# Chapter 1

# DIFFERENTIAL GEOMETRY OF REAL MANIFOLDS

## 1.1 Simplest Applications of Structure Equations

### 1.1.1 Moving Frames in Euclidean Spaces

We equip $\mathbb{R}^N$ with the standard inner product $<,>$. By a *moving frame* in $U \subseteq \mathbb{R}^N$ we mean a choice of orthonormal bases $\{e_1(x), \cdots, e_N(x)\}$ for all $\mathcal{T}_xU$, $x \in U$. Taking exterior derivatives we obtain

$$dx = \sum_A \omega_A e_A, \quad de_A = \sum_B \omega_{BA} e_B \tag{1.1.1}$$

where $\omega_A$'s and $\omega_{AB}$'s are 1-forms. Since $\omega_A$ and $\omega_{AB}$ depend on the point $x$ and the choice of the moving frame $\{e_1, \cdots, e_N\}$, their natural domain of definition is the principal bundle $\mathcal{F}_g \rightarrow U$ of orthonormal frames on $U$. However, due to the functorial property of the exterior derivative ($f^\star(d\eta) = df^\star(\eta)$), the actual domain is immaterial for many calculations and sometimes we use local parametrizations in our computations. The orthonormality condition implies $0 = d < e_A, e_B > = < de_A, e_B > + < e_A, de_B >$ and consequently

$$\omega_{AB} + \omega_{BA} = 0. \tag{1.1.2}$$

That is, the matrix valued 1-form $\omega = (\omega_{AB})$ takes values in the Lie algebra $\mathcal{SO}(N)$. From $ddx = 0$ and $dde_A = 0$ we obtain

$$d\omega_A + \sum_B \omega_{AB} \wedge \omega_B = 0; \quad d\omega_{AB} + \sum_C \omega_{AC} \wedge \omega_{CB} = 0. \tag{1.1.3}$$

These equations are often called the *structure equations* for Euclidean space or more precisely for the group of rigid motions of Euclidean space. The second set of equations is also known as the *structure equations* for the (proper) orthogonal group. These equations are a special case of Maurer-Cartan equations as discussed in chapter 1, and here we have given another derivation of them. In fact, by fixing an origin and an orthonormal frame the set of (positively oriented) frames on $\mathbb{R}^n$ can be identified with the group of (proper) Euclidean motions, and (1.1.3) becomes identical with the Maurer-Cartan equations where we have represented the $(N+1)\times(N+1)$ matrix $U^{-1}dU$ in the form (see chapter 1, §3.5)

$$\begin{pmatrix} \omega_{11} & \cdots & \omega_{1N} & \omega_1 \\ \vdots & \ddots & \vdots & \vdots \\ \omega_{N1} & \cdots & \omega_{NN} & \omega_N \\ 0 & \cdots & 0 & 0 \end{pmatrix}$$

### 1.1.2 Curves in the Plane

Geometry of curves in the plane is the simplest and oldest area of differential geometry. Let us interpret the 1-form $\omega_{12}$ in the context of plane curves. Choose the orthonormal moving frame $e_1, e_2$ such that $e_1$ is the unit tangent vector field to $\gamma$ and $e_1, e_2$ is positively oriented for the standard orientation of the plane. Then we have $de_1 = \omega_{21}e_2$ and $de_2 = \omega_{12}e_1$. The 1-form $\omega_{21}$ or $\omega_{21}(e_1)$ has a familiar geometric interpretation. Set $\gamma(t) = (x(t), y(t))$ and let $\dot{x} = \frac{dx}{dt}$ etc. Then

$$\omega_{21} = \frac{\dot{x}\ddot{y} - \ddot{x}\dot{y}}{\dot{x}^2 + \dot{y}^2}dt = d\phi, \tag{1.1.4}$$

where $\tan\phi = \frac{\dot{y}}{\dot{x}}$, i.e., $\tan\phi$ is the slope of the tangent to $\gamma$ at $\gamma(t)$. Therefore $\omega_{21}(e_1)$ is the curvature $\kappa$ of the plane curve as defined in elementary calculus. (This interpretation of $\omega_{21}$ may be somewhat misleading in higher dimensions as we shall see later.) The quantity $|\omega_{21}(e_1)|$ is independent of the parametrization of $\gamma$ although the sign of $\omega_{21}(e_1)$ depends on the choice of orientation for $\mathbb{R}^2$ and whether we are traversing the curve in the counterclockwise or the clockwise direction. Since it is conventional to assign positive curvature to the circle, we use the standard orientation for $\mathbb{R}^2$ and move counterclockwise along the curve. It is also convenient to parametrize $\gamma$ by its arc-length $s$, i.e. $t = s$, so that $\dot{x}^2 + \dot{y}^2 = 1$. We shall do so for the remainder of this subsection. We also recall from our treatment of immersions of the circle into $\mathbb{R}^2$ (chapter 1, §5.5 (????)) that if $\mathsf{G} : C \to S^1 \subset \mathbb{C}$ is defined by $\mathsf{G}(s) = e_i(s)$, $i = 1$ or 2, and $d\theta$ denotes the standard measure on the circle, then

$$\mathsf{G}^\star(d\theta) = \kappa(s)ds. \tag{1.1.5}$$

Therefore $\int_C \omega_{12}$ is the winding number of the curve $C$.

**Example 1.1.1** Let $\gamma : I \to \mathbb{R}^2$ be a curve with curvature $\kappa$ and assume $\gamma$ is parametrized by arc-length $s$. We give a geometric interpretation of curvature of $\Gamma$ which is sometimes useful. Assume $\kappa \neq 0$ and by reversing the orientation of $\gamma$ we assume $\kappa > 0$. Let $e_1, e_2$ be moving frame with $e_1$ (resp. $e_2$) tangent (resp. normal) to the curve. Consider the mapping $\Phi : I \times [0, \epsilon) \to \mathbb{R}^2$

$$\Phi(s,t) = \gamma(s) + te_2(s).$$

From calculus we know that if $\epsilon < \min_s \frac{1}{|\kappa(s)|}$, then the mapping $\Phi$ is injective. Furthermore, the image of $(0,1) \times (0,\epsilon)$ is an open subset of $\mathbb{R}^2$. A general point in $U = \Phi((0,1) \times (0,\epsilon))$ has a unique representation $q = \gamma(s) + te_2$ and

$$dq = [ds + t\omega_{21}]e_1(s) + dte_2(s).$$

We obtain

$$dv_U = ds \wedge dt + t\omega_{21} \wedge dt$$

for the volume element on $U \subset \mathbb{R}^2$. This expression for the volume element is no longer valid if $\epsilon$ is large. In fact, if $\kappa(s) > 0$ (for the chosen orientation of the curve), and the point $y$ along the normal $e_3$ is at a distance $> \frac{1}{\kappa(s)}$ from $\gamma(s)$, then there are points $s' \neq s$ arbitrarily close to $s$ such that

$$d(y, \gamma(s)) \geq d(y, \gamma(s')), \tag{1.1.6}$$

and the map $\Phi$ is no longer a diffeomorphism. This observation will be useful later. ♠

Let $\gamma : [0, L] \to \mathbb{R}^2$ be a simple closed curve parametrized by arc length $s$ and denote its image by $\Gamma$. Such a curve decomposes the plane into two connected components which are the *interior* $\Gamma_i$ and the *exterior* $\Gamma_e$ of $\Gamma$. This is a topological fact which is so familiar from experience that assuming its validity will not be a cause for concern. For a $C^1$ simple closed curve, we can define the interior $\Gamma_i$ as the set of points $q \notin \Gamma$ such that a generic ray (i.e., half infinite straight line) starting at $q$, intersects $\Gamma$ is an odd number of points. Later, we will discuss a generalization and rigorous proof of this fact. We say $\Gamma$ is *convex* if $\Gamma_i$ is a convex set. It is not difficult to show that the convexity of $\Gamma$ is equivalent to any one of the following conditions:

1. For every tangent line $T$ to $\Gamma$, $\Gamma_i \cap T = \emptyset$.

2. For every tangent line $T$, $\Gamma_i$ lies on one side of $T$.

3. The intersection of any straight line with $\Gamma$ has at most two connected components.

These descriptions of convexity in the plane are so familiar that if necesary we will make use of them without further ado.

There is also a differential geometric condition which implies convexity. Intuitively, this condition says that if the curvature of $\Gamma$ does not vanish, then the angle that the inward (or outward) normals to $\Gamma$ make with a fixed direction, e.g. the positive $x$-axis gives a parametrization of the curve $\Gamma$ and $\Gamma$ is convex, in other words,

**Lemma 1.1.1** *The Gauss map $\mathsf{G}: \Gamma \rightarrow S^1$ of a simple closed curve with nowhere vanishing curvature is a diffeomorphism, and $\Gamma$ is convex.*

**Proof** - Since the winding number of a simple closed curve is $2\pi$ (see chapter 1, §5.5) we know that $\mathsf{G}$ is onto. If $\mathsf{G}(s_1) = \mathsf{G}(s_2)$ then by Rolle's theorem $\mathsf{G}'(t)$ vanishes for some $t$ and consequently there is a point with zero curvature. Similarly, if $\Gamma$ does not lie on one side of a tangent line $T$, then it intersects $\Gamma$ in at least two distinct points $q_1, q_2$. By Rolle's theorem there is a point between $q_1$ and $q_2$ where the tangent line is parallel to $T$. This contradicts the first assertion of the lemma proving convexity of $\Gamma$. ♣

It is customary to refer to a simple closed curve $\Gamma \subset \mathbb{R}^2$ with nowhere vanishing curvature as *strictly convex*. Since lemma 1.1.1 shows that the unit circle parametrizes a simple closed strictly convex curve, we ask whether any positive function $\kappa$ on $S^1$ can be realized as the curvature of such a curve $\Gamma$ with $\kappa(\theta)$ the curvature at the unique point on $\Gamma$ with normal $e_2 = e^{i\theta}$. The following proposition shows that there is a necessary condition to be satisfied:

**Proposition 1.1.1** *Let $e_1, e_2$ denote the unit tangent and normal to the simple closed curve $\Gamma$ given by an embedding $\gamma : S^1 \rightarrow \mathbb{R}^2$, and $\tilde{\kappa}$ denote the curvature of $\Gamma$ regarded as a function on $S^1$. Then*

$$\int_{S^1} \frac{1}{\tilde{\kappa}} e_2 d\theta = 0 \quad \text{or equivalently} \quad \int_{S^1} \frac{1}{\tilde{\kappa}} e_1 d\theta = 0.$$

*If $\tilde{\kappa}$ is positive, then this condition is also sufficient for the existence of a strictly convex simple closed curve with curvature $\tilde{\kappa}$ relative to the parametrization of lemma 1.1.1*

**Proof** - Since $\gamma$ is a simple closed curve, it can be given as an embedding of $S^1$ into $\mathbb{R}^2$. Let $s$ be the arc length along $\Gamma$ and $\theta$ the parameter along $S^1$. Then

$$\frac{d\gamma}{d\theta} = \frac{d\gamma}{ds}\frac{ds}{d\theta}. \tag{1.1.7}$$

Therefore $d\gamma = \frac{1}{\tilde{\kappa}} e_1 d\theta$ and the necessity follows by integration $\int_{S^1} d\gamma = 0$ and the fact that $e_1$ and $e_2$ differ by the constant rotation through $\frac{\pi}{2}$. To prove the sufficiency assertion we integrate the equation

$$\frac{d\gamma}{d\theta} = \frac{1}{\tilde{\kappa}} e_1, \quad \text{where} \quad e_1 = e^{i(\frac{\pi}{2}+\theta)},$$

on the circle $S^1$. The periodicity of the solution $\gamma$ follows from the hypothesis. From (1.1.7) we see that $\tilde{\kappa}$ is the curvature of and $e_1$ is the unit tangent vector field to $\gamma$. From the positivity of $\tilde{\kappa}$ it follows that for $0 \leq \alpha \leq \beta \leq \pi$

$$\int_{\alpha}^{\beta} \frac{1}{\tilde{\kappa}} e^{i(\frac{\pi}{2}+\theta)} d\theta = 0$$

is not possible unless $\alpha = \beta$. Let $\alpha, \beta \in S^1$. Since one of the two circular arcs $(\alpha, \beta)$ or $(\beta, \alpha)$ is $\leq \pi$, $\gamma(\alpha) = \gamma(\beta)$ is not possible unless $\alpha = \beta$. Therefore $\gamma$ is an embedding and sufficiency follows. ♣

Probably the best known classical result in the geometry of simple closed curves in the plane is the *Four Vertex* or *Mukhopadhyaya's theorem*, namely,

**Proposition 1.1.2** *The curvature function of a simple closed curve in the plane has at least two maxima and two minima.* (If $\kappa$ is constant on any arc, then by convention it has infinitely many maxima and minima.)

**Proof** - Assume $\kappa$ has only one maximum and one minimum which occur at points with parameter values $\theta_+$ and $\theta_-$. It follows from the Intermediate Value theorem that there is a pair of antipodal points $e^{\pm i\theta_\circ}$ such that

$$\tilde{\kappa}(\theta_\circ) = \tilde{\kappa}(-\theta_\circ).$$

This gives the decomposition of the circle into two semi-circles such that the curvature on one is everywhere greater than on the other which contradicts the necessary condition of proposition 1.1.1. The number of maxima and minima necessarily being equal, we obtain the required result. ♣

**Exercise 1.1.1** *Show that the conclusion of proposition 1.1.2 may not be valid for an immersion of $S^1$ into $\mathbb{R}^2$.*

**Remark 1.1.1** Proposition 1.1.1 has a higher dimensional analogue which will be discussed in the subsection on Christoffel, Minkowski and Weyl problems. If we relax the particular parametrization defined by the Gauss map $\mathsf{G}$ and only require that $\kappa$ be the the curvature after some diffeomorphism of $\Gamma$ onto $S^1$, then any positive function with at least two maxima and two minima can be realized as the curvature function of a simple closed strictly convex curve. Proving this requires studying the diffeomorphism group of the circle and will not be pursued here. The Four Vertex theorem can be proven without reference to proposition 1.1.1, however, the above proof is preferable since it relates it to the Minkowski problem. In [Oss1] an estimate for the number of critical points of the curvature of a simple closed curve is given. ♡

In the above analysis we used integration to deduce from the local data $\kappa(s)$ a global result, namely, the existence of four critical points. The use of integration in going from local information to global consequences is a common occurrence in differential geometry. Formulating a problem as the solution to a variational problem (i.e., existence of critical points) is another general device (besides integration) for obtaining global geometric information. There are many examples of this kind of argument in geometry and physics. Example 1.1.2 below, due to Tabachnikov, demonstrates this general principle, in the context of convex curves in the plane, in an elementary yet elegant manner. First we need to recall an observation from plane geometry. For vectors $OA = (a_1, b_1)$ and $OB = (a_2, b_2)$ in the plane we define

$$OA * OB = \det \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix}. \tag{1.1.8}$$

From elementary geometry we recall that $OA * OB$ is the signed area of the parallelogram determined by the vectors $OA$ and $OB$ or equivalently twice the signed area of the triangle $OAB$. The sign is positive or negative according as the vectors $OA, OB$ form a positively or negatively oriented basis.

**Example 1.1.2** Let $\Gamma$ be a simple closed convex curve in $\mathbb{R}^2$ and assume that its curvature is nowhere zero. For an angle $\phi \in S^1$ we let $\phi$ also denote the unique point on $\Gamma$ with $\mathsf{G}(\phi) = \phi$ which causes no confusion in view of lemma 1.1.1. We seek points $\phi \in \Gamma$ such that the normals to $\Gamma$ at $\phi - \frac{2\pi}{3}, \phi$ and $\phi + \frac{2\pi}{3}$ are concurrent. We refer to such a configuration of normals as a *tripod*. It is clear that for such $\phi \in \Gamma$ (if exists), the three normals intersect at angles of $\pm\frac{2\pi}{3}$ at their common point of interesection. For arbitrary $\phi$, the intersections of the three normals to $\Gamma$ at $\phi - \frac{2\pi}{3}, \phi$ and $\phi + \frac{2\pi}{3}$ form the vertices of an equilateral triangle, and we will show as $\phi$'s moves along the curve this triangle degenerates (at least twice) into a point and the three normals become concurrent. To prove the existence of a (or two) tripod(s) fix an origin $O$ not lying on the curve $\Gamma$ and it is perhaps less confusing (although unimportant) if we take the origin to be in the exterior of the curve. Let $p = (x, y)$ denote the vector from $O$ to a point with coordinates $(x, y)$ on $\Gamma$. We make the convention that $p(\phi)$ denotes the vector from $O$ to the point on $C$ corresponding to the parameter value $\phi \in S^1$ as described above, but $p'(\phi)$ and $p''(\phi)$ denote the first and second derivatives of $p$ (at $\phi$) relative to the arc length $s$ on $\Gamma$. Consider the function

$$F(\phi) = p(\phi - \frac{2\pi}{3}) * p'(\phi - \frac{2\pi}{3}) + p(\phi) * p'(\phi) + p(\phi + \frac{2\pi}{3}) * p'(\phi + \frac{2\pi}{3})$$

defined on $\Gamma$. Then

$$\frac{dF}{ds}(\phi) = p(\phi - \frac{2\pi}{3}) * p''(\phi - \frac{2\pi}{3}) + p(\phi) * p''(\phi) + p(\phi + \frac{2\pi}{3}) * p''(\phi + \frac{2\pi}{3}). \tag{1.1.9}$$

As noted earlier, the three terms on the right hand side of (1.1.9) are twice the signed areas of the triangles $OAB, OAC$ and $OBC$ (see figure XXXX) and their sum is twice the area of the triangle $ABC$ where $A, B$ and $C$ are the intersections of the normals to the curve $\Gamma$ at the points $\phi - \frac{2\pi}{3}, \phi$ and $\phi + \frac{2\pi}{3}$. The function $F$ has at least two critical points and at these critical points, the triangle $ABC$ degenerates into a point (since the area of the equilateral triangle $ABC$ becomes zero) and we obtain the desired tripods. Tabachnikov also established a similar property for convex polygons. For this and other material on the Four Vertex theorem see [Tab] and references thereof. ♠

A curve in the plane is completely determined, up to Euclidean motion, by its curvature. In fact we have the differential equation $de_1 = \kappa e_2$ where the vector $e_2$ is uniquely determined by the requirement that $e_1, e_2$ is positively oriented orthonormal frame. The differential equation is uniquely solvable once the initial point and initial direction are specified. This observation is both local and global and can be stated as follows:

**Lemma 1.1.2** *Let $\gamma, \gamma' : [0, L] \rightarrow \mathbb{R}^2$ be two $C^3$ plane curves parametrized by arc length $s$, and assume their curvatures as equal as a function of $s$. Then $\gamma$ and $\gamma'$ differ by a Euclidean motion.*

### 1.1.3 Curves in Space

In order to make use of the structure equations to study geometry of curves in space, we make a special choice for the frame $e_1, e_2, e_3$. Consider a curve $\Gamma \subset \mathbb{R}^3$, and choose the frame $\{e_1, e_2, e_3\}$ such that $e_1$ is the unit tangent to $\Gamma$ and set

$$de_1 = \frac{1}{\rho} e_2 ds, \quad de_2 = -\frac{1}{\rho} e_1 ds + \tau e_3 ds, \quad de_3 = -\tau e_2 ds. \tag{1.1.10}$$

The quantities $\kappa = \frac{1}{\rho}$ and $\tau$ are called the *curvature* and *torsion* of the curve. The frame $\{e_1, e_2, e_3\}$ is called the *Frenet frame* for the curve $\Gamma$. The notation $\frac{1}{\rho}$ implicitly assumes that the curve $\Gamma$ is generic in the sense that its curvature is nowhere zero. At a point where the curvature is non-zero, there is an ambiguity of $\pm$ in the choice of $e_2$ while at point where curvature vanishes, $e_2$ can be any unit vector normal to $e_1$. In the former case the ambiguity can be removed by the requirement that $\frac{1}{\rho} > 0$. Note that the sign of the curvature of a space curve cannot be intrinsically defined. In fact, since a reflection in the plane can be extended to an element of $SO(3)$, simple examples show that there is no continuous function $\kappa$ on $\Gamma$ with the following properties:

1. In the limit of a plane curve, $\kappa$ tends to the curvature of the plane curve;

2. $\kappa$ is $SO(3)$-invariant.

On the other hand, if we stipulate on the positivity of the curvature of a space curve, then as we pass through a point of zero curvature, the vector field $e_2$ may undergo a discontinuity. To circumvent this difficulty, at least in the case where the curvature vanishes only at isolated points, we choose Frenet frames on open connected subsets $\Gamma_1, \Gamma_2, \cdots$ of $\Gamma$ where $\frac{1}{\rho} \neq 0$. By appropriate choice of $\pm$ sign of $e_2$ on each $\Gamma_i$ we can obtain a smooth Frenet frame on the entire curve. But by doing so we allow the curvature to take negative values as well. One can remove any ambiguity in the sign of $e_2$ or curvature by the (non-canonical) requirement of positivity of curvature at one point in $\cup\Gamma_i$. No confusion should arise as long as one keeps these issues in mind. The analogue of lemma 1.1.2 is also valid for space curves:

**Lemma 1.1.3** *Let $\gamma, \gamma' : [0, L] \rightarrow \mathbb{R}^3$ be two $C^4$ curves in space parametrized by arc length $s$, and assume their curvatures and torsions as equal as a function of $s$. If the curvature of $\gamma$ (or $\gamma'$) vanishes nowhere, then $\gamma$ and $\gamma'$ differ by a Euclidean motion.*

**Proof** - Let $A = (e_1, e_2, e_3)$ be the $3 \times 3$ matrix denoting the Frenet frame. Then the solution to the differential equation

$$\frac{dA}{ds} = A \begin{pmatrix} 0 & -\kappa & 0 \\ \kappa & 0 & \tau \\ 0 & -\tau & 0 \end{pmatrix}$$

lies in $O(3)$ in view of skew-symmetry of the matrix $A^{-1}\frac{dA}{ds}$. By choosing the initial condition to be a positively oriented orthonormal frame we ensure $A \in SO(3)$. Solving the differential equation $\frac{dp}{ds} = e_1$ we obtain a curve $\gamma''$ which will coincide with either $\gamma$ or $\gamma'$ by a judicious choice of the initial conditions, i.e. a Euclidean motion. ♣

**Exercise 1.1.2** *How should lemma 1.1.3 be modified if the curvature is allowed to vanish at isolated points.*

**Example 1.1.3** As an application of the Frenet frame we calculate the volume of a tube of small radius $r > 0$ around a curve $\Gamma$ in $\mathbb{R}^3$. The tube of radius $r > 0$ around $\Gamma$ is

$$\tau_r(\Gamma) = \{p + t_2e_2 + t_3e_3 | p \in \Gamma, \; t_2^2 + t_3^2 < r\}.$$

Denoting a generic point on $\tau_r(\Gamma)$ by $q = p + t_2e_2 + t_3e_3$ we obtain

$$dq = (ds - t_2\kappa ds)e_1 + (dt_2 - t_3\tau ds)e_2 + (dt_3 + t_2\tau ds)e_3,$$

so that the volume element on $\tau_r(\Gamma)$ is

$$ds \wedge dt_2 \wedge dt_3 - t_2 \kappa ds \wedge dt_2 \wedge dt_3.$$

It is clear that because of the factor $t_2$, the integral on $\tau_r(\Gamma)$ of the second term vanishes. Therefore

$$\mathrm{vol}(\tau_r(\Gamma)) = \pi r^2(\mathrm{length}(\Gamma)). \tag{1.1.11}$$

Thus the volume of the tube depends only on the length of the curve and $r$ but not the curvature or torsion of $\Gamma$. Of course this simple formula is valid for curves in $\mathbb{R}^N$ if we replace $\pi r^2$ by the volume of the ball of radius $r > 0$ in $\mathbb{R}^{N-1}$. Formula (1.1.11) is valid only for small values of $r > 0$ since the parametrization $q = p + t_2 e_2 + t_3 e_3$ is valid only for small $r > 0$. One may be tempted to assume that the right hand side of (1.1.11) gives an upper bound for the volume of $\tau_r(\Gamma)$ for all $r$, however simple examples show that this is not true. ♠

Let $\gamma : [0, L] \to \mathbb{R}^2$ be a curve in the plane (parametrized by arc length for convenience) and $M_\gamma$ be the cylinder based on $\gamma$, i.e., $\{(x, y, z) \mid (x, y) \in \mathrm{Im}(\gamma)\}$. Consider the mapping

$$\Phi : (-1, 1) \times [0, L] \to \mathbb{R}^3, \quad \Phi(u, t) = (\gamma(t), u)$$

Let $e_1, e_2, e_3$ be a positively oriented moving frame in $\mathbb{R}^3$ with $e_3$ normal to $M_\gamma$ and $e_1$ tangent to $\mathrm{Im}(\gamma)$. A possible choice is to take $e_1$ parallel to the $(x, y)$-plane and $e_2$ in the direction of $z$-axis. Then

$$d\gamma = \omega_1 e_1, \quad \omega_2 = du, \quad \omega_{12} = 0.$$

The last equality follows from the fact that $e_2 \cdot de_1 = 0$ since $\gamma$ is a plane curve. Let $\delta(t) = (\delta_1(t), \delta_2(t), \delta_3(t))$ be a curve such that $(\delta_1(t), \delta_2(t)) = \gamma(t)$ and $\delta_3(t) > 0$ for all $t \in [0, L]$. Modify the frame to $e_1', e_2', e_3$ differing from $e_1, e_2, e_3$ by an element $SO(2)$ acting on the $e_1, e_2$ vectors so that $e_1$ is tangent to both $\gamma$ and $\delta$ curves. Then it is a simple calculation that $\omega_{12}'$ is related to $\omega_{12}$ by

$$\omega_{12}' = \omega_{12} + d\theta = d\theta,$$

where $\theta$ is the angle of rotation relating the frames under consideration. Therefore $d\omega_{12}' = 0$ and it follows from Stokes' theorem that

$$\int_\circ^L \gamma^\star(\omega_{12}') - \int_\circ^L \delta^\star(\omega_{12}') = \Lambda_{\gamma\delta}, \tag{1.1.12}$$

where $\Lambda_{\gamma\delta}$ is the contribution of the line integrals of $\omega'_{12}$ along the vertical lines joining the initial and end points of $\gamma$ to those of $\delta$. We refer to $\omega'_{12}(e'_1)$ (on $\delta$) as the curvature $\kappa_\delta$ of the space curve $\delta$. It is convenient to write the integrals on the left hand side of (1.1.12) as $\int_\gamma \omega'_{12}$ and $\int_\delta \omega'_{12}$, and refer to these quantities as the total curvatures of $\gamma$ and $\delta$.

Now assume $\delta$ is a simple closed curve in $\mathbb{R}^3$, i.e., $\delta$ is a diffeomorphism of the circle onto its image. One refers to such a curve as a *knot*. Let $\gamma$ denote the orthogonal projection of $\delta$ in the $(x, y)$-plane. Now $\gamma$ generally has self-intersections which we may assume (by transversality or making small perturbations) are of the form of two branches passing through a point, i.e., no triple or higher intersections. We can break up $\gamma$ into a union of curves $\gamma_1, \cdots, \gamma_N$ with no self intersections, and accordingly decompose $\delta$ into a union $\delta_1, \cdots, \delta_N$ with $\delta_j$ lying vertically above $\gamma_j$. It follows from (1.1.12) that

$$\int_\gamma \omega'_{12} - \int_\delta \omega'_{12} = \sum_j \Lambda_{\gamma_j\delta_j}.$$

Since $\delta$ is a closed curve, the sum $\sum_j \Lambda_{\gamma_j\delta_j}$ vanishes and

$$\int_\gamma \omega'_{12} = \int_\delta \omega'_{12}. \tag{1.1.13}$$

The left (resp. right) hand side of (1.1.13) is the integral of the curvature of the space $\gamma$ (resp. plane $\delta$). We want to obtain an estimate for $\int_\gamma |\omega'_{12}|$ where absolute value sign means we are calculating the integral of the absolute value of the curvature. The principle is best demonstrated by looking at an example. Consider the knot $\delta$ in figure (XXXX) known as the *trefoil* knot. This is the simplest non-trivial knot. Non-trivial means it cannot be deformed into a circle without crossing itself. In chapters 4 and 6 we will make a systematic study of knots, but for the time being the intuitive notions will suffice. Now break up the orthogonal projection $\gamma$ of the trefoil knot in the plane into three simple closed curves as shown in the figure and denote them by $C_1, C_2$ and $C_3$. We orient $C_i$'s in the counterclockwise direction. Each simple closed curve in the plane has total curvature $2\pi$ as noted in chapter 1, §5.3. In replacing $\gamma$ with three simple closed curves we created two issues which have to be addressed, viz.,

1. There are additional contributions to the total curvature by twice the sum of the angles of the triangle $ABC$ (see figure XXXX).

2. The orientation of of portions of the curves were reversed and consequently the curvature was multiplied by -1 on these sections.

The first point implies that if we estimate $\int_\gamma |\omega'_{12}|$ by relating it to $\sum \int_{C_i} \omega'_{12}$, then $2\pi$ should be substracted from it to compensate for this additional contribution. The second issue is addressed by noting we are looking at $\int_\gamma |\omega'_{12}|$ and therefore we have the inequality

$$\int_\gamma |\omega'_{12}| \geq \sum_{i=1}^{3} \int_{C_i} \omega'_{12} - 2\pi = 4\pi. \tag{1.1.14}$$

This simple argument can be applied to any knot to give a lower bound for $\int_\gamma |\omega'_{12}|$, but carrying out the details rigorously involves a technical examination of knot crossings is not very interesting. The reader should experiment with more complex knots to be convinced of the validity of

$$\int_\delta |\omega'_{12}| \geq 4\pi \tag{1.1.15}$$

for any non-trivial knot $\delta$. An elegant and simple proof of it, based on Crofton's formula for the sphere, is given in the next subsection. The inequality (1.1.15) is known as the *Fary-Milnor* theorem.

### 1.1.4 Integral Geometry in Dimension 2

To further demonstrate the use of moving frames and how the group of proper motions of Euclidean space enters into geometric problems we consider some problems in integral geometry in the plane and on the unit sphere $S^2$. These examples will not be used in the discussion of Riemannian geometry and the reader may directly proceed to the next section on Riemannian geometry. Let $C$ and $C'$ be curves in the plane $\mathbb{R}^2$, and pose the following questions:

1. What is the *average* number of intersections of $C$ and $g(C')$ as $g$ ranges over $SE(2)$ (the group of proper Euclidean motions of $\mathbb{R}^2$?
2. What is the *average* number of intersections of $C$ and an affine line in $\mathbb{R}^2$?

In both of these problems we have to give a meaning to the word *average*. Let $N(g)$, $g \in SE(2)$, denote the number of points of intersection of $g(C')$ and $C$. Then the desired average is $\int N(g)$ where the integration is over the space of all possible $g(C')$. To make this more precise let $SE(C, C') = \{g \in SE(2) | g(C') \cap C \neq \emptyset\}$, and let $m(SE(C, C'))$ denote the measure of this set relative to the kinematic density $dv_{SE(2)}$. Denote the lengths of $C$ and $C'$ by $l$ and $l'$, and let $s$ and $s'$ be the arc-length along these curves. Consider the mapping

$$F : [0, l] \times [0, l'] \times [-\pi, \pi) \to SE(2),$$

where $F(s, s', \theta)$ is the proper Euclidean motion which translates the point with coordinate $s'$ on $C'$ to the origin, followed by rotation through angle $\theta$ and translation of the origin to the point with coordinate $s$ on $C$. Then computing $m(SE(C, C'))$ can be restated as integrating $F^\star(dv_{SE(2)})$ on $[0, l] \times [0, l'] \times [-\pi, \pi)$. Notice that in this calculation, distinct points of intersection of the curves $g(C')$ and $C$ correspond to different values in the domain of $F$. Therefore we have

$$\int N(g) = \int_0^l \int_0^{l'} \int_{-\pi}^{\pi} F^\star(dv_{SE(2)}).$$

To evaluate this integral let $(x_1(s), x_2(s))$ and $(y_1(s'), y_2(s'))$ be parametrizations of $C$ and $C'$ by arc length. Then

$$F(s, s', \theta) = \begin{pmatrix} \cos\theta & -\sin\theta & x_1(s) - y_1(s')\cos\theta + y_2(s')\sin\theta \\ \sin\theta & \cos\theta & x_2(s) - y_1(s')\sin\theta - y_2(s')\cos\theta \\ 0 & 0 & 1 \end{pmatrix}$$

Therefore

$$F^\star(dv_{SE(2)}) = dv_1 \wedge dv_2 \wedge d\theta = -[(x_1'y_1' + x_2'y_2')\sin\theta + (x_1'y_2' - x_2'y_1')\cos\theta] ds \wedge ds' \wedge d\theta.$$

Let $\alpha$ and $\alpha'$ be the angles of the tangents to $C$ and $C'$ with $x_1$-axis at the points corresponding to $s$ and $s'$ respectively. Then $x_1' = \cos\alpha$, $x_2' = \sin\alpha$ etc. and we obtain

$$(x_1'y_1' + x_2'y_2')\sin\theta - (x_1'y_2' - x_2'y_1')\cos\theta = -\sin(\theta - \alpha' + \alpha)$$

Note that $\varphi \equiv \theta - \alpha' + \alpha$ is the angle between the curves $C$ and $F(s, s', \theta)(C')$. Since $\alpha$ and $\alpha'$ depend only on $s$ and $s'$ respectively, we have the expression

$$F^\star(dv_{SE(2)}) = \pm \sin\varphi ds \wedge ds' \wedge d\varphi, \tag{1.1.16}$$

for the pull-back of the kinematic density. (The reason for the ambiguity in sign is that we have not specified orientations when measuring the angles; we require the measure to be positive which determines the sign.) Therefore

$$\text{meas}(S(C, C')) = \int_0^l \int_0^{l'} \int_{-\pi}^{\pi} |\sin\varphi| d\varphi ds' ds = 4ll', \tag{1.1.17}$$

and

$$\int N(g) = 4ll'. \tag{1.1.18}$$

The equation (1.1.18) is often called *Poincaré's formula.* The second problem is only slightly different. The set of all affine lines in $\mathbb{R}^2$ is the inhomogeneous Grassmann manifold $\tilde{\mathbf{G}}^{\circ}_{1,1}(\mathbb{R}) = SE(2)/H$, where $H = \mathbb{R}\cdot O(1) = \mathbb{R}\cdot\mathbb{Z}/2$ is the group Euclidean motions of $\mathbb{R}$. The desired average is $\int n(L)$ where $n(L)$ is the number of intersections of the line $L \in \tilde{\mathbf{G}}^{\circ}_{1,1}$ with the curve $C$, and integration is over $\tilde{\mathbf{G}}^{\circ}_{1,1}$. We consider the map $F : [0,l] \times [0,\pi) \rightarrow \tilde{\mathbf{G}}_{1,1}(\mathbb{R})$ where $F(s,\theta)$ is the coset $gH$ and $g \in SE(2)$ is rotation through angle $\theta$ followed by translation of the origin to the point corresponding to the point with parameter $s$ on $C$. By a reasoning as before

$$\int n(L) = \int_0^l \int_0^\pi F^\star(dv_{\tilde{\mathbf{G}}^{\circ}_{1,1}}).$$

From example **??** of chapter 1 we have $dv_{\tilde{\mathbf{G}}^{\circ}_{1,1}} = \omega_2 \wedge \omega_{12}$ which gives

$$F^\star(dv_{\tilde{\mathbf{G}}^{\circ}_{1,1}}) = \pm \cos\varphi ds \wedge d\varphi,$$

where $\varphi$ is the angle between the line $L$ and the curve $C$. Consequently

$$\int n(L) = 2l. \tag{1.1.19}$$

This is the simplest of a class of equations known as *Crofton's formula(e).* The important feature of (1.1.19) and (1.1.18) is that the right hand side is proportional to the length(s) of the curve(s). In our computations we used a mapping $F$ to pull-back a canonically defined form on a group or homogeneous space and then integrated it over the parameter space. This kind of reasoning occurs frequently in differential geometry.

