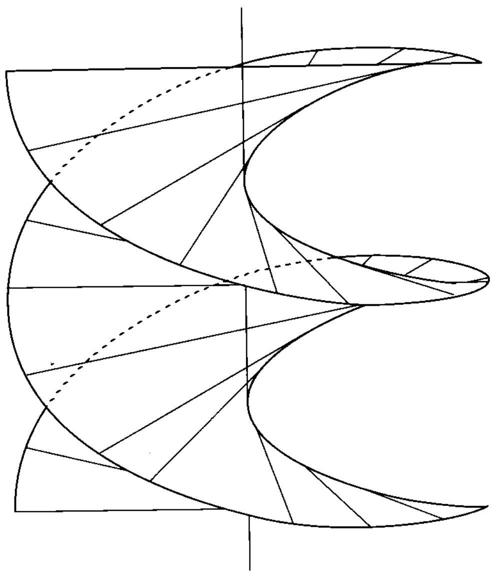
The tangent vectors at A and B are negatives of each other, and the plane P in which they lie is the osculating plane of c at these points, so that c'' lies in P at A and B. We then consider the portion of the tangent developable of c which is formed by the positive multiples of the tangent vectors. From this surface we can cut out a strip, as shown in the second part of the figure, which is homeomorphic to the Möbius strip. If we choose c so that all $c^{(k)}$, $k > 2$ vanish at A and B, then this surface will be C^∞. Notice that the resulting picture is a "back-view" of the Möbius strip shown above.

In Chapter 5 we will mention a way of constructing an analytic flat Möbius strip.

2. *(Right) Helicoid*

This surface is generated by a line which moves along the z-axis in such a way that it remains parallel to the (x, y)-plane, and passes through the points of a circular helix (in other words, the surface is generated by a line perpendicular to the z-axis under a "screw" motion). It is thus given by

$$f(s,t) = (t\cos s, t\sin s, bs), \qquad b \neq 0.$$



The lines t = constant are helices (compare pg. II. 32). Computing from (A) and (B), we find that

$$f_1(s,t) = (-t\sin s, t\cos s, b) \qquad f_2(s,t) = (\cos s, \sin s, 0)$$
$$f_{11}(s,t) = (-t\cos s, -t\sin s, 0) \qquad f_{22}(s,t) = (0,0,0)$$
$$f_{12}(s,t) = (-\sin s, \cos s, 0)$$
$$E = b^2 + t^2, \quad F = 0, \quad G = 1; \quad \sqrt{EG - F^2} = \sqrt{b^2 + t^2}$$
$$l = n = 0, \quad m = \frac{b}{\sqrt{b^2 + t^2}}$$
$$K = -\frac{b^2}{(b^2 + t^2)^2}, \qquad H = 0.$$

The helices t = constant intersect the rulings s = constant in right angles ($F = 0$). They also point in the asymptotic directions ($l = n = 0$), so H must be 0.

III. QUADRIC SURFACES

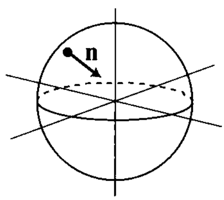
These are the surfaces of the form $W^{-1}(0)$, where

$$W(x_1, x_2, x_3) = \sum a_{ij} x_i x_j + \sum b_i x_i + c.$$

Standard arguments (Problem 6) show that, aside from trivial cases, they are all one of the following (up to rotations and translations); some of them are old friends of ours.

0. *Sphere*

This is a special case of the surfaces of the first group, but it surely deserves special mention. As we know, for the sphere of radius R, the normal map ν is just $-1/R$ times the identity (choosing the *inward* pointing normal, as on pg. II.52); every point is umbilic, the principal curvatures are $1/R$, and $K = 1/R^2$, $H = 1/2R$.

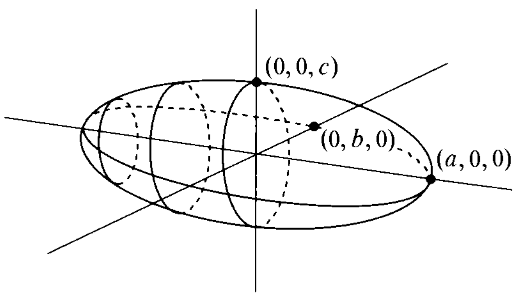


1. *Ellipsoid*

This surface has the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

The planes perpendicular to an axis intersect the surface in a family of similar ellipses.



Choosing

$$W(x, y, z) = \frac{1}{2}\left(\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} - 1\right)$$

$$W_1(x, y, z) = \frac{x}{a^2} \qquad W_2(x, y, z) = \frac{y}{b^2} \qquad W_3(x, y, z) = \frac{z}{c^2}$$

$$(W_{ij}) = \begin{pmatrix} 1/a^2 & 0 & 0 \\ 0 & 1/b^2 & 0 \\ 0 & 0 & 1/c^2 \end{pmatrix}$$

and applying (***) on page 140, we have

$$K = \frac{1}{a^2b^2c^2} \cdot \left(\frac{x^2}{a^4} + \frac{y^2}{b^4} + \frac{z^2}{c^4}\right)^{-2}.$$

Problem 7 casts this into a more useful form, which among other things allows us to see immediately that the maximum and minimum Gaussian curvatures occur at the expected places.

Using (D″), we find that (x, y, z) is an umbilic if and only if

$$\det \begin{pmatrix} x/a^2 & a_1/a^2 & a_1 \\ y/b^2 & a_2/b^2 & a_2 \\ z/c^2 & a_3/c^2 & a_3 \end{pmatrix} = 0 \text{ for all } a_1, a_2, a_3 \text{ with } \frac{a_1x}{a^2} + \frac{a_2y}{b^2} + \frac{a_3z}{c^2} = 0.$$

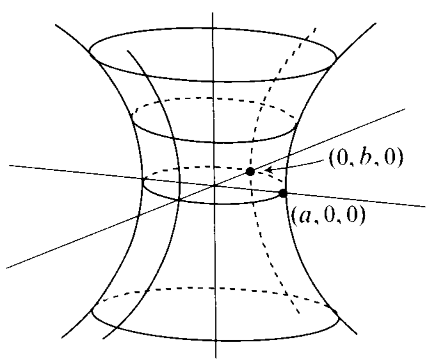
If $a > b > c > 0$, then there turn out to be exactly four umbilics on M (Problem 9). If, on the other hand, we are dealing with an ellipse of rotation, then there will be two whole circles of umbilics.

2. *Elliptic Hyperboloid (of one sheet)*

The equation now is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1.$$

Planes perpendicular to the z-axis intersect the surface in similar ellipses, while planes perpendicular to the other axes intersect it in hyperbolas. When $a = b$, it may be obtained by revolving a hyperbola around the z-axis (hyperboloid of revolution).



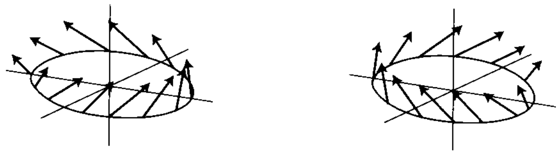
This surface is a *ruled surface*! In fact, it is doubly ruled: it may be parameterized as

$$f(s,t) = (a\cos s, b\sin s, 0) + t(-a\sin s, b\cos s, c)$$

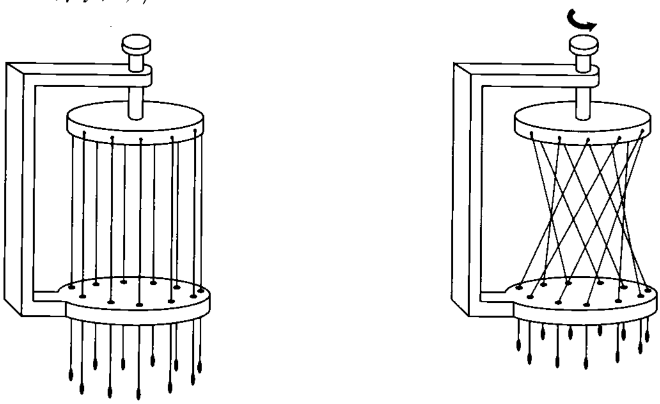
or

$$f(s,t) = (a\cos s, b\sin s, 0) + t(a\sin s, -b\cos s, c).$$

In each case the rulings pass through the ellipse $x^2/a^2 + y^2/b^2 = 1$, $z = 0$ and are perpendicular to the radius vector to that point. For a hyperboloid of



revolution, it is possible to demonstrate this ruling dramatically with apparatus like that pictured below. (The general elliptic hyperboloid must then also be ruled, since it is the image of an hyperboloid of revolution under a linear map $(x, y, z) \mapsto (\alpha x, \beta y, z)$.)



To compute K, we choose a W similar to that for the ellipsoid, and obtain

$$W_1 = \frac{x}{a^2} \qquad W_2 = \frac{y}{b^2} \qquad W_3 = -\frac{z}{c^2}$$

$$(W_{ij}) = \begin{pmatrix} 1/a^2 & 0 & 0 \\ 0 & 1/b^2 & 0 \\ 0 & 0 & 1/c^2 \end{pmatrix}.$$

Then K turns out to be precisely

$$K = \frac{-1}{a^2b^2c^2}\left(\frac{x^2}{a^4} + \frac{y^2}{b^4} + \frac{z^2}{c^4}\right)^{-2}.$$

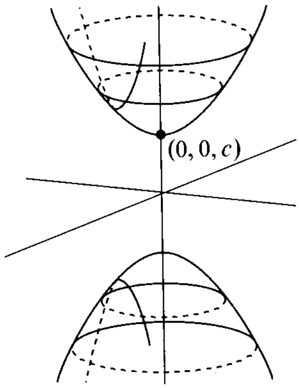
Since K is negative, there are, of course, no umbilics.

3. *Elliptic Hyperboloid (of two sheets)*

The equation is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = -1.$$

It is still true that planes perpendicular to the z-axis intersect the surface in ellipses (when they intersect it at all), while planes perpendicular to the other axes intersect it in hyperbolas. However, the surface looks quite different.



To compute K we choose a W which differs only by a constant from the W for the elliptic hyperboloid of one sheet. The computations are then precisely the same, except that the factor $x^2/a^2 + y^2/b^2 - z^2/c^2$ which appears is now equal to -1. So we get

$$K = \frac{1}{a^2b^2c^2}\left(\frac{x^2}{a^4} + \frac{y^2}{b^4} + \frac{z^2}{c^4}\right)^{-2}.$$

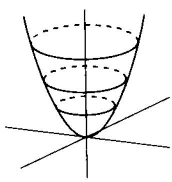
Again there are four umbilics (Problem 9).

4. *Elliptic Paraboloid*

The equation is simply

$$z = \frac{x^2}{a^2} + \frac{y^2}{b^2}.$$

Planes perpendicular to the z-axis intersect the surface in similar ellipses. Planes perpendicular to the other axes intersect it in parabolas. When $a = b$ we have a paraboloid of revolution.



Computations are especially simple:

$$W(x,y,z) = \frac{1}{2}\left(\frac{x^2}{a^2} + \frac{y^2}{b^2} - z\right)$$

$$W_1 = \frac{x}{a^2} \qquad W_2 = \frac{y}{b^2} \qquad W_3 = -\frac{1}{2}$$

$$(W_{ij}) = \begin{pmatrix} 1/a^2 & 0 & 0 \\ 0 & 1/b^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

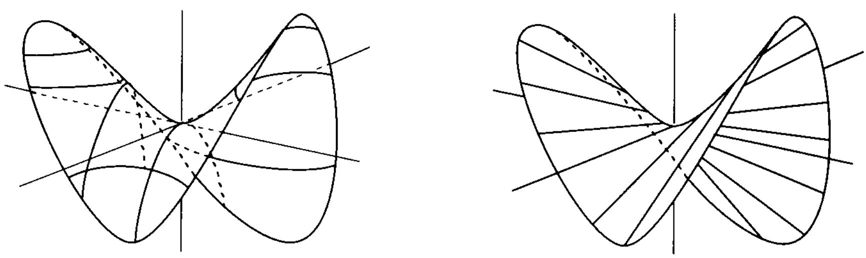
$$K = \frac{1}{4a^2b^2}\left(\frac{x^2}{a^4} + \frac{y^2}{b^4} + \frac{1}{4}\right)^{-2}.$$

There are two umbilics (Problem 10).

5. *Hyperbolic Paraboloid*
Here the equation is

$$z = \frac{x^2}{a^2} - \frac{y^2}{b^2}.$$

Planes perpendicular to the y-axis intersect the surface in parabolas, and planes perpendicular to the x-axis intersect the surface in parabolas pointing the other way. Planes perpendicular to the z-axis intersect the surface in hyperbolas pointing in one direction when the plane lies above the (x,y)-plane, and in the other direction when the plane lies below the (x,y)-plane; the (x,y)-plane itself intersects the surface in two intersecting straight lines.



This surface is also doubly ruled. It may be parameterized as

$$f(s,t) = (as, 0, s^2) + t(a, b, 2s) \qquad \text{or} \qquad f(s,t) = (as, 0, s^2) + t(a, -b, 2s).$$

[It is a classical result that all doubly ruled surfaces with $K < 0$ are quadratic. An elementary, somewhat unsatisfying proof is given in Problem 11; a nice proof can be given (Problem 4-16) by means of affine surface theory, which shouldn't be too surprising, since the property of being doubly ruled is invariant under affine maps.]

For computations we choose

$$W(x, y, z) = \frac{1}{2}\left(\frac{x^2}{a^2} - \frac{y^2}{b^2} - z\right)$$

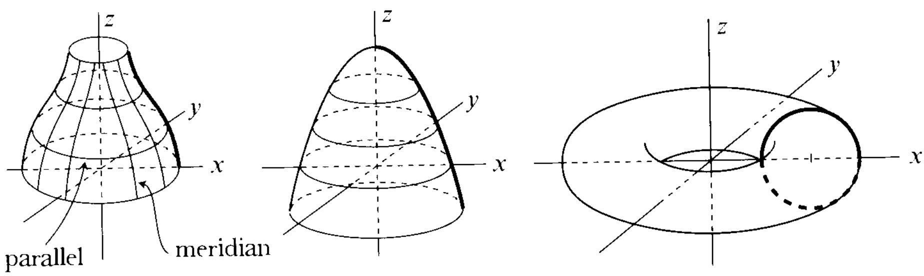
$$W_1 = \frac{x}{a^2} \qquad W_2 = -\frac{y}{b^2} \qquad W_3 = -\frac{1}{2}$$

$$(W_{ij}) = \begin{pmatrix} 1/a^2 & 0 & 0 \\ 0 & -1/b^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$K = \frac{-1}{4a^2b^2}\left(\frac{x^2}{a^4} + \frac{y^2}{b^4} + \frac{1}{4}\right)^{-2}.$$

IV. SURFACES OF REVOLUTION

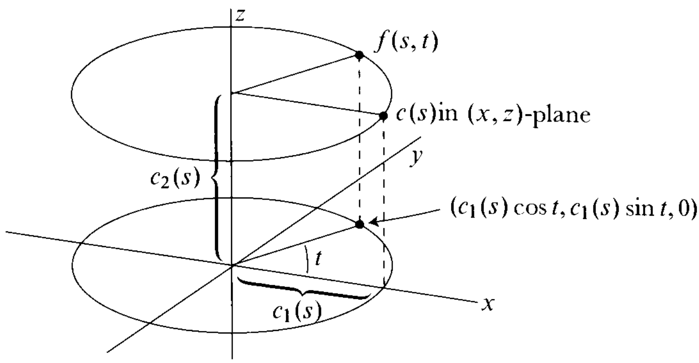
These are the surfaces obtained by starting with a curve (the **profile** curve), lying in the right half of the (x, z)-plane, and revolving it about the z-axis. If



the curve intersects the z-axis, it must do so at a right angle. As illustrated in the figure at the top of the next page, the surface is parameterized by

$$f(s, t) = \big(c_1(s)\cos t,\ c_1(s)\sin t,\ c_2(s)\big).$$

The curves s = constant are called **parallels** and the curves t = constant are called **meridians**.



By (A) we compute

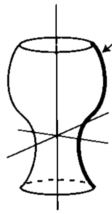
(1)
$$\begin{array}{ll}
f_1 = (c_1' \cos t,\ c_1' \sin t,\ c_2') & f_2 = (-c_1 \sin t,\ c_1 \cos t,\ 0) \\
f_{11} = (c_1'' \cos t,\ c_1'' \sin t,\ c_2'') & f_{22} = (-c_1 \cos t,\ -c_1 \sin t,\ 0) \\
f_{12} = (-c_1' \sin t,\ c_1' \cos t,\ 0) &
\end{array}$$
$$E = (c_1')^2 + (c_2')^2 \qquad F = 0 \qquad G = {c_1}^2$$
$$\sqrt{EG - F^2} = c_1 \sqrt{(c_1')^2 + (c_2')^2}$$
$$l = \frac{c_1'c_2'' - c_2'c_1''}{\sqrt{(c_1')^2 + (c_2')^2}} \qquad m = 0 \qquad n = \frac{c_1 c_2'}{\sqrt{(c_1')^2 + (c_2')^2}}$$

Since $F = m = 0$, the tangent vectors of parallels and meridians point in the directions of the principal curvatures; we can use equations (G) to find directly that the principal curvatures are

(2)
$$k_{\text{meridian}} = \frac{l}{E} = \frac{c_1'c_2'' - c_2'c_1''}{[(c_1')^2 + (c_2')^2]^{3/2}}$$
$$k_{\text{parallel}} = \frac{n}{G} = \frac{c_2'}{c_1\,[(c_1')^2 + (c_2')^2]^{1/2}}$$
$$K = k_{\text{meridian}} \cdot k_{\text{parallel}} = \frac{c_2'(c_1'c_2'' - c_2'c_1'')}{c_1\,[(c_1')^2 + (c_2')^2]^2}$$
$$H = \frac{1}{2}\left(k_{\text{meridian}} + k_{\text{parallel}}\right).$$

It will be useful for us to consider the special case where $c(s) = (g(s), s)$. Then we find that

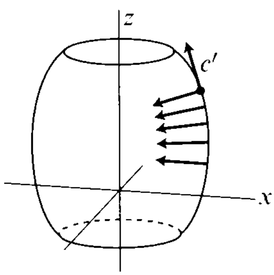
(3)
$$k_{\text{meridian}} = \frac{-g''}{[1+(g')^2]^{3/2}}$$
$$k_{\text{parallel}} = \frac{1}{g[1+(g')^2]^{1/2}}$$
$$K = \frac{-g''}{g[1+(g')^2]^2}$$
$$H = \frac{1+(g')^2 - gg''}{2g[1+(g')^2]^{3/2}}$$



It is also useful to consider the **canonical parameterization**, where $|c'|^2 = (c_1')^2 + (c_2')^2 = 1$. Then also $c_1'c_1'' + c_2'c_2'' = 0$, so we find

(4)
$$E = 1 \qquad F = 0 \qquad G = {c_1}^2$$
$$K = \frac{c_2'(c_1'c_2'' - c_2'c_1'')}{c_1} = \frac{c_1'(c_2'c_2'') - c_1''(c_2')^2}{c_1}$$
$$= \frac{c_1'(-c_1'c_1'') - c_1''[1-(c_1')^2]}{c_1}$$
$$= -\frac{c_1''}{c_1}$$

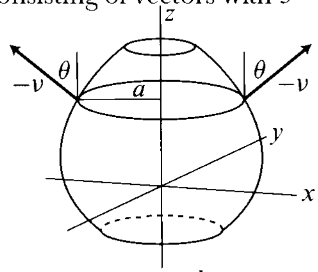
It is interesting to note that we can obtain all these results in a purely geometric way without any calculations. We first observe that the normals to the surface along the profile curve c lie in the (x, z)-plane. This means that

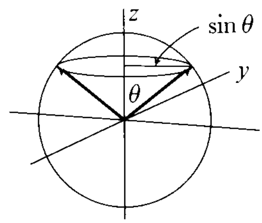


the derivative of ν along the profile curve also lies in the (x, z)-plane. Since it must also be tangent to the surface, it is a multiple of c'. Thus c' is one eigenvector for $-d\nu$. The corresponding eigenvalue is also easy to find. We notice first that our choice of N as $(f_1 \times f_2)/|f_1 \times f_2|$ makes ν inward pointing;

so along c, the vector ν is just the normal $\mathbf{n}$ of c. Therefore $-d\nu(c') = -\mathbf{n}' = \kappa \cdot c'$ (from the Serret-Frenet formulas for plane curves), and the eigenvalue is just the curvature of c. Comparing the formula for k_{meridian} in (2) with the formula on pg. II.8, we see that this is precisely what we have obtained in the calculations.

We next consider the outward pointing normal $-\nu$ along a parallel. This makes a constant angle θ with the z-axis, so in S^2 it traces out a circle, of radius $\sin\theta$, consisting of vectors with 3rd component constant. So the derivative of $-\nu$



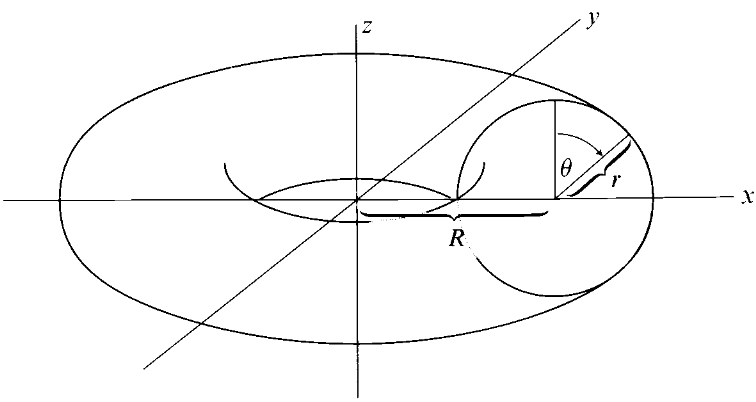


will be a vector with 3rd component 0, and perpendicular to the radius vector. It is therefore a multiple of the tangent vector of the parallel. If we parameterize our parallel so that we go once around in time 2π, then its tangent vector has length a = radius of the parallel. In the same time, the vector $-\nu$ goes once around a circle of radius $\sin\theta$, so its tangent vector has length $\sin\theta$. This shows that the corresponding eigenvalue is $\frac{1}{a}\sin\theta$, which is precisely what the formula for k_{parallel} in (2) gives.

Now let us take some particular cases. We begin with the most familiar example (after cylinders, cones, and spheres).

1. *Torus*

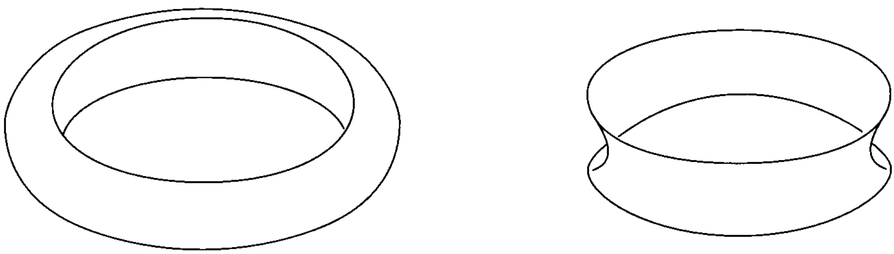
We rotate a circle of radius r around the z-axis so that its center traces out a circle of radius R. Measuring angles θ around the little circle clockwise from



the z-axis, we see that the principal curvatures are:

$$\frac{1}{r} = \text{curvature of circle of radius } r$$
$$\frac{\sin\theta}{R + r\sin\theta}, \quad \text{since} \quad a = R + r\sin\theta.$$

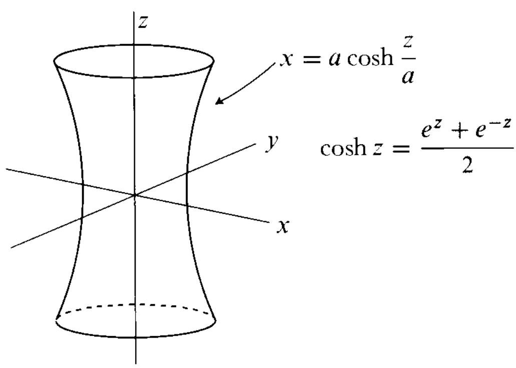
In particular, as suspected, $K > 0$ on the outer half of the torus, and $K < 0$ on



the inner half. Notice that the principal curvatures can never be equal, so there are no umbilics.

2. *Catenoid*

This surface, which we have already met in Volume I, Chapter 9, but not yet been formally introduced to, is obtained by revolving a **catenary**, with equation $x = a\cosh(z/a) = g(z)$, around the z-axis.



Since

$$\cosh' z = \frac{e^z - e^{-z}}{2} = \sinh z, \qquad \cosh'' z = \frac{e^z + e^{-z}}{2} = \cosh z$$
$$1 + (\cosh' z)^2 = 1 + \frac{e^{2z} - 2 + e^{2z}}{4} = \frac{1}{2} + \frac{e^{2z}}{2} = (\cosh z)^2,$$

formulas (3) give us

$$k_{\text{meridian}} = \frac{-\frac{1}{a}\cosh\frac{z}{a}}{\left[\cosh^2\frac{z}{a}\right]^{3/2}} = \frac{-1}{a\cosh^2\frac{z}{a}}$$

$$k_{\text{parallel}} = \frac{1}{a\cosh\frac{z}{a}\left[\cosh^2\frac{z}{a}\right]^{1/2}} = \frac{1}{a\cosh^2\frac{z}{a}}$$

$$H = 0, \qquad K = \frac{-1}{a^2\cosh^4\frac{z}{a}}.$$

Clearly $-1/a^2 \leq K < 0$, with $K = -1/a^2$ on the inner circle $z = 0$, and $K \to 0$ as $z \to \pm\infty$.

It is also useful to find the canonical parameterization for the catenoid; we take the case $a = 1$. For $c(u) = (\cosh u, u)$, we have

$$\text{length of } c \text{ from 0 to } u = \int_0^u \sqrt{1 + (\cosh' v)^2}\,dv = \int_0^u \cosh v\,dv = \sinh u,$$

so we want to take the curve

$$\gamma(s) = c(\sinh^{-1}(s)) = \big(\cosh(\sinh^{-1}(s)),\ \sinh^{-1}(s)\big)$$
$$= \big(\sqrt{1+s^2},\ \sinh^{-1}(s)\big) \qquad \text{[see Problem I.9-20(d)]}.$$

We then have, by formulas (4),

$$E = 1 \qquad F = 0 \qquad G = 1 + s^2$$
$$K = \frac{-1}{(1+s^2)^2}.$$

3. *Rotation Surfaces of Constant Curvature*

We consider surfaces of revolution with canonical parameterization

$$f(s,t) = \big(g(s)\cos t,\ g(s)\sin t,\ h(s)\big), \qquad (g')^2 + (h')^2 = 1,$$

and among these seek the ones with constant K. According to equations (4) we have

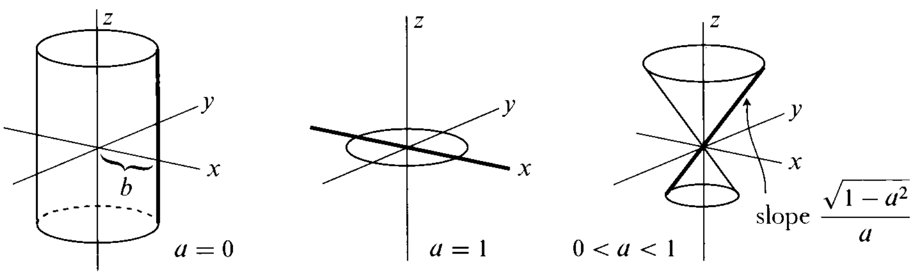
$$K = -\frac{g''}{g}.$$

Case 1. $K = 0$

Then $g(s) = as + b$. If $a \neq 0$, then we can assume that $b = 0$, since this merely amounts to renaming the parameter s. To have $(g')^2 + (h')^2 = 1$ we take

$$\left.\begin{aligned} g(s) &= as \\ h(s) &= \int_0^s \sqrt{1-a^2}\,dt = \pm s\sqrt{1-a^2} \end{aligned}\right\} \quad \text{or} \quad \left\{\begin{aligned} g(s) &= 0\cdot s + b \\ h(s) &= s \end{aligned}\right.$$

(changing h by a constant merely amounts to translating the profile curve along the z-axis); clearly we must have $|a| \leq 1$. For $a = 0$ we obtain a cylinder, for $|a| = 1$ a plane, and for $0 < |a| < 1$ a cone.



Case 2. $K > 0$

For simplicity, we take the case $K = 1$. We have to find g satisfying $g'' + g = 0$. The general solution, $g(s) = a_1 \cos s + a_2 \sin s$, can also be written

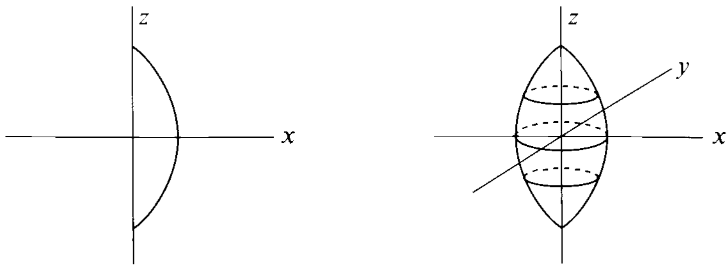
$$g(s) = a\cos(s + b).$$

We can always assume $b = 0$, and we take

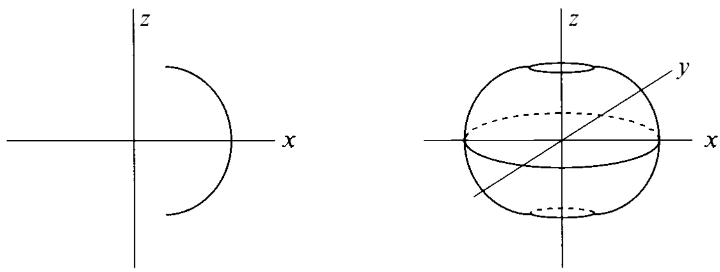
$$\begin{aligned} g(s) &= a\cos s \\ h(s) &= \pm\int_0^s \sqrt{1 - g'(t)^2}\,dt = \pm\int_0^s \sqrt{1 - a^2\sin^2 t}\,dt. \end{aligned}$$

For $a = 1$ we obtain the sphere of radius 1.

For $a < 1$, the integrand in the expression for h is always real, and the only restriction on our formulas is that $g(s)$ must be ≥ 0. We can take $0 \leq s < \pi/2$; the resulting profile curve can be expressed in terms of elliptic integrals.



For $a > 1$, we must restrict s to $0 \leq s \leq \arcsin 1/a$ for the integrand to be real. At the endpoint of the interval, h' is 0, so the profile curve has horizontal tangents; once again, elliptic integrals are involved.



Case 3. $K < 0$

We take the case $K = -1$. The general solution of $g'' - g = 0$ is

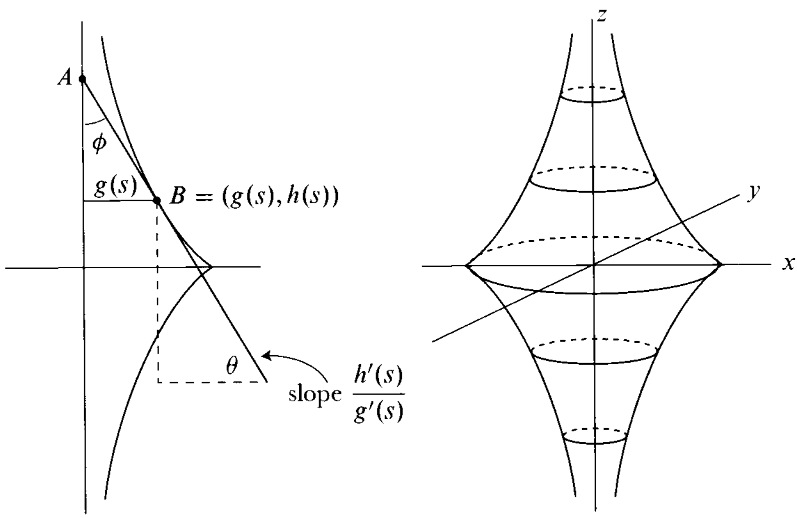
$$g(s) = ae^s + be^{-s}.$$

Suppose first that one of a, b is 0. We can assume that $b = 0$, since changing s to $-s$ interchanges a and b. We might as well assume $a > 0$, since changing a to $-a$ just changes the profile curve to its mirror image. Finally, we can assume $a = 1$, since changing s to $s + s_0$ multiplies a by e^{s_0}. So we take

$$g(s) = e^s$$
$$h(s) = \pm \int_0^s \sqrt{1 - e^{2t}}\, dt;$$

we clearly need $e^{2s} \leq 1$, and therefore $g(s) \leq 1$. The resulting surface is called a **pseudosphere**. Its profile curve was known to mathematicians long before the advent of differential geometry. Since s is the parameterization by arclength, the angle ϕ in the picture below satisfies

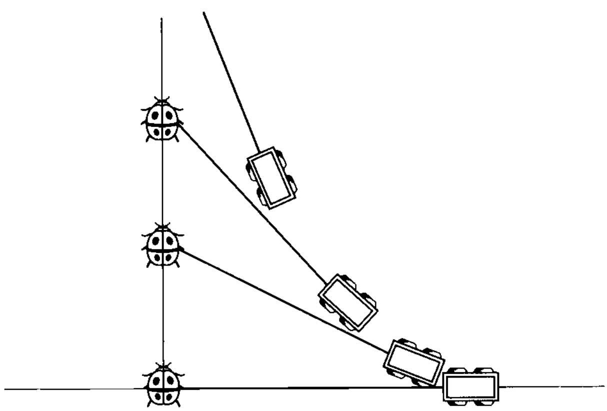
$$\sin\phi = \cos\theta = \frac{g'(s)}{\sqrt{[g'(s)]^2 + [h'(s)]^2}} = g'(s) = e^s.$$



So AB has constant length

$$\overline{AB} = \frac{g(s)}{\sin\phi} = \frac{e^s}{e^s} = 1.$$

If one started at $(0,0)$ and walked along the y-axis pulling a wagon that started at $(1,0)$ and had a handle of length 1, then the wagon would follow this curve, which is therefore called a **tractrix** (Latin: *trahere, tractum* to draw). The upper



half of the tractrix is the graph of the function

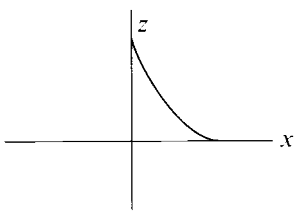
$$\begin{aligned} f(x) &= \int_0^{\log x} \sqrt{1 - e^{2t}}\ dt \\ &= \int_1^x \sqrt{1 - u^2} \cdot \frac{1}{u}\, du \\ &= \sqrt{1 - x^2} - \cosh^{-1} \frac{1}{x}. \end{aligned}$$

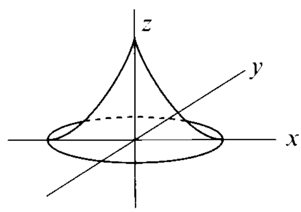
Now suppose that $a, b \neq 0$. Since changing s to $s + s_0$ multiplies a and b by different constants, we can assume that either $a = -b$ or $a = b$.

In the case $a = -b$ we can assume $a > 0$ (by changing s to $-s$ and thereby interchanging a and b). So we take

$$\begin{aligned} g(s) &= a(e^s - e^{-s}) = 2a \sinh s \\ h(s) &= \pm \int_0^s \sqrt{1 - 4a^2 \cosh^2 t}\ dt; \end{aligned}$$

we need $0 < 2a < 1$ and $1 \leq \cosh s \leq 1/2a$, so that $0 \leq s \leq \cosh^{-1} 1/2a$ and $0 \leq g(s) \leq \sqrt{1 - 4a^2}$. These functions can also be expressed in terms of elliptic integrals.

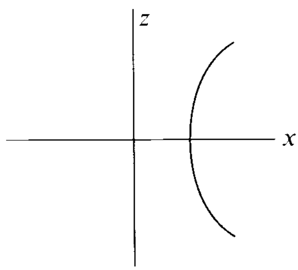


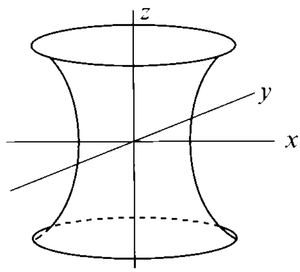


In case $a = b$, we can assume both are positive, since changing the sign of both changes the profile curve to its mirror image. So we take

$$\begin{aligned} g(s) &= 2a \cosh s \\ h(s) &= \pm \int_0^s \sqrt{1 - 4a^2 \sinh^2 t}\ dt; \end{aligned}$$

we need $|\sinh s| \leq 1/2a$ and thus $2a \leq g(s) \leq \sqrt{1+4a^2}$. Elliptic integrals are again required.



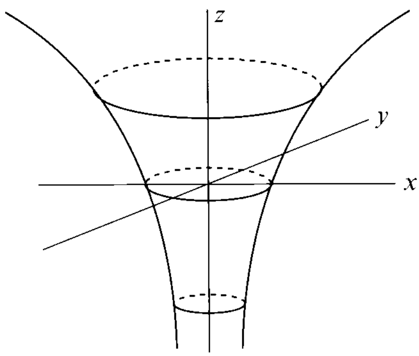


These results about surfaces of revolution may be compared with the remarks made by Riemann in section II.5 of his Inaugural Lecture (pg. II.159).

4. *A Classical Counterexample*

Consider the surface of revolution

$$f(s,t) = (s \sin t,\ s \cos t,\ \log s).$$



Formulas (1) and (2) give

$$E = 1 + \frac{1}{s^2} \qquad F = 0 \qquad G = 1$$

$$k_{\text{meridian}} = \frac{-1/s^2}{\left[1 + \frac{1}{s^2}\right]^{3/2}} \qquad k_{\text{parallel}} = \frac{1/s}{s\left[1 + \frac{1}{s^2}\right]^{1/2}}$$

$$K = \frac{-1}{(1+s^2)^2}.$$

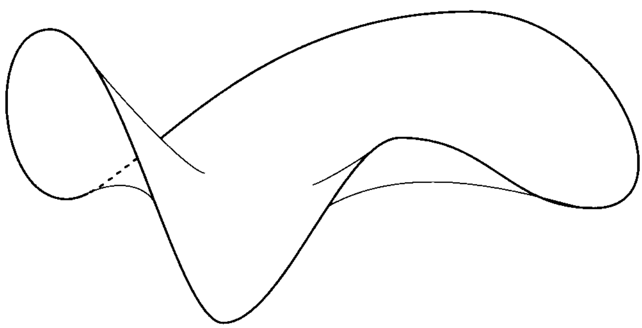
On the other hand, interchanging the role of s and t in the parameterization of the helicoid (page 150), we see that $g(s,t) = (s\cos t,\ s\sin t,\ t)$ has

$$E = 1 \qquad F = 0 \qquad G = 1 + s^2$$
$$K = \frac{-1}{(1+s^2)^2}.$$

Consequently, the map $f(s,t) \mapsto g(s,t)$ preserves K, but is not an isometry; in fact, there is clearly no local isometry between the two surfaces, since the s-parameters would have to correspond to preserve K, and then E would not be preserved.

V. MINIMAL SURFACES

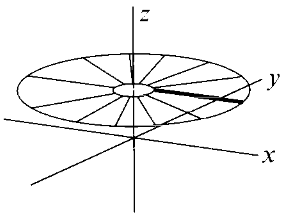
A whole branch of mathematics is devoted to the study of surfaces with mean curvature $H = 0$. As we shall show in Chapter 9, this condition is precisely the one which must be satisfied by a surface which is a critical point for the area function among all surfaces with the same boundary curve c. In particular, if a

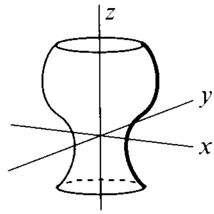


surface has minimum area among those with c as boundary, then it must satisfy the condition $H = 0$. For this reason, surfaces with mean curvature $H = 0$ are called **minimal**.

We have already met two minimal surfaces in our survey, the helicoid and the catenoid. They, in fact, were the first two non-planar minimal surfaces to be discovered (by Meusnier), and we are led to them directly if we seek minimal surfaces among other known classes of surfaces.

Let us consider first the surfaces of revolution. If our profile curve is a straight line perpendicular to the z-axis, we obtain a plane. Otherwise, some portion of the curve can be represented by $c(s) = (g(s), s)$. Using formulas (3) on





page 158, we find that $H = 0$ when

$$1 + (h')^2 - hh'' = 0.$$

This is precisely the equation we obtained on pg. I.321, when we were actually finding the minimum area of a surface of revolution. We found there that the solutions are

$$g(z) = a \cosh \frac{z+b}{a}.$$

This result applies only to a portion of the surface, but a simple least upper bound argument shows that if a connected profile curve has this form somewhere, then it must have it everywhere (and in particular cannot also contain part of a line perpendicular to the z-axis). We have thus shown that:

Any connected minimal surface of revolution is part of a plane or a catenoid.

We consider next the ruled surface

$$f(s,t) = \sigma(s) + t\delta(s)$$
$$|\delta| = |\delta'| = 1, \qquad \langle \sigma', \delta' \rangle = 0$$
$$[\delta, \sigma' \text{ linearly independent}].$$

Then we have

$$f_1 = \sigma' + t\delta' \qquad f_2 = \delta$$
$$f_{11} = \sigma'' + t\delta'' \qquad f_{22} = 0$$
$$f_{12} = \delta'$$

$$F = \langle \delta, \sigma' \rangle \qquad G = 1; \qquad \text{let } W = \sqrt{EG - F^2}$$

$$l = \frac{1}{W} \det \begin{pmatrix} \sigma'' + t\delta'' \\ \sigma' + t\delta' \\ \delta \end{pmatrix} \qquad m = \frac{1}{W} \det \begin{pmatrix} \delta' \\ \sigma' + t\delta' \\ \delta \end{pmatrix} \qquad n = 0.$$

So equations (B) show that $H = 0$ when

$$0 = -2\langle \delta, \sigma' \rangle \det \begin{pmatrix} \delta' \\ \sigma' + t\delta' \\ \delta \end{pmatrix} + \det \begin{pmatrix} \sigma'' + t\delta'' \\ \sigma' + t\delta' \\ \delta \end{pmatrix}.$$

Using multilinearity of det as a function of its rows, and noting that the coefficient of each power of t must vanish, we obtain

$$(1)\ \langle \delta, \sigma' \rangle \det \begin{pmatrix} \delta' \\ \sigma' \\ \delta \end{pmatrix} = 0 \qquad (2)\ \det \begin{pmatrix} \sigma'' \\ \delta' \\ \delta \end{pmatrix} + \det \begin{pmatrix} \delta'' \\ \sigma' \\ \delta \end{pmatrix} = 0 \qquad (3)\ \det \begin{pmatrix} \delta'' \\ \delta' \\ \delta \end{pmatrix} = 0.$$

Equation (3) shows that δ'' is a linear combination of δ and δ'. But we also have

$$\langle \delta', \delta' \rangle = 1 \implies \langle \delta', \delta'' \rangle = 0$$
$$\langle \delta, \delta \rangle = 1 \implies \langle \delta, \delta' \rangle = 0 \implies \langle \delta', \delta' \rangle + \langle \delta, \delta'' \rangle = 0 \implies \langle \delta, \delta'' \rangle = -1,$$

which shows that

$$\delta'' = -\delta.$$

This means that $-\delta$ is the normal $\mathbf{n}$ of the curve δ, and that the curvature of δ is $\kappa = 1$. Also, $\mathbf{b} = \mathbf{t} \times \mathbf{n} = \delta' \times -\delta$, so

$$\mathbf{b}' = -(\delta' \times \delta)' = -(\delta' \times \delta') - (\delta'' \times \delta) = 0.$$

Thus $\tau = 0$, and δ is a plane curve. Since it lies in S^2, and has curvature 1, it must be a circle of radius 1. We can assume therefore that

$$\delta(s) = (\cos s, \sin s, 0).$$

Now in equation (2), the second determinant is already 0, so we find that σ'' is a linear combination of δ, δ', which means that σ'' lies in the (x, y)-plane. So σ must be of the form

$$\sigma(s) = (\alpha(s), \beta(s), bs + a) \implies \sigma'(s) = (\alpha'(s), \beta'(s), b).$$

We might as well assume that $a = 0$, since this just amounts to a translation along the z-axis.