**Exercise 1.1.3** *With the notation and framework of example* **??**, *let $\beta_j$ be the angle between $g(C')$ and $C$ at the $j^{th}$ point of intersection. Show that*

$$\int \sum_j \beta_j = 2\pi l l',$$

*where the integral is over all $g \in SE(2)$ such that $g(C') \cap C \neq \emptyset$.*

**Example 1.1.4** We continue with the notation of example **??**. Let $K$ be a compact subset of $\mathbb{R}^2$ with piece-wise smooth boundary and consider the problem of estimating the number of intersections in the interior of $K$ of $n$ lines (in general position) in $\mathbb{R}^2$. Let

$dv_n = \underbrace{dv_{\tilde{\mathbf{G}}^\circ_{1,1}} \wedge \cdots \wedge dv_{\tilde{\mathbf{G}}^\circ_{1,1}}}_{n \text{ copies}}$ be the invariant volume element on the product of $n$ copies of $\tilde{\mathbf{G}}^\circ_{1,1}$. Define $\epsilon_{ij} : \tilde{\mathbf{G}}^\circ_{1,1} \times \cdots \times \tilde{\mathbf{G}}^\circ_{1,1} \to \mathbb{R}$ by

$$\epsilon_{ij}(L_1, \cdots, L_n) = \begin{cases} 1 & \text{if} L_i \cap L_j \cap K \neq \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

For $n = 2$ we obtain from Crofton's formula (1.1.19)

$$\int_{\tilde{\mathbf{G}}^\circ_{1,1} \times \tilde{\mathbf{G}}^\circ_{1,1}} \epsilon_{12}(L, L') dv_{\tilde{\mathbf{G}}^\circ_{1,1}} \wedge dv_{\tilde{\mathbf{G}}^\circ_{1,1}} = 2 \int_{\tilde{\mathbf{G}}^\circ_{1,1}} l_L dv_{\tilde{\mathbf{G}}^\circ_{1,1}}(L),$$

where $l_L$ denotes the length of the segment $L \cap K$. Now it is trivial to show that $\int_{\tilde{\mathbf{G}}^\circ_{1,1}} l_L dv_{\tilde{\mathbf{G}}^\circ_{1,1}}(L) = 2a_K$, where $a_K$ is the area of $K$. Consequently

$$\int_{\tilde{\mathbf{G}}^\circ_{1,1} \times \tilde{\mathbf{G}}^\circ_{1,1}} \epsilon_{12}(L, L') dv_{\tilde{\mathbf{G}}^\circ_{1,1}} \wedge dv_{\tilde{\mathbf{G}}^\circ_{1,1}} = 4a_K. \tag{1.1.20}$$

Let $\epsilon = \sum_{i<j} \epsilon_{ij}$, $U_K = \{L \in \tilde{\mathbf{G}}^\circ_{1,1} | L \cap K \neq \emptyset\}$, and $M(n, K)$ be the measure of the set of $n$-tuples of lines such that every pair intersect inside of $K$. Then

$$N(n, K) = \int_{U_K \times \cdots \times U_K} \epsilon dv_n = 2n(n-1)a_K \int_{U_K \times \cdots \times U_K} dv_{n-2}.$$

Another application of Crofton's formula gives

$$N(n, K) = 2^{n-1} n(n-1) a_K l^{n-2}_{\partial K}.$$

where $l_{\partial K}$ is the length of the boundary curve $\partial K$. To obtain the *average* $\mathsf{N}(n, K)$ of the number of intersections, we have to properly normalize the quantity $N(n, K)$. A natural normalization is by dividing $N(n, K)$ by the measure of the set of $n$-tuples of lines intersecting $K$. From Crofton's formula the latter quantity is $2^n l^n_{\partial K}$. Therefore we obtain

$$\mathsf{N}(n, K) = \frac{n(n-1)a_K}{2l^2_{\partial K}}$$

for the desired average $\mathsf{N}(n, K)$. ♠

One should exercise caution in the probabilistic interpretation of our calculation of averages as *expected values.* Our approach was based on the integration on a space (e.g., $SE(2)$) of infinite volume, but the integrals were convergent due to compactness of the domain of integration. To give strict probabilistic interpretation it is necessary to clarify the undelying probability space. We shall not pursue this issue here.

We define the *signed total curvature* of a real plane curve as $\kappa_C = \int_C d\phi = \int_C \kappa(s)ds$ (see also notion of winding number in exercise **??**). In interpreting this quantity one should be cognizant of the fact that as one moves along a curve the angle $\phi$ can exceed $2\pi$ and the signed total curvature of a general real plane curve can be any real number. If the curve is only piece-wise smooth, again the same definition is applicable with the proviso that at points of nondifferentiability, the derivative $d\phi/ds$ is a delta function which is equal to the angle between the curves as the first rotates counterclockwise onto the second.

**Example 1.1.5** Let $D_1$ and $D_2$ be open relatively compact regions in $\mathbb{R}^2$ with piecewise smooth boundaries $\partial D_i$. Assume the boundary curves are parametrized by arc lengths $s_i$, and let $\phi_i$ be the corresponding angle. We apply proper Euclidean motions $g \in SE(2)$ to $D_1$ and look at the intersections $D(g) = g(D_1) \cap D_2$ which is an open relatively compact region with piecewise smooth boundary. Let $\kappa_i$ and $\kappa_g$ denote the signed total curvatures of the curves $\partial D_i$ and $\partial D(g)$ respectively. Just as in examples **??** and 1.1.4 we want to compute the integral of $\kappa_g$ as $g$ ranges over $SE(2)$. To do so we consider the mappings

$$\Phi_1 : \partial D_1 \times [0, 2\pi) \times D_2 \longrightarrow \partial D_1 \times SE(2), \quad \text{and} \quad \Phi_2 : \partial D_2 \times [0, 2\pi) \times D_1 \longrightarrow \partial D_2 \times SE(2),$$

where $\Phi_1(s_1, \theta, x_1, x_2) = (d\phi(s_1)/ds_1, F(s_1, \theta, x_1, x_2))$ and $F(s_1, \theta, x_1, x_2)$ is the proper Euclidean motion of $D_1$ which translates the point with parameter $s_1$ on $\partial D_1$ to the origin, then rotates the translate of $D_1$ through angle $\theta$ and then translates it so that the point with parameter $s_1$ will coincide with the point $(x_1, x_2) \in D_2$. $\Phi_2$ is similarly defined. It is a simple calculation that

$$\Phi_1^\star(\frac{d\phi_1}{ds_1} \wedge dv_{SE(2)}) = \frac{d\phi_1}{ds_1} ds_1 \wedge d\theta \wedge dx_1^2 \wedge dx_2^2, \quad \text{and} \quad \Phi_2^\star(\frac{d\phi_2}{ds_2} \wedge dv_{SE(2)}) = \frac{d\phi_2}{ds_2} ds_2 \wedge d\theta \wedge dx_1^1 \wedge dx_2^1,$$

where $dx_1^i \wedge dx_2^i$ is Euclidean volume element on the domain $D_i$. The required average is

$$\sum_{i=1}^{2} \int_{\partial D_1 \times [0,2\pi) \times D_2} \Phi_i^\star(\frac{d\phi_i}{ds} \wedge dv_{SE(2)}) + \int \sum \beta_j,$$

where $\beta_j$ is the angle between $g(\partial D_1)$ and $\partial D_2$ at the $j^{th}$ intersection point and the integral is over all $g \in SE(2)$ such that $g(\partial D_1) \cap D_2 \neq \emptyset$. Notice that the reason for symmetrizing with respect to $i = 1, 2$ is that $\partial(g(D_1) \cap D_2)$ consists of two parts coming from $g(\partial D_1)$ and $\partial D_2$. From exercise 1.1.3 we have $\int \sum \beta_j = 2\pi l_1 l_2$ and it is trivial to see that the first sum is $2\pi(\kappa_1 a_2 + \kappa_2 a_1)$ where $a_i$ is the area of the region $D_i$. Therefore we have shown that the average of the signed total curvature is

$$\int \kappa_g = 2\pi(\kappa_1 a_2 + \kappa_2 a_1 + l_1 l_2). \tag{1.1.21}$$

This equation is known as *Blaschke's formula*. For a simple closed curve $C$, $\kappa_C = 2\pi$. Therefore if $D_1$ and $D_2$ are relatively compact convex domains in $\mathbb{R}^2$ with piecewise smooth boundary, then for all $g \in SE(2)$, $\partial(g(D_1) \cap D_2)$ is a simple closed curve. Blaschke's formula then implies for

$$\mathrm{meas}(\{g \in SE(2) | g(D_1) \cap D_2 \neq \emptyset\}) = 2\pi(a_1 + a_2) + l_1 l_2,$$

under the additional hypothesis that $D_i$'s are convex. ♠

Formulae of Crofton, Poincaré and Blaschke demonstrated the use of the group of Euclidean motions of the plane in geometric problems dealing with averages. We now show that the latter two imply the isoperimetric inequality in the plane which is independent of the averages. Let $C = \partial D$ be a simple closed curve of length $l_{\partial D}$ in the plane bounding a region $D$. As noted above $\kappa_C = 2\pi$. Applying the formulae of Blaschke and Poincaré to the case where $D_i = D$ and $C = C' = \partial D$ we obtain

$$\int N(g) = 4l^2_{\partial D}, \quad \text{and} \quad \frac{1}{2\pi}\int \kappa_g = 4\pi a_D + l^2_{\partial D}$$

where the integrals are taken over the set of $g \in SE(2)$ such that $g(D) \cap D \neq \emptyset$. Assume furthermore that $D$ is convex so that $\partial(g(D) \cap D)$ is a simple closed curve (if nonempty). Let $m_j$ be the measure of the set of $g \in SE(2)$ such that $g(C)$ and $C$ intersect at exactly $i$ points. Then

$$4l^2_{\partial D} = \int N(g) = \sum i m_i, \quad \text{and} \quad 4\pi a_D + l^2_{\partial D} = \frac{1}{2\pi}\int \kappa_g = \sum m_i.$$

Since $m_i$'s are non-negative quantities and obviously $m_1 = 0$ (in fact, $m_{2i-1} = 0$) we obtain

$$l^2_{\partial D} - 4\pi a_D \geq 0, \tag{1.1.22}$$

for $D$ compact convex with piecewise smooth boundary. For $D$ non-convex let $D'$ be its convex closure. Since $a_D \leq a_{D'}$ and $l_{\partial D'} \leq l_{\partial D}$, the assumption of convexity in (1.1.22) is unnecessary.

While the isoperimetric inequality is sharp, the following clever idea gives an estimate for the defect in the inequality for convex domains with piecewise smooth boundaries. Let $r_i$ (resp. $r_e$) denote the radius of the largest (resp. smallest) circle inscribed in (resp. circumscribed about) the compact convex domain $D = D_1$. Let $r_i \leq r \leq r_e$ and $D_2$ denote the disc of radius $r$. The inequalities of Poincaré and Blaschke imply

$$8\pi r l_{\partial D} = \int N(g) = \sum i m_i, \quad \text{and} \quad 2\pi(a_D + \pi r^2 + r l_{\partial D}) = \frac{1}{2\pi}\int \kappa_g = \sum m_i.$$

Therefore, proceeding as before, we obtain

$$2\pi r l_{\partial D} - 2\pi(a_D + \pi r^2) \geq 0.$$

Now we write

$$l_{\partial D}^2 - 4\pi a_D = (l_{\partial D} - 2\pi r)^2 + 2[2\pi r l_{\partial D} - 2\pi(a_D + \pi r^2)] \geq (l_{\partial D} - 2\pi r)^2,$$

whence, by averaging,

$$l_{\partial D}^2 - 4\pi a_D \geq \frac{1}{2}[(l_{\partial D} - 2\pi r_e)^2 + (l_{\partial D} - 2\pi r_i)^2]. \tag{1.1.23}$$

Inequality (1.1.23) is called *Bonnesen inequality.*

Finally we derive an analogue of Crofton's formula for the unit sphere $S^2$. To formulate the problem let $\gamma : [0, l_\gamma] \rightarrow S^2$ be a curve of length $l_\gamma$ which we assume is parametrized by arc length, and for every $p \in S^2$ let $C_p$ be the oriented great circle on $S^2$ which is the equator relative to the north pole $p$. The set of $C_p$'s is the homogeneous space $SO(3)/SO(2) \simeq S^2$. Let $N_\gamma(p)$ be the number of intersections of the great circle $C_p$ with the curve $\gamma$. The problem is to calculate the average number of these interesections. More precisely we will prove

$$\int_{SO(3)/SO(2)} N_\gamma(p) dv(p) = 4 l_\gamma, \tag{1.1.24}$$

where $dv$ is the invariant volume element on $SO(3)/SO(2)$. Let $s$ denote arc length along $\gamma$ and $\gamma(s), e_2(s), e_3(s)$ form a positively oriented orthonormal frame. A great circle $C_p$ passes through $\gamma(s)$ if and only if the vector $p$ lies in plane spanned by $e_2(s)$ and $e_3(s)$. Therefore $p$ has can be written as

$$p = p_{s,\tau} = \cos\tau e_2 + \sin\tau e_3,$$

and $F(s,\tau) = p_{s,\tau}$ gives a (local) parametrization of $S^2$. The structure equations for $SO(3)$ imply

$$\frac{d}{ds}\begin{pmatrix}\gamma(s)\\ e_2(s)\\ e_3(s)\end{pmatrix} = A\begin{pmatrix}\gamma(s)\\ e_2(s)\\ e_3(s)\end{pmatrix},$$

where $A$ is a skew symmetric matrix depending on $s$ and the hypothesis that $\gamma$ is parametrized by arc length implies $A_{12}^2 + A_{13}^2 = 1$ and therefore we have $A_{12} = \cos\phi$, $A_{13} = \sin\phi$. Taking exterior derivative of $p_{s,\tau}$ and using the structure equations we obtain

$$dp_{s,\tau} = (-\sin\tau e_2 + \cos\tau e_3)(d\tau + A_{23}ds) - (\cos\phi\cos\tau + \sin\phi\sin\tau)ds.$$

Therefore the volume element of $S^2$ in $(\tau, s)$ coordinates is

$$F^\star(dv_{S^2}) = \cos(\tau-\phi)d\tau ds.$$

The desired average is the integral of $|F^\star(dv_{S^2})| = |\cos(\tau-\phi)|d\tau ds$ on $[0,l_\gamma]\times[0,2\pi)$ where the absolute value is necessary to make sure cancelations due to the signs of intersections do not occur. We obtain

$$\int_\circ^{l_\gamma} ds \int_\circ^{2\pi} |\cos(\tau-\phi)|d\tau = 4l_\gamma,$$

which is the desired formula (1.1.24).

As an application of Crofton's formula for the sphere, we give a simple proof of the Fary-Milnor theorem (1.1.15). Consider a knot $\delta : S^1 \to \mathbb{R}^3$ and let $\gamma : [0,l] \to S^2$ be the unit tangent vector field to the knot. It is no loss of generality to assume that $\gamma$ is an immersion of $S^1$ into $S^2$ and $l = l_\gamma$ is the length of the curve on $S^2$ traced out by $\gamma$. Let $\kappa$ denote the curvature of $\delta$ and first assume $\delta$ is only an arc on which $\kappa$ is positive. Then

$$\int \kappa ds = \int \frac{d\gamma(s)}{ds}ds = l_\gamma.$$

It follows from Crofton's formula (1.1.24) that

$$\int_\delta \kappa ds = \frac{1}{4}\int N_\gamma(p)dv(p). \tag{1.1.25}$$

Clearly this formula remains valid if we break up $\delta$ into subsets where $\kappa$ does not change sign and replace $\kappa$ by $|\kappa|$. Now observe that $N_\gamma(p)$ is the number of critical points of the function $f_p(t) =< p, \delta(t) >$ defined on the knot $\delta$ where $< ., . >$ denotes the standard inner product on $\mathbb{R}^3$. ($f_p$ is the *height function* in the direction of $p \in S^2$ and is the projection of $\delta(t)$ on the line through $p$.) It is elementary that the number of critical points of $f_p$ is even. If the total curvature of the knot $\delta$ is $< 4\pi$, then (1.1.25) implies that there is $p \in S^2$ such that $f_p$ has only two critical points, namely, a maximum $p_{\max}$ and a minimum $p_{\min}$. Therefore $\delta$ is divided into two arcs where along one $f_p$ is increasing and is decreasing along the other. This implies that planes perpendicular to the direction $p$ (and between the planes corresponding to $p_{\min}$ and $p_{\max}$) intersect the knot $\mathrm{Im}(\delta)$ in exactly two points. The union of straight line segments joining these pairs of points exhibit the knot $\mathrm{Im}(\delta)$ as the boundary of a disc which means $\delta$ is not knotted (see example **??** in chapter 1).

**Exercise 1.1.4** *Let $\delta : [0, L] \rightarrow \mathbb{R}^3$ be a simple closed curve, and $\gamma(s)$ denote the unit tangent vector field to $\delta$. Show that the curve $s \rightarrow \gamma(s)$ intersects every great circle on $S^2$ at least twice. Deduce that*

$$\int |\kappa| ds \geq 2\pi,$$

*where $\kappa$ denotes the curvature of $\delta$.*

For an extensive discussion of integral geometry and its applications see [S] and references thereof.

## 1.2 Riemannian Geometry

### 1.2.1 Basic Concepts

We introduce the fundamental concepts of Riemannian geometry by first looking at Euclidean space and its submanifolds, and determining which notions are dependent or independent of the embedding. This special case, besides being of intrinsic interest, will serve as a good example for the more abstract development of the general case. Let $M \subset U$ be a submanifold. To adapt the moving frame to $M$, we assume that $x$ ranges over $M$ and $e_1(x), \cdots, e_m(x)$ form an orthonormal basis for $\mathcal{T}_x M$. To simplify notation, we make the following convention on indices:

$$1 \le A, B, C, \cdots \le N, \quad 1 \le i, j, k, \cdots \le m, \quad m+1 \le a, b, p, q, \cdots \le N.$$

Since $x$ ranges over $M$, $\omega_p = 0$, and hence $dx = \sum_i \omega_i e_i$. This simply expresses the fact that $\mathcal{T}_x M$ is spanned by $e_1, \cdots, e_m$. In a more cumbersome language this can be rephrased as follows: If $f : M \to U$ is a submanifold, then $f^\star(\omega_p)$ vanishes identically. By writing $\omega_p = 0$ we emphasize the point of view that $M$ is regarded as the solutions to the Pfaffian system

$$\omega_{m+1} = 0, \ \cdots, \ \omega_N = 0.$$

The first set of structure equations becomes

$$d\omega_i + \sum_j \omega_{ij} \wedge \omega_j = 0, \quad \sum_i \omega_{pi} \wedge \omega_i = 0, \qquad \text{on } M. \tag{1.2.1}$$

A fundamental property of the $\omega_i$'s is that relative to the Riemannian metric induced on $M$, the metric has the form $ds^2 = \sum_i \omega_i^2$. This is essentially obvious since for any curve $\gamma : I \to M$, the element of arc length is $ds^2(\dot\gamma) = < dx(\dot\gamma), dx(\dot\gamma) >= \sum_i \omega_i(\dot\gamma)\omega_i(\dot\gamma)$, where $\dot\gamma$ is the tangent vector to the curve. In practice, $\omega_i$'s are often computed from the relation $ds^2 = \sum_i \omega_i^2$.

It is convenient to decompose the matrix $(\omega_{AB})$ in the form

$$\tilde\omega = \begin{pmatrix} (\omega_{ij}) & (\omega_{ip}) \\ (\omega_{pi}) & (\omega_{pq}) \end{pmatrix}.$$

The $m \times m$ matrix $\omega = (\omega_{ij})$ is called the *Levi-Civita connection* for the induced metric on $M \subset U$. Let us see how the connection $\omega$ transforms under a change of orthonormal frame. Let $A = (A_{ij})$ be an orthogonal matrix, and the frames $\{e_i\}$ and $\{f_i\}$ be related by the orthogonal transformation $e_j = \sum_i A_{ij} f_i$. Setting $f_p = e_p$ for $m+1 \leq p \leq N$, and denoting the connection form relative to the $f_A$'s by $\omega'$, we obtain after a simple calculation

$$\omega = A^{-1}\omega' A + A^{-1}dA \tag{1.2.2}$$

Notice that because of the additive factor $A^{-1}dA$, the connection $\omega$ is not a tensor but a collection of 1-forms transforming according to (1.2.2). As noted earlier, because of the dependence of $\omega$ on the choice of frame, its natural domain of definition is the principal bundle of orthonormal frames, however, we shall not dwell on this point. The matrix-valued function $A$ effecting a change of frames is generally called a *gauge transformation.* Since the entries of $A^{-1}dA$ contain a basis for left invariant 1-forms on the special orthogonal group, for every point $p \in M$ there is a gauge transformation $A$ defined in a neighborhood of $p$ such that $\omega'$ vanishes at $p \in M$. In general, one cannot force $\omega'$ to vanish in a neighborhood of $p \in M$.

Before giving the formal definition(s) of curvature, let us give some general motivation for the approach we are taking. In analogy with the definition of the curvature of a curve in the plane, it is reasonable to try to define the curvature of a hypersurface in $\mathbb{R}^{m+1}$, or more generally of submanifolds of Euclidean spaces, by taking exterior derivatives of the normal vectors $e_p$. We shall show below that the exterior derivative $de_p$ determines an $m \times m$ symmetric matrix $\mathsf{H}_p = (\mathsf{H}^p_{ij})$ for every direction $e_p$. The matrix $\mathsf{H}_p$ depends also on the choice of the frame $e_1, \cdots, e_m$ for the tangent spaces $\mathcal{T}_xM$ and therefore the individual components $\mathsf{H}^p_{ij}$ are not of geometric interest. However, the eigenvalues of $\mathsf{H}_p$ and their symmetric functions such as trace and determinant are independent of the choice of frames $e_1, \cdots, e_m$. Our first notions of curvature will be the trace and determinant of the matrices $\mathsf{H}_p$. For the case of surfaces $M \subset \mathbb{R}^3$, Gauss made the fundamental observation (*Theorema Egregium*) that $\det(\mathsf{H}_3)$ (there is only one normal direction $e_3$) is computable directly in terms of the coefficients of the metric tensor $ds^2$ which is only the necessary data for calculating lengths of curves on the surface $M$. Gauss' theorem was taken up by Riemann who founded Riemannian geometry on the basis of the tensor $ds^2$ thus completely freeing the notion (or more precisely some notions) of curvature from the embedding. To achieve this fundamental point of view, we make use of the fact, which is far from obvious without hindsight, that the structure equations $d\omega_{AB} + \sum \omega_{AC} \wedge \omega_{CB} = 0$ express *flatness* (vanishing of curvature which will be elaborated on below) of Euclidean spaces, and the 2-forms $d\omega_{ij} + \sum_k \omega_{ik} \wedge \omega_{kj}$ (recall $1 \leq i, j \leq m$) which quantify the deviation of structure equations from being valid on $M$, contain much of the information about the curvature of the submanifold $M \subset \mathbb{R}^N$. The 2-form $d\omega_{ij} + \sum_k \omega_{ik} \wedge \omega_{kj}$ reduces to $d\omega_{12}$ for surfaces in $\mathbb{R}^3$ and it will be demonstrated shortly that

$$d\omega_{12} = -\det(\mathsf{H}_3)\omega_1 \wedge \omega_2. \tag{1.2.3}$$

The point is that once a Riemannian metric is specified, one can calculate the quantities $\omega_i$ and $\omega_{ij}$ although they depend on the choice of frames for the tangent spaces $\mathcal{T}_x M$ (see subsection on Levi-Civita Connection below). Therefore (1.2.3) contains Theorema Egregium. It should be pointed out that $\mathrm{Tr}(\mathsf{H}_3)$ is not computable from the data $ds^2$ alone, and it contains siginifcant geometric information which will be discussed in this chapter. In view of these facts, any quantity which is expressible in terms of $\omega_i$'s and $\omega_{ij}$'s is called *intrinsic* to a Riemannian manifold $M$, and quantities which necessarily involve $\omega_p$'s or $\omega_{Ap}$'s are called *extrinsic* in the sense that they depend on the embedding. Our immediate goal in this subsection is to make mathematics out of these remarks and specialize them to the case of surfaces in $\mathbb{R}^3$. Various notions of curvature, based on the above comments, will be introduced in the following subsections. We begin with the following algebraic lemma:

**Lemma 1.2.1** (Cartan's Lemma) - *Let $v_1, \cdots, v_m$ be linearly independent vectors in a vector space $V$, and $w_1, \cdots, w_m$ be vectors such that*

$$v_1 \wedge w_1 + \cdots + v_m \wedge w_m = 0.$$

*Then $w_j = \sum \mathsf{H}_{ij} v_i$ with $\mathsf{H}_{ij} = \mathsf{H}_{ji}$. The converse is also true.*

**Proof** - Let $\{v_1, \cdots, v_m, \cdots, v_N\}$ be a basis for $V$, and set $w_j = \sum_i \mathsf{H}_{ij} v_i + \sum_p \mathsf{H}_{pj} v_p$. Then

$$\sum_{i=1}^m v_i \wedge w_i = \sum_{i,j=1}^m (\mathsf{H}_{ji} - \mathsf{H}_{ij}) v_i \wedge v_j + \sum_{i=1}^m \sum_{p=k+1}^N \mathsf{H}_{pi} v_i \wedge v_p.$$

Therefore $\mathsf{H}_{ij} = \mathsf{H}_{ji}$ and $\mathsf{H}_{pi} = \mathsf{H}_{ip}$. The converse statement is trivial. ♣

Applying Cartan's lemma to the second equation of (1.2.1), we can write

$$\omega_{ip} = \sum_j \mathsf{H}^p_{ij} \omega_j, \tag{1.2.4}$$

where $(\mathsf{H}^p_{ij})$ is a symmetric matrix. The *Second Fundamental Form* of the submanifold $M$ in the direction $e_p$ is the quadratic differential given by

$$\mathsf{H}_p = \sum_{i,j} \mathsf{H}^p_{ij} \omega_i \omega_j \tag{1.2.5}$$

This means that the value of $\mathsf{H}_p$ on a tangent vector $\xi \in \mathcal{T}_x M$ is $\sum_{i,j} \mathsf{H}^p_{ij} \omega_i(\xi) \omega_j(\xi)$. The reason for regarding $\mathsf{H}_p$ as a quadratic differential (i.e., a section of the second symmetric power of $\mathcal{T}^\star M$) is its transformation property which descibed below. (The *First Fundamental Form* is the metric $ds^2$.) Clearly $\mathsf{H}_p$ may also be regarded as the symmetric linear transformation, relative to the inner product induced from $\mathbb{R}^N$, of $\mathcal{T}_xM$ defined by the matrix $(\mathsf{H}^p_{ij})$ with respect to the basis $\{e_1, \cdots, e_m\}$. Note that there is a second fundamental form for every normal direction to $M$.

Let us see how the second fundamental form transforms once we make a change of frames. First assume that $e_{m+1}, \cdots, e_N$ are kept fixed but $e_1, \cdots, e_m$ are subjected a transformation $A \in O(m)$. From the transformation property of the matrix $(\omega_{AB})$ we obtain the transformation

$$\begin{pmatrix} \omega_{1p} \\ \vdots \\ \omega_{mp} \end{pmatrix} \longrightarrow A' \begin{pmatrix} \omega_{1p} \\ \vdots \\ \omega_{mp} \end{pmatrix}.$$

It follows that for fixed $e_{m+1}, \cdots, e_N$ the symmetric matrix $\mathsf{H}_p = (\mathsf{H}^p_{ij})$ transforms according

$$\mathsf{H}_p \longrightarrow A'\mathsf{H}_pA. \tag{1.2.6}$$

This transformation property justifies regarding the second fundamental form as a quadratic differential on $M$. Similarly, if we fix $e_1, \cdots, e_m$ and subject $e_{m+1}, \cdots, e_N$ to a transformation $A \in O(N-m)$, then the matrices $\mathsf{H}_p$ transform according as

$$\mathsf{H}_p \longrightarrow \sum_q A_{qp}\mathsf{H}_q. \tag{1.2.7}$$

While the matrix $(\mathsf{H}^p_{ij})$ depends on the choice of the orthonormal basis for $\mathcal{T}_xM$, the symmetric functions of its characteristic values depend only on the direction $e_p$ and not on the choice of basis for $\mathcal{T}_xM$. For example, the *mean curvature* in the direction $e_p$ defined by $\mathsf{H}_p = \frac{1}{m}\mathrm{trace}(\mathsf{H}^p_{ij}) = \sum_i \mathsf{H}^p_{ii}$ expresses a geometric property of the manifold $M \subset \mathbb{R}^N$ which we will discuss later especially in the codimension one case for surfaces. For a hypersurface $M \subset \mathbb{R}^{m+1}$, there is only one normal direction and we define the *Gauss-Kronecker curvature* at $x \in M$ as $K(x) = (-1)^{m+1}\det(\mathsf{H}_{ij})$ (in case $m = 2$ one simply refers to $K$ as *curvature*). The eigenvalues of $\mathsf{H}$ are called the *principal curvatures* and are often denoted as $\kappa_1 = \frac{1}{R_1}, \cdots, \kappa_m = \frac{1}{R_m}$. If the eigenvalues of $\mathsf{H}$ are distinct, then (locally) we have $m$ orthonormal vector fields on $M$ diagonalizing the second fundamental form. The directions determined by these vector fields are called the *principal directions*, and an integral curve for such a vector field is called a *line of curvature*. Note that in the case of hypersurfaces the second fundamental form can also be written in the form

$$\mathsf{H} = - < dx, de_{m+1} > . \tag{1.2.8}$$

**Example 1.2.1** Consider the sphere $S^n_r \subset \mathbb{R}^{n+1}$ of radius $r > 0$. Taking orthonormal frames as prescribed above, we obtain $x = re_{n+1}$, and consequently $\omega_{in+1} = \frac{1}{r}\omega_i$, $\mathsf{H}_{ij} = -\frac{\delta_{ij}}{r}\omega_i$, and $\Pi = -\frac{1}{r}\sum_i \omega_i^2$. Therefore the Gauss-Kronecker curvature of $S^n_r$ is $K(x) = \frac{1}{r^n}$. ♠

**Example 1.2.2** A simple case of a submanifold of codimension one is that of a surface $M \subset \mathbb{R}^3$. In this case the Levi-Civita connection is the matrix

$$\omega = \begin{pmatrix} 0 & \omega_{12} \\ -\omega_{12} & 0 \end{pmatrix}$$

The symmetric matrix $(\mathsf{H}_{ij})$ in the definition of second fundamental form is defined by

$$\omega_{13} = \mathsf{H}_{11}\omega_1 + \mathsf{H}_{12}\omega_2, \quad \omega_{23} = \mathsf{H}_{12}\omega_1 + \mathsf{H}_{22}\omega_2.$$

Therefore

$$d\omega_{12} = -\omega_{13} \wedge \omega_{32} = (\mathsf{H}_{11}\mathsf{H}_{22} - \mathsf{H}_{12}^2)\, \omega_1 \wedge \omega_2. \tag{1.2.9}$$

Therefore the measure of the deviation of the quantity $d\omega_{ij} + \sum \omega_{ik} \wedge \omega_{kj}$ from vanishing, which we had alluded to earlier, is the curvature $K$. It should be emphasized that the second fundamental form was obtained by restricting $\omega_p$ to $M$ and therefore (1.2.9) is valid as an equation on $M$. Note that we have arrived at the curvature $K$ of the surface via two different routes. The intrinsic approach where it is defined by $d\omega_{12} = K\omega_1 \wedge \omega_2$ (or the deviation of $d\omega_{12}$ from vanishing), and the extrinsic approach as the determinant of the matrix $H$ of the second fundamental form. ♠

We have emphasized that the 1-form $\omega_{12}$ depends on the choice of the frame and therefore is naturally defined on the bundle of frames $\mathcal{P}M$. By fixing a frame (locally) we can express $\omega_{12}$ as a 1-form on $M$[1]. We can use this fact to advantage and deduce interesting geometric information as demonstrated in the following example:

**Example 1.2.3** Consider a compact surface $M \subset \mathbb{R}^3$ without boundary and assume that $\xi$ is nowhere vanishing vector field on $M$. From $\xi$ we obtain a unit tangent vector field $e_1$ globally defined on $S^2$ and let $e_2$ be the unit tangent vector field to $M$ such that $e_1, e_2$ is a positively oriented orthonormal frame. Let $\omega_{12}$ be the Levi-Civita connection expressed relative to the moving frame $e_1, e_2$ which is a 1-form on $M$. Since $\partial M = \emptyset$, Stokes' theorem implies

$$\int_M d\omega_{12} = 0.$$

[1] In more sophisticated language, the frame $e_1, e_2$ is a global section of the bundle of frames $\mathcal{P}M$ and $\omega_{12}$, which is naturally defined on it, is pulled back to $M$ by this section.

On the other hand, $d\omega_{12} = K\omega_1 \wedge \omega_2$, and therefore

$$\int_M K\omega_1 \wedge \omega_2 = 0. \tag{1.2.10}$$

If we let $M = S^2$ be a sphere, then $K$ is a positive constant and therefore (1.2.10) cannot hold. Therefore $S^2$ does not admit of a nowhere vanishing vector field $\xi$. On the other hand, it is easy to see that the torus $T^2$ admits of a nowhere vanishing vector field, and therefore no matter what embedding of $T^2$ in $\mathbb{R}^3$ we consider, still relation (1.2.10) remains valid. We shall return to this issue in the next chapter. ♠

The intrinsic description of the Gauss-Kronecker curvature $K$ via the formula $d\omega_{12} = K\omega_1 \wedge \omega_2$ reduces the computation of $K$ to straightforward algebra once the metric $ds^2$ is explicitly given. In fact, we have

**Exercise 1.2.1** (a) - *Let $ds^2 = P^2(u,v)du^2 + Q^2(u,v)dv^2$. Show that the connection and curvature are given by*

$$\omega_{12} = \frac{1}{Q}\frac{\partial P}{\partial v}du - \frac{1}{P}\frac{\partial Q}{\partial u}dv, \quad K = -\frac{1}{PQ}\{\frac{\partial}{\partial v}(\frac{1}{Q}\frac{\partial P}{\partial v}) + \frac{\partial}{\partial u}(\frac{1}{P}\frac{\partial Q}{\partial u})\}.$$

(b) - *Let $M \subset \mathbb{R}^3$ be a surface, and $L$ be a line of curvature on $M$. Show that the surface formed by the normals to $M$ along $L$ has zero curvature.*

In view of the above considerations it is reasonable to define the *curvature matrix* $\Omega = (\Omega_{ij})$ *of a submanifold* $M \subset \mathbb{R}^N$ as

$$\Omega_{ij} = d\omega_{ij} + \sum_{k=1}^{m} \omega_{ik} \wedge \omega_{kj}.$$

For hypersurface $M \subset \mathbb{R}^{m+1}$, the curvature matrix $\Omega$ is then related to the second fundamental form by the important relation

$$\Omega_{ij} = -\omega_{im+1} \wedge \omega_{m+1j}, \tag{1.2.11}$$

for a $M \subset \mathbb{R}^3$. This formula follows immediately from the structure equations and the definition of $\Omega_{ij}$. The definition of the curvature matrix will be extended and discussed in the following subsections.

Related to (1.2.11) is the concept of Gauss mapping which will be used exensively. Let $M \subset \mathbb{R}^{m+1}$ be a hypersurface and consider the mapping $\mathsf{G} : M \to S^m$ given by $\mathsf{G}(x) = e_{m+1}(x)$ called the *Gauss mapping.* Since $de_{m+1} = \sum \omega_{im+1} e_i$, we easily obtain

$$\mathsf{G}^\star(dv_{S^m}) = \omega_{1m+1} \wedge \cdots \wedge \omega_{mm+1} = (-1)^m \det(\mathsf{H})\omega_1 \wedge \cdots \wedge \omega_m. \tag{1.2.12}$$

In following two examples the concepts of metric, curvature etc. are related to their classical (and maybe more familiar) form for surfaces:

**Example 1.2.4** Assume a surface $M \subset \mathbb{R}^3$ is described parametrically by a map $f$ from the $(u, v)$-plane to the $(x, y, z)$-space, i.e., $M$ is given by $(x(u, v), y(u, v), z(u, v))$, then from the prescription in calculus texts for the computation of the arc-length it is evident that the metric is given by the symmetric positive definite matrix

$$(Df)'Df = \begin{pmatrix} (\frac{\partial x}{\partial u})^2 + (\frac{\partial y}{\partial u})^2 + (\frac{\partial z}{\partial u})^2 & \frac{\partial x}{\partial u}\frac{\partial x}{\partial v} + \frac{\partial y}{\partial u}\frac{\partial y}{\partial v} + \frac{\partial z}{\partial u}\frac{\partial z}{\partial v} \\ \frac{\partial x}{\partial u}\frac{\partial x}{\partial v} + \frac{\partial y}{\partial u}\frac{\partial y}{\partial v} + \frac{\partial z}{\partial u}\frac{\partial z}{\partial v} & (\frac{\partial x}{\partial v})^2 + (\frac{\partial y}{\partial v})^2 + (\frac{\partial z}{\partial v})^2 \end{pmatrix}$$

where $Df$ is the derivative of $f$ and superscript $'$ denotes the transposed matrix. It is customary to set

$$E = (\frac{\partial x}{\partial u})^2 + (\frac{\partial y}{\partial u})^2 + (\frac{\partial z}{\partial u})^2;\ F = \frac{\partial x}{\partial u}\frac{\partial x}{\partial v} + \frac{\partial y}{\partial u}\frac{\partial y}{\partial v} + \frac{\partial z}{\partial u}\frac{\partial z}{\partial v};\ G = (\frac{\partial x}{\partial v})^2 + (\frac{\partial y}{\partial v})^2 + (\frac{\partial z}{\partial v})^2.$$

so that the metric becomes $ds^2 = Edu^2 + 2Fdudv + Gdv^2$. The unit normal to the surface $M \subset \mathbb{R}^3$ is the vector

$$e_3 = (\frac{1}{J}\det\begin{pmatrix} \frac{\partial y}{\partial u} & \frac{\partial z}{\partial u} \\ \frac{\partial y}{\partial v} & \frac{\partial z}{\partial v} \end{pmatrix}, \frac{1}{J}\det\begin{pmatrix} \frac{\partial z}{\partial u} & \frac{\partial x}{\partial u} \\ \frac{\partial z}{\partial v} & \frac{\partial x}{\partial v} \end{pmatrix}, \frac{1}{J}\det\begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{pmatrix}).$$

where $J = \sqrt{EG - F^2}$. Denoting the components of $e_3$ by $\xi, \eta,$ and $\zeta$ respectively, we obtain

$$L = -\frac{\partial \xi}{\partial u}\frac{\partial x}{\partial u} - \frac{\partial \eta}{\partial u}\frac{\partial y}{\partial u} - \frac{\partial \zeta}{\partial u}\frac{\partial z}{\partial u},\ M = -\frac{\partial \xi}{\partial u}\frac{\partial x}{\partial v} - \frac{\partial \eta}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial \zeta}{\partial u}\frac{\partial z}{\partial v},\ N = -\frac{\partial \xi}{\partial v}\frac{\partial x}{\partial v} - \frac{\partial \eta}{\partial v}\frac{\partial y}{\partial v} - \frac{\partial \zeta}{\partial v}\frac{\partial z}{\partial v},$$

where the expression $Ldu^2 + 2Mdudv + Ndv^2$ is the second fundamental form. ♠

**Exercise 1.2.2** *Show that if a surface is given as $z = z(x, y)$, then the coefficients of the first and second fundamental form are*

$$\begin{array}{ccc} E = 1 + p^2, & F = pq, & G = 1 + q^2 \\ L = \frac{r}{\sqrt{1+p^2+q^2}}, & M = \frac{s}{\sqrt{1+p^2+q^2}}, & N = \frac{t}{\sqrt{1+p^2+q^2}} \end{array}$$

*where* $p = \frac{\partial z}{\partial x}$, $q = \frac{\partial z}{\partial y}$, $r = \frac{\partial^2 z}{\partial x^2}$, $s = \frac{\partial^2 z}{\partial x \partial y}$, and $t = \frac{\partial^2 z}{\partial y^2}$. *The Gauss-Kronecker curvature* $K$ *and the mean curvature* $H$ *are*

$$K = \frac{rt - s^2}{(1 + p^2 + q^2)^2}, \quad H = \frac{(1 + q^2)r - 2pqs + (1 + p^2)t}{2(1 + p^2 + q^2)^{3/2}}.$$

*Deduce that the paraboloid* $z = x^2 - y^2$ *and the hyperboloid* $x^2 + y^2 - z^2 = 1$ *have everywhere negative curvature, however, the curvature tends to zero as one moves to infinity.* (This is typical in the sense that there are no complete surfaces in $\mathbb{R}^3$ with curvature bounded above by a negative constant.)