Now consider equation (1), which says that a certain product is 0. If the second factor is 0 for some s_0, then $\sigma'(s_0)$ must be a linear combination of $\delta(s_0), \delta'(s_0)$, so $b = 0$. In this case, all rulings $\sigma(s) + t\delta(s)$ lie in a plane, and our surface is just the plane. If $b \neq 0$, then for all s we must have

$$0 = \langle \delta(s), \sigma'(s) \rangle = \alpha'(s) \cos s + \beta'(s) \sin s$$
$$0 = \langle \delta'(s), \sigma'(s) \rangle = -\alpha'(s) \sin s + \beta'(s) \cos s.$$

So $\alpha' = \beta' = 0$, i.e., α and β are constants, and our surface is given by

$$f(s,t) = (\alpha + t\cos s,\ \beta + t \sin s,\ bs).$$

A translation in the (x, y)-plane changes α and β to 0, and we obtain the helicoid.

Our analysis has left a few points in doubt, because the standard parameterization with which we began is possible only when the directions of the rulings of our ruled surface are always changing. When the directions of the rulings are *never* changing we obtain a generalized cylinder, which is minimal only if it is a plane. As before, a least upper bound argument shows that if $\delta' \neq 0$ on some interval, so that we do have a helicoid on this interval, then $\delta' \neq 0$ everywhere. We have thus shown that:

Any connected minimal ruled surface is part of a plane or a helicoid.

It is, of course, not particularly surprising that each of these families of surfaces contains only one non-planar minimal surface. But it is rather surprising that these two surfaces, *the catenoid and the helicoid, are locally isometric.* To prove this, we merely recall (page 167) that

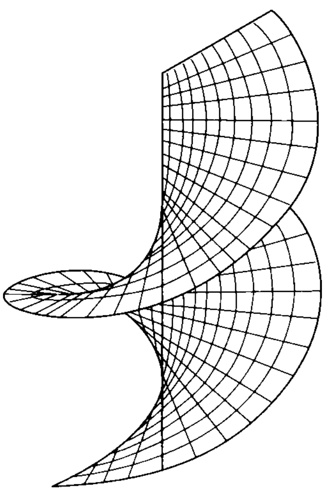
$$\text{the helicoid } f(s,t) = (s\cos t, s\sin t, t) \text{ has } \quad E = 1 \quad F = 0 \quad G = 1 + s^2,$$

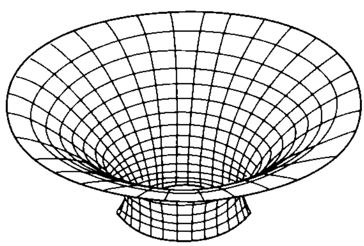
while (page 161)

$$\text{the catenoid } g(s,t) = \big(\sqrt{1+s^2}\,\cos t,\ \sqrt{1+s^2}\,\sin t,\ \sinh^{-1}(s)\big) \text{ also has}$$
$$E = 1 \quad F = 0 \quad G = 1 + s^2.$$

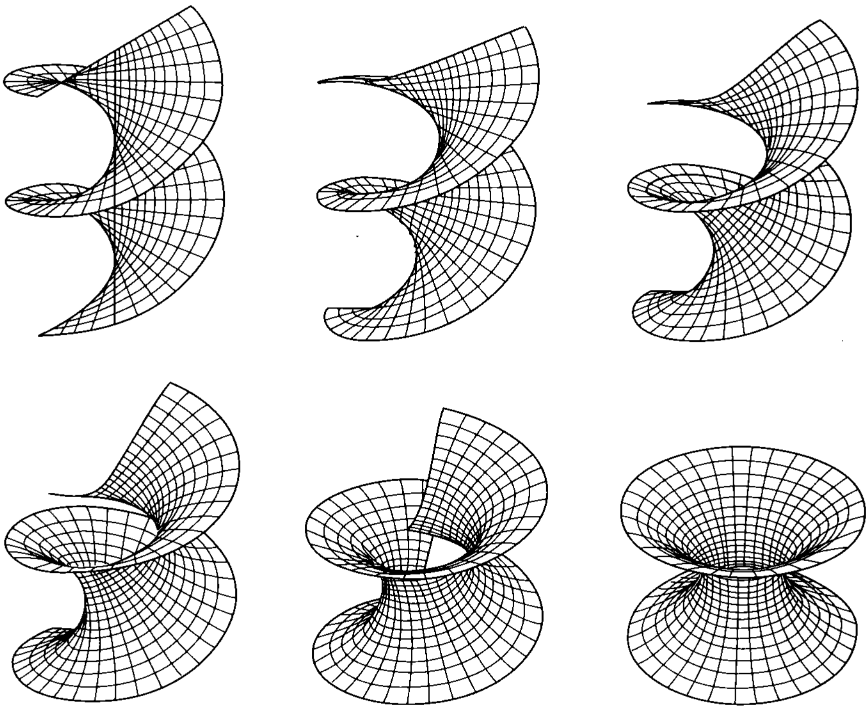
The isometry, taking $f(s,t)$ to $g(s,t)$, carries

rulings of the helicoid to *meridians* of the catenoid (t constant)
helices of the helicoid to *parallels* of the catenoid ($s \neq 0$ constant)
z*-axis* of the helicoid to *center circle* of the catenoid ($s = 0$).





Actually, we can do a lot better than this: it is possible to deform one of these surfaces into the other by means of a 1-parameter family of isometric surfaces:



Although we could write down an explicit formula for this 1-parameter family, it will come out very naturally in Chapter 9.

The example of the helicoid and catenoid also shows that two immersions, $f, \tilde{f}: \mathbb{R}^2 \to \mathbb{R}^3$ with the same g_{ij} (and therefore also the same K) as well as the same H need not differ by a Euclidean motion of $\mathbb{R}^3$—one needs to know the l_{ij} themselves, not just trace(l_{ij}) and $\det(l_{ij})$ [or equivalently the eigenvalues of (l_{ij})], in order to determine the surface.

The first minimal surface to be discovered after the catenoid and helicoid was

1. *Scherk's Minimal Surface*

This is the surface M defined by

$$e^z \cos x = \cos y.$$

If we define $W : \mathbb{R}^3 \to \mathbb{R}$ by

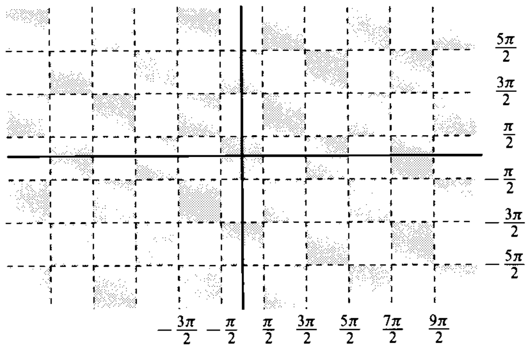
$$W(x, y, z) = e^z \cos x - \cos y,$$

then the three vectors

$$W_1(x, y, z) = -e^z \sin x \qquad W_2(x, y, z) = \sin y \qquad W_3(x, y, z) = e^z \cos x$$

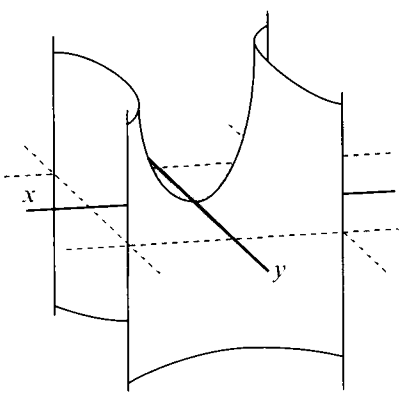
are never all 0, so $W^{-1}(0)$ is a surface, which is orientable since it has a well-defined normal, as on page 138. The lines $x = \pi/2 + m\pi$ and $y = \pi/2 + m\pi$, for $m \in \mathbb{Z}$, divide $\mathbb{R}^2$ into squares, and those where $\cos x \cos y > 0$ form a checkerboard pattern. Since $e^z > 0$, there are clearly no points of M over the

1. P-Q4, N-KB3
2. N-Q2, P-K4
3. P×P, N-N5
4. P-KR3, N-K6
5. Resign

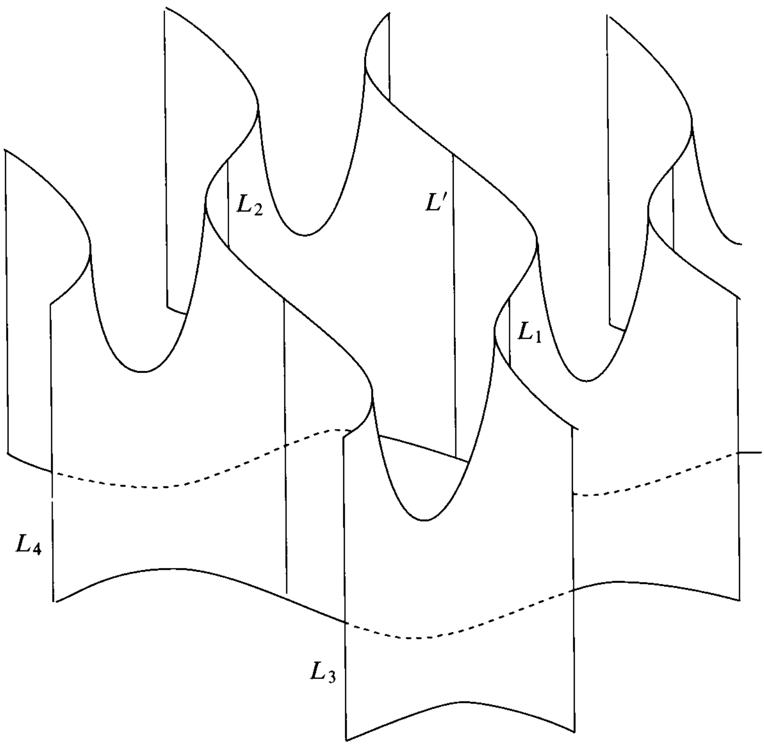


"white" squares. Over the vertices of the squares, where $\cos x = \cos y = 0$, we have perpendicular lines, since z can have any value. Over the "black" squares, we can solve explicitly for z,

$$z = \log\left(\frac{\cos y}{\cos x}\right).$$



The surface is made up of infinitely many of these structural units.



The small portion of the surface pictured above is simply homeomorphic to a cylinder (with a very floppy side), and there is an obvious closed curve c connecting the four saddle points which represents a non-trivial element of the fundamental group. Moreover, this curve c does not disconnect the (complete) surface. To see this, note that just as the vertical line L_1 can be connected to the line L_2 by a curve lying in the two adjacent units which share the common edge L', so lines L_3 and L_4 can be connected by a curve lying in two adjacent units (not drawn in the picture). But L_3 and L_4 can be connected to points which lie on (apparently) opposite sides of the curve c.

Similarly, we easily see that the surface actually has "infinite genus" (infinitely many closed curves may be removed from it without disconnecting it). It also has only one end. But it is known that two orientable surfaces with the same genus and the same space of ends are homeomorphic, so our surface must be homeomorphic to surface (A) in Problem I.1-20.

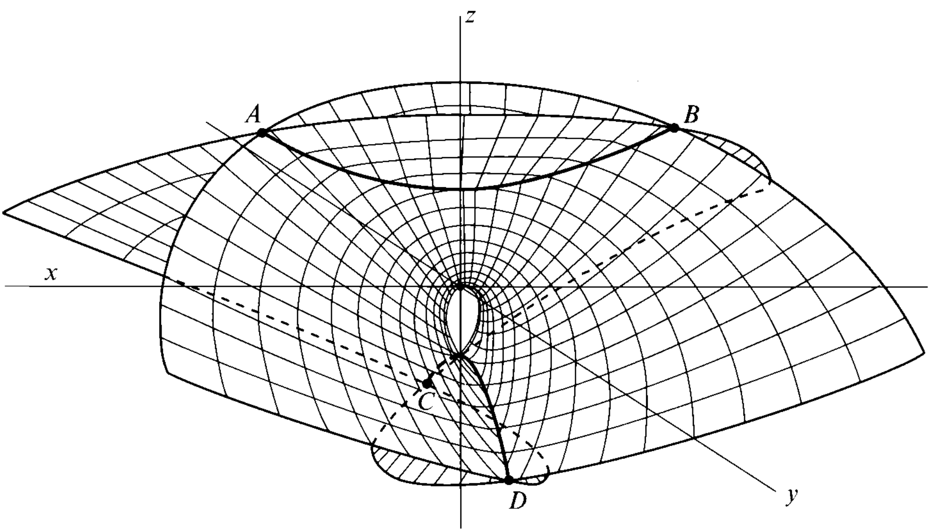
2. *Enneper's Minimal Surface*

This surface is parameterized by

$$f(u,v) = \left(\frac{u}{2} - \frac{u^3}{6} + \frac{uv^2}{2},\ -\frac{v}{2} + \frac{v^3}{6} - \frac{u^2v}{2},\ \frac{u^2}{2} - \frac{v^2}{2}\right).$$

A computation from equations (A) and (B) shows that $H = 0$. Of course, at present it is pretty hard to see how any one ever thought of this example, but in Chapter 9 we will see that in a certain sense it is the simplest minimal surface.

In the figure below, showing the image of f on $[-2.5, 2.5] \times [-2.5, 2.5]$, the top portion is seen almost completely from the side, while the bottom portion

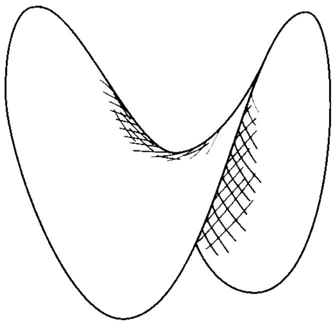


is seen almost head on. The surface is taken into itself by the map

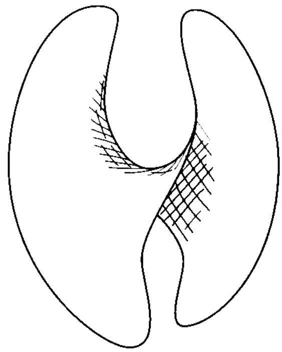
$$(x, y, z) \mapsto (y, x, -z),$$

with the line AB corresponding to the line CD.

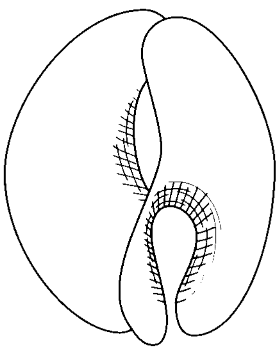
Since the surface is the graph of a function, it is merely an immersed plane, and its structure can perhaps be better understood from the series of pictures on the next page, which show a saddle surface being deformed into a surface of the same type as Enneper's surface.



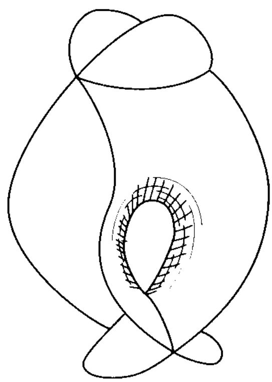
(a)



(b)



(c)

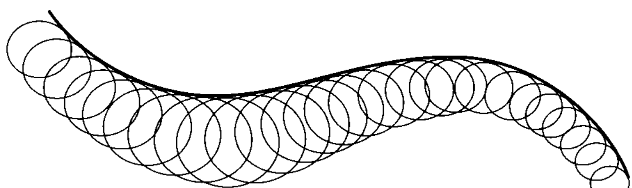


(d)

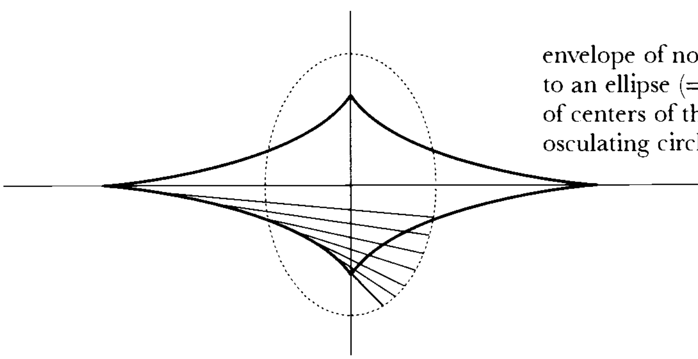
ADDENDUM
ENVELOPES OF 1-PARAMETER FAMILIES OF PLANES

In classical differential geometry, a central role was played by the notion of the envelope of a family of curves or surfaces. A careful treatment of this topic involves many delicate points, which to be sure were rather indelicately handled by classical geometers. However, the study of envelopes played such an important role in the evolution of the concept of a connection that a sketch of its essential features seems in order; the ideas developed here will also be used on a couple of later occasions.

Consider first a 1-parameter family $\bar{\alpha}$ of curves in the plane, given by $\bar{\alpha}(u) = t \mapsto \alpha(u,t)$ for some C^∞ function $\alpha\colon [0,1] \times [0,1] \to \mathbb{R}^2$. An **envelope** of this family is defined to be a curve c which is *not* a member of this family but which is tangent to some member of the family at every point. Unfortunately, it often

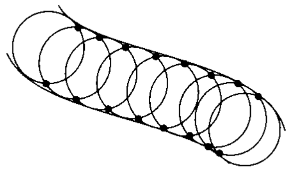


turns out that the envelope of a perfectly nice family of curves has a cusp or something worse; but for the time being we won't worry too much about this.

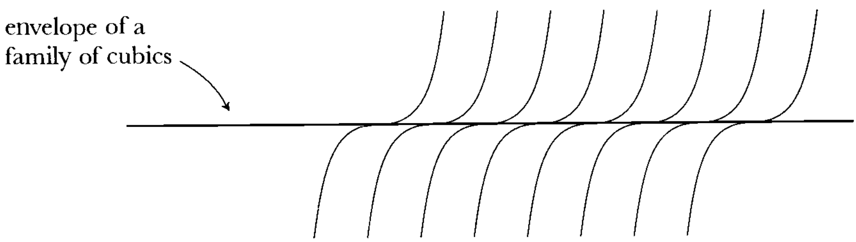


envelope of normals to an ellipse (= locus of centers of the osculating circles)

The classical way of finding the envelope of α was very geometric. For each u, we let $c(u)$ be the limit, as $\varepsilon \to 0$, of the intersection of $\bar{\alpha}(u)$ and $\bar{\alpha}(u+\varepsilon)$: the

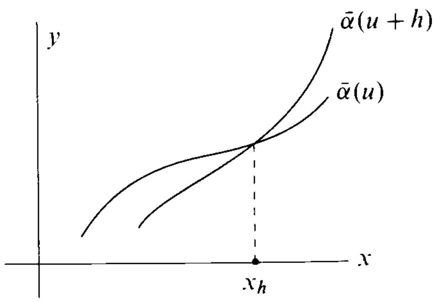


envelope consists of the "intersections of members of the family with another member infinitely close to it". The picture below shows that this idea can run into some serious difficulties. Nevertheless, it often works out rather well in



particular cases, and even in the general case it leads us to the proper analytic condition, when we argue as follows.

Let us consider first the case where our curves $\bar{\alpha}(u)$ are all expressed as the graphs of functions; thus there is a function $(u,x) \mapsto f(u,x)$ such that $\alpha(u,t) = (t, f(u,t))$. Suppose that the curve $\bar{\alpha}(u)$ intersects the curve $\bar{\alpha}(u+h)$ at the point



$$(x_h, f(u, x_h)) = (x_h, f(u + h, x_h)).$$

Then we have

$$0 = \frac{f(u + h, x_h) - f(u, x_h)}{h}.$$

Assuming that x_h approaches a number $x(u)$ as $h \to 0$, we find that $x(u)$ must be a point for which

$$(*) \qquad D_1 f(u, x(u)) = 0.$$

If we find the points $x(u)$ for all u, then the envelope should be the curve consisting of all points $(x(u), f(u, x(u)))$.

If we are given a general family $\bar{\alpha}$, not necessarily expressed as graphs of functions, then we can introduce the function f in two steps. We first determine $t(u, x)$ so that

$$(1) \qquad \alpha^1(u, t(u, x)) = x,$$

and then define

$$(2) \qquad f(u, x) = \alpha^2(u, t(u, x)).$$

Then equation $(*)$ becomes

$$(3) \qquad 0 = D_1\alpha^2(u, t(u, x)) + D_2\alpha^2(u, t(u, x)) \cdot D_1 t(u, x),$$

while equation (1) gives

$$D_1\alpha^1(u, t(u, x)) + D_2\alpha^1(u, t(u, x)) \cdot D_1 t(u, x) = 0,$$
$$D_1 t(u, x) = -\frac{D_1\alpha^1}{D_2\alpha^1}(u, t(u, x)).$$

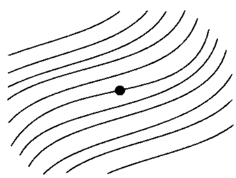
Substituting this into (3), we obtain

$$[D_1\alpha^2 \cdot D_2\alpha^1 - D_1\alpha^1 \cdot D_2\alpha^2](u, t(u, x)) = 0.$$

Thus we find that the envelope should consist of points $\alpha(u, t)$ where (u, t) satisfies

$$(**) \qquad \det(D_i\alpha^j(u, t)) = 0.$$

Now even without resorting to the motivating geometric construction, it is clear that if there is an envelope of the family $\bar{\alpha}$, then it must be a subset of the points $\alpha(u,t)$ for which (u,t) satisfies $(**)$. For, if the determinant in $(**)$ is non-zero, then α is an immersion at (u,t), and the curves $\bar{\alpha}(u)$ form a foliation of a neighborhood of $\alpha(u,t)$; consequently, the only curve through $\alpha(u,t)$ which



is tangent to some curve of the family at each point is $\bar{\alpha}(u)$ itself, which means that $\alpha(u,t)$ cannot be a point of an envelope.

Similar considerations will apply to 1-parameter families of surfaces in space. In particular, the geometric construction will be found quite useful when we consider 1-parameter families of planes. A plane P can be described as

$$\{x \in \mathbb{R}^3 : a_1x_1 + a_2x_2 + a_3x_3 = \langle a, x\rangle = d\}$$

for some number d, and some $(a_1, a_2, a_3) \neq 0$, which we might as well assume is a unit vector. (Choosing the point $(x_1, x_2, x_3) \in P$ closest to 0, and noting that it must be a multiple of a, we see that d is just the distance from 0 to P, provided that a is picked so that it points in the direction of points further from 0). So a 1-parameter family of planes amounts to two functions $a\colon \mathbb{R} \to S^2$ and $d\colon \mathbb{R} \to \mathbb{R}$. Obviously if $a' = 0$, so that a is constant, then we obtain a family of parallel planes, and there is no envelope. Let us assume that, in fact, a' is *never* 0. Then any two nearby planes, corresponding to $u < v$, must intersect in a straight line, and all x on this line satisfy

$$\sum_i a_i(u)x_i - d(u) = \sum_i a_i(v)x_i - d(v) = 0.$$

Applying Rolle's Theorem to

$$s \mapsto \sum_i a_i(u + s[v-u])x_i - d(u + s[v-u]) \qquad \text{on } [0, 1],$$

we find that

$$(1) \qquad \sum_i a_i{}'(\xi)x_i - d'(\xi) = 0$$

for some ξ (depending on x) between u and v.

It follows that as $v \to u$, this line approaches the line satisfying

$$(*) \qquad \begin{cases} \langle a(u), x\rangle = d(u) \\ \langle a'(u), x\rangle = d'(u) \end{cases}$$

(note that $a \in S^2$ implies that a' is perpendicular to a, so the planes given by these two equations are not parallel).

The line determined by $(*)$ is called the **characteristic line** of the plane determined by u; its direction is $a(u) \times a'(u)$. If $(a \times a')' = 0$, so that all characteristic lines are parallel, then the envelope will be the generalized cylinder formed by all these characteristic lines, provided this actually exists (it could happen, for example, that all characteristic lines are the same, in which case no envelope would exist). Let us assume that, in fact, $(a \times a')'$ is *never* 0. Then any three nearby planes, corresponding to $u < v < w$, must have linearly independent $a(u), a(v), a(w)$, so the planes corresponding to u, v, w must intersect at a point (x_1, x_2, x_3). This point must satisfy (1) and an analogous condition for v, w:

$$\sum_i a_i'(\bar{\xi})x_i - d'(\bar{\xi}) = 0$$

for some $\bar{\xi}$ between v and w.

Arguing as before, we see that

$$\sum_i a_i''(\eta)x_i - d''(\eta) = 0$$

for some η between ξ and $\bar{\xi}$.

It follows that as $v, w \to u$, this point approaches the point $c(u)$ satisfying

$$(**) \qquad \begin{cases} \langle a(u), c(u)\rangle = d(u) \\ \langle a'(u), c(u)\rangle = d'(u) \\ \langle a''(u), c(u)\rangle = d''(u). \end{cases}$$

The point $c(u)$ determined by $(**)$ is called the **characteristic point** of the plane determined by u, and lies on its characteristic line. It is possible that $c' = 0$, so that c is a point. In this case, the envelope is the generalized cone formed by all the characteristic lines with c as vertex. Let us assume that, in fact, c' exists and is *never* 0. Differentiating the first two equations of $(**)$ gives

$$\langle a'(u), c(u)\rangle + \langle a(u), c'(u)\rangle = d'(u), \qquad \text{hence} \quad (2) \ \ \langle a(u), c'(u)\rangle = 0$$

$$\langle a''(u), c(u)\rangle + \langle a'(u), c'(u)\rangle = d''(u), \qquad \text{hence} \quad (3) \ \ \langle a'(u), c'(u)\rangle = 0.$$

Differentiating (2) then gives

$$\langle a'(u), c'(u)\rangle + \langle a(u), c''(u)\rangle = 0,$$

which together with (3) gives

$$(4) \qquad \langle a(u), c''(u)\rangle = 0.$$

We thus have:

$c(u)$ is the characteristic point of the plane determined by u
$c'(u)$ has the same direction as the characteristic line (*)
[by (2) and (3)]
$c''(u)$ is in a plane parallel to the plane determined by u
[by (4)].

So *the plane determined by* u *is the osculating plane of* c, *and the tangent developable of* c *is the envelope of the family*. Each plane of the family is tangent to this developable along the points where it intersects it, namely along its characteristic line. (For all this to work, of course, we need c'' to be non-zero.)

To sum things up, a 1-parameter family of planes "in general" has an envelope, which is either a generalized cylinder, a generalized cone, or the tangent developable of a curve. This well-known fact from classical differential geometry was precisely what gave Levi-Civita the clue for defining parallel translation in a Riemannian-manifold. He first observed that since generalized cylinders, generalized cones, and tangent developables are locally isometric to the plane, it makes sense to talk about parallel vector fields in these surfaces—they are just the images, under the local isometry, of parallel vector fields in the plane.

Now suppose that we are given a curve c on an arbitrary surface M. Consider the 1-parameter family of planes formed by the tangent planes $M_{c(u)}$. This family "generally speaking" has an envelope N, which is a generalized cylinder or cone, or the tangent developable to some (other) curve; and the tangent space of N is the same as that of M all along the curve c. So we can define a vector field V_u to be parallel along c in M if it is parallel along c in N. Once Levi-Civita had this definition of parallel vector fields along a curve c in M it was not hard to derive the usual equation for it, in terms of the Christoffel symbols. This equation shows that parallelism does not depend on the imbedding, and can be used to define parallel vector fields along a curve in any Riemannian manifold, of any dimension.

PROBLEMS

1. If $a_1 f_1 + a_2 f_2$ is a principal vector, then

$$\begin{pmatrix} G & -F \\ -F & E \end{pmatrix} \begin{pmatrix} l & m \\ m & n \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \lambda \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

for some λ. Write the two equations which this gives, and divide by a_1 and a_2, respectively, to obtain two expressions for λ. From this derive equation (D) (and then check that it holds also for a_1 or $a_2 = 0$).

2. Let $f\colon V \to \mathbb{R}$ be a linear function on a (possibly infinite-dimensional) vector space V.

(a) If $v_1, v_2 \in V$, then $f(v_1)v_2 - f(v_2)v_1 \in \ker f$.
(b) We can write $V = \ker f \oplus W$, where W is 1-dimensional.
(c) If $g\colon V \to \mathbb{R}$ is a linear function with $\ker f \subset \ker g$, then $g = \lambda f$ for some $\lambda \in \mathbb{R}$.
(d) If $g, f_1, \ldots, f_k\colon V \to \mathbb{R}$ and $\bigcap_i \ker f_i \subset \ker g$, then $g = \sum_i \lambda_i f_i$ for some $\lambda_i \in \mathbb{R}$.

3. Let the Jacobian of $f\colon \mathbb{R}^n \to \mathbb{R}^k$ have rank k on $f^{-1}(0)$, so that $M = f^{-1}(0)$ is a submanifold of $\mathbb{R}^n$ of dimension $n-k$. Let $g\colon M \to \mathbb{R}$ be differentiable, and suppose that g has a maximum at $p \in M$.

(a) $M_p = \bigcap_{i=1}^k \ker df^i$, where $df^i\colon {\mathbb{R}^n}_p \to \mathbb{R}$.
(b) If $X_p \in M_p$ then $dg(X_p) = 0$. *Hint*: $X_p = c'(0)$ for some curve c in M.
(c) Use Problem 2 to conclude that there are $\lambda_1, \ldots, \lambda_k$ with

$$D_j g = \sum_{i=1}^k \lambda_i D_j f^i \qquad \text{for } j = 1, \ldots, n.$$

4. (a) For the ruled surface $f(s,t) = c(s) + t\delta(s)$, show that $m = \langle c', \delta \times \delta' \rangle$, and

$$EG - F^2 = \langle \delta, \delta \rangle \cdot \langle c' + t\delta', c' + t\delta' \rangle - \langle c' + t\delta', \delta \rangle^2.$$

(b) If θ is the angle between δ and $c' + t\delta'$, then

$$\langle \delta, \delta \rangle \cdot \langle c' + t\delta', c' + t\delta' \rangle \cdot \cos^2\theta = \langle \delta, c' + t\delta' \rangle^2.$$

So $EG - F^2 = EG \sin^2\theta = |(c' + t\delta') \times \delta|^2$.
(c) For the standard parameterization, show that

$$\sigma' \times \delta = \langle \sigma', \delta \times \delta' \rangle \cdot \delta',$$

and deduce the formula for K on page 148.

5. (a) Let $a, b \in \mathbb{R}^3$ and let $v, w \in \mathbb{R}^3$ be non-parallel vectors. If $a + t_0 v$ and $b + t_1 w$ are the points on the lines $\{a + tv\}$ and $\{b + tw\}$ which are closest to each other, then the line from $a+t_0 v$ to $b+t_1 w$ must be perpendicular to both v and w. Conclude that

$$t_0 = \frac{\langle w, w\rangle \cdot \langle a-b, v\rangle \; - \; \langle w, v\rangle \cdot \langle a-b, w\rangle}{\langle v, w\rangle^2 \; - \; \langle w, w\rangle \cdot \langle v, v\rangle}.$$

(b) Consider the ruled surface $c(s) + t\delta(s)$ with $|\delta| = 1$. If the point $P(\varepsilon)$ on page 147 is $c(s) + t(\varepsilon)\delta(s)$, then

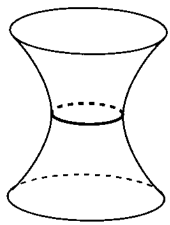
$$t(\varepsilon) = \frac{\langle c(s) - c(s+\varepsilon), \delta(s)\rangle \; - \; \langle \delta(s), \delta(s+\varepsilon)\rangle \cdot \langle c(s) - c(s+\varepsilon), \delta(s+\varepsilon)\rangle}{\langle \delta(s), \delta(s+\varepsilon)\rangle^2 \; - \; 1}.$$

Use L'Hôpital's rule to show that

$$\lim_{\varepsilon \to 0} t(\varepsilon) = \frac{-\langle c'(s), \delta'(s)\rangle}{\langle \delta'(s), \delta'(s)\rangle}.$$

(c) The striction curve of the tangent developable of c is c.
(d) The striction curve of the hyperboloid of revolution

$$\frac{x^2}{a^2} + \frac{y^2}{a^2} - \frac{z^2}{c^2} = 1$$



is the central circle. (Notice that in each case the tangent vector of the striction curve at a point is not perpendicular to the generator through that point, even though the striction curve is the limit of common perpendiculars to generators.)

6. (a) Modify Proposition II.4-14 as follows: If $\langle\ ,\ \rangle$ is any (possibly degenerate) inner product on $\mathbb{R}^n$, then there is a basis of $\mathbb{R}^n$ which is orthogonal for $\langle\ ,\ \rangle$ and orthonormal with respect to the usual inner product.
(b) Consider a quadric $\left\{(x_1, x_2, x_3) : \sum_{ij} a_{ij}x_i x_j + \sum_i b_i x_i + c = 0\right\}$. Show that some rotation of $\mathbb{R}^3$ takes this into a set of the form

$$\left\{(x_1, x_2, x_3) : \sum_i \alpha_i {x_i}^2 + \sum_i \beta_i x_i + \gamma = 0\right\}.$$

(c) Show that a translation can be used to make $\beta_i = 0$ if $\alpha_i \neq 0$. Conclude that the quadric is an ellipsoid or hyperboloid of one or two sheets (or $\emptyset$), if all $\alpha_i \neq 0$; and it is an elliptic or hyperbolic paraboloid if just one $\alpha_i = 0$. Show that all other cases are lines, planes, or cylinders over parabolas, ellipses and hyperbolas.

7. Let $M \subset \mathbb{R}^3$ be a surface with unit normal $\nu: M \to \mathbb{R}^3$. We define the **support function** h of M by

$$h(p) = -\langle p, \nu(p)\rangle.$$

(a) Show that $|h(p)| =$ distance from 0 to M_p, and that $h(p) > 0$ if and only if $\nu(p)$ points toward the side of M_p which contains 0.
(b) For the ellipsoid $W^{-1}(0)$, where $W(x, y, z) = \frac{1}{2}(x^2/a^2 + y^2/b^2 + z^2/c^2 - 1)$, show that

$$h(x, y, z) = \frac{-1}{|Z|} \qquad \text{for } Z = (W_1, W_2, W_3)(x, y, z).$$

Conclude that

$$K = \frac{h^4}{a^2b^2c^2},$$

and locate the points of maximum and minimum curvature.
(c) For the elliptic hyperboloids of one and two sheets, show that

$$K = \frac{-h^4}{a^2b^2c^2} \qquad \text{and} \qquad K = \frac{h^4}{a^2b^2c^2}, \qquad \text{respectively.}$$

8. Let $M \subset \mathbb{R}^3$ be an imbedded surface such that $\nu: M \to S^2$ is one-one. For $\xi \in S^2$, let

$$p(\xi) = h(\nu^{-1}(\xi)),$$

where h is the support function of M.

(a) The tangent plane at $\nu^{-1}(\xi)$ is

$$\left\{x \in \mathbb{R}^3 : \sum_{j=1}^{3} \xi^j x^j = p(\xi)\right\},$$

so

$$\sum_{j=1}^{3} \xi^j [\nu^{-1}(\xi)]^j = p(\xi).$$

(b) For $x \in \mathbb{R}^3 - \{0\}$, let

$$P(x) = |x| \cdot p\left(\frac{x}{|x|}\right).$$

Then

$$\frac{\partial P(x)}{\partial x^i} = \left[\nu^{-1}\left(\frac{x}{|x|}\right)\right]^i + \sum_{j=1}^{3} x^j \cdot \frac{\partial\left[\nu^{-1}\left(\frac{x}{|x|}\right)\right]^j}{\partial x^i} = \left[\nu^{-1}\left(\frac{x}{|x|}\right)\right]^i.$$

Hint: The vanishing of the second term is equivalent to the assertion that $\partial[\nu^{-1}(x/|x|)]/\partial x^i$ is tangent to M at $\nu^{-1}(x/|x|)$. Note that $\nu^{-1}(x/|x|) \in M$ for all x.

9. (a) The determinant on page 152 equals

$$xa_2a_3(c^2 - b^2) + ya_1a_3(a^2 - c^2) + za_1a_2(b^2 - a^2).$$

By sign considerations, show that for $a > b > c > 0$ there is no umbilic with $y \neq 0$. Then show that there are four umbilics, with coordinates

$$x = \pm a\left(\frac{a^2 - b^2}{a^2 - c^2}\right)^{1/2} \qquad z = \pm c\left(\frac{b^2 - c^2}{a^2 - c^2}\right)^{1/2}.$$

(b) Similarly, find the four umbilics on the elliptic hyperboloid of two sheets.

10. Find the umbilics on the elliptic paraboloid by using formulas (E′). [There are two if $a \neq b$, and one if $a = b$.]

11. (a) Let M be a doubly ruled non-flat surface, and choose three mutually skew straight lines L_1, L_2, L_3 from the first family of rulings. Show that there is a unique family of straight lines which intersect all three of L_1, L_2, L_3.
(b) Show that three mutually skew lines L_1, L_2, L_3 lie on some quadric, and conclude that M is this quadric.

12. Let $M \subset \mathbb{R}^3$ be a surface with normal map ν. Then $\{p + \varepsilon\nu(p) : p \in M\}$ is called a **parallel surface** $\bar{M}$ of M. We have a map $f\colon M \to \bar{M}$ given by $f(p) = p + \varepsilon\nu(p)$.

(a) If X is a tangent vector of M, then $f_*X = X + \varepsilon d\nu(X)$ [identifying tangent vectors with elements of $\mathbb{R}^3$ as usual]. Hence f is an immersion if $\varepsilon \neq 1/k_i$ for either principal curvature k_i at any point p of M. In particular, if M is compact, then $\bar{M}$ is a surface for small enough ε. (One could also use Theorem I.9-20.)
(b) The normal $\bar{\nu}$ at $p + \varepsilon\nu(p)$ is just $\nu(p)$. (One could also use a generalization of Problem I.9-28.)
(c) The principal curvatures of $\bar{M}$ are

$$\frac{k_i}{1 + \varepsilon k_i},$$

so the Gaussian and mean curvatures of $\bar{M}$ are

$$\bar{K} = \frac{K}{1 + \varepsilon H + \varepsilon^2 K} \qquad \bar{H} = \frac{H + 2\varepsilon K}{1 + \varepsilon H + \varepsilon^2 K}.$$

(d) If M has constant Gaussian curvature $K > 0$, then some parallel surface has constant mean curvature, and if M has constant mean curvature $H \neq 0$, then some parallel surface has constant Gaussian curvature (Bonnet).

(e) The volume element $d\bar{A}$ of $\bar{M}$ is related to the volume element dA of M by

$$f^*(d\bar{A}) = (1 + 2\varepsilon H + \varepsilon^2 K)\, dA.$$

(f) If M has mean curvature $H = 0$, and M is not part of a plane, then the area of $\bar{M}$ is *smaller* than the area of M (Steiner).

13. Let c be an arclength parameterized curve. Recall that the "rectifying plane" of c is spanned by the tangent $\mathbf{t}$ and binomial $\mathbf{b}$. Suppose that the family of rectifying planes has an envelope M. Show that c is a geodesic on M. (This is the reason for the word rectifying—the curve c is "made straight" or "rectified" on M.)

CHAPTER 4

CURVES ON SURFACES

In classical surface theory, a great deal of emphasis was placed on special curves lying within a surface. In addition, several new invariants can be defined for a curve c on a surface, apart from the curvature κ and torsion τ which c has as a curve in $\mathbb{R}^3$. The total corpus of accumulated results exhibits—to me at least—the unappealing weightiness of a massive treatise on the conic sections. So in this chapter we will give the definitions and then explore only a few of the pertinent results, concentrating on those which are of importance later on.

Let $M \subset \mathbb{R}^3$ be an oriented surface with corresponding unit normal field ν, and let c be an arclength parameterized curve in M. Then we can consider the normal and tangential components of $c''(s)$,

$$\perp c''(s) = \langle c''(s), \nu(c(s))\rangle \cdot \nu(c(s)),$$

$$\mathsf{T}c''(s) = \frac{D}{ds}c'(s), \qquad \text{by Corollary 1-2.}$$

The normal component $\perp c''(s)$ is sometimes called the **normal curvature vector** of c at s, and

(1) $$\kappa_n(s) = \langle c''(s), \nu(c(s))\rangle$$

is called the **normal curvature** of c at s; it is the signed length of the normal curvature vector. As we mentioned in Chapter 1, the tangential component $\mathsf{T}c''(s)$ is called the **geodesic curvature vector**. Using the orientation of M we can assign a sign to the length of this vector. To do this, we first choose the unit vector $\mathbf{u}(s) \in M_{c(s)}$ perpendicular to $c'(s)$ for which $(c'(s), \mathbf{u}(s))$ is positively oriented, so that

$$c'(s) \times \mathbf{u}(s) = \nu(c(s)).$$

Then we note that

$$\langle c'(s), c'(s)\rangle = 1 \implies \left\langle \frac{D}{ds}c'(s), c'(s)\right\rangle = 0,$$

so that $\mathsf{T}c''(s) = D/ds(c'(s))$ must be a multiple of $\mathbf{u}(s)$. So we can define the **geodesic curvature** $\kappa_g(s)$ of c at s by

(2) $$\begin{aligned}\mathsf{T}c''(s) &= \kappa_g(s) \cdot \mathbf{u}(s)\\ &= \kappa_g(s) \cdot \nu(c(s)) \times c'(s).\end{aligned}$$

Thus κ_g is the signed length of $\mathsf{T}c''$. Obviously we have

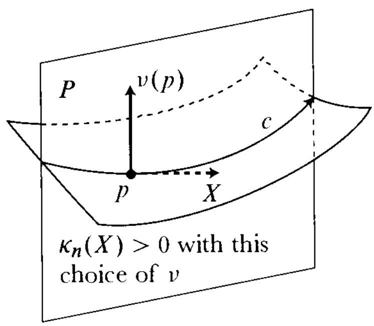
$$(3) \qquad \kappa = \sqrt{\kappa_n{}^2 + \kappa_g{}^2}.$$

We will see that κ_n and κ_g have quite different properties.

We begin by recalling a few facts from Chapter II.3B. The equation $0 = \langle c'(s), \nu(c(s))\rangle$ implies that

$$(4) \quad \kappa_n(s) = \langle c''(s), \nu(c(s))\rangle = -\left\langle \frac{d\nu(c(s))}{ds}, c'(s)\right\rangle = -\langle d\nu(c'(s)), c'(s)\rangle$$
$$= \mathrm{II}(c'(s), c'(s)).$$

Thus $\kappa_n(s)$ *depends only on the direction* ($c'(s)$ or $-c'(s)$) of c at s, and otherwise not on the curve c itself, so we can write $\kappa_n(X)$ for a unit vector X. Now for a given unit vector $X \in M_p$, there is a natural choice for a curve c in M with $c'(0) = X$, namely the arclength parameterized curve which is cut out on M by the plane P containing $\nu(p)$ and X. Then $\mathrm{II}(X, X) = \kappa_n(X)$ is

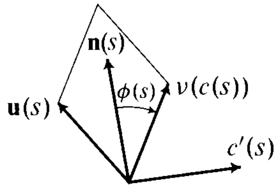


the signed curvature of this curve, when $(X, \nu(p))$ is chosen as the positive orientation for P. If $k_i = \kappa_n(X_i)$ are the minimum and maximum of these signed curvatures, so that X_i are eigenvectors of $-d\nu$, with eigenvalues k_i, then X_1 and X_2 are orthogonal, and if $X = (\cos\theta)X_1 + (\sin\theta)X_2$ is any unit vector, then (Euler's Theorem)

$$(5) \qquad \kappa_n(X) = k_1 \cos^2\theta + k_2 \sin^2\theta.$$

Equation (4) can just as well be used to relate κ_n and κ for any curve c in M. Note that the normal $\mathbf{n}(s)$ of c at s is in the plane spanned by $\mathbf{u}(s)$ and $\nu(c(s))$.

Choosing $(\mathbf{u}(s), \nu(c(s)))$ as the positive orientation of this plane, we define $\phi(s)$



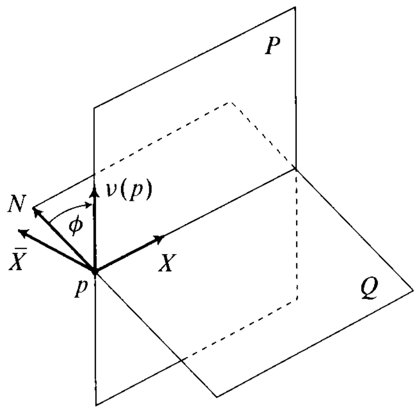
to be the oriented angle from $\mathbf{n}(s)$ to $\nu(c(s))$, so that

(6) $$\mathbf{n}(s) = \sin\phi(s) \cdot \mathbf{u}(s) + \cos\phi(s) \cdot \nu(c(s)).$$

Then equation (4) implies that

(7) $$\kappa_n(s) = \mathrm{II}(c'(s), c'(s)) = \kappa(s) \cdot \cos\phi(s),$$

where κ (> 0) is the curvature of c. (If $\kappa(s) = 0$, then $\mathbf{n}(s)$ is undefined, so $\phi(s)$ is also; but equation (7) then holds with any choice of $\phi(s)$.) In particular, suppose that Q is any plane containing a unit vector $X \in M_p$. Let $\bar{X} \in M_p$ be the unit vector perpendicular to X for which $(X, \bar{X})$ is positively oriented, and



let $N \in \mathbb{R}^3{}_p$ be a unit vector in Q which is perpendicular to X. Choose (X, N) as the positive orientation for Q, and let ϕ be the oriented angle from N to $\nu(p)$ when $(\bar{X}, \nu(p))$ is chosen as the positive orientation for the plane perpendicular to X (changing N to $-N$ reverses the orientation of Q, and changes ϕ to $\phi - \pi$). If c_ϕ is the arclength parameterized curve cut out on M by Q, then its curvature κ_ϕ is given by

(7′) $$\kappa_\phi \cdot \cos\phi = \kappa(X).$$

Equation (7) or (7′) is known as *Meusnier's Theorem*; another formulation of this theorem appears in Problem 1.

We can already state a result which is trivial, but which we will need to refer to later on.

1. PROPOSITION. Let c be an arclength parameterized curve in a surface $M \subset \mathbb{R}^3$, with $c(0) = p$, and $c'(0) = X \in M_p$.

(1) If X is an asymptotic vector, then either $\kappa(0) = 0$ or $\mathbf{n}(0)$ is perpendicular to $\nu(p)$.

(2) If X is not an asymptotic vector, then c cannot have curvature 0 at p, nor can $\mathbf{n}(0)$ be perpendicular to $\nu(p)$; if $\kappa_n(X) > 0$, then the angle between $\mathbf{n}(0)$ and $\nu(p)$ is acute, and if $\kappa_n(X) < 0$ it is obtuse (in the picture on page 188 the angle is 0).

PROOF. Everything follows from equation (4). ❖

Note, finally, that if c is not parameterized by arclength, then

(8) $$\kappa_n(t) = \frac{\mathrm{II}(c'(t), c'(t))}{|c'(t)|^2},$$

since both numerator and denominator are multiplied by the same number under change of parameter.

Now consider the geodesic curvature $\kappa_g(s)$ of c at s, given by equation (2),

$$\mathsf{T}c''(s) = \kappa_g(s) \cdot \mathbf{u}(s).$$

Since we also have

$$c'' = \kappa \cdot \mathbf{n},$$

equation (6) immediately implies that

(9) $$\kappa_g = \kappa \cdot \sin\phi.$$

(Again, this equation holds for any choice of ϕ when $\kappa = 0$.) Unlike κ_n, the quantity κ_g = signed length of $D/ds(c'(s))$ is *intrinsic*—it can be calculated directly from E, F, G using the formula on pg. II.232; one has to be careful to parameterize by arclength. On the other hand, $\kappa_g(s)$ does not depend only on c'. Indeed, κ_g is identically zero if c is a geodesic, and there are geodesics with arbitrary unit tangent vectors at any point.

Moving frame freaks will be happy to learn that κ_n and κ_g arise naturally when one chooses the appropriate moving frame along c. In fact, the original use of the moving frame was by Darboux, whose monumental 4 volume work on surfaces includes incredibly detailed investigations of curves on surfaces. Given an arclength parameterized curve c in $M \subset \mathbb{R}^3$, we define the **Darboux frame** of c on M to be the moving frame

$$\mathbf{t}(s) = c'(s), \qquad \mathbf{u}(s), \qquad \mathbf{v}(s) = \mathbf{t}(s) \times \mathbf{u}(s) = \nu(c(s)),$$

as opposed to the **Frenet frame t, n, b**. The Darboux frame is defined at all points of c, even those where $\kappa = 0$. Since the Darboux frame is also orthonormal, the expression for $(\mathbf{t}, \mathbf{u}, \mathbf{v})'$ is given by a skew-symmetric matrix times $(\mathbf{t}, \mathbf{u}, \mathbf{v})$:

$$\begin{aligned} \mathbf{t}' &= \qquad\quad \kappa_g \mathbf{u} + \kappa_n \mathbf{v} \\ \mathbf{u}' &= -\kappa_g \mathbf{t} \qquad\quad + \tau_g \mathbf{v} \\ \mathbf{v}' &= -\kappa_n \mathbf{t} - \tau_g \mathbf{u} \end{aligned} \tag{10}$$

From the moving frame point of view, the functions $\kappa_g, \kappa_n, \tau_g$ appearing here are defined by these equations, although it is clear from the first equation that κ_g and κ_n are the same as previously defined. The function τ_g is called the **geodesic torsion** of c. We proceed to indicate how these functions are analyzed using moving frames.

We first observe that $\mathbf{v}'(0)$ depends only on $\mathbf{t}(0)$, since

$$\mathbf{v}'(0) = \left.\frac{d\nu(c(s))}{ds}\right|_{s=0} = d\nu(\mathbf{t}(0)). \tag{11}$$

The third equation in (10) then shows immediately that κ_n and τ_g depend only on $\mathbf{t}$, so that we can write $\kappa_n(X)$ and $\tau_g(X)$ for unit vectors X. In fact, if $\bar{X} \in M_p$ is the unit vector perpendicular to X with $(X, \bar{X})$ positively oriented, then equations (10) and (11) give

$$\begin{aligned} \kappa_n(X) &= -\langle d\nu(X), X\rangle = \mathrm{II}(X, X) \\ \tau_g(X) &= -\langle d\nu(X), \bar{X}\rangle = \mathrm{II}(X, \bar{X}). \end{aligned} \tag{12}$$

Now let $X_1, X_2 \in M_p$ be principal directions with (X_1, X_2) positively oriented, and let k_1, k_2 be the corresponding principal curvatures. If θ is the oriented angle from X_1 to a unit vector $X \in M_p$, then we have

$$\begin{aligned} \kappa_n(X) &= -\langle d\nu(X), X\rangle \\ &= \langle k_1(\cos\theta)X_1 + k_2(\sin\theta)X_2,\ (\cos\theta)X_1 + (\sin\theta)X_2\rangle \\ &= k_1 \cos^2\theta + k_2 \sin^2\theta. \end{aligned}$$

Here we have merely rederived Euler's Theorem. But exactly the same procedure gives an explicit expression for τ_g:

2. PROPOSITION. Let $X_1, X_2 \in M_p$ be principal directions, with (X_1, X_2) positively oriented, and let k_1, k_2 be the corresponding principal curvatures at p. If θ is the oriented angle from X_1 to a unit vector $X \in M_p$, then

$$\tau_g(X) = (k_2 - k_1)\sin\theta\cos\theta.$$

PROOF. Equation (12) gives

$$\begin{aligned}\tau_g(X) &= -\langle d\nu(X), \bar{X}\rangle \\ &= \langle k_1(\cos\theta)X_1 + k_2(\sin\theta)X_2,\ -(\sin\theta)X_1 + (\cos\theta)X_2\rangle \\ &= (k_2 - k_1)\sin\theta\cos\theta. \ ❖\end{aligned}$$

Now let ϕ be the oriented angle from $\mathbf{n}$ to $\mathbf{v}$, as on page 189, so that

$$\begin{aligned}\mathbf{n} &= \sin\phi\cdot\mathbf{u} + \cos\phi\cdot\mathbf{v} \\ \mathbf{b} &= -\cos\phi\cdot\mathbf{u} + \sin\phi\cdot\mathbf{v},\end{aligned}$$

and hence

$$\begin{aligned}\mathbf{u} &= \sin\phi\cdot\mathbf{n} - \cos\phi\cdot\mathbf{b} \\ \mathbf{v} &= \cos\phi\cdot\mathbf{n} + \sin\phi\cdot\mathbf{b}.\end{aligned}$$

Using the first equation in (10), and the Serret-Frenet formulas, we have

$$\begin{aligned}\kappa_g &= \langle\mathbf{u}, \mathbf{t}'\rangle = \langle(\sin\phi)\mathbf{n} - (\cos\phi)\mathbf{b}, \kappa\mathbf{n}\rangle \\ &= \kappa\cdot\sin\phi \\ \kappa_n &= \langle\mathbf{v}, \mathbf{t}'\rangle = \langle(\cos\phi)\mathbf{n} + (\sin\phi)\mathbf{b}, \kappa\mathbf{n}\rangle \\ &= \kappa\cdot\cos\phi,\end{aligned}$$

as before. (If $\kappa = 0$, then ϕ is undefined, but it follows immediately from the first equation of (10) that also $\kappa_g = \kappa_n = 0$.) We still have to make use of the second equation in (10); it will give us the geometric interpretation of τ_g. Now we have

$$\tau_g = \langle\mathbf{v}, \mathbf{u}'\rangle = \left\langle(\cos\phi)\mathbf{n} + (\sin\phi)\mathbf{b},\ \frac{d}{ds}[(\sin\phi)\mathbf{n} - (\cos\phi)\mathbf{b}]\right\rangle;$$

using the Serret-Frenet formulas

$$\frac{d\mathbf{n}}{ds} = \mathbf{n}' = -\kappa\mathbf{t} + \tau\mathbf{b} \qquad \text{and} \qquad \frac{d\mathbf{b}}{ds} = -\tau\mathbf{n},$$

we end up with

$$\text{(13)} \qquad \tau_g = \tau + \frac{d\phi}{ds}, \qquad \text{whenever } \tau \text{ is defined.}$$

In particular we have

3. PROPOSITION.
(a) If $X \in M_p$ is a unit vector, then $\tau_g(X)$ is the torsion $\tau(0)$ of the geodesic γ with $\gamma'(0) = X$.

(b) The geodesics pointing in a principal direction at a point have torsion 0 at that point; in particular, all geodesics have torsion 0 at an umbilic.

(c) If $X, Y \in M_p$ are perpendicular unit vectors, then $\tau_g(X) = -\tau_g(Y)$; thus orthogonal geodesics through a point have torsions at that point which are negatives of each other.

Remark: These statements hold only with the additional *proviso* that the torsions in question exist. Problem 3 considers what happens otherwise.

PROOF. Part (a) follows from equation (13), since for the geodesic γ we have $\phi = 0$ or $\phi = \pi$ for all s. Then parts (b) and (c) follow from Proposition 2. ❖

We note in passing that changing the direction of a curve c in M changes $\mathbf{t}$ to $-\mathbf{t}$, and $\mathbf{u}$ to $-\mathbf{u}$, but leaves $\mathbf{v}$ fixed. So equations (10) show that κ_g changes sign, while κ_n and τ_g remain the same. For κ_n this follows from equation (7), since $\mathbf{n}$ is changed to $-\mathbf{n}$, so ϕ is changed to $\phi - \pi$. It also follows from the interpretation of $\kappa_n(X)$ as the signed curvature of the curve cut out on M by the plane P through X and $\nu(p)$—normally, reversing the curve would change the curvature, but in this case, since we change X to $-X$, we also change the orientation of P. The fact that τ_g remains the same follows from the fact that reversing the direction of a curve does not change its torsion.

One other fact about our new invariants will be of interest. An old theorem of Laguerre says that, like κ_n and τ_g, the quantity

$$\frac{d\kappa_n(s)}{ds} - 2\tau_g(s)\kappa_g(s)$$

also depends only on $c'(s)$; using equations (7), (9), and (13), this quantity can be written in Laguerre's formulation, involving ϕ, κ, and τ, as

$$\frac{d\kappa(s)}{ds}\cos\phi(s) - \left(3\frac{d\phi(s)}{ds} + 2\tau(s)\right)\kappa(s)\sin\phi(s),$$

which makes the result seem even more mysterious. Élie Cartan observed that just as κ_n and τ_g can be expressed in terms of the tensor II, this new expression, and others like it, can be expressed in terms of the *covariant derivatives* of II. Recall that by Corollary II.6-5 the tensor $(\nabla_{Z_p}\mathrm{II})(X_p, Y_p) = (\nabla\mathrm{II})(X_p, Y_p, Z_p)$ satisfies

$$(\nabla_{Z_p}\mathrm{II})(X_p, Y_p) = Z_p(\mathrm{II}(X, Y)) - \mathrm{II}(\nabla_{Z_p}X, Y_p) - \mathrm{II}(X_p, \nabla_{Z_p}Y),$$

for all vector fields X, Y, Z extending X_p, Y_p, Z_p. This immediately yields

4. PROPOSITION. For all arclength parameterized curves c in $M \subset \mathbb{R}^3$ with the same tangent vector $c'(0) \in M_p$, the quantity

$$\kappa_n{}'(s) - 2\tau_g(s)\kappa_g(s)$$

has the same value at $s = 0$. The same is true for

$$\tau_g{}'(s) + 2[\kappa_n(s) - H(c(s))]\kappa_g(s).$$

PROOF. Let X be a unit vector field on M which extends $\mathbf{t}$, and let $\bar{X}$ be the perpendicular unit vector field with $(X, \bar{X})$ positively oriented. Then equation (2), and the fact that $\mathsf{T}c''(s) = D/ds(c'(s))$, shows that

$$\nabla_X X = \kappa_g \cdot \bar{X} \qquad \text{along } c,$$

so by equation (12) we have

$$\mathrm{II}(\nabla_X X, X) = \kappa_g \cdot \tau_g \qquad \text{along } c.$$

Hence

$$\begin{aligned}(\nabla_X \mathrm{II})(X, X) &= X(\mathrm{II}(X, X)) - 2\mathrm{II}(\nabla_X X, X)\\ &= \kappa_n{}' - 2\kappa_g\tau_g \qquad \text{along } c.\end{aligned}$$

This shows that the first expression depends only on $X = c'$.

We also have

$$\begin{aligned}\langle X, \bar{X}\rangle = 0 &\Longrightarrow \langle\nabla_X \bar{X}, X\rangle = -\langle\bar{X}, \nabla_X X\rangle = -\kappa_g,\\ \langle\bar{X}, \bar{X}\rangle = 1 &\Longrightarrow \langle\nabla_X \bar{X}, \bar{X}\rangle = 0,\end{aligned}$$

which implies that

$$\nabla_X \bar{X} = -\kappa_g X \qquad \text{along } c.$$

So we obtain

$$\begin{aligned}(\nabla_X \mathrm{II})(X,\bar{X}) &= X(\mathrm{II}(X,\bar{X})) - \mathrm{II}(\nabla_X X,\bar{X}) - \mathrm{II}(X,\nabla_X \bar{X}) \\ &= \tau_g{}' - \kappa_g \mathrm{II}(\bar{X},\bar{X}) + \kappa_g \mathrm{II}(X,X) \\ &= \tau_g{}' + \{2\mathrm{II}(X,X) - [\mathrm{II}(X,X) + \mathrm{II}(\bar{X},\bar{X})]\}\kappa_g \\ &= \tau_g{}' + 2(\kappa_n - H)\kappa_g \qquad \text{along } c. \ ❖\end{aligned}$$

We are now ready to consider the three main classes of curves on a surface $M \subset \mathbb{R}^3$. The curve c is called a **line of curvature** (or **principal curve**) if c' always points along a principal direction. This means that

$$-d\nu(c') = k \cdot c'$$

for some function k, where $k(t)$ must be a principal curvature at $c(t)$. We can also write this as

$$\frac{-d\nu(c(t))}{dt} = k(t)\frac{dc}{dt} \qquad \text{or} \qquad \frac{d\nu}{dt} + k\frac{dc}{dt} = 0, \quad \text{in the classical manner.}$$

Oddly enough, the last equation has a special name, **Rodrigues' formula**. A more interesting characterization of principal curves can be given in terms of one of our invariants—the third equation in (10), together with (11), shows that c is a line of curvature if and only if τ_g is identically zero.

A curve c in M is called an **asymptotic curve** if c' always points along an asymptotic direction. Equation (4) [or (8)] shows that c is an asymptotic curve if and only if $\mathbf{n}(t)$ is perpendicular to $\nu(c(t))$ at all points t where $\kappa(t) \neq 0$. Equivalently, c is an asymptotic curve if and only if $\mathbf{n}(t)$ lies in $M_{c(t)}$ whenever $\kappa(t) \neq 0$, or yet again, if and only if the osculating plane of c at t coincides with $M_{c(t)}$ whenever $\kappa(t) \neq 0$. Equation (4), or the third equation in (10), together with (11), shows that c is an asymptotic curve if and only if κ_n is identically zero. Equation (3), or the first equation of (10), then shows that c is an asymptotic curve if and only if $\kappa = \pm\kappa_g$ everywhere. Moreover, at points where $\kappa \neq 0$, we then have $\mathbf{n} = \pm\mathbf{u}$, so $\mathbf{b} = \pm\mathbf{v}$ (the same sign holding in both cases), and the third equation of (10) shows that $\tau = \tau_g$; this also follows from equation (13), since $\phi = \pi/2$ or $3\pi/2$ for all t.

Finally, we have the geodesics, which may be defined as curves c with $c''(t)$ always perpendicular to $M_{c(t)}$, so that $\mathbf{n}$ is perpendicular to M at points where $\kappa \neq 0$, or as curves with κ_g identically zero.

To summarize, we have

$$
\begin{aligned}
c \text{ is a line of curvature} &\iff -d\nu(c') = k \cdot c' \\
&\iff \tau_g = 0 \\
c \text{ is an asymptotic curve} &\iff \mathbf{n}, \begin{array}{l}\text{when defined, is}\\ \text{always tangent to } M\end{array} \\
&\iff \begin{array}{l}\text{the osculating plane, when defined,}\\ \text{always coincides with the}\\ \text{tangent plane of } M\end{array} \\
&\iff \kappa_n = 0 \\
&\iff \kappa = \pm\kappa_g \\
&\implies \tau = \tau_g, \text{ when } \tau \text{ is defined} \\
c \text{ is a geodesic} &\iff \mathbf{n}, \begin{array}{l}\text{when defined, is}\\ \text{always perpendicular to } M\end{array} \\
&\iff \kappa_g = 0 \\
&\implies \tau = \tau_g, \text{ when } \tau \text{ is defined.}
\end{aligned}
$$

Since κ_n and τ_g actually depend only on the direction of c at a point, it also makes sense to talk about a curve c being "asymptotic at t" ($\kappa_n(t) = 0$) or "principal at t" ($\tau_g(t) = 0$); this just means that $c'(t)$ points in an asymptotic direction or in a principal direction. The equivalences given above for lines of curvature and asymptotic curves can all be replaced by corresponding equivalences for curves which are principal at a point or asymptotic at a point; however the conclusion $\tau(t) = \tau_g(t)$ does not follow from the mere assumption that c is asymptotic at t.

We will begin our study of these special curves by considering some very general properties. Taking the asymptotic curves first, we note that they can exist only in regions where $K \leq 0$. This already leads to another simple

5. PROPOSITION. A straight line on a surface is an asymptotic curve, so the curvature K of the surface satisfies $K \leq 0$ along any straight line lying in it. The curvature is everywhere 0 along the straight line if and only if the normal ν is constant along the line (equivalently: if and only if the tangent space is parallel along the line).

PROOF. The first assertion follows immediately from equation (4). To prove the second, let c be any parameterization of the straight line (with $c'(t)$ always

$\neq 0$). If the normal is constant along c, then

$$0 = \frac{d\nu(c(t))}{dt} = d\nu(c'(t)),$$

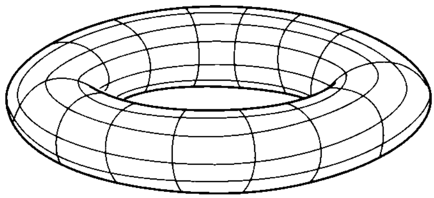
and this can happen only if $d\nu$ has determinant $K = 0$. Conversely, suppose $K = 0$ at each point $c(t)$, so that $d\nu$ has at least one eigenvalue 0 at $c(t)$. If both eigenvalues are 0, then certainly $d\nu(c'(t)) = 0$. On the other hand, it is easy to see that if one eigenvalue is non-zero, then the only asymptotic vectors are multiples of the eigenvector with eigenvalue 0; in other words, the asymptotic vector $c'(t)$ must satisfy $d\nu(c'(t)) = 0$. ❖

Remark: Actually, we have proved a more general result a long time ago (Corollary 1-7), but I thought it would be nice to include the classical proof also. As before, we can state a slightly more precise result: $K = 0$ at a point of a straight line if and only if at this point the normal ν has derivative 0 along the line.

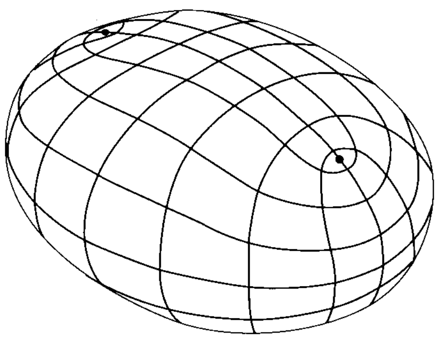
As an immediate corollary of Proposition 5, notice that a ruled surface must have $K \leq 0$ everywhere, as we found by computation in Chapter 3. We also see that $K = 0$ everywhere on a ruled surface if and only if the normal ν is constant on each generator. On page 146 we found that the tangent plane at $f(s,t)$ is spanned by $c'(s) + t\delta'(s)$ and $\delta(s)$. This is independent of t if and only if $\delta(s), \delta'(s), c'(s)$ are linearly dependent. Our formula for K on page 147 shows that this is indeed true if and only if $K = 0$ everywhere. Classically, the ruled surfaces with $K = 0$ everywhere, i.e., the ruled surfaces with constant normals along each generator, were called **developable surfaces**; in the next chapter we will consider them in greater detail.

In a small region of a surface where $K < 0$, we can choose two linearly independent asymptotic unit vectors X_p, Y_p at each point p. It is easy to show that $p \mapsto X_p$ and $p \mapsto Y_p$ will be C^∞ vector fields; the asymptotic curves are just the integral curves of these C^∞ vector fields—as one approaches a parabolic or planar point these integral curves can run together in complicated ways. In Chapter 2 we pointed out that the asymptotic directions are perpendicular precisely when the mean curvature $H = 0$. So on a minimal surface without planar points, the asymptotic curves are everywhere orthogonal. This is illustrated on the right helicoid by the rulings and the helices.

In contrast to the asymptotic curves, lines of curvature can exist in regions of any sort, and it is only umbilics which cause problems. In a small region free of umbilics, we can choose two linearly independent principal unit vectors X_p, Y_p at each point p. As in the case of asymptotic directions, it is easy to show that $p \mapsto X_p$ and $p \mapsto Y_p$ are C^∞ vector fields (on regions where $K < 0$ this is also a consequence of the fact that the principal directions bisect the asymptotic directions); the lines of curvature are the integral curves of these vector fields. We have already seen that on a torus of revolution, with no umbilics, the

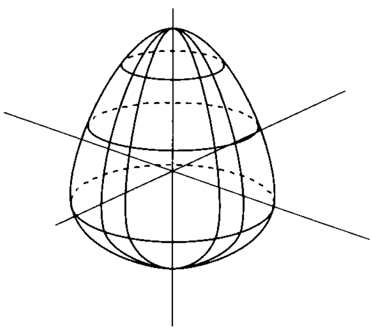


lines of curvature are the parallels and meridians. By contrast, the following famous picture shows how the lines of curvature behave in a neighborhood of the umbilic points of an ellipsoid.



Notice that if a surface M has no umbilics, like the torus, then there is a C^∞ 1-dimensional distribution on M—at each point p we choose the set of all vectors in M_p which are eigenvalues for the larger principal curvature, say. Equivalently, we can pick out two units vectors $X_p, -X_p \in M_p$. Now we can construct a 2-fold covering space $\pi\colon \widetilde{M} \to M$ by choosing the two points in $\pi^{-1}(p)$ to correspond to these two vectors. There is then an obvious nowhere zero vector field $\widetilde{X}$ on $\widetilde{M}$ (if $q \in \pi^{-1}(p)$ is the point corresponding to $X_p \in M_p$, then $\pi_*\widetilde{X}_q = X_p$). On the other hand, Theorem I.11-30 tells us that for the compact orientable surface $\widetilde{M}$, this can happen only when the Euler characteristic $\chi(\widetilde{M}) = 0$, which implies that $\chi(M) = 0$, since $\chi(\widetilde{M}) = 2\chi(M)$ (Problem 7). But the torus is the *only* compact orientable surface with Euler characteristic $= 0$ (Problem I.11-2(c)). So *any compact surface in* $\mathbb{R}^3$ *not homeomorphic to the torus must have at least one umbilic.*

Carathéodory conjectured that every compact convex surface in $\mathbb{R}^3$ must have at least *two* umbilics. Weird as this may seem, there is a natural way to try to prove it. On any compact surface $M \subset \mathbb{R}^3$ with only finitely many umbilics $p_1, \ldots, p_k$ we can choose a C^∞ 1-dimensional distribution on $M - \{p_1, \ldots, p_k\}$. There is a way of defining the index of this distribution at each point p_i, in much the same way that we defined the index of a vector field in Chapter I.11, except that now the index can take on half-integer values. Moreover, it turns out that the sum of the indices of the distribution is again the Euler characteristic $\chi(M)$. The precise definition of the index, and the proof of this result are given in Addendum 2. Now $\chi(S^2) = 2$, so Carathéodory's conjecture could be proved by showing that at an umbilic the index of our particular distribution cannot be equal to 2. For the analytic case, Hamburger [1] gave a proof of this which is 183 pages long! Bol [1] then gave a proof that is only 22 pages long, although it requires a correction (Klotz [1]). After all this work, it still seems that nothing is known when the surface is not analytic, or when it is not convex, even if it is homeomorphic to S^2. I also know of no example where there are only two umbilics. On compact surfaces of revolution the lines of curvature have only two singularities, at the two poles, but I suspect that there will always be at least one whole parallel of umbilics in addition.



Finally, we have the geodesics. They, of course, not only exist in any sort of region, but can be found in any direction. On the other hand, just as the asymptotic lines intersect orthogonally only when $H = 0$, orthogonality of two families of geodesics implies that $K = 0$; in fact, even more is true:

6. PROPOSITION. If two families of geodesics intersect at a constant angle everywhere on M, then M is flat.

PROOF. Let X [or Y] be the vector field of unit tangents to the curves of the first [or second] family. Then X is parallel along the integral curves of X.

Since the angle between X and Y is constant, and since we are on a surface, the vector field Y must also be parallel along the integral curves of X. The same argument holds with X and Y interchanged. We therefore have

$$0 = \nabla_X X = \nabla_Y Y = \nabla_X Y = \nabla_Y X = [X, Y].$$

Consequently,

$$R(X, Y)Y = \nabla_X(\nabla_Y Y) - \nabla_Y(\nabla_X Y) - \nabla_{[X,Y]}Y = 0. \text{ ❖}$$

Proposition 6, of course, is not really a theorem about surfaces in $\mathbb{R}^3$ at all—it is actually a theorem about the intrinsic geometry of surfaces (I do not know whether any analogue holds for higher dimensional manifolds).

With this very general discussion of the behavior of our three classes of curves out of the way, we proceed to the main results about each class. For asymptotic curves this result is

7. THEOREM (BELTRAMI-ENNEPER). If c is an asymptotic curve with $c(0) = p$, and $\kappa(0) \neq 0$, then

$$|\tau(0)| = \sqrt{-K(p)}.$$

Moreover, if $K(p) < 0$ and the two distinct asymptotic curves through p both have non-zero curvature at p, then their torsions at p are negatives of each other.

FIRST (SEMI-) PROOF. Parameterize c by arclength. We know that both $\mathbf{t}$ and $\mathbf{n}$ lie in the tangent space of M, and we can let $\nu = \mathbf{b} = \mathbf{t} \times \mathbf{n}$. Then the Serret-Frenet formulas give

$$-d\nu(\mathbf{t}(s)) = -\mathbf{b}'(s) = \tau(s) \cdot \mathbf{n}(s).$$

So the matrix of the self-adjoint transformation $-d\nu\colon M_p \to M_p$, with respect to the orthonormal basis $\mathbf{t}(0), \mathbf{n}(0)$ must be the symmetric matrix

$$\begin{pmatrix} 0 & \tau(0) \\ \tau(0) & 0 \end{pmatrix},$$

with determinant $K(p) = -\tau(0)^2$. This proof does not give any information about the sign of τ.

SECOND PROOF. We know that $\tau = \tau_g$ for the asymptotic curve c. So Proposition 2 gives

$$(1) \qquad \tau(0) = (k_2 - k_1)\sin\theta\cos\theta,$$

where θ is the oriented angle from the principal vector X_1 to $X = c'(0)$. On the other hand, since X is an asymptotic vector, Euler's formula [equation (5)] gives

$$(2) \qquad 0 = k_1\cos^2\theta + k_2\sin^2\theta,$$

so

$$(3) \qquad k_2 = -k_1\frac{\cos^2\theta}{\sin^2\theta}.$$

Substituting into (1) we have

$$\begin{aligned}\tau(0) &= -k_1\left(\frac{\cos^2\theta}{\sin^2\theta} + 1\right)\sin\theta\cos\theta \\ &= -k_1\cdot\frac{\cos\theta}{\sin\theta},\end{aligned}$$

while

$$K(p) = k_1k_2 = -k_1{}^2\frac{\cos^2\theta}{\sin^2\theta} \qquad \text{by (3).}$$

Hence $\tau(0)^2 = -K(p)$ [we divided by $\sin\theta$, but if $\sin\theta = 0$, we could instead solve for k_1 in terms of k_2; alternatively, we can simply note that if $\sin\theta = 0$, then we have $k_1 = 0$ from (2) and $\tau(0) = 0$ from (1)].

Equation (1) also shows that $\tau(0)$ changes sign when we change θ to $-\theta$, which gives the second part of the theorem. ❖

As an example, on the right helicoid we can choose as one family of asymptotic curves the helices $c(s) = (t\cos s, t\sin s, bs)$, for fixed t. On pg. II.33 we found that

$$\tau = \frac{b}{b^2 + t^2},$$

in agreement with our formula for K (page 150). In this example the other family of asymptotic curves are straight lines, with vanishing curvature.

The Beltrami-Enneper theorem raises a natural question, one so natural that no one since Darboux seems to have considered it. Given a unit vector $X \in M_p$, we can consider the line of curvature, the asymptotic line, and the geodesic which have tangent vector X at p (for simplicity we assume p is not an umbilic for the case of lines of curvature, and $K(p) < 0$ for the case of an asymptotic curve). We know that

the torsion of the asymptotic curve c with $c'(0) = X$ is $\pm\sqrt{-K(p)}$,
the torsion of the geodesic c with $c'(0) = X$ is $\tau_g(X) = \mathrm{II}(X, \bar{X})$,
the curvature of the geodesic c with $c'(0) = X$ is $|\kappa_n(X)| = |\mathrm{II}(X, X)|$;

the first statement is Theorem 7, the second is Proposition 3, and the third follows from the equation $\kappa_n{}^2 + \kappa_g{}^2 = \kappa^2$, since $\kappa_g = 0$ for a geodesic. Now it is just as reasonable to ask for the curvature of the asymptotic curve c with $c'(0) = X$. To determine it, we can use one of the invariants of Proposition 4. We saw, in the proof of that Proposition, that

$$(\nabla_X \mathrm{II})(X, X) = \kappa_n'(0) - 2\tau_g(0)\kappa_g(0)$$

for *any* curve c with $c'(0) = X$. Now if c is an asymptotic curve, then κ_n is identically zero, while $\kappa = \pm\kappa_g$ and $\tau = \tau_g$, so we have

$$\begin{aligned}(\nabla_X \mathrm{II})(X, X) &= 0 \mp 2\tau(0)\kappa(0) \\ &= \mp 2\sqrt{-K(p)}\kappa(0), \qquad \text{by Theorem 7.}\end{aligned}$$

Hence, if $K(p) \neq 0$, then

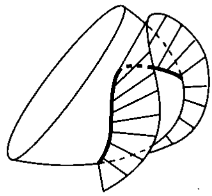
the curvature of the asymptotic curve c with $c'(0) = X$ is

$$\frac{|(\nabla_X \mathrm{II})(X, X)|}{2\sqrt{-K(p)}}.$$

We can also ask for the curvature and torsion of a line of curvature c. Here the situation is somewhat different, since $c'(0) = X \in M_p$ is already essentially determined. We have $\tau_g = 0$ for lines of curvature, so the second invariant of Proposition 4 gives us the value of $[\kappa_n(X) - H(p)]\kappa_g(0)$, and hence of $\kappa_g(0)$. Then we can determine $\kappa = \sqrt{\kappa_g{}^2 + \kappa_n{}^2}$ at 0. To compute the torsion $\tau(0)$, we have to use yet another invariant, involving *second* derivatives (Problems 5 and 6).

To study lines of curvature, we begin with a pretty, though not very useful, criterion for such curves.

8. THEOREM (BONNET). A curve c in M is a line of curvature if and only if the surface S formed by the normals to the surface along c is flat.



PROOF. The surface S is the ruled surface parameterized by

$$f(s,t) = c(s) + t\nu(c(s)) = c(s) + t\delta(s), \qquad \text{say.}$$

From page 147 we see that S is flat if and only if

$$0 = \langle c'(s), \delta(s) \times \delta'(s) \rangle = \left\langle c'(s),\ \nu(c(s)) \times \frac{d\nu(c(s))}{ds} \right\rangle,$$

which is true if and only if $d\nu(c(s))/ds$ is a multiple of $c'(s)$. ❖

In the case of a surface of revolution, Theorem 8 shows that meridians and parallels must be lines of curvature, since the corresponding surfaces S are planes and cones. (Of course, we have already argued in essentially just this way on pages 158–159.) The next theorem can also be used to find the lines of curvature on a surface of revolution.

9. THEOREM (TERQUEM-JOACHIMSTHAL). Let c be a curve in $M_1 \cap M_2$ which is a line of curvature in M_1. Then c is a line of curvature in M_2 if and only if M_1 and M_2 intersect at a constant angle along c (i.e., the normals of M_1 and M_2 have the same angle along c).

PROOF. If ν_i is the unit normal field on M_i, then

$$\begin{aligned}
\frac{d}{ds}\langle \nu_1(c(s)), \nu_2(c(s)) \rangle &= \left\langle \frac{d\nu_1(c(s))}{ds}, \nu_2(c(s)) \right\rangle + \left\langle \nu_1(c(s)), \frac{d\nu_2(c(s))}{ds} \right\rangle \\
&= \left\langle -k(s)\frac{dc}{ds}, \nu_2(c(s)) \right\rangle + \left\langle \nu_1(c(s)), \frac{d\nu_2(c(s))}{ds} \right\rangle, \\
&\qquad \text{since } c \text{ is a line of curvature of } M_1 \\
&= 0 + \left\langle \nu_1(c(s)), \frac{d\nu_2(c(s))}{ds} \right\rangle,
\end{aligned}$$

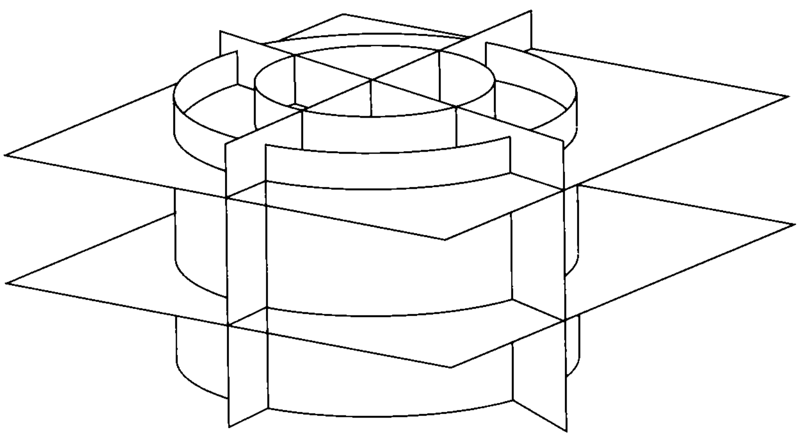
since c is a curve in M_2. If c is a line of curvature in M_2, then the remaining term is similarly 0, so $\langle \nu_1(c(s)), \nu_2(c(s)) \rangle$ is constant.

Conversely, if this quantity has derivative 0, then $d\nu_2(c(s))/ds$ is perpendicular to $\nu_1(c(s))$. On the other hand, it is also perpendicular to $\nu_2(c(s))$. If $\nu_1(c(s))$ and $\nu_2(c(s))$ are linearly independent, then $d\nu_2(c(s))/ds$ must be a multiple of $c'(s)$, and consequently c is a line of curvature in M_2. If $\nu_1(c(s))$ and $\nu_2(c(s))$ are *not* linearly independent, then we must have $\nu_1(c(s)) = \pm\nu_2(c(s))$ for all s (since $\langle \nu_1(c(s)), \nu_2(c(s)) \rangle$ is constant). In this case there is nothing left to prove. ❖

The one really interesting result about lines of curvature concerns **triply orthogonal systems** of surfaces—these are triples of 1-parameter families of surfaces with the property that at each point the tangent planes of the surfaces from any two families are perpendicular. The simplest examples of triply orthogonal systems are the following:

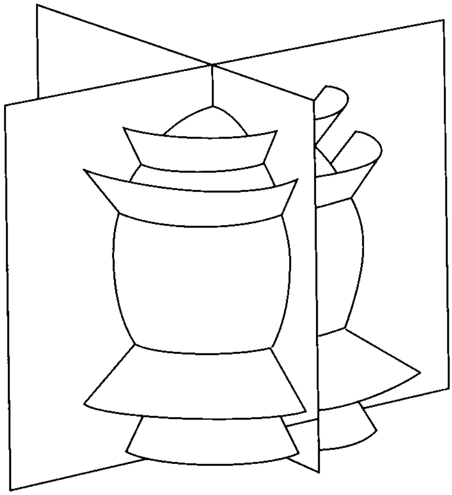
(1) Each family consists of all the planes that are parallel to one of the coordinate planes.

(2) The first family consists of all planes parallel to the (x, y)-plane; the second family consists of all the circular cylinders having the z-axis as their common axis; the third family consists of all planes that pass through the z-axis.



(3) The first family consists of all the concentric spheres around the point 0; the second family consists of all planes that pass through the z-axis; the

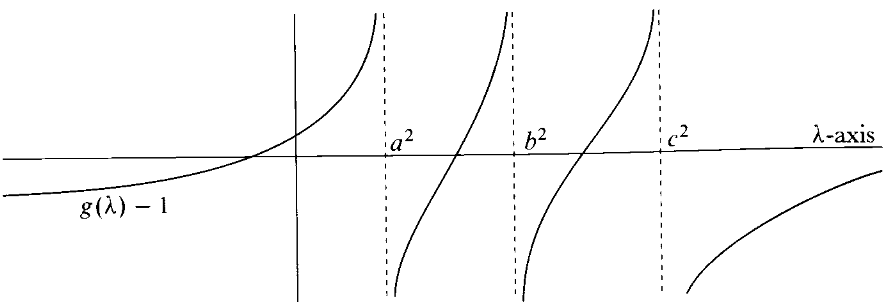
third family consists of cones, each cone being formed by the all the lines through 0 that make some fixed angle with the z-axis.



The one other, less trivial, standard example is formed by the set of all surfaces satisfying the equation

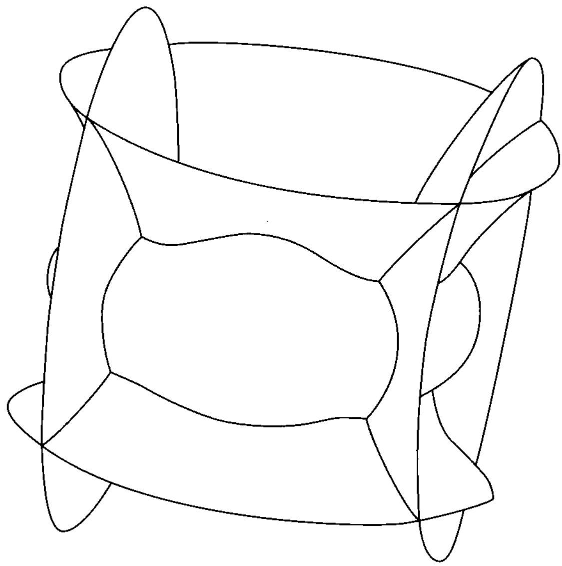
$$(*) \qquad g(\lambda) = \frac{x^2}{a^2-\lambda} + \frac{y^2}{b^2-\lambda} + \frac{z^2}{c^2-\lambda} = 1, \qquad 0 < a^2 < b^2 < c^2.$$

For $\lambda < a^2$ we obtain ellipsoids, for $a^2 < \lambda < b^2$ hyperboloids of one sheet, and for $b^2 < \lambda < c^2$ hyperboloids of two sheets. For any (x, y, z) with $x, y, z \neq 0$, the function $\lambda \mapsto g(\lambda) - 1$ is continuous except at a^2, b^2, c^2; it clearly jumps from $+\infty$ to $-\infty$ as we pass from the left of one of these points to the right of it, and $g(\lambda) - 1 \to -1$ as $\lambda \to -\infty$. Consequently, $g(\lambda) - 1$ must be 0 for at



least one $\lambda_1 < a^2$, one λ_2 with $a^2 < \lambda_2 < b^2$, and one λ_3 with $b^2 < \lambda_3 < c^2$.

There are only 3 roots, since $g(\lambda)-1=0$ is equivalent to a cubic equation in λ. Thus one surface from each family passes through each such point (x,y,z). At



a point (x,y,z) on the surface $g(\lambda_i)=1$, the normal vector has the direction

$$\frac{1}{2}(D_1g(\lambda_i), D_2g(\lambda_i), D_3g(\lambda_i)) = \left(\frac{x}{a^2-\lambda_i}, \frac{y}{b^2-\lambda_i}, \frac{z}{c^2-\lambda_i}\right).$$

At a point (x,y,z) on the two surfaces $g(\lambda_i)=1$ and $g(\lambda_j)=1$, the inner product of the two normal vectors is therefore

$$\frac{x^2}{(a^2-\lambda_i)(a^2-\lambda_j)} + \frac{y^2}{(b^2-\lambda_i)(b^2-\lambda_j)} + \frac{z^2}{(c^2-\lambda_i)(c^2-\lambda_j)}$$

which can be written as

$$\frac{g(\lambda_i)-g(\lambda_j)}{\lambda_j-\lambda_i}=0.$$

Thus our system is orthogonal. Since we can always imbed a given ellipsoid $x^2/a^2+y^2/b^2+z^2/c^2=1$ in a system of the form (*), the following result enables us to describe the lines of curvature on an ellipsoid.