**Example 1.2.5** Let $\Gamma$ be a curve in the first quadrant of $(x, z)$-plane described by $(x(t), z(t))$ where $t$ is the arc length. Rotating $\Gamma$ around the $z$-axis generates a surface of revolution $M \subset \mathbb{R}^3$. The Riemannian metric on $M$ is $dt^2 + x(t)^2 d\theta^2$ relative to the $(t, \theta)$ coordinates where $\theta$ is the angle of rotation. Therefore we set $\omega_1 = dt$ and $\omega_2 = x d\theta$. The Levi-Civita connection and Gaussian curvature of $M$ are easily computed to obtain (see exercise 1.2.1):

$$\omega_{12} = -\frac{dx}{dt} d\theta, \qquad K = -\frac{1}{x}\frac{d^2 x}{dt^2}.$$

Thus the curvature of $M$ depends only on the variable $t$ which reflects its invariance under rotations around the $z$-axis. Now if we specify any function of one variable $K(t)$, we can solve the ordinary differential equation $\frac{1}{x}\frac{d^2x}{dt^2} = -K$ locally. To make sure that this is the Gaussian curvature of a surface of revolution we have to demonstrate the existence of a function $z(t)$ such that

$$\left(\frac{dx}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2 = 1, \tag{1.2.13}$$

so that $t$ becomes arc length along the curve $(x(t), z(t))$. Locally we can always accomplish this by making sure that $|\frac{dx}{dt}| < 1$, so that in a neighborhood of the initial point we can set $\frac{dz}{dt} = \sqrt{1 - \left(\frac{dx}{dt}\right)^2}$ and solve for $z(t)$ to obtain the desired curve. Thus for an arbitrarily function of one variable we can construct a surface of revolution with the given function as the Gaussian curvature. In particular, if we set $K$ equal to a positive constant then we obtain spheres as surfaces of revolution of constant Gaussian curvature $K$. For $K$ a negative constant, we can still obtain a local solution in terms of hyperbolic functions. But this solution cannot be continued to exist for all $t$ since the relation (1.2.13) will be violated. This is no accident and will be elaborated on later. ♠

**Example 1.2.6** Roughly speaking, any differentiable function of one variable is the mean curvature function of a surface of revolution locally. To make this statement more precise, let

$\Gamma$ be the graph of a function $y = f(x)$ of one variable and assume $\Gamma$ lies in the first quadrant of the $(x, y)$-plane. The the portion of the surface $M$, lying the half plane $z > 0$, obtained by rotating $\Gamma$ around the $x$-axis is the graph of the function $z = \sqrt{f(x)^2 - y^2}$. Exercise 1.2.2 provides us with a formula for the mean curvature of $M$. The formula involves computing second partial derivatives of the function $z$ with respect to $x$ and $y$. Since the the surface $M$ is invariant under rotations in the $(y, z)$-plane, we can set $y = 0$ in the expression for the mean curvature of a surface given in exercise 1.2.2. In fact, we obtain after a straightforward calculation

$$f'' = \frac{1 + f'^2}{f} - 2H\left(1 + f'^2\right)^{\frac{3}{2}}, \tag{1.2.14}$$

where the mean curvature $H$ is a function of $x$ (and the sign of $H$ depends of the direction of the unit normal $e_3$) only. In principle, this differential equation can be solved for any differentiable function $H$ to obtain a surface of revolution with prescribed mean curvature function $H(x)$. ♠

**Example 1.2.7** Let $M \subset \mathbb{R}^3$ be a surface. A point $x \in M$ is called an *umbilical point* if the principal curvatures $\kappa_1$ and $\kappa_2$ are equal at $x$. Clearly, every point of the sphere $S^2 \subset \mathbb{R}^3$ is an umbilical point. In this example, we show that there are no umbilical points on the (standard) torus. Let $0 < r < 1$ and consider the circle

$$\Gamma : \quad x = 1 + r\cos\theta, \quad y = r\sin\theta,$$

in the $xy$-plane. Rotating $\Gamma$ around the $y$-axis, we obtain the torus $M$ given parametrically as

$$(\theta, \phi) \longrightarrow ((1 + r\cos\theta)\cos\phi, r\sin\theta, (1 + r\cos\theta)\sin\phi).$$

The unit tangent vectors

$$e_1 = (-\sin\theta\cos\phi, \cos\theta, -\sin\theta\sin\phi), \quad e_2 = (-\sin\phi, 0, \cos\phi),$$

give a trivialization of the tangent bundle $\mathcal{T}M$ of $M$. The corresponding basis of 1-forms are

$$\omega_1 = rd\theta, \quad \omega_2 = (1 + r\cos\theta)d\phi$$

The unit normal is $e_3 = (\cos\theta\cos\phi, \sin\theta, \cos\theta\sin\phi)$, and so $de_3 = (d\theta)e_1 + (\cos\theta d\phi)e_2$. Consequently,

$$\omega_{13} = \frac{1}{r}\omega_1, \quad \omega_{23} = \frac{\cos\theta}{1 + r\cos\theta}\omega_2.$$

Therefore the second fundamental form is diagonalized and the principal curvatures are

$$\kappa_1 = \frac{1}{r}, \quad \kappa_2 = \frac{\cos\theta}{1 + r\cos\theta}.$$

Clearly $\kappa_1 = \kappa_2$ is not possible and there are no umbilical points on the torus. In chapter 6, we will show that, for topological reasons, the only compact surfaces in $\mathbb{R}^3$ with no umbilical points are tori.

**Example 1.2.8** Let $M \subset \mathbb{R}^3$ be a surface all whose points are umbilics. This means on $M$ we have

$$\omega_{31} = a\omega_1, \quad \omega_{32} = a\omega_2,$$

for a function $a$ on $M$. It follows that

$$d\omega_{13} = da \wedge \omega_1 - a\omega_{12} \wedge \omega_2.$$

Comparing with $d\omega_{13} = -\omega_{12} \wedge \omega_{23} = -a\omega_{12} \wedge \omega_2$, we obtain

$$da \wedge \omega_1 = 0, \quad da \wedge \omega_2 = 0,$$

where the second identity is obtained by a similar argument. Therefore $a$ is a constant, and $M$ is a subset of a sphere. To prove the latter assertion, note that from $de_3 = \omega_{13}e_1 + \omega_{23}e_2$ it follows that

$$de_3 = -a(\omega_1 e_1 + \omega_2 e_2) = -adp = d(-ap)$$

where $p$ denotes a generic point on $M$. Therefore $d(e_3 + ap) = 0$ and after a translation we can assume $e_3 = -ap$. Thus if $M$ is defined by an equation $F(x_1, x_2, x_3) = 0$, then we have

$$\frac{\partial F}{\partial x_1} = \rho x_1, \quad \frac{\partial F}{\partial x_2} = \rho x_2, \quad \frac{\partial F}{\partial x_3} = \rho x_3,$$

for some function $\rho$. Computing $\frac{\partial^2 F}{\partial x_i \partial x_j}$ from the above equations we obtain the system of linear equations

$$x_1\frac{\partial \rho}{\partial x_2} - x_2\frac{\partial \rho}{\partial x_1} = 0,\ x_2\frac{\partial \rho}{\partial x_3} - x_3\frac{\partial \rho}{\partial x_2} = 0,\ x_3\frac{\partial \rho}{\partial x_1} - x_1\frac{\partial \rho}{\partial x_3} = 0.$$

Therefore $\frac{\partial \rho}{\partial x_i} = 0$ and $\rho$ is a constant. We easily integrate to obtain

$$F(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - c$$

proving that $M$ is a subset of a sphere. In particular, a compact surface $M \subset \mathbb{R}^3$ all whose points are umbilics is necessarily a sphere. ♠

**Remark 1.2.1** Let $M \subset \mathbb{R}^3$. We want to derive (a little heuristically) the analogue of the Euler-Lagrange equations for the critical points of the variation of the element of area for $M$ in the language of moving frames. Taking exterior derivative of the element of area $\omega_1 \wedge \omega_2$ we obtain

$$d(\omega_1 \wedge \omega_2) = (\omega_{13} \wedge \omega_2 - \omega_{23} \wedge \omega_1) \wedge \omega_3 = (-(A_{11} + A_{22})\omega_1 \wedge \omega_2) \wedge \omega_3,$$

where $A = (A_{ij})$ is the matrix of the second fundamental form. In order for the variation of the element of area to be critical, the variation in the normal direction $e_3$, or the coefficient of $\omega_3$, should vanish. Therefore vanishing of the mean curvature $\frac{1}{2}(\mathsf{H}_{11} + \mathsf{H}_{22})$ is the Euler-Lagrange equations for the element of area. For this reason, surfaces with vanishing mean curvature are called *minimal surfaces.* More generally, surfaces in $\mathbb{R}^N$ for which the mean curvature $\mathsf{H}_p$ vanishes for every normal direction $e_p$, are also called minimal surfaces and by a similar argument the terminology can be justified.

**Exercise 1.2.3** *With the hypothesis and notation of exercise 1.2.2 show that the element of area for the surface $z = z(x, y)$ is given by*

$$\sqrt{1 + p^2 + q^2}dx \wedge dy.$$

*Applying the Euler-Lagrange equation from the Calculus of Variations, deduce that the critical points for the area of surfaces of a given boundary satisfy $(1 + q^2)r - 2pqs + (1 + p^2)t = 0$, i.e., mean curvature should vanish.* (Compare with the preceding remark.)

The following exercise shows that one can obtain a solution to the minimal surface equation in $\mathbb{R}^3$ by separation of variables.

**Exercise 1.2.4** *Substituting $z = f(x) + h(y)$ in the minimal surface equation $(1 + q^2)r - 2pqs + (1 + p^2)t = 0$, show that it reduces to two ordinary differential equations*

$$\frac{f''}{1 + f'^2} = a = -\frac{h''}{1 + h'^2},$$

*where $a$ is a constant and superscript $'$ denotes differentiation. For $a \neq 0$ derive the solution*

$$z = d + \frac{1}{a}[\log\cos(ax + b) \ - \ \log\cos(ay + c)],$$

*where $a \neq 0, b, c$ and $d$ are arbirary constants, and the domains of $x$ and $y$ are appropriately restricted.* (This surface is called *Scherk surface.* Surfaces representable as $x(u, v) = f_1(u) + \mathsf{H}_1(v)$, $y(u, v) = f_2(u) + \mathsf{H}_2(v)$, $z(u, v) = f_3(u) + \mathsf{H}_3(v)$, are called *surfaces of translation.*)

**Exercise 1.2.5** *Consider a surface $M \subset \mathbb{R}^3$, and let $e_1, e_2, e_3$ be a moving frame with $e_3$ is normal to the surface $M$. Writing a generic point of the tube $\tau_r(M)$ of radius $r > 0$ ($r$ small) as $q = p + te_3$ with $p \in M$ and $|t| < r$, show that the volume element on $\tau_r(M)$ is*

$$dv_{\tau_r(M)} = (\omega_1 + t\omega_{13}) \wedge (\omega_2 + t\omega_{23}) \wedge dt.$$

*Deduce that for $r > 0$ small*

$$\mathrm{vol}(\tau_r(M)) = 2r\mathrm{vol}(M) + \frac{2}{3}r^3 \int_M K\omega_1 \wedge \omega_2.$$

### 1.2.2 Levi-Civita Connection

The Levi-Civita connection $(\omega_{ij})$ for a submanifold $M \subset \mathbb{R}^N$ is an anti-symmetric matrix with the property $d\omega_i + \sum \omega_{ij} \wedge \omega_j = 0$ (1.2.1). In general, for 1-forms $\theta_1, \cdots, \theta_m$ spanning the cotangent spaces to $M$, we can only assert the existence of a matrix of 1-forms $(\theta_{ij})$ such that $d\theta_i + \sum \theta_{ij} \wedge \theta_j = 0$. A remarkable consequence of an inner product on $\mathbb{R}^N$ was that if we set $\theta_i = \omega_i$ then the matrix $(\theta_{ij})$ can be replaced by the anti-symmetric matrix $(\omega_{ij})$, i.e., a matrix of 1-forms taking values in the Lie algebra of $SO(m)$. The following proposition shows that the existence of a Riemanniann metric on $M$ (and not an embedding) is all that is needed to ensure the existence and uniqueness of the matrix $(\omega_{ij})$ with the required properties[2]:

**Proposition 1.2.1** *Let $\omega_1, \cdots, \omega_m$ be a basis of one forms reducing the Riemannian metric to the identity matrix, i.e., $ds^2 = \sum_i \omega_i^2$. Then there is a unique skew-symmetric matrix $\omega = (\omega_{ij})$* (called the *Levi-Civita connection* for the given Riemannian metric) *such that*

$$d\omega_i + \sum_j \omega_{ij} \wedge \omega_j = 0.$$

**Proof** - We have

$$d\omega_i = \sum_{j,k} a_{ijk}\omega_j \wedge \omega_k,$$

where the coefficients $a_{ijk}$ satisfy the anti-symmetry condition

$$a_{ijk} + a_{ikj} = 0.$$

[2]The remarkable property of the Levi-Civita connection becomes more evident when one studies geometric structures corresponding subgroups other than the orthogonal groups.

Since $a_{jki} + a_{kji}$ is symmetric in the indices $(j,k)$, we have

$$\sum_{j,k} \omega_j \wedge (a_{jki} + a_{kji})\omega_k = 0,$$

and consequently

$$d\omega_i = \sum_{j,k} \omega_j \wedge (a_{ijk} + a_{jki} + a_{kji})\omega_k.$$

Now set

$$\omega_{ij} = \sum_k (a_{ijk} + a_{jki} + a_{kji})\omega_k$$

which satisfies the requirements of the proposition. To prove uniqueness, let $(\omega'_{ij})$ be another such matrix, and set $\theta_{ij} = \omega_{ij} - \omega'_{ij}$. Applying Cartan's lemma to $\sum \omega_i \wedge \theta_{ij} = 0$, we obtain

$$\theta_{ij} = \sum_k b_{kij}\omega_k, \qquad b_{kij} = b_{ikj}.$$

On the other hand by anti-symmetry of $\theta_{ij}$, we have $b_{kij} = -b_{kji}$. It follows easily that $b_{ijl} = 0$ thus completing the proof of the proposition. ♣

**Exercise 1.2.6** *For the metric $ds^2$ in the diagonal form $ds^2 = \sum_i g_{ii} dx_i^2$, show that the Levi-Civita connection is given by*

$$\omega_{ij} = \frac{1}{\sqrt{g_{jj}}} \frac{\partial \log \sqrt{g_{ii}}}{\partial x_j}\omega_i - \frac{1}{\sqrt{g_{ii}}} \frac{\partial \log \sqrt{g_{jj}}}{\partial x_i}\omega_j,$$

*where $\omega_i = \sqrt{g_{ii}}dx_i$.*

The connection $\omega$ enables us to differentiate vector fields. More precisely, let $e_1, \cdots, e_m$ be an orthonormal frame on the Riemannian manifold $M$, and $(\omega_{ij})$ be the Levi-Civita connection for the Riemannian metric $g$. Define

$$\nabla e_i = \sum_j \omega_{ji} e_j, \tag{1.2.15}$$

and we extend $\nabla$ to a vector field $\xi = \sum_i b_i e_i$ by

$$\nabla \sum_i b_i e_i = \sum_{ij} \omega_{ji} b_i e_j + \sum_i db_i e_i. \tag{1.2.16}$$

The quantity $\nabla\xi$ is called the *covariant derivative* of the vector field $\xi$. For a vector field $\eta$

$$\nabla_\eta \xi = \sum_{ij} \omega_{ji}(\eta) b_i e_j + \sum_i db_i(\eta) e_i,$$

is the *covariant derivative of* $\xi$ *in the direction* $\eta$. It is not difficult to verify that $\nabla_\eta\xi$ and $\nabla\xi$ are independent of the choice of orthonormal frame $e_1, \cdots, e_m$.

Another very useful operation on tensor is contraction. For every pair $(i,j)$, with $1 \leq i \leq m$ and $1 \leq j \leq n$ the *contraction* operator

$$C_{ij} : \underbrace{V \otimes \cdots \otimes V}_{m \text{ times}} \otimes \underbrace{V^\star \otimes \cdots \otimes V^\star}_{n \text{ times}} \longrightarrow \underbrace{V \otimes \cdots \otimes V}_{m-1 \text{ times}} \otimes \underbrace{V^\star \otimes \cdots \otimes V^\star}_{n-1 \text{ times}}$$

is defined by

$$C_{ij}(v_1 \otimes \cdots \otimes v_m \otimes \xi_1 \otimes \cdots \otimes \xi_n) = \xi_j(v_i) v_1 \otimes \cdots \otimes \hat{v}_i \otimes \cdots \otimes v_m \otimes \xi_1 \otimes \cdots \otimes \hat{\xi}_j \otimes \cdots \otimes \xi_n,$$

where $\hat{v}_i$ means $v_i$ is omitted.

We now can extend covariant differentiation to a derivation on the space of tensors by the requirements

1. $\nabla f = df$ for a smooth function $f$;

2. $\nabla$ commutes with contractions.

An immediate consequence is

$$0 = dg(e_i, e_j) = \nabla(g)(e_i, e_j) + g(\nabla(e_i), e_j) + g(e_i, \nabla(e_j)) = \nabla(g)(e_i, e_j) + \omega_{ji} + \omega_{ij} = \nabla(g)(e_i, e_j).$$

Therefore

$$\nabla g = 0, \quad \text{or equivalently} \quad dg(\xi, \zeta)(\eta) = g(\nabla_\eta \xi, \zeta) + g(\xi, \nabla_\eta \zeta). \tag{1.2.17}$$

This equation expresses a fundamental property of the Levi-Civita connection.

**Remark 1.2.2** We have followed the mathematical tradition of only considering Riemannian rather than *indefinite* metrics by which we mean the condition of positive definiteness of the symmetric matrix $g = (g_{ij})$ is replaced by that of nondegeneracy. We shall see in subsections on spaces of constant curvature and homogeneous spaces that indefinite metrics, besides being of intrinsic interest in physics, are useful in understanding the behavior of Riemannian metrics. For an indefinite metric $ds^2$ with $r$ positive and $m-r$ negative eigenvalues we consider frames (also call them orthonormal) with the property

$$ds^2(e_i, e_j) = \pm\delta_{ij},$$

where $+$ or $-$ sign is chosen according as $i \leq r$ or $r+1 \leq i \leq m$. The definition of Levi-Civita connection $\omega_{ij}$ is the same except that instead of skew symmetry we require $(\omega_{ij})J + J(\omega_{ij}) = 0$ where $J$ is the diagonal matrix whose first $r$ diagonal entries are 1, and the remaining diagonal entries are $-1$. In other words $(\omega_{ij})$ takes values in the Lie algebra of the orthogonal group of $J$. The existence and uniqueness of the Levi-Civita connection is the same as in the Riemannian case. ♡

### 1.2.3 Parallel Translation and the Gauss-Bonnet Theorem

The notion of parallel translation plays a fundamental role in differential geometry. Let $M$ be a Riemannian manifold with the Levi-Civita connection $(\omega_{ij})$, $\gamma : I \rightarrow M$ a curve in $M$ and assume that $\gamma$ is parametrized by its arc-length, so that the tangent vectors $\dot{\gamma}(t) = D\gamma(t)(1)$ have length 1. Consider the system of ordinary differential equations

$$\gamma^\star(\omega_{ij})(t)(1) = \omega_{ij}(\dot{\gamma}(t)) = 0. \tag{1.2.18}$$

What this system specifies is how the frame $\{e_1, \cdots, e_m\}$ should be chosen so that the connection form $(\omega_{ij})$ vanishes along $\gamma$ when evaluated on the tangents to $\gamma$. Recall that by a gauge transformation we can make the Levi-Civita connection vanish at one point. The differential equations of parallel translation describe a frame along a curve relative to which the Levi-Civita connection vanishes when evaluated on the tangent field to the curve. This is really the best one can do in general to simplify the expression for the Levi-Civita connection. The precise geometric meaning of parallel translation in the context of surfaces in $\mathbb{R}^3$ and its relationship to parallel translation in the Euclidean plane is explained in the subsection Flat Surfaces and Parallel Translation below. This condition is independent of the parametrization of $\gamma$. Once the frame is specified at a point, say $x = \gamma(0) \in M$, then the ordinary differential equations (1.2.18) completely determine it along the curve $\gamma$. Let the frame $\{e_1, \cdots, e_m\}$ be so determined along $\gamma$. Let $f_1, \cdots, f_m$ be another frame which differs from $e_1, \cdots, e_m$ by a gauge transformation $A$. From the transformation formula (1.2.2) it follows that $f_1, \cdots, f_m$ is parallel along $\gamma$ if and only if $A^{-1}dA(\dot{\gamma})$ vanishes. This means that the gauge transformation $A$ is constant along $\gamma$. A vector field $\xi = \sum_i \xi_i e_i$ along $\gamma$ is *parallel* if the coefficients $\xi_i$ are constants along $\gamma$ where we are assuming that $e_1, \cdots, e_m$ is parallel along $\gamma$. This condition is equivalent to $\nabla_{\dot{\gamma}}\xi = 0$. We say a curve $\gamma$ is a *geodesic* if $\nabla_{\dot{\gamma}}\dot{\gamma} = 0$. The condition of geodesy implies that $g(\dot{\gamma}, \dot{\gamma})$ is constant along $\gamma$ which implies that a geodesic is necessarily parametrized by a multiple of arc-length (with arbitrary initial point).

For a surface $M$, we can give a simple geometric interpretation to $\omega_{12}(\dot{\gamma}(t))$. Let $\gamma$ be a curve parametrized by its arc length $s$, and assume the frame $e_1, e_2$ is such that $e_1$ is tangent to $\gamma$. We want to choose a new frame $f_1, f_2$ such that if $\omega'_{12}$ denotes the connection form relative to $f_1, f_2$, then $\omega'_{12}(\dot\gamma(s)) = 0$. Let $\omega_{12}$ denote the connection form relative to $e_1, e_2$. Then along $\gamma$ we can write $\omega_{12} = \alpha(s)ds$, i.e., $\omega_{12}(\dot\gamma(s)) = \alpha(s)$. Denote by $\phi$ the angle between the tangent to $\gamma$ and the vector $f_1$. Then from (1.2.2) and the requirement on the frame $f$ we see that $d\phi - \alpha(s)ds = 0$. Therefore the rate $d\phi/ds$ of the rotation of the angle $\phi$ is given by $\omega_{12}(\dot\gamma(s))$, and we have

$$\int_\gamma \omega_{12}(\dot\gamma(t)) = \int_\gamma d\phi. \tag{1.2.19}$$

The right hand side of (1.2.19) is simply the total angle through which the tangent vector to $\gamma$ rotates through relative to $f_1$ as one traverses the curve. This interpretation has significant implications as will be shown momentarily.

As an application of the concept of parallel translation, we discuss the Gauss-Bonnet theorem for surfaces[3]. Let $M$ be a surface with a Riemannian metric $ds^2$, and $\Delta \subset M$ an open relatively compact subset with connected boundary consisting of finitely many smooth curves. We think of $\Delta$ as a polygon but the the boundary curves need not be geodesic segments. Let $\gamma_1, \cdots, \gamma_n$ be the smooth boundary curves of $\Delta$ ordered according to their indices so that $\gamma_i$ intersects $\gamma_{i+1}$ at the vertex $v_i$ of $\Delta$ and $\gamma_n$ intersects $\gamma_1$ at the vertex $v_n$. Denote the exterior angle between $\gamma_i$ and $\gamma_{i+1}$ by $\alpha_i$ (the exterior angle between $\gamma_n$ and $\gamma_1$ is denoted by $\alpha_n$.) Applying Stokes' theorem and using $d\omega_{12} + K\omega_1 \wedge \omega_2 = 0$ we obtain

$$\int_\Delta K\omega_1 \wedge \omega_2 = -\int_{\partial\Delta} \omega_{12} \tag{1.2.20}$$

To correctly understand the meaning of the right hand side of (1.2.20) let us assume that on $\partial\Delta$ and in the complement of its vertices $e_1$ is tangent to the boundary curve and $e_2$ points to the interior of $\Delta$ (see also remark 1.2.3 below). Then to evaluate the right hand side of (1.2.20), we recall that $\int_{\gamma_i} \omega_{12}$ measures the rotation of the tangent vector to $\gamma_i$ relative to a parallel vector field $f_1$ along $\gamma_i$ as one traverses the curve (see 1.2.19). To understand this better first consider the case where the boundary curves $\gamma_i$ are geodesics. Then $\omega_{12}(\dot\gamma_i) = 0$, and the contributions to the integral come from the rotation of the tangent vector to the terminal point of $\gamma_i$ to the tangent vector to the initial point of $\gamma_{i+1}$, i.e., $\alpha_i$. Therefore, if $\partial\Delta$ consists of (broken) geodesics, (1.2.20) can be written in the form

$$\int_\Delta K\omega_1 \wedge \omega_2 = 2\pi - \sum_i \alpha_i. \tag{1.2.21}$$

[3]The version of the Gauss-Bonnet theorem for surfaces as we know it today, appears to be substanially due to Blaschke.

Notice that in the plane $\gamma_i$'s are straight line segments and $\sum_i \alpha_i = 2\pi$, however, for a general surface $\sum_i \alpha_i \neq 2\pi$. In fact, the deviation $2\pi - \sum \alpha_i$ is accurately measured by the the integral of the Gaussian curvature according to (1.2.21). Clearly if the boundary curves are not geodesics, then (1.2.21) should be modified as follows:

$$\int_\Delta K\omega_1 \wedge \omega_2 = 2\pi - \sum_i \alpha_i - \sum_i \int_{\gamma_i} \omega_{12}(\dot{\gamma}_i(t)). \tag{1.2.22}$$

We can rephrase the derivation of the right hand side of of (1.2.22) by saying that $\int_{\partial\Delta} \omega_{12}$ consists of two kinds of contributions, namely, the individual contributions of the smooth curves $\gamma_i$ and the exterior angles which reflect the discontinuities in the tangent vector field to the boundary. The quantity $\omega_{12}(\dot{\gamma}(t))$ is called the *geodesic curvature* of $\gamma$, and $2\pi - \sum_i \alpha_i$ the *excess of the sum of exterior angles*. Formulae (1.2.21) and (1.2.22) are versions of the *Gauss-Bonnet* theorem which we summarize as a proposition for future reference:

**Proposition 1.2.2** (Gauss-Bonnet) *Let $\Delta \subset M$ be an open relatively compact subset with connected boundary consisting of finitely many smooth curves and exterior angles $\alpha_i$ as described above. If the boundary curves are geodesics then (1.2.21) expresses the excess of the exterior angles of $\Delta$ as an integral of the Gaussian curvature $K$. If the boundary curves are not necessarily geodesic segments, then (1.2.21) should be replaced by (1.2.22).*

**Remark 1.2.3** Note that in the above discussion we intuitively thought of $\Delta$ as a polygonal region in $\mathbb{R}^2$ where $\mathbb{R}^2$ is given a Riemannian metric $ds^2$. It is instructive to examine this point a little closely. Assume for example that $\gamma$ is a small circle of (Euclidean) radius $\sin\alpha > 0$, $\alpha$ small, of the unit sphere $S^2$. Then $\gamma$ disconnects $S^2$ into two parts $\Delta_1$ (the small part) and $\Delta_2$ (the big part). In the application of (1.2.22) we have an ambiguity as to $\Delta$ being $\Delta_1$ or $\Delta_2$. Equation (1.2.22) gives in the case of $\Delta_1$ (with $e_1, e_2$ positively oriented and $e_2$ pointing to the interior of $\Delta_1$)

$$\int_{\Delta_1} K\omega_1 \wedge \omega_2 = 2\pi - \int_\gamma \omega_{12}(\dot{\gamma}(t)).$$

In exercise 1.2.8 below it is shown that

$$\int_\gamma \omega_{12}(\dot{\gamma}(t)) = 2\pi \cos\alpha.$$

Thus as $\alpha \to 0$ we obtain the obvious relation $\int K\omega_1 \wedge \omega_2 = 0$. On the other hand, the application of Stokes'theorem to $\Delta_2$ gives

$$\int_{\Delta_2} K\omega_1 \wedge \omega_2 = 2\pi + 2\pi\cos\alpha.$$

Therefore if $\alpha \to 0$ then we get $\int K\omega_1 \wedge \omega_2 = 4\pi$ which is of course valid. We can do a similar thing on any compact orientable surface $M$ with a Riemannian metric, by taking a small simple closed curve $\gamma$ disconnecting $M$ into a small part $\Delta_1$ and a big part $\Delta_2$. Applying (1.2.22) to the small part and shrinking $\gamma$ to a point we obtain the trivial relation $0 = 0$. On the other hand, if we apply it to the big part we obtain a possibly non-zero quantity. The meaning of this quantity will be discussed in detail in the next chapter after the introduction of homology. ♡

**Exercise 1.2.7** *Consider the upper half plane $\mathcal{H} = \{z = x + iy \in \mathbb{C} | y > 0\}$ with the Riemannian metric*

$$ds^2 = \frac{dx^2 + dy^2}{y^2}.$$

*Prove that the curvature of the hyperbolic plane is -1, and consequently the sum of the angles of a geodesic triangle is less than $\pi$.* ($\mathcal{H}$ with this metric is called the *hyperbolic* or *Poincaré plane*. See also subsection Spaces of Constant Curvature below.)

**Exercise 1.2.8** *Let $C$ be a (Euclidean) circle of radius $0 < \sin\alpha < 1$ on the unit sphere $S^2$, and $e_1, e_2$ be an orthonormal frame at $p \in C \subset S^2$. Parallel translate $e_1, e_2$ along $C$. Show that upon first return to the initial point $p$, the new frame makes an angle of $2\pi\cos\alpha$ radians with original frame.*

### 1.2.4 Geodesics

We noted earlier that parallel translation is specified by the system of ordinary differential equations (1.2.18). To understand this better for geodesics, we work in an open subset of $U \subset \mathbb{R}^m$ with a fixed Riemannian metric $g$ and denote the Levi-Civita connection by $(\omega_{ij})$. Let $e_1, \cdots, e_m$ be an orthonormal moving frame on $U$ and $\epsilon_1, \cdots, \epsilon_m$ be the standard basis for $\mathbb{R}^m$. Then we have

$$\epsilon_j = \sum_i \beta_{ij} e_i.$$

Let $\gamma(t) = (x_1(t), \cdots, x_m(t))$, $0 \le t \le 1$ be a curve in $U$, then

$$\dot\gamma(t) = (\frac{dx_1}{dt}, \cdots, \frac{dx_m}{dt}) = \sum_{i,j} \frac{dx_j}{dt}\beta_{ij}(t) e_i.$$

Noting that $(\beta_{ij})$ is an invertible matrix, we easily see that the condition for being a geodesic, i.e. $\nabla_{\dot\gamma}\dot\gamma = 0$, becomes a system of second order nonlinear ordinary differential equations:

$$\frac{d^2x_i}{dt^2} + \Phi_i(x, \frac{dx_1}{dt}, \cdots, \frac{dx_m}{dt}) = 0. \tag{1.2.23}$$

Here $\Phi_i$ are locally defined functions. Note the variable $t$ does not explicitly appear in $\Phi$'s. At the end of this subsection we explicitly exhibit the system of ordinary differential equations characterising geodesics. From the existence and uniqueness theorem for ordinary differential equations we immediately obtain

**Corollary 1.2.1** *Let $M$ be a Riemannian manifold, $p \in M$ and $\xi_p \in \mathcal{T}_pM$. Then there is a unique geodesic $\gamma = \gamma_{p,\xi_p} : (-\epsilon, \epsilon) \to M$ with $\gamma(0) = p$ and $\dot\gamma(0) = \xi_p$.*

The corollary enables us to define the *exponential map*, $\mathrm{Exp}_p : \mathcal{T}_pM \to M$ as $\mathrm{Exp}_p(\xi_p) = \gamma_{p,\xi_p}(1)$. Of course there is no reason why $\gamma_{p,\xi_p}(1)$ should even be defined since, for example, by taking a point out of a manifold many geodesics will terminate after finite time. It is a simple matter to see that

$$D\mathrm{Exp}_p(0) = \mathrm{Identity}$$

and therefore $\mathrm{Exp}_p$ is a diffeomorphism of a neighborhood of $0 \in \mathcal{T}_pM$ onto a neighborhood of $p$ in $M$. Let $B_p(r) \subset \mathcal{T}_pM$ be the ball of radius $r > 0$ and $\mathcal{B}_p(r) = \mathrm{Exp}_p(B_p(r))$. We set $\mathcal{S}_p(r) = \partial\mathcal{B}_p(r) = \mathrm{Exp}_p(S_p(r))$ where $S_p(r) = \partial B_p(r)$. Note that the length of a geodesic joining $p$ to a point on $\mathcal{S}_p(r)$ is $r$.

**Example 1.2.9** Let $M$ be a Riemannian manifold and $\tau : M \to M$ be an *isometry* of $M$, i.e., $\tau^\star(ds^2) = ds^2$ or the Riemannian metric is invariant under $\tau$. Assume $M^\tau = \{p \in M | \tau(p) = p\}$ is a submanifold. Let $p \in M^\tau$ and $\xi \in \mathcal{T}_pM^\tau$, then the geodesic in $M$ given by $\mathrm{Exp}_p(t\xi)$ remains in $M^\tau$, since otherwise using $\tau$ we obtain two geodesics in $M$ with the same initial point and tangent vector. Applying this observation to the sphere $S^m \subset \mathbb{R}^{m+1}$ and allowing $\tau$ to be reflections relative to the coordinate hyperplanes, we see that great circles are geodesics on the sphere. Similarly, the intersections of the ellipsoid or more generally the locus defined by

$$\frac{x_1^2}{a_1} + \frac{x_2^2}{a_2} + \frac{x_3^2}{a_3} = 1$$

with the coordinate hyperplanes are geodesics. In this way we obtain three closed on the ellipsoid. Obviously similar considerations apply to ellipsoids in higher dimensions. In the subsection on quadrics, we shall see that there is in fact a continuum of closed geodesics on the ellipsoid. In the case of the sphere $S^m$, invariance of the metric under $SO(m+1)$ and the uniqueness property of corollary 1.2.1 imply that all geodesics are great circles. ♠

It is sometimes useful to make use of the notion of *partial parallel translation* which is defined as follows: Let $\psi$ be a function on $U \subset M$ so that $M_c : \psi = c$ is a hypersurface. Let $\gamma$ be a curve in $U$ and $e_1, e_2, \cdots, e_m$ a moving frame such that $e_2, \cdots, e_m$ are tangent to the hypersurfaces $M_c$. Let $A$ be a gauge transformation given by the proper orthogonal matrix

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{m2} & \cdots & a_{mm} \end{pmatrix}$$

relative to the frame $e_1, e_2, \cdots, e_m$ where $a_{ij}$'s are functions on $U$. Let $(\omega_{ij})$ denote the connection form relative to the moving frame $e_1, A(e_2), \cdots, A(e_m)$. Then the system of ordinary differential equations

$$\omega_{ij}(\dot{\gamma}(t)) = 0, \quad \text{for } \ 2 \leq i, j \leq m$$

can be solved for $A$ along $\gamma$. This means that we can choose $A$ such that $A(e_2), \cdots, A(e_m)$ remain tangent $M_c$ and the coefficients $\omega_{ij}(\dot{\gamma}(t))$ for $2 \leq i, j \leq m$ vanish along $\gamma$.

Trying to understand the behavior of geodesics on a Riemannian manifold by directly solving the ordinary differential equations (1.2.23) is generally an exercise in futility. One should employ more clever ideas in recognizing and investigating geodesics. To this end we begin with the following observation: Assume that the metric has the form

$$ds^2 = g_{11} dx_1^2 + \sum_{i,j=2}^{m} g_{ij} dx_i dx_j. \tag{1.2.24}$$

Here $g_{ij}$'s are functions of $x_1, x_2, \cdots, x_m$. Then the curves $\Gamma_\gamma : x_2 = \gamma_2, \cdots, x_m = \gamma_m$, are orthogonal to the hypersurfaces $M_c : x_1 = c$. Let $e_1, \cdots, e_m$ be a moving frame such that $e_1$ is the unit tangent vector field to the curves $\Gamma_\gamma$. Then $\omega_1 = \sqrt{g_{11}} dx_1$, where $\omega_1, \cdots, \omega_m$ is the dual coframe. We furthermore assume that along each $\Gamma_\gamma$ the frame $e_1, e_2, \cdots, e_m$ is given by partial parallel translation along $\Gamma_\gamma$ so that $\omega_{ij}(e_1) = 0$ for $2 \leq i, j \leq m$.