10. THEOREM (DUPIN). The lines of intersection of the surfaces of a triply orthogonal system are lines of curvature on the surfaces.

PROOF. Let Δ_i be the distribution formed by the tangent planes to the i^{th} family of surfaces. Pick unit vector fields X, Y, Z with

$$X \in \Delta_1 \cap \Delta_3, \quad Y \in \Delta_2 \cap \Delta_3, \quad Z \in \Delta_1 \cap \Delta_3.$$

Letting ∇' be the ordinary directional derivative in $\mathbb{R}^3$, we have

$$\nabla'_X Y - \nabla'_Y X = [X, Y] \in \Delta_3, \qquad \text{since } \Delta_3 \text{ is integrable (pg. I.192).}$$

Using orthogonality, we conclude that

$$\text{(1)} \qquad \langle \nabla'_X Y, Z \rangle = \langle \nabla'_Y X, Z \rangle,$$

and of course we can permute X, Y, Z in this equation. On the other hand, we also have

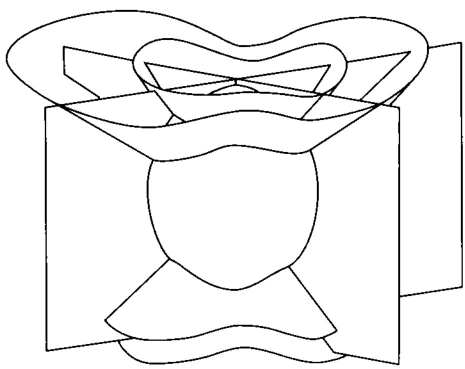
$$\text{(2)} \qquad 0 = X(\langle Y, Z \rangle) = \langle \nabla'_X Y, Z \rangle + \langle Y, \nabla'_X Z \rangle,$$

together with the equations obtained by permuting X, Y, Z. From the equations comprised in (1) and (2) we can conclude, for example, that

$$0 = \langle \nabla'_X Y, Z \rangle = \langle \nabla'_X Y, Y \rangle,$$

so that $\nabla'_X Y$ must be a multiple of X. It follows that the line of intersection of two surfaces in the first and third family is a line of curvature on the surface in the first family, since Y is the normal along this line of intersection. ❖

An exact converse of Dupin's theorem is not true—the lines of intersection of the surfaces of a triple family may be lines of curvature on all the surfaces, even though the surfaces are not orthogonal. For example, the first family may consist of concentric spheres around 0, the second family of planes through the z-axis, and the third family of non-circular cones. On the other hand, if none



of the surfaces in question have umbilics, so that the lines of curvature on each are orthogonal, then the surfaces are clearly orthogonal. The following is a less trivial converse to Theorem 10.

11. THEOREM (DARBOUX). If two families of surfaces are orthogonal, and the intersections are lines of curvature on both, then there exists a third family of surfaces orthogonal to the first two families.

PROOF. Let Δ_1 and Δ_2 be the distributions formed by the tangent planes to the first and second family of surfaces, respectively. Let Δ_3 be the 2-dimension distribution which is everywhere perpendicular to both Δ_1 and Δ_2. Pick unit vector fields X, Y, Z with

$$X \in \Delta_1 \cap \Delta_3, \quad Y \in \Delta_2 \cap \Delta_3, \quad Z \in \Delta_1 \cap \Delta_3.$$

The hypotheses imply that $\nabla'_X Y = \lambda X$ and $\nabla'_Y X = \mu Y$ for certain functions λ and μ. So

$$[X, Y] = \nabla'_X Y - \nabla'_Y X = \lambda X - \mu Y \in \Delta_3.$$

This shows that Δ_3 is integrable, so the third family of surfaces exists, by the Frobenius integrability theorem. ❖

Dupin's Theorem has as a consequence a geometric proof of a theorem about maps $f: U \to V$ from an open set $U \subset \mathbb{R}^3$ to an open set $V \subset \mathbb{R}^3$ which are conformal (angle preserving). In the case of maps $f: U \to V$ with $U, V \subset \mathbb{R}^2 = \mathbb{C}$, it is well-known (Problem 9) that these are precisely the maps which are complex analytic or whose conjugates are. In $\mathbb{R}^3$ the situation is quite different. One class of conformal maps are the **similarities**, the compositions of translations, orthogonal maps, and multiplications by non-zero constants. There is also an analogue for $\mathbb{R}^3$ of the complex analytic map $z \mapsto 1/z$ from $\mathbb{C} - \{0\}$ to $\mathbb{C} - \{0\}$. The analogue is easiest to see if we compose this map with conjugation (= reflection through the real axis), so that we obtain the conformal map

$$z \mapsto \frac{1}{\bar{z}} = \frac{z}{|z|^2}.$$

The same formula

$$I(x) = \frac{x}{|x|^2} \qquad x \in \mathbb{R}^3 - \{0\},$$

where $|x|$ denotes the norm of x, defines a conformal map (Problem 10), called inversion with respect to the unit sphere. The conformal map

$$x \mapsto r^2 \frac{x}{|x|^2}$$

is called inversion with respect to the sphere of radius r about 0; it keeps points on this sphere fixed, and in general x and $f(x)$ lie on the same line through 0 and $|x| \cdot |f(x)| = r^2$. Of course, we can also consider the inversion

$$x \mapsto x_0 + r^2 \frac{x - x_0}{|x - x_0|^2}$$

with respect to the sphere of radius r about x_0. Notice that any inversion I' satisfies $I' \circ I' =$ identity (on its domain).

12. THEOREM (LIOUVILLE). Every conformal map $f\colon U \to V$ from a connected open subset U of $\mathbb{R}^3$ to an open subset V of $\mathbb{R}^3$ is the restriction to U of a composition of similarities and inversions (in fact, at most one of each).

PROOF. Let $S \subset U$ be any connected surface which is part of a plane or a sphere. We can find a triply orthogonal family of surfaces, with S contained in one of the families, such that the lines of intersection with S are curves with any desired tangent vector at any given point. The image of this triple family under f is again orthogonal, since f is conformal. So by Dupin's Theorem, the lines of intersections of this new family with $f(S)$ are lines of curvature on $f(S)$. Therefore we can find lines of curvature pointing in all directions at any point of $f(S)$. So all points of $f(S)$ are umbilics, and by Theorem 2-2 the surface $f(S)$ is either part of a sphere or part of a plane. We now use

13. LEMMA (MÖBIUS). If $U, V \subset \mathbb{R}^3$ are open sets, with U connected, and $f\colon U \to V$ is a map which takes portions of planes and spheres to portions of planes and spheres, then f is the restriction to U of a composition of similarities and immersions (in fact, at most one of each).

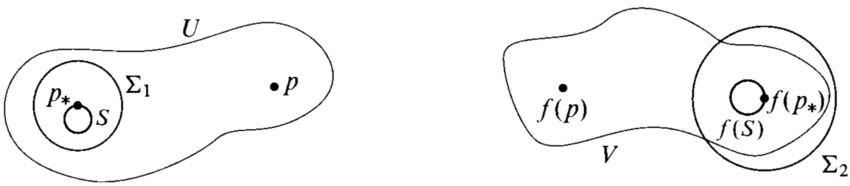
PROOF. We begin with a preliminary observation. Let I' be an inversion with respect to a sphere around p, and let S be a sphere with $p \in S$. Then $I'(S - \{p\})$ is a plane. This can be verified by direct calculation, or one can use the following argument: By what we have just shown, $I'(S - \{p\})$ is part of a plane or sphere. It is also easy to see that $I'(S - \{p\})$ is complete, but not compact (for it becomes compact if we add in the point at infinity). So $I'(S - \{p\})$ must be a plane.

Similarly, if P is a plane not containing p, then one can verify by direct calculation that $I'(P)$ is $S - \{p\}$, or one can use the following argument: By what we have just shown, $I'(P)$ is part of a plane or sphere, and contains points arbitrarily close to P. If $I'(P)$ were part of a plane Q, then Q would have to go through p. But I' keeps planes through p fixed, so we would have $P = I'(I'(P)) \subset I'(Q) = Q$, contradicting the fact that P does not contain p.

So $I'(P)$ must be part of a sphere S through p. Since we already know that $I'(S - \{p\})$ is a plane, we easily conclude that $I'(P)$ is all of $S - \{p\}$.

Now to prove the Lemma it obviously suffices to prove that f has the desired form in a neighborhood of any point p, for then f must be analytic, and consequently equal everywhere to any one of these compositions. In particular, we may assume that f is one-one.

Let p_* be a point of U distinct from p, and let Σ_1 be a sphere around p_* such that all points in the ball B bounded by Σ_1 are in U, but $p \notin B$. Let Σ_2 be any sphere around $f(p_*)$. Let $I_1 : \mathbb{R}^3 - \{p_*\} \to \mathbb{R}^3 - \{p_*\}$ be the inversion

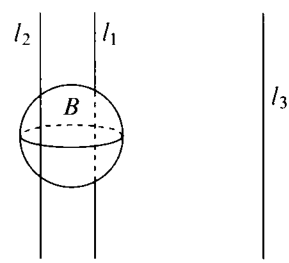


with respect to Σ_1, and let $I_2 : \mathbb{R}^3 - \{f(p_*)\} \to \mathbb{R}^3 - \{f(p_*)\}$ be the inversion with respect to Σ_2. Then

$$F = I_2 \circ f \circ I_1 : \mathbb{R}^3 - B \to \mathbb{R}^3$$

is defined everywhere on $\mathbb{R}^3 - B$, has p in its domain, and takes portions of planes and spheres to portions of planes and spheres.

Now if S is any sphere inside Σ_1 with $p_* \in S$, then $f(S)$ must be a sphere in V with $f(p_*) \in f(S)$, so $I_2(f(S) - \{f(p_*)\})$ is a plane. It follows that F takes planes in $\mathbb{R}^3 - B$ into planes of $\mathbb{R}^3$. It also follows that F takes straight lines in $\mathbb{R}^3 - B$ into straight lines of $\mathbb{R}^3$, since a straight line is the intersection of two planes. We claim that F also preserves parallelism of straight lines. This is clear if l_1 and l_2 are parallel lines lying in a plane $P \subset \mathbb{R}^3 - B$, for then $F(l_1)$ and $F(l_2)$ are disjoint straight lines in $F(P)$. For the case of two parallel lines l_1 and l_2 lying on opposite sides of B, we choose a straight line l_3 parallel

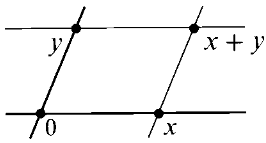


to l_1 and l_2 such that l_1 and l_3 lie in a plane $P_1 \subset \mathbb{R}^3 - B$ while l_2 and l_3 lie in a plane $P_2 \subset \mathbb{R}^3 - B$. Then $F(l_1)$ is parallel to $F(l_3)$ and $F(l_2)$ is parallel to $F(l_3)$, so again $F(l_1)$ is parallel to $F(l_2)$.

Now let T_q denote the translation $x \mapsto x + q$, and consider the map

$$G = T_{-F(p)} \circ F \circ T_p,$$

defined in some convex neighborhood $\mathcal{U}$ of 0. This map takes 0 to 0, and also takes straight lines to straight lines and preserves parallelism. From the parallelogram construction of the sum of two vectors, it is clear that we must



have $G(x + y) = G(x) + G(y)$ whenever x and y are linearly independent vectors with $x, y, x + y \in \mathcal{U}$. The same result holds for linearly dependent x and y, by continuity. From this we easily see that $G(\alpha x) = \alpha G(x)$ for all $\alpha \in \mathbb{R}$ with $x, \alpha x \in \mathcal{U}$. So G is linear, and thus (Problem I.3-31) a composition of an orthogonal map and a self-adjoint map. But G also takes small spheres around 0 to spheres. So we easily see that the self-adjoint factor must be a multiple of the identity, and consequently $G = T_{-F(p)} \circ I_2 \circ f \circ I_1 \circ T_p$ is a similarity in a neighborhood of 0. It follows that f is a composition of similarities and inversions in a neighborhood of p.

To show that f is actually a composition of at most one similarity and inversion, we regard f as extended to the "conformal space" $\mathbb{R}^3 \cup \{\infty\}$, where all similarities are defined at all points, and repeat the proof, choosing $p_* = \infty$. Then the inversion I_1 around p_* is just a similarity on $\mathbb{R}^3$, so we obtain a composition of a similarity and one inversion. (If $f(\infty) = \infty$, then I_2 is also a similarity, and our composition reduces to a similarity.)

This completes the proof of the Lemma, and the Theorem. ❖

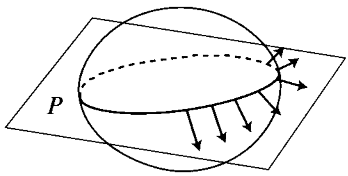
We already have many results about geodesics, which we obtained in our study of intrinsic Riemannian geometry. The following result, though not at all hard, has always seemed to me particularly nice, because of the way that intrinsic and extrinsic notions are intermingled.

14. THEOREM. Let M be a connected surface in $\mathbb{R}^3$ such that every geodesic of M is a plane curve. Then M is part of a plane or a sphere.

PROOF. According to Theorem 2-2, it suffices to show that every point $p \in M$ is an umbilic. We can assume that p is not a planar point, since these are automatically umbilics. Then it certainly suffices to show that any non-asymptotic unit vector $X \in M_p$ is a principal vector. To do this, let c be the geodesic with $c'(0) = X$, lying in the plane P. Proposition 1 shows that the curvature of c is non-zero at 0, and hence in a whole neighborhood of 0. Therefore the desired result follows from

15. LEMMA. If c is a geodesic in M which lies in a plane P and has nowhere 0 curvature, then c is a line of curvature.

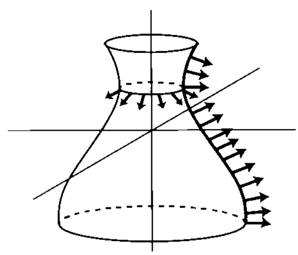
PROOF. Since $c'' \neq 0$ lies in P, and is also perpendicular to the surface, we see that the normal ν to the surface along c lies in P. Hence $d\nu(c(s))/ds$ lies



in P, which means that it must be a multiple of $c'(s)$. [Alternate proof: use Theorem 8.] ❖

To complete our study of geodesics on a surface, we will consider the special case of surfaces of revolution, where the generally intractable differential equations for geodesics reduce to an equation with a simple geometric interpretation. Suppose that our surface is parameterized by

$$f(u,v) = (\rho_1(u)\cos v,\ \rho_1(u)\sin v,\ \rho_2(u)),$$



for a curve $\rho = (\rho_1, \rho_2)$ in the (x,z)-plane. Before we do any computations at all, we notice that by Corollary 1-3 the meridians are geodesics, while the parallel at height $\rho_2(u)$ is a geodesic if and only if $\rho_1{}'(u) = 0$. We will not use all the information given in equations (1) on page 157, but only the fact that the metric has the form

$$\begin{aligned} g_{11}(u,v) &= E(u) \\ g_{12}(u,v) &= 0 \\ g_{22}(u,v) &= G(u) \end{aligned} \qquad \Longrightarrow \qquad g^{11} = \frac{1}{E}, \quad g^{12} = 0, \quad g^{22} = \frac{1}{G}.$$

We then compute the Christoffel symbols (as in Chapter 2, the symbol $[ij,k]$ now denotes the Christoffel symbols for the metric $f^*\langle\ ,\ \rangle$ with respect to the usual coordinate system on $\mathbb{R}^2$):

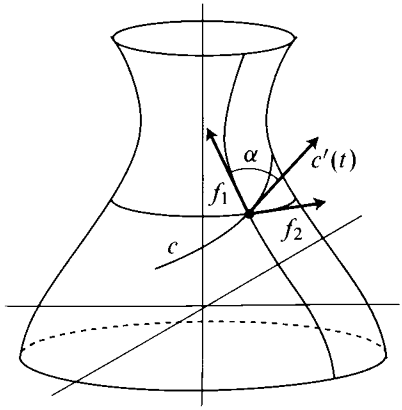
$$[11,1] = \frac{1}{2}E'$$
$$[12,2] = [21,2] = -[22,1] = \frac{1}{2}G'$$
$$\text{all other } [ij,k] = 0;$$
$$\Gamma_{11}^1 = \frac{E'}{2E}, \quad \Gamma_{22}^1 = -\frac{G'}{2E}, \quad \Gamma_{12}^2 = \Gamma_{21}^2 = \frac{G'}{2G}$$
$$\text{all other } \Gamma_{ij}^k = 0.$$

If $t \mapsto \gamma(t) = (\gamma_1(t), \gamma_2(t))$ is a geodesic in $\mathbb{R}^2$ with the metric $f^*\langle\ ,\ \rangle$, then the equations on pg. I.329 give

$$(*) \qquad \frac{d^2\gamma_2}{dt^2} + \frac{G'}{G}(\gamma_1(t))\frac{d\gamma_1}{dt}\frac{d\gamma_2}{dt} = 0.$$

Now for any curve c on a surface of revolution, let $r(t)$ be the distance from $c(t)$ to the axis of revolution, and let $\alpha(t)$ be the angle between c and the meridian curve that it crosses at time t. More precisely, $\alpha(t)$ is the oriented angle from the meridian tangent vector $f_1(c(t))$ to $c'(t)$ when (f_1, f_2) is chosen as the positive orientation for the tangent space of M, so that

$$\frac{c'}{|c'|} = (\cos\alpha)\frac{f_1}{|f_1|} + (\sin\alpha)\frac{f_2}{|f_2|}.$$



Then equation $(*)$ immediately leads to

16. THEOREM (CLAIRAUT). A geodesic c on a surface of revolution satisfies the equation

$$r(t) \cdot \sin \alpha(t) = A$$

for some constant A. Conversely, if c satisfies this equation and is not a parallel, then c is a geodesic, provided that it is parameterized by arclength.

PROOF. We can write c as $c = f \circ \gamma$ for some curve γ which is a geodesic in $\mathbb{R}^2$ with the metric $f^*\langle\ ,\ \rangle$. Equation (*) gives

$$\begin{aligned} 0 &= G(\gamma_1(t)) \cdot \gamma_2''(t) + G'(\gamma_1(t)) \cdot \gamma_1'(t)\gamma_2'(t) \\ &= (G \circ \gamma_1)(t) \cdot \gamma_2''(t) + (G \circ \gamma_1)'(t) \cdot \gamma_2'(t) \\ &= [(G \circ \gamma_1) \cdot \gamma_2']'(t), \end{aligned}$$

so $G(\gamma_1(t)) \cdot \gamma_2'(t)$ is constant.

On the other hand, since $g_{12} = 0$, we have

$$\begin{aligned} G(\gamma_1(t)) \cdot \gamma_2'(t) &= \langle f_1(\gamma(t)) \cdot \gamma_1'(t) + f_2(\gamma(t)) \cdot \gamma_2'(t),\ f_2(\gamma(t)) \rangle \\ &= \langle c'(t), f_2(\gamma(t)) \rangle \\ &= |c'(t)| \cdot |f_2(\gamma(t))| \cdot \sin \alpha(t) \\ &= |c'(t)| \cdot \sqrt{G(\gamma_1(t))} \cdot \sin \alpha(t), \end{aligned}$$

so $\sqrt{G(\gamma_1(t))} \cdot \sin \alpha(t)$ is constant. But the formulas on page 157 show that

$$\sqrt{G(\gamma_1(t))} = \rho_1(\gamma_1(t)),$$

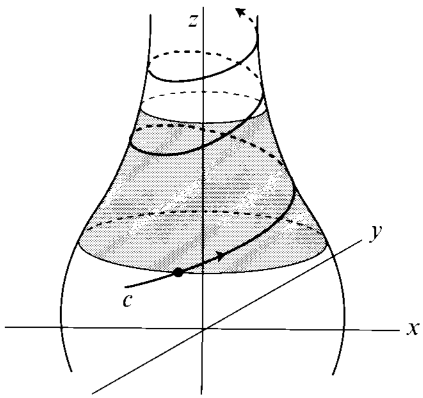
which is exactly the distance of $c(t)$ from the z-axis.

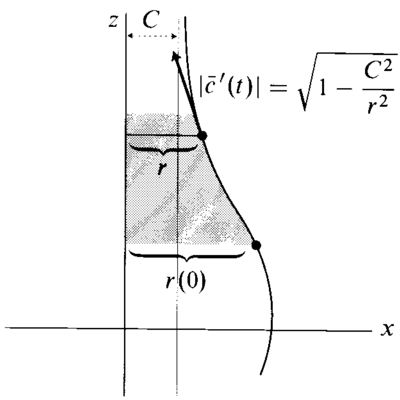
The proof of the converse is left to the reader. ❖

Clairaut's Theorem allows us to give a very complete description of the global behavior of geodesics on surfaces of revolution. Let ρ be the profile curve of the surface, parameterized by arclength, and let $\bar{c}$ denote the projection of c on the (x, z)-plane, so that $\bar{c}$ lies along the image of ρ. Let us suppose that the geodesic c is also parameterized by arclength, and, for concreteness, that $c'(0)$ is pointing upwards. If our geodesic c satisfies $r(t) \cdot \sin \alpha(t) = C$, then the length of $\bar{c}'(t)$ is

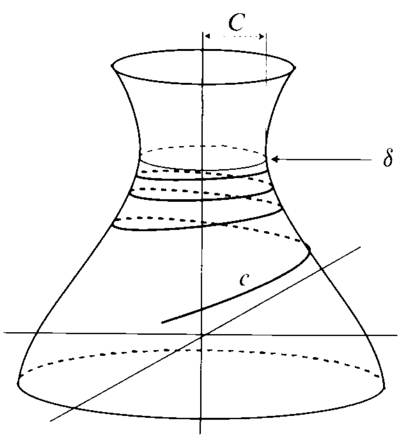
$$\begin{aligned} |\bar{c}'(t)| &= \langle \rho'(t), c'(t) \rangle = \cos \alpha(t) \\ &= \sqrt{1 - \frac{C^2}{r(t)^2}}. \end{aligned}$$

From this we see that so long as c lies in a region where $r(t)$ is bounded away from C, the tangent vector $\bar{c}'(t)$ will have length bounded away from 0. It is now easy to deduce the following: If the profile curve ρ never comes within distance C of the z-axis, as we traverse it in the direction of $\bar{c}$, then $\bar{c}$ must traverse the whole of ρ in this direction.





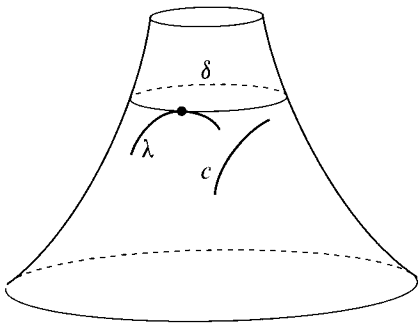
Now suppose that ρ does come within distance C of the z-axis, and let δ be the first meridian above the one at $r(0)$ which has radius C. Then c clearly must come arbitrarily close to δ. If δ happens to be a geodesic, then c cannot intersect δ, for we would then have $\alpha = \pi/2$, which would mean that c' would point along a tangent vector of δ.



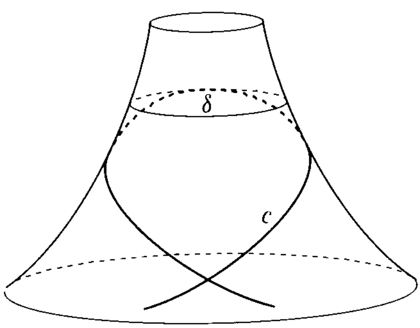
Finally, suppose that δ is *not* a geodesic. Consider a geodesic λ starting at a point of δ, with tangent vector pointing along δ. If β is the angle which λ makes with the meridian, then λ satisfies

$$r(t) \cdot \sin\beta(t) = \text{ constant},$$

and $r(t) = C$ when $\beta(t) = \pi/2$, so the constant must also be C. Since δ is not a geodesic, the region directly above δ has $r < C$, and λ cannot go into it. Consequently, it enters the region where c is. A rotation about the axis will



bring λ into coincidence with c since they are both determined by the same constant C, and naturally the rotated curve λ is still a geodesic. In other words, we have shown that c eventually hits δ. Moreover, c' points along δ' at the intersection point (as it must, since $r(t) \cdot \sin\alpha(t) = C$). In addition, c must bounce off δ and proceed downwards. Naturally, the shape of the part going downwards must differ from that of the part going upwards only by a reflection.



ADDENDUM 1
SPECIAL PARAMETER CURVES

Proposition I.5-18 immediately implies

17. COROLLARY. Let p be a point on a surface M in $\mathbb{R}^3$.

(a) If p is not an umbilic point, then there is an imbedding $f: U \to M$, with $p \in f(U)$, whose parameter curves are lines of curvature.

(b) If $K(p) < 0$, then there is an imbedding $f: U \to M$, with $p \in f(U)$, whose parameter curves are asymptotic curves.

It is sometime useful, especially in the next chapter, to write some of our formulas in terms of these and other special coordinate systems; readers may check for themselves that the following formulas are correct.

A. *The parameter lines are orthogonal.*

Then $F = 0$, and the formula in Problem 13 becomes (subscripts denoting partial derivatives)

$$K = -\frac{1}{2\sqrt{EG}}\left[\left(\frac{E_2}{\sqrt{EG}}\right)_2 + \left(\frac{G_1}{\sqrt{EG}}\right)_1\right].$$

B. *The parameter lines are lines of curvature.*

In this case, of course, we still have the equation from (A). We also have

$$l = k_1 E, \quad n = k_2 G, \quad m = 0, \quad F = 0.$$

So the Codazzi-Mainardi equations (page 56) become

$$l_2 = \frac{E_2}{2}\left(\frac{l}{E} + \frac{n}{G}\right)$$

$$n_1 = \frac{G_1}{2}\left(\frac{l}{E} + \frac{n}{G}\right).$$

C. *The parameter lines are asymptotic curves.*

We have $l = n = 0$, and the Codazzi-Mainardi equations become

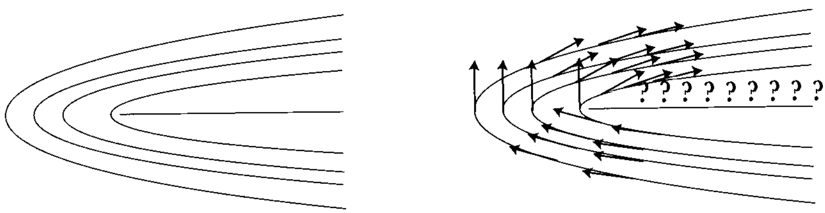
$$m_1 = \frac{\left[\frac{1}{2}(EG - F^2)_1 + FE_2 - EG_1\right]}{EG - F^2} \cdot m$$

$$m_2 = \frac{\left[\frac{1}{2}(EG - F^2)_2 + FG_1 - GE_2\right]}{EG - F^2} \cdot m.$$

ADDENDUM 2
SINGULARITIES OF LINE FIELDS

In Chapter I.11 we defined the index of an isolated zero of a vector field, and we proved (Theorem I.11-30) that if a vector field on a compact oriented manifold M has only isolated zeros, then the sum of the indices of these zeros is the Euler characteristic of M.

Now consider the situation where we have a 1-dimensional distribution Δ defined in a neighborhood of a point p of a 2-dimensional manifold M, except at the point p itself. As is easily seen from the pictures below, it may not be



possible to find a nowhere zero vector field X such that $\Delta(q)$ is always spanned by $X(q)$. Nevertheless, we will define an index of Δ at p.

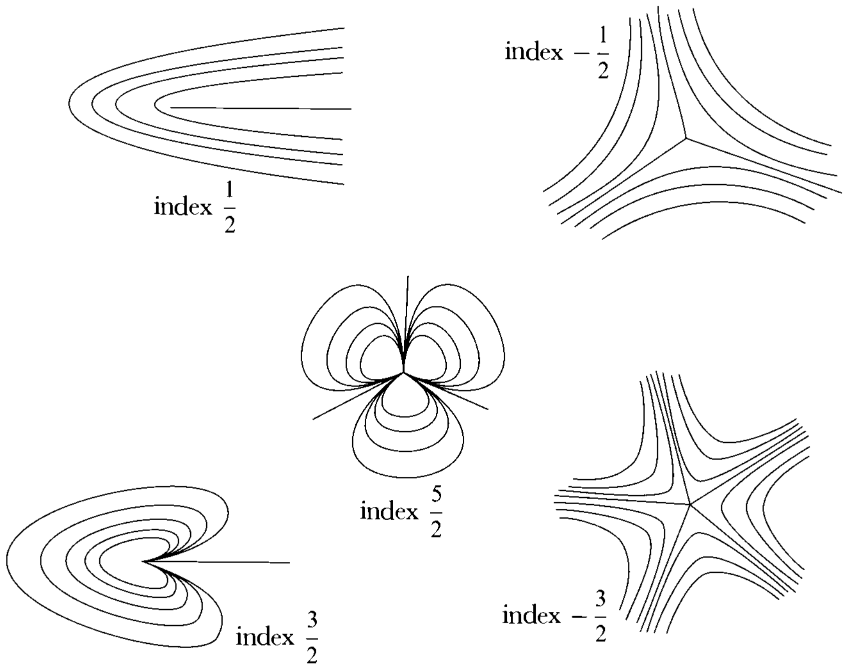
As in the case of a vector field, we first suppose that the distribution Δ is defined on $U - \{0\}$, for U a neighborhood of $0 \in \mathbb{R}^2$. We introduce the projective line $\mathbb{P}^1$, which is S^1 with all pairs of antipodal points x and $-x$ identified. Alternatively, $\mathbb{P}^1$ is the set of all lines through $0 \in \mathbb{R}^2$, and thus the set of all directions in $\mathbb{R}^2$. Then we have a map $f_\Delta \colon U - \{0\} \to \mathbb{P}^1$ defined by $f_\Delta(q) =$ the direction of $\Delta(q)$. If we let $i \colon S^1 \to U$ be $i(x) = \varepsilon x$ for some $\varepsilon > 0$, then we have the map $f_\Delta \circ i \colon S^1 \to \mathbb{P}^1$. But $\mathbb{P}^1$ is homeomorphic to S^1—we can define a homeomorphism $\alpha \colon \mathbb{P}^1 \to S^1$ by noting that $\mathbb{P}^1$ is the same as a semi-circle with end points identified. Thus we have a map

$$\alpha \circ f_\Delta \circ i \colon S^1 \to S^1,$$

and this map has a certain degree. We define the **index** of Δ at p to be

$$\text{index of } \Delta \text{ at } p = \frac{1}{2}\,\text{degree}(\alpha \circ f_\Delta \circ i).$$

The same arguments which we used in Chapter I.11 allow us to extend this definition from $\mathbb{R}^2$ to any arbitrary surface. Some examples of these indices are given below.



If Δ happens to be of the form $\Delta(q) =$ space spanned by $X(q)$, for a vector field X, and $f_X : U - \{0\} \to S^1$ is the map taking q to $X(q)/|X(q)| \in S^1$, then we have the commutative diagram

$$\begin{array}{ccc} S^1 & \xrightarrow{f_X \circ i} & S^1 \\ & {\scriptstyle f_\Delta \circ i} \searrow & \downarrow \pi \\ & & \mathbb{P}^1 \\ & & \downarrow \alpha \\ & & S^1 \end{array}$$

where π is the natural projection. Since $\alpha \circ \pi : S^1 \to S^1$ has degree 2, it follows that we have

$$\begin{aligned}\text{index of } \Delta \text{ at } p &= \frac{1}{2}\,\text{degree}(\alpha \circ f_\Delta \circ i)\\ &= \frac{1}{2}\,\text{degree}(\alpha \circ \pi \circ f_X \circ i)\\ &= \text{degree}(f_X \circ i)\\ &= \text{index of } X \text{ at } p.\end{aligned}$$

In particular, the index of Δ at p is an integer in this case.

Conversely, if the index of Δ at p is an integer, then a suitable vector field X can be found. The easiest way to see this is to give an alternative description of the index. Let $c\colon [0,1] \to \mathbb{R}^2$ be the curve $c(t) = \varepsilon(\cos 2\pi t, \sin 2\pi t)$. The angle between the x-axis and the direction of $\Delta(q)$ is defined only up to a multiple of π [while the angle between the x-axis and a vector is defined only up to a multiple of 2π], but we can find a continuous function $\theta\colon [0,1] \to \mathbb{R}$ such that $\theta(t)$ is an angle between the x-axis and the direction of $\Delta(c(t))$. Then (compare Proposition II.1-6) we have

$$\text{index of } \Delta \text{ at } p = \frac{1}{2\pi}[\theta(1) - \theta(0)].$$

If this index is an integer, so that $\theta(1) - \theta(0)$ is a multiple of 2π, then we can pick out X along c by letting $X(c(t))$ be the unit vector in $\Delta(c(t))$ such that $\theta(t)$ is an angle between $X(c(t))$ and the x-axis.

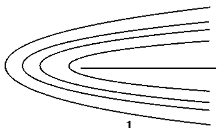
Given any 1-dimensional distribution Δ defined in $M - \{p_1, \ldots, p_k\}$, we can define a 2-fold covering space $\varpi\colon M' \to M - \{p_1, \ldots, p_k\}$ just as on page 198: we let the two points of $\varpi^{-1}(q)$ correspond to the two unit vectors in $\Delta(q)$. If U is an open ball around p_i, then $\varpi^{-1}(U)$ is either two disjoint copies of $U - \{p\}$, or else it is connected and $\varpi|\varpi^{-1}(U)$ looks like the map $z \mapsto z^2$ taking

$$\{z \in \mathbb{C}\colon 0 < |z| < 1\} \to \{z \in \mathbb{C}\colon 0 < |z| < 1\}.$$

The first case occurs when Δ comes from a vector field, and the second case occurs when it does not. In the former case, we will add two new points to M', one for each of the disjoint copies of $U - \{p_i\}$, define ϖ of each of these two new points to be p_i, and define the neighborhoods of these new points in the obvious way so that $\varpi^{-1}(U)$ now consists of two disjoint copies of U. In the second case, we will add just *one* new point p_i^* whose neighborhoods we define to be the sets $\{p_i^*\} \cup \varpi^{-1}(A - \{p_i\})$ for A a neighborhood of p_i in U. Let $\widetilde{M}$ be M' with all new points added. Then $\widetilde{M}$ is a manifold, but the map

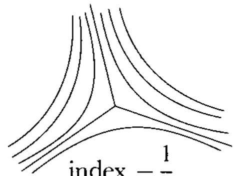
$\varpi: \widetilde{M} \to M$ is not a 2-fold covering space, for over certain points p_i it looks like the map $z \mapsto z^2$. These points p_i are called "branch points" of the 2-fold "branched covering space" $\varpi: \widetilde{M} \to M$. There is clearly a distribution $\widetilde{\Delta}$ on $\varpi^{-1}(M - \{p_1, \ldots, p_k\})$ with $\varpi_* \widetilde{\Delta} = \Delta$; moreover $\widetilde{\Delta}$ obviously comes from a vector field.

If our original Δ looks like



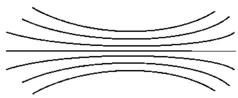
index $\frac{1}{2}$

or



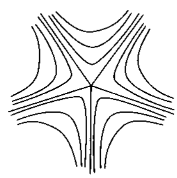
index $-\frac{1}{2}$

in a neighborhood U of p_i, then $\widetilde{\Delta}$ looks something like



index 0

or



index -2

(to see this, just note that if we wrap the bottom pictures twice around the origin, by $z \mapsto z^2$, then the images cover the top pictures). In general,

18. LEMMA. Let $\varpi: \widetilde{M} \to M$ be a 2-fold branched covering space, and let p be a branch point. Let Δ be a 1-dimensional distribution defined on a neighborhood of p, except at p itself, and let $\widetilde{\Delta}$ be a 1-dimensional distribution defined on a neighborhood of $\varpi^{-1}(p)$, except at $\varpi^{-1}(p)$ itself, such that $\varpi_* \widetilde{\Delta} = \Delta$. Then the index $\tilde{\imath}$ of $\widetilde{\Delta}$ at $\varpi^{-1}(p)$ is related to the index i of Δ at p by

$$\tilde{\imath} = 2i - 1.$$

PROOF. Regard both p and $\varpi^{-1}(p)$ as $0 \in \mathbb{C}$, and ϖ as the map $z \mapsto z^2$. Let $c: [0,1] \to \mathbb{R}^2$ be the semi-circle $c(t) = \varepsilon(\cos \pi t, \sin \pi t)$, and let $\theta: [0,1] \to \mathbb{R}$ be a continuous function such that $\theta(t)$ is an angle between the x-axis and the direction of $\widetilde{\Delta}(c(t))$. Note that by our construction of $\widetilde{\Delta}$ we have

$$\tilde{\imath} = 2 \cdot \frac{1}{2\pi}[\theta(1) - \theta(0)].$$

Now the curve $\varpi \circ c\colon [0,1] \to \mathbb{R}^2$ is the circle

$$\varpi \circ c(t) = \varepsilon^2(\cos 2\pi t, \sin 2\pi t).$$

It is easy to see that the function

$$\phi(t) = \theta(t) + \pi t$$

gives an angle between the x-axis and the direction of $\Delta(\varpi \circ c(t))$. So

$$\begin{aligned} i &= \frac{1}{2\pi}[\phi(1) - \phi(0)] = \frac{1}{2\pi}[\theta(1) - \theta(0)] + \frac{1}{2} \\ &= \frac{\tilde{i}}{2} + \frac{1}{2}. \text{ ❖} \end{aligned}$$

It is also easy to find the Euler characteristic $\chi(\tilde{M})$.

19. LEMMA. If $\varpi\colon \tilde{M} \to M$ is a 2-fold branched covering space with l branch points $p_1, \dots, p_l$, then

$$\chi(\tilde{M}) = 2\chi(M) - l.$$

FIRST PROOF. We will use a triangulation of M (Problem 17 suggests a simple proof that any compact surface can be triangulated). Choose the triangulation so that $p_1, \dots, p_l$ are included among the V vertices (0-simplexes), and let there be E edges (1-simplexes) and F faces (2-simplexes). There is an obvious triangulation of $\tilde{M}$ with 2 vertices over each vertex of M, except for the vertices $p_1, \dots, p_l$ over which there is only 1 vertex. So the number $\tilde{V}$ of vertices of $\tilde{M}$ is

$$\tilde{V} = 2V - l,$$

while we have

$$\tilde{E} = 2E, \quad \tilde{F} = 2F.$$

So

$$\chi(\tilde{M}) = \tilde{V} - \tilde{E} + \tilde{F} = 2V - l - 2E + 2F = 2\chi(M) - l.$$

SECOND PROOF. Let X be a vector field on M with only finitely many zeros $q_1, \dots, q_k$, and let ι_j be the index of X at q_j. We might as well assume that the p_i are contained among the q's, for at any point we can always introduce a new zero of X (with index 0). Then there is a vector field $\tilde{X}$ on $\varpi^{-1}(M - \{q_1, \dots, q_k\})$ with $\varpi_* \tilde{X} = X$. If q_j is a branch point, then by Lemma 18,

$$\text{index of } \tilde{X} \text{ at } \varpi^{-1}(q_j) = 2\iota_j - 1.$$

If q_j is not a branch point, then $\varpi^{-1}(q_j)$ consists of two points q'_j, q''_j, and

$$\text{index of } \widetilde{X} \text{ at } q'_j \text{ or } q''_j = \iota_j.$$

Then Theorem I.11-30 gives

$$\begin{aligned}\chi(\widetilde{M}) &= \text{ sum of the indices of } \widetilde{X} \\ &= 2\sum \iota_j - \text{number of branch points} \\ &= 2\chi(M) - l. \quad ❖\end{aligned}$$

Exactly the same sort of reasoning which was used in this second proof leads us to our main result.

20. THEOREM. Let M be a compact oriented surface, and Δ a 1-dimensional distribution on $M - \{p_1, \ldots, p_k\}$. Then the sum of the indices of Δ at the p_i is $\chi(M)$.

PROOF. Consider the 2-fold branched covering $\varpi\colon \widetilde{M} \to M$ constructed previously, and the distribution $\widetilde{\Delta}$ on $\widetilde{M}$. If p_i is a branch point, then by Lemma 18

$$\text{index of } \widetilde{\Delta} \text{ at } \varpi^{-1}(p_i) = 2(\text{index of } \Delta \text{ at } p_i) - 1.$$

If p_i is not a branch point, then $\varpi^{-1}(p_i)$ consists of two points p'_i, p''_i, and

$$\text{index of } \widetilde{\Delta} \text{ at } p'_i \text{ or } p''_i = \text{index of } \Delta \text{ at } p_i.$$

It follows that

(1) $$\begin{aligned}\text{sum of the indices of } \widetilde{\Delta} = 2(&\text{sum of the indices of } \Delta) \\ &- \text{number of branch points.}\end{aligned}$$

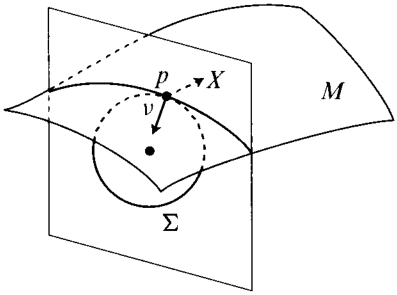
Since $\widetilde{\Delta}$ comes from a vector field, it follows from Theorem I.11-30 that

(2) $$\text{sum of the indices of } \widetilde{\Delta} = \chi(\widetilde{M}).$$

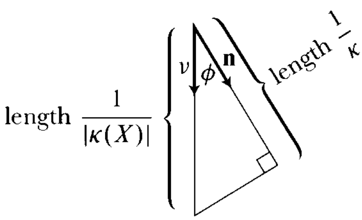
The result now follows from (1), (2), and Lemma 19. ❖

PROBLEMS

1. Let $M \to \mathbb{R}^3$ be a surface, and $X \in M_p$ a unit vector. Let Σ be the circle in the plane perpendicular to X which is tangent to M at p and has radius $1/|\kappa(X)|$. Show that for every curve c in M with $c'(0) = X$, the center of the



osculating circle of c at 0 lies on Σ. The picture below is a hint.



2. Let $c\colon \mathbb{R} \to \mathbb{R}^3$ be a curve which lies on a sphere of radius r.

(a) We have $\kappa_n = 1/r$, and consequently $\kappa = \sqrt{1/r^2 + \kappa_g{}^2} > 0$.
(b) If $\mathbf{t}, \mathbf{n}, \mathbf{b}$ is the Frenet frame of c, and $\mathbf{v}(s) = \nu(c(s))$, where ν is the unit normal of the sphere in which c lies, then

$$\begin{aligned} 0 = \langle \mathbf{t}, \mathbf{v}\rangle' &= \kappa\langle \mathbf{n}, \mathbf{v}\rangle + 1/r \\ \langle \mathbf{n}, \mathbf{v}\rangle' &= \tau\langle \mathbf{b}, \mathbf{v}\rangle \\ \langle \mathbf{b}, \mathbf{v}\rangle' &= -\tau\langle \mathbf{n}, \mathbf{v}\rangle. \end{aligned}$$

(c) If $\kappa' = 0$, then $\tau = 0$. If κ' is nowhere 0, then τ is nowhere 0 and

$$\frac{\tau}{\kappa} + \left[\frac{1}{\tau}\left(\frac{1}{\kappa}\right)'\right]' = 0.$$

(d) Conversely, if this condition holds, then c lies on some sphere.
(e) For convenience, say $r = 1$, so that

$$\kappa = \sqrt{1 + {\kappa_g}^2}.$$

Let $\mathbf{t}, \mathbf{u}, \mathbf{v}$ be the Darboux frame for c. By differentiating the equations

$$\mathbf{t}' = \kappa_g \mathbf{u} + \mathbf{v}$$
$$\mathbf{t}' = \kappa \mathbf{n},$$

show that

$$\tau = \frac{{\kappa_g}'}{1 + {\kappa_g}^2}.$$

3. (a) Let $X \in M_p$ be a unit vector, and suppose that the geodesic γ with $\gamma'(0) = X$ has $\kappa(0) = 0$. Then

$$|\tau_g(X)| = \sqrt{-K(p)}.$$

(b) If γ is a geodesic with $\gamma'(0) \in M_p$ a principal vector and $\kappa(0) = 0$, then $K(p) = 0$.
(c) If γ_i are geodesics with perpendicular tangent vectors ${\gamma_i}'(0) \in M_p$, and $\kappa_1(0) = 0$, but $\kappa_2(0) \neq 0$, then the torsion $\tau_2(0)$ of γ_2 satisfies

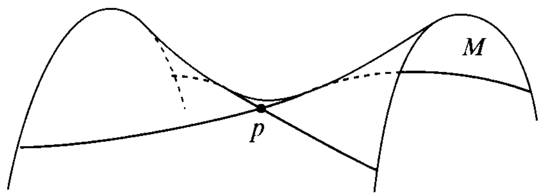
$$|\tau_2(0)| = \sqrt{-K(p)}.$$

4. (a) Let c be an arclength parameterized curve on a surface $M \subset \mathbb{R}^3$ such that $c'(0) \in M_p$ is an asymptotic direction. If c is *not* asymptotic at p [$c''(0) \notin M_p$], then $\kappa(0) = 0$.
(b) Now suppose that $c''(0) \in M_p$. Let $\bar{c}$ be the arclength parameterized asymptotic curve with $\bar{c}'(0) = c'(0)$, and denote the curvature and torsion of $\bar{c}$ by $\bar{\kappa}$ and $\bar{\tau}$.

Using Laguerre's formulation of Proposition 4, and equation (13) on page 192, show that

$$\kappa(0)[3\bar{\tau}(0) - \tau(0)] = 2\bar{\tau}(0)\bar{\kappa}(0)$$
$$\Longrightarrow \kappa(0)\big[\pm 3\sqrt{-K(p)} - \tau(0)\big] = \pm 2\sqrt{-K(p)}\,\bar{\kappa}(0).$$

(c) Show that at a point $p \in M$ where $K(p) < 0$, the intersection of M and the tangent plane at p consists of two curves which cross each other at p, and which point in the asymptotic directions. For either of these curves, show that



the curvature at p is 2/3 times the curvature of the corresponding asymptotic curve through p (Beltrami).

5. Let $M \subset \mathbb{R}^3$ be a surface, and let X_1, X_2 be an orthonormal moving frame, in terms of which we write

$$\mathrm{II} = \sum_{i,j} l_{ij}\theta^i \otimes \theta^j.$$

(a) By Problem 2-5 we have

$$\nabla \mathrm{II} = \sum_{i,j,k} l_{ij;k}\theta^i \otimes \theta^j \otimes \theta^k$$

where

$$\sum_k l_{ij;k}\theta^k = dl_{ij} - \sum_\rho l_{\rho j}\omega_i^\rho - \sum_\rho l_{i\rho}\omega_j^\rho.$$

If X_1, X_2 is the frame $X, \bar{X}$ in the proof of Proposition 4, apply this equation to X_1 to obtain

$$(1) \qquad l_{11;1} = \frac{dl_{11}}{ds} - 2l_{12} \cdot \omega_1^2(X_1),$$

and deduce the first part of Proposition 4. Deduce the second part similarly.
(b) Show that

$$d\psi_i^3 = \sum_j dl_{ij} \wedge \theta^j - \sum_{j,\rho} l_{ij}\omega_\rho^j \wedge \theta^\rho.$$

Conclude from equation (1) that

$$(l_{i2;1} - l_{i1;2})\theta^1 \wedge \theta^2 = 0.$$

(c) If

$$\nabla\nabla \mathrm{II} = \sum_{i,j,k,h} l_{ij;kh}\theta^i \otimes \theta^j \otimes \theta^k \otimes \theta^h,$$

show that

$$l_{11;11} = \frac{dl_{11;1}}{ds} - 3l_{12;1}\omega_1^2(X_1)$$

$$l_{12;11} = \frac{dl_{12;1}}{ds} - (2l_{22;1} - l_{11;1})\omega_1^2(X_1)$$

$$= \frac{dl_{12;1}}{ds} - \left(2\left[2\frac{dH}{ds} - l_{11;1}\right] - l_{11;1}\right)\omega_1^2(X_1).$$

(d) For all arclength parameterized curves c in M with the same tangent vector $c'(0) \in M_p$ the quantity

$$\kappa_n''(s) - 2\tau_g(s)\kappa_g'(s) - 5\tau_g'(s)\kappa_g(s) - 6[\kappa_n(s) - H(c(s))]\kappa_g^2(s)$$

has the same value at $s = 0$. The same is true for

$$\tau_n''(s) + [2\kappa_n(s) - H(c(s))]\kappa_g'(s)$$
$$+ \left[5\kappa_n'(s) - 6\tau_g(s)\kappa_g(s) - 6\frac{dH(c(s))}{ds}\right]\kappa_g(s).$$

6. Let $X \in M_p$ be a principal vector, and c the principal curve with $c'(0) = X$.

(a) Determine $\kappa_n'(0)$.
(b) Use Problem 5 to determine $\kappa_g'(0)$ [in terms of $X(H)$].
(c) Show how to determine $\kappa'(0)$.
(d) Show how to use equation (7) on page 189 to determine $\phi'(0)$, and then how to find $\tau(0)$.

7. Let $\pi : \tilde{M} \to M$ be an n-fold covering of a compact orientable manifold M.

(a) Let X be a vector field on M with only finitely many zeros. Show that there is a vector field $\tilde{X}$ on $\tilde{M}$ with n zeros of index ι for every zero of X with index ι. Conclude that $\chi(M) = n \cdot \chi(M)$.
(b) Also prove this result by finding a triangulation of M for which there is a corresponding triangulation of $\tilde{M}$ with n k-simplexes for each k-simplex of M.

8. Equation (∗) on page 205 is the basis for the best coordinate system for ellipsoids and hyperboloids of one or two sheets. For given (x, y, z) let $\phi(\lambda)$ be the cubic

(1) $$\phi(\lambda) = (a^2 - \lambda)(b^2 - \lambda)(c^2 - \lambda)(g(\lambda) - 1),$$

with roots

$$\lambda_1 < a^2, \qquad a^2 < \lambda_2 < b^2, \qquad b^2 < \lambda_3 < c^2.$$

Clearly

$$\phi(\lambda) = (\lambda_1 - \lambda)(\lambda_2 - \lambda)(\lambda_3 - \lambda).$$

(a) Substituting for the expression for $\phi(\lambda)$ from (1), and choosing $\lambda = a^2$, b^2, and c^2, obtain

$$x^2 = \frac{(a^2 - \lambda_1)(a^2 - \lambda_2)(a^2 - \lambda_3)}{(a^2 - b^2)(a^2 - c^2)}$$

$$y^2 = \frac{(b^2 - \lambda_1)(b^2 - \lambda_2)(b^2 - \lambda_3)}{(b^2 - a^2)(b^2 - c^2)}$$

$$z^2 = \frac{(c^2 - \lambda_1)(c^2 - \lambda_2)(c^2 - \lambda_3)}{(c^2 - a^2)(c^2 - b^2)}.$$

(b) Setting one λ_i = constant, these equations give a parameterization of the surface $g(\lambda_i) = 1$ by means of the two other variables λ_j, λ_k. Setting

$$a^2 - \lambda_i = \alpha, \qquad b^2 - \lambda_i = \beta, \qquad c^2 - \lambda_i = \gamma$$

$$\lambda_j - \lambda_i = u, \qquad \lambda_k - \lambda_i = v,$$

we have the surface

$$\frac{x^2}{\alpha} + \frac{y^2}{\beta} + \frac{z^2}{\gamma} = 1$$

parameterized by

$$x = \sqrt{\frac{\alpha(\alpha - u)(\alpha - v)}{(\alpha - \beta)(\alpha - \gamma)}}$$

$$y = \sqrt{\frac{\beta(\beta - u)(\beta - v)}{(\beta - \alpha)(\beta - \gamma)}}$$

$$z = \sqrt{\frac{\gamma(\gamma - u)(\gamma - v)}{(\gamma - \alpha)(\gamma - \beta)}}.$$

Note that the u- and v-parameter lines are clearly lines of curvature. Calculate that

$$E = \frac{u(u-v)}{f(u)} \qquad F = 0 \qquad G = \frac{v(v-u)}{f(v)}$$

where $f(t) = 4(\alpha - t)(\beta - t)(\gamma - t)$.

9. Let $f\colon \mathbb{C} \to \mathbb{C}$ be

$$f(x,y) = f(x+iy) = (u(x,y), v(x,y)) = u(x,y) + iv(x,y)..$$

(a) Calculate that

$$f^*(dx \otimes dx + dy \otimes dy) = ({u_x}^2 + {v_x}^2)\, dx \otimes dx + ({u_y}^2 + {v_y}^2)\, dy \otimes dy \\ + (u_x u_y + v_x v_y)(dx \otimes dy + dy \otimes dx).$$

Conclude that f is conformal if and only if

$${u_x}^2 + {v_x}^2 = {u_y}^2 + {v_y}^2 \quad \text{and} \quad u_x u_y + v_x v_y = 0.$$

(b) Show that f is conformal if and only if

$$u_x = \pm v_y, \quad u_y = \mp v_x.$$

Hint: Multiply the first equation of part (a) by ${v_y}^2$.

10. Consider the map

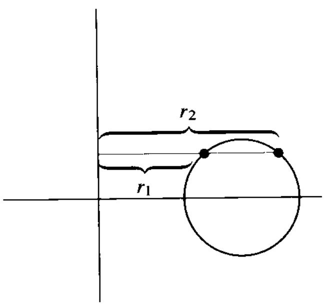
$$I(x) = \frac{x}{|x|^2} \qquad x \in \mathbb{R}^3 - \{0\}.$$

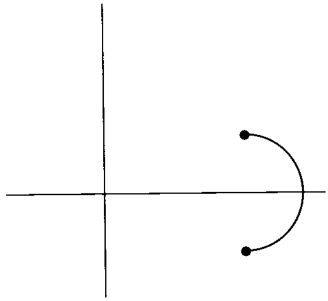
(a) If $S^2(r)$ is the 2-sphere around 0 of radius r, and $X \in S^2(r)_p$, then $I_*(X) \in S^2(\frac{1}{r})_{I(p)}$ is parallel to X, and $|I_*(X)| = \frac{1}{r}|X|$.
(b) If ν is the unit normal to $S^2(r)$ at p, then $I_*(\nu)$ is $\frac{1}{r}$ times the unit normal to $S^2(\frac{1}{r})$ at $I(p)$.
(c) I is conformal.

11. Note that Lemma 13 is valid also for $U, V \subset \mathbb{R}^2$, where $f\colon U \to V$ takes portions of straight lines and circles to portions of straight lines and circles. Conclude that if f is orientation preserving, then f is of the form

$$f(z) = \frac{az+b}{cz+d}, \qquad a,b,c,d \in \mathbb{C}.$$

12. Use Clairaut's Theorem to analyze the geodesics on a torus. (The fact that points with different values of r can have the same height is confusing, but irrelevant; it may help to think of the torus in terms of the profile curve on the right, but with the end points identified.)





Answer: All geodesics, except the inner parallel, intersect the outer parallel. Those which intersect at a small angle θ describe a sine-like curve between two parallels at equal distances from the outer one. For a certain angle θ we obtain a sine-like curve between the top and bottom parallels. For somewhat larger θ the curve flops over the top and bottom and bounces between two parallels on the inner part of the torus. For a certain angle θ we obtain a curve which approaches the inner parallel asymptotically. For slightly larger θ the geodesic hits the inner parallel, at a small angle, after going around many times. As θ increases, the geodesic hits the inner parallel at larger angles, after going around fewer times, until, at $\theta = \pi/2$, we obtain a meridian circle.

13. From pg. II.131 we have the formula

$$W^4 K = \det\begin{pmatrix} -\frac{1}{2}G_{11} + F_{12} - \frac{1}{2}E_{22} & \frac{1}{2}E_1 & F_1 - \frac{1}{2}E_2 \\ F_2 - \frac{1}{2}G_1 & E & F \\ \frac{1}{2}G_2 & F & G \end{pmatrix}$$

$$- \det\begin{pmatrix} 0 & \frac{1}{2}E_2 & \frac{1}{2}G_1 \\ \frac{1}{2}E_2 & E & F \\ \frac{1}{2}G_1 & F & G \end{pmatrix},$$

where $W = \sqrt{EG - F^2}$. Verify the following (seemingly non-rational) expression for K, due to Frobenius:

$$K = -\frac{1}{4W^4}\det\begin{pmatrix} E & E_1 & E_2 \\ F & F_1 & F_2 \\ G & G_1 & G_2 \end{pmatrix} + \frac{1}{2W}\left[\left(\frac{F_2 - G_1}{W}\right)_1 + \left(\frac{F_1 - E_2}{W}\right)_2\right]$$

14. Let $f : U \to M$ be an imbedding whose parameter curves are asymptotic curves, so that

$$f_{11} = Af_1 + Bf_2, \quad f_{22} = Cf_1 + Df_2.$$

Choose the orientation so that $\det(f_1, f_2, f_{12}) > 0$.

(a) For the affine first fundamental form I_f we have

$$g_{11} = g_{22} = 0 \qquad \text{and} \qquad g_{12} = \sqrt{\det(f_1, f_2, f_{12})} = \alpha, \text{ say.}$$

(b) The Christoffel symbols for I_f are

$$[11, 2] = \alpha_1, \quad [22, 1] = \alpha_2, \qquad \text{all other } [ij, k] = 0.$$

$$\Gamma^1_{11} = \frac{\alpha_1}{\alpha}, \qquad \Gamma^2_{22} = \frac{\alpha_2}{\alpha}, \qquad \text{all other } \Gamma^k_{ij} = 0,$$

so

$$\nabla_{f_1} f_1 = \frac{\alpha_1}{\alpha} f_1, \qquad \nabla_{f_2} f_2 = \frac{\alpha_2}{\alpha} f_2.$$

(c) We have

$$\mathfrak{s}(f_i, f_j) = f_{ij} - \nabla_{f_i} f_j \implies \ell_{ijk} = \{f_{ij} - \nabla_{f_i} f_j, f_k\},$$

and consequently

$$\ell_{111} = B\alpha, \quad \ell_{112} = A\alpha - \alpha_1, \quad \ell_{221} = D\alpha - \alpha_2, \quad \ell_{222} = C\alpha.$$

(d) Show that

$$(\ell_{111})^2 = \det(f_1, f_{11}, f_{111})$$
$$-(\ell_{222})^2 = \det(f_2, f_{22}, f_{222}).$$

[This gives another proof to the second part of Theorem 7.]
(e) Show that

$$\ell_{112} = \ell_{221} = 0.$$

Thus

$$\mathrm{II}_f = \sqrt{\det(f_1, f_{11}, f_{111})}\, ds^1 \otimes ds^1 \otimes ds^1 + \sqrt{-\det(f_2, f_{22}, f_{222})}\, ds^2 \otimes ds^2 \otimes ds^2.$$

15. (a) For the parameterization in Problem 14, show that the Pick invariant is

$$J = \frac{\ell_{111}\ell_{222}}{(g_{12})^3}.$$

(b) On a region of M where $\ell_{111} = 0$, we have $f_{11} = af_1$. Hence M is a ruled surface. Similarly on a region where $\ell_{222} = 0$. Conversely, a ruled surface has $J = 0$.

16. As an alternative to the approach taken in Problem 3-11, use Problem 14 to show that a doubly ruled surface with $K < 0$ has $\mathrm{I\!I} = 0$, so that it is quadratic, by Proposition 2-19.

17. Show that a compact Riemannian 2-manifold can be triangulated by choosing the vertices to be an ε-dense set, where every point has a geodesically convex neighborhood of radius $> \varepsilon$, and choosing the edges to be geodesic segments.

CHAPTER 5

COMPLETE SURFACES OF CONSTANT CURVATURE

We have already seen, in Chapter 3, that there are many surfaces with constant curvature. On the other hand, few of our examples were complete manifolds. In this chapter we will determine precisely which surfaces in $\mathbb{R}^3$ can be obtained by isometrically immersing complete manifolds with constant curvature $K > 0$, $K = 0$, or $K < 0$.

In the case of complete surfaces of constant curvature $K > 0$ we will actually assume that the surface is compact; in Chapter 8 (Theorem 8-17), however, we will see that this additional hypothesis is superfluous. By Hadamard's Theorem, our surface is an imbedded submanifold $M \subset \mathbb{R}^3$.

1. LEMMA (HILBERT). Let M be a surface immersed in $\mathbb{R}^3$, and let $p \in M$ be a non-umbilic point. Let $k_1 \geq k_2$ be the two principal curvatures on M and suppose that k_1 has a local maximum at p, and k_2 has a local minimum at p. Then $K(p) \leq 0$.

PROOF. According to Addendum 1 to Chapter 4, we can choose an imbedding $f : U \to M$, with $p \in f(U)$, whose coordinate lines are the lines of curvature. Then Gauss' equation and the Codazzi-Mainardi equations become [subscripts, except those on k_1 and k_2, denote partial derivatives]

$$K = -\frac{1}{2\sqrt{EG}}\left[\left(\frac{E_2}{\sqrt{EG}}\right)_2 + \left(\frac{G_1}{\sqrt{EG}}\right)_1\right] \tag{1}$$

$$l_2 = \frac{E_2}{2}\left(\frac{l}{E} + \frac{n}{G}\right) = \frac{E_2}{2}(k_1 + k_2) \tag{2}$$

$$n_1 = \frac{G_1}{2}\left(\frac{l}{E} + \frac{n}{G}\right) = \frac{G_1}{2}(k_1 + k_2); \tag{3}$$

the second equalities in (2) and (3) follow from the fact that

$$l = k_1 E, \qquad n = k_2 G.$$

Moreover, differentiation of these last two equations yields

$$l_2 = \frac{\partial k_1}{\partial t} E + k_1 E_2, \qquad n_1 = \frac{\partial k_2}{\partial s} G + k_2 G_1.$$

(The functions k_i are differentiable near p, since the functions H and K are differentiable, and $k_i = H \pm \sqrt{H^2 - K}$, where $H^2 - K > 0$ in a neighborhood of the non-umbilic point p.) Together with (2) and (3) we then have

$$(2') \qquad E_2 = -\frac{2E}{k_1 - k_2} \cdot \frac{\partial k_1}{\partial t}$$

$$(3') \qquad G_1 = \frac{2G}{k_1 - k_2} \cdot \frac{\partial k_2}{\partial s}.$$

Substituting (2′), (3′) into (1) gives

$$(1') \qquad K = -\frac{1}{2EG}\left[-\frac{2E}{k_1 - k_2} \cdot \frac{\partial^2 k_1}{\partial t^2} + \frac{2G}{k_1 - k_2} \cdot \frac{\partial^2 k_2}{\partial s^2}\right] + (\text{something continuous}) \cdot \frac{\partial k_1}{\partial t} + (\text{something continuous}) \cdot \frac{\partial k_2}{\partial s}.$$

Since k_1 has a local maximum at p, and k_2 a local minimum, we have

$$\frac{\partial k_1}{\partial t}(p) = \frac{\partial k_2}{\partial s}(p) = 0, \qquad \frac{\partial^2 k_1}{\partial t^2}(p) \le 0, \qquad \frac{\partial^2 k_2}{\partial s^2}(p) \ge 0.$$

Together with (1′) this shows that $K(p) \le 0$. ❖

2. THEOREM. If M is a compact connected surface in $\mathbb{R}^3$ with constant curvature $K > 0$, then M is a sphere.

PROOF. Let $k_1 \ge k_2$ be the principal curvatures on M, and let p be a point where k_1 achieves its maximum. Then $k_2 = K/k_1$ has its minimum at p. If we had $k_1(p) > k_2(p)$, so that p was not an umbilic, then the Lemma would imply that $K(p) \le 0$, a contradiction. Hence $k_1(p) = k_2(p)$. Moreover, for any point $q \in M$ we then have

$$k_1(p) \ge k_1(q) \ge k_2(q) \ge k_2(p) = k_1(p),$$

so also $k_1(q) = k_2(q)$. Thus all points of M are umbilics, and Theorem 2-2 applies. ❖

We also have another result of interest:

3. THEOREM. If M is a compact connected surface in $\mathbb{R}^3$, with K everywhere > 0, and constant mean curvature H, then M is a sphere.

PROOF. As before, if k_1 achieves its maximum at p, then $k_2 = H - k_1$ achieves its minimum. Since we are assuming $K(p) > 0$, we find that $k_1(p) = k_2(p)$, and the rest of the proof proceeds as before. ❖

We now turn our attention to immersed surfaces M in $\mathbb{R}^3$ which are flat, that is, which have $K = 0$ everywhere. We know that if a connected open set $U \subset M$ consists entirely of planar points, then U is part of a plane. Let us consider a point of M which is *not* a planar point, and hence a parabolic point. In a neighborhood U of this point we can choose a C^∞ unit vector field X_1 such that each $X_1(q)$ is a principal vector with principal curvature $k_1(q) = 0$, and another C^∞ unit vector field X_2, orthogonal to X_1, such that each $X_2(q)$ is a principal vector with principal curvature $k_2(q) \neq 0$.

4. PROPOSITION. The integral curves of X_1 are straight line segments. Consequently, every non-planar point in a flat surface has a neighborhood which is a ruled surface.

PROOF. Let $\theta^i, \omega_1^2, \psi_i^3$ be the forms associated with the adapted orthonormal moving frame (X_1, X_2, ν). Since

$$\nabla'_{X_i}\nu = \begin{cases} 0 & i = 1 \\ k_2 X_2 & i = 2, \end{cases}$$

we have

$$\psi_1^3(X_i) = -\psi_3^1(X_i) = -\langle X_1, \nabla'_{X_i}\nu\rangle = 0$$
$$\psi_2^3(X_i) = -\psi_3^2(X_i) = -\langle X_2, \nabla'_{X_i}\nu\rangle = -k_2\theta^2(X_i).$$

In short, we have

$$\psi_1^3 = 0, \quad \psi_2^3 = -k_2\theta^2,$$

so one of the Codazzi-Mainardi equations (page 70) gives

$$0 = d\psi_1^3 = -\psi_2^3 \wedge \omega_1^2 = k_2\theta^2 \wedge \omega_1^2.$$

Since k_2 is never 0, this can happen only if ω_1^2 is always a multiple of θ^2. This implies that

$$0 = \omega_1^2(X_1) = \langle \nabla'_{X_1} X_1, X_2\rangle,$$

while we also have

$$0 = \psi_1^3(X_1) = \langle \nabla'_{X_1} X_1, X_3 \rangle.$$

Therefore $\nabla'_{X_1} X_1 = 0$, which means that the integral curves of X_1 are straight lines in $\mathbb{R}^3$. ❖

We still haven't said what happens at a planar point which is a limit of parabolic points, but before worrying about such points, we will first obtain some information about flat ruled surfaces, classically known as **developable surfaces**.

Consider a ruled surface

$$f(s,t) = c(s) + t\delta(s), \tag{1}$$

where we assume $|\delta| = 1$ for convenience, but do not necessarily insist on the canonical parameterization (since it is not always possible to introduce it). As we have seen (page 147 and page 197), this surface is flat precisely when c', δ, δ' are everywhere linearly dependent. Let us first consider an open interval for s on which δ, δ' alone are *everywhere linearly dependent.* Since $|\delta| = 1$, we have $\langle \delta, \delta' \rangle = 0$, so actually δ' must be 0, and δ is constant. We then have a portion of a cylinder. Next let us consider an interval on which δ, δ' are *everywhere linearly independent.* Then there are unique C^∞ functions α, β with

$$c'(s) = \alpha(s)\delta(s) + \beta(s)\delta'(s). \tag{2}$$

Let

$$c^*(s) = c(s) - \beta(s)\delta(s). \tag{3}$$

Then

$$\begin{aligned} c^{*\prime}(s) &= c'(s) - \beta(s)\delta'(s) - \beta'(s)\delta(s) \\ &= [\alpha(s) - \beta'(s)]\delta(s). \end{aligned} \tag{4}$$

Again, we will consider only two special cases. On an interval where $\alpha(s) - \beta'(s)$ is *always* 0, we have $c^*(s) =$ constant vector c_0^*, and by (1) and (3) our surface is

$$f(s,t) = c_0^* + (t + \beta(s))\delta(s),$$

which is a portion of the cone

$$g(s,t) = c_0^* + t\delta(s).$$

On the other hand, on an interval where $\alpha(s) - \beta'(s)$ is *never* 0, we have

(5) $$\delta(s) = \frac{c^{*\prime}(s)}{\alpha(s) - \beta'(s)},$$

so by (1) and (3) our surface is

$$\begin{aligned} f(s,t) &= c(s) + t\delta(s) = c^*(s) + (t + \beta(s))\delta(s) \\ &= c^*(s) + \left[\frac{t + \beta(s)}{\alpha(s) - \beta'(s)}\right] c^{*\prime}(s), \end{aligned}$$

which is a portion of the tangent developable

$$g(s,t) = c^*(s) + tc^{*\prime}(s)$$

of the curve c^*. [Notice that by (4) we have

$$c^{*\prime\prime}(s) = (\alpha(s) - \beta'(s))\delta'(s) + (\alpha'(s) - \beta''(s))\delta(s);$$

since we are on an interval where δ' and δ are linearly independent and $\alpha - \beta'$ is nowhere 0, this shows that the curve c^* does indeed have non-vanishing curvature on the interval.]

The discussion in the preceding paragraph constitutes the classical "classification" of developable surfaces, which was commonly expressed by saying that all developables are planes, cylinders, cones, or tangent developables. We have clearly not proved any such result, since we have only considered special intervals on which certain conditions hold. Nowadays people tend to say: Oh well, the classical classification of developables was really for analytic surfaces—one ought to say that a connected *analytic* developable surface is either a plane, cylinder, cone, or tangent developable. But *even this* is not true. It is true that if a connected analytic developable surface contains a plane, cylinder or cone, then it must be a plane, cylinder, or cone; for planes, cylinders, and cones are the surfaces which arise in our analysis when certain functions are zero on a whole interval. But an analytic developable surface can also be made up of several tangent developables joined together along a line belonging to neither. For example, consider the analytic function $\delta\colon \mathbb{R} \to S^2$ defined by

(1) $$\delta(s) = \frac{1}{\sqrt{(1+s^2)^2 + (1+s^3)^2 + s^8}} \cdot (1+s^2, 1+s^3, s^4).$$

We clearly have

(2) $$\delta'(s) = sA(s)$$

for some analytic function A, and we easily check that

(3) $$A(0) \quad \text{is not a multiple of} \quad \delta(0).$$

Now let $c\colon \mathbb{R} \to \mathbb{R}^3$ be an analytic curve with $c(0) = 0$ and

(4) $$c'(s) = \delta(s) + A(s).$$

Since $c'(0) = \delta(0) + A(0)$ is linearly independent of $\delta(0)$ [by (3)], the map

$$f(s,t) = c(s) + t\delta(s)$$

is an immersion at $(0,t)$ for all t. We claim that f is flat at all points, i.e., that $c'(s), \delta(s), \delta'(s)$ are linearly dependent for all s. This is clear for $s = 0$, since $\delta'(0) = 0$, while for $s \neq 0$ we have, by (4),

$$c'(s) = 1 \cdot \delta(s) + \frac{1}{s}\delta'(s).$$

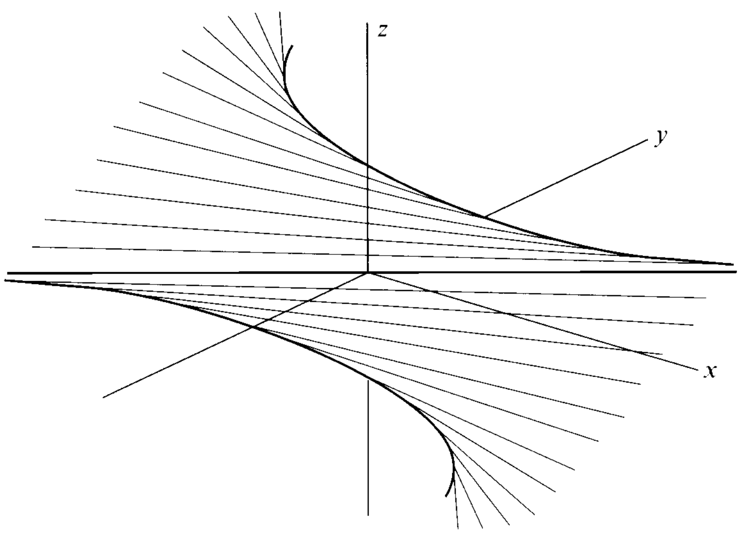
This equation, together with equation (3) on page 236, shows that for $s \neq 0$ our surface is the tangent developable of the curve

$$c^*(s) = c(s) - \frac{1}{s}\delta(s).$$

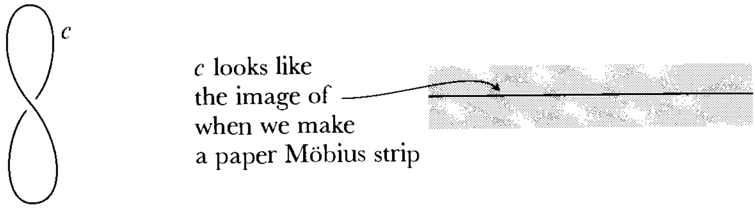
Since c and δ are analytic, c^* is definitely *not* analytic at 0. Instead we have two different curves for $s > 0$ and $s < 0$; it is easy to see that

$$c^*(s) \to (-\infty, -\infty, 0) \quad \text{as} \quad s \to 0^+$$
$$c^*(s) \to (\infty, \infty, 0) \quad \text{as} \quad s \to 0^-.$$

Our surface looks something like the following picture:

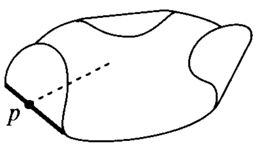


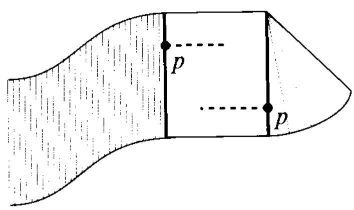
The possible complexity of analytic developables is perhaps most strikingly illustrated by the fact that there is an *analytic* developable surface which is homeomorphic to the Möbius strip; it may be constructed as follows (see Wunderlich [1] for details). We take an analytic closed curve c which looks like the center line of the paper Möbius strip. In the paper model, this line is a geo-



desic, since it comes from a straight line in the flat piece of paper which we bent to form the Möbius strip. So we want a developable surface on which c is a geodesic. This can be obtained by taking the rectifying developable of c (Problem 3-13); because of the way c twists, its rectifying developable twists so as to be homeomorphic to the Möbius strip.

Since so many complexities arise even for analytic developables, it might seem hopeless to say anything at all about flat surfaces which are merely C^∞, and which may contain planar points (in which case we cannot even be sure that they are ruled surfaces). To see how C^∞ flat surfaces can be different from analytic ones, consider the two surfaces pictured below. The first is obtained by





rolling up three pieces of a disc (choosing an appropriate profile for the rolled up portions, so that the surface is C^∞), the second by gluing a cylinder and a cone to a plane. In Chapter 3, our construction of a C^∞ flat Möbius strip gave another example, in which two tangent developables were glued together in a C^∞, but non-analytic, way. In the above pictures we have singled out certain points p which are planar points, but at the same time the limit of parabolic points. In each case there are segments (indicated by dashed lines) which have p as an endpoint, but which cannot be extended past p in the other direction.

On the other hand, these rays do not go through parabolic points; in fact, most of the points on them lie in completely planar regions. Moreover, in each case there is also a segment (indicated by heavy lines) containing p in its interior. We will show that this situation is completely typical. The main tool is a Lemma which is obtained by following the philosophically prescribed route:

5. LEMMA. Let p be a parabolic point on a flat surface M immersed in $\mathbb{R}^3$, let $L_p \subset \mathbb{R}^3$ be the straight line containing the integral curve through p of the vector field of Proposition 4, and let O_p be the component containing p of the set of points in $L_p \cap M$ where $k_2 \neq 0$. Let c be the arclength parameterization of L_p, with $c(0) = p$, and let $k(s) = k_2(c(s))$. Then on O_p the function k is of the form

$$k(s) = \frac{1}{As + B}$$

for some constants A and B.

PROOF. We keep the same notation as in the proof of Proposition 4, so that we have

$$(1) \quad \psi_1^3 = 0, \qquad\qquad (2) \quad \psi_2^3 = -k_2\theta^2.$$

We have already found, using one of the Codazzi-Mainardi equations, that ω_1^2 is always a multiple of θ^2. This means that on the region where $k_2 \neq 0$ it is also a multiple of ψ_2^3, say

$$(3) \qquad \omega_1^2 = g\psi_2^3.$$

Now we use the *other* Codazzi-Mainardi equation, to obtain

$$(4) \qquad d\psi_2^3 = -\psi_1^3 \wedge \omega_2^1 = 0 \qquad \text{by (1)}.$$

Thus

$$\begin{aligned} 0 = d\psi_2^3 &= -dk_2 \wedge \theta^2 - k_2\, d\theta^2 \qquad \text{by (2)} \\ &= -dk_2 \wedge \theta^2 + k_2\omega_1^2 \wedge \theta^1, \end{aligned}$$

and therefore

$$dk_2 \wedge \theta^2 = -k_2\theta^1 \wedge \omega_1^2.$$

Applying this to (X_1, X_2) gives

$$(*) \qquad \begin{aligned} X_1(k_2) &= -k_2\omega_1^2(X_2) \\ &= -k_2 g\psi_2^3(X_2) \qquad \text{by (3)} \\ &= (k_2)^2 g \qquad \text{by (2)}. \end{aligned}$$

We also want to use the Gauss equation $d\omega_1^2 = -K\theta^1 \wedge \theta^2$ (page 69). Since $K = 0$, this says that

$$d\omega_1^2 = 0. \tag{5}$$

Hence

$$\begin{aligned} 0 = d\omega_1^2 &= d(g\psi_2^3) && \text{by (3)} \\ &= dg \wedge \psi_2^3 + 0 && \text{by (4)} \\ &= -k_2\, dg \wedge \theta^2 && \text{by (2).} \end{aligned}$$

Applying this to (X_1, X_2) gives $dg(X_1) = 0$. In other words,

(**) g is constant along the integral curves of X_1.

From (*) and (**) we see that the function $k(s) = k_2(c(s))$ satisfies the differential equation

(***) $k'(s) = -Ak(s)^2$ for some constant A.

We can solve this explicitly: Since

$$-\frac{k'}{k^2} = \left(\frac{1}{k}\right)',$$

we have

$$k(s) = \frac{1}{As + B}.$$

More precisely, by suitable choice of B we obtain any desired initial condition $k(0)$ except $k(0) = 0$—the solution with this initial condition is simply $k = 0$. But $k(0) \neq 0$ by assumption, so our k has the above form. ❖

6. COROLLARY. Let p be a parabolic point on a flat surface M immersed in $\mathbb{R}^3$. Then there is a unique straight line L_p through p such that the component C_p of $M \cap L_p$ which contains p is an interval (possibly infinite) with p in its interior. All the points of C_p are also parabolic. Moreover, C_p cannot end in M; that is, if C_p has an endpoint q, then q is not in M.

PROOF. Existence of L_p follows from Proposition 4, and uniqueness is obvious,* since there cannot be two distinct asymptotic directions at p. If there were non-parabolic points of C_p, then one of them would be an endpoint of the interval O_p of Lemma 5. If this point is $q = c(s_0)$, then we would have

$$0 = k_2(q) = k_2(c(s_0)) = \lim_{s \to s_0} \frac{1}{As + B},$$

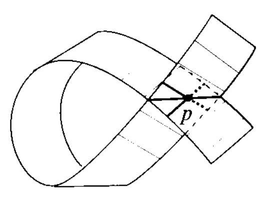
which is impossible. Thus all points of C_p are parabolic, and $C_p = O_p$. Similarly, if C_p had an endpoint $q \in M$, then $k_2(q)$ could not be 0, so q would also be a parabolic point, and a neighborhood of q would be a ruled surface. From this it is clear that C_p could be extended to include q in its interior, which gives a contradiction. ❖

Once we have this Corollary, the next two follow by completely elementary argumentation.

7. COROLLARY. Let M be a flat surface immersed in $\mathbb{R}^3$, and let $p \in M$ be a planar point which is a limit of parabolic points p_n. Then the conclusion of Corollary 6 still holds, except that all points of C_p are now planar points which are limits of parabolic points.

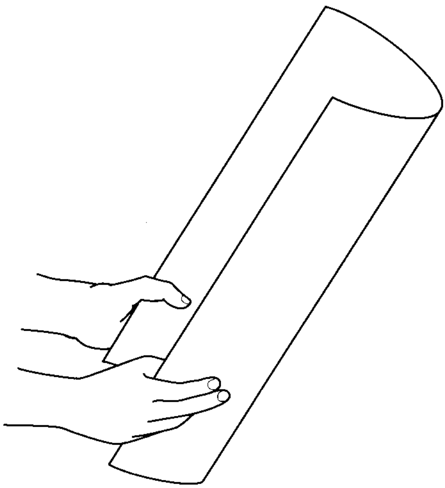
PROOF. Some subsequence of the straight lines L_{p_n} have a limiting direction. Let L_p be the straight line through p with this limiting direction. The components C_{p_n} have p_n in their interior; moreover, Corollary 6 shows that the lengths of the C_{p_n} are bounded away from 0 in each direction from p_n. It follows that C_p has p in its interior. The assertion about endpoints of C_p is an immediate consequence of the same property for the C_{p_n}. It is also clear that all points of C_p are limits of parabolic points, since all points of each C_{p_n} are parabolic. If some point of C_p itself were a parabolic point, then, by Corollary 6, all points of C_p would be parabolic points, including p itself, a contradiction. To prove uniqueness, notice that another straight line L' through p would have to intersect C_{p_n} for large enough n; thus L' would contain parabolic points, so all points on L' would be parabolic points, including p itself, a contradiction. ❖

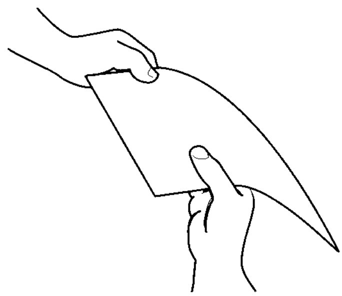
*A "counterexample" is shown in the figure on the right: since we are dealing with immersed surfaces, the uniqueness has to be given a careful formulation, which is left to the reader. We will also be somewhat sloppy about such questions in the sequel, since the precise statements are always clear, but irritatingly messy to state.



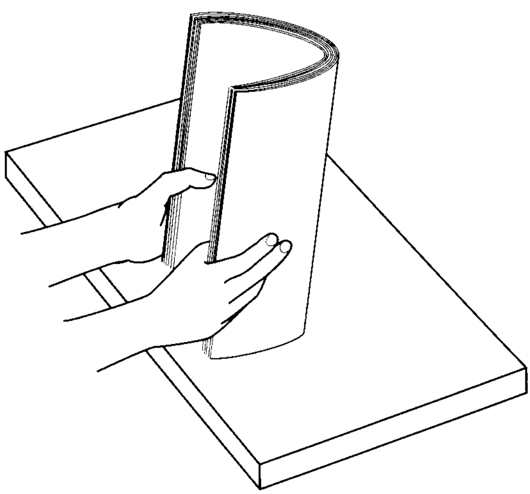
8. COROLLARY. Every point p of a flat surface M immersed in $\mathbb{R}^3$ is contained in the interior of some line segment lying in M. This segment is unique unless some neighborhood of p is a plane, and the only segments which can end in M are those whose interior points are of this type.

R. Malz likes to point out that this result has a very important application in everyday life. If one holds a piece of paper in the shape of a cylinder, then





it will stay stiff even if it is very long; however, as soon as it is allowed to be planar, it will flop down under gravity. That is why people always curl up a pile of papers when they try to align it by tapping it on a desk top.

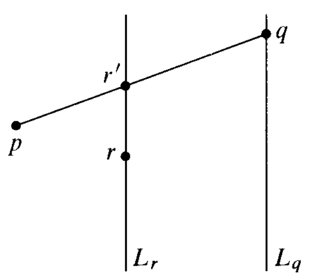


At this point it is almost clear that a *complete* flat surface M immersed in $\mathbb{R}^3$ must be a generalized cylinder. For an open dense subset of M will be a union of pieces of planes, cylinders, cones and tangent developables, where we can arrange for the latter 3 types to contain only parabolic points. Because of completeness, Corollaries 6 and 7 show that the generators of these cylinders, cones, and tangent developables must be infinite straight lines. But in the case of cones and tangent developables this is simply impossible, for there would definitely be singularities at the vertex or edge of regression. So an open dense subset of M consists of cylinders (some possibly degenerating to planes). It seems fairly clear that these cylinders must all have parallel generators if they and a nowhere dense closed set are somehow going to make up a smooth surface M; but proving this might become quite sticky. Fortunately, there is a direct proof of the global result which makes no use whatsoever of the classical local classification.

9. THEOREM. If M is a complete flat surface and $f\colon M \to \mathbb{R}^3$ is an isometric immersion, then $f(M) \subset \mathbb{R}^3$ is a generalized cylinder.

FIRST PROOF. We can assume that M is simply-connected (by applying the result to $f \circ \pi$ where $\pi\colon \tilde{M} \to M$ is the universal covering space). Then (Problem 1-5) M is isometric to $\mathbb{R}^2$, with its usual Riemannian metric.

We claim first that if $f(M)$ is not simply a plane in $\mathbb{R}^3$, then for every point $p \in M$ there is a *unique* infinite straight line L_p through p which is contained in $f(M)$. Corollaries 6 and 7, together with completeness, show that this is true if p is parabolic or a limit of parabolic points. Now consider a point $p \in f(M)$ which has a whole neighborhood contained in a plane P. Let $Q \subset P$ be the component of $f(M) \cap P$ containing p. If Q is not all of P, let q be a boundary point of Q in P such that all points of the segment $\overline{pq}$ other than q itself lie in the interior of Q (relative to P). The point q must lie in $f(M)$, by completeness,



so q is a planar point which is a limit point of parabolic points. Therefore q is on a unique straight line L_q in $f(M)$. The line L_q must lie in P, since P is the tangent space of q, by continuity of the tangent spaces; moreover, all points of L_q are planar points which are limits of parabolic points.

We claim that all points between p and L_q lie in Q also. Otherwise, there is a point r between p and L_q such that all points of $\overline{pr}$ are in P, but r is also a limit of parabolic points.

There is a corresponding line L_r, and it must be parallel to L_q, since points on L_q and L_r have only one straight line passing through them. Thus L_r intersects $\overline{pq}$ at r'. Then r' is also a limit of parabolic points, which is absurd, since r' is in the interior of Q, a component of $M \cap P$.

The same arguments may be applied if there are any boundary points of Q on the other side of p. Consequently, either $Q = P$, or Q is the part of P bounded by L_q, or Q is the part of P bounded by two parallel lines $L_q, L_{q'}$. Leaving aside the case where $Q = P$ (which occurs only if $f(M) = P$), we see that there is a unique infinite straight line through p in M, namely the one parallel to L_q [and $L_{q'}$]. This proves our claim.