**Lemma 1.2.2** *With the above notation and hypotheses on the metric (1.2.24) and the coframe $\omega_1, \cdots, \omega_m$, there is a non-singular matrix $(c_{ik})_{i,k=2,\cdots,m}$ of functions such that*

$$\omega_{1k} = (\sum_{i=2}^{m} c_{ki} \frac{\partial \log g_{11}}{\partial x_i}) \omega_1 + \ f_{k2} \omega_2 + \cdots + f_{km} \omega_m,$$

*for some functions $f_{jk}$.*

**Proof** - We have

$$d\omega_1 = \sum_{j=2}^{m} \frac{1}{\sqrt{g_{11}}} \frac{\partial \sqrt{g_{11}}}{\partial x_j} dx_j \wedge \omega_1. \tag{1.2.25}$$

There is a non-singular $(m-1) \times (m-1)$ matrix $(q_{jk})_{j,k=2,\cdots,m}$ such that

$$dx_j = \sum_{k=2}^{m} q_{kj}\omega_k.$$

Substituting in (1.2.25) and using the relation $d\omega_i + \sum \omega_{ij} \wedge \omega_j = 0$, we see that $\omega_{1k}$ is of the required form. ♣

**Proposition 1.2.3** *Let $M$ be a Riemannian manifold with metric $ds^2$ of the form (1.2.24). Then the orthogonal trajectories to the hypersurfaces $x_1 = c$ are geodesics, after possibly reparametrizing by arc-length, if and only if $g_{11}$ is a function of $x_1$ only.*

**Proof** - Since $2 \leq i, j \leq m$, $\omega_{ij}(e_1) = 0$, it follows from lemma 1.2.2 that the curves $\Gamma_\gamma$ are geodesics (after reparametrization by arc-length), i.e., $\omega_{ij}(e_1) = 0$, if and only if $g_{11}$ depends only on $x_1$. ♣

**Exercise 1.2.9** *Let $\Gamma$ be a curve in the $xz$-plane described by $(x(s), z(s))$ where $s$ is the arc length. Rotating $\Gamma$ around the $z$-axis gives a surface of revolution $S$ in $\mathbb{R}^3$. Show that the metric on $S$ is given by $ds^2 + x^2 d\theta^2$ with respect to the coordinates $(s, \theta)$ where $\theta$ is the angle of revolution, and deduce that the curves $\theta = c$ are geodesics for this metric.*

**Exercise 1.2.10** *Consider the upper half space $\{x = (x_1, \cdots, x_m) \in \mathbb{R}^m | x_m > 0\}$ with the Riemannian metric $ds^2 = g(x_m)(dx_1^2 + \cdots + dx_m^2)$ where $g(x_m) > 0$. Show that the orthogonal trajectories to $x_m = \text{const.}$, that is, the lines $x_1 = c_1, \cdots, x_{m-1} = c_{m-1}$ are geodesics after parametrization by arc-length.*

**Exercise 1.2.11** *Consider a disc $D \subset \mathbb{R}^m$ centered at the origin with a spherically symmetric metric $ds^2 = g(r)(dx_1^2 + \cdots + dx_m^2)$ where $r^2 = x_1^2 + \cdots + x_m^2$ and $g(r) > 0$. Let $\varphi_1, \cdots, \varphi_{m-1}$ be polar coordinates on $S_r = \{x \in D | x_1^2 + \cdots + x_m^2 = r^2\}$ for each $r > 0$. Show that the straight lines through the origin, i.e., the lines $\varphi_1 = c_1, \cdots, \varphi_{m-1} = c_{m-1}$, are geodesics up to reparametrization.*

To make the above observation more useful we proceed as follows: Let $\psi$ be a smooth function on $U \subseteq M$ where $M$ is a Riemannian manifold. Assume that the subsets defined by $M_c : \psi(x) = c$ are submanifolds of codimension one. We want to investigate the condition for the local existence of functions $\psi_2, \cdots, \psi_m$ such that

1. The subsets defined by $\Gamma_\gamma : \psi_2 = \gamma_2, \cdots, \psi_m = \gamma_m$ are the orthogonal trajectories to the submanifolds $M_c$;

2. $\Gamma_\gamma$'s are geodesics after reparametrization by arc-length;

3. $d\psi \wedge d\psi_2 \wedge \cdots \wedge d\psi_m \neq 0$, i.e., $\psi, \psi_2, \cdots, \psi_m$ is a coordinate system.

Let $\Psi = (\frac{\partial\psi}{\partial x_1}, \cdots, \frac{\partial\psi}{\partial x_m})$ where $x_1, \cdots, x_m$ are coordinate functions on $U$. Let $g = (g_{ij})$ be the matrix representation of the metric $ds^2$ relative to the coordinates $x_1, \cdots, x_m$. We show

**Lemma 1.2.3** *With the above notation, a necessary and sufficient condition for the existence of functions $\psi_2, \cdots, \psi_m$ satisfying conditions 1, 2 and 3 is*

- *$\Psi g^{-1}\Psi'$ is expressible as a function of $\psi$ only*[4]*,*

*where superscript $'$ denotes transpose.*

**Proof** - First we show the necessity. Let $h = (h_{ij})$ denote the the transformed metric relative to the coordinate system $\psi, \psi_2, \cdots, \psi_m$. Since the orthogonal trajectories $\psi_2 = \gamma_2, \cdots, \psi_m = \gamma_m$ are geodesics (after reparametrization) the metric $h = (h_{ij})$ has the property

1. $h_{11}$ is a function of $\psi$ only;

2. $h_{1i} = h_{i1} = 0$.

It follows from the transformation property of the metric that the 11-entry of the symmetric matrix $h = A^{-1}gA'^{-1}$ is expressible as function of only $\psi$ and its $1i$ entries vanish for $i > 1$. Here $A$ denotes the matrix

$$A = \begin{pmatrix} \frac{\partial\psi}{\partial x_1} & \frac{\partial\psi_2}{\partial x_1} & \cdots & \frac{\partial\psi_m}{\partial x_1} \\ \frac{\partial\psi}{\partial x_2} & \frac{\partial\psi_2}{\partial x_2} & \cdots & \frac{\partial\psi_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial\psi}{\partial x_m} & \frac{\partial\psi_2}{\partial x_m} & \cdots & \frac{\partial\psi_m}{\partial x_m} \end{pmatrix}.$$

Taking inverse of $A^{-1}gA'^{-1}$ we obtain the necessity. To prove sufficiency let $\psi_2, \cdots, \psi_m$ be a coordinate system on $M_c$ ($c$ fixed). We want to transport this coordinate system to $M_b$ for

[4]In other words, $||\mathrm{grad}\psi||$ is constant on each hypersurface $\psi = c$ in which case $\Gamma_\gamma$'s are the trajectories of the vector field $\mathrm{grad}\psi$. This formulation of the geodesic equations is the Hamilton-Jacobi view point which was introduced in the subsection on symplectic and contact structures in chapter 1. We prove the validity of this formulation directly here. A reason for usefulness of the Hamilton-Jacobi formulation is that the parametrization by arc-length, which is incorporated in the ordinary differential equations describing geodesics, is removed here.

all $b$ in a neighborhood of $c \in \mathbb{R}$ such that conditions 1, 2 and 3 are satisfied. We use the flow $\varphi_t$ of grad($\psi$) (which is orthogonal to the submanifolds $M_b$) to transport the coordinate functions $\psi_2, \cdots, \psi_m$ and show that the condition $\bullet$ ensures that $\psi(\varphi_t(x))$ depends only on $t$ (and $c$) but not on $x \in M_c$, i.e. $\varphi_t$ maps $M_c$ to $M_b$. This will prove the required sufficiency. Recall from chapter 1, example 1.8 that the vector field grad($\psi$) is expressible as the system of ordinary differential equations

$$\frac{dx_i}{dt} = \sum \tilde{g}_{ij} \frac{\partial \psi}{\partial x_j}, \quad i = 1, \cdots, m,$$

where $\tilde{g} = g^{-1}$. Therefore, by $\bullet$,

$$\frac{d\psi}{dt} = \sum_j \frac{\partial \psi}{\partial x_j} \frac{dx_j}{dt} = \sum_{j,k} \frac{\partial \psi}{\partial x_j} \tilde{g}_{jk} \frac{\partial \psi}{\partial x_k} = f(\psi).$$

Now $\frac{d\psi}{dt} = f(\psi)$ is a first order ordinary differential equation on the line which implies that $\psi(\varphi_t(x))$ depends only on $t$ (and $c$) but not on $x \in M_c$ as required. ♣

We can now complete the local picture for geodesics on a surface of revolution which we only partially discussed in exercise 1.2.9.

**Exercise 1.2.12** *With notation of exercise 11, let $\psi(s, \theta)$ be a function on the surface of revolution $S \subset \mathbb{R}^3$. Show that orthogonal trajectories to $\psi(s, \theta) = c$ are geodesics after reparametrization if and only if*

$$(\frac{\partial \psi}{\partial s})^2 + \frac{1}{x(s)^2}(\frac{\partial \psi}{\partial \theta})^2 = F(\psi),$$

*where $F$ is a function of one variable. Setting $F(\psi) = \psi^2$ and separating variables by setting $\psi(s, theta) = \psi_1(s)\psi_2(\theta)$ and $F(\psi) = \psi^2$ prove that the condition $\bullet$ becomes*

$$(\frac{d}{ds} \log \psi_1(s))^2 + \frac{1}{x(s)^2}(\frac{d}{d\theta} \log \psi_2(\theta))^2 = 1.$$

*Show that this differential equation can be explicitly integrated by for example setting $\psi_2(\theta) = e^{a\theta}$ to obtain geodesics on a surface of revolution.*

A particularly important application of the criterion $\bullet$ is

**Example 1.2.10** Fix a point $p \in M$ and let $r(x)$ be the distance of $x$ from $p$, i.e. the length of the shortest geodesic $\gamma_x$ joining $p$ to $x$. Define the function $E$ in a neighborhood of $p$ as $E(x) = \frac{1}{2}r(x)^2$. We want to verify condition $\bullet$ for the function $E$. Let $e_1, \cdots, e_m$ be a moving frame with $e_1$ the tangent vector field to $\gamma$ which we assume is parametrized by arc-length. Let $\omega_1, \cdots, \omega_m$ be the dual coframe as usual. Set $dE = \sum E_i\omega_i$, then

$$\frac{1}{2}r(x)^2 = \int_\gamma dE = \int_0^{r(x)} E_1(\gamma(t))dt.$$

It follows that $E_1(x) = r(x)$ and consequently $\mathrm{grad}(E) = r(x)e_1$. Therefore $dE(\mathrm{grad}(E)) = 2E$ and condition $\bullet$ is verified. This implies that there are functions $\psi_2, \cdots, \psi_m$ on the hypersurfaces $M_c : E(x) = c$ such that the Riemannian metric on $M$ takes the form

$$ds^2 = \frac{1}{2E}(dE)^2 + \sum_{i,j=2}^{m} \mathsf{H}_{ij}d\psi_i d\psi_j.$$

Equivalently, one can express $E$ in terms of $r$ to obtain the following which maybe regarded as the polar coordinate expression for the Riemannian metric:

$$ds^2 = dr^2 + \sum_{i,j=2}^{m} \mathsf{H}_{ij}d\psi_i d\psi_j. \tag{1.2.26}$$

We have shown that every Riemannian metric can be locally put in the form (1.2.24) with $g_{11}$ depending only on $x_1$, in fact the constant 1. Both functions $E$ and $r$ are important in Riemannian geometry. We could have used only the function $r$ in this example in which case $dr = \sum r_i\omega_i$ and $r_1 = 1$, $r_i = 0$ for $i > 1$ relative to the above choice of (co)frame. This expression for $dr$ is a version of the *First Variation Formula.* ♠

It is now trivial to prove the local length minimizing property of geodesics. In fact, we may assume the metric is of the form (1.2.24) with $g_{11}$ depending only on $x_1$. Then if $\gamma$ is a curve along the $x_1$ parameter curve joining $p = (c_1, c_2, \cdots, c_m)$ to $q = (c_1', c_2, \cdots, c_m)$ and $\delta$ is any other curve joining $p$ to $q$, then

$$\int \sqrt{ds^2(\dot\gamma, \dot\gamma)}dt = \int \sqrt{g_{11}}|\frac{dx_1}{dt}|dt \geq \int \sqrt{ds^2(\dot\delta, \dot\delta)}dt.$$

The same argument applies to $\int E(\dot\gamma)$ so that geodesics are also the critical points of the *Energy Functional* $\int E(\dot\gamma)$. A consequence of the above considerations is

**Exercise 1.2.13** *The geodesics emanating from $p$ intersect $\mathcal{S}_p(r)$ orthogonally for $r > 0$ small.* (This is sometimes called *Gauss Lemma*).

**Example 1.2.11** Let $M$ be a surface of negative curvature. By example 1.2.10 we may assume the metric has the form $ds^2 = dr^2 + G^2(r,\theta)d\theta^2$ where $G(r,\theta) \geq 0$ in a neighborhood $U$ of $x \in M$. We want to compare non-Euclidean distances of points in $U$ with with those in $\mathcal{T}_xM \simeq \mathbb{R}^2$. It is no loss of generality to assume that the restriction of the Riemannian metric to $\mathcal{T}_xM$ is the standard inner product on $\mathbb{R}^2$. Furthermore, by continuity, the area of a small ball of radius $r > 0$ around $x \in M$ tends to $\pi r^2$ as $r \to 0$. Therefore $G(r,\theta) = r + O(r^2)$ as $r \to 0$ by Taylor expansion, and

$$\lim_{r\to 0}\frac{G(r,\theta)}{r} = \lim_{r\to 0}\frac{\partial G}{\partial r} = 1.$$

Now $K = -\frac{1}{G}\frac{\partial^2 G}{\partial r^2} < 0$ which implies $\frac{\partial G}{\partial r}$ is increasing along each ray, and consequently $G(r,\theta) \geq r$. The Euclidean metric on $\mathbb{R}^2$ in polar coordinates is $dr^2 + r^2d\theta^2$ which we can now compare to $ds^2 = dr^2 + G^2(r,\theta)d\theta^2$. It follows that if $P$ and $Q$ are points near $x \in M$ with coordinates $(r_1,\theta_1)$ and $(r_2,\theta_2)$, then

$$d(P,Q) \;\geq\; \sqrt{r_1^2 + r_2^2 - 2r_1r_2\cos(\theta_2 - \theta_1)}. \tag{1.2.27}$$

This expresses the important geometric fact that on surfaces of negative curvature, geodesics locally diverge. This divergence property holds in any neighborhood $U$ of $x \in M$ where the representation of the metric as $ds^2 = dr^2 + G^2(r,\theta)d\theta^2$ is valid and there is a geodesic joining $P$ to $Q$ in $U$ and realizing $d(P,Q)$. An equivalent way of stating the divergence property (1.2.27) is the following inequality which replaces the law of cosines in plane geometry:

$$c^2 \geq a^2 + b^2 - 2ab\cos C, \tag{1.2.28}$$

where $a, b$ and $c$ are the lengths of the sides of a geodesic triangle with the vertices $A$, $B$ and $C$. ♠

**Example 1.2.12** We discuss a useful variation of example 1.2.11. We may assume the metric, in a small ball $U$ centered at $x \in M$, has the form $ds^2 = dr^2 + G^2(r,\theta)d\theta^2$ where $G(r,\theta) \geq 0$, and $G(r,\theta) = r + O(r^2)$. Then the rays through the origin (i.e., defined by $\theta = \text{constant}$) are geodesics and let $\gamma_1, \gamma_2$ be two rays intersecting at an angle $0 < \phi < \pi$ at the origin. Let $p_i$ be the point along $\gamma_i$ a distance $\delta$ from the origin. Then comparing the the metric $ds^2 = dr^2 + G^2(r,\theta)d\theta^2$ with the Euclidean metric $ds^2 = dr^2 + r^2d\theta^2$ we deduce that the length of the (Euclidean) line segment joining $p_1$ to $p_2$ is

$$2\delta\sqrt{1-\epsilon} + O(\delta^2),$$

where $0 < \epsilon < 1$ depends on the angle $\phi$. It follows that the distance between $p_1, p_2$ satisfies the strict inequality

$$d(p_1, p_2) < 2\delta, \tag{1.2.29}$$

for $\delta > 0$ sufficiently small. ♠

An examination of the argument in example 1.2.11 shows

**Exercise 1.2.14** *For a surface $M$ of positive curvature, the reverse of the inequality (1.2.28) is valid, i.e., $c^2 \leq a^2 + b^2 - 2ab\cos C$.*

**Exercise 1.2.15** *Show that the curvature of a surface determines the metric locally. Let $M$ be a surface with the metric $ds^2 = (a(t)+b(t)u)^2dt^2 + du^2$ relative to $(t,u)$ coordinates. Show that $M$ is flat. Describe explicitly (by integrations and other change of variables) how by a diffeomorphism the metric can be put in the Euclidean form $dx^2 + dy^2$.*

Let $M$ be a Riemannian manifold and $p, q \in M$. We define the distance between $p$ and $q$ as

$$d(p,q) = \inf_{\delta} L(\delta),$$

where $\delta : [0,1] \rightarrow M$ is a piece-wise $C^1$ path with $\delta(0) = p$ and $\delta(1) = q$, and inf is taken over all such paths. It is easy to show that $(M, d)$ is a metric space. In general, one cannot assert that inf can be replaced by min since by taking a point out of the plane we see that there are many pairs of points which cannot be connected by a straight line. To get around such anomalies, we introduce the notion of a complete Riemannian manifold. We say $M$ is *complete* if every geodesic $\gamma$ is defined for all parameter values $t \in \mathbb{R}$ ($t$ arc length). We have to see how this notion of completeness is related to the completeness of the metric space $(M, d)$, and whether for complete Riemannian manifolds, the inf can be realized by a geodesic. The answer to these questions is given by

**Proposition 1.2.4** (Hopf-Rinow) *For a complete Riemannian manifold $M$, and $p, q \in M$ there is a geodesic (not necessarily unique) $\gamma : [0,1] \rightarrow M$ connecting $p$ and $q$ and of length $d(p,q)$. The Riemannian manifold $M$ is complete if and only if $(M,d)$ is complete as a metric space.*

**Proof** - Assume every geodesic can be continued indefinitely. Let $d(p,q) = r$. From the locally length minimizing property of geodesics it follows that for $\epsilon > 0$ sufficiently small, there is $x \in S_p(\epsilon)$ (sphere of radius $\epsilon > 0$ centered at $p$) such that $d(x,q) = r - \epsilon$. Let $\gamma$ be the geodesic starting at $p$ and going through $x$. Let $T$ be defined by

$$T = \{t | d(\gamma(t), q) = r - t, 0 \leq t \leq r\}$$

It is trivial that $T$ is a non-empty closed subinterval $[0, t_\circ] \subseteq [0, r]$. We want to show $T = [0, r]$; so assume $t_\circ < r$. Consider $S_{\gamma(t_\circ)}(\delta)$ where $\delta > 0$ is small. Then there is $z \in S_{\gamma(t_\circ)}(\delta)$ such that $d(z,q) = r - t_\circ - \delta$, and $\gamma'$ be the geodesic joining $\gamma(t_\circ)$ to $z$. In view of the strict inequality (1.2.29) the distance between $\gamma(t_\circ - \delta)$ and $z = \gamma'(\delta)$ is strictly less than $2\delta$ which implies $d(p,q) < r$ contradicting the hypothesis. Therefore $z$ lies on the continuation of the geodesic $\gamma$ and $T = [0, r]$ proving that there is a geodesic of length $d(p,q) = r$ joining $p$ to $q$. Assuming every geodesic can be continued indefinitely, we show that $M$ is a complete as a metric space. Let $x_\circ, x_1, x_2, \cdots$ be a Cauchy sequence and let $\gamma_j$ be the geodesic joining $x_\circ$ to $x_j$. For fixed $\epsilon > 0$ (small number) the sequence of points $y_j = \gamma_j(\epsilon)$ has a convergent subsequence and by passing to a subsequence we can assume $y_j \rightarrow y$. Let $\gamma$ be the geodesic joining $x_\circ$ to $y$, then it easily verified that $x_j$ converge to a point on $\gamma$. Conversely, assume $M$ is complete as a metric space. Let $\gamma$ be a geodesic and assume $\gamma(t)$ is defined for $t < t_\circ$ but $\gamma(t_\circ)$ is not defined. Let $t_j \rightarrow t_\circ$. Then the Cauchy sequence $x_j = \gamma(t_j)$ will not converge. ♣

In view of proposition 1.2.4, the the usage of complete for two *a priori* different concepts will not cause any confusion.

**Remark 1.2.4** A related property of Riemannian manifolds is that every point $p$ has a *geodesically convex* neighborhood $U_p$ (in fact of the form $\mathcal{S}_p(\epsilon)$). This means that for every pair of points $x, y \in U_p$ there is a unique geodesic segment $\gamma$ of minimal length joining $x$ to $y$ and it necessarily lies entirely within $U_p$. This property will be useful, for example, in connection with de Rham cohomology. The method of proof is not related to the techniques emphasized in this book and is therefore omitted (see [Wh]). ♡

**Exercise 1.2.16** *Let $z_1, \cdots, z_n \in \mathbb{C}$ and consider the metric*

$$ds^2 = \frac{1}{\prod |z - z_i|^2} dz d\bar{z}$$

*on $M = \mathbb{C} \setminus \{z_1, \cdots, z_n\}$. Show that $ds^2$ is a metric of negative curvature on $M$. Is this metric complete?*

To derive the explicit form of the differential equations describing geodesics, it is convenient to work in the framework of symplectic geometry. For a manifold $M$ with a Riemannian metric $ds^2$, the symplectic structure on $\mathcal{T}^\star M$ can be transported to $\mathcal{T}M$ by invoking the isomorphism $\mathcal{T}_x^\star M \xrightarrow{\sim} \mathcal{T}_x M$ induced by $ds^2$. More precisely, the linear functions $\theta_i$ are transported to the tangent space $\mathcal{T}_x M$ to obtain $\phi_i = \sum_j g_{ij}\theta_j$. Now set $\tilde{\varepsilon} = \sum_i \phi_i dx_i$, then the symplectic form on the tangent bundle is

$$\tilde{\omega} = -d\tilde{\varepsilon} = \sum_i dx_i \wedge d\phi_i. \tag{1.2.30}$$

The system of ordinary differential equations characterizing geodesics on a Riemannian manifold $M$ can be studied in the framework of symplectic geometry. In the context of chapter 1 §3.5(???) we want to show that this system is of the form (**??**) relative to the symplectic structure on the tangent bundle of $M$. This description of geodesics will facilitate their study. To express the equations of geodesics as a Hamiltonian system on the tangent bundle we follow the procedure familiar from classical mechanics. Consider the function $E : \mathcal{T}M \to \mathbb{R}$ defined by $2E(\xi_x) = ds^2(\xi_x, \xi_x)$, where $\xi_x \in \mathcal{T}_x M$. Let $L = \tilde{\varepsilon} - Edt$, and note that substituting $\xi_i = \frac{dx_i}{dt}$ in $L$ we obtain $L = E$ for a curve described by $\gamma(t) = (x_1(t), \cdots, x_n(t))$. Therefore geodesics are critical points (i.e., curves $\gamma(t)$) of the functional $\int L$. By the standard procedure of Calculus of Variations, we take exterior derivative of the integrand $L$ to obtain

$$\int dL = \int \big(\sum_i (\frac{dx_i}{dt} - \frac{\partial E}{\partial \phi_i}) d\phi_i\big) \wedge dt - \int \big(\sum_i (\frac{d\phi_i}{dt} + \frac{\partial E}{\partial x_i}) dx_i\big) \wedge dt.$$

Setting the integrands equal to zero we obtain the differential equations characterizing critical curves:

$$\frac{dx_i}{dt} = \frac{\partial E}{\partial \phi_i}, \quad \frac{d\phi_i}{dt} = -\frac{\partial E}{\partial x_i}. \tag{1.2.31}$$

Denoting the transpose of a matrix by the superscript $'$ and setting $\xi = (\xi_1, \cdots, \xi_n)$, $\phi = (\phi_1, \cdots, \phi_n)$, we obtain $E = \frac{1}{2}\xi g \xi' = \frac{1}{2}\phi g^{-1}\phi'$. Therefore

$$dE = -\frac{1}{2}\phi g^{-1}(dg)g^{-1}\phi' + \xi d\phi'.$$

Substituting in (1.2.31) we obtain

$$\frac{dx_i}{dt} = \xi_i, \quad \frac{d\phi_i}{dt} = \frac{1}{2}\sum_{j,k} \frac{\partial g_{jk}}{\partial x_i}\xi_j \xi_k, \qquad i = 1, \cdots, m. \tag{1.2.32}$$

We refer to (1.2.32) or its equivalent form

$$\frac{d}{dt}(\sum_j g_{ij}\frac{dx_j}{dt}) = \frac{1}{2}\sum_{j,k} \frac{\partial g_{jk}}{\partial x_i}\frac{dx_j}{dt}\frac{dx_k}{dt}, \qquad i = 1, \cdots, m \tag{1.2.33}$$

as the *symplectic form of the equations of geodesics.*

**Example 1.2.13** Let $\gamma : S^1 \to \mathbb{R}^2$ be a smooth simple closed curve. Let the coordinates in $\mathbb{R}^2$ be $(x, z)$ and we assume

1. Either the image $\gamma$ lies in the half plane $x > 0$;

2. Or image of $\gamma$ is invariant under the symmetry $(x, z) \to (-x, z)$.

Rotation of the image of $\gamma$ around the $z$-axis yields in the first case a surface diffeomorphic to a torus, and in the second case a surface diffeomorphic to the sphere. Denote this surface by $M$, and assume $\gamma$ is parametrized by its arclength $s$. Let us show that the geodesic flow on $\mathcal{T}M$ is a completely integrable Hamiltonian system. The metric on $M$ is diagonal with $g_{11} = 1$ and $g_{22} = x(s)^2$ (see exercise 1.2.9) and consequently the symplectic form of the equations of geodesics is

$$\frac{d^2s}{dt^2} = x(s)x'(s)(\frac{d\theta}{dt})^2, \quad \frac{d}{dt}(x(s)^2\frac{d\theta}{dt}) = 0.$$

The second equation suggests that the function $G = x(s)^2\frac{d\theta}{dt}$ together with the Riemannian metric or the function $E$ implement complete integrability. The second equation shows that $G$ is invariant under the geodesic flow. It then follows from the definition of Poisson bracket and $E$ and $G$ are involution. The rank condition is also easily verified to hold and we have established completely integrability of the geodesic flow on a surface of revolution. The tori $N_c$ have also a very simple description in this case. The differential equations for geodesics (up to reparametrization) are given by the vector field grad$\psi$, i.e.,

$$\frac{ds}{dt} = \sqrt{1 - \frac{a^2}{x(s)^2}}\psi, \quad \frac{d\theta}{dt} = \frac{a\psi}{x(s)^2},$$

where $\psi$ is the function constructed in example **??**. The second equation shows that $\psi$ is a constant times $G$ and therefore each torus $N_c$ is simply the orthogonal trajectories to the level curves $\psi =$ const. (Note that $\psi$ depends on $a$ so that, in general, different values of $a$ give distinct tori.) ♠

### 1.2.5 Curvature

We stated earlier that the deviation of the quantity $d\omega_{ij} + \sum \omega_{ik} \wedge \omega_{kj}$ from vanishing reflects the curvature of the space, and can be calculated from the metric $ds^2$ alone. We set $\Omega_{ij} = d\omega_{ij} + \sum \omega_{ik} \wedge \omega_{kj}$, and call the matrix $\Omega = (\Omega_{ij})$ the *curvature form.* Our immediate goal in this subsection is to study $\Omega$ and in the next subsection we relate it to the second fundamental form(s) just as we did in the case of surfaces in $\mathbb{R}^3$. $\Omega$ is a skew symmetric matrix and depends on the choice of frame. Therefore it is defined on the bundle $\mathcal{F}_g$ of orthonormal frames and its individual entries are not of geometric interest. From (1.2.2) it follows easily that the dependence of $\Omega$ on the choice of frame is given by

$$\Omega = A^{-1}\Omega' A. \tag{1.2.34}$$

Since the entries of $\Omega$ are 2-forms, and 2-forms commute, we can manipulate the matrix $\Omega$ as if it were a matrix of scalars. Thus, for example, the various symmetric functions of the characteristic roots of $\Omega$, which are polynomials in $\Omega_{ij}$'s, are independent of the choice of the frame and are defined on the manifold $M$. This observation plays an important role in the differential geometry of Riemannian manifolds and understanding the connection between geometry and topology.

The identity $dd = 0$ implies certain relations among $\omega_i, \omega_{ij}$ and $\Omega_{ij}$. Indeed $dd\omega_i = 0$ implies the *first Bianchi* identity:

$$\sum_j \Omega_{ij} \wedge \omega_j = 0. \tag{1.2.35}$$

Similarly the relation $dd\omega_{ij} = 0$ implies the *second Bianchi* identity:

$$d\Omega_{ij} = \sum_k \omega_{ik} \wedge \Omega_{kj} - \sum_k \Omega_{ik} \wedge \omega_{kj}. \tag{1.2.36}$$

We set

$$2\Omega_{ij} = \sum_{k,l} R_{ijkl}\omega_k \wedge \omega_l.$$

The scalar $R_{ijij}$ is called the *sectional curvature* of the plane determined by the vectors $e_i,\ e_j$. $R_{ijkl}$ is called the *curvature tensor*. The curvature tensor satisfies the relations

$$R_{ijkl} + R_{iklj} + R_{iljk} = 0, \quad R_{ijkl} = -R_{jikl} = R_{jilk}, \quad R_{ijkl} = R_{klij}. \tag{1.2.37}$$

The first identity is a consequence of (1.2.35), the second and third equations are trivial and the last equality follows from the preceding ones by simple manipulations. These properties of the curvature tensor can be re-stated in the framework of group representations and will be again discussed in the subsection Representations of Curvature Tensor.

The curvature of a Riemannian manifold may be interpreted as an endomorphism in more than one way. The endomorphism of $\mathcal{T}_xM$, for each $x \in M$, defined by $S_{ik}(e_l) = 2\sum_{j=1}^m \Omega_{ij}(e_k, e_l)e_j$ has trace $R_{ik}$ (called *Ricci tensor*)

$$R_{ik} = \mathrm{Tr}(S_{ik}) = 2\sum_{j,l=1}^{m} < \Omega_{ij}(e_k, e_l)e_j, e_l >_x = \sum_{l=1}^{m} R_{ilkl}$$

where $<,>_x$ denotes the inner product (Riemannian metric) on $\mathcal{T}_xM$. Thus each component of the Ricci tensor at $x \in M$ is the trace of an endomorphism of $\mathcal{T}_xM$ (see also subsections on Young Diagrams in Chapter 1, and on the Structure of Curvature Tensor below). Clearly the Ricci tensor is a symmetric matrix. It is a simple exercise to see that under a gauge transformation $A$, the Ricci tensor transforms according

$$(R_{ik}) = A^{-1}(R^{\star}_{ik})A, \tag{1.2.38}$$

where $(R^{\star}_{ik})$ denotes the Ricci tensor relative to the new orthonormal basis.

It is customary to define the curvature operator $R$ as

$$R(e_i, e_j)e_k = 2\sum_{l=1}^{m} \Omega_{lk}(e_i, e_j)e_l.$$

It is useful to regard the curvature operator $R$ and the curvature tensor $R_{ijkl}$ as multilinear functions on $\mathcal{T}_xM$ or elements of the tensor algebra on $\mathcal{T}_xM$. For instance, if $v = \sum v_i e_i$ and $w = \sum w_i e_i$, then

$$R(v, w) = \sum_{i,j} v_i w_j R(e_i, e_j).$$

Similarly if $v' = \sum v'_i e_i$ and $w' = \sum w'_i e_i$, then

$$R(v, w, v', w') = \sum_{i,j,k,l} v_i w_j v'_k w'_l R_{ijkl}.$$

With this interpretation it is immediate that the sectional curvature can be regarded as the assignment of a number to each 2-plane in $\mathcal{T}_xM$. If $V \subset \mathcal{T}_xM$ is a 2-plane, and $e_1, \cdots, e_m$ is a basis for $\mathcal{T}_xM$ with $e_1, e_2$ spanning $V$, then

$$R_{1212} = \frac{R(v_1, v_2, v_1, v_2)}{g_x(v_1, v_1)g_x(v_2, v_2) - (g_x(v_1, v_2))^2}, \tag{1.2.39}$$

where $v_1, v_2$ is any basis for $V$ and $g_x$ denotes the inner product on $\mathcal{T}_xM$ (the Riemannian metric). The curvature tensor may be regarded as an element of $S^2(\bigwedge^2 W)$ where $W = \mathcal{T}^{\star}_x M$ (symmetric bilinear form on the second exterior power). Since a symmetric bilinear $B$ form is uniquely determined by its values on the diagonal, i.e.,

$$2B(u, v) = B(u + v, u + v) - B(u, u) - B(v, v),$$

the curvature tensor is determined by the sectional curvatures.

A Riemannian manifold $M$ is called *Einstein* if its Ricci tensor, when expressed relative to an orthonormal frame, is a multiple of the identity. This condition is equivalent to the requirement that relative to a coordinate system the Ricci tensor is multiple of the metric $ds^2$. In view of the transformation property (1.2.38), the Einstein property is independent of the choice of orthonormal frame. It expresses an intrinsic geometric property of the Riemannian manifold which is not as restrictive as being of constant sectional curvature.

**Example 1.2.14** In this example we investigate the Einstein condition in the special case where $\dim M = 4$. We fix an orthonormal frame $\{e_1, \cdots, e_4\}$, and recall that $R_{ij} = \sum_k R_{ikjk}$. In particular, for an Einstein manifold we have

$$\sum_k R_{ikik} - \sum_k R_{jkjk} = 0,$$

for all $i, j$. This is a homogeneous system of three linear equations in six unknowns $R_{ikik}$. It is a simple matter to see that the solutions to this system are characterized by

$$R_{1212} = R_{3434},\ R_{1313} = R_{2424},\ R_{1414} = R_{2323}.$$

In other words, sectional curvatures of the planes determined by $\{e_1, e_2\}$ and $\{e_3, e_4\}$ are equal, etc. In view of the independence of the Einstein condition from the choice of frame and the transformation property (1.2.38), this conclusion can be restated as a four dimensional Riemannian manifold is Einstein if and only if its sectional curvatures are identical on orthogonal planes. ♠

**Exercise 1.2.17** *By emulating the argument of example (1.2.14) show that for an Einstein manifold of dimension 3, sectional curvatures at a point $x \in M$ do not depend on the choice of the planes in $\mathcal{T}_xM$, and $R_{1213} = 0$ etc. Thus $\Omega_{ij} = R(x)\omega_i \wedge \omega_j$ for some function $R : M \to \mathbb{R}$.*

The following example shows how part of exercise 1.2.17 generalizes to higher dimensions:

**Example 1.2.15** Let $M$ be a Riemannian manifold of dimension $\geq 3$ and assume that the sectional curvatures at $x \in M$ do not depend on the choice of the plane (spanned by $e_i, e_j$). We show that the symmetries of the curvature tensor imply that $M$ necessarily has constant curvature. Let $e_1, \cdots, e_m$ be a moving frame for $M$, and set

$$e_1^\theta = \cos\theta e_1 + \sin\theta e_3, \quad e_3^\theta = -\sin\theta e_1 + \cos\theta e_3.$$

Then $e_1^\theta, e_2, e_3^\theta, e_4, \cdots$ is a moving frame for $M$, and let $\omega_1^\theta, \omega_2, \omega_3^\theta, \omega_4, \cdots$ be the dual coframe. We denote the curvature form relative to this frame by $(\Omega_{ij}^\theta)$. Let $R_{1212}^\theta$ denote the coefficient of $\omega_1^\theta \wedge \omega_2$ in $\Omega_{12}^\theta$. By 4-linearity of the curvature tensor

$$R_{1212}^\theta = \cos^2\theta R_{1212}^\circ + \sin^2\theta R_{3232}^\circ + \sin 2\theta R_{1232}^\circ.$$

The hypothesis implies that $R_{1212}^\theta = R_{3232}^\theta$ and is independent of $\theta$, and consequently $R_{1232}^\circ = 0$. In other words, $R_{ijkl} = R_{ijkl}^\circ = 0$ if exactly three of the indices $i, j, k, l$ are distinct. Similarly by looking at the coefficient of $\omega_1^\theta \wedge \omega_4$ in $\Omega_{12}^\theta$ and using $R_{1214} = 0$ etc. we obtain $R_{1234}^\circ + R_{3214}^\circ = 0$, or

$$R_{ijkl}^\circ + R_{kjil}^\circ = 0 \tag{1.2.40}$$

This relation together with the first Bianchi identity $R_{4321}^\circ + R_{4213}^\circ + R_{4231}^\circ = 0$ imply

$$R_{4213}^\circ + 2R_{4132}^\circ = 0. \tag{1.2.41}$$

Equation (1.2.40) and skew symmetry of the curvature tensor in the last two indices imply

$$R_{4132} = -R_{3142} = -R_{4231} = R_{4213}$$

Substituting in (1.2.41) we obtain $R_{4213} = 0$. It follows that the curvature tensor form $\Omega_{ij} = \Omega_{ij}^\circ$ is of the form

$$\Omega_{ij} = R(x)\omega_i \wedge \omega_j. \tag{1.2.42}$$

Taking exterior derivative of $\Omega_{ij}$, using the second Bianchi identity and substituting from (1.2.42) we obtain

$$\sum dR \wedge \omega_i \wedge \omega_j = 0,$$

which implies that $dR = 0$ and $M$ has constant curvature. This example is due to *Schur*. ♠

**Example 1.2.16** A consequence of example 1.2.15 is the extension of example **??** to higher dimensions, i.e., we show that a hypersurface all whose points are umbilics is necessarily part of a sphere or a hyperplane. Let $m \geq 3$ and $M \subset \mathbb{R}^{m+1}$ be a hypersurface. Let $e_1, \cdots, e_{m+1}$ be a be an orthonormal frame with $e_{m+1}$ a unit normal vector field to $M$. Just as in the case of a surface in $\mathbb{R}^3$, the hypothesis that every point is an umbilic implies

$$\omega_{1\ m+1} = a\omega_1, \cdots, \omega_{m\ m+1} = a\omega_m, \tag{1.2.43}$$

where $a$ is a function on $M$. It follows that sectional curvatures of $M$ are the same for all the planes in $\mathcal{T}_xM$ and therefore $M$ has constant sectional curvature. Consequently it is part of a sphere, a hyperplane or a space of constant negative curvature[5]. In the latter case, not all the principal curvatures can be equal and therefore does not occur. ♠

**Example 1.2.17** Let $U$ be the group of $3 \times 3$ upper triangular matrices with 1's along the diagonal. A general matrix in $U$ will be denoted by $u = (u_{ij})$, and a basis for left invariant 1-forms on $U$ is given by the three non-zero entries of the matrix $u^{-1}du$ as explained in chapter 1. Therefore such a basis of left invariant 1-forms is

$$\omega_1 = du_{12}, \quad \omega_2 = du_{23}, \quad \omega_3 = du_{13} - u_{12}du_{23}.$$

Consequently a left invariant Riemannian metric is $ds^2 = A^2\omega_1^2 + B^2\omega_2^2 + C^2\omega_3^2$ for any positive numbers $A, B$, and $C$. Using the defining property of the Levi-Civita connection, i.e., $\omega_{ij} + \omega_{ji} = 0$ and $d\omega_i + \sum \omega_{ij} \wedge \omega_j = 0$, one obtains after a simple calculation

$$(\omega_{ij}) = \begin{pmatrix} 0 & -\frac{C^2}{2AB}\omega_3 & -\frac{C}{2AB}\omega_2 \\ \frac{C^2}{2AB}\omega_3 & 0 & \frac{C}{2AB}\omega_1 \\ \frac{C}{2AB}\omega_2 & -\frac{C}{2AB}\omega_1 & 0 \end{pmatrix}.$$

Thus

$$\Omega = \frac{C^2}{4A^2B^2} \begin{pmatrix} 0 & -3\omega_1 \wedge \omega_2 & \omega_1 \wedge \omega_3 \\ 3\omega_1 \wedge \omega_2 & 0 & \omega_2 \wedge \omega_3 \\ -\omega_1 \wedge \omega_3 & -\omega_2 \wedge \omega_3 & 0 \end{pmatrix}.$$

is the curvature form of $ds^2$. ♠

**Exercise 1.2.18** *Compute the Ricci tensor of $U$ relative to $ds^2$ of example 1.2.17, and show that it has two negative and one positive eigenvalue.*

**Example 1.2.18** Let $M = SU(2) \simeq S^3$. We have seen that $M$, being a compact simple analytic group, admits of a unique (up to scalar multiplication) Riemannian metric which is invariant under left and right translations. This metric is in fact the natural metric induced on $S^3$ from $\mathbb{R}^4$. For $u$ a variable point in $SU(2)$, we write the matrix $u^{-1}du$ which is skew-hermitian and traceless, as

$$u^{-1}du = \begin{pmatrix} i\omega_1 & \omega_2 + i\omega_3 \\ -\omega_2 + i\omega_3 & -i\omega_1 \end{pmatrix}.$$

[5] The classification of spaces of constant curvature is discussed in the next section and in chapter 4 where the vailidity of this assertion becomes evident.