Since f is an isometry, each L_p is the image under f of a geodesic (or possibly many geodesics) in $\mathbb{R}^2$; these geodesics are just ordinary straight lines. In this family of straight lines, distinct lines are disjoint, so our family is the set of lines in $\mathbb{R}^2$ parallel to a fixed line; for convenience we assume that they are all parallel to the y-axis.

Now by completeness, the functions k of Lemma 5 are defined for all s. But this can happen only if $A = 0$. So we see that all k are constants. In other words,

$$k_2(x, y) = \kappa(x)$$

for some function κ. On the other hand, consider the map $f\colon \mathbb{R}^2 \to \mathbb{R}^3$ defined by

$$g(x, y) = (c_1(x), c_2(x), y),$$

where c is a curve in $\mathbb{R}^2$ with curvature function κ. We easily compute that the maps f and g both have second fundamental forms with components

$$l = \kappa, \qquad m = 0, \qquad n = 0.$$

So by the fundamental theorem of surface theory, f differs from the generalized cylinder g by a Euclidean motion.

SECOND PROOF. We replace the last argument, using the fundamental theorem of surface theory, with some very elementary geometry. Any two parallel lines L_1 and L_2 of our family have the property that the function

$$p \mapsto d(p, L_2)$$

is constant on L_1, where $d(p, L_2)$ is the distance (in $\mathbb{R}^2$) from p to L_2. Since f is an isometry, the function $q \mapsto \bar{d}(q, f(L_2))$ must be constant on $f(L_1)$, where $\bar{d}$ denotes the distance in $f(M)$. Now if $f(L_1)$ and $f(L_2)$ were skew lines, or lines intersecting at just one point, then the function

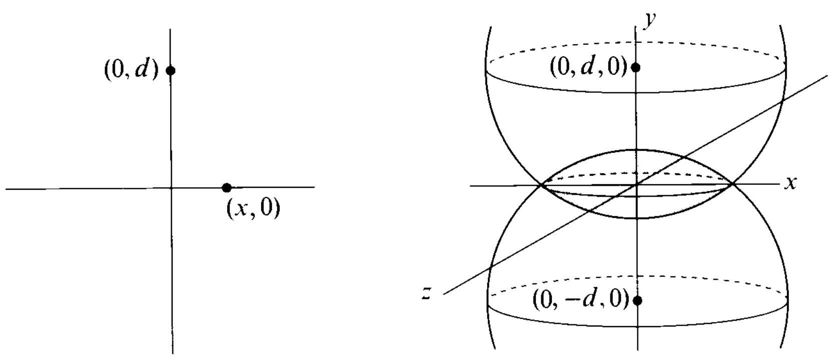
$$q \mapsto \text{Euclidean distance from } q \text{ to } f(L_2)$$

would $\to \infty$ as $q \to \infty$ along L_1. Since

$$\bar{d}(q, f(L_2)) \geq \text{Euclidean distance from } q \text{ to } f(L_2),$$

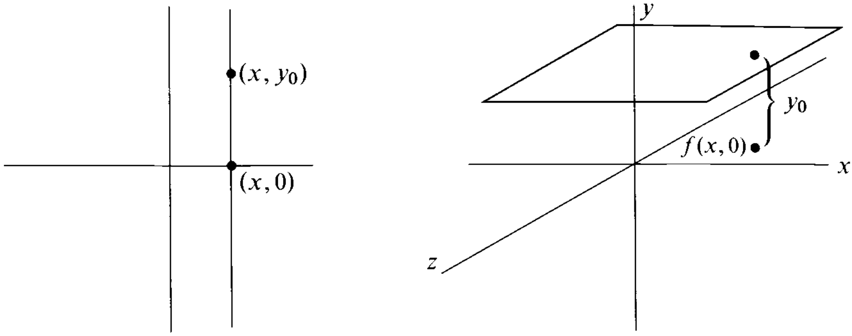
the same would be true for $\bar{d}$, so $\bar{d}$ could not be constant. Thus $f(L_1)$ and $f(L_2)$ must be parallel (or equal). So $f(M)$ is a generalized cylinder.

THIRD PROOF. This time we reduce almost everything to elementary geometry. We merely note that either all points of our surface are planar points, or by Corollaries 6 and 7, and completeness, *some* straight line of $\mathbb{R}^2$ maps to a straight line in $\mathbb{R}^3$; for simplicity we assume that $(0, y) \mapsto (0, y, 0)$. Consider first a point $(x, 0)$ of $\mathbb{R}^2$. Its distance from $(0, d)$ is $\sqrt{x^2 + d^2}$. Since this must be the distance from $f(x, 0)$ to $f(0, d)$ in $f(M)$, the point $f(x, 0)$ must lie in the Euclidean ball around $(0, d, 0)$ of radius $\sqrt{x^2 + d^2}$. Similarly, $f(x, 0)$ must



lie in the Euclidean ball around $(0, -d, 0)$ of radius $\sqrt{x^2 + d^2}$. The intersection of these two balls is a lens-shaped region of height $2(\sqrt{x^2 + d^2} - d)$. Since this $\to 0$ as $d \to \infty$, the point $f(x, 0)$ must lie in the plane $y = 0$.

Now the same argument shows that $f(x, y_0)$ must lie in the plane $y = y_0$.



But $f(x, y_0)$ must also lie in the Euclidean ball of radius y_0 around $f(x,0)$. So $f(x, y_0)$ must be on the line through $f(x,0)$ parallel to the y-axis. ❖

The remainder of this chapter is devoted to the proof of Hilbert's theorem that there are *no* complete surfaces of constant negative curvature K immersed in $\mathbb{R}^3$. There is no loss of generality in considering only the case $K = -1$, since similarities of $\mathbb{R}^3$ multiply K by a (positive) constant. We will actually give two proofs of this result. The second is related to, but considerably simpler than, the original proof of Hilbert [1], while the first is an alternative to Hilbert's argument, due to Holmgren [1]. The proofs depend on several classical formulas for surfaces of constant negative curvature, so in each case a few preparatory results are in order.

10. LEMMA. Let M be a 2-dimensional immersed submanifold of $\mathbb{R}^3$ with constant curvature $K < 0$. Then for every point $p \in M$ there is a diffeomorphism

$$g: (-\varepsilon, \varepsilon) \times (-\varepsilon, \varepsilon) \to M,$$
$$g(0,0) = p,$$

whose parameter curves are asymptotic curves *parameterized by arclength.*

A CLASSICAL PROOF. In Addendum 1 to Chapter 4 we found that for every $p \in M$ there is a diffeomorphism $g: (-\varepsilon, \varepsilon) \times (-\varepsilon, \varepsilon) \to M$, with $g(0,0) = p$, such that the parameter curves are asymptotic curves. By a suitable reparameterization we can clearly arrange that the two parameter curves through $p = g(0,0)$ are parameterized by arclength. Thus we have

(1) $$E(s,0) = 1, \qquad G(0,t) = 1.$$

We now claim that *all* parameter curves are parameterized by arclength. To prove this we note that the Codazzi-Mainardi equations on page 217 can be written

$$(2)\qquad \begin{aligned} (m^2)_1 &= 2\frac{\left[\frac{1}{2}(EG-F^2)_1 + FE_2 - EG_1\right]}{EG-F^2}m^2 \\ (m^2)_2 &= 2\frac{\left[\frac{1}{2}(EG-F^2)_2 + FG_1 - GE_2\right]}{EG-F^2}m^2. \end{aligned}$$

But we also have

$$K = \frac{ln - m^2}{EG-F^2} = \frac{-m^2}{EG-F^2},$$

so that

$$m^2 = (-K)(EG-F^2), \qquad \text{where } K \text{ is a constant.}$$

Substituting in the first equation of (2), we get

$$(-K)(EG-F^2)_1 = 2(-K)\left[\frac{1}{2}(EG-F^2)_1 + FE_2 - EG_1\right],$$

which becomes simply

$$EG_1 - FE_2 = 0.$$

Similarly,

$$-FG_1 + GE_2 = 0.$$

Since $EG - F^2 \neq 0$, this set of linear equations is satisfied only if $E_2 = 0$ and $G_1 = 0$. Together with (1), this shows that $E = 1$ and $G = 1$ everywhere.

SECOND PROOF. The desired result obviously amounts to the following: If Y_1 and Y_2 are linearly independent unit asymptotic vector fields, then $[Y_1, Y_2] = 0$.

First consider an adopted orthonormal moving frame (X_1, X_2, X_3) for which X_1 and X_2 are principal directions, with corresponding principal curvatures k_1 and k_2. Then the forms ψ_i^3 satisfy

$$\psi_i^3 = k_i\theta^i.$$

We also have the Codazzi-Mainardi equations

$$d\psi_1^3 = \omega_1^2 \wedge \psi_2^3, \qquad d\psi_2^3 = -\omega_1^2 \wedge \psi_1^3.$$

Taking the exterior derivative of the equation $\psi_1^3 = k_1\theta^1$ thus yields

$$dk_1 \wedge \theta^1 + k_1\, d\theta^1 = d\psi_1^3 = \omega_1^2 \wedge \psi_2^3 = k_2\omega_1^2 \wedge \theta^2,$$

so that

$$dk_1 \wedge \theta^1 + k_1\omega_1^2 \wedge \theta^2 = k_2\omega_1^2 \wedge \theta^2.$$

Applying this to (X_1, X_2), we find that

$$-X_2(k_1) = (k_2 - k_1) \cdot \omega_1^2(X_1).$$

Similarly, exterior differentiation of the equation $\psi_2^3 = k_2\theta^2$ leads to

$$-X_1(k_2) = (k_2 - k_1) \cdot \omega_1^2(X_2).$$

Thus

$$\omega_1^2 = -\frac{X_2(k_1)}{k_2 - k_1}\theta^1 - \frac{X_1(k_2)}{k_2 - k_1}\theta^2. \tag{1}$$

Now consider a unit vector field

$$\alpha_1 X_1 + \alpha_2 X_2,$$

with

$$(\alpha_1)^2 + (\alpha_2)^2 = 1. \tag{2}$$

For this to be an asymptotic vector field we need

$$k_1(\alpha_1)^2 + k_2(\alpha_2)^2 = 0.$$

For simplicity we now take the case $K = -1$, so that $k_1k_2 = -1$. Then $\alpha_1 X_1 + \alpha_2 X_2$ is asymptotic if and only if

$$(\alpha_2)^2 = -\frac{k_1}{k_2}(\alpha_1)^2 \implies (\alpha_1)^2 - \frac{k_1}{k_2}(\alpha_1)^2 = 1 \qquad \text{by (2)}$$

$$\implies (\alpha_1)^2 = \frac{1}{1 - \dfrac{k_1}{k_2}}$$

$$= \frac{1}{1 + k_1{}^2}, \quad \text{since } k_1k_2 = -1.$$

So our unit asymptotic vector fields must be

$$Y_1 = \pm\alpha_1 X_1 \pm \alpha_2 X_2$$

$$Y_2 = \pm\alpha_1 X_1 \pm \alpha_2 X_2,$$

where the α_i are given by

$$(3) \qquad \alpha_i = \frac{1}{\sqrt{1+k_i{}^2}}.$$

It is convenient to note that we can write equation (1) in terms of the α_i as

$$(4) \qquad \omega^i_j = \frac{X_j(\alpha_i)}{\alpha_i}\theta^i - \frac{X_i(\alpha_j)}{\alpha_j}\theta^j.$$

Now to show that $[Y_1, Y_2] = 0$ we just need to show that $[\alpha_1 X_1, \alpha_2 X_2] = 0$. But we have (see pg. I.215)

$$\begin{aligned}
\theta^i([\alpha_1 X_1, \alpha_2 X_2]) &= \alpha_1 X_1(\theta^i(\alpha_2 X_2)) - \alpha_2 X_2(\theta^i(\alpha_1 X_1)) - d\theta^i(\alpha_1 X_1, \alpha_2 X_2) \\
&= \delta_{i2}\alpha_1 X_1(\alpha_2) - \delta_{i1}\alpha_2 X_2(\alpha_1) + \sum_j (\omega^i_j \wedge \theta^j)(\alpha_1 X_1, \alpha_2 X_2) \\
&= \delta_{i2}\alpha_1 X_1(\alpha_2) - \delta_{i1}\alpha_2 X_2(\alpha_1) \\
&\quad + \delta_{i1}\alpha_2 X_2(\alpha_1) - \delta_{i2}\alpha_1 X_1(\alpha_2) \qquad \text{using (4)} \\
&= 0. \text{ ❖}
\end{aligned}$$

For any 2-dimensional Riemannian manifold M (not necessarily immersed in $\mathbb{R}^3$), an immersion $g\colon (a,b) \times (c,d) \to M$ is called a **Tschebyscheff net** if all parameter curves are parameterized by arclength. If we think of the domain $(a,b) \times (c,d)$ as a piece of cloth woven from fibres parallel to the axes, then the immersion g doesn't stretch any fibres. So the surface can be outfitted in a sexy tight fitting suit if we can find Tschebyscheff nets around each point (we might have to sew a lot of pieces together). Lemma 10 shows that this can always be done on a submanifold of $\mathbb{R}^3$ with constant negative curvature. The notion of a Tschebyscheff net is an intrinsic one, however, and our next result is also.

11. LEMMA. Let M be a 2-dimensional Riemannian manifold and $g\colon (a,b) \times (c,d) \to M$ a Tschebyscheff net. Define $\omega\colon (a,b) \times (c,d) \to \mathbb{R}$ as follows: $\omega(s_0, t_0)$ is the unique number with $0 < \omega(s_0, t_0) < \pi$ such that $\omega(s_0, t_0)$ is an angle between

$$\left.\frac{dg(s,t_0)}{ds}\right|_{s=s_0} \quad \text{and} \quad \left.\frac{dg(s_0,t)}{dt}\right|_{t=t_0}$$

Then ω satisfies the differential equation

$$\frac{\partial^2 \omega}{\partial s \partial t} = (-K)\sin\omega.$$

PROOF. We have $E = G = 1$, and

$$F = \cos\omega, \qquad W = \sqrt{EG - F^2} = \sin\omega.$$

From the equation in Problem 4-13 we obtain

$$\begin{aligned} K &= \frac{1}{2W}\left[\frac{\partial}{\partial t}\left(\frac{F_1}{W}\right) + \frac{\partial}{\partial s}\left(\frac{F_2}{W}\right)\right] \\ &= \frac{1}{2W}\left[\frac{\partial}{\partial t}\left(\frac{-\sin\omega\dfrac{\partial\omega}{\partial s}}{W}\right) + \frac{\partial}{\partial s}\left(\frac{-\sin\omega\dfrac{\partial\omega}{\partial t}}{W}\right)\right] \\ &= \frac{1}{2\sin\omega}\left[\frac{\partial}{\partial t}\left(-\frac{\partial\omega}{\partial s}\right) + \frac{\partial}{\partial s}\left(-\frac{\partial\omega}{\partial t}\right)\right] \\ &= -\frac{\dfrac{\partial^2\omega}{\partial s\partial t}}{\sin\omega}. \end{aligned}$$ ❖

We are now ready to prove the theorem, which still requires quite a bit of argument. We will use the term **asymptotic Tschebyscheff net** for a Tschebyscheff net of the sort constructed in Lemma 10, with all parameter curves being asymptotic curves.

12. THEOREM. A complete surface M with constant curvature $K = -1$ cannot be immersed in $\mathbb{R}^3$.

PROOF. The proof depends on establishing two facts:

(A) Suppose that M could be immersed in $\mathbb{R}^3$. Then there would be a Tschebyscheff net $f\colon \mathbb{R}^2 \to M$, from the whole plane to M, and the function ω, defined on all of $\mathbb{R}^2$, which gives the angle between the first and second parameter lines would satisfy

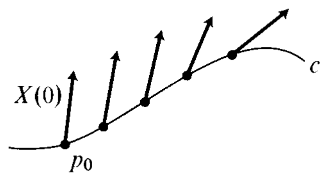
$$\frac{\partial^2\omega}{\partial s\partial t} = \sin\omega, \qquad 0 < \omega < \pi.$$

(B) There is no function $\omega\colon \mathbb{R}^2 \to \mathbb{R}$ satisfying

$$\frac{\partial^2\omega}{\partial s\partial t} = C\sin\omega, \qquad 0 < \omega < \pi,$$

where $C > 0$ is any constant.

PROOF OF (A). Select a point $p_0 \in M$. Let $c\colon \mathbb{R} \to M$ be an asymptotic curve, parameterized by arclength, with $c(0) = p_0$; since c is locally an integral curve of a *unit* vector field, it can be defined on all of $\mathbb{R}$ since M is complete [compare Problem 1-5(c)]. Let $X(0)$ be a unit asymptotic vector at p_0 which is linearly independent of $c'(0)$, and let $X(t)$ be the unique continuous vector field X along c such that $X(t) \in M_{c(t)}$ is a unit asymptotic vector linearly independent



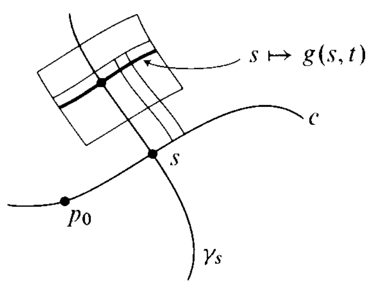
of $c'(t)$. The vector field X along c is just a device to enable us to distinguish a direction for each parameter value of c. We now define $f\colon \mathbb{R}^2 \to M$ as follows:

$$f(s,t) = \gamma_s(t),$$

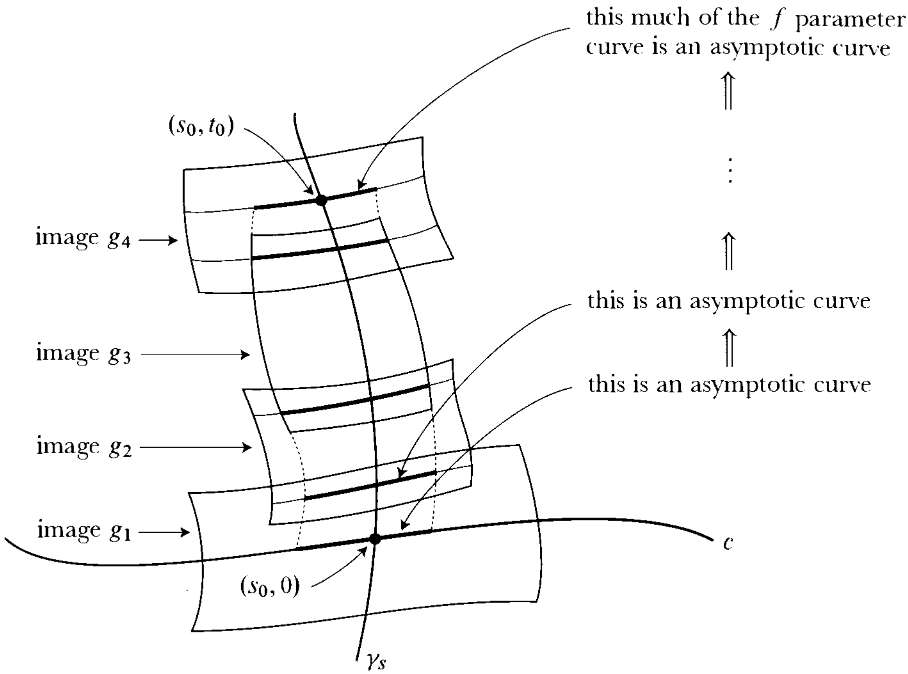
where γ_s is the unique asymptotic curve, parameterized by arclength, with $\gamma_s(0) = c(s)$ and $\gamma_s{}'(0) = X(s)$.

What we have to show is that each curve $s \mapsto f(s,t)$ is an asymptotic curve. This depends on the existence, as guaranteed by Lemma 10, of asymptotic Tschebyscheff nets $g\colon (-\varepsilon,\varepsilon) \times (-\varepsilon,\varepsilon) \to M$ around any point. From the very definition of f the following at least is clear:

> *Observation*: If for some $t \in (-\varepsilon,\varepsilon)$ the parameter curve $s \mapsto g(s,t)$ lies along a parameter curve $s \mapsto f(s,\bar{t})$, then *all* parameter curves $s \mapsto g(s,t)$ lie along parameter curves of f.



Now for any (s_0, t_0) we can find a finite number of asymptotic Tschebyscheff nets $g_1, \ldots, g_k$ whose images cover $\{f(s_0, t): 0 \le t \le t_0\}$. Arranging the g_i as in the picture below, so that the images of consecutive ones overlap, noting that



$s \mapsto f(s, 0)$ is an asymptotic curve by definition, and applying the Observation repeatedly, we see that $s \mapsto f(s, t_0)$ is an asymptotic curve for s sufficiently close to s_0, which is what we wanted.

Our equation for ω then follows immediately from Lemma 11.

PROOF OF (B). Suppose we had a function $\omega: \mathbb{R}^2 \to \mathbb{R}$ satisfying

(1) $$\frac{\partial^2 \omega}{\partial s \partial t} = C \sin \omega, \qquad 0 < \omega < \pi,$$

for a constant $C > 0$, and hence, in particular,

(2) $$\frac{\partial^2 \omega}{\partial s \partial t} > 0.$$

This implies that $\partial \omega / \partial s$ is increasing as a function of t, so that

(3) $$\frac{\partial \omega}{\partial s}(s, t) > \frac{\partial \omega}{\partial s}(s, 0) \qquad \text{for } t > 0.$$

Consequently, for $t > 0$ we have

$$\int_a^b \frac{\partial\omega}{\partial s}(s,t)\,ds > \int_a^b \frac{\partial\omega}{\partial s}(s,0)\,ds,$$

so that

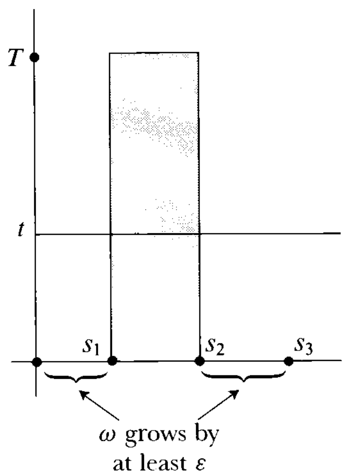
$$\text{(4)} \qquad \omega(b,t) - \omega(a,t) > \omega(b,0) - \omega(a,0) \qquad \text{for } t > 0 \text{ and } a < b.$$

Now we can't have $\partial\omega/\partial s = 0$ everywhere, so we can assume (changing our coordinates by a translation) that $\partial\omega/\partial s(0,0) \neq 0$. Since the function $(s,t) \mapsto \omega(-s,-t)$ also satisfies (1), we can even assume that $\partial\omega/\partial s(0,0) > 0$. Choose three fixed numbers

$$\text{(5)} \qquad 0 < s_1 < s_2 < s_3 \qquad \text{with} \qquad \frac{\partial\omega}{\partial s}(s,0) > 0 \quad \text{for } 0 \le s \le s_3,$$

and let

$$\varepsilon = \min \begin{cases} \omega(s_3,0) - \omega(s_2,0) \\ \omega(s_1,0) - \omega(0,0). \end{cases}$$



Then for all $t > 0$ and all $s \in [0, s_3]$ we have the following:

$$\begin{aligned} &\omega(s,t) \text{ is increasing in } s, && \text{by (3) and (5)} \\ &\left.\begin{aligned} &\omega(s_1,t) - \omega(0,t) > \varepsilon \\ &\omega(s_3,t) - \omega(s_2,t) > \varepsilon \end{aligned}\right\}, && \text{by (4) and the definition of } \varepsilon \\ &0 < \omega(s,t) < \pi. \end{aligned}$$

Putting these together, we conclude that

$$\varepsilon \leq \omega(s,t) \leq \pi - \varepsilon \qquad \text{for } s \in [s_1, s_2] \text{ and } t \geq 0,$$

and hence

$$\sin \omega(s,t) \geq \sin \varepsilon \qquad \text{for } s \in [s_1, s_2] \text{ and } t \geq 0. \tag{6}$$

But suppose we integrate equation (1) over the rectangle $[s_1, s_2] \times [0, T]$. We obtain

$$\begin{aligned} C \int_0^T \int_{s_1}^{s_2} \sin \omega(s,t)\, ds\, dt &= \int_0^T \int_{s_1}^{s_2} \frac{\partial^2 \omega}{\partial s \partial t}\, ds\, dt \\ &= [\omega(s_2, T) - \omega(s_1, T) - \omega(s_2, 0) + \omega(s_1, 0)], \end{aligned}$$

or

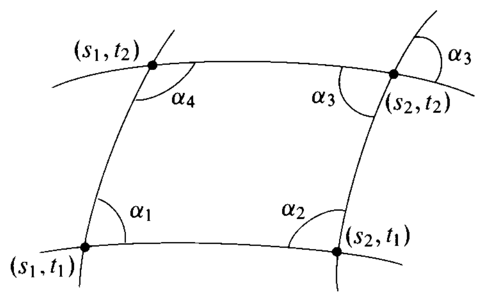
$$\begin{aligned} \omega(s_2, T) - \omega(s_1, T) &= \omega(s_2, 0) - \omega(s_1, 0) + C \int_0^T \int_{s_1}^{s_2} \sin \omega(s,t)\, ds\, dt \\ &\geq \omega(s_2, 0) - \omega(s_1, 0) + CT(s_2 - s_1) \sin \varepsilon, \qquad \text{by (6).} \end{aligned}$$

Taking T large enough, we get a contradiction, since the left side is $< \pi$. ❖

For the second proof of Hilbert's theorem, we need an observation which follows directly from Lemma 11.

13. LEMMA (HAZZIDAKIS' FORMULA). Let M be a 2-dimensional Riemannian manifold of constant curvature $K < 0$, and let $g\colon (a,b) \times (c,d) \to M$ be a Tschebyscheff net. Then any quadrilateral Q formed by parameter curves has area

$$\begin{aligned} \text{area}(Q) &= \frac{1}{-K}\left(\sum_{i=1}^{4} \alpha_i - 2\pi\right) \\ &\leq \frac{2\pi}{-K}, \end{aligned}$$



where $\alpha_i \in (0, \pi)$ are the interior angles of Q.

PROOF. Introducing the function ω of Lemma 11, we have

$$dA = W\,ds \wedge dt = \sin\omega\, ds \wedge dt$$
$$\sin\omega = \frac{1}{-K}\frac{\partial^2\omega}{\partial s\partial t}.$$

So

$$\begin{aligned}
\text{area}(Q) &= \int_Q dA = \int_Q \sin\omega\, ds \wedge dt \\
&= \frac{1}{-K}\int_Q \frac{\partial^2\omega}{\partial s\partial t}\, ds \wedge dt \\
&= \frac{1}{-K}\int_{s_1}^{s_2}\int_{t_1}^{t_2} \frac{\partial^2\omega}{\partial s\partial t}\, ds\, dt \\
&= \frac{1}{-K}[\omega(s_2,t_2) - \omega(s_1,t_2) - \omega(s_2,t_1) + \omega(s_1,t_1)] \\
&= \frac{1}{-K}[\alpha_3 - (\pi - \alpha_4) - (\pi - \alpha_2) + \alpha_1] \\
&= \frac{1}{-K}\left(\sum_{i=1}^{4}\alpha_i - 2\pi\right). \quad ❖
\end{aligned}$$

On the other hand, consider the upper half-plane $\mathcal{H}^2 \subset \mathbb{R}^2$ with the Riemannian metric

$$\langle\ ,\ \rangle = \frac{dx \otimes dx + dy \otimes dy}{y^2}.$$

This is a complete 2-dimensional Riemannian manifold of constant curvature $K = -1$ (see pg. II.301 and Problem I.9-41). We have

$$E = G = \frac{1}{y^2} \qquad\qquad F = 0$$
$$dA = \sqrt{EG - F^2}\, dx \wedge dy = \frac{1}{y^2}\, dx \wedge dy,$$

so the total area of $\mathcal{H}^2$ is

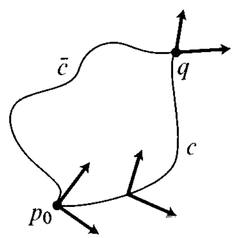
$$\int_{-\infty}^{\infty}\int_0^{\infty} \frac{1}{y^2}\, dy\, dx = \infty.$$

Thus Lemma 13 shows that we *cannot* parameterize all of $\mathcal{H}^2$ by a Tschebyscheff net. This is the basis of our second proof of

12. THEOREM. A complete surface M of constant curvature $K = -1$ cannot be immersed in $\mathbb{R}^3$.

PROOF. As in the proof of Theorem 9, we can assume that M is simply-connected. Then (Problem 1-5) M is isometric to $(\mathcal{H}^2, \langle\ ,\ \rangle)$.

We claim first that there are two linearly independent unit asymptotic vector fields Y_1, Y_2 defined *on all of* M. The proof uses the fact that M is simply-connected, and follows a standard procedure. We arbitrarily select $(Y_1(p_0), Y_2(p_0))$ for some $p_0 \in M$. Then for every curve $c\colon [0,1] \to M$ with $c(0) = p_0$, there will be a unique possible continuous choice of $(Y_1(c(t)), Y_2(c(t)))$ for all $t \in [0,1]$ which extends $(Y_1(p_0), Y_2(p_0))$. If $\bar{c}$ is another such curve with corresponding



$(\bar{Y}_1(c(t)), \bar{Y}_2(c(t)))$, and if moreover $\bar{c}(1) = c(1) = q$, then $\bar{Y}_i(q) = Y_i(q)$; the proof uses the usual argument involving a contraction to the constant path at p_0 of the curve c followed by $\bar{c}$ in the reverse direction. A nicer argument is the following. For each p there are 4 possible choices for $Y_1(p)$, and then 2 possible choices for $Y_2(p)$, thus 8 possible choices for $(Y_1(p), Y_2(p))$. The set of all such pairs, for all $p \in M$, obviously forms an 8-fold covering space of M. Since M is simply-connected, this covering space consists of 8 components; any component gives us the desired pair of vector fields.

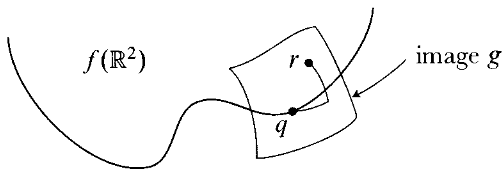
We now claim that the Tschebyscheff net $f\colon \mathbb{R}^2 \to M$ which we constructed in the first proof is actually a *diffeomorphism*. To prove this, we first describe f a little differently. Let $\{\phi_t^i\}$ be the 1-parameter group of diffeomorphisms generated by Y_i (recall this means that $t \mapsto \phi_t^i(p)$ is the integral curve of Y_i through p, for each p); the ϕ_t^i can be defined for all $t \in \mathbb{R}$ since M is complete and Y_i are unit vector fields (as in our first proof). We now pick a point $p_0 \in M$ and define

$$f(s,t) = \phi_t^2(\phi_s^1(p_0)). \tag{1}$$

Since $[Y_1, Y_2] = 0$ (Lemma 10), the 1-parameter groups ϕ^1, ϕ^2 commute (Lemma I.5-13), so we easily find that

$$(2) \qquad f(s+s', t+t') = \phi^2_{t'}(\phi^1_{s'}(f(s,t))).$$

We claim first that f is onto M. Otherwise, there is a point $q \notin f(\mathbb{R}^2)$ with q on the boundary of $f(\mathbb{R}^2)$. Now there is an asymptotic Tschebyscheff net $g: (-\varepsilon, \varepsilon) \times (-\varepsilon, \varepsilon) \to M$ with $g(0,0) = q$, and there is some point $r \in$



(image g) $\cap f(\mathbb{R}^2)$. Then q must be of the form

$$\begin{aligned} q &= \phi^2_{t'}(\phi^1_{s'}(r)) && \text{for some } s', t' \\ &= \phi^2_{t'}(\phi^1_{s'}(f(s,t))) && \text{for some } s, t \\ &= f(s+s', t+t') && \text{by (2),} \end{aligned}$$

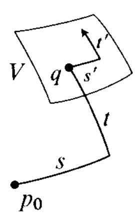
so actually $q \in f(\mathbb{R}^2)$, a contradiction.

Now we claim that $f: \mathbb{R}^2 \to M$ is actually a covering map. For any point $q \in M$, choose an asymptotic Tschebyscheff net $g: (-2\varepsilon, 2\varepsilon) \times (-2\varepsilon, 2\varepsilon) \to M$ with $g(0,0) = q$ such that g is a diffeomorphism onto some open subset $V \subset M$. We can assume that the s [and t] parameter curves of g lie along the s [and t] parameter curves of f. Suppose that $(s,t) \in f^{-1}(q)$. Consider the map

$$\phi: V \to (s - 2\varepsilon, s + 2\varepsilon) \times (t - 2\varepsilon, t + 2\varepsilon) \subset \mathbb{R}^2$$

defined by

$$g(s', t') \mapsto (s + s', t + t').$$



Using equation (2), we see that $f\colon (s-2\varepsilon, s+2\varepsilon)\times(t-2\varepsilon, t+2\varepsilon)\to V$ is a diffeomorphism with inverse ϕ. Now let

$$W = g\big((-\varepsilon,\varepsilon)\times(-\varepsilon,\varepsilon)\big)$$

and for each $(s,t)\in\mathbb{R}^2$ let

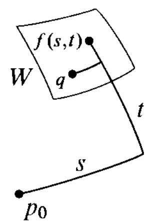
$$W_{(s,t)} = (s-\varepsilon, s+\varepsilon)\times(t-\varepsilon, t+\varepsilon).$$

We claim that

$$f^{-1}(W) = \bigcup_{(s,t)\in f^{-1}(q)} W_{(s,t)}.$$

In fact, if

$$(s,t)\in f^{-1}(W),$$



then we have

$$f(s,t) = g(s',t') \qquad \text{for some} \quad (s',t')\in(-\varepsilon,\varepsilon)\times(-\varepsilon,\varepsilon),$$

and by equation (2),

$$\begin{aligned} f(s-s', t-t') &= \phi^2_{-t'}(\phi^1_{-s'}(f(s,t))) \\ &= \phi^2_{-t'}(\phi^1_{-s'}(g(s',t'))) \\ &= q, \end{aligned}$$

which proves the claim. Since each of the $W_{(s,t)}$ is mapped diffeomorphically onto W, the proof that f is a covering map will be complete once we show that any two distinct such rectangles $W_{(s_1,t_1)}$ and $W_{(s_2,t_2)}$ are disjoint. Now if $W_{(s_1,t_1)}\cap W_{(s_2,t_2)}\neq\emptyset$, then

$$(s_2,t_2)\in(s_1-2\varepsilon, s_1+2\varepsilon)\times(t_1-2\varepsilon, t_1+2\varepsilon).$$

But we know that f is a diffeomorphism on this rectangle. Since $f(s_1,t_1) = q = f(s_2,t_2)$, this means that $(s_1,t_1)=(s_2,t_2)$, so that $W_{(s_1,t_1)}$ and $W_{(s_2,t_2)}$ are actually the same. Thus f is indeed a covering map.

Now by the simple-connectivity of M we conclude that f is actually a diffeomorphism. Consequently, we can exhaust M by quadrilaterals formed by the parameter curves of f. Lemma 13 then implies that M has area $\leq 2\pi$, while we computed that M has infinite area. This contradiction establishes the theorem. ❖

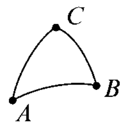
It is easy to find a complete surface $M \subset \mathbb{R}^3$ with non-constant curvature $K < 0$ everywhere—for example, the elliptic hyperboloid of one sheet has this property. But in this example K comes arbitrarily close to 0. In 1964, Efimov [1] proved, by a lengthy ingenious argument, using no particularly sophisticated machinery, that there are no complete surfaces $M \subset \mathbb{R}^3$ with curvature $K < 0$ bounded away from 0; an exposition of the proof may be found in Klotz [2].

CHAPTER 6

THE GAUSS-BONNET THEOREM AND RELATED TOPICS

In Volume II we presented Gauss' proof that if $\triangle ABC$ is a geodesic triangle on a surface with Gaussian curvature K, then

$$\int_{\triangle ABC} K\,dA = \angle A + \angle B + \angle C - \pi.$$

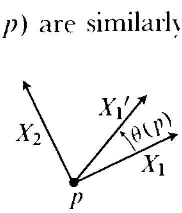


At that time we also cast a suspicious glance at Stokes' Theorem, which seemed to be lurking in the background, and promised to present a proof which would implicate it more fully. We are now in a position to redeem that pledge. It is almost a foregone conclusion that moving frames will play a leading role in the proceedings, since this is the only treatment of curvature in which differential forms appear explicitly. Actually, we are going to generalize Gauss' result, and in two quite different directions. On the one hand, we will allow polygons with any number of sides, and we will not require the sides to be geodesics. On the other hand, we will also have something to say about the integral of K over a whole surface. We begin much more modestly, however, with quite elementary considerations.

Suppose that $\mathbf{X} = (X_1, X_2)$ and $\mathbf{X}' = (X'_1, X'_2)$ are two orthonormal moving frames on a 2-dimensional Riemannian manifold M. We have already found the relationship between the matrices of 1-forms $\omega = (\omega^i_j)$ and $\omega' = (\omega'^i_j)$ associated to these moving frames: If $\mathbf{X}' = \mathbf{X} \cdot a$ for an orthonormal matrix function a, then (see pg. II.280) we have

$$(1) \qquad \omega' = a^{-1}\,da + a^{-1}\omega a.$$

Of course, in a 2-dimensional manifold this relationship can be expressed much more simply. If $\mathbf{X}(p)$ and $\mathbf{X}'(p)$ are similarly oriented, then the matrix $a(p)$



is just

$$a(p) = \begin{pmatrix} \cos\theta(p) & \sin\theta(p) \\ -\sin\theta(p) & \cos\theta(p) \end{pmatrix},$$

where $\theta(p)$ is the oriented angle between $X_1(p)$ and $X'_1(p)$.

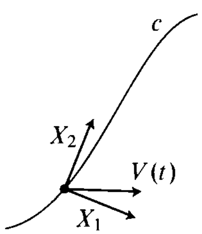
Usually we define the "angle" between two vectors to be a number between $-\pi$ and π, but then the function θ need not be continuous. Locally, we can make θ differentiable by allowing other values of θ. In the next result, it does not matter that this θ is not well-defined, because the form $d\theta$ still is.

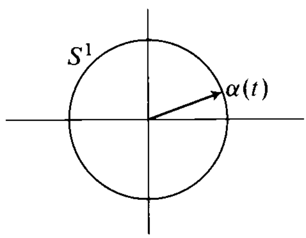
1. PROPOSITION. Let X_1, X_2 and X'_1, X'_2 be two similarly oriented orthonormal moving frames on a 2-dimensional Riemannian manifold M, and let ω_1^2, ω'^2_1 be the corresponding connection forms. Let θ be a differentiable choice of the angle between X_1 and X'_1. Then

$$\omega'^2_1 = \omega_1^2 + d\theta.$$

PROOF. It is easily checked that this is precisely what equation (1) comes down to. It is probably also a good exercise to derive the whole thing from scratch, using properties of ∇, or by reproving Proposition II. 7-14 for 2-manifolds. ❖

Now consider a curve $c\colon [a,b] \to M$ which lies in a region on which we have an orthonormal moving frame X_1, X_2. Suppose that V is a unit vector field along c. We can then define the angle between V and X_1 in an even more precise manner, based on the constructions on pp. II. 16–18. We first define a map $\alpha\colon [a,b] \to S^1$ by letting $\alpha(t)$ be the image of $V(t)$ under the unique linear map $M_{c(t)} \to \mathbb{R}^2$ which takes $X_1(c(t))$ to e_1. We then have, by





Proposition II. 1-5, a continuous map $\phi\colon [a,b] \to \mathbb{R}$ with

$$\alpha(t) = (\cos\phi(t), \sin\phi(t)),$$

and any two such maps differ by a multiple of 2π. We will refer to any such ϕ as a continuous choice of the angle between X_1 and V. It is easy to see that ϕ is actually C^∞ if V is a C^∞ vector field along c.

2. COROLLARY. Let M be a 2-dimensional Riemannian manifold, and let $c\colon [a,b] \to M$ be an immersed curve which lies in a region on which we have a positively oriented orthonormal moving frame X_1, X_2. Let V be a C^∞ unit vector field along c, and let ϕ be a continuous choice of the angle between X_1 and V. Then V is parallel along c if and only if

$$\omega_1^2(c'(t)) + \phi'(t) = 0.$$

PROOF. Locally we can find an orthonormal moving frame X'_1, X'_2 oriented similarly to X_1, X_2, with $X'_1 = V$ along c. We can choose the angle $\theta(p)$ between $X'_1(p)$ and $X_1(p)$ so that

$$\theta(c(t)) = \phi(t).$$

Then V is parallel along c if and only if

$$\begin{aligned} 0 = \langle \nabla_{c'(t)} X'_1, X'_2 \rangle &= \omega'^2_1(c'(t)) \\ &= \omega_1^2(c'(t)) + d\theta(c'(t)) \qquad \text{by Proposition 1} \\ &= \omega_1^2(c'(t)) + (\theta \circ c)'(t) \\ &= \omega_1^2(c'(t)) + \phi'(t). \end{aligned}$$ ❖

In Chapter 4 we defined the geodesic curvature κ_g of a curve c in an oriented surface $M \subset \mathbb{R}^3$. We also noted that κ_g is intrinsic. Indeed, for any arclength parameterized curve c in any Riemannian manifold $(M, \langle\ ,\ \rangle)$ we can define κ_g (≥ 0) as the norm of $Dc'(s)/ds$; if M is an oriented 2-dimensional Riemannian manifold, then we defined the signed geodesic curvature κ_g by

$$\kappa_g(s) = \left\langle \frac{Dc'(s)}{ds}, \mathbf{u}(s) \right\rangle,$$

where $\mathbf{u}(s) \in M_{c(s)}$ is the unit vector perpendicular to $c'(s)$ with $(c'(s), \mathbf{u}(s))$ positively oriented.

3. COROLLARY. Let M be an oriented 2-dimensional Riemannian manifold, and let $c\colon [a,b] \to M$ be a curve, parameterized by arclength, which lies in a region on which we have a positively oriented orthonormal moving frame X_1, X_2. Let ϕ be a continuous choice of the angle between X_1 and $c'(s)$. Then the signed geodesic curvature κ_g of c is given by

$$\kappa_g(s) = \omega_1^2(c'(s)) + \phi'(s).$$

PROOF. Locally we can find a positively oriented orthonormal moving frame X'_1, X'_2 with $X'_1 = c'$ along c. So we can choose the angle $\theta(p)$ between $X'_1(p)$ and $X_1(p)$ so that

$$\theta(c(s)) = \phi(s).$$

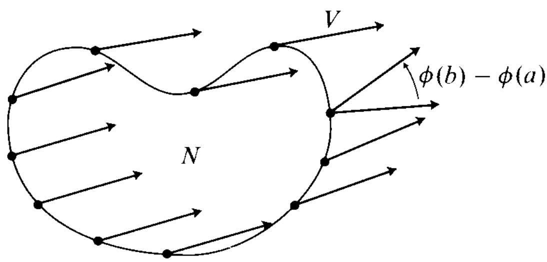
Then

$$\begin{aligned}\kappa_g(s) &= \langle \nabla_{X'_1} X'_1, X'_2\rangle(c(s)) = \omega'^2_1(c'(s))\\ &= \omega_1^2(c'(s)) + d\theta(c'(s)) \qquad \text{by Proposition 1}\\ &= \omega_1^2(c'(s)) + \phi'(s).\end{aligned}$$ ❖

Each of our Corollaries can be used to obtain an interesting result. The first theorem partially fulfills a promise made on pg. II.243, for it gives a quantitative description of the fact that parallel translation along a closed curve generally does not bring a vector back to itself.