The entries $\omega_i$ are left invariant 1-forms, and the real symmetric 2-tensor $\sum \omega_j^2$ is invariant under the adjoint action of $SU(2)$. Therefore it is the bi-invariant metric on $M$. The corresponding Levi-Civita connection $\omega_{jk}$ is computed from the equation $d(u^{-1}du) = -u^{-1}du \wedge u^{-1}du$ which yields

$$(\omega_{jk}) = \begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix}.$$

The corresponding curvature form is $\Omega_{jk} = \omega_j \wedge \omega_k$. For $A, B$ and $C$ positive numbers

$$ds^2 = A^2\omega_1^2 + B^2\omega_2^2 + C^2\omega_3^2$$

is a left invariant metric on $M$ which is not right invariant unless $A = B = C$. In this example, we investigate the curvature of this metric. Let $\theta_1 = A\omega_1, \theta_2 = B\omega_2$ and $\theta_3 = C\omega_3$. Using its defining property, it is a simple calculation to derive the following expression for the Levi-Civita connection $(\theta_{jk})$:

$$(\theta_{jk}) = \begin{pmatrix} 0 & -\frac{A^2+B^2-C^2}{ABC}\theta_3 & \frac{A^2-B^2+C^2}{ABC}\theta_2 \\ \frac{A^2+B^2-C^2}{ABC}\theta_3 & 0 & -\frac{-A^2+B^2+C^2}{ABC}\theta_1 \\ -\frac{A^2-B^2+C^2}{ABC}\theta_2 & \frac{-A^2+B^2+C^2}{ABC}\theta_1 & 0 \end{pmatrix}.$$

It is convenient to set

$$\alpha_1 = \frac{-A^2+B^2+C^2}{ABC}, \quad \alpha_2 = \frac{A^2-B^2+C^2}{ABC}, \quad \alpha_3 = \frac{A^2+B^2-C^2}{ABC}.$$

Then the curvature form $\Theta_{jk} = d\theta_{jk} + \sum \theta_{jl} \wedge \theta_{lk}$ is the skew symmetric matrix

$$\begin{pmatrix} 0 & (-\alpha_1\alpha_2 + \alpha_1\alpha_3 + \alpha_2\alpha_3)\theta_1 \wedge \theta_2 & (\alpha_1\alpha_2 - \alpha_1\alpha_3 + \alpha_2\alpha_3)\theta_1 \wedge \theta_3 \\ -(-\alpha_1\alpha_2 + \alpha_1\alpha_3 + \alpha_2\alpha_3)\theta_1 \wedge \theta_2 & 0 & (\alpha_1\alpha_2 + \alpha_1\alpha_3 - \alpha_2\alpha_3)\theta_2 \wedge \theta_3 \\ -(\alpha_1\alpha_2 - \alpha_1\alpha_3 + \alpha_2\alpha_3)\theta_1 \wedge \theta_3 & -(\alpha_1\alpha_2 + \alpha_1\alpha_3 - \alpha_2\alpha_3)\theta_2 \wedge \theta_3 & 0 \end{pmatrix}.$$

It follows that the Ricci tensor $K = (K_{jk})$ is diagonal with diagonal entries given by

$$K_{11} = \alpha_2\alpha_3, \quad K_{22} = \alpha_1\alpha_3, \quad K_{33} = \alpha_1\alpha_2.$$

In particular, the metric is not Einstein unless $A = B = C$. ♠

Two Riemannian metrics differing by multiplication by a positive function, are called *conformally equivalent* since the measure of angles between tangent vectors is the same relative to such metrics. Conformally equivalenmt metrics, besides being of obvious geometric interest are also significant in physics. The following exercise describes the simplest situation where they naturally occur:

**Exercise 1.2.19** *Let $M \subset \mathbb{R}^3$ be a surface not passing through the origin, and $\bar{M}$ be the surface obtained from $M$ via inversion relative to the origin with inversion parameter $c$. This means that if $\mathbf{x} = (x_1, x_2, x_3)$ describes the surface $M$, then $\bar{M}$ is given by*

$$\bar{\mathbf{x}} = \frac{c^2}{x_1^2 + x_2^2 + x_3^2}\mathbf{x}.$$

*Show that the induced metric on $\bar{M}$ is related to that of $M$ by*

$$d\bar{s}^2 = \frac{c^4}{(x_1^2 + x_2^2 + x_3^2)^2} ds^2,$$

*Therefore $M$ and $\bar{M}$ are conformally equivalent. By looking at the second fundamental form of $\bar{M}$ show that the lines of curvature of $M$ are mapped to those of $\bar{M}$ through inversion. Let $p =< \mathbf{x}, e_3 >$. Show that the principal curvatures of $\bar{M}$ are related to those of $M$ by*

$$\bar{\kappa}_i = -\frac{x_1^2 + x_2^2 + x_3^2}{c^2}\kappa_i - \frac{2p}{c^2},$$

*and relate the Gaussian and mean curvatures of $\bar{M}$ and $M$.*

**Example 1.2.19** Let us compute the curvature of the metric $du^2 = e^{2\sigma}(dx_1^2 + \cdots + dx_m^2)$, where $\sigma$ is a function of $x = (x_1, \cdots, x_m)$. Our problem is to compute the curvature of a conformally flat (or Euclidean) metric. Now $\theta_1 = e^\sigma dx_1, \cdots, \theta_m = e^\sigma dx_m$ is an orthonormal coframe for $du^2$, and define $\sigma_j$ by

$$d\sigma = \sum \sigma_j \theta_j,$$

i.e., $\sigma_j = e^{-\sigma}\frac{\partial \sigma}{\partial x_j}$. Therefore

$$d\theta_l = \sum \theta_j \wedge (\sigma_j \theta_l),$$

and the Levi-Civita connection for $du^2$ is

$$\theta_{lj} = \sigma_j \theta_l - \sigma_l \theta_j,$$

where the addition of the term $-\sigma_l\theta_j$ is to make $\theta_{lj}$ anti-symmetric. To compute the curvature we first calculate the second derivative of $\sigma$ relative to the coframe $\theta_1, \cdots, \theta_m$ as follows: From $dd = 0$ we obtain $d(\sum \sigma_k \theta_k) = 0$ and consequently

$$0 = \sum d\sigma_k \wedge \theta_k + \sum \sigma_k d\theta_k = \sum_k (d\sigma_k - (\sum_l \sigma_l \theta_{lk})) \wedge \theta_k.$$

By Cartan's lemma

$$d\sigma_k - \sum_l \sigma_l \theta_{lk} = \sum \sigma_{kl}\theta_l,$$

where $\sigma_{lk} = \sigma_{kl}$. (The expression for $\sigma_{kl}$ in terms of coordinates is given in the exercise 1.2.20 below, but will not be used.) Let $\Theta_{jl} = d\theta_{jl} + \sum_i \theta_{ji}\wedge\theta_{il}$ denote the curvature form. Substituting and applying exterior derivative, we obtain after some calculation

$$\Theta_{jl} = \sum_k (\sigma_{kj} - \sigma_j\sigma_k)\theta_k \wedge \theta_l + \sum_k (\sigma_{kl} - \sigma_k\sigma_l)\theta_j \wedge \theta_k + (\sum_k \sigma_k^2)\theta_j \wedge \theta_l. \tag{1.2.44}$$

In terms of components $2\Theta_{ij} = \sum_{k,l} S_{ijkl}\theta_k \wedge \theta_l$, this translates into

$$S_{ijkl} = -\delta_{jk}(\sigma_{li}-\sigma_l\sigma_i)+\delta_{jl}(\sigma_{ki}-\sigma_k\sigma_i)+\delta_{ki}(\sigma_{lj}-\sigma_l\sigma_j)-\delta_{il}(\sigma_{kj}-\sigma_k\sigma_j)+(\delta_{ki}\delta_{lj}-\delta_{il}\delta_{jk})(\sum_i \sigma_k^2). \tag{1.2.45}$$

Finally

$$L_{ij} = (m-2)(\sigma_{ij} - \sigma_i\sigma_j) + (m-2)\delta_{ij}(\sum_k \sigma_k^2) + \delta_{ij}(\sum_k \sigma_{kk}). \tag{1.2.46}$$

is the Ricci tensor $L_{ij} = \sum_k S_{ikjk}$ of the metric $du^2$. ♠

**Exercise 1.2.20** *Show that*

$$\sigma_{kl} = -2\sigma_k\sigma_l + \delta_{kl}(\sum_i \sigma_i^2) + e^{-2\sigma}\frac{\partial^2\sigma}{\partial x_l \partial x_k}.$$

The significance of this example goes beyond giving explicit formulae for the curvature of a special metric. It is a simple calculation that if $dv^2 = e^{2\sigma}ds^2$ with Levi-Civita connections $\phi_{ij}$ (for $dv^2$) and $\omega_{ij}$ (for $ds^2$), then

$$\phi_{lj} = \sigma_j\phi_l - \sigma_l\phi_j + \omega_{lj}. \tag{1.2.47}$$

Repeating the calculation in the example almost verbatim (here the connection forms appear, but the terms explicitly involving the connections $\omega_{ij}$ and $\phi_{ij}$ cancel out[6]), we obtain the remarkable fact that the curvature of the metric $dv^2 = e^{2\sigma}ds^2$ is the sum of the curvatures

[6]The fact that the terms involving the connection forms cancel out is not surprising in view of the different transformation properties of the connection and curvature forms, and the fact that one can make the connection vanish at one point by a gauge transformation. This phenomenon is sometimes very useful.

of $ds^2$ and $du^2 = e^{2\sigma}(dx_1^2 + \cdots + dx_m^2)$ formally. By *formally* we mean that the expressions are the sum of the corresponding expressions, however, we should note that the quantities $\sigma_i$ and $\sigma_{ij}$ are calculated relative to the metric $dv^2$ and the connection $\phi_{ij}$ and not $du^2$ and $\theta_{ij}$. We refer to this property as the *formal additivity* of the curvature tensor under conformal change of metric. Summarizing

**Proposition 1.2.5** *Let $dv^2 = e^{2\sigma}ds^2$ be conformally equivalent Riemannian metrics. Then their Levi-Civita connections $\phi_{ij}$ and $\omega_{ij}$ are related by (1.2.47), and the corresponding curvatures are*

$$\Phi_{jl} - \Omega_{jl} = \sum_k (\sigma_{kj} - \sigma_j\sigma_k)\phi_k \wedge \phi_l + \sum_k (\sigma_{kl} - \sigma_k\sigma_l)\phi_j \wedge \phi_k + (\sum_k \sigma_k^2)\phi_j \wedge \phi_l.$$

*Writing $2\Phi_{ij} = \sum_{k,l} F_{ijkl}\phi_k \wedge \phi_l$, this translates into*

$$\begin{aligned} F_{ijkl} - R_{ijkl} &= -\delta_{jk}(\sigma_{li} - \sigma_l\sigma_i) + \delta_{jl}(\sigma_{ki} - \sigma_k\sigma_i) + \delta_{ki}(\sigma_{lj} - \sigma_l\sigma_j) \\ & \quad -\delta_{il}(\sigma_{kj} - \sigma_k\sigma_j) + (\delta_{ki}\delta_{lj} - \delta_{il}\delta_{jk})(\textstyle\sum_i \sigma_k^2). \end{aligned}$$

*Denoting the Ricci tensor for $dv^2$ and $ds^2$ by $\mathsf{H}_{ij}$ and $K_{ij}$ respectively we obtain*

$$\mathsf{H}_{ij} - K_{ij} = (m-2)(\sigma_{ij} - \sigma_i\sigma_j) + (m-2)\delta_{ij}(\sum_k \sigma_k^2) + \delta_{ij}(\sum_k \sigma_{kk}).$$

(We emphasize that the quantities $\sigma_i$ and $\sigma_{ij}$ are calculated relative to the metric $dv^2$ and the connection $\phi_{ij}$.)

The following proposition relates the curvature of the Levi-Civita connection to covariant differentiation, and it is sometimes used as the definition of curvature. While not essential for the development of the theory in our framework, it may help the reader relate the material here to other points of view.

**Proposition 1.2.6** *For vector fields $\xi$ and $\eta$ on the Riemannian manifold $M$ we have*

$$\nabla_\xi\eta - \nabla_\eta\xi - [\xi,\eta] = 0, \quad \nabla_\eta\nabla_\xi - \nabla_\xi\nabla_\eta - \nabla_{[\eta,\xi]} = 2\Omega(\xi,\eta).$$

(Notice that the right hand side of the second equation does not involve differentiation and is a purely algebraic pointwise operation. This equation should be interpreted as the result of applying the differential operators of the left hand side to the a vector field $\zeta$ is identical as that of applying the matrix $2\Omega(\xi,\eta)$ to $\zeta$.)

**Proof** - Let $\{e_1, \cdots, e_m\}$ be an orthonormal moving frame on $M$ and $\{\omega_1, \cdots, \omega_m\}$ be the dual coframe. Set $\xi = \sum h_i e_i$ and $\eta = \sum g_i e_i$. Then the first assertion reduces to showing

$$\nabla_{e_k} e_i - \nabla_{e_i} e_k - [e_k, e_i] = 0.$$

Now by (1.2.15) we have

$$\nabla_{e_k} e_i - \nabla_{e_i} e_k = \sum_j (\omega_{ji}(e_k) - \omega_{jk}(e_i))e_j.$$

Then from proposition 1.2.1 and the formula for exterior derivative we obtain

$$\omega_j([e_i, e_k]) = 2d\omega_j(e_k, e_i) = \omega_j(\nabla_{e_k} e_i - \nabla_{e_i} e_k),$$

from which the first formula follows. To prove the second formula let $\zeta = \sum f_l e_l$. After a straightforward computation we see that it suffices to establish

$$\nabla_{e_k}\nabla_{e_i} e_l - \nabla_{e_i}\nabla_{e_k} e_l - \nabla_{[e_k,e_i]} e_l = 2\sum_j \Omega_{jl}(e_k, e_i)e_j.$$

We have

$$\nabla_{e_k}\nabla_{e_i} e_l = \sum_j e_k(\omega_{jl}(e_i))e_j + \sum_{j,n} \omega_{jl}(e_i)\omega_{nj}(e_k)e_n.$$

Therefore

$$\begin{aligned}
\nabla_{e_k}\nabla_{e_i} e_l - \nabla_{e_i}\nabla_{e_k} e_l &= \textstyle\sum_j (e_k(\omega_{jl}(e_i)) - e_i(\omega_{jl}(e_k)))e_j \\
&\quad \textstyle+\sum_{j,n} (\omega_{jl}(e_i)\omega_{nj}(e_k) - \omega_{jl}(e_k)\omega_{nj}(e_i))e_n \\
&= \textstyle\sum_j (e_k(\omega_{jl}(e_i)) - e_i(\omega_{jl}(e_k)))e_j \\
&\quad \textstyle+2\sum_{j,n} \omega_{jl} \wedge \omega_{nj}(e_i, e_k)e_n \\
&= \textstyle\sum_j (e_k(\omega_{jl}(e_i)) - e_i(\omega_{jl}(e_k)))e_j + 2\sum_j \Omega_{jl}(e_k, e_i)e_j - 2\sum_j d\omega_{jl}(e_k, e_i)e \\
&= \textstyle\sum_j (\omega_{jl}([e_k, e_i]))e_j + 2\sum_j \Omega_{jl}(e_k, e_i)e_j \\
&= \textstyle\nabla_{[e_k,e_i]} e_l + 2\sum_j \Omega_{jl}(e_k, e_i)e_j,
\end{aligned}$$

and the required result follows. ♣

### 1.2.6 Curvature and Second Fundamental Forms

In this subection we investigate some of the basic relations between curvature and the second fundamental forms in higher (co)dimensions. We begin with an interpretation of the eigenvalues of the second fundamental form of a hypersurface in terms of the curvature of plane curves.

**Example 1.2.20** Let $M \subset \mathbb{R}^{m+1}$ be a smooth hypersurface and $e_1, \cdots, e_{m+1}$ a moving frame near $M$ with $e_{m+1}$ normal to $M$. Let $P_x$ be a plane containing the vector $e_{m+1}(x)$, then the intersection $P_x \cap M$ is a curve $\Gamma$ in the plane $P_x$. Let $\kappa_1 < \cdots < \kappa_m$ be the principal curvatures of $M$. Therefore if $e_1, \cdots, e_m$ are along the principal directions then the matrix of the second fundamental form becomes diagonal with diagonal entries $\kappa_1, \cdots, \kappa_m$. Now let $f_1 \cdots, f_{m+1} = e_{m+1}$ be a moving frame with $f_1$ along tangent to the curve $\Gamma$. Let $\tilde{\omega}_A$ and $\tilde{\omega}_{AB}$ be 1-forms defined by $f_1, \cdots, f_{m+1}$ defined via $df_A = \sum \tilde{\omega}_{BA} f_B$ etc. Then we have

$$-\tilde{\omega}_{1\ m+1} = \mathsf{H}_{11}\tilde{\omega}_1 + \cdots + \mathsf{H}_{1m}\tilde{\omega}_m, \ \ \cdots, \ \ -\tilde{\omega}_{m\ m+1} = \mathsf{H}_{m1}\tilde{\omega}_1 + \cdots + \mathsf{H}_{mm}\tilde{\omega}_m,$$

where the matrix $\mathsf{H} = (\mathsf{H}_{ij})$ is symmetric. The curvature of the plane curve $\Gamma$ is $\kappa_\Gamma = \tilde{\omega}_{1\ m+1}(f_1) = \mathsf{H}_{11}$. Since the matrix of the second fundamental form transforms according $\mathsf{H} \to A'\mathsf{H}A$ where $A$ is an orthogonal matrix, we obtain

$$a = A_{11}^2\kappa_1 + \cdots + A_{m1}^2\kappa_m. \tag{1.2.48}$$

Therefore, by orthogonality of $A$, the curvature of $\Gamma$ is a convex combination of the principal curvatures of $M$. ♠

Let $M \subset \mathrm{R}^{m+1}$ be a hypersurface and $\mathsf{H} = (\mathsf{H}_{ij})$ denote the matrix of the second fundamental form relative to an orthonormal frame $e_1, \cdots, e_{m+1}$ with $e_{m+1}$ normal to $M$. It follows from the strucure equations that

$$\Omega_{ij} = -\omega_{im+1} \wedge \omega_{m+1j}, \tag{1.2.49}$$

for a hypersurface $M \subset \mathbb{R}^{m+1}$. Recall that the second fundamental form for a hypersurface $M$ is the symmetric matrix $\mathsf{H} = (\mathsf{H}_{ij})$ where

$$\omega_{m+1i} = \sum_{j=1}^{m} \mathsf{H}_{ij}\omega_j.$$

Substituting in (1.2.49) we see that that the sectional curvature $-R_{ijij}$ is given by the principal minor

$$R_{ijij} \ = \ -\det\begin{pmatrix} \mathsf{H}_{ii} & \mathsf{H}_{ij} \\ \mathsf{H}_{ij} & \mathsf{H}_{jj} \end{pmatrix} \tag{1.2.50}$$

which is the generalization to hypersurfaces of (1.2.9) for surfaces.

The fact that sectional curvatures are expressible in terms of $2 \times 2$ minors of the second fundamental form can be extended to submanifolds of arbitrary codimension. Let $M \subset \mathrm{R}^N$ be a submanifold of codimension $N - m$, and $\mathsf{H}_p = (\mathsf{H}_{ij}^p)$ denote the matrix of the second fundamental form in the direction of the normal vector $e_p$ ($p \geq m+1$). From the structure equations we see that the analogue of (1.2.49) for submanifolds is

$$\Omega_{ij} = -\sum_{p=m+1}^{N} \omega_{ip} \wedge \omega_{pj}. \tag{1.2.51}$$

The symmetric matrix $\mathsf{H}_p = (\mathsf{H}^p_{ij})$ is determined by

$$\omega_{pi} = \sum_{j=1}^{m} \mathsf{H}_{ij}\omega_j.$$

Substituting in (1.2.51) we see that that the sectional curvature $-R_{ijij}$ is given by the principal minor

$$R_{ijij} \;=\; -\sum_{p=m+1}^{N} \det\begin{pmatrix} \mathsf{H}^p_{ii} & \mathsf{H}^p_{ij} \\ \mathsf{H}^p_{ij} & \mathsf{H}^p_{jj} \end{pmatrix}, \tag{1.2.52}$$

thus generalizing Theorema Egregium to submanifolds of arbitrary codimension. It follows that the sum on the right hand side of (1.2.52) is independent of the choice of orthonormal frame $e_{m+1}, \cdots, e_N$.

In example 1.2.20 the principal curvatures of a hypersurface were related to the curvature of plane curves on a surface. The following example relates the sectional curvature to the Gaussian curvature of certain embedded surfaces. This example will be derived in a more elaborate way in example 1.2.25 below, but it is included here since it is an instructive demonstration of the use of moving frames.

**Example 1.2.21** Let $M$ be a Riemannian manifold $\gamma : I = (-1,1) \to M$ a geodesic. Let $e_1, \cdots, e_m$ be an orthonormal frame at $p = \gamma(0)$ with $e_1$ tangent to $\gamma$. Parallel translate the frame $e_1, \cdots, e_m$ along $\gamma$. For a small number $\epsilon > 0$ and a unit vector field $\xi$ along $\gamma$ orthogonal to $e_1$, we let $N_\xi$ be the surface

$$N_\xi = \cup_{t\in(-\epsilon,\epsilon)} \mathrm{Exp}_{\gamma(s)} t\xi_\gamma(s).$$

From smoothness of the dependence of solutions of an ordinary differential equation on initial conditions we deduce that $N_\xi$ is a surface. We parallel translate $e_1, e_2$ *relative to* $N_\xi$ along the geodesics $t \to \mathrm{Exp}_{\gamma(s)} t\xi_\gamma(s)$ and extend it to a moving frame. As $\xi$ varies we obtain a moving frame in an open set containing the image of $\gamma$. Note that the vector field $e_1$ may not be parallel along $t \to \mathrm{Exp}_{\gamma(s)} t\xi_\gamma(s)$ relative to $M$ although it is so relative to $N_\xi$. For definiteness let $\xi = e_2$, set $N = N_{e_2}$. We relate the the Gaussian curvature of $N$ to the sectional curvature of the Riemannian manifold $M$. To do so we use the superscript $N$ for quantities referring to $N$. We have

$$d\omega_1^N + \omega_{12}^N \wedge \omega_2^N = 0, \quad d\omega_2^N + \omega_{21}^N \wedge \omega_1^N = 0.$$

Comparing with

$$d\omega_1 + \omega_{12} \wedge \omega_2 + \sum_{k=3}^{m} \omega_{1k} \wedge \omega_k = 0, \quad d\omega_1 + \omega_{21} \wedge \omega_1 + \sum_{k=3}^{m} \omega_{2k} \wedge \omega_k = 0,$$

and noting that on $N$

$$\omega_1^N = \omega_1, \quad \omega_2^N = \omega_2$$

we obtain

$$\omega_{12}^N = \omega_{12}, \qquad \text{on} \quad N. \tag{1.2.53}$$

By the assumption of parallel translation $\omega_{ij}(e_1) = 0$ along $\gamma$ and consequently

$$\omega_{1k} \wedge \omega_{k2} = 0 \quad \text{on } N \text{ along } \gamma. \tag{1.2.54}$$

The curvature forms are

$$\Omega_{12}^N = d\omega_{12}^N, \quad \Omega_{12} = d\omega_{12} + \sum_{k=3}^{m} \omega_{12} \wedge \omega_{k2}.$$

Since $d\omega_{12}^N = d\omega_{12}$ on $N$, (1.2.54) implies

$$\Omega_{12}^N = \Omega_{12} \quad \text{on } N \text{ along } \gamma. \tag{1.2.55}$$

In other words, the sectional curvatures of $M$ relative to the planes spanned by $e_1, e_2$ along $\gamma$ are identical with the Gaussian curvature of $N$ along $\gamma$. This does not mean that for $x \in N$, the sectional curvature of $M$ relative the plane $\mathcal{T}_x N$ is equal to the Gaussian curvature of $N$ at $x$. We are ensured of equality only for points on $\gamma$. The essential property of $N$ that we used in this example was that it contains the geodesic $\gamma$ and $e_2$ was parallel in $M$ along $\gamma$. ♠

**Example 1.2.22** Consider a surface $M \subset \mathbb{R}^4$ and let $\tau_r(M)$ denote the tube of radius $r$ around $M$. We want to calculate $\mathrm{vol}(\tau_r(M))$ for small $r > 0$. Proceedings as in the case of curves in $\mathbb{R}^3$ we note

$$\tau_r(M) = \{p + t_3e_3 + t_4e_4 \mid p \in M,\ t_3^2 + t_4^2 < r\}.$$

Denoting a generic point of $\tau_r(M)$ by $q = p + t_3e_3 + t_4e_4$ we obtain

$$dq = (\omega_1 + t_3\omega_{13} + t_4\omega_{14})e_1 + (\omega_2 + t_3\omega_{23} + t_4\omega_{24})e_2 + dt_3e_3 + dt_4e_4.$$

Implicit in this representation is the local parametrization of $\tau_r(M)$ as $M \times B^2(r)$ where $B^2(r)$ denotes the disc of radius $r$ in $\mathbb{R}^2$. To obtain a useful expression for the volume element on $\tau_r(M)$, the terms $\omega_{AB}$ should be expressed in terms of $\omega_1, \omega_2$ by restriction to $M$. It follows that the volume element on $\tau_r(M)$ can be written as

$$dv = (\omega_1 + t_3\omega_{13} + t_4\omega_{14}) \wedge (\omega_2 + t_3\omega_{23} + t_4\omega_{24}) \wedge dt_3 \wedge dt_4. \tag{1.2.56}$$

To integrate this 4-form we note that any term involving an odd power of $t_3$ or $t_4$ will vanish after integration over $\tau_r(M)$. Therefore

$$\int_{\tau_r(M)} dv = \int_{\tau_r(M)} (\omega_1 \wedge \omega_2 + t_3^2 \omega_{31} \wedge \omega_{32} + t_4^2 \omega_{41} \wedge \omega_{42}) \wedge dt_3 \wedge dt_4.$$

Making the change of variable to polar coordinates

$$t_3 = \rho \cos\phi, \quad t_4 = \rho \sin\phi,$$

and carrying out the integration in $(t_3, t_4)$, we obtain

$$\int_{\tau_r(M)} dv = \pi r^2 \mathrm{vol}(M) + \frac{\pi r^4}{4} \int_M (\omega_{31} \wedge \omega_{32} + \omega_{41} \wedge \omega_{42}).$$

In view of (1.2.52) we obtain

$$\int_{\tau_r(M)} dv = \pi r^2 \mathrm{vol}(M) + \frac{\pi r^4}{4} \int_M K \omega_1 \wedge \omega_2, \tag{1.2.57}$$

for the volume of tube of small radius $r > 0$ around $M$. Here $K$ denotes the Gaussian curvature of the surface $M$. Note that the volume of tube does not involve the mean curvature or any quantity which depends on the embedding, and is expressed in terms of the volume and Gaussian curvature which can be calculated from the knowledge of $ds^2$. This reflects a general phenomenon about the volume of tubes around submanifolds. The point is that the integrals of terms involving odd powers of $t_j$ vanish for reasons of symmetry, and the coefficients of terms involving only even powers of $t_j$'s can be expressed in terms of quantities intrinsic to $M$, i.e., $\omega_j$'s and $\Omega_{ij}$'s. For an extensive discussion of volumes of tubes see [Gr]. ♠

**Example 1.2.23** It is useful to see how the second derivative or the Hessian of a function is calculated in the context of moving frames. Let $f : M \to \mathbb{R}$ be a smooth function on the Riemannian manifold $M$. Relative to a coframe $\omega_1, \cdots, \omega_m$ we have

$$df = \sum_i f_i \omega_i.$$

The $f_i$'s may be regarded as partial derivative relative to the coframe $\omega_i$. The relation $ddf = 0$ implies

$$\sum_k \left[ df_k - \sum_i f_i \omega_{ik} \right] \wedge \omega_k = 0$$

By Cartan's lemma we have

$$df_k - \sum_i f_i\omega_{ik} = \sum f_{jk}\omega_j, \qquad \text{with} \;\; f_{jk} = f_{kj}.$$

The matrix $(f_{jk})$ is the Hessian of $f$ relative to the coframe $\omega_1, \cdots, \omega_m$. Let $M \subset \mathbb{R}^{m+n}$ be a submanifold (embedded) and consider the function $f(x) = \frac{1}{2} < x, x >$ on $M$ where $< ., . >$ is the standard inner product on $\mathbb{R}^{m+n}$. We compute the Hessian of $f$ relative a moving frame $e_1, \cdots, e_m$ Extend $e_1, \cdots, e_m$ to an (orthonormal) moving frame $e_1, \cdots, e_{m+n}$ for $\mathbb{R}^{m+n}$. Set $y_i =< x, e_i >$, $i = 1, \cdots, m$ and $y_a =< x, e_a >$, $a = m+1, \cdots, m+n$ to obtain $df = \sum_i y_i\omega_i$ where we emphasize that $f$ is regarded as a function on $M$. Then $ddf = 0$ becomes

$$\sum_k dy_k \wedge \omega_k - \sum_{i,k} y_i\omega_{ik} \wedge \omega_k - \sum_{i,a} y_i\omega_{ia} \wedge \omega_a = 0.$$

On $M$ we have $\omega_a = 0$. Therefore substituting $de_i = \sum_j \omega_{ji}e_j + \sum_a \omega_{ai}e_a$ in the above equation we obtain

$$\sum_{a=m+1}^{m+n} (\omega_i + y_a\omega_{ai}) \wedge \omega_i = 0.$$

Since $\omega_{ai} = \sum_k \mathsf{H}^a_{ik}\omega_k$ with $\mathsf{H}^a_{ik} = \mathsf{H}^a_{ki}$ (second fundamental form), the above equation becomes

$$\sum_i \left[\delta_{ik} + \sum_a y_a\mathsf{H}^a_{ik}\right]\omega_k \wedge \omega_i = 0,$$

and $f_{ik} = \delta_{ik} + \sum_a y_a\mathsf{H}^a_{ik}$ gives the Hessian of $f$. ♠

**Example 1.2.24** The conclusion of the example 1.2.23 for the function $f(x) = \frac{1}{2} < x, x >$ for a hypersurface $M \subset \mathbb{R}^{m+1}$ can be stated as

$$< x, e_{m+1} > (\mathsf{H}_{ij}) = (f_{ij}) - I. \tag{1.2.58}$$

Since $R_{ijij} = \mathsf{H}_{ii}\mathsf{H}_{jj} - \mathsf{H}^2_{ij}$ and $2\times 2$ principal minors of a negative definite matrix are positive definite, the sectional curvatures of a compact hypersurface are positive at some point, viz., the maxima of $f$. (In view of example **??** we may assume $f$ is a Morse function so that the matrix $(f_{ij})$ is negative definite at a maximum.) Shortly we will see how this argument can be vastly generalized to obtain non-isometric embedding theorems for negatively curved Riemannian manifolds. Under certain circumstances one can establish the existence of an approximate maximum for the function $f(x) = \frac{1}{2} < x, x >$ on a noncompact hypersurface. Then one can conclude that there are regions where the sectional curvatures of the hypersurface are positive. This kind of argument occurs in connection with the proof of the non-existence of complete minimal surfaces in the interior of certain regions of $\mathbb{R}^3$ since the curvature of a minimal surface is everywhere non-positive. The analytical argument establishing the existence of an approximate maximum is known as *Omori's lemma* (See [O]). ♠

It is reasonable to surmise that the relationship between the second fundamental form and curvature can be utilized to gain some insight into how curvature affects isometric embedding of a Riemannian manifold $M$. To develop this theme we introduce some algebraic definitions. For $x \in M$ let $V_x \subset \mathcal{T}_xM$ be the linear subspace spanned by unit vectors $e_1$ such that $\Omega_{1j}$ vanishes identically at $x$. This is equivalent to the statement $R_{1j1j} = 0$ for all $j = 2, \cdots, m$. It is clear that $V_x$ is the maximal linear subspace of $\mathcal{T}_xM$ such that if $e_1 \in V_x$ is a unit vector, then $\Omega_{1j}$ or equivalently the sectional curvatures $R_{1j1j}$ vanish identically at $x \in M$. We set $\mu_x = \dim V_x$, and refer to $\mu_x$ as the *index of nullity* at $x$. Now assume $M \subset \mathbb{R}^{m+n}$ with the Riemannian metric induced from the ambient Euclidean space. We let $W_x \subset \mathcal{T}_xM$ be the linear span of all unit vectors $e_1$ such that for all normal directions $e_p \in (\mathcal{T}_xM)^\perp$ the linear form $\mathsf{H}_p(e_1, .)$ vanishes identically at $x$. It is immediate that $W_x$ is the maximal linear subspace of $\mathcal{T}_xM$ such that for all unit vectors $e_1 \in W_x$ we have $\mathsf{H}_p(e_1, .) = 0$. We set $\nu_x = \dim W_x$ and refer to it as the *index of relative nullity*. In view of (1.2.52) we have

$$\nu_x \leq \mu_x. \tag{1.2.59}$$

Next we introduce a bilinear mapping

$$\alpha_x : \mathcal{T}_xM \times \mathcal{T}_xM \to (\mathcal{T}_xM)^\perp, \quad \alpha_x(\xi, \eta) = \sum_{p=m+1}^{m+n} \mathsf{H}_p(\xi, \eta)e_p,$$

where $e_{m+1}, \cdots, e_{m+n}$ is an orthonormal basis for $(\mathcal{T}_xM)^\perp$. It is immediate from the transformation property (1.2.4) of the second fundamental form that $\alpha_x$ is meaningfully defined. Let $S_x$ denote the orthogoanl complement of $W_x$ in $V_x$. Then by restriction $\alpha_x$ induces a bilinear map

$$\tilde{\alpha}_x : S_x \times S_x \longrightarrow (\mathcal{T}_xM)^\perp.$$

We have

**Lemma 1.2.4** *Let $\beta : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^n$ be a bilinear pairing. If $k > n$, then there are vectors $u, v \in \mathbb{R}^k$, not both zero, such that*

$$\beta(v, v) = \beta(u, u), \quad \beta(v, u) = 0.$$

**Proof** - Regarding $\beta$ as a bilinear pairing of complex vector spaces, we deduce the existence of a non-zero vector $w = u + iv$ such that $\beta(w, w) = 0$ since $k > n$. Expanding into real and imaginary parts gives the required result. ♣

Then the identity (1.2.52) can be stated in terms of $\alpha$ as

$$R_{ijij} =< \alpha(e_i, e_i), \alpha(e_j, e_j) > - < \alpha(e_i, e_j), \alpha(e_i, e_j) > . \tag{1.2.60}$$

Now we complement inequality (1.2.59) with

**Lemma 1.2.5** *With the above notation we have*

$$\nu_x \leq \mu_x \leq \nu_x + n,$$

*where $n$ is the codimension of the embedding of $M$.*

**Proof** - It remains to prove the second inequality which is equivalent to $\dim(S_x) \leq n$. Assume $\dim(S_x) > n$, then applying lemma 1.2.4 to the bilinear map $\tilde{\alpha}$ we obtain vectors $u, v \in \mathcal{T}_x M$ such that

$$\tilde{\alpha}(u, u) = \tilde{\alpha}(v, v), \quad \tilde{\alpha}(u, v) = 0.$$

We may assume both $u$ and $v$ are non-zero and linearly independent. Let $e_1, e_2, \cdots$ be an orthonorml basis such that $e_1, e_2$ span the plane spanned by $u$ and $v$. Then the identity (1.2.60) and $R_{1212} = 0$ imply $\tilde{\alpha}(u, u) = 0 = \tilde{\alpha}(v, v)$. Since $R_{1j1j} = R_{2j2j} = 0$ for $e_j \in \mathcal{T}_x M$, we obtain from (1.2.60)

$$\tilde{\alpha}(e_1, e_j) = 0 = \tilde{\alpha}(e_2, e_j).$$

It follows that $e_1, e_2$ and therefore $u, v$ are in $V_x$ contrary to the hypothesis. ♣

In example **??** of chapter 1 we showed that if $M \subset \mathbb{R}^{m+n}$ is a compact submanifold then there is $x \in M$ such that

$$f : M \to \mathbb{R}, \quad f(x) = \frac{1}{2} < x - p, x - p >$$

is a Morse function for almost all $p \in \mathbb{R}^{m+n}$. Therefore we may assume that the function $f(x) = \frac{1}{2} < x, x >$ is a Morse function on $M$ (after a translation). In view of example 1.2.23 the Hessian of $f$ is given by $(\delta_{ik} + \sum_a y_a \mathsf{H}^a_{ik})$. In particular at a point where $f$ is a maximum the matrix $(\sum_a y_a \mathsf{H}^a_{ik})$ is negative definite. This implies

**Lemma 1.2.6** *Let $M \subset \mathbb{R}^{m+n}$ be a compact submanifold, and $f(x) = \frac{1}{2} < x, x >$. Then, after possibly a translation of $M$, at a maximum of $f$ we have $\nu_x = 0$.*

We now use the above observations to relate the codimension of a compact embedded submanifold and its sectional curvatures.