4. THEOREM. Let M be an oriented 2-dimensional Riemannian manifold, with Gaussian curvature K, and volume element dA. Let $N \subset M$ be a compact 2-dimensional manifold-with-boundary whose boundary ∂N is connected, let $c\colon [a,b] \to \partial N$ be a closed curve such that $c'(t)$ is positively oriented (with respect to the induced orientation on ∂N), and let V be a parallel unit vector field along c. If X_1, X_2 is a positively oriented moving frame defined on N, and $\phi\colon [a,b] \to \mathbb{R}$ is a continuous choice of the angle between X_1 and V, then

$$\phi(b) - \phi(a) = \int_N K\, dA.$$



PROOF. By the equations on page 69 we have

$$\begin{aligned}\int_N K\,dA &= \int_N K\theta^1 \wedge \theta^2 = -\int_N d\omega_1^2 \\ &= -\int_{\partial N} \omega_1^2 \qquad \text{by Stokes' Theorem} \\ &= -\int_a^b \omega_1^2(c'(t))\,dt \\ &= \int_a^b \phi'(t)\,dt \qquad \text{by Corollary 2} \\ &= \phi(b) - \phi(a). \ ❖\end{aligned}$$

Notice that in order to measure the change in V effected by parallel translation, we made use of an orthonormal moving frame X_1, X_2 defined on all of N. Such a moving frame always exists, because a unit vector field X_1 exists on the *bounded* manifold N (see Problem I.11-13(e)), and X_2 is then determined by the orientation.

In our next theorem, the region N must be very special.

5. THEOREM. Let M be an oriented 2-dimensional Riemannian manifold, with Gaussian curvature K, and volume element dA. Let $N \subset M$ be a compact 2-dimensional manifold-with-boundary which is diffeomorphic to a subset of $\mathbb{R}^2$, and whose boundary is connected. Let ds be the volume element of ∂N (determined by the induced Riemannian metric and induced orientation of ∂N), and let κ_g be the signed geodesic curvature of ∂N (on which we have a direction determined by the induced orientation). Then

$$\int_N K\,dA = -\int_{\partial N} \kappa_g\,ds + 2\pi.$$

PROOF. Because of our hypotheses on N, we might as well assume that M is an open subset of $\mathbb{R}^2$. On M we define a positively oriented orthonormal moving frame X_1, X_2 by requiring X_1 to be a positive multiple of $\partial/\partial x^1$. Let $c\colon [a,b] \to \partial N$ be a closed curve, parameterized by arclength σ, such that $c'(\sigma)$ is positively oriented, and let ϕ be a continuous choice of the angle between X_1

and $c'(\sigma)$. Then

$$\begin{aligned}\int_N K\,dA &= \int_N K\theta^1\wedge\theta^2 = -\int_N d\omega_1^2 = -\int_{\partial N}\omega_1^2\\ &= -\int_a^b \omega_1^2(c'(\sigma))\,d\sigma\\ &= -\int_a^b \kappa_g(\sigma)\,d\sigma + \int_a^b \phi'(\sigma)\,d\sigma \qquad \text{by Corollary 3}\\ &= -\int_{\partial N}\kappa_g\,ds + \phi(b) - \phi(a).\end{aligned}$$

To complete the proof, we have to show that $\phi(b) - \phi(a) = 2\pi$. This is done by noting three things.

(1) The number $\phi(b) - \phi(a)$ is a multiple of 2π, since $\phi(b)$ and $\phi(a)$ are both choices of the angle for $c'(a) = c'(b)$.

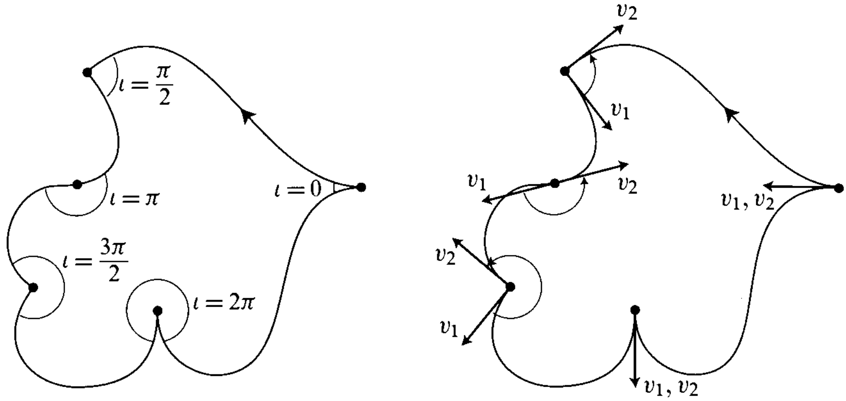
(2) Let $\langle\ ,\ \rangle$ be the Riemannian metric on M, and let $\langle\!\langle\ ,\ \rangle\!\rangle$ be the usual Riemannian metric on $\mathbb{R}^2$. It is easily checked that for each $t \in [0,1]$, the tensor

$$\langle\ ,\ \rangle^t = t\langle\ ,\ \rangle + (1-t)\langle\!\langle\ ,\ \rangle\!\rangle$$

is also a Riemannian metric. Let ${X^t}_1, {X^t}_2$ be a positively oriented moving frame which is orthonormal with respect to $\langle\ ,\ \rangle^t$, and for which ${X^t}_1$ is a positive multiple of $\partial/\partial x^1$; then let ϕ^t be a continuous choice of the angle between ${X^t}_1$ and $c'(\sigma)/\|c'(\sigma)\|^t$. The choice ϕ^t clearly depends continuously on t (if we make $\phi^t(a)$ vary continuously). Consequently, $\phi^t(b) - \phi^t(a)$ varies continuously with t. Since it is always a multiple of 2π, it must be constant.

(3) When $t = 0$, we have ${X^t}_1 = \partial/\partial x^1$ and ${X^t}_2 = \partial/\partial x^2$, and consequently ϕ^0 is just a choice of the angle between the x-axis and c'. So $\phi^0(b) - \phi^0(a)$ is the total curvature of c, as defined on pg. II.18. By the *Hopf Umlaufsatz* (Theorem II.1-7), this total curvature is 2π. ❖

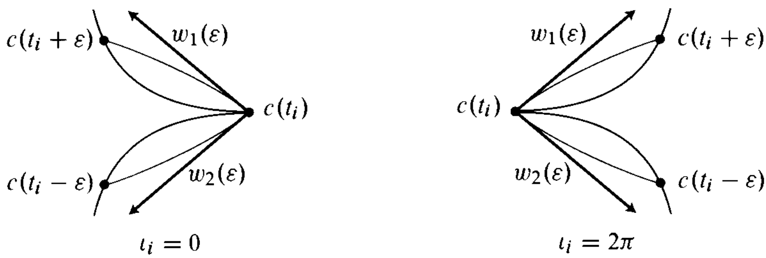
We would like to generalize Theorem 5 slightly, so that the boundary of N need not be smooth, but only piecewise smooth. The proof itself will go through almost precisely as before, and the real problem is to formulate the definitions and state the results correctly (something almost no one ever bothers to do). We will say that a compact 2-dimensional manifold-with-boundary $N \subset M$ is a **polygon** if ∂N is connected and if there is a simple closed curve $c\colon [a,b] \to N$ such that c is a smooth imbedding on each interval $[t_{i-1}, t_i]$ of some partition $a = t_0 < \cdots < t_{n+1} = b$. Thus $c'(t)$ exists for all $t \neq t_i$, and the right and left hand derivatives $c'(t_i{}^+)$ and $c'(t_i{}^-)$ exist for all t_i. It will be convenient to work only with curves c such that $c'(b^-) = c'(a^+)$. The **vertices** of c will then be $t_1, \dots, t_n$; for each such vertex t_i, we would like to define its **interior angle** $\iota_i \in [0, 2\pi]$, as shown below. To do this, we first choose c so that $c'(t)$ is always



positively oriented. Let $v_1 = c'(t_i{}^+)$ and $v_2 = -c'(t_i{}^-)$. If v_1 and v_2 do not point in the same direction, we define

$$\iota_i = \text{oriented angle (between 0 and } 2\pi\text{) from } v_1 \text{ to } v_2.$$

This still leaves us with the problem of defining ι_i when v_1 and v_2 point in the same direction. To treat this case, let $w_1(\varepsilon)$ be the tangent vector of the geodesic from $c(t_i)$ to $c(t_i + \varepsilon)$, and let $w_2(\varepsilon)$ be the tangent vector of the geodesic from $c(t_i)$ to $c(t_i - \varepsilon)$. For sufficiently small $\varepsilon > 0$, the vectors $w_1(\varepsilon)$

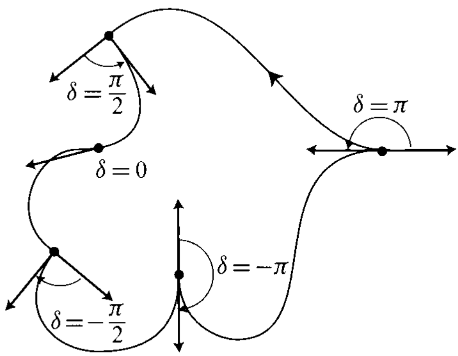


and $w_2(\varepsilon)$ are nearly in the same direction, so, in particular, they do not point in opposite directions. Therefore, the orientation of $(w_1(\varepsilon), w_2(\varepsilon))$ cannot change, since $w_1(\varepsilon)$ and $w_2(\varepsilon)$ are always distinct for small ε. We define ι_i to be 0 if $(w_1(\varepsilon), w_2(\varepsilon))$ is positively oriented, and 2π if it is negatively oriented. [The formula

$$\iota_i = \lim_{\varepsilon \to 0} \{\text{oriented angle from } w_1(\varepsilon) \text{ to } w_2(\varepsilon)\}$$

could be used as a general definition that would work in all cases.] Now that we have successfully defined the interior angle ι_i, we define the **discontinuity** δ_i of c' at t_i by

$$\delta_i = \pi - \iota_i \in [-\pi, \pi].$$



Remark: It is easy to see that if ϕ is an angle between $c'(t_i{}^-)$ and some vector $X \in M_{c(t_i)}$, then $\phi + \delta_i$ is an angle between $c'(t_i{}^+)$ and X.

6. THEOREM (THE GAUSS-BONNET FORMULA). Let M be an oriented 2-dimensional Riemannian manifold, with Gaussian curvature K, and volume element dA. Let $N \subset M$ be a polygon which is diffeomorphic to a subset of $\mathbb{R}^2$, let ds be the volume element of ∂N, and let κ_g be its signed geodesic curvature. Suppose that ∂N has vertices $t_1, \dots, t_n$, with discontinuities $\delta_1, \dots, \delta_n$. Then

$$\begin{aligned}\int_N K\,dA &= -\int_{\partial N} \kappa_g\,ds - \sum_{i=1}^{n} \delta_i + 2\pi \\ &= -\int_{\partial N} \kappa_g\,ds + \sum_{i=1}^{n} \iota_i + (2-n)\pi.\end{aligned}$$

PROOF. As in the proof of the previous theorem, we assume that M is an open subset of $\mathbb{R}^2$, and we define X_1, X_2 exactly as before. Choose the curve $c\colon [a, b] \to \partial N$ to be parameterized by arclength σ, and so that $c'(\sigma)$ is positively oriented. By our Remark, we can choose $\phi_i : [t_{i-1}, t_i] \to \mathbb{R}$ so that each ϕ_i is a continuous choice of the angle between X_1 and $c'(\sigma)$ on (t_{i-1}, t_i), and so that

$$(*) \qquad \phi_{i+1}(t_i) - \phi_i(t_i) = \delta_i \qquad i = 1, \dots, n.$$

Then

$$\begin{aligned}
\int_N K\,dA &= \int_N K\,d\theta^1 \wedge d\theta^2 = -\int_N d\omega_1^2 = -\int_{\partial N} \omega_1^2 \\
&= -\sum_{i=1}^{n+1} \int_{t_{i-1}}^{t_i} \omega_1^2(c'(\sigma))\,d\sigma \\
&= -\sum_{i=1}^{n+1} \left[\int_{t_{i-1}}^{t_i} \kappa_g(\sigma)\,d\sigma - \int_{t_{i-1}}^{t_i} \phi_i{}'(\sigma)\,d\sigma \right] \qquad \text{by Corollary 3} \\
&= -\int_{\partial N} \kappa_g\,ds + \sum_{i=1}^{n+1} \phi_i(t_i) - \phi_i(t_{i-1}) \\
&= -\int_{\partial N} \kappa_g\,ds - \sum_{i=1}^{n} \delta_i + [\phi_{n+1}(b) - \phi_1(a)], \qquad \text{using } (*).
\end{aligned}$$

To complete the proof, we have to show that $\phi_{n+1}(b) - \phi_1(a) = 2\pi$. We do this by showing that the three observations in the previous proof now hold for $\phi_{n+1}(b) - \phi_1(a)$.

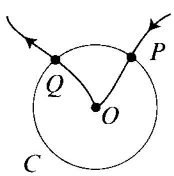
(1) and (2) are obvious.

(3) We are now dealing with a piecewise smooth simple closed curve c in $\mathbb{R}^2$. We want to show that

$$2\pi = \phi_{n+1}(b) - \phi_1(a) = \sum_{i=1}^{n+1} \phi_i(t_i) - \phi_i(t_{i-1}) + \sum_{i=1}^{n} \delta_i.$$

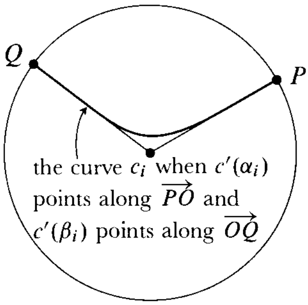
To prove this generalization of the Hopf Umlaufsatz, we proceed as follows.

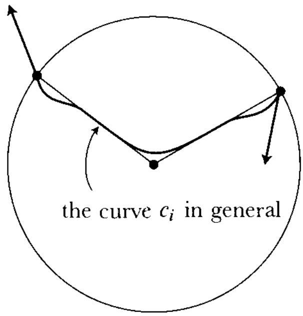
Draw a small circle C around $O = c(t_i)$. Let $P = c(\alpha_i)$ be the last point of $c|[t_{i-1}, t_i]$ on C and let $Q = c(\beta_i)$ be the first point of $c|[t_i, t_{i+1}]$ on C. The



oriented angle from $\overrightarrow{OQ}$ to $\overrightarrow{OP}$ approaches ι_i as the radius of C approaches 0. Also, the direction of $c'(\alpha_i)$ is nearly that of $\overrightarrow{PO}$, while the direction of $c'(\beta_i)$

is nearly that of $\overrightarrow{OQ}$. The picture below shows a curve c_i from P to Q which begins in the direction of $c'(\alpha_i)$, ends in the direction of $c'(\beta_i)$, and stays inside C except at P and Q. The total change in angle of the tangent vector $c_i{}'$





is easily seen to be very close to $\delta_i = \pi - \iota_i$. For each i, let us replace the portion of c between α_i and β_i by the curve c_i. If the circles C are small enough (so that curve does not enter the circle around $c(t_i)$ on any interval other than $[t_{i-1}, t_{i+1}]$), then the new curve, $\bar{c}$, is simple. The total change in angle of the tangent vector $\bar{c}'$ is therefore 2π, by the Hopf Umlaufsatz. On the other hand, the total change along the portions c_i adds up to something very close to $\sum_i \delta_i$, while the total change along the other portions adds up to something very close to $\sum_i \phi_i(t_i) - \phi_i(t_{i-1})$. Therefore the number

$$\sum_{i=1}^{n+1} \phi_i(t_i) - \phi_i(t_{i-1}) + \sum_{i=1}^{n} \delta_i = \phi_{n+1}(b) - \phi_1(a)$$

must be close to 2π. Therefore it must be exactly 2π. ❖

7. COROLLARY. If the sides of the polygon N in Theorem 6 are geodesics, then

$$\int_N K\,dA = -\sum_{i=1}^{n} \delta_i + 2\pi = \sum_{i=1}^{n} \iota_i + (2-n)\pi.$$

In particular, for a geodesic triangle we have

$$\int_N K\,dA = \iota_1 + \iota_2 + \iota_3 - \pi.$$

We are now in a position to find the integral of $K\,dA$ over all of M. The first method of doing this will use a triangulation $\{\sigma_i\}$ of M by 2-simplexes σ_i. Triangulations are defined in the "optional" Chapter 11 of Volume I (see pg. I.426), and it is not easy to prove that they always exist on C^∞ manifolds; however, the proof for compact 2-manifolds is fairly easy (Problem 4-17). For a given triangulation of M we will let

$$V = \text{number of 0-simplexes ("vertices")}$$
$$E = \text{number of 1-simplexes ("edges")}$$
$$F = \text{number of 2-simplexes ("faces").}$$

The number

$$V - E + F = \chi(M)$$

is called the **Euler characteristic** of M. According to Theorem I.11-5, we have

$$\chi(M) = \dim H^0(M) - \dim H^1(M) + \dim H^2(M),$$

so actually $\chi(M)$ does not depend on the triangulation. However, it is not necessary to know this fact in order to follow the next proof.

8. THE GAUSS-BONNET THEOREM. Let M be a compact oriented 2-dimensional Riemannian manifold, with Gaussian curvature K, and volume element dA. Then

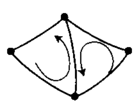
$$\int_M K\,dA = 2\pi \cdot \chi(M).$$

PROOF. Consider a triangulation $\sigma_1, \dots, \sigma_F$ of M. Let A_j, B_j, C_j be the three interior angles of σ_j. Then Theorem 6 gives

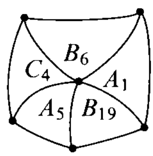
$$\begin{aligned}\int_M K\,dA &= \sum_{j=1}^{F} \int_{\sigma_j} K\,dA \\ &= \sum_{j=1}^{F} \left(\int_{\partial\sigma_j} \kappa_g\,ds \right) + \sum_{j=1}^{F} (A_j + B_j + C_j) - \sum_{j=1}^{F} 3\pi + \sum_{j=1}^{F} 2\pi.\end{aligned}$$

Now we note the following.

(1) The sum $\sum \int \kappa_g\,ds$ is 0, because each edge of the triangulation appears twice, with opposite orientations.



(2) The sum $\sum(A_j + B_j + C_j)$ is $2\pi V$, since the sum of all interior angles occurring at each vertex is exactly 2π.



(3) The sum $-\sum 3\pi$ is just $-3F\pi$. On the other hand, we clearly have $3F = 2E$, since $3F$ is the total number of edges of all faces, each edge being counted twice since it is in two faces. So $-\sum 3\pi = 2\pi(-E)$.

(4) The sum $\sum 2\pi$ is just $2\pi F$.

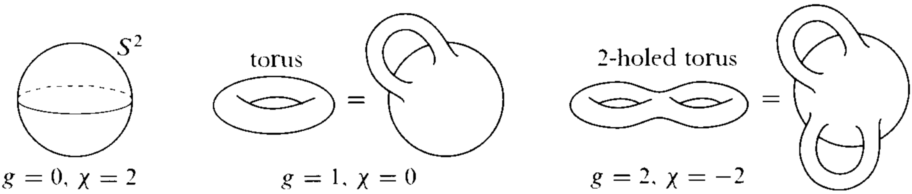
Therefore, our final sum is $2\pi(V - E + F) = 2\pi \cdot \chi(M)$. ❖

A whole slew of consequences follows immediately from this spectacular theorem, which expresses a differential-geometric quantity $\int_M K\,dA$ in terms of a number which has nothing at all to do with curvature, or even with a Riemannian metric. Note, first of all, that we can restate our result in a way which does not involve $\chi(M) = \dim H^0(M) - \dim H^1(M) + \dim H^2(M)$. We have shown that for *any* triangulation of M we have

$$\int_M K\,dA = 2\pi(V - E + F).$$

Picking a fixed triangulation, this shows that $\int_M K\,dA$ is independent of the metric. On the other hand, picking a fixed metric, we obtain an "elementary" proof, without using cohomology, that $V - E + F$ is independent of the triangulation.

The simplest consequences of the Gauss-Bonnet Theorem involve the sign of K on various compact oriented surfaces. All such surfaces are homeomorphic to the surface obtained by adding $g \geq 0$ handles to S^2, and the Euler characteristic is then given by $\chi = 2 - 2g$. The latter result may be proved as



in Problem I.11-2, or by considering triangulations (we just have to check that $V - E + F$ changes by -2 when we add one more handle). Therefore,

$$\int_M K\,dA > 0 \quad \text{if} \quad M \text{ is homeomorphic to } S^2$$

$$\int_M K\,dA = 0 \quad \text{if} \quad M \text{ is homeomorphic to a torus}$$

$$\int_M K\,dA < 0 \quad \text{if} \quad M \text{ is any other compact oriented surface.}$$

It follows, in particular, that if there is a metric on a compact oriented 2-manifold M with $K > 0$ everywhere, then M must be homeomorphic to S^2. This result is rather different from Theorem 2-11, because we do not assume that the metric comes from an imbedding in $\mathbb{R}^3$ (whether such a metric does, in fact, always come from an imbedding is a question which we will mention only later, in Chapter 11). On the other hand, if there is a metric on M with $K = 0$ everywhere, then M must be homeomorphic to the torus. As we have already seen (pg. II.179), such a flat metric does indeed exist on the torus. It is a good exercise to compute that $\int_M K\,dA = 0$ when M is the torus on page 159. Finally, if there is a metric on M with $K < 0$ everywhere, then M must be a sphere with $g \geq 2$ handles. Of course, we can never find an imbedding of such a surface M into $\mathbb{R}^3$ with $K < 0$ everywhere (by Proposition 2-8). But there is, nevertheless, an abstract Riemannian metric on M which has $K < 0$ everywhere; in fact, there is a Riemannian metric on M with $K = -1$ everywhere. The construction of such metrics is carried out in Addendum 1.

Our next consequence of the Gauss-Bonnet Theorem involves the notion of the index of a vector field, as defined in Chapter I.11.

9. THEOREM. Let M be a compact oriented 2-dimensional Riemannian manifold with Gaussian curvature K, and volume element dA. Let X be a vector field on M with only finitely many zeros. Then

$$\int_M K\,dA = 2\pi \cdot (\text{sum of indices of } X).$$

FIRST PROOF. We just combine the Gauss-Bonnet Theorem,

$$\int_M K\,dA = 2\pi \cdot \chi(M),$$

with the Poincaré-Hopf Theorem (I.11-30),

$$\chi(M) = \text{sum of indices of } X.$$

SECOND PROOF. This proof does not involve the intermediary $\chi(M)$, which appears nowhere in the statement of the theorem, nor will a triangulation be invoked. Let $p_1, \dots, p_r$ be the zeros of X, and choose disjoint closed discs D_i containing p_i. Each D_i is diffeomorphic to $D = \{x \in \mathbb{R}^2 : |x| \leq 1\}$, and we will let $D_i(\varepsilon)$ denote the set corresponding to $\{x \in \mathbb{R}^2 : |x| \leq \varepsilon\}$. Let

$$N(\varepsilon) = M - \left(\bigcup_i \text{interior } D_i(\varepsilon)\right).$$

On $N(\varepsilon)$ there is a positively oriented orthonormal moving frame X_1, X_2 with $X_1 = X/\|X\|$. Then

$$\begin{aligned} \int_{N(\varepsilon)} K\, dA &= -\int_{N(\varepsilon)} d\omega_1^2 \\ &= -\sum_i \int_{\partial D_i(\varepsilon)} \omega_1^2 \qquad \text{by Stokes' Theorem.} \end{aligned}$$

For the moment consider one particular i. Let X'_1, X'_2 be a fixed positively oriented orthonormal moving frame on D_i. On D_i minus a line segment we have a differentiable choice θ of the angle between X_1 and X'_1, and by Proposition 1 we have

$$\begin{aligned} -\int_{\partial D_i(\varepsilon)} \omega_1^2 &= \int_{\partial D_i(\varepsilon)} d\theta - \int_{\partial D_i(\varepsilon)} \omega'^2_1 \\ &= (\text{index of } X \text{ at } p_i) - \int_{\partial D_i(\varepsilon)} \omega'^2_1 \\ &\qquad [\text{where } \omega'^2_1 \text{ really depends on } i]. \end{aligned}$$

Therefore,

$$\lim_{\varepsilon \to 0} -\int_{\partial D_i(\varepsilon)} \omega_1^2 = (\text{index of } X \text{ at } p_i) - 0.$$

Consequently,

$$\int_M K\, dA = \lim_{\varepsilon \to 0} \int_{N(\varepsilon)} K\, dA = \sum_i \text{index of } X \text{ at } p_i. \text{ ❖}$$

Notice that the second proof of Theorem 9 reproves the fact that the sum of the indices of a vector field X is independent of X (it also reproves the fact that $\int_M K\,dA$ is independent of the metric). In conjunction with the proof of the Gauss-Bonnet Theorem, it reproves the fact that the sum of the indices of a vector field is $V - E + F$.

In contrast to the previous consequence of the Gauss-Bonnet Theorem, in which $\chi(M)$ does not appear, in the next consequence K does not appear.

10. THEOREM. Let $M \subset \mathbb{R}^3$ be a compact oriented 2-dimensional manifold, and let $\nu\colon M \to S^2$ be the normal map. Then

$$\text{degree of } \nu = \frac{\chi(M)}{2}.$$

PROOF. By the definition of $\deg \nu$ (pg. I.275), we have

$$\int_M \nu^*\omega = \deg \nu \cdot \int_{S^2} \omega$$

for all 2-forms ω on S^2. Choosing ω to be the volume element da of S^2, this means that

$$\begin{aligned}(4\pi)\deg \nu &= \int_M \nu^*(da)\\ &= \int_M K\,dA,\end{aligned}$$

where K is the curvature for the induced metric. Together with the Gauss-Bonnet Theorem, this yields the desired result. ❖

Just as in the case of Theorem 9, one would expect to find a proof of Theorem 10 which does not use the intermediary K. In fact, the same statement can be proved for hypersurfaces of $\mathbb{R}^n$, and this result played an important role in generalizing the Gauss-Bonnet Theorem to higher dimensions (see Chapter 13). Because the proof is differential-topological in nature, rather than differential-geometric, it has been shunted off to Addendum 2.

Our next result is merely a curiosity, an alternative proof of Theorem 2-11 which avoids covering spaces.

11. PROPOSITION. Let M be a compact connected 2-manifold, and let $f: M \to \mathbb{R}^3$ be an immersion with $K(p) > 0$ for all $p \in M$. Then M is orientable, the normal map $N: M \to S^2 \subset \mathbb{R}^3$ is a diffeomorphism, the map $f: M \to \mathbb{R}^3$ is an imbedding, and $f(M)$ is convex.

PROOF. Orientability of M is trivial, as in the proof of Theorem 2-11. The Gauss-Bonnet Theorem then shows that

$$2\pi \cdot \chi(M) = \int_M K\,dA > 0.$$

The only possibility is that M is homeomorphic to S^2, with $\chi(M) = 2$; so

$$\int_M K\,dA = 4\pi.$$

Since $N(M) \subset S^2$ is closed (by compactness of M) and also open (as $K(p) \neq 0$ means that N is regular at p), the map N is onto S^2. To prove that N is one-one, suppose instead that $N(p) = N(q)$ for some $p \neq q \in M$. Then there is an open set $U \ni q$ such that $N(M - \bar{U}) = S^2$. If da is the volume element on S^2, then for any open set $V \subset M - \bar{U}$ on which N is one-one we have

$$\int_V K\,dA = \int_V N^*(da) = \int_{N(V)} da, \qquad \text{since } N \text{ is orientation preserving.}$$

It follows that

$$\int_{M-\bar{U}} K\,dA \geq \int_{S^2} da = 4\pi.$$

Therefore

$$\int_M K\,dA = \int_{M-\bar{U}} K\,dA + \int_U K\,dA > 4\pi,$$

a contradiction. So N is one-one. The remainder of the proof is the same as for Theorem 2-11. ❖

Finally, here's a result that isn't about surfaces at all! (For a history of this result, see McCleary [1].)

12. THEOREM (JACOBI; 1842). Let c be a closed curve in $\mathbb{R}^3$ with nowhere vanishing curvature κ, and let $\mathbf{n}$ be its normal map, into S^2. If $\mathbf{n}$ is a simple closed curve on S^2, then it divides S^2 into two regions of equal area.

PROOF. Let B be one of the regions into which $\mathbf{n}$ divides S^2, and orient $c\colon [a,b] \to S^2$ so that the corresponding orientation for $\mathbf{n}$ coincides with the induced orientation for ∂B. Theorem 5 gives

$$\int_B dA = -\int_{\partial B} \kappa_g \, d\sigma + 2\pi,$$

where κ_g is the geodesic curvature of $\mathbf{n}$ on S^2, and σ is the arclength function of $\mathbf{n}$. Since S^2 has area 4π, it suffices to prove that the integral of κ_g is 0. Now for any arclength parameterized curve γ on a surface M, the definition of its geodesic curvature κ_g is

$$\mathsf{T}\gamma'' = \kappa_g \cdot \nu \times \gamma',$$

where ν is the normal to M. This implies that

$$\kappa_g = \langle \gamma'', \nu \times \gamma' \rangle = \langle \gamma' \times \gamma'', \nu \rangle.$$

When γ is not parameterized by arclength, we have

$$\kappa_g(t) = \langle \gamma'(t) \times \gamma''(t), \nu(\gamma(t)) \rangle / |\gamma'(t)|^3.$$

Applying this to our curve $\mathbf{n}$ on S^2 we obtain

$$\kappa_g(s) = \langle \mathbf{n}'(s) \times \mathbf{n}''(s), \mathbf{n}(s) \rangle \bigg/ \left(\frac{d\sigma}{ds}\right)^3.$$

The Serret-Frenet formulas for c allow us to write this as

$$\kappa_g(s) = [\kappa(s)\tau'(s) - \kappa'(s)\tau(s)] \bigg/ \left(\frac{d\sigma}{ds}\right)^3,$$

where κ and τ are the curvature and torsion of c. Now since

$$\mathbf{n}' = -\kappa \mathbf{t} + \tau \mathbf{b},$$

we have

$$\frac{d\sigma}{ds} = |\mathbf{n}'(s)| = \sqrt{\kappa^2(s) + \tau^2(s)}.$$

So

$$\kappa_g(s) = \frac{\kappa\tau' - \kappa'\tau}{\kappa^2 + \tau^2} \bigg/ \frac{d\sigma}{ds} = \frac{d}{ds}\left(\arctan\frac{\tau}{\kappa}\right) \frac{d\sigma}{ds}$$
$$= \frac{d}{d\sigma}\left(\arctan\frac{\tau}{\kappa}\right).$$

Since c is a closed curve, this gives

$$\int_{\partial B} \kappa_g \, d\sigma = \int_a^b d\left(\arctan\frac{\tau}{\kappa}\right) = 0. \; ❖$$

We will conclude this Chapter with some considerations which, although they do not bear directly on the Gauss-Bonnet Theorem, are nevertheless related to the integral of the curvature. As we have already noted in the proof of Theorem 10, for an immersion $i\colon M \to \mathbb{R}^3$ of a compact oriented 2-manifold M into $\mathbb{R}^3$, the integral of its curvature K is

$$\int_M K\,dA = (4\pi)\deg N,$$

where $N\colon M \to S^2$ is the normal map. This degree of N is simply the "signed" number of points in $N^{-1}(p)$ for any regular value $p \in S^2$ of N (Theorem I.8-12). Now let us consider the actual number $\#(p) \geq 0$ of points in $N^{-1}(p)$. We would like to look at

$$\int_{S^2} \#\cdot da,$$

where da is the volume element of S^2; for technical reasons we will instead consider

$$\int_{S^2-C} \#\cdot da,$$

where C is the set of critical values of N (which has measure 0 by Sard's Theorem). Notice that for some $p \in S^2$ the set $N^{-1}(p)$ might be infinite; but such p are clearly contained in C and therefore do not bother us. Moreover, on S^2-C the function # is locally constant, so it is certainly continuous, and therefore the integral makes sense. Any point of $S^2 - C$ has a neighborhood U on which # has a constant value m and such that $N^{-1}(U) \subset M$ is the disjoint union of m open sets $V_1, \dots, V_m$ on each of which $N\colon V_\alpha \to U$ is a diffeomorphism. Since $N^*(da) = K\,dA$, we have

$$\int_{V_\alpha} K\,dA = \begin{cases} \displaystyle\int_U da & N \text{ orientation preserving, i.e., } K > 0 \\ \displaystyle -\int_U da & N \text{ orientation reversing, i.e., } K < 0. \end{cases}$$

So

$$\int_{V_\alpha} |K|\,dA = \int_U da.$$

From this we readily see that

$$\int_M |K|\,dA = \int_{S^2-C} \#\,da.$$

This number is called the **total absolute curvature** of the immersion $i\colon M \to \mathbb{R}^3$.

13. THEOREM. For any immersion $i\colon M \to \mathbb{R}^3$ of a compact oriented 2-manifold in $\mathbb{R}^3$, the total absolute curvature is $\geq 4\pi$. In fact,

$$\int_{\{p\in M:K(p)>0\}} |K|\,dA = \int_{\{p\in M:K(p)>0\}} K\,dA \geq 4\pi.$$

If the total absolute curvature of the immersion $i\colon M \to \mathbb{R}^3$ equals 4π, then i is an imbedding and $i(M)$ is convex.

PROOF. Let $M' = \{p \in M : K(p) \geq 0\}$. Since

$$\int_{\{p\in M:K(p)>0\}} K\,dA = \int_{M'} K\,dA = \int_{M'} N^*(da),$$

the first part of the theorem will certainly be proved if we show that $N(M') = S^2$. For any $v \in S^2$, consider a plane $P \subset \mathbb{R}^3$ perpendicular to v and far away from $i(M)$ in the direction of v. Move this plane towards 0 until it first touches $i(M)$ at a point p. Then $N(p) = v$. Moreover, $K(p) \geq 0$, since M does not lie on both sides of its tangent plane P_0 at p. This proves the first part of the Theorem.

Now suppose that the total absolute curvature of the immersion $i\colon M \to \mathbb{R}^3$ equals 4π. We will first show that $i(M)$ lies on one side of each of its tangent planes. Suppose that M_p cuts $i(M)$, i.e., that $i(M)$ lies on both sides of its tangent plane M_p. The points furthest from M_p on each side will have normals which are negatives of each other and both perpendicular to M_p. Since $N(p)$ is also perpendicular to M_p, the point $v = N(p)$ has $\#(v) \geq 2$. It is clear that $i(M)$ also lies on both sides of the tangent planes M_q for q in a neighborhood U of p. Now if $K(p) \neq 0$, then we can also assume that n is a diffeomorphism on U, by making U smaller if necessary. Then $\# \geq 2$ on the whole open set $N(U)$. Since we have already shown that $\# \geq 1$ on S^2, this shows that

$$\int_M |K|\,dA = \int_{S^2} \#\,da > 4\pi,$$

a contradiction. So to complete the proof that $i(M)$ lies on one side of M_p, we just have to show that if $K(p) = 0$, then there would be some other point $\bar{p}$ such that $i(M)$ lies on both sides of $M_{\bar{p}}$, and also $K(\bar{p}) \neq 0$.

We will first find a point p' whose tangent plane cuts $i(M)$ and such that p' is not a planar point (i.e., either $K(p') \neq 0$ or p' is a parabolic point). There is certainly no problem doing this if there are non-planar points arbitrarily close to p. So suppose that all points in a neighborhood of p are planar points; the image of this neighborhood under the immersion i then lies in M_p. Consider the set of all points in M whose image lies in M_p, and the component of this set which contains p. Let q be a boundary point of this component. Then $M_q = M_p$, so M_q also cuts $i(M)$. Moreover, q cannot have a whole neighborhood of planar points (for then it would not be a boundary point of the component). So there are points p' arbitrarily close to q such that p' is not a planar point. By choosing p' close enough, we can insure that $M_{p'}$ cuts $i(M)$.

If the point p' which we have produced satisfies $K(p') \neq 0$, we are done. Suppose instead that p' is a parabolic point. Let $L_{p'}$ be the straight line given by Corollary 5-6. Consider the set of points on $L_{p'}$ where $K = 0$; it is a certain closed interval I (which might be just $\{p'\}$). Along I the tangent plane of $i(M)$ is constant, by Proposition 4-5. So if q' is an endpoint of I, then $M_{q'} = M_{p'}$ cuts $i(M)$. Now q' cannot have a whole neighborhood of parabolic points, so there are points $\bar{p}$ arbitrarily close to q' with $K(\bar{p}) \neq 0$. By choosing $\bar{p}$ close enough, we can insure that $M_{\bar{p}}$ cuts $i(M)$.

Thus we have shown that $i(M)$ lies on one side of each of its tangent planes. Let $C \subset \mathbb{R}^3$ be the intersection of the closed half-spaces which are bounded by the planes M_p and contain the points of $i(M)$. Then C is a compact convex set with non-empty interior, and $i(M) \subset$ boundary C. Since $i\colon M \to \mathbb{R}^3$ is an immersion, the set $i(M)$ is both open and closed in boundary C, hence $i(M) =$ boundary C. By Problem 2-4, the map $i\colon M \to$ boundary $C \approx S^2$ is a covering map. So i must be a homeomorphism. ❖

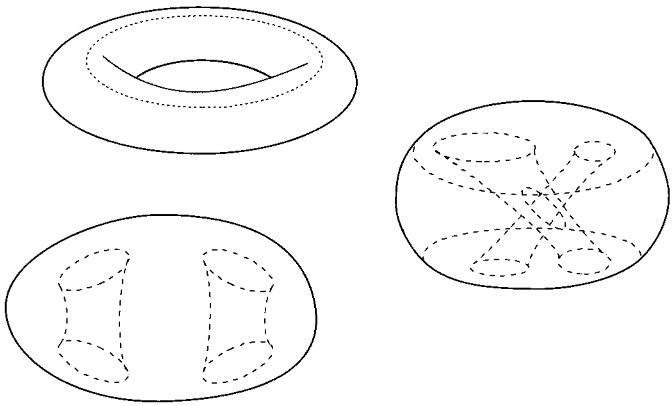
The first part of the preceding proof actually shows more than the asserted inequality, for we clearly have strict inequality if any open subset of S^2 is covered twice. Thus,

$$(*) \qquad \int_{\{p\in M:K(p)>0\}} K\,dA = 4\pi \iff N \text{ is one-one on } \{p \in M : K(p) > 0\}.$$

[Condition $(*)$ can also be expressed in terms of the total absolute curvature, for

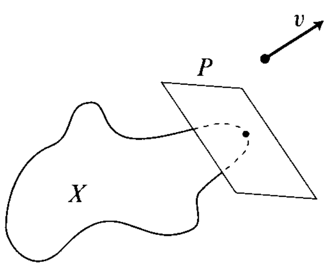
$$\begin{aligned}\int_M |K|\,dA &= \int_{\{p\in M:K(p)>0\}} K\,dA - \int_{\{p\in M:K(p)<0\}} K\,dA \\ &= 2\int_{\{p\in M:K(p)>0\}} K\,dA - \int_M K\,dA \\ &\geq 8\pi - 2\pi\cdot\chi(M) \qquad \text{by the Gauss-Bonnet Theorem,}\end{aligned}$$

with equality if and only if $(*)$ holds.] The pictures below illustrate some immersed surfaces with property $(*)$. In each case, the region of positive curvature



is a subset of a convex surface, bounded by convex plane curves. We will show, by elementary but involved arguments, that this is always so.

Given any set $X \subset \mathbb{R}^3$, let $H(X)$ be its convex hull, the smallest convex set containing X. Elementary considerations (Problem 1) show that if X is compact, then so is $H(X)$. For any unit vector $v \in S^2 \subset \mathbb{R}^3$, consider the subset of X where the function $\iota_v(x) = \langle x, v\rangle$ has its maximum value on X. This will be a subset of the plane P perpendicular to v which is furthest from the origin and still hits X. It is called the "topset" of X in the direction v, and is clearly a subset of the boundary $\partial H(X)$ of $H(X)$.



14. LEMMA. Let $i: M \to \mathbb{R}^3$ be an immersion of a compact oriented surface satisfying $(*)$. If $i(p) \in \partial H(i(M))$, then $K(p) \geq 0$; and conversely if we have the strict inequality $K(p) > 0$, then $i(p) \in \partial H(i(M))$.

PROOF. The first assertion is clear, since $i(M)$ has no support plane at $i(p)$ if $K(p) < 0$, while $H(i(M))$ has a support plane at every point of its boundary.

For each $v \in S^2$, consider the topset of $i(M)$ in the direction v. It is a closed convex subset of a plane. Suppose it contains at least 2 points $i(p_1), i(p_2)$ for $p_1, p_2 \in M$. We clearly have $K(p_1), K(p_2) \geq 0$. Condition $(*)$ then shows that we must have either $K(p_1) = 0$ or $K(p_2) = 0$. Thus $v \in N(\{p \in M : K(p) =\})$ $= C$, say. In other words, if $v \in S^2 - C$, then the topset in the direction v contains only one point. This point of $\partial H(i(M))$ must be $i(p)$ for some $p \in M$, and clearly $N(p) = v$ and $K(p) \geq 0$. Thus

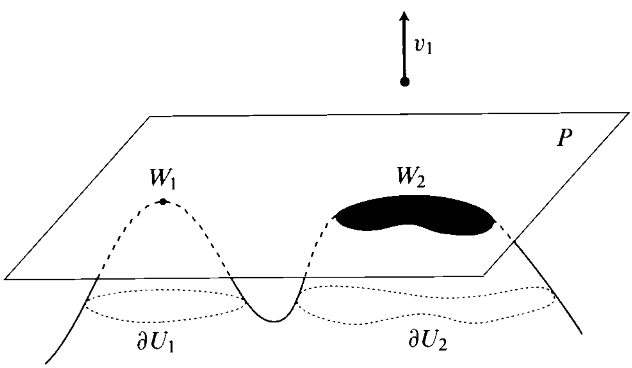
$$N(\{p \in M : i(p) \in \partial H(i(M)) \text{ and } K(p) \geq 0\}) \supset S^2 - C,$$

which has area 4π. So we cannot have $K(p) > 0$ for any other points $p \in M$. ❖

The main part of the argument goes into the following

15. LEMMA. Let $i \colon M \to \mathbb{R}^3$ be an immersion of a compact oriented surface satisfying $(*)$. Then for any $v_1 \in S^2$, the topset of $i(M)$ in the direction v_1 is connected. Moreover, if $v_2 \in S^2$ is perpendicular to v_1, then the topset in the direction v_2 of {the topset of $i(M)$ in the direction v_1} is connected.

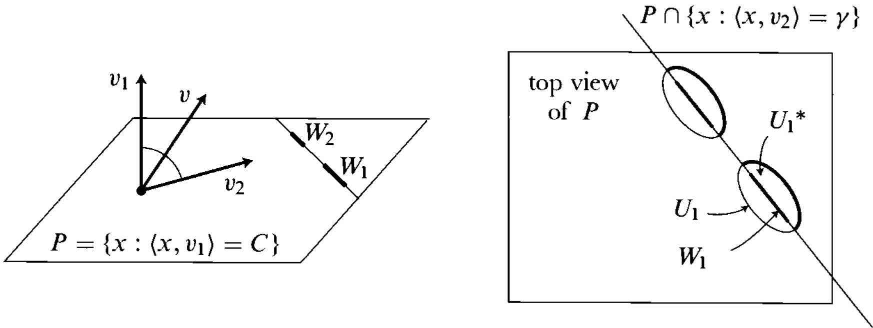
PROOF. If the topset in the direction v_1 is disconnected, then it is the disjoint union of two closed sets W_1, W_2, which are also closed as subsets of $i(M)$, since the topset is a closed set. Let U_1, U_2 be disjoint open neighborhoods of W_1 and W_2 in $i(M)$. On the compact set ∂U_i we have $\langle x, v_1 \rangle$ strictly smaller than



the value of $\langle x, v_1 \rangle$ for $x \in W_i$. So the same is true for all $v \in S^2$ sufficiently close to v_1. This means that for all such v, the topset of U_i in the direction v contains a point $i(p) \in U_i$ not on the boundary of U_i. This implies that $v = N(p)$, and also that $K(p) \geq 0$. So for all v in a whole neighborhood of v_1, with the exception of a set of measure 0, we have $v = N(p)$ where $i(p) \in U_i$ and $K(p) > 0$. Since the U_i are disjoint, this contradicts condition $(*)$.