**Proposition 1.2.7** *Let $M$ be a compact Riemannian manifold. Assume that for every $x \in M$ there is a subspace $V_x \subset \mathcal{T}_x M$ of dimension $\geq q \geq 2$ such that the curvature of any 2-plane contained in $V_x$ is non-positive. Then the codimension of an isometric embedding of $M$ in $\mathbb{R}^N$ is at least $n$.*

**Proof** - Let $\psi : M \to \mathcal{R}^N$ be an isometric embedding. After a possible translation, we may assume $f(x) = \frac{1}{2} < x, x >$ is a Morse function on $M$, and $\nu_x = 0$ at a maximum of $f$ on $M$ by lemma 1.2.6. Therefore it suffices to show $\mu_x \geq q$ by lemma 1.2.5. If $\mu_x < q$, then by lemma 1.2.4 there are vectors $u, v \in V_x$ such that $\tilde{\alpha}(u, u) = \tilde{\alpha}(v, v) \neq 0$ and $\tilde{\alpha}(u, v) = 0$. It follows from (1.2.60) that the curvature of plane spanned by $u$ and $v$ is positive contrary to hypothesis. ♣

**Corollary 1.2.2** *A compact flat Riemannian manifold $M$ of dimension $m$ cannot be isometrically embedded in $\mathbb{R}^{2m-1}$.*

The flat $m$-dimensional torus can be isometrically embedded in $\mathbb{R}^{2m} = \mathbb{C}^m$ as

$$\{(e^{i\theta_1}, \cdots, e^{i\theta_m}) \mid \theta_j \in [0, 2\pi)\}.$$

Therefore the conclusion of proposition 1.2.7 is sharp. The Riemannian metric of an a submanifold $\psi : M \to \mathbb{R}^N$ is given by the $m \times m$ matrix $(D\psi)' D\psi$. Therefore the existence problem for isometric immersions/embeddings of Riemannian manifolds hinges on the existence of solutions to the system of nonlinear partial differential equations $(D\psi)' D\psi = g$. This is a difficult problem in analysis and is inapropriate in the context of this volume.

**Exercise 1.2.21** *Let $M \subset \mathbb{R}^{m+1}$ be a hypersurface. Show that if the rank of the second fundamental form is $\geq 2$, then $\nu_x = \mu_x$.*

While we have emphasized the the geometry of submanifolds of Euclidean space, the basic concept of second fundamental form can be defined for submanifolds of Riemannian manifolds. Let $e_1, \cdots, e_n$ be a moving frame for the Riemannian manifold $N$ and $\omega_1, \cdots, \omega_n$ the corresponding dual coframe. Assume the submanifold $M$ is defined by the Pfaffian system

$$\omega_{m+1} = 0, \cdots, \omega_n = 0.$$

Let $\omega_{AB}$, $1 \leq A, B \leq n$, denote the Levi-Civita connection for $N$ and $\Omega_{AB}$ be the corresponding curvature forms. Since $\omega_a = d\omega_a = 0$ on $M$, the application of Cartan's lemma to $\sum_i \omega_{ai}\wedge\omega_i = 0$ on $M$ implies the existence of the symmetric matrix $(\mathsf{H}^a_{ij})$ (the second fundamental form in the direction $e_a$) such that

$$\omega_{ai} = \sum_j \mathsf{H}^a_{ij}\omega_j.$$

Let superscripts $^M$ and $^N$ signify reference to the manifolds $M$ and $N$ respectively. Then

$$\Omega^N_{ij} = d\omega_{ij} + \sum_{k=1}^m \omega_{ik}\wedge\omega_{kj} + \sum_{a=m+1}^n \omega_{ia}\wedge\omega_{aj} = \Omega^M_{ij} + \sum_{a=m+1}^n \omega_{ia}\wedge\omega_{aj}.$$

Therefore

$$\Omega^N_{ij} = \Omega^M_{ij} - \sum_{a=m+1}^n \sum_{k,l=1}^m \left[\mathsf{H}^a_{ik}\mathsf{H}^a_{jl} - \mathsf{H}^a_{il}\mathsf{H}^a_{jk}\right]\omega_k\wedge\omega_l, \tag{1.2.61}$$

which relates the curvature tensors of $M$ and $N$ to the second fundamental form of $M\subset N$. In particular the sectional curvatures are related by

$$R^M_{ijij} = R^N_{ijij} + \sum_{a=m+1}^n \left[\mathsf{H}^a_{ii}\mathsf{H}^a_{jj} - (\mathsf{H}^a_{ij})^2\right]. \tag{1.2.62}$$

**Exercise 1.2.22** *Let $N$ be a Riemannian manifold and $M$ a submanifold. Let $e_1,\cdots,e_n$ be a moving frame on $N$ with $e_1,\cdots,e_m$ tangent to $M$. Show that*

$$\nabla^N_{e_k}e_i = \nabla^M_{e_k}e_i + \alpha(e_i,e_k),$$

*where $\alpha(e_i,e_j) = \sum_{a=m+1}^n \mathsf{H}^a(e_i,e_j)e_a$ and $\mathsf{H}^a$ is regarded as a bilinear form on $\mathcal{T}_xM$.*

**Example 1.2.25** In this example we use (1.2.62) to improve on example 1.2.21. Let $M\subset N$ be a surface in a Riemannian manifold $N$ and let $\gamma$ be a geodesic (segment) of $N$ contained entirely in $M$. Since $\gamma$ is a geodesic we can choose a (co)frame in a neighborhood of $\gamma$ in $N$ with $e_1=\gamma'$, $e_2$ tangent to $M$ and relative to which the connection form $\omega_{AB}$ satisfies

$$\omega_{AB}(e_1) = 0, \qquad \text{for } \ A,B\neq 2, \tag{1.2.63}$$

along $\gamma$. The indices $A,B$ range over $1,\cdots,N$ and we have excluded only $A$ or $B=2$. That this is possible is just like making partial parallel translation which amounts to the existence of solution to the system of ordinary differential equations (1.2.63). The vanishing of $\omega_{1a}$ for $a\geq 3$ implies that $\mathsf{H}^a_{11}=0$ and consequently from (1.2.62) we obtain

$$K_M = R^M_{1212} = R^N_{1212} - \sum_{a=3}^n (\mathsf{H}^a_{12})^2. \tag{1.2.64}$$

This proves that the curvature of $M$ along $\gamma$ is bounded above by sectional curvature of $N$ relative to the plane spanned by $e_1, e_2$. It is also clear that if $e_2$ were parallel along $\gamma$ then we can assume the frame is such that $\omega_{AB}(e_1) = 0$ (along $\gamma$) for all $A, B$, and consequently $\mathsf{H}^a_{12} = 0$ which once more proves the asertion of example 1.2.21. We can say something more which will be useful in connection with the discussion of Jacobi's equation. Assume (for simplicity) that sectional cuvatures $R^N_{1212}$ along $\gamma$ are positive. If $e_2$ is *not* parallel along $\gamma$ then we have strict inequality $K_M < R^N_{1212}$. This is clear unless $\mathsf{H}^a_{12} = 0$ for all $a \geq 3$. But in this case $\det(\mathsf{H}^a) = 0$ proving that $K_M = 0 < R^N_{1212}$. ♠

# 1.3 Special Properties of Riemannian Manifolds

## 1.3.1 Spaces of Constant Curvature

A Riemannian manifold $M$ such that the sectional curvatures $R_{ijij}$ are independent of the indices $i, j$ and constant on $M$ is called a space of *constant curvature*. Spaces of constant curvature are the simplest non-Euclidean spaces and have special properties which warrants their separate investigation. Spheres, with metric induced from the ambient Eulidean spaces, are spaces of constant positive curvature $R_{ijij} > 0$. We now construct Riemannian manifolds of constant negative curvature. In this construction we make use of the Lorentz metric which is of interest in relativity as well. By a *Lorentz metric* on a manifold $M$ we mean a symmetric contravariant 2-tensor which is everywhere nondegenrate and has signature $(m-1, 1)$ (i.e., $m-1$ positive and one negative eigenvalue). The simplest example of a Lorentz metric is $d\sigma^2 = dx_1^2 + \cdots + dx_m^2 - dx_{m+1}^2$ on $\mathbb{R}^{m+1}$. Consider the hypersurface $\mathcal{H}'$ (sometimes called the *hypersphere*) in $\mathbb{R}^{m+1}$ defined by

$$x_1^2 + \cdots + x_m^2 - x_{m+1}^2 + 1 = 0.$$

$\mathcal{H}'$ has two connected components corresponding to $x_{m+1} > 0$ and $< 0$. Let $\mathcal{H}_m$ denote either component, say for definiteness $x_{m+1} > 0$. First we show that $d\sigma^2_{|\mathcal{H}_m}$ is a Riemannian metric. By a simple application of the implicit function theorem the vectors

$$(1, 0, \cdots, 0, \frac{x_1}{x_{m+1}}), \cdots, (0, \cdots, 0, 1, \frac{x_m}{x_{m+1}}),$$

form a basis for the tangent space to $\mathcal{H}_m$ at $x = (x_1, \cdots, x_{m+1})$. Therefore writing a general tangent vector to $\mathcal{H}_m$ in the form $\tau = (x_{m+1}\xi_1, \cdots, x_{m+1}\xi_m, x_1\xi_1 + \cdots + x_m\xi_m)$ we obtain

$$\begin{aligned} d\sigma^2(\tau, \tau) &= x_{m+1}^2(\xi_1^2 + \cdots + \xi_m^2) - (x_1\xi_1 + \cdots + x_m\xi_m) \\ &\geq x_{m+1}^2(\xi_1^2 + \cdots + \xi_m^2) - (x_1^2 + \cdots + x_m^2)(\xi_1^2 + \cdots + \xi_m^2) \\ &= \xi_1^2 + \cdots + \xi_m^2 \end{aligned}$$

which proves positive definiteness of $d\sigma^2_{|\mathcal{H}_m}$.

In analogy with the Riemannian case we consider moving frames $e_1, \cdots, e_{m+1}$ which are orthonormal relative to the Lorentz metric, i.e.,

$$d\sigma^2(e_A, e_B) = 0 \quad \text{if } A \neq B, \quad d\sigma^2(e_i, e_i) = 1, \quad d\sigma^2(e_{m+1}, e_{m+1}) = -1;$$

where we recall the notational convention $1 \leq A, B, \cdots \leq m+1$, and $1 \leq i, j, \cdots \leq m$. Now assume that the vectors $e_1, \cdots, e_m$ are tangent to $\mathcal{H}_m$. Then proceeding as in the Riemannian case we have

$$dx = \sum_i \omega_i e_i, \quad de_A = \sum_B \omega_{BA} e_B.$$

The 1-forms $\omega_{AB}$ satisfy the identities

$$\omega_{ij} + \omega_{ji} = 0; \quad \omega_{i\ m+1} = \omega_{m+1\ i}, \quad \omega_{AA} = 0;$$

and since $\omega_{m+1}$ vanishes on $\mathcal{H}_m$, we have the following identities on $\mathcal{H}_m$:

$$d\omega_i + \sum_j \omega_{ij} \wedge \omega_j = 0, \qquad \sum_i \omega_{m+1\ i} \wedge \omega_i = 0.$$

Similarly the curvature of $\mathcal{H}_m$ is given by

$$\Omega_{ij} = d\omega_{ij} + \sum_k \omega_{ik} \wedge \omega_{kj} = -\omega_{i\ m+1} \wedge \omega_{m+1\ i}.$$

Now it is a simple exercise to prove that the unit normal to $\mathcal{H}_m$ at a point $x = (x_1, \cdots, x_{m+1})$, relative to the Lorentz metric, is the vector $x$. Therefore the situation is entirely analogous to that of the ordinary sphere in Euclidean space where $e_{m+1} = x$, $\omega_{i\ m+1} = \omega_i$ and by (1.2.11)

$$\Omega_{ij} = -\omega_{i\ m+1} \wedge \omega_{m+1\ j} = \omega_i \wedge \omega_j.$$

Therefore $\mathcal{H}_m$ has constant negative sectional curvature $-1$.

**Exercise 1.3.1** *By considering the stereographic projection of $\mathcal{H}_m$ onto the unit disc, or otherwise, show that the sectional curvatures of the metric*

$$4\frac{dx_1^2 + \cdots + dx_m^2}{(1 - (x_1^2 + \cdots + x_m^2))^2}$$

*are $-1$. More generally, show that all sectional curvatures of the metric $4\frac{dx_1^2+\cdots+dx_m^2}{(1+K(x_1^2+\cdots+x_m^2))^2}$ are the constant $K$. Generalize exercise 1.3.6 by showing that the volume of the ball in $\mathcal{H}_m$ increases exponentially with radius. Describe the geodesics through* $\mathbf{0}$.

**Example 1.3.1** As another application of the Lorentz metric we derive the fundamental formulae of hyperbolic trigonometry. First we derive a formula for the length of the arc of the hyperbola $\mathcal{H}_1 : z^2 - x^2 = 1$ between two points $A$ and $B$ relative to the Lorentz metric $dx^2 - dz^2$ in the plane. Since the hyperbola $\mathcal{H}_1$ and the Lorentz metric are invariant under the transformations $\begin{pmatrix} \cosh\psi & \sinh\psi \\ \sinh\psi & \cosh\psi \end{pmatrix}$, we may assume $A = (0, 1)$ and $B = (\sinh\beta, \cosh\beta)$. It is a simple calculation that the desired arc length is $\beta$. Denote the Lorentz inner product on $\mathbb{R}^3$ by $\ll, \gg$, i.e., for $v = (x, y, z)$, $v' = (x', y', z')$, then $\ll v, v' \gg = xx' + yy' - zz'$, and let $SO(2,1)$ be the group of linear transformations leaving $\ll , \gg$ invariant. Now $\mathcal{H}_2$ and the Lorentz metric $d\sigma^2 = dx^2 + dy^2 - dz^2$ are also invariant under $SO(2,1)$. Let $A, B \in \mathcal{H}_2$. We show that the distance between $A$ and $B$ relative to the Riemannian metric $d\sigma^2_{\mathcal{H}_2}$ is given by

$$d(A,B) = \cosh^{-1}(- \ll A, B \gg). \tag{1.3.1}$$

We may assume $A = (0,0,1)$ and $B = (\sinh\beta, 0, \cosh\beta)$ in view of the invariance of both sides under $SO(2,1)$. Since geodesics through $A$ are, after proper parametrization, intersections of the planes $ax + by = 0$ with $\mathcal{H}_2$ (use e.g. idea of example 1.2.9), the preceding calculation for the hyperbola $\mathcal{H}_1$ is applicable and the required formula for the distance follows. Now let $ABC$ be a geodesic triangle (i.e., the sides are geodesics). We may assume $A = (0,0,1)$, $B = (\sinh\psi, 0, \cosh\psi)$ and $C = (\cos\theta\sinh\phi, \sin\theta\sinh\phi, \cosh\phi)$ after transformation by an element of $SO(2,1)$. Set $a = d(B,C)$ then from (1.3.1)

$$\cosh a = \cosh\phi\cosh\psi - \cos\theta\sinh\phi\sinh\psi. \tag{1.3.2}$$

It is trivial to see that $\cos\theta = \cos A$. Since $\phi = d(A,C)$ and $\psi = d(A,B)$ by (1.3.1), it is customary to replace $\phi$ by $b$ and $\psi$ by $c$ so that (1.3.2) takes the familiar form (*law of cosines for hyperbolic triangles*):

$$\cosh a = \cosh b\cosh c - \cos A\sinh b\sinh c. \tag{1.3.3}$$

From (1.3.3) it follows that

$$\sin^2 A = \frac{1 - \cosh^2 a - \cosh^2 b - \cosh^2 c + 2\cosh a\cosh b\cosh c}{\sinh^2 b\sinh^2 c},$$

whence

$$\frac{\sin A}{\sinh a} = \frac{\sin B}{\sinh b} = \frac{\sin C}{\sinh c}$$

which is the *law of sines for hyperbolic triangles.* ♠

Similarly, one proves

**Exercise 1.3.2** *Prove the fundamental formulae of spherical trigonometry, namely,*

$$\cos a = \cos b\cos c + \cos A\sin b\sin c, \qquad \frac{\sin A}{\sin a} = \frac{\sin B}{\sin b} = \frac{\sin C}{\sin c}.$$

A consequence of the law of sines for hyperbolic triangles is

**Corollary 1.3.1** *Similar geodesic triangles in the hyperbolic plane are congruent.*

**Proof** - Let $ABC$ and $A'B'C'$ be *similar* geodesic triangles in the hyperbolic planes, i.e., their angles are equal. From the law of sines it follows that

$$\sinh a' = \lambda \sinh a, \quad \sinh b' = \lambda \sinh b, \quad \sinh c' = \lambda \sinh c,$$

where $a', b'$ and $c'$ are the lengths of the sides opposite to $A', B'$ and $C'$ respectively. This implies that if by an isometry we move the vertex $A'$ to $A$ such that the sides $b'$ and $c'$ are along the same geodesics as $b$ and $c$ respectively[7], then one of the triangles $ABC$ and $A'B'C'$ will contain the other depending on whether $\lambda \geq 1$ or $\lambda \leq 1$. By the Gauss-Bonnet theorem (curvature is constant)

$$\int_{ABC} dv = \int_{A'B'C'} dv$$

since each side is negative the excess of the sum of the exterior angles of the corresponding triangle. Therefore $\lambda = 1$ ♣

Corollary 1.3.1 is in sharp constrast to the case of Euclidean space where there is a profusion of similar triangles. The same conclusion holds for geodesic triangles on the surface of a sphere (see exercise 1.3.3 below).

**Exercise 1.3.3** *Show that similar geodesic triangles on the surface of a sphere are congruent.*

**Exercise 1.3.4** *Generalize the formula (1.3.1) of example 1.3.1 to $\mathcal{H}_m$.*

Using example **??** it is not difficult to extend the divergence property of geodesics on surfaces of negative curvature to general Riemannian manifolds:

**Exercise 1.3.5** *Extend the inequality (1.2.28) to general Riemannian manifolds of non-positive curvature ($R_{ijij} \leq 0$ for all $i, j$).*

[7]By a fractional linear transformation we can only ascertain that $b'$ lies along the same geodesic as $b$ or $c$ and $c'$ along the other. If $b'$ lies along $c$ we compose the isometry with the reflection with respect to the bisector of the angle at $A$. If $A$ is the point $i$ in the upper half plane and the bisector of the angle is the $y$-axis, then the reflection is given by $x + iy \to -x + iy$. Since a conformal orientation preserving automorphism of the upper half plane is a fractional linear transformation, the group of isometries of $\mathcal{H}$ contains $SL(2, \mathbb{R})/\pm I \simeq SO(2, 1)$ as a subgroup of index two.

**Example 1.3.2** Consider the upper half plane $\mathcal{H} = \{z = x + iy \in \mathbb{C} | y > 0\}$ with the Poincaré metric

$$ds^2 = \frac{dx^2 + dy^2}{y^2}.$$

This metric is invariant under the action of $SL(2, \mathbb{R})$ through fractional linear transformations

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} : z \longrightarrow \frac{az + b}{cz + d}.$$

Therefore $SL(2, \mathbb{R})/\pm I$ is a group of isometries of the hyperbolic plane. Note also that the matrix $\begin{pmatrix} \sqrt{y} & x/\sqrt{y} \\ 0 & 1/\sqrt{y} \end{pmatrix}$ maps the point $i$ to the point $z = x + iy$ so that $SL(2, \mathbb{R})$ acts transitively on $\mathcal{H}$. The isotropy subgroup at $i$ is the rotation group $SO(2) = \{\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\}$, and therefore $\mathcal{H} = SL(2, \mathbb{R})/SO(2)$. From exercise 1.2.10 we know that straight lines orthogonal to the real axis are geodesics, and more precisely the curves $t \to c + ie^t$ are are geodesics. From elementary geometry we know that under fractional linear transformations straight lines and circle are mapped to each other and (Euclidean) angles are preserved. Therefore the straight lines $t \to c + ie^t$ and semi-circles (after parametrization by a multiple of arc-length) intersecting the real axis orthogonally are geodesics. Since every tangent vector at $z \in \mathcal{H}$ is tangent to a semi-circle through $z$ intersecting the $x$-axis orthogonally or to the the straight line $t \to \Re(z) + ie^t$, all geodesics are of this form. This in particular implies that $\mathcal{H}$ is complete. Another consequence of this observation is that the isotropy subgroup at $i$, namely $SO(2)$, acts transitively on the set of geodesics through $i$. Consequently, $SL(2, \mathbb{R})$ acts transitively on the set of geodesics of $\mathcal{H}$, and the isotropy subgroup at the geodesic $t \to ie^t$ is $\pm I$. Therefore we can identify $SL(2, \mathbb{R})/\pm I$ with the set of geodesics of $\mathcal{H}$ or equivalently with the unit tangent bundle of $\mathcal{H}$ (the *unit tangent bundle* of a Riemannian manifold $M$ is $\mathcal{T}_1 M = \{(x, \xi) \in \mathcal{T}M | ds^2(\xi, \xi) = 1\}$).

From this description of geodesics it is trivial to see directly that every pair of points $z, w \in \mathcal{H}$ can be joined by a unique geodesic $\gamma_{z,w} : [0, 1] \to \mathcal{H}$ with $\gamma_{z,w}(0) = z$ and $\gamma_{z,w}(1) = w$. We denote the length of this geodesic by $d(z, w)$ call it the *hyperbolic distance* between $z$ and $w$. It is trivial to see that $d(i, iy) = \log y$ for $y > 1$. To compute $d(z, w)$ we first make the observation that the cross ratio

$$c(z, w) = \frac{(z - w)(\bar{z} - \bar{w})}{(\bar{z} - w)(z - \bar{w})}$$

is invariant under fractional linear transformations. Then it is trivial to see that

$$d(z,w)=\log\frac{1-\sqrt{c(z,w)}}{1+\sqrt{c(z,w)}}.$$

The transformation $z\rightarrow\frac{z-i}{z+i}$ maps $\mathcal{H}$ onto the unit disc $\mathcal{D}\subset\mathbb{C}$ and transforms the Riemannian metric into

$$ds^2=4\frac{dzd\bar{z}}{(1-|z|^2)^2}.$$

One refers to $\mathcal{D}$ with this metric as the *hyperbolic disc*. Clearly the geodesics in the hyperbolic disc are straight lines through the origin and semi-circles orthogonal to the unit circle (which is the boundary of $\mathcal{D}$). In terms of the coordinates of the hyperbolic disc, the length of the geodesic joining 0 to $z\in\mathcal{D}$ is $d(0,z)=\log\frac{1+|z|}{1-|z|}$, and the

$$d(z,w)=\log\frac{1+t}{1-t},$$

where $t=\frac{|w-z|}{|1-\bar{z}w|}$. ♠

**Exercise 1.3.6** *Prove that the area of the disc of radius $r>0$ in $\mathcal{D}$ is $4\pi\sinh^2\frac{r}{2}$.* (This is a special case of the fact that the volume of a ball in a Riemannian manifold with sectional curvatures bounded above by a negative constant increases exponentially with radius.)

**Exercise 1.3.7** *Let $z\in\mathcal{H}$. Let $\mathcal{S}_z(r)$ denote the non-Euclidean circle of radius $r>0$ centered at $z$. Show that $\mathcal{S}_z(r)$ is a circle in the sense of Euclidean geometry and determine its Euclidean radius and center.*

The following two exercises appear in the work of Lobachevsky and are of some interest in Gromov's theory.

**Exercise 1.3.8** *Let $L$ be a complete geodesic in the hyperbolic plane (i.e., a semi-circle orthogonal to the $x$-axis or a straight line parallel to the $y$-axis), and $r_1$ and $r_2$ be the end points of $L$. Let $z\notin L$ and $L_1$ and $L_2$ be the unique geodesics through $z$ tending to $r_1$ and $r_2$ respectively (i.e., $L_1$ and $L_2$ are the extreme geodesics through $z$ that do not intersect $L$). Let $\alpha$ be the angle of intersection of $L_1$ and $L_2$ at $z$, and $\delta$ be the distance of of $z$ to $L$. Show that*

$$\tan\frac{\alpha}{2}=\frac{1}{\sinh\delta}.$$

(Use fractional linear transformations to put $L$ and $z$ in nice locations.)

**Exercise 1.3.9** *Let $\Delta$ be a geodesic triangle in the upper half plane with edges $\alpha, \beta$ and $\gamma$. For $p \in \alpha$, let $d(p) = \inf_q d(p,q)$ where $\inf$ is taken on all $q \in \beta \cup \gamma$. Show that there is a number $\rho < \infty$ and independent of $\Delta$ such that*

$$d(p) \leq \rho.$$

**Exercise 1.3.10** *Let $z, w \in \mathcal{D}$ and denote the intersections of the geodesic connecting $z$ and $w$ with the unit circle (i.e., $\partial\mathcal{D}$) by $z'$ and $w'$. Let us assume that moving from one end of the geodesic to the other we encounter these points in the order $w', z, w, z'$. Define*

$$D(z,w,z',w') = \frac{(z-z')(w-w')}{(z-w')(w-z')}.$$

*Show that $d(z,w) = \log D(z,w,z',w')$.*

### 1.3.2 Decomposition of the Curvature Tensor

The curvature tensor admits of certain decompositions which conceptually are easier to understand in the context of representations $GL(m,\mathbb{R})$, $O(m)$ and $SO(m)$ which we alluded to in chaper 1. With the usual notation, let $V \simeq \mathbb{R}^m$, $T^k(V)$ denote the $k^{\text{th}}$ tensor power of $V$, and $S^k(V)$ the $k^{\text{th}}$ symmetric power of $V$. We fix an inner product $<,>$ on $V$ which we may assume to be the standard one on $\mathbb{R}^m$. Clearly $S^2(\bigwedge^2 V) \subset T^4(V)$, and $S^2(\bigwedge^2 V)$ is invariant under the induced action of $GL(m,\mathbb{R})$. For $\dim V = 4$, the structure of $S^2(\bigwedge^2 V)$ as a $GL(4,\mathbb{R})$, $O(4)$ or $SO(4)$-module was analyzed in examples **??** and **??** of chapter 1. Let $W$ be the representation space for the irreducible representation $\rho_T$ of $GL(m,\mathbb{R})$ determined by the Young diagram or partition $T : 4 = 2+2$. Then $\dim(W) = \frac{m^2(m^2-1)}{12}$. Denote by $T'$ the partition $4 = 1+1+1+1$, so that $\rho_{T'}$ is the fourth exterior power representation of the natural representation $\rho_1$ of $GL(m,\mathbb{R})$. We have

**Lemma 1.3.1** *$S^2(\bigwedge^2 V)$ has the decomposition*

$$S^2(\bigwedge^2 V) \simeq W \oplus \bigwedge^4 V,$$

*into irreducible $GL(m,\mathbb{R})$-modules via the representations $\rho_T$ and $\rho_{T'}$.*

**Proof** - It is not difficult to see that $S^2(\bigwedge^2 V)$ contains a copy of $W$ (consider e.g., $(e_1 \wedge e_2) \otimes (e_1 \wedge e_2)$) which we again denote by $W$, and $\bigwedge^4 V$ since $(x \wedge y) \wedge (z \wedge w) = (z \wedge w) \wedge (x \wedge y)$. The required result follows for dimension reasons. ♣

The symmetries of the curvature tensor $R_{ijkl}$, i.e., $R_{ijkl} + R_{jikl} = 0$ and $R_{ijkl} = R_{klij}$ suggest that we should arrange the components $R_{ijkl}$ as an $\frac{m(m-1)}{2} \times \frac{m(m-1)}{2}$ symmetric matrix where the rows (or columns) of the matrix are enumerated as $(1,2), (1,3), \cdots, (m-1,m)$ corresponding to the basis $e_1 \wedge e_2, e_1 \wedge e_3, \cdots, e_{m-1} \wedge e_m$ of $\wedge^2 V$.

We can use our knowledge of representations of $GL(m,\mathbb{R})$ and $\mathcal{S}_4$ to give a group theoretic interpretation to the first Bianchi identity. Let $\beta$ be the irreducible representation of degree two of $\mathcal{S}_4$ corresponding to the partition $4 = 2 + 2$ (see example **??** of chapter 1). Since $\deg(\beta) = 2$, the subspace $Z_T$ of theorem **??** of chapter 1 is

$$Z_T \simeq W \otimes \mathbb{R}^2 \simeq W \oplus W',$$

where $W'$ is a complementary subspace to $W$ in $Z_T$, and $W$ and $W'$ are necessarily isomorphic as $GL(m,\mathbb{R})$-modules. Observe that the eigenvalues of the matrices $\beta((123))$ or $\beta((132))$ are the third roots of unity $\zeta \neq 1$ and $\zeta^2$. It follows that for every vector $v \in \mathbb{R}^2$ we have

$$v \ + \ \beta((123))v \ + \ \beta((132))v \ = \ \mathbf{0}, \tag{1.3.4}$$

which is also a consequence of Schur's orthogonality relations. Therefore the first Bianchi identity is satisfied for every vector $w \in W \oplus W'$. On the other hand, since a cyclic permutation of three letters has sign $+1$, the first Bianchi identity is not valid for non-zero elements of $\bigwedge^4 V$. Therefore we have the following interpretation of first Bianchi identity:

**Lemma 1.3.2** *The curvature tensor takes values in the irreducible $GL(m,\mathbb{R})$-module $W$.*

Since $\bigwedge^4 V = 0$ for a vector space $V$ of dimension three, Bianchi identity for three dimensional Riemannian manifolds is a consequence of the symmetry properties of the curvature tensor. Of course, this fact is easily verified directly.

To decompose the curvature tensor we look at the action of the orthogonal group. Recall that for every pair of indices $i \neq j$ we defined in chapter 1, §5.2, the $O(V)$-equivariant trace map $\mathrm{Tr}_{ij} : T^k(V) \rightarrow T^{k-2}(V)$. In view of the symmetries of the curvature tensor $R = (R_{ijkl})$ we have

$$\mathrm{Tr}_{12}(W) = \mathrm{Tr}_{34}(W) = 0 \quad \text{and} \quad \mathrm{Tr}_{13}(R) = \mathrm{Tr}_{24}(R) = K,$$

where $K = (K_{ij})$ is the Ricci tensor. We denote the restriction of $\mathrm{Tr}_{13} = \mathrm{Tr}_{24}$ to $S^2(\wedge^2 V)$ and $W$ by the same letter $\kappa$ and call it the *Ricci map*. The symmetry properties of the curvature tensor imply that $\kappa$ takes values in the space of symmetric matrices. In view of the decomposition $S^2(V) = \mathbb{R} \oplus S^2_\circ(V)$, where $S^2_\circ(V)$ is the space of symmetric trace zero matrices, the Ricci tensor $K$ admits of the further $O(V)$-equivariant decomposition

$K = \frac{\mathrm{Tr}(K)}{m} I + K'$ where $\mathrm{Tr}(K') = 0$. We call $R = \mathrm{Tr}(K)$ the *scalar curvature* and $K'$ the *traceless Ricci tensor*.

We want to construct an $O(V)$-equivariant section $\lambda : S^2(V) \to W_1$. (By a section we mean $\kappa\lambda(K) = K$ for every symmetric tensor (matrix) $K$.) The reason for constructing such a section is to gain a better understanding of the curvature tensor under a conformal change of the metric, as will be shown shortly. We noted in chapter 1, §5.2, that $\ker(\kappa)$ is an $O(V)$-irreducible module, and we have the decomposition

$$W_1 \simeq \ker(\kappa) \oplus \mathbb{R} \oplus S^2_\circ(V) \tag{1.3.5}$$

into inequivalent $O(V)$-irreducible subspaces. It follows that the desired section $\lambda$ is unique. To explicitly construct $\lambda$ it is convenient to write $\lambda = \lambda_1 \oplus \lambda_2$ where $\lambda_1$ (resp. $\lambda_2$) is defined on $\mathbb{R}$ (resp. $S^2_\circ(V)$).

The sections $\lambda_i$ are easily constructed by using the notion of Young symmetrizer. Naturally we consider the Young symmetrizer for the partition $T : 4 = 2 + 2$. Now

$$\begin{aligned}\mathbb{C}_T \;=\;& (1) - (12) - (34) + (12)(34) + (13)(24) - (1324) - (1423) + (14)(23) + \\ & (13) - (132) - (143) + (1432) + (24) - (124) - (234) + (1234).\end{aligned}$$

To construct the sections $\lambda_i$ we take the $O(V)$-fixed nontrivial 2-tensor $I = (\delta_{ij})$, tensor it with $K'$ and $I$, and apply the Young symmetrizer $\mathbb{C}_T$ to get it into $W$. This procedure gives an $O(V)$-equivariant linear mapping into $W$. In other words, we set

$$\lambda_1(I) = c_1 \mathbb{C}_T(I \otimes I), \qquad \lambda_2(K') = c_2 \mathbb{C}_T(I \otimes K'), \tag{1.3.6}$$

where the constants $c_i$ will be chosen suitably to make $\lambda_i$'s sections. In terms of components, $\lambda_i$'s are

$$\begin{aligned}\lambda_1(I)_{i_1i_2i_3i_4} &= 8c_1(\delta_{i_1i_3}\delta_{i_2i_4} - \delta_{i_1i_4}\delta_{i_2i_3}), \\ \lambda_2(K')_{i_1i_2i_3i_4} &= 4c_2(\delta_{i_1i_3}K'_{i_2i_4} + \delta_{i_2i_4}K'_{i_1i_3} - \delta_{i_2i_3}K'_{i_1i_4} - \delta_{i_1i_4}K'_{i_2i_3}).\end{aligned}$$

It is a simple calculation that for

$$c_1 = \frac{1}{8(m-1)}, \quad c_2 = \frac{1}{4(m-2)}, \tag{1.3.7}$$

$\lambda_1$ and $\lambda_2$ are sections. We fix these values for $c_1$ and $c_2$.

As an application of the above algebraic analysis we discuss the Weyl conformal curvature tensor. Recall from example 1.2.19 that the curvature of the conformally flat metric $du^2 = e^{2\sigma}(dx_1^2 + \cdots + dx_m^2)$ is given by (1.2.45) and (1.2.46). It is a simple calculation that

$$\lambda(L)_{ijkl} = S_{ijkl}, \tag{1.3.8}$$

in the notation of (1.2.45) and (1.2.46). This means that the change in curvature due to conformal change of the metric lies entirely in $\mathbb{R} \oplus S^2_\circ(V)$ in the notation of (1.3.5). From formal additivity of the curvature tensor under conformal change of metric and (1.3.8) it follows that

$$C_{ijkl} = R_{ijkl} - \lambda(K)_{ijkl} \tag{1.3.9}$$

is invariant under a conformal change of the metric. $C = (C_{ijkl})$ is called the *Weyl conformal (curvature) tensor* and is the conformally invariant part of the curvature tensor. For future reference we summarize the above analysis as

**Lemma 1.3.3** *With the above notation, the Weyl conformal tensor is invariant under a conformal change of the Riemannian metric $ds^2$.*

**Exercise 1.3.11** *Show that for three dimensional Riemannian manifolds, the Weyl conformal tensor vanishes (for dimension reasons).*

**Example 1.3.3** The analysis of the representation $\rho_T$ for the case $\dim(V) = 4$ given in the subsection on Young diagrams in chapter 1, §5.2, has geometric implications. The map $\kappa$ of example **??** of chapter 1 is identical with the Ricci map. The decomposition of $\ker(\kappa) \simeq W_1' \oplus W_1''$ into $\pm 1$ eigenspaces of $E$ implies that the Weyl conformal curvature tensor $C$ admits of the decomposition $C = C^+ + C^-$. $C^+$ and $C^-$ are called the *selfdual* and *anti-selfdual* components of the Weyl conformal tensor. The change of bases described in examples **??** and **??** of chapter 1 imply that the same transformation puts the curvature tensor represented as the $6 \times 6$ symmetric matrix $(R_{ijkl})$ in the form

$$\begin{pmatrix} I & E_1 \\ -E_1 & I \end{pmatrix} (R_{ijkl}) \begin{pmatrix} I & -E_1 \\ E_1 & I \end{pmatrix} = \begin{pmatrix} \mathcal{R}_1 & \mathcal{R}_2 \\ \mathcal{R}_2' & \mathcal{R}_3 \end{pmatrix},$$

where the $3 \times 3$ matrix $\mathcal{R}_2$ is completely determined by the traceless Ricci tensor, and the symmetric matrices $\mathcal{R}_1$ and $\mathcal{R}_3$ have the properties

$$\mathrm{Tr}(\mathcal{R}_1) = \mathrm{Tr}(\mathcal{R}_3); \ \mathcal{R}_1 - \frac{R}{6} I = 4C^+; \ \mathcal{R}_3 - \frac{R}{6} I = 4C^-.$$

In view of the expression for $B$ in terms of $\kappa_{ij}$'s in example **??** of chapter 1, Einstein condition for a Riemannian manifold of dimension four is equivalent to the vanishing of the matrix $\mathcal{R}_2$. While the matrices $\mathcal{R}_1 - \frac{R}{6}I$, $\mathcal{R}_3 - \frac{R}{6}I$ and $\mathcal{R}_2$ are obtained from the $C^+$, $C^-$ and $K'$, one should exercise some caution in identifying them with the Weyl conformal and Ricci tensors since, normally, a tensor is expressed relative to a frame on the base manifold, but here we have made a change of bases for $\wedge^2 V$. ♠

### 1.3.3 Some Homogeneous Spaces

Let $G \subset GL(n, \mathbb{R})$ be an analytic group and $K \subset G$ a closed connected subgroup[8]. For certain compact subgroups $K \subset G$ we would like to investigate the curvature properties of the coset space $M = G/K$ relative to a suitable Riemannian metric. Just as in the case of spaces of constant curvature, it is useful to introduce an indefinite metric. We assume there is an indefinite metric $d\sigma^2$ on $G$ which is $G$-bi-invariant, i.e., invariant under left translations and the adjoint action of $G$. The metric $d\sigma^2$ as an inner product on a subspace of $\mathcal{G}$ is denoted by $<,>$. $<,>$ is required to have the invariance property

$$< \mathrm{Ad}(g)\xi, \mathrm{Ad}(g)\eta >=< \xi, \eta >, \quad \text{or infinitesimally} \quad < [\zeta, \xi], \eta > + < \xi, [\zeta, \eta] >= 0,$$

for all $g \in G$ and left invariant vector fields $\xi, \eta, \zeta \in \mathcal{G}$.

We assume that we have orthogonal direct sum decomposition $\mathcal{G} = \mathcal{K} \oplus \mathcal{M}$ relative to $<,>$, where $\mathcal{M}$ is a subspace with the following properties:

$$[\mathcal{M}, \mathcal{M}] \subseteq \mathcal{K}, \quad \mathrm{Ad}(K)\mathcal{M} = \mathcal{M}. \tag{1.3.10}$$

We furthermore assume that $<,>$ is positive definite on $\mathcal{M}$, however, on $\mathcal{K}$ it is either positive definite or negative definite ($K$ is compact).

Let $\dim \mathcal{G} = N$, $\{e_1, \cdots, e_m\}$ be an orthonormal basis for $\mathcal{M}$ and $\{e_{m+1}, \cdots, e_N\}$ orthonormal basis for for $\mathcal{K}$. Let $\{\omega_A\}$ be the dual basis of left invariant 1-forms. The Levi-Civita connection for the indefinite metric is defined to be a matrix of 1-forms $\omega = (\omega_{AB})$ which is skew-symmetric relative to the inner product $<,>$ and such that $d\omega_A + \sum_B \omega_{AB} \wedge \omega_B = 0$. To compute the Levi-Civita connection let $\gamma^A_{BC}$ be the structure constants of the Lie algebra $\mathcal{G}$, i.e.,

$$[e_A, e_B] = \sum_C \gamma^C_{AB} e_C.$$

In terms of $\gamma^C_{AB}$, Jacobi identity is given by

$$\sum_D (\gamma^D_{AB}\gamma^E_{CD} + \gamma^D_{CA}\gamma^E_{BD} + \gamma^D_{BC}\gamma^E_{AD}) = 0. \tag{1.3.11}$$

[8] In practice, the condition of connectedness is a little too restrictive, but often we may assume finiteness of the number of connected components. Connectedness allows one to reduce many considerations to the level of Lie algebras. When $K$ is not connected we may have to impose the additional requirement of invariance under the finite group $K/K^\circ$ where $K^\circ$ denotes the connected component of $K$, after reduction to Lie algebras.

Since $e_A$'s are left invariant, $2d\omega(e_A, e_B) = -\omega([e_A, e_B])$. From this and the invariance property of $<,>$ it follows easily that

$$\omega_{AB} = \sum_C \gamma^A_{CB}\omega_C$$

is the Levi-Civita connection. By (1.2.15) we have

$$\nabla_{e_A} e_B = \frac{1}{2}[e_A, e_B], \tag{1.3.12}$$

and consequently by proposition 1.2.6 and the Jacobi identity we obtain

$$2\Omega(e_A, e_B)e_C = -\frac{1}{4}[e_C, [e_A, e_B]]. \tag{1.3.13}$$

Equivalently, substituting in $\Omega_{AB} = d\omega_{AB} + \sum \omega_{AC} \wedge \omega_{CB}$ and using (1.3.11) we obtain

$$\Omega_{AB} = -\sum_{D,E}(\sum_C \gamma^C_{DB}\gamma^A_{EC})\omega_D \wedge \omega_E. \tag{1.3.14}$$

In view of the $G$-invariance of the metric the Ricci curvature is given by

$$R_{AB} = \frac{1}{4}\sum_C < [e_A, e_C], [e_B, e_C] > \tag{1.3.15}$$

Our calculations were done at the level of the group $G$ or its Lie algebra, rather than on the homogeneous space $M = G/K$. The algebraic structure of $G$ enabled us to carry out these calculations very simply by using the inner product $<,>$ on $\mathcal{G}$. To obtain the curvature of $M$ simply restrict the indices $1 \leq A, B, C, \cdots \leq N$ to the range $1 \leq i, j, l, \cdots \leq m$. Notice for example that $[e_i, e_j] \in \mathcal{K}$ so that the extension of $<,>$ to $\mathcal{G}$ is essential.

**Exercise 1.3.12** *Show that if the bi-invariant metric $d\sigma^2$ is positive definite then sectional curvatures of $M$ are non-negative. Furthermore, if $d\sigma^2$ is negative definite on $\mathcal{K}$, the sectional curvatures of $M$ are non-positive.*

Let us apply these considerations to some concrete cases.

**Example 1.3.4** Let $G = U(n+k)$ and $K = U(k) \times U(n)$ so that $M = G/K$ is the complex Grassman manifold $\mathbf{G}_{k,n}$. Here $G$ is compact, and the indefinite $d\sigma^2$ on $\mathcal{G}$ given by

$$< \xi, \eta >= -\frac{1}{2}\mathrm{Tr}(\xi\eta),$$

where $\xi$ and $\eta$ are identified with skew hermitian matrices, is already Riemannian. Let $E_{jk}$ be the matrix with 1 at the $(j,k)^{\text{th}}$ spot and zeros elsewhere, then the matrices

$$i(E_{jp}+E_{pj}), \quad E_{jp}-E_{pj} \quad \text{for} \;\; 1 \leq j \leq k,\; k+1 \leq p \leq n+k,$$

form a basis for $\mathcal{M}$. For $e_A = i(E_{jp}+E_{pj})$ and $e_B = E_{jp}-E_{pj}$ we see

$$< \Omega(e_A, e_B)e_A, e_B >= -4,$$

i.e., sectional curvature of the plane spanned by $e_A$ and $e_B$ is 4. Similarly, sectional curvatures of the planes spanned by the vectors $\{i(E_{jp}+E_{pj}), E_{jq}-E_{qj}\}$, $\{i(E_{jp}+E_{pj}), i(E_{jq}+E_{qj})\}$, $\{E_{jp}-E_{pj}, E_{jq}-E_{qj}\}$, $\{i(E_{jp}+E_{pj}), E_{lp}-E_{pl}\}$, $\{i(E_{jp}+E_{pj}), E_{jq}-E_{qj}\}$, or $\{E_{jp}-E_{pj}, E_{jq}-E_{qj}\}$, for $p \neq q$ and $j \neq l$ is 1. The sectional curvature of plane spanned by $\{i(E_{jp}+E_{pj}), E_{lq}-E_{ql}\}$ for $p \neq q$ and $j \neq l$ is 0. In particular, for the complex projective space (i.e., $k=1$) the sectional curvatures are either 1 or 4. To compute the Ricci tensor we make use of (1.2.38). $G$-invariance of the metric implies that the Ricci tensor is fixed by transformations $A \in U(k) \times U(n)$, i.e., $\rho(A^{-1})(R_{ik})\rho(A) = (R_{ik})$ where $\rho$ denotes the adjoint action of $K = U(k) \times U(n)$ on $\mathcal{M}$. Since $\rho$ is irreducible (as a complex representation), $(R_{ik})$ is a multiple of identity. Now it is an easy computation to see that

$$(R_{ik}) = 2(2+k+n)I.$$

In particular $M$ is Einstein with scalar curvature $2(2+k+n)$. ♠

**Example 1.3.5** Let $J_{k,n} = \begin{pmatrix} -I_k & 0 \\ 0 & I_n \end{pmatrix}$ where $I_k$ is the $k \times k$ identity matrix, $G = U(k,n)$ be the unitary group of $J_{k,n}$, i.e., the set of complex invertible matrices $U \in GL(n+k;\mathbb{C})$ such that $\bar{U}'J_{k,n}U = J_{k,n}$. (Here bar and prime denote complex conjugate and transpose of the matrix.) This condition is equivalent to the relations

$$-\bar{A}'A + \bar{C}'C = -I_k, \quad -\bar{B}'A + \bar{D}'C = \mathbf{0}, \quad -\bar{B}'B + \bar{D}'D = I_n, \tag{1.3.16}$$

where $U = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$. We set $K = U(k) \times U(n) \subset G$, then $M = G/K$ can be identified with the generalized disc $D_{k,n}$ of $k \times n$ matrices $Z$ such that $I_n - \bar{Z}'Z$ is positive definite. In fact for an element $Z \in D_{k,n}$ the action of $G$ is given by

$$Z \longrightarrow (AZ+B)(CZ+D)^{-1}.$$

The isotropy subgroup of zero matrix $\mathbf{0} \in D_{k,n}$ is $K$, and transitivity of the action of $G$ on $D_{k,n}$ is a simple exercise in linear algebra. However, it remains to show that $CZ+D$ is invertible and $D_{k,n}$ is invariant under $G$. Let $W = (AZ+B)(CZ+D)^{-1}$, then using (1.3.16) it is easily verified that

$$(\bar{Z}'\bar{C}' + \bar{D})^{-1}(I_n - \bar{Z}'Z)(CZ + D)^{-1} = (I_n - \bar{W}'W)$$

which is valid for the open dense set $\det(CZ+D) \neq 0$. Since $D_{k,n}$ is a bounded domain, $Z \rightarrow \det(CZ+D)^{-1}$ is bounded for an open dense subset of $D_{k,n}$, and therefore $\det(CZ+D) \neq 0$ for all $Z \in D_{k,n}$ and $G$ maps $D_{k,n}$ to itself. Proceeding as in the preceding example, we define the indefinite metric $d\sigma^2$ on the Lie algebra $\mathcal{G}$ of $G$ by

$$< \xi, \eta >= \frac{1}{2}\mathrm{Tr}(\xi\eta),$$

$\mathcal{G}$ consists of matrices skew hermitian relative to $J_{k,n}$, i.e., matrices $X$ satisfying $\bar{X}'J_{k,n} + J_{k,n}X = \mathbf{0}$. Let $\mathcal{K}$ be the Lie algebra of $K$ and $\mathcal{M}$ be its orthogonal complement. Then

$$E_{jp} + E_{pj}, \quad i(E_{jp} - E_{pj}) \quad for \;\; 1 \leq j \leq k, \; k+1 \leq p \leq n+k,$$

is a basis for $\mathcal{M}$, $d\sigma^2$ is positive definite on $\mathcal{M}$ and negative definite on $\mathcal{K}$. Now computing as before we see that the sectional curvatures of $D_{k,n}$ are 0, -1 or -4, and its Ricci tensor is the $(R_{ik}) = -2(2+k+n)I$. For $k = 1$ the sectional curvatures are either -1 or -4. ♠

Since it is not our purpose to give an account of the theory of symmetric spaces, we mention the following example in the form of an exercise and refer the reader to [H] or [KN] for an extensive discussion of differential geometry of symmetric spaces:

**Exercise 1.3.13** *Let $G = SP(n;\mathbb{R})$ the (symplectic group) be the set of (invertible) $2n \times 2n$ real matrices $U = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ such that $U'JU = J$ where $J = \begin{pmatrix} 0 & -I_n \\ I_n & 0 \end{pmatrix}$ is the standard skew symmetric matrix. Consider the action of the $G$ on the Siegel Upper Plane*

$$\mathcal{P} = \{Z = X + iY | Z \text{ complex symmetric } n \times n \text{ matrix}, \; Y \text{ positive definite}\},$$

*given by $Z \rightarrow (AZ+B)(CZ+D)^{-1}$. Show that the isotropy subgroup at $Z = iI$ is isomorphic to the unitary group $K = U(n)$, and the action of $G$ is transitive on $\mathcal{P}$. Prove that the mapping*

$$Z \longrightarrow (I + iZ)(I - iZ)^{-1}$$

*maps $\mathcal{P}$ onto the set of complex symmetric matrices $V$ such that $I - \bar{V}V$ is positive definite. Imitating the argument of example 1.3.5 show that $CZ + D$ is invertible and $\mathcal{P}$ is in fact invariant under $G$. Thus $\mathcal{P} \simeq G/K$. Obtain the decomposition $\mathcal{G} = \mathcal{K} \oplus \mathcal{M}$. Define the bi-invariant indefinite metric on $\mathcal{G}$ by $< \xi, \eta >= \mathrm{Tr}(\xi\eta)$, and compute the sectional curvatures and the Ricci tensor of $\mathcal{P}$.*

**Example 1.3.6** Since quadrics are a rich source of examples in geometry, we consider the curvature properties of the standard complex quadric $Q \subset \mathbb{C}P(n)$ defined by the single quadratic equation $z_\circ^2 + \cdots + z_n^2 = 0$. (See also subsection on quadrics above.) It is a straighforward exercise to show that $Q$ is a connected complex manifold. The group $U(n{+}1)$ acts on $S^{2n+1} \subset \mathbb{C}^{n+1}$ and the action induces action on $\mathbb{C}P(n)$. Now $SO(n{+}1) \subset U(n{+}1)$ and the quadric $Q$ is invariant under the action of $SO(n+1)$. Let $v = (\frac{1}{\sqrt{2}}, \frac{i}{\sqrt{2}}, 0, \cdots, 0) \in S^{2n+1}$ then the image $[v]$ of $v$ in $\mathbb{C}P(n)$ lies on $Q$ and isotropy subgroup at $[v]$ is the subgroup $K$ of matrices of the form

$$\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & A \end{pmatrix},$$

where $A \in SO(n-1)$. Since $SO(n+1)/K$ is a compact submanifold of $Q$ of real dimension $2n = \dim_{\mathbb{R}}(Q)$, $Q = SO(n{+}1)/K$. Relative to the positive definite inner product $<\xi, \eta> = -\frac{1}{2}\mathrm{Tr}(\xi\eta)$ on the Lie algebra $\mathcal{U}(n+1)$, the orthogonal complement of $\mathcal{K}$ is the subspace

$$\mathcal{M} = \left\{ \begin{pmatrix} 0 & 0 & x \\ 0 & 0 & y \\ -x' & -y' & 0 \end{pmatrix} \right\},$$

where $x$ and $y$ are real row vectors. It is convenient therefore to represent elements of $\mathcal{M}$ as $\xi = (x, y)$. We thus obtain the decomposition $\mathcal{U}(n+1) = \mathcal{K} \oplus \mathcal{M}$ and the condition (1.3.10) is satisfied. Therefore we can proceed as before for the computation of the curvature of $Q$. Then for $\xi = (x, y),\ \eta = (u, v),\ \zeta \in \mathcal{M}$ we obtain $2\Omega(\xi, \eta)\zeta = \frac{1}{4}[[\xi, \eta], \zeta]$ and

$$[\xi, \eta] = 2 \begin{pmatrix} xu' - ux' & 0 & 0 \\ 0 & yv' - vy' & 0 \\ 0 & 0 & x'u + y'v - u'x - v'y \end{pmatrix}.$$

Furthermore, the Ricci tensor is $4(n-1)I$. ♠

To understand the structure of geodesics on the homogeneous spaces $M = G/K$ considered above, it is convenient to make use of the bi-invariant indefinite metric $d\sigma^2$ introduced above. First we note that the notion of parallel translation is defined relative to the Levi-Civita connection, and is therefore identical with the case of a Riemannian metric. Covariant derivative is also defined by the same formula as in the case of a Riemannian metric. Similarly a curve $\gamma$ is a geodesic if $\nabla_{\dot\gamma}\dot\gamma = 0$. However, the notion of distance and length minimizing property of geodesics do not carry over to the indefinite case. The essential observation is the following proposition:

**Proposition 1.3.1** *With respect to the indefinite metric $d\sigma^2$, the 1-parameter subgroups $\gamma_\xi(t) = \exp(t\xi)$ are the geodesics through $e \in G$. The geodesic through $g \in G$ with tangent vector at $g$ the left invariant vector field $\xi$ at $g$, is $g.\gamma_\xi(t)$. In particular, geodesics exist for all values of the parameter $t$.*

**Proof** - From the fact that $t \rightarrow \exp(t\xi)$ is a homomorphism it follows that the tangent vector field to the curve $t \rightarrow g\exp(t\xi)$ is the left invariant vector field $\xi$. Now relative to the Levi-Civita connection for the bi-invariant metric $d\sigma^2$, $\nabla_\eta\xi = \frac{1}{2}[\eta, \xi]$ for left invariant vector fields $\eta, \xi$. Therefore $\nabla_\xi\xi = 0$ and the proof of the proposition is complete. Q E D

Now assume that there is an involution $\theta : G \rightarrow G$ whose fixed point set is $K$ (see remark 1.3.1 below) and $\mathcal{M}$ is the eigenspace corresponding to eigenvalue -1 for the induced action of $\theta$ on $\mathcal{G}$. For instance, in examples 1.3.4 and 1.3.5 above the involution $\theta$ is given by $\theta(g) = J_{k,n}gJ_{k,n}$, and for the symplectic group $\theta(g) = JgJ^{-1}$. We use $\theta$ to embed $M = G/K$ into $G$. In fact, consider the mapping $\jmath : G \rightarrow G$ given by $g \rightarrow g\theta(g)^{-1}$. Clearly $G$ acts transitively on $\mathrm{Im}(\jmath)$ which makes it into a homogeneous space for $G$. Now $e \in \mathrm{Im}(\jmath)$ and the isotropy subgroup at $e$ is the fixed point set of $\theta$ which is $K$. Therefore $\mathrm{Im}(\jmath) \simeq M = G/K$ and we identify $M$ with $\mathrm{Im}(\jmath)$. $M$ is a *totally geodesic* submanifold of $G$ relative to the bi-invariant metric $d\sigma^2$, i.e., every geodesic emanating from a point in $M$ and initial tangent vector tangent to $M$, remains in $M$. Since geodesics in $G$ are left translates of 1-parameter subgroups, to show that $M$ is totally geodesic it suffices to show that for $\xi \in \mathcal{M}$ the curve $\gamma_\xi(t) = \exp(t\xi)$ lies in $M$. Since $\xi \in \mathcal{M}$, $\xi = \frac{\xi}{2} - \theta(\frac{\xi}{2})$ and $\theta(\frac{\xi}{2})$ and $\frac{\xi}{2}$ commute. Consequently

$$\gamma_\xi(t) = \exp(t\frac{\xi}{2})\exp(-t\theta(\frac{\xi}{2})) = h\theta(h)^{-1},$$

where $h = \exp(t\frac{\xi}{2})$. This shows that $M$ is a totally geodesic submanifold of $G$. The restriction of the indefinite metric $d\sigma^2$ to $M$ is Riemannian, and is precisely the metric considered earlier in examples 1.3.4 and 1.3.5. ♣

**Remark 1.3.1** The condition that $K$ is the fixed point set of $\theta$ is in general too restrictive. Normally one only requires $K$ to lie between the fixed point set of $\theta$ and its connected component which has finite index in the former group. This implies that $M$ is a finite covering of $\mathrm{Im}(\jmath)$ in the sense of chapter 4, and the local conclusions about curvature etc. remain valid. Exercise 1.3.15 below gives an example of a situation in which the fixed point set of $\theta$ has actually two connected components. ♡

**Exercise 1.3.14** *Let $H$ be a compact analytic group. Show that $H$ maybe regarded as a homogeneous space of the form $H \simeq G/K$ as described above by setting $G = H \times H$ and $K = \{(h, h)|h \in H\}$.*

**Exercise 1.3.15** *Let* $\mathbf{G}^{\circ}_{k,n}(\mathbb{R})$ *be the Grassman manifold of oriented* $k$*-planes in* $\mathbb{R}^{k+n}$ (see example of chapter 1). *Then* $\mathbf{G}^{\circ}_{k,n}(\mathbb{R}) \simeq SO(k+n)/SO(k)\times SO(n)$ *is the realization of this space as* $M = G/K$ *as above. Obtain the decomposition* $\mathcal{G} = \mathcal{K} \oplus \mathcal{M}$ *using the inner product (positive definite)* $-\frac{1}{2}\mathrm{Tr}(\xi\eta)$. *Compute the sectional curvatures and Ricci tensor of* $\mathbf{G}^{\circ}_{k,n}(\mathbb{R})$.

**Exercise 1.3.16** *By realizing the complex projective space* $\mathbb{C}P(n)$ *as* $U(n+1)/U(1)\times U(n)$ *show that the geodesics in* $\mathbb{C}P(n)$ *relative to the metric* $-\frac{1}{2}\mathrm{Tr}(\xi\eta)$ *are of the form*

$$t \longrightarrow g.[\cos t, \sin t, 0, \cdots, 0],$$

*in homogeneous coordinates, where* $g \in U(n+1)$.

**Exercise 1.3.17** *Show that a geodesic in the Siegel upper half plane can be put in the form*

$$t \longrightarrow i\exp(tD),$$

*where* $D$ *is a diagonal matrix, by a symplectic transformation* $g \in SP(n,\mathbb{R})$. *Obtain a similar result for the generalized unit disc* $D_{k,n}$.

**Exercise 1.3.18** *Let* $\mathcal{P}$ *be the space of* $n\times n$ *symmetric positive definite real matrices of determinant 1. Show that the mapping* $A \rightarrow AA'$, *where* $'$ *denotes transpose, gives the realization*

$$\mathcal{P} \simeq SL(n,\mathbb{R})/SO(n),$$

*as a homogeneous space. Show that a bi-invariant indefinite metric for* $SL(n,\mathbb{R})$ *is* $\mathrm{Tr}(\xi\eta)$. *Obtain the decomposition* $\mathcal{G} = \mathcal{K} \oplus \mathcal{M}$ *for this case and show that the relations (1.3.10) are valid. Deduce that every geodesic in* $\mathcal{P}$ *is of the form*

$$t \longrightarrow g.e^{tA}.g',$$

*where* $A$ *is a diagonal matrix with* $\mathrm{Tr}(A) = 0$, *and* $g \in SL(n,\mathbb{R})$. *Show also that the sectional curvatures of* $\mathcal{P}$ *are non-positive. Prove that the Riemannian metric on* $\mathcal{P}$ *has coordinate expression* $ds^2 = \mathrm{Tr}((UdU)^2)$ *where* $U$ *runs over symmetric positive definite matrices of determinant 1.* (Essentially the same assertions are valid for the homogeneous space $\mathcal{P}' \simeq GL(n,\mathbb{R})/O(n)$ of symmetric positive definite real matrices.)

### 1.3.4 The Laplacian

There are various ways of defining the Laplace operator $\Delta$ on a Riemannian manifold $M$. First we confine ourselves to real or complex valued forms on $M$. Let $M$ be an oriented Riemannian manifold (without boundary) and $dv$ denote the Riemannian volume element. On the space of real or complex valued compactly supported smooth functions on $M$ we define the the inner product

$$< \phi, \psi >= \int_M \phi(x)\overline{\psi(x)} dv.$$

The completion of the space of smooth functions under $< ., . >$ is the Hilbert space $L^2(M, dv)$. We want to extend the notion of inner product $< ., . >$ to forms on $M$. Let $\omega_1, \cdots, \omega_m$ denote an orthonormal coframe for $M$. Then locally a $p$-form $\beta$ is a linear combination of expressions of the form $\omega_{i_1} \wedge \cdots \wedge \omega_{i_p}$. Define the star operator $\star$ (see also chapter 1, §6.2):

$$\star(\omega_{i_1} \wedge \cdots \wedge \omega_{i_p}) = \epsilon_{i_1 \cdots i_p j_1 \cdots j_{m-p}} \omega_{j_1} \wedge \cdots \wedge \omega_{j_{m-p}}$$

where $\{j_1, \cdots, j_{m-p}\}$ is the complement of $\{i_1, \cdots, i_p\}$ in $\{1, 2, \cdots, m\}$ and $\epsilon_{i_1 \cdots i_p j_1 \cdots j_{m-p}}$ is the sign of the permutation $1 \to i_1, \cdots, m \to j_{m-p}$. It is readily verified that $\star$ is independent of the choice of orthonormal coframe $\omega_1, \cdots, \omega_m$. Clearly $\star$ extends linearly to $p$-forms, and we define the inner product of two $p$-forms $\alpha$ and $\beta$ as

$$< \alpha, \beta >= \int_M \alpha \wedge \star\beta.$$

This definition is compatible with the inner product of two functions regarded as 0-forms. If $\alpha$ is a $q$-form and $p \neq q$ then we set $< \alpha, \beta >= 0$.

**Exercise 1.3.19** *Show that for a function $\phi$ on a Riemannian manifold $M$ we have*

$$ds^2(\text{grad}\phi, \text{grad}\phi) dv = d\phi \wedge \star d\phi = ds^2(d\phi, d\phi) dv,$$

*where $dv$ denotes the volume element.*

Define the operator $\delta$ mapping a $p$-form on $M$ to a $(p-1)$-form by

$$\delta = (-1)^{mp+m+1} \star d \star .$$

Since $\star\star = \pm\text{Id.}$, $\delta\delta = 0$. Furthermore, for compactly supported $C^1$ forms $\alpha$ and $\beta$ we have

$$< d\alpha, \beta >=< \alpha, \delta\beta > . \tag{1.3.17}$$

In terms of $d$ and $\delta$ the Laplace operator $\Delta$ on forms is defined as

$$-\Delta = d\delta + \delta d.$$

The symmetric character of the Laplace operator is evident from its definition. It is easily verified that the application of Laplace operator to functions on a Riemannian manifold is given by

$$\Delta f = \mathrm{div}\ \mathrm{grad}(f) \tag{1.3.18}$$

The fact that $\Delta$ is defined on forms (not just functions) has many geometric applications some of which we will discuss in connection with cohomology. The operator $-\Delta$ is positive semi-definite on compactly supported forms since

$$< -\Delta\alpha, \alpha >=< d\alpha, d\alpha > + < \delta\alpha, \delta\alpha >\geq 0.$$

Since $\Delta$ is negative semi-definite on compactly supported functions, it is customary to refer to $-\Delta$ as the *positive Laplacian.* A $p$-form $\beta$ such that $\Delta\beta = 0$ is called *harmonic.*

**Exercise 1.3.20** *Let $M$ be a surface with Riemannian metric $ds^2 = e^{2\sigma}(du^2 + dv^2)$, and $f$ a real or complex valued function on $M$. Show that*

$$\Delta f = e^{-2\sigma}\Big(\frac{\partial^2 f}{\partial u^2} + \frac{\partial^2 f}{\partial v^2}\Big).$$

**Exercise 1.3.21** *Show that for the usual metric $ds^2 = d\varphi^2 + \sin^2\varphi d\theta^2$ on $S^2$ the Laplacian is given by*

$$\Delta f = \frac{\partial^2 f}{\partial\varphi^2} + \cot\varphi\frac{\partial f}{\partial\varphi} + \frac{1}{\sin^2\varphi}\frac{\partial^2 f}{\partial\theta^2}.$$

**Exercise 1.3.22** *Let $(g_{ij})$ be the matrix of a Riemannian metric relative to a coordinate system $(x_1, \cdots, x_m)$ on a Riemannian manifold. Denote the determinant and inverse of $(g_{ij})$ by $g$ and $(g^{ij})$ respectively. Show that (1.3.18) becomes*

$$\Delta f = \frac{1}{\sqrt{g}}\sum_k \frac{\partial}{\partial x_k}\Big(\sum_j g^{jk}\sqrt{g}\frac{\partial f}{\partial x_j}\Big).$$

To study the Laplace operator on the sphere $S^m\imath\mathbb{R}^{m+1}$ we consider the moving coframe on $\mathbb{R}^{m+1}$

$$\omega_1 = r\tilde{\omega}_1,\ \ \cdots,\ \omega_m = r\tilde{\omega}_m,\ \omega_{m+1} = dr$$

where $ds^2_{\mathbb{R}^{m+1}} = \omega_1^2 + \cdots + \omega_{m+1}^2$ and the 1-forms $\tilde{\omega}_j$ are a moving frame for the unit sphere $S^m$ and therefore do not depend on $r$. For a function $f$ on $\mathbb{R}^{m+1}$ we have

$$df = \sum_{i=1}^{m} f_i r\tilde{\omega}_i + \frac{\partial f}{\partial r} dr.$$

Therefore $\star df = \sum f_i\beta_i + \frac{\partial f}{\partial r} r^m \tilde{\omega}_1 \wedge \cdots \wedge \tilde{\omega}_m$ where the $m$-forms $\beta_i$ contain $dr$. Consequently,

$$\Delta_{\mathbb{R}^{m+1}} f = -\star d \star df = \big[\frac{\partial^2 f}{\partial r^2} + \frac{m}{r}\frac{\partial f}{\partial r}\big] + \frac{1}{r^2}\Delta_{S^m} f,$$

where $\Delta_{S^m}$ denotes the Laplacian on the unit sphere. Therefore

$$\Delta_{S^m} f = r^2 \Delta_{\mathbb{R}^{m+1}} f - \big[r^2 \frac{\partial^2 f}{\partial r^2} + mr\frac{\partial f}{\partial r}\big]. \tag{1.3.19}$$

An immediate consequence of (1.3.19) is that if $f$ is a homogeneous polynomial of degree $n$ on $\mathbb{R}^{m+1}$ such that $\Delta f = 0$ then $f$ is an eigenfunction of the Laplacian on the unit sphere with eigenvalue $-n(n+m-1)$. These polynomials or their restrictions to the unit sphere are called *spherical harmonics*. Naturally by the *degree* of a spherical harmonic we mean its degree as a polynomial. It is an important theorem in analysis that spherical harmonics contain an orthonormal basis for $L^2(S^m, dv)$ where $dv$ denotes the invariant measure on the sphere. We summarize the basic facts regarding this in the following theorem and refer to [SW] for the proof and some applications:

**Theorem 1.3.1** *The space $\mathcal{A}_k$ of spherical harmonics of degree $k$ has dimension $\binom{m+k}{k} - \binom{m+k-2}{k-2}$, where the binomial coefficient $\binom{a}{b} = 0$ for $b < 0$. $\mathcal{A}_k$ and $\mathcal{A}_l$ are orthogonal for $k \neq l$ relative to the standard $L^2$ inner product on $S^m$. Finite linear combinations of elements of $\bigcup_k \mathcal{A}_k$ are dense in $L^2(S^m, dv)$ and in the space of continuous functions on $S^m$ relative to the sup norm.*

**Example 1.3.7** To gain some understanding of spherical harmonics consider the two dimensional case. The Laplacian on $S^1$ is $\Delta = \frac{d^2}{d\theta^2}$ and its eigenfunctions are $1, \cos n\theta, \sin n\theta$. Writing $x_1 = \cos\theta$ and $x_2 = \sin\theta$, from the standard expansions $\cos n\theta$ and $\sin n\theta$ we obtain

$$\cos n\theta = x_1^n - \binom{n}{2} x_1^{n-2}x_2^2 + \binom{n}{4} x_1^{n-4}x_2^4 + \cdots, \quad \sin n\theta = \binom{n}{1} x_1^{n-1}x_2 - \binom{n}{3} x_1^{n-3}x_2^3 + \cdots$$

Regarding $x_1, x_2$ as independent Cartesian coordinates in $\mathbb{R}^2$, denoting the above expressions for $\cos n\theta$ and $\sin n\theta$ by $P_n(x_1, x_2)$ and $Q_n(x_1, x_2)$ and applying the Laplacian $\Delta = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2}$ we obtain

$$\Delta P_n = 0, \quad \Delta Q_n = 0.$$

Now assume $F(x_1, x_2) = \sum_k c_k x_1^{n-k} x_2^k$ is a homogeneous polynomial of degree $n$ with $\Delta F = 0$. The latter relation implies the recursion

$$c_{k+2} = -\frac{(n-k)(n-k-1)}{(k+1)(k+2)} c_k$$

The initial conditions $c_\circ = 1, c_1 = 0$ and $c_\circ = 0, c_1 = 1$ lead to the polynomials $P_n$ and $Q_n$. Consequently polynomial solutions of $\Delta$ in the plane are linear combinations of $P_n$'s and $Q_n$'s. Since classically harmonics refer to the trigonometric functions $\cos n\theta$ and $\sin n\theta$, this example should motivate the use of the terminology "spherical harmonic". ♠

There are a number of applications of spherical harmonics and in particular of Fourier series to geometry. Perhaps Hurwitz' proof of the isoperimetric inequality is first deep application of Fourier series to a geometric problem. Since we have already given two proofs of this fundamental result, we delegate this proof to the following exercise:

**Exercise 1.3.23** *Let $\gamma : [0, 2\pi] \rightarrow \mathbb{R}^2$ be a simple closed curve of length $L$. Assume the parameter $t \in [0, 2\pi]$ is a multiple of arc-length and $\gamma$ is oriented counterclockwise. Denote the image of $\gamma$ by $\Gamma$ and let $C$ be the region enclosed by $\Gamma$. Let $A$ denote the area of $C$.*

1. *Using Stokes' theorem show that*

$$\begin{aligned} L^2 - 4\pi A &= 2\pi \int_\circ^{2\pi} [(\frac{dx_1}{dt})^2 + (\frac{dx_2}{dt})^2 + 2x_2 \frac{dx_1}{dt})dt \\ &= 2\pi \int_\circ^{2\pi} (\frac{dx_1}{dt} + x_2)^2 dt + 2\pi \int_\circ^{2\pi} [(\frac{dx_2}{dt})^2 - x_2^2] dt, \end{aligned}$$

*where $\gamma(t) = (x_1(t), x_2(t))$.*

2. *Let $f$ be a periodic function of period $2\pi$ whose zeroth Fourier coefficient vanishes. Show that*

$$\int_\circ^{2\pi} (\frac{df}{dt})^2 dt \geq \int_\circ^{2\pi} f(t)^2 dt.$$

(This inequality is often called *Poincaré inequality.*)

3. *Since $L^2 - 4\pi A$ is invariant under translations we may assume $\int_\circ^{2\pi} x_2(t)dt = 0$. Applying (2) to $f = x_2$ and using (1) deduce the isoperimetric inequality $L^2 - 4\pi A \geq 0$.*

4. *Deduce that the circle is the only curve for which $L^2 - 4\pi A = 0$.*

**Exercise 1.3.24** *Consider the two dimensional torus $T^2 = \mathbb{R}^2/L$ where $L \subset \mathbb{R}^2$ is the lattice with basis $(\alpha, 0), (\beta, \gamma)$. Show that the functions*

$$1, \ \sin\frac{2\pi my}{\gamma}, \ \cos\frac{2\pi my}{\gamma}, \ \sin 2\pi m(\frac{x}{\alpha} - \frac{\beta y}{\alpha\gamma}), \ \cos 2\pi m(\frac{x}{\alpha} - \frac{\beta y}{\alpha\gamma}).$$

*are eigenfunctions for $\Delta$, and they form a basis for $L^2(T^2, dxdy)$. Find all eigenvalues for $\Delta$.*

**Exercise 1.3.25** *Let $M = S^m$ or $T^m$, the $m$-dimensional flat torus. Let $N(\lambda)$ denote the number of eigenvalues of $-\Delta$ which are $< \lambda$. Verify the validity of Weyl's asymptotic formula*

$$N(\lambda) \sim \frac{v_m \mathrm{vol}(M)}{(2\pi)^m}\lambda^{\frac{m}{2}},$$

*in these cases. Here $v_m$ denotes the Euclidean volume of the $m$-dimensional ball.*

A basic problem about the Laplace operator on a Riemannian manifold is the determination of its spectrum which depends strongly on the Riemannian structure. Of course the exact determination of the spectrum, except in very special cases, is not within the range of the present knowledge. Typical issues on which progress has been made are

1. For the positive Laplacian $-\Delta$ acting functions on a compact orientable Riemannian manifold, the smallest eigenvlaue is $\lambda_\circ = 0$ which occurs with mutiplicity 1. Can we give upper and lower bounds for the next eigenvalue in terms of the Riemannian structure?