Now consider the topset in the direction v_2 of the topset of $i(M)$ in the direction v_1; say that ι_{v_2} has the value γ on this topset. Suppose this topset is disconnected and write it as the disjoint union of closed sets W_1 and W_2. Choose U_1, U_2 to be disjoint open neighborhoods of W_1, W_2 in $i(M)$, and let

$$U_i^* = \{x \in U_i : \langle x, v_2 \rangle \geq \gamma\}.$$



One part of ∂U_i^* is the set

$$A_i = \partial U_i \cap \{x : \langle x, v_2 \rangle \geq \gamma\}$$

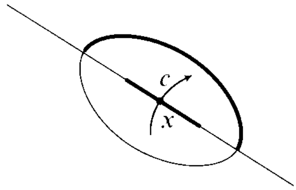
(indicated by a heavy line in the figure). On this compact set we have $\langle x, v_1 \rangle$ strictly smaller than the value $\langle x, v_1 \rangle$ for $x \in W_i$. So the same is true for all unit v close to v_1. The other part of ∂U_i^* is

$$B_i = U_i \cap \{x : \langle x, v_2 \rangle = \gamma\}.$$

On this set, the function $\langle x, v_1 \rangle$ takes on its maximum at some $x \in W_i$. Suppose we choose our unit vector v in the plane spanned by v_1 and v_2. Then for $x \in B_i$ we have

$$\langle x, v \rangle = (\text{constant}) \cdot \langle x, v_1 \rangle + \text{constant},$$

so the maximum still occurs on W_i. Thus, for a unit vector v_0 close enough to v_1, and in the plane of v_1 and v_2, the maximum of $\langle x, v_0 \rangle$ on ∂U_i^* occurs at a point $x \in W_i$. Now we can choose a curve c in $i(M)$ with $c(0) = x$ and $c'(0) = v_2$, since the tangent plane at x is clearly the plane P which contains



the topset in the direction v_1. If

$$v_0 = av_1 + bv_2, \qquad a, b > 0,$$

then

$$\left.\frac{d}{dt}\right|_{t=0} \langle c(t), v_0\rangle = \langle c'(0), v_0\rangle$$
$$= \langle v_2, v_0\rangle = b > 0.$$

So $\langle c(t), v_0\rangle$ has values greater than $\langle c(0), v_0\rangle = \langle x, v_0\rangle$ for small $t > 0$. This shows that the maximum of $\langle x, v_0\rangle$ on $U_i{}^*$ occurs at a point not on the boundary. Hence the same is true of $\langle x, v\rangle$ for all v sufficiently close to v_0. As before, this leads to a contradiction. ❖

16. THEOREM. Let $i\colon M \to \mathbb{R}^3$ be an immersion of a compact oriented surface such that

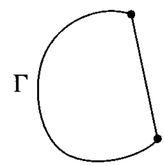
$$\int_{\{p\in M : K(p)>0\}} K\, dA = 4\pi.$$

Then there exist disjoint open sets U and V in M such that M is the union of U, V, and their common boundary, and such that

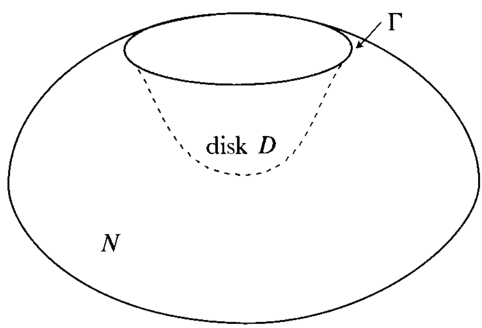
(1) $K \geq 0$ on U and $K \leq 0$ on V.

(2) $i\colon U \to i(U)$ is a diffeomorphism, and $i(U)$ is an open subset of the set $\partial H(i(M))$, bounded by a finite number of convex plane curves, each plane being the common tangent plane of $i(M)$ along the curve.

PROOF. Consider a topset of $i(M)$: it is of the form $P \cap M$ for some support plane P of $H(i(M))$, and $C = P \cap H(i(M))$ is a convex set. If this convex set contains only one point, then this point is in $i(M)$. If the convex set is a line segment, then the endpoints must be in $i(M)$, so the whole segment must be in $i(M)$, by the first part of Lemma 15. Otherwise, the convex set is bounded by a curve Γ in P. We claim that all points of Γ are in $i(M)$. This is clear for all points of Γ which are extreme points of C (points which are not between other points of C). So the only possible exceptions are points on straight line segments of Γ. But the endpoints of such segments must be in $i(M)$, and then the whole segment must, by the second part of Lemma 15.



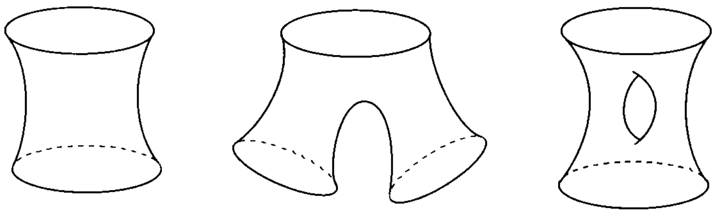
It is possible that Γ bounds a disc on $i(M)$, for the whole interior of Γ in P may be part of $i(M)$. But it is not possible that Γ is the boundary of a disc with $K \leq 0$ everywhere and $K < 0$ somewhere. For suppose it were. The disc D would be tangent to P along Γ. Let N be a surface tangent to P along Γ



such that the union of N and the disc which Γ bounds in P is convex. Then $N' = N \cup D$ is homeomorphic to S^2, but

$$\begin{aligned}\int_{N'} K\,dA &= \int_N K\,dA + \int_D K\,dA \\ &= 4\pi + \int_D K\,dA \\ &< 4\pi,\end{aligned}$$

a contradiction. So if the interior of Γ is not part of $i(M)$, then Γ must be joined to a similar curve Γ', or to several such curves. Since M is compact,

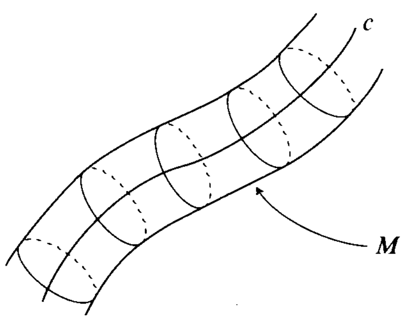


there can be only finitely many such curves $\Gamma_1, \dots, \Gamma_k$. Then $H(i(M))$ minus the discs bounded by $\Gamma_1, \dots, \Gamma_k$ can be taken to be $i(U)$. ❖

The study of total curvature (for submanifolds of Euclidean spaces in general) has recently grown into a little field of its own. However, almost all the results require methods from topology or Morse Theory, which are beyond our reach. We will end instead with a few somewhat older results concerning immersions $i \colon S^1 \to \mathbb{R}^3$, or equivalently, closed curves $c \colon [a,b] \to \mathbb{R}^3$. If c is parameterized by arclength, we define the **total curvature** of c to be

$$\int_a^b \kappa(s)\, ds = \int_a^b |\mathbf{t}(s)|\, ds = \text{length of the curve } \mathbf{t} \colon [a,b] \to S^2,$$

where $\kappa \geq 0$ is the ordinary curvature, and $\mathbf{t}$ is the unit tangent vector. [This definition should not be confused with that given on pg. II.18, where we were concerned only with plane curves, and consequently allowed κ to be both positive and negative.] This total curvature can be related to the total absolute curvature of a certain surface M in $\mathbb{R}^3$, the "canal surface" formed from the union of circles which bound discs of radius ε perpendicular to c (the Adden-

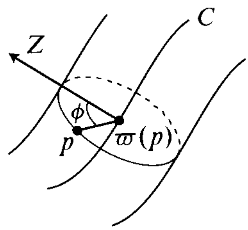


dum to Chapter I.9 can be used to show that M is an immersed surface for sufficiently small ε).

17. PROPOSITION. If M is a canal surface of the closed curve $c\colon [a,b] \to \mathbb{R}^3$, then

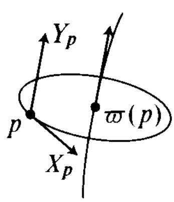
$$\int_a^b \kappa(s)\,ds = \frac{1}{4}\int_M |K|\,dA = \frac{1}{2}\int_{\{p\in M:K(p)>0\}} K\,dA.$$

PROOF. Let $C = c([a,b])$ and $\varpi\colon M \to C$ the projection with $\varpi(p) = q$ when $p \in M$ is on the circle of radius ε perpendicular to C at q. Choose a unit vector field Z along C which is everywhere normal to C. On $M - \varepsilon\cdot Z$ we can



define a function ϕ, with values in $(0, 2\pi)$, giving the angle between p and Z on the circle $\varpi^{-1}(\varpi(p))$. If we regard the arclength s as a function on C (or $C - c(a)$, to be more rigorous), and also let s denote the function $s\circ\varpi$ on M, then (ϕ, s) is a coordinate system on M (minus a set of measure 0).

For each $p \in M$, let X_p be the unit vector tangent to the circle $\varpi^{-1}(\varpi(p))$ at p, and let Y_p be the unit vector at p which is parallel to the tangent vector dc/ds at $\varpi(p)$. The normal $N(p)$ at p points in the direction from $\varpi(p)$ to p



(Problem 2), so X_p and Y_p are tangent to M and p. It is easy to see that

$$\left.\begin{aligned} d\phi(X) &= 1\\ ds(X) &= 0\\ ds(Y) &= 1 \end{aligned}\right\} \implies d\phi\wedge ds(X,Y) = 1.$$

This shows that

$$d\phi \wedge ds = dA, \qquad \text{the volume element on } M.$$

So

$$\int_M |K|\, dA = \int_M |K|\, d\phi \wedge ds = \int_a^b \left(\int_0^{2\pi} |K(\phi, s)|\, d\phi \right) ds.$$

Now

$$dN(X_p) = X_p, \qquad \text{since } N \text{ is the identity map on the circle } \varpi^{-1}(\varpi(p)),$$

so if p has coordinates (ϕ, s), then

(1) $$\begin{aligned} K(\phi, s) &= -\langle dN(Y_p), Y_p \rangle = -\langle \nabla'_{Y_p} N, Y_p \rangle \\ &= \langle N(\phi, s), \nabla'_{Y_p} Y \rangle \qquad \text{since } Y \text{ is a unit vector field} \\ &= \langle N(\phi, s), \kappa(s) \cdot \mathbf{n}(s) \rangle \\ &= \kappa(s) \cdot \{\text{cosine of the angle between } N(\phi, s) \text{ and } \mathbf{n}(s)\} \\ &= \kappa(s) \cdot \cos(\phi - \phi_0), \text{ say.} \end{aligned}$$

(When $\kappa(s) = 0$ and $\mathbf{n}(s)$ is undefined, this formula still holds, with an arbitrary choice of ϕ_0). Thus we have

$$\int_0^{2\pi} |K(\phi, s)|\, d\phi = \kappa(s) \int_0^{2\pi} |\cos(\phi - \phi_0)| = 4\kappa(s),$$

which shows that

$$\int_M |K|\, dA = \int_a^b 4\kappa(s)\, ds.$$

The second equality can easily be deduced from the fact that

$$\int_M K\, dA = 0.$$

which follows from the Gauss-Bonnet Theorem. since M is a torus. We can also deduce the result directly from equation (1). for if $\kappa(s) \neq 0$. then $K(\phi, s) \geq 0$ for $\cos(\phi - \phi_0) \geq 0$. and the integral of $\cos(\phi - \phi_0)$ over the subset of $[0, 2\pi]$ where it is ≥ 0 is exactly 2. ❖

18. COROLLARY (FENCHEL). The total curvature of any closed curve $c\colon [a,b] \to \mathbb{R}^3$ is $\geq 2\pi$. Equality holds if and only if c is a plane convex curve.

PROOF. The inequality follows directly from Theorem 13 and Proposition 17.

Now suppose that equality holds. In the proof of Theorem 13 we actually showed that

$$\int_{\widetilde{M}} K\,dA \geq 4\pi,$$

where $\widetilde{M}$ is the set of all points $p \in M$ such that M lies on one side of M_p. We clearly also have

$$\int_{\widetilde{\widetilde{M}}} K\,dA \geq 4\pi,$$

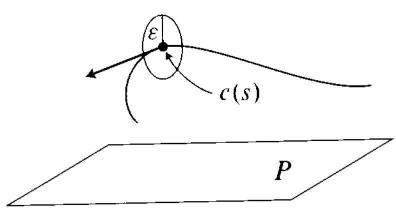
where $\widetilde{\widetilde{M}} = \widetilde{M} \cap \{p \in M : K(p) > 0\}$. Now if $\{p \in M : K(p) > 0\} - \widetilde{\widetilde{M}} \neq \emptyset$, then we would have

$$\int_{\{p\in M:K(p)>0\}} K\,dA > 4\pi,$$

since $\widetilde{\widetilde{M}}$ is a closed subset of $\{p \in M : K(p) > 0\}$. This would contradict the assumption that the total curvature of c is 2π, by Proposition 17. So we see that

$$K(p) > 0 \implies M \text{ lies on one side of } M_p.$$

Now consider a point $c(s)$ with $\kappa(s) \neq 0$. On the circle $\varpi^{-1}(c(s))$ there is an open semi-circle with $K > 0$. At the two endpoints of this semi-circle the tangent planes P_1, P_2 of M are parallel. Moreover, M lies on one side of each of these planes, since the points are limits of points where $K > 0$, and hence of points p such that M lies on one side of M_p. So M lies entirely within the space between the parallel planes P_1, P_2. These planes are at distance 2ε if M is the canal surface formed by discs of radius ε. We claim that C must lie entirely on the plane P midway between P_1 and P_2. For consider a point $c(s)$ of C furthest



from P. The tangent vector $c'(s)$ must be parallel to P, which means that the disc of radius ε through $c(s)$ and perpendicular to c is also perpendicular to P. But then some points of this disc will lie outside of the region bounded by P_1 and P_2 unless $c(s)$ is on P. Thus c is a plane curve. The proof that c is convex is left to the reader. ❖

This proof, due to Voss [1], is quite different from the original proof of Fenchel [1]. For the sake of completeness, we offer the following extremely simple proof of Horn [1], where references to many other proofs may be found.

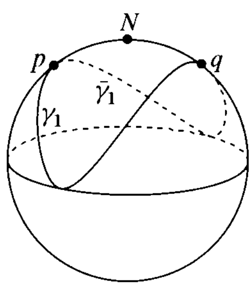
SECOND PROOF. We claim that if the closed curve $\mathbf{t}\colon [a,b] \to S^2$ lies in a closed hemisphere of S^2, then c must be a plane curve (this implies, moreover, that $\mathbf{t}$ cannot lie in an open hemisphere). Indeed, suppose that $\mathbf{t}$ lies in a closed hemisphere; without loss of generality we can assume that it is the northern hemisphere, so that the third component $\mathbf{t}^3$ of $\mathbf{t}$ satisfies $\mathbf{t}^3 \geq 0$. Then

$$0 = c^3(b) - c^3(a) = \int_a^b \mathbf{t}^3(s)\,ds \geq 0,$$

which implies that $\mathbf{t}^3(s) = 0$ for all s, and hence c is a plane curve. We now appeal to a simple

19. LEMMA. If γ is a closed curve on S^2 of length $< 2\pi$, then γ is contained in some open hemisphere of S^2; if γ has length 2π, then γ is contained in some closed hemisphere.

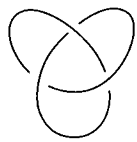
PROOF. Choose points p, q on γ which divide it into two arcs, γ_1 and γ_2, of equal length. Rotate γ so that the north pole $N = (0,0,1)$ lies on the midpoint of the shorter arc of the great circle joining p and q. If arc γ_1 intersects the



equator at some point, let $\bar{\gamma}_1$ be the arc from p to q which is symmetric to γ_1 with respect to N. Then the closed curve made up of γ_1 and $\bar{\gamma}_1$ has the same length as γ, and also *contains two antipodal points.* This shows that the length of γ

is $\geq 2\pi$, and strict inequality holds if γ_1 actually enters the southern hemisphere. Since the same considerations hold for the other arc γ_2, the lemma is proved, and with it the Theorem. ❖

Our last result concerns knotted closed curves $c\colon [a,b] \to \mathbb{R}$. There are several possible ways to define when a closed curve is knotted; for our purposes it will be simplest to say that c is **unknotted** if $c([a,b])$ is the boundary of an imbedded disc, and **knotted** otherwise. The following closed curve is knotted, although proving this fact requires considerable work.



20. THEOREM (FARY, MILNOR). The total curvature of a knotted closed curve $c\colon [a,b] \to \mathbb{R}^3$ is $\geq 4\pi$.

PROOF. Suppose that the total curvature of c is $< 4\pi$. If M is a canal surface of c, then by Proposition 17

$$\int_M |K|\, dA < 16\pi \implies \int_{S^2} \#\, da < 16\pi,$$

where da is the volume element of S^2. This means that some point $v \in S^2$ is the image of at most 3 points of M, which is equivalent to saying that $c'(s)$ is perpendicular to v for at most 3 values of s. For simplicity, say that $v = (0,0,1)$. Since

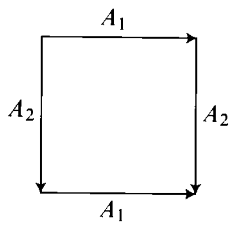
$$\langle v, c'(s)\rangle = \frac{d}{ds}\langle v, c(s)\rangle,$$

the function $s \mapsto \langle v, c(s)\rangle =$ (height of $c(s)$ above (x,y)-plane) has derivative 0 for at most 3 values of s. Since the number of relative maxima or minima of the height function is even, there must be just one of each. Therefore the curve c must consist of two arcs joining the lowest and highest points, each arc having monotonically increasing height. Each plane parallel to the (x,y)-plane between these lowest and highest points intersects the curve in 2 points. Joining each such pair by the line segment between them, we obtain an imbedded disc whose boundary is the curve. ❖

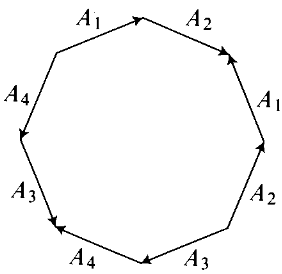
ADDENDUM 1

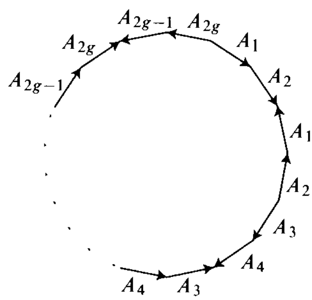
COMPACT SURFACES WITH CONSTANT NEGATIVE CURVATURE

In order to construct a Riemannian metric $\langle\ ,\ \rangle$ with $K = -1$ on any compact surface M of genus $g \geq 2$, we first need to consider the standard topological way of obtaining these surfaces. For simplicity we will work only with the oriented ones. We have often used the fact that the torus, with genus $g = 1$, can be obtained by identifying sides of a square according to the scheme shown below.

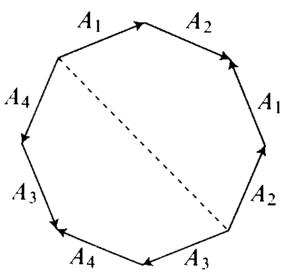


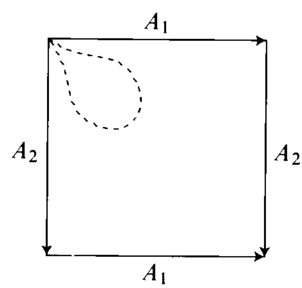
In general, the "g-holed torus", with genus $g \geq 1$, can be obtained by identifying sides of a $4g$-gon according to the following scheme:



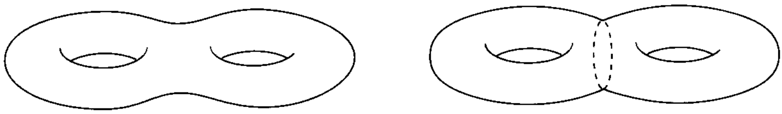


If your powers of visualization are much better than mine, you may be able to literally see that this is the case. Otherwise, the following argument should convince you. The dashed line in the first figure below is a circle, because, as

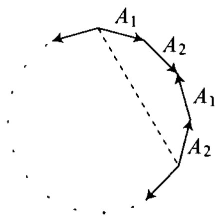




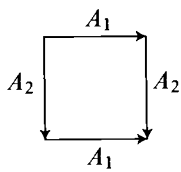
a quick check shows, the required identifications of sides forces all vertices to be identified to one point. This circle divides the identification space into two parts. The right half of the figure shows that each part, together with the bounding circle, is homeomorphic to a torus with an open disc removed. So the identification space is homeomorphic to the space obtained by removing a disc from each of two tori, and then identifying the boundary circles; hence



it is a 2-holed torus. A similar argument can be used to treat the general case, by induction.



Now the flat metric on the torus is just obtained from the flat metric on the square by performing the required identifications. The fact that we actually get



a Riemannian metric on the identification space depends on two circumstances: first, the opposite sides are of equal length, and second, the sum of the angles at the four vertices is exactly 2π. It is the failure of this second condition in larger polygons which prevents us from getting a flat metric on the orientable surfaces of higher genus. For surfaces with $g > 1$ we will construct a metric with $K = -1$ by obtaining $4g$-gons in the *non-Euclidean plane* whose angles add up to exactly 4π.

Our model for the non-Euclidean plane will be the Poincaré upper half-plane $\mathcal{H}^2 = \{(x, y) \in \mathbb{R}^2 : y > 0\}$ with the Riemannian metric

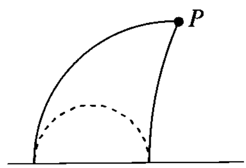
$$\langle\ ,\ \rangle = \frac{dx \otimes dx + dy \otimes dy}{y^2};$$

this manifold has constant curvature $K = -1$ (compare Problem I.9-41 and pg. II.301). It would be possible to check that if we define the "straight lines" of $\mathcal{H}^2$ to be the geodesics for $\langle\ ,\ \rangle$, then the Poincaré upper half-plane satisfies the axioms for non-Euclidean geometry, and then conclude, by a theorem of non-Euclidean geometry, that the sum of the angles of a geodesic triangle is always $< \pi$. Happily, we can also reach this conclusion simply by applying the Gauss-Bonnet Formula. Indeed we find (Corollary 7) that for a geodesic triangle Δ with interior angles $\iota_1, \iota_2, \iota_3$ we have

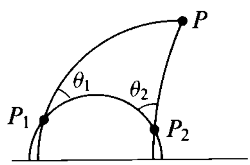
$$\pi - (\iota_1 + \iota_2 + \iota_3) = -\int_{\Delta} -1\, dA = \text{area}(\Delta) > 0.$$

This also shows us that small triangles are very close to Euclidean triangles: the sum $\iota_1 + \iota_2 + \iota_3$ can be made as close to π as we like. It follows that the interior angles of an n-gon can be made as close to $(n-1)\pi$ as we like.

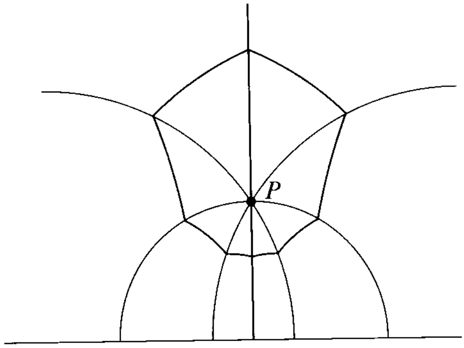
In contrast to the situation for small triangles, let us consider what happens when we take 2 fixed geodesic rays from a point P and join points far out on each of the two sides with a third geodesic. The figure below shows two geodesic rays, both of which are portions of circles or straight lines which meet the x-axis at right angles (Problem I.9-41). The dashed line is another geodesic



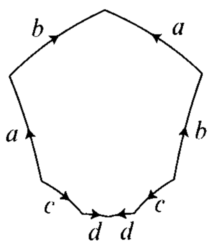
which "meets these at 0 angle". Of course, this third geodesic doesn't really meet either of the first two in $\mathcal{H}^2$. But it is clear from this picture that if we take points P_1 and P_2 far enough out on the two original geodesics, then the angles θ_1 and θ_2 between the unique geodesic through P_1 and P_2 and our given geodesics can be made as small as we like.



Now let us take a point P in $\mathcal{H}^2$, and draw $4g$ geodesic rays from P, the angle between two successive rays being $2\pi/4g$. For each $r > 0$, consider the equilateral $4g$-sided geodesic polygon obtained by joining the points on these rays which are at distance r from P. Let $\Sigma(r)$ be the sum of the interior angles



of this polygon; clearly $\Sigma(r)$ is a continuous function of r. But we have seen that $\Sigma(r) \to 0$ as $r \to \infty$, while as $r \to 0$ we have $\Sigma(r) \to (4g-1)\pi > 2\pi$ for $g \geq 2$. So if $g \geq 2$, then there is an r with $\Sigma(r) = 2\pi$. This gives us an equilateral $4g$-gon with the sum of the interior angles $= 2\pi$. After identifying



pairs of sides according to the scheme on page 292, we then have a metric on the surface of genus g with constant curvature $K = -1$.

We add here some supplementary remarks which may be of interest to readers familiar with complex analysis. The necessary identification of pairs of sides can easily be accomplished by means of one-one complex analytic maps of $\mathcal{H}^2$ onto itself, and it is thus easy to see that M can be made into a Riemann surface (a complex manifold of complex dimension 1), meaning that there is a collection $\mathcal{A}$ of homeomorphisms $f: U \to \mathbb{C} = \mathbb{R}^2$, from open subsets $U \subset M$ onto open subsets of $\mathbb{C}$, such that

(i) the union of the domains of all $f \in \mathcal{A}$ covers M

(ii) if $f: U \to \mathbb{C}$ and $g: V \to \mathbb{C}$ are in $\mathcal{A}$, then

$$g \circ f^{-1}: f(U \cap V) \to g(U \cap V)$$

is complex analytic.

[In fact, it is possible to establish even more. Since each vertex of our polygon has angle $2\pi/4g$, we can arrange exactly $4g$ polygons congruent to it around every vertex. The same construction can be carried out at the new vertices of the new polygons, and the construction can then be repeated indefinitely. In this way we arrive at a "tiling" of the hyperbolic plane by equilateral $4g$-gons. The one-one complex analytic maps of $\mathcal{H}^2$ onto itself which are required for identifying the sides of our original polygon all preserve the tiling, and generate a group $\mathcal{G}$, with M homeomorphic to the quotient space $\mathcal{H}^2/\mathcal{G}$ obtained by identifying z and $g(z)$ for all $z \in \mathcal{H}^2$ and $g \in \mathcal{G}$. The map $\mathcal{H}^2 \to \mathcal{H}^2/\mathcal{G}$ defined by taking z into its equivalence class in $\mathcal{H}^2/\mathcal{G}$ is a covering map. Thus we obtain an explicit construction of a covering map $\pi: \mathcal{H}^2 \to M$, which shows that $\mathcal{G}$ must be isomorphic to $\pi_1(M)$, and that the universal covering space of M is homeomorphic to $\mathbb{R}^2$. (The latter fact could also have been deduced from purely topological considerations, since all simply-connected (paracompact) 2-manifolds are homeomorphic to either $\mathbb{R}^2$ or S^2, and S^2 cannot be the universal covering space of M since S^2 is compact, while $\pi_1(M)$ is infinite.)]

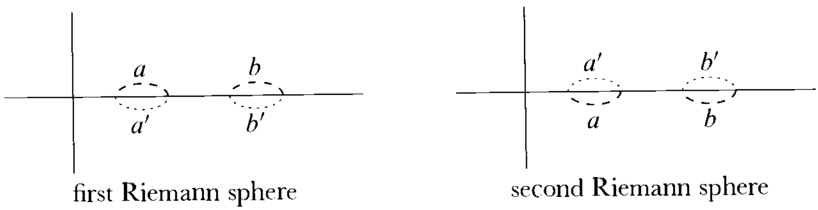
There are also two other methods which we can use to put a complex manifold structure on M.

Method A. Choose an arbitrary Riemannian metric $\langle\!\langle\ ,\ \rangle\!\rangle$ for M. We recall (pg. II.296) that a map f between Riemannian manifolds is **conformal** if each f_* is angle preserving. For any point p of a 2-dimensional Riemannian manifold $(M, \langle\!\langle\ ,\ \rangle\!\rangle)$, there is a neighborhood U of p and a conformal diffeomorphism $f: U \to \mathbb{R}^2$ onto an open subset of $\mathbb{R}^2$ with its usual Riemannian metric; this fact was mentioned in Volume II, and a proof will be found in Addendum 1 to Chapter 9 of these Volumes. It is also an elementary fact (Problem 4-9) that a diffeomorphism $f: W \to \mathbb{C}$, from an open set $W \subset \mathbb{C}$ onto an open subset of $\mathbb{C}$, is complex analytic if and only if f is orientation preserving and conformal with respect to the usual metric on $\mathbb{R}^2$. We can therefore define $\mathcal{A}$ to be the collection of all conformal orientation preserving $f: U \to \mathbb{R}^2$; if $f, g \in \mathcal{A}$, then $g \circ f^{-1}$ is orientation preserving and conformal, so it is complex analytic.

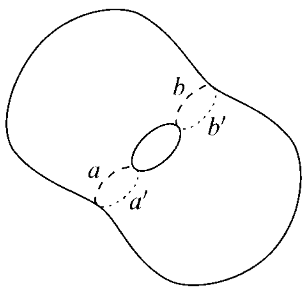
Method B. This method involves the Riemann surface, including branch points, of a "complete analytic function". From its definition, it is clear that this surface is a Riemann surface in the sense of being a complex manifold. On the other hand, the usual method for visualizing the Riemann surface of

$$\sqrt{(z-1)(z-2)\dots(z-2(g+1))}$$

is to take two copies of the Riemann sphere, make "cuts" from 1 to 2, from 3 to 4, ... , and from $2g+1$ to $2g+2$, and identify the corresponding cuts in the



two different spheres. This is homeomorphic to the g-holed torus.



If we use one of these two methods for putting a complex manifold structure on a compact oriented surface M of genus $g \geq 2$, then we can give another construction of a metric on M with constant curvature $K = -1$, provided that we are willing to use yet more machinery. We consider the universal covering space $\pi : \widetilde{M} \to M$, and give it the structure of a Riemann surface in the obvious way. Then M is the quotient of $\widetilde{M}$ by the group of covering transformations, each of which is a one-one complex analytic map of $\widetilde{M}$ onto itself. Now we expect that $\widetilde{M}$ is $\mathcal{H}^2$ (as we have already mentioned at the beginning of this whole discussion). We can establish this fact independently by using the "general uniformization theorem", which tells us that the simply-connected Riemann surface $\tilde{M}$ must be analytically equivalent to either $\mathbb{C}$ or the upper half-plane $\mathcal{H}^2$. Now all one-one complex analytic maps of $\mathbb{C}$ onto itself are of the form $z \mapsto az+b$. This group, and hence any subgroup, is abelian, while the fundamental group of M is non-abelian if it has genus > 2. So if M is a compact surface of genus $g \geq 2$, then $\tilde{M}$ is $\mathcal{H}^2$. But the only one-one complex analytic maps of $\mathcal{H}^2$ onto itself are of the form $z \mapsto (az+b)/(cz+d)$, and (Problem I.9-47, or Problem 7-6) these maps are isometries of $\mathcal{H}^2$ with the metric $(dx \otimes dx + dy \otimes dy)/y^2$. Consequently, M has a metric $\langle\ ,\ \rangle$ such that $\pi^*\langle\ ,\ \rangle = (dx \otimes dx + dy \otimes dy)/y^2$; clearly $(M, \langle\ ,\ \rangle)$ has constant curvature -1.

ADDENDUM 2
THE DEGREE OF THE NORMAL MAP

Let $M \subset \mathbb{R}^m$ be a compact hypersurface with normal map $\nu: M \to S^{m-1}$; we want to show that the degree of ν is $\chi(M)/2$. We will actually prove a more general result, involving any compact orientable manifold $M^n \subset \mathbb{R}^m$. From the Addendum to Chapter I.9 we know that for sufficiently small $\varepsilon > 0$, the set $N = \{q \in \mathbb{R}^m : d(q, M) \leq \varepsilon\}$ is a compact m-dimensional manifold-with-boundary, which is a tubular neighborhood under a projection map $\pi: N \to M$. We give N the usual orientation and ∂N the induced orientation, so that the corresponding normal map $\nu: \partial N \to S^{m-1}$ is outward pointing. We will show that:

The degree of the normal map $\nu: \partial N \to S^{m-1}$ is $\chi(M)$.

[In the special case where $M \subset \mathbb{R}^m$ is a hypersurface, the manifold ∂N consists of two components each homeomorphic to M, and the degree of $\nu: \partial N \to S^{m-1}$ is just twice the degree of the normal map of M, so it will follow that this degree is $\chi(M)/2$.]

We will follow the exposition in Milnor {1}. The first step is a simple

21. LEMMA. Let $N \subset \mathbb{R}^m$ be a compact m-dimensional manifold-with-boundary, and let $\nu: \partial N \to S^{m-1}$ be the normal map. Let X be a vector field on N with isolated zeros, and suppose that X is outward pointing on ∂N. Then the degree of $\nu: \partial N \to S^{m-1}$ is equal to the sum of the indices of X.

PROOF. We regard X as a map $X: N \to \mathbb{R}^m$. Let $p_1, \dots, p_k$ be the zeros of X, and let $U_1, \dots, U_k$ be open ε-balls around $p_1, \dots, p_k$ with all $\overline{U_i} \subset$ interior N. Then $\mathcal{B} = N - (U_1 \cup \dots \cup U_k)$ is a compact manifold-with-boundary, and $\bar{X} = X/|X|: \mathcal{B} \to S^{m-1}$. Now if ω is any $(m-1)$-form on S^{m-1}, then

$$\int_{\partial\mathcal{B}} \bar{X}^*(\omega) = (\text{degree } \bar{X}|\partial\mathcal{B}) \cdot \int_{S^{m-1}} \omega.$$

By Stokes' Theorem we have

$$\int_{\partial\mathcal{B}} \bar{X}^*(\omega) = \int_{\mathcal{B}} d\bar{X}^*(\omega) = \int_{\mathcal{B}} \bar{X}^*(d\omega) = 0,$$

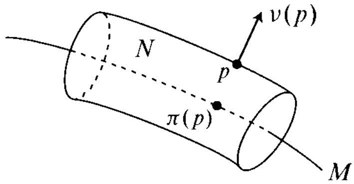
so the degree of $\bar{X}|\partial\mathcal{B}$ must be 0. But $\bar{X}|\partial N$ is smoothly homotopic to ν, while the degrees on the other components of $\partial\mathcal{B}$ are the negatives of the indices of X at the p_i (the minus signs come from the fact that the orientations on these components are the negatives of the usual ones). Thus

$$\deg \nu - (\text{sum of indices of } X \text{ at the } p_i) = 0. \ ❖$$

Now suppose we have a compact oriented manifold $M^n \subset \mathbb{R}^m$ and we choose an arbitrary vector field X on M with only isolated zeros; according to the Poincaré-Hopf Theorem (I.11-30) the sum of its indices is $\chi(M)$. We can extend X to a vector field Y on the tubular neighborhood N in a rather obvious way, by defining

$$(*) \qquad Y(p) = [p - \pi(p)] + X(\pi(p)) \qquad \text{(considered as a vector at } p\text{)}.$$

Since the normal $\nu(p)$ points in the direction from $\pi(p)$ to p (Problem 2), and the vector $p - \pi(p)$ is perpendicular to M_p, it is clear that Y points outward



along ∂N, and that the zeros of Y in all of N are precisely the zeros of X in M. If we could show that the index of Y at such a zero equals the index of X, then the desired result would follow from Lemma 21. However, a direct analysis of the index of Y turns out to be very difficult, so we take a slight detour.

Consider first a vector field X on $\mathbb{R}^n$, with an isolated zero at $0 \in \mathbb{R}^n$. We regard X as a function $X: \mathbb{R}^n \to \mathbb{R}^n$, and we define X to be **non-degenerate** at 0 if the derivative $DX(0): \mathbb{R}^n \to \mathbb{R}^n$ is non-singular.

22. LEMMA. If X has a non-degenerate zero at 0, then the index of X at 0 is $+1$ or -1, depending on whether $\det DX(p) > 0$ or < 0.

PROOF. We can assume $p = 0$. Then X is a diffeomorphism on some convex open neighborhood U of 0. In the proof of Lemma I.11-27 we saw that if X is orientation preserving, then X is smoothly homotopic to the identity via maps which have no zeros. So the index is 1. Similarly, if X reverses orientation, then it is smoothly homotopic to a reflection, and has index -1. ❖

Now consider an oriented n-manifold M, and a vector field X on M with an isolated zero at $p \in M$. Let $f: U \to \mathbb{R}^n$ be an orientation preserving diffeomorphism, where $U \subset M$ is an open set containing p, and $f(p) = 0$. Then f_*X is a vector field on $\mathbb{R}^n$, and we define X to be **non-degenerate** at p if f_*X is non-degenerate at 0. If $M \subset \mathbb{R}^m$ is a submanifold of $\mathbb{R}^m$, and we regard the vector field X on M as a function $X\colon M \to \mathbb{R}^m$, then $DX(p)\colon \mathbb{R}^m \to \mathbb{R}^m$ will take M_p to M_p. In fact, for the function $X \circ f^{-1}\colon \mathbb{R}^n \to M$ we have $(X \circ f^{-1})_*\colon \mathbb{R}^n{}_0 \to M_p$; but $(f^{-1})_*(\mathbb{R}^n{}_0) = M_p$, while X_* is the same as DX on M_p. It is easy to see from Lemma 22 that if p is a non-degenerate 0, then the index of X at 0 is $+1$ or -1, depending on whether $\det DX(p)\colon M_p \to M_p$ is > 0 or < 0.

23. THEOREM. Let $M^n \subset \mathbb{R}^m$ be a compact oriented manifold with a closed tubular neighborhood $\pi\colon N \to M$. Then the degree of the normal map $\nu\colon \partial N \to S^{m-1}$ is $\chi(M)$.

PROOF. We will show that there is a vector field X on M with only non-degenerate zeros. Assuming this fact for the moment, we use the vector field X to define a vector field Y on N by (∗). It is easy to see that for $p \in M$ we have

$$DY(p) = \begin{cases} DX(p) & \text{on } M_p \\ I & \text{on } {M_p}^{\perp}. \end{cases}$$

Therefore $\det DY(p) = \det DX(p)\colon M_p \to M_p$. So the index of Y at a zero p equals the index of X at p. Then Lemma 21 implies that the degree of $\nu\colon \partial N \to S^{m-1}$ is the sum of the indices of X, which equals $\chi(M)$ by the Poincaré-Hopf Theorem.

To obtain the desired vector field on M, we first choose a vector field X with only finitely many (possibly degenerate) zeros (as on pp. I.449–450). It obviously suffices to show that for each zero p we can find a new vector field which equals X outside of a small neighborhood U of p, and has only non-degenerate zeros inside the neighborhood. It obviously suffices to work on $\mathbb{R}^n$. Given a neighborhood U of a zero $p \in \mathbb{R}^n$, let $f\colon \mathbb{R}^n \to [0,1]$ be a C^∞ function which is 1 on an open set W with $p \in W \subset \overline{W} \subset U$, and 0 on $\mathbb{R}^n - U$. By Sard's Theorem, there is a regular value X_0 of X arbitrarily close to 0. Consider the vector field

$$\bar{X} = X - f \cdot X_0.$$

Within W this vector field is 0 only at points q with $X(q) = X_0$, so all zeros in W are non-degenerate; on the other hand, in $U - W$ there are no zeros at all if X_0 is sufficiently small. ❖

As a final remark, we note that if we choose U to be a closed manifold-with-boundary, then the sum of the indices of $\bar{X}$ at zeros within U is just the degree of $\bar{X}/|\bar{X}|\colon \partial U \to S^{m-1}$, as in the proof of Lemma 21. But this map is the map $X/|X|\colon \partial U \to S^{m-1}$, whose degree is, by definition, the index of X at p. So, without using the Poincaré-Hopf Theorem, we have reproved the fact that for any vector field X on M with only isolated zeros, the sum of its indices is a constant, namely $\deg \nu\colon \partial N \to S^{m-1}$. We could then identify this constant with $\chi(M)$ as we did in Chapter I.11, when we originally proved the Poincaré-Hopf Theorem.

PROBLEMS

1. Let $X \subset \mathbb{R}^m$ be any set.

(a) The convex hull $H(X)$ is the union of the convex hulls of all finite subsets of X.
(b) The convex hull of $m + 2$ points in $\mathbb{R}^m$ is the union of the convex hulls of all its subsets of $m + 1$ points.
(c) If X is compact, then so is $H(X)$.

2. Prove the assertions about the normals to a canal surface on page 287, and the normals of the tubular neighborhood, on page 300, by using the argument in Problem 3-12. Also generalize the argument of Problem I.9-28.

MINI-BIBLIOGRAPHY FOR VOLUME III

Brackets [ ] indicate journal articles, braces { } indicate books.

Bol, G.

[1] *Über Nabelpunkte auf einer Eifläche*, Math. Z. **49** (1944), 389–410.

Efimov, N. V.

[1] *Generation of singularities on surfaces of negative curvature* (Russian), Mat. Sb. **64** (1964), 286–320.

Fenchel, W.

[1] *Über Krummung und Windung geschlossener Raumkurven*, Math. Ann. **101** (1929), 238–252.

Hamburger, H.

[1] *Beweis einer Carathéodoryschen Vermutung. Teil I*, Ann. of Math. **41** (1940), 63–86; *II*, Acta Math. **73** (1941), 175–228; *III*, Acta Math. **73** (1941), 229–332.

Hilbert, D.

[1] *Ueber Flächen von constanter Gausscher Krümmung*, Trans. Amer. Math. Soc. **2** (1901), 87–99.

Holmgren, E.

[1] *Sur les surfaces à courbure constante négative*, C. R. Acad. Sci. Paris Ser. A-B **134** (1902), 740–743.

Horn, R. A.

[1] *On Fenchel's theorem*, Amer. Math. Monthly **78** (1971), 380-381.

Klotz, T.

[1] *On G. Bol's Proof of Carathéodory's conjecture*, Comm. Pure Appl. Math. **12** (1959), 277–311.

Klotz Milnor, T.

[2] *Efimov's Theorem about complete immersed surfaces of negative curvature*, Advances in Math. **8** (1972), 474–543.

McCleary, J.

{1} *On Jacobi's remarkable curve theorem*, Historia Math. **21** (1994), 377–385.

Milnor, J. W.

{1} *Topology from the Differentiable Viewpoint*, University Press of Viriginia, Charlottesville, Virginia, 1965.

Voss, K.

[1] *Eine Bemerkung über die Totalkrümmung geschlossener Raumkurven*, Arch. Math. (Basel) **6** (1955), 259–263.

Wunderlich, W.

[1] *Über ein abwickelbares Möbiusband*, Monatsch. Math. **66** (1962), 276–289.

NOTATION INDEX

INDEX