2. Let $M$ be a compact orientable Riemannian manifold. For a positive real number $\lambda$, let $N(\lambda)$ denote the number of eigenvalues of $-\Delta$ less than $\lambda$. While an exact and practical formula for $N(\lambda)$ may not be feasible, much is known about its asymptotic behavior.

3. Does the spectrum of $-\Delta$ determine the Riemannian metric? The answer is negative and we will say something about this in chapter 4.

4. Let $M \subset \mathbb{R}^2$ be a bounded doamin with smooth boundary. The eigenvalues of $-\Delta$ on the space functions vanishing on the boundary determine the frequencies of the sound generated by a drum in the shape of $M$? Therefore problem 3 in this case becomes "Can we hear the shape of a drum?"

5. A generalization of problem 4 is to what extent an unknown entity (e.g. a Riemannian metric, an obstacle, a potential etc.) is determined by the observables, i.e., the spectrum. This is a fundamental problem of physics which may be classified as the Inverse Spectral or the Inverse Scattering theory.

The literature on this subject is very extensive, and our goal here is to explain some simple ideas which have proven fruitful in gaining some insight into these issues. Naturally, all methods rely heavily on techniques from analysis which are not appropriate for this text. We limit ourselves to cases where the analysis is of a rather elementary nature, can be described without invoking advanced analytical techniques, and has finite dimensional analogues which are easy to explain.

For a symmetric linear operator on a finite dimensional real Hilbert space $V$, the eigenvalues are real and denote them by $\lambda_\circ < \lambda_1 < \lambda_2 < \cdots$. The smallest eigenvalue $\lambda_\circ$ is

$$\lambda_\circ = \inf_{0\neq v\in V} \frac{|< v, Av >|}{< v, v >}, \tag{1.3.20}$$

where $< .,. >$ denotes the inner product. Let $V_j$ denote the eigenspace corresponding to eigenvalue $\lambda_j$ and $V_j^\perp$ denote its orthogonal complement. The second smallest eigenvalue $\lambda_1$ is then given by

$$\lambda_1 = \inf_{0\neq v\in V_\circ^\perp} \frac{|< v, Av >|}{< v, v >}. \tag{1.3.21}$$

Similarly by replacing $V_\circ^\perp$ with the orthogonal complement of $V_\circ \oplus \cdots \oplus V_{j-1}$ in (1.3.21) we obtain a formula for $\lambda_j$. This description of the eigenvalues of a symmetric linear operator (in finite or infinite dimensions) is known as the *variational characterization* of eigenvalues.

The variational characterization of eigenvalues of a symmetric operator in the finite dimensional case generalizes to the infinite dimensional case. We note the following plausible facts about the unbounded operator $\Delta$:

**Proposition 1.3.2** *The positive Laplacian $-\Delta$ acting on $C^2$ functions on the compact Riemannian manifold $M$ has the following properties:*

1. *The eigenvalues of $\Delta$ form a discrete set $0 = \lambda_\circ < \lambda_1 < \lambda_2 < \cdots \subset \mathbb{R}$.*

2. $\lim_{j\to\infty} \lambda_j = \infty$.

3. *The eigenspace $V_j$ corresponding to eigenvalue $\lambda_j$ is finite dimensional.*

4. *$V_\circ$ is one dimensional and consists of constants.*

5. *There is an orthonormal basis for $L^2(M, dv)$ consisting of eigenfunctions of* $\Delta$ (Completeness property).

6. *The characterizations of of eigenvalues of symmetric operator given by (1.3.20) and (1.3.21) (and similar formulae for all $\lambda_j$'s) remain valid for $\Delta$. (Orthogonal complements are relative to the inner product on $L^2(M)$.)*

Except for the facts that $\lambda_\circ$ may be $> 0$ and $\dim V_\circ$ is not necessarily 1, all the properties enunciated in proposition 1.3.2 are valid for the Laplacian acting on forms. The interpretation of $V_\circ$ for forms is mentioned in connection with cohomology in another volume. A detailed proof of this proposition involves some standard analysis of elliptic operators. Since the method of proof depends on techniques different from those emphasized in this volume, we simply accept the validity of proposition 1.3.2 and proceed from there. We also note the important fact from analysis that the eigenfunctions of $-\Delta$ are necessarily smooth (even analytic). Such results are often called *regularity theorems* in partial differential equations.

The variational characterization of eigenvalues of $-\Delta$ can be utilized to obtain the first term in the asymptotic expansion of $N(\lambda)$ for a bounded domain in $\mathbb{R}^m$. The asymptotic formula given in exercise **??** for torii and spheres, although valid for general compact Riemannian manifolds is considerably deeper. To formulate the problem for a bounded domain $U \subset \mathbb{R}^m$, with piecewise $C^2$ boundary, requires introducing boundary conditions:

1. (*Dirichlet Boundary Condition*) - We look for eigenvalues $\lambda$, $-\Delta\phi = \lambda\psi$, where $\phi$ is a function on $U$, continuous up to the boundary, and vanishing on $\partial U$.

2. (*Neumann Boundary Condition*) - We look for eigenvalues $\lambda$, $-\Delta\phi = \lambda\psi$, where $\phi$ is a function on $U$, continuously differentiable up to the boundary, with $\frac{\partial\phi}{\partial\nu} = 0$ on $\partial U$ where $\frac{\partial}{\partial\nu}$ denotes the derivative in the direction normal to the boundary. The boundary requirement is only on the open dense portion of $\partial U$ which is $C^2$.

**Exercise 1.3.26** *Consider the rectangle $R \subset \mathbb{R}^m$ with sides of lengths $l_1, \dots, l_m$. Show that the eigenfunctions of $-\Delta$ acting on functions on $R$ with Dirichlet boundary condition are products of the form*

$$\sin\frac{\pi k_1 x_1}{l_1} \dots \sin\frac{\pi k_m x_1}{l_m},$$

*where $k_1, \dots, k_m$ are positive integers, and the corresponding eigenvalues are $\pi^2[\frac{k_1^2}{l_1^2} + \dots + \frac{k_m^2}{l_m^2}]$. Similarly, for the Neumann boundary condition the eigenfunctions are of the same form with* *sines replaced by cosines and* $k_1, \ldots, k_m$ *non-negative integers. The corresponding eigenvalues are* $\pi^2[\frac{k_1^2}{l_1^2} + \ldots + \frac{k_m^2}{l_m^2}]$. *Deduce the same asymptotic expansions*

$$N^{\mathcal{D}}(\lambda) \sim \frac{v_m \mathrm{vol}(U)}{(2\pi)^m}\lambda^{\frac{m}{2}}, \quad N^{\mathcal{N}}(\lambda) \sim \frac{v_m \mathrm{vol}(U)}{(2\pi)^m}\lambda^{\frac{m}{2}}$$

*in both cases.*

The key idea in the application of the variational characterization of eigenvalues to the computation of $N(\lambda)$ is via the *domain monotonicity* property which is described in lemmas 1.3.4 and 1.3.5 below. We let $\lambda_1 \leq \lambda_2 \leq \ldots$ denote the eigenvalues of $-\Delta$ acting on $L^2$ functions on $U$ with Dirichlet boundary condition. The sign $\leq$ means that the eigenvalues are arranged in increasing order and each is repeated as many times as its multiplicity. Whenever necessary to emphasize the distinction between Dirichlet and Neumann boundary conditions we use the superscripts $\mathcal{D}$ with $\mathcal{N}$.

**Lemma 1.3.4** *Let* $U_1, \ldots, U_n$ *be mutually disjoint open subsets of* $U$ *with piecewise smooth boundaries. Let* $\delta_1 \leq \delta_2 \leq \ldots$ *be the eigenvalues of* $-\Delta$ *acting on* $L^2$ *functions on* $U_1 \cup \ldots \cup U_n$ *vanishing on* $\partial U_1 \cup \ldots \cup \partial U_n$. *Then* $\lambda_k \leq \delta_k$.

**Lemma 1.3.5** *Let* $U_1, \ldots, U_n$ *be mutually disjoint open subsets of* $U$ *with piecewise smooth boundaries and assume* $U = \overline{U_1} \cup \ldots \cup \overline{U_m}$. *Let* $\eta_1 \leq \eta_2 \leq \ldots$ *be the eigenvalues of* $-\Delta$ *acting on* $L^2$ *functions on* $U_1 \cup \ldots \cup U_n$ *with vanishing normal derivatives on* $\partial U_1 \cup \ldots \cup \partial U_n$. *Then* $\eta_k \leq \lambda_k$.

Before giving the proof of the lemmas let us see how they imply

**Proposition 1.3.3** *For a bounded domain* $U \subset \mathbb{R}^m$ *with piecewise* $C^2$ *boundary,* $N^{\mathcal{D}}(\lambda)$, *the number of eigenvalues* $\leq \lambda$ *for the Dirichlet boundary condition, satisfies*

$$N^{\mathcal{D}}(\lambda) \sim \frac{v_m \mathrm{vol}(U)}{(2\pi)^m}\lambda^{\frac{m}{2}},$$

*where* $v_m$ *is the volume of the unit ball in* $\mathbb{R}^m$.

**Proof** - Partition the space with equi-spaced hyperplanes orthogonal to the coordinate axes, and let $U_1, \ldots, U_n$ be the open cubes of the partition that lie entirely in $U$ and $N_j^{\mathcal{D}}(\lambda)$ be the number of eigenvalues $\leq \lambda$ for the Dirichlet boundary value problem on $U_j$. If the length of a side of the cube $U_j$ is $\alpha$ then the corresponding eigenfunctions are given as products of sine functions and the eigenvalues are

$$\pi^2\frac{k_1^2+\ldots+k_m^2}{\alpha^2},$$

where $k_1,\ldots,k_m$ range over positive integers (see exercise 1.3.26 above.) The validity of the assertion of the proposition for $N_j^{\mathcal{D}}(\lambda)$ follows by an elementary counting argument. Lemma 1.3.4 implies

$$N^{\mathcal{D}}(\lambda)\geq\sum N_j^{\mathcal{D}}(\lambda).$$

This inequality together with the validity of the desired estimate for $U_j$'s imply

$$\liminf_{\lambda\to\infty}\frac{N^{\mathcal{D}}(\lambda)}{\lambda^{\frac{m}{2}}}\geq\frac{v_m\mathrm{vol}(U)}{(2\pi)^m}. \tag{1.3.22}$$

To prove the converse inequality, let $U_1,\ldots,U_N$ be the cubes of the partition such that $\overline{U_j}\cap\bar{U}\neq\emptyset$ so that $U\subset\mathrm{int}(\overline{U_1}\cup\ldots\cup\overline{U_N})$. Let $V=\mathrm{int}(\overline{U_1}\cup\ldots\cup\overline{U_N})$. Then by lemma 1.3.4 $N_U^{\mathcal{D}}(\lambda)\leq N_V^{\mathcal{D}}(\lambda)$ and By lemma 1.3.5 $N_V^{\mathcal{D}}(\lambda)\leq\sum_{j=1}^N N_j^{\mathcal{N}}(\lambda)$. Therefore

$$N_U^{\mathcal{D}}(\lambda)\leq\sum_{j=1}^N N_j^{\mathcal{N}}(\lambda) \tag{1.3.23}$$

The eigenvalues and eigenfunctions for the Neumann boundary condition on $U_j$'s are given as products of cosine functions and the eigenvalues are

$$\pi^2\frac{k_1^2+\ldots+k_m^2}{\alpha^2},$$

where $k_1,\ldots,k_m$ range over the non-negative integers. It follows easily that the estimate of the proposition is valid for $N_j^{\mathcal{N}}$ as well. Substituting in (1.3.23) we obtain

$$\limsup_{\lambda\to\infty}\frac{N^{\mathcal{D}}(\lambda)}{\lambda^{\frac{\lambda}{2}}}\leq\frac{v_m\mathrm{vol}(U)}{(2\pi)^m}. \tag{1.3.24}$$

(1.3.22) and (1.3.24) imply the required result. ♣

It remains to proves the lemmas.

**Proof of Lemma 1.3.4** - Let $\psi_j$ be the eigenfunction corresponding to $\delta_j$ on a subset $U_{j_l}$, and extend $\psi_j$ by 0 outside $U_{j_l}$. Then for every $k$ we may assume $\psi_1,\ldots,\psi_k$ are an orthonormal sequence in $L^2(U)$. Let $\phi_1, \ldots, \phi_{k-1}$ be eigenfunctions for the $-\Delta$ with Dirichlet boundary values on $U$ corresponding to eigenval;ues $\lambda_1 \leq \ldots \leq \lambda_{k-1}$. Then the orthogonal projections $P(\psi_1), \ldots, P(\psi_k)$ on the span of $\phi_1, \ldots, \phi_k k-1$ are linearly dependent and consequently there are scalars $\beta_1, \ldots, \beta_k$, not all zero, such that

$$\sum_{j=1}^{k} \beta_j < \psi_j, \phi_l >= 0, \quad \text{for} \quad l = 1, \ldots, k-1.$$

Set $f = \sum \beta_j \psi_j$. Then $f$ is orthogonal to the span $\phi_1, \ldots, \phi_{k-1}$ and by the variational characterization of eigenvalues

$$\begin{aligned} \lambda_k ||f||^2 &\leq < df, df > \\ &= -\int_U (f\Delta f) dx \\ &= \sum_{j=1}^{k} \delta_j \beta_j^2 \\ &= \delta_k ||f||^2, \end{aligned}$$

proving the lemma. ♣

**Proof of Lemma 1.3.5** - Let $\psi_j$ be the eigenfunction (Neumann boundary condition) for eigenvalue $\eta_j$ on a subset $U_{j_l}$ and as before extend it by 0 to outside of $U_{j_l}$. If $f$ is orthogonal to the span of $\psi_1, \ldots, \psi_{k-1}$ in $L^2(U)$, then by the variational characterization of eigenvalues

$$\begin{aligned} < df, df > &= -\sum_{j=1}^{m} \int_{U_j} (f\Delta f) dx \\ &\geq \sum_{j=1}^{m} \eta_k \int_{U_j} |f|^2 \\ &= \eta_k ||f||^2. \end{aligned}$$

Let $f = \sum_{j=1}^{k} \gamma_j \phi_j$ be any non-zero element orthogonal to $\psi_1, \ldots, \psi_{k-1}$ (which clearly exists). Then

$$< df, df > \leq \lambda_k ||f||^2,$$

which implies $\lambda_k \leq \eta_k$. ♣

The asymptotic formula of proposition 1.3.3 is also valid for Neumann boundary condition and can be proven by more or less similar arguments. It is possible to extend the asymptotic formula for $N(\lambda)$ to a compact Riemannian manifold $M$ (with or without boundary) but the above method based on the variational characterization does not seem to generalize. For the remainder $R(\lambda) = N(\lambda) - \frac{v_m \mathrm{vol}(M)}{(2\pi)^m}\lambda^{\frac{m}{2}}$ we have

$$R(\lambda) = O(\lambda^{\frac{m-1}{2}}).$$

That this estimate for the remainder is sharp can be established by elementary arguments using the explicit knowledge of eigenvalues on the sphere, however the proof . It is remarkable that the remainder is related to the existence of periodic geodesics on $M$. In fact on manifolds where the geodesic flow is not periodic the estimate can be improved. For a discussion of the remainder the reader is referred to [Ho] and [DG]. For manifolds with boundary the standard conjecture for the asymptotic distribution of eigenvalues was

$$N(\lambda) = \frac{v_m \mathrm{vol}(M)}{(2\pi)^m}\lambda^{\frac{m}{2}} - \frac{c_m \mathrm{vol}(\partial M)}{(2\pi)^{m-1}}\lambda^{\frac{m-1}{2}} + o(\lambda^{\frac{m-1}{2}},$$

where $c_m$ is a constant depending only on $m$. That this formula is not valid was established by R. Melrose et al. For an account of $N(\lambda)$ for Riemannian manifolds with boundary see [Iv], [Pet] and references thereof.

Note that the analysis in the finite dimensional case is basic linear algebra. To make a story in the finite dimensional case, we replace the compact manifold $M$ with a finite graph. Let $\mathcal{V}$ be the set of vertices and $\mathcal{E}$ the set of edges of a finite graph $\Gamma$ (with no loops, i.e., an edge joining a vertex to itself; and no multiple edges). If two vertices $u, v$ are connected an edge, we write $u \leftrightarrow v$. For $v \in \mathcal{V}$ let $\delta_v$ denote the number of vertices $u$ such that $v \leftrightarrow u$. Let $\mathcal{L}$ denote the set of real or complex valued functions on $\mathcal{V}$ which is a finite dimensional vector space. One may define the Laplacian on $\mathcal{L}$ as

$$\Delta\varphi(u) = -\varphi(u) + \frac{1}{\sqrt{\delta_u}} \sum_{v, v\leftrightarrow u} \frac{\varphi(v)}{\sqrt{\delta_v}}.$$

With this definition (or some generalizations of it) on may transport a portion of the theory of the Laplacian in differential geometry or analysis to the context of graphs and Markov chains. For a discussion of this aspect of the subject see [Chu].

**Example 1.3.8** While the eigenvalue $\lambda_1 > 0$, exercise 1.3.24 shows that it can be arbitrarily small by taking $\gamma$ large. We now give a class of examples of compact surfaces for which $\lambda_1 > 0$ is arbitrarily small and sheds some light on how to obtain a lower bound for $\lambda_1$ which depends on geometric data. Let $M_i$, $i = 1, 2$, be a compact surfaces with Riemannian metrics $ds_i^2$. Assume there are small discs $D_j \subset M_j$ where $ds_j^2$ is flat. Join the surfaces by a cylinder of $P$

length $l$ and radius $r$ which intersects $M_j$ inside $D_j$ as shown in Figure XXXX, and smooth it out to obtain a surface $M$. By a slight modification of the metrics around $D_j$'s we extend it to a metric on $M$ which is the standard flat metric on $P$. Let $\phi$ be a function which is equal to $c>0$ on $M_1\backslash D_1$, $-c$ on $M_2\backslash D_2$, and decreases linearly on $P$ from $c$ to $-c$, where $c$ is to be determined later. It is clear from the construction that after a small perturbation of $\phi$ we may assume $\int_M \phi dv = 0$. We have the approximations

$$ds^2_M(d\phi, d\phi) \simeq \begin{cases} 0 & \text{on} M_1 \cup M_2, \\ \frac{4c^2}{l^2} & \text{on the cylinder} P. \end{cases}$$

It follows that

$$\lambda_1 \leq \frac{< -\Delta\phi, \phi >}{< \phi, \phi >} \simeq \frac{8\pi c^2 r}{l < \phi, \phi >}.$$

Let us assume $\mathrm{vol}(M_1) = \mathrm{vol}(M_2)$. Now let $r = \epsilon^2$, $l = \frac{1}{\epsilon}$ and determine $c > 0$ so that $< \phi, \phi >= 1$. $c > 0$ depends on $\epsilon > 0$, however, since area of the cylinder $P$ tends to 0 with $\epsilon \to 0$, $c$ remains bounded as $\epsilon \to 0$. It follows that $\lambda_1 \to 0$ as $\epsilon \to 0$. ♠

Example 1.3.8 suggests that if $S \subset M$ is a hypersurface decomposing $M$ into two pieces $M_1$ and $M_2$, then the ratio $\frac{\mathrm{Area}(S)}{\mathrm{vol}(M_i)}$ may play a role in how small $\lambda_1$ can be. To make this precise we define

$$h = \inf_S \frac{\mathrm{Area}(S)}{\min(\mathrm{vol}(M_1), \mathrm{vol}(M_2))}, \tag{1.3.25}$$

where the infimum is taken over all hypersurfaces $S$ which decompose $M$ into two disjoint submanifolds $M_1$ and $M_2$. Naturally $\mathrm{Area}(S)$ refers to the $(m-1)$-dimensional volume of $S$ relative to the volume element of $S$ obtained from the Riemannian metric on $M$. The quantity $h$, called *Cheeger's constant*, can be defined for non-compact Riemannian manifolds by a slight modification of (1.3.25). In fact we let the infimum be over all relatively compact open subsets $U \subset M$ with smooth boundary $\partial U = S$ and replace the denominator by $\mathrm{vol}(U)$. The quantity $h$ is clearly geometric in character and the fact that it gives a lower bound for $\lambda_1$ is confirmed by the proposition 1.3.4 below. First we need an observation about the volume element. For hypersurfaces $S_r$ defined by $\phi = r$ let $\omega_1, \cdots, \omega_m$ be such that $\omega_1, \cdots, \omega_{m-1}$ form orthonormal coframes for $S_r$'s. Then $\omega_m = 0$ defines the family of hypersurfaces $S_r$. Since $\omega_m$ has unit length

$$d\phi = \gamma\omega_m, \quad \text{with} \ \ \gamma = \sqrt{ds^2(d\phi, d\phi)}. \tag{1.3.26}$$

Therefore the volume element on $M$ can be written as

$$dv = \frac{1}{\gamma}\omega_1 \wedge \cdots \omega_{m-1} \wedge d\phi. \tag{1.3.27}$$

This formula enables one to relate integration on $M$ to that relative to $r \in \mathbb{R}$. In fact, if $A(r)$ denotes the volume of $S_r$ (relative to $\omega_1 \wedge \cdots \wedge \omega_{m-1}$) and $U_a^b(M)$ denotes the portion of $M$ defined by the inequalities $a \leq \phi(x) \leq b$, then

$$\int_{U_a^b(M)} \sqrt{ds^2(d\phi, d\phi)}dv = \int_a^b A(r)dr. \tag{1.3.28}$$

It is customary to refer to (1.3.27) or its integrated form (1.3.28) as the *co-area* formula.

**Proposition 1.3.4** *Let $M$ be a compact Riemannian manifold, then $\lambda_1 \geq \frac{1}{4}h^2$.*

**Proof** - For a $C^2$ function $\phi$ on $M$ let $U_+(r) = \{x \in M \mid \phi(x) \geq r\}$ and $U_-(r) = \{x \in M \mid \phi(x) \leq r\}$. If $r$ is a regular value then $S_r = U_+(r) \cap U_-(r)$ is a hypersurface decomposing $M$ into two pieces. Let $\phi$ be an eigenfunction for eigenvalue $\lambda_1$, then

$$\lambda_1 = \frac{< d\phi, d\phi >}{< \phi, \phi >} \geq \frac{[\int_M |\phi(x)| \sqrt{ds^2(d\phi(x), d\phi(x))}dv]^2}{< \phi, \phi >^2},$$

where $\geq$ follows from the Cauchy-Schwartz inequality. Since $d\phi^2 = 2\phi d\phi$ we obtain

$$\lambda_1 \geq \frac{1}{4}\frac{[\int_M \sqrt{ds^2(d\phi^2, d\phi^2)}dv]^2}{< \phi, \phi >^2}. \tag{1.3.29}$$

Assume 0 is a regular value for $\phi$, $A(r)$ be the area of the submanifold of $U_+(0)$ defined by $\phi^2 = r$ and $V(r)$ denote the volume of the portion of $U_+(r)$. Then

$$\begin{array}{rcl}
\int_{U_+(0)} \sqrt{ds^2(d\phi^2, d\phi^2)} & = & \int_\circ^\infty A(r)dr \\
\text{(by definition of } h) & \geq & h \int_\circ^\infty V(r)dr \\
\text{(integration by parts)} & = & -h \int_\circ^\infty rV'(r)dr \\
(-V' = \omega_1 \wedge \cdots \wedge \omega_{m-1}) & = & h \int_{U_+(0)} \phi^2 \omega_1 \wedge \cdots \wedge \omega_m \\
 & = & h < \phi, \phi > .
\end{array}$$

We obtain a similar inequality by looking at $U_-(0)$. The assumption that 0 is a regular value is inessential, since by looking at $U_\pm(\epsilon)$ the same inequalities can be proven. ♣

Upper bounds for $\lambda_1$ involving the Cheeger constant are more subtle and will not be discussed here. Cheeger's constant is reminiscent of the isoperimetric inequality, however, there is one essentail difference, namely, the numerator and denominator in Cheeger's constant have different dimensions while in the isoperimetric inequality $\frac{L^2}{A} \geq 4\pi$, they have the same dimension. This suggests that one should attempt to obtain lower bounds for $\lambda_1$ in terms of

$$\inf_S \frac{(\text{Area}(S))^{\frac{m}{m-1}}}{\min(\text{vol}(M_1), \text{vol}(M_2))}. \tag{1.3.30}$$

This leads to the concepts of the isoperimetric and Sobolev constants which we shall not pursue any further here since it involves more analysis than we would like to invoke at this stage.

The Laplace operator $\Delta$ admits of a nonlinear extension to mappings of Riemannian manifolds $f : M \to N$ which has proven to be geometrically significant. We use moving frames to describe this generalization. Let $f : M \to N$ be a smooth mapping of Riemannian manifolds, and let $\omega_1, \cdots, \omega_m$, and $\theta_1, \cdots, \theta_n$ be orthonormal coframes reducing the Riemannian metrics on $M$ and $N$ to the identity. Let $1 \leq i, j, \cdots \leq m$ and $1 \leq a, b, \cdots \leq n$ be the range of the indices in this subsection. We let $(\omega_{ij})$ and $(\theta_{ab})$ denote the corresponding Levi-Civita connections. Set

$$f^\star(\theta_a) = \sum f_i^a \omega_i. \tag{1.3.31}$$

Taking exterior derivatives of (1.3.31) and making use of the structure equations we obtain:

$$\sum_j \big(df_j^a + \sum_i f_i^a \omega_{ji} + \sum_b f_j^b f^\star(\theta_{ab})\big) \wedge \omega_j = 0 \tag{1.3.32}$$

Therefore by Cartan's lemma

$$df_j^a + \sum_i f_i^a \omega_{ji} + \sum_b f_j^b f^\star(\theta_{ab}) = \sum_k f_{jk}^a \omega_k, \tag{1.3.33}$$

where $f_{jk}^a = f_{kj}^a$. The *Laplacian* of $f$ is by definition the collection

$$\Delta f = \{\sum_j f_{jj}^a\}_{a=1,\cdots,n}. \tag{1.3.34}$$

A map $f : M \to N$ is *harmonic* if $\sum_j f_{jj}^a = 0$ for all $a$. While the entries of the matrix $(f_{ij}^a)$ depend on the choice of the frames, the vanishing of the traces $\sum_i f_{ii}^a$, for all $a$, is independent of these choices. We omit the verification of this fact.

Harmonic maps of Riemannian manifolds have interesting features, however, investigating their properties often requires the introduction of analytical techniques which are postponed to another volume. Here we only discuss some elementary aspects of harmonic maps and give some examples.

**Exercise 1.3.27** *Show that for an $\mathbb{R}^k$-valued function $f$ on a Riemannian manifold the two definitions of $\Delta f$ given above are identical.* ($\mathbb{R}^k$ is endowed with the standard flat Euclidean metric.)

**Exercise 1.3.28** *Assume $M$ has dimension 1, so that $M$ is either an open interval in $\mathbb{R}$ or the circle. Show that $f : M \to N$ is harmonic if and only if $f$ is a geodesic.*

**Example 1.3.9** Let $\jmath : M \to \mathbb{R}^N$ be an isometric immersion so that $M$ maybe locally regarded as a submanifold of $\mathbb{R}^N$. We want to see when the mapping $\jmath$ is harmonic. We choose moving frames on $\mathbb{R}^N$ such that $e_1, \cdots, e_m$ are tangent to $\jmath(M)$ and also use $e_1, \cdots, e_m$ as a moving frame on $M$. Denoting the coframes forms on $M$ and $\mathbb{R}^N$ by $\omega_i$ and $\theta_A$, and the connection forms by $\omega_{ij}$ and $\theta_{AB}$, we obtain

$$\jmath^\star(\theta_i) = \omega_i, \quad \jmath^\star(\theta_p) = 0, \quad \jmath^\star(\theta_{ij}) = \omega_{ij}. \tag{1.3.35}$$

It follows that $\jmath_i^k = \delta_i^k$ and $\jmath_i^p = 0$. (Recall the index convention $1 \leq i, j, \cdots \leq m$, $m+1 \leq p, q, \cdots \leq N$.) Therefore (1.3.32) becomes

$$\sum_j \jmath^\star(\theta_{pj}) \wedge \omega_j = 0,$$

which, by means of Cartan's lemma, determines $\jmath_{ij}^p$. Comparing with the definition of second fundamental form it follows that the symmetric matrix $(\jmath_{ij}^p)$ is the matrix of the second fundamental form of $\jmath(M)$ in the normal direction $e_p$, and

$$\sum_j \jmath_{jj}^p = mH_p, \tag{1.3.36}$$

where $H_p$ is mean curvature in the direction $e_p$. The same calculation carried out for a tangential direction as well. In fact going through the calculation of the Laplacian $\jmath_{jk}^i$ we see that, for each $i \leq m$, the matrix $(\jmath_{jk}^i)$ is determined by Cartan's lemma and the equation

$$\sum_k (\omega_{ik} - \jmath^\star(\theta_{ik})) \wedge \omega_k = 0.$$

In other words,

$$\omega_{ik} - \jmath^\star(\theta_{ik}) = \sum_{l=1}^m \jmath_{kl}^i \omega_l.$$

By (1.3.35) the left hand side vanishes and therefore $\jmath^i_{kl} = 0$. The above calculations can be summarized as

$$\Delta\jmath = m \sum_{p=m+1}^{N} H_p e_p. \tag{1.3.37}$$

Thus $\jmath$ is harmonic if and only if the mean curvature $H_p$ vanishes for every normal direction $e_p$. ♠

**Example 1.3.10** Continuing with the notation and hypotheses of example 1.3.9, we assume $N = m+1$ and set $\jmath(x) = (\jmath_1(x), \cdots, \jmath_{m+1}(x))$ relative to the standard coordinates on $\mathbb{R}^{m+1}$. Then $\jmath^2(x) = \sum_A \jmath_A^2(x) =< \jmath(x), \jmath(x) >$ is a real valued function on $M$. We want to calculate its Laplacian[9]. We have

$$d\jmath^2 = 2 < d\jmath, \jmath >= \sum_{j=1}^{m} (\jmath^2)_j \omega_j.$$

Since $\jmath^2$ is real valued we have omitted dependence on the index $a$ which refers to a frame on $\mathbb{R}$. Applying $- \star d\star$ we obtain

$$\begin{aligned} - \star d \star d\jmath^2 &= 2\sum_k < \Delta\jmath, \jmath > \\ &+ 2\star\left(\sum_k (-1)^k < e_k, \sum_l e_l\omega_l > \wedge\tilde{\omega}_k\right) \\ &+ 2\star\left(\sum_k (-1)^k < e_k, \jmath > d\tilde{\omega}_k\right), \end{aligned}$$

where $dv_M$ is the volume element on $M$ and

$$\tilde{\omega}_k = \omega_1 \cdots \wedge \omega_{k-1} \wedge \omega_{k+1} \wedge \cdots \wedge \omega_m.$$

In view of example 1.3.9 we have

$$2\sum_k < \Delta\jmath, \jmath > dv_M = 2mH < e_{m+1}, \jmath >,$$

[9]This calculation was carried out in example 1.2.23 since the calculation of $\Delta\jmath^2$ is the same as computing the trace of the Hessian of the function $\jmath^2$ on $M$. We will do this calculation one more time using the definition of $\Delta$ as $- \star d \star d$ and notice that the answers are identical!

and it is clear that

$$2\bigg(\sum_k(-1)^k < e_k, \sum_l e_l\omega_l > \wedge\tilde{\omega}_k\bigg) = 2mdv_M.$$

Expanding $< e_k, \jmath > d\tilde{\omega}_k$ we obtain a sum of terms each of which contains the connection form $\omega_{ij}$. However the quantities $-\star d\star d\jmath^2$ and the first and second sums in the expansion of $-\star d\star d\jmath^2$ are defined on $M$ independently of the choice of frame. Since by appropriate choice of frame we can make all $\omega_{ij}$'s vanish at any given point, the third sum $\sum_k(-1)^k < e_k, \jmath > d\tilde{\omega}_k$ vanishes identically on $M$. Therefore we have

$$\Delta\jmath^2 = 2m(1 + H < e_{m+1}, \jmath >). \tag{1.3.38}$$

We will give some applications of this formula. ♠

As an application of (1.3.38) we derive some integral formulae for compact hypersurfaces. Let $\jmath : M \to \mathbb{R}^{m+1}$ be an isometric immersion so that $M$ may be locally regarded as a hypersurface. In developing local expressions for various geometric quantities, we omit any reference to the immersion $\jmath$ and work on $M$ directly. Let $e_1, \cdots, e_{m+1}$ be a moving frame with $e_{m+1}$ a unit normal vector field to $M$. For $t$ a small fixed real number, consider the hypersurfaces

$$M_t : \quad x - te_{m+1}, \qquad \text{where} \;\; x \in M.$$

Let $y^t = x - te_{m+1}$, then

$$dy^t = \sum_{i=1}^m (\omega_i - t\omega_{i\ m+1})e_i.$$

This relation implies that $e_{m+1}$ is normal to $M_t$ and the 1-forms $\omega_i - t\omega_{i\ m+1}$ form an orthonormal coframe for $M_t$. The second fundamental form $\mathsf{H}^t$ of $M_t$ is calculated from

$$-\omega_{i\ m+1} = \sum_{j=1}^m \mathsf{H}^t_{ij}(\omega_j - t\omega_{j\ m+1}).$$

Since $-\omega_{i\ m+1} = \sum \mathsf{H}_{ij}\omega_j$ we obtain after a simple calculation

$$\mathsf{H}^t = \mathsf{H}\big[I + t\mathsf{H}\big]^{-1},$$

where $\mathsf{H} = \mathsf{H}^\circ$ denotes the second fundamental form of $M$. Therefore the principal curvatures of $M_t$ are related to those of $M$ by[10]

$$\kappa_i^t = \frac{\kappa_i}{1 + t\kappa_i}. \tag{1.3.39}$$

The normalized $k^{\text{th}}$ elementary symmetric function of the principal curvatures will be denoted by $H_{(k)}$, i.e.,

$$H_{(k)} = \frac{1}{\binom{m}{k}} \sum_{i_1,\cdots,i_k} \kappa_{i_1}^t \cdots \kappa_{i_k}^t,$$

where the summation is over distinct indices $i_1, \cdots, i_k$. Therefore the volume element for $M_t$ can be written as

$$dv_{M_t} = (\omega_1 - t\omega_{1\ m+1}) \wedge \cdots \wedge (\omega_m - t\omega_{m\ m+1}) = \left(\prod_{j=1}^m (1 + t\kappa_j)\right) dv_M. \tag{1.3.40}$$

Set $P(t) = \prod_j (1 + t\kappa_j)$. Then the mean curvature of $M_t$ is given by

$$H^t = \frac{1}{m}\mathrm{Tr}\mathsf{H}^t = \frac{P'(t)}{mP(t)}, \tag{1.3.41}$$

where $'$ denotes differentiation relative to $t$. We can now prove

**Proposition 1.3.5** (Minkowski) *Let $M$ be a compact orientable Riemannian manifold and $\jmath : M \to \mathbb{R}^{m+1}$ an isometric immersion. Then for $k \leq m-1$ we have*

$$\int_M \big(H_{(k)} + H_{(k+1)} < e_{m+1}, \jmath > \big) dv_M = 0.$$

**Proof** - With the notation of the paragraph preceding the proposition, we let $\jmath_t(x) = \jmath(x) - te_{m+1}$. Then for $t$ small, $\jmath_t$ is also an immersion. It follows from (1.3.38) and Stokes' theorem that

$$\int \big(1 + H^t < e_{m+1}, \jmath_t > \big) dv_{M_t} = 0. \tag{1.3.42}$$

[10] Recall that the signs of the principal curvatures depend on the direction of the unit normal $e_{m+1}$. In order for the principal curvatures of the sphere to be positive, one should use the inward pointing unit normal vector field. This explains the + sign in the denominator of (1.3.39).

From (1.3.40) and (1.3.41) and a straightforward calculation, (1.3.42) simplifies to

$$\int_M \big(mP(t) - tP'(t) + P'(t) < e_{m+1}, \jmath > \big) dv_M = 0.$$

The left hand side is a polynomial in $t$ and the conclusion follows from the vanishing of the coefficient of $t^k$. ♣

We conclude this subsection with another example of harmonic maps.

**Example 1.3.11** Let $M\imath\mathbb{R}^{m+1}$ be a hypersurface and $\mathsf{G} : M \to S^m\imath\mathbb{R}^{m+1}$ the corresponding Gauss map. Thus if $e_1, \cdots, e_{m+1}$ is a moving frame with $e_{m+1}$ a unit normal vector field to $M$, then $\mathsf{G}(x) = e_{m+1}$. Let $\omega_1, \cdots, \omega_{m+1}$ be the dual coframe. We may regard $e_1, \cdots, e_{m+1}$ as a moving frame in a neighborhood of $S^m$ with $e_{m+1}$ the unit normal to $S^m$. Denote the corresponding coframe for $S^m$ by $\theta_1, \cdots, \theta_m, \theta_{m+1}$, then $\theta_1^2 + \cdots + \theta_m^2$ is the standard Riemannian metric on $S^m\imath\mathbb{R}^{m+1}$. Since $\theta_i =< de_{m+1}, e_i >$ we obtain

$$\mathsf{G}^\star(\theta_i) = \omega_{i\ m+1} = \sum_{j=1}^m \mathsf{H}_{ij}\omega_i. \tag{1.3.43}$$

Thus with the notation of this subsection (see formula (1.3.31)) we have $\mathsf{G}_j^i = m\mathsf{H}_{ij}$. From the structure equations we have

$$d\mathsf{G}^\star(\theta_i) + \sum_{j=1}^m \mathsf{G}^\star(\theta_{ij}) \wedge \mathsf{G}^\star(\theta_j) = 0,$$

where $(\theta_{ij})$ is the Levi-Civita connection for $S^m$. Taking exterior derivative of $\mathsf{G}^\star(\theta_i)$, using the structure equations, (1.3.43) and proceeding to compute the Laplacian of $\mathsf{G}$ we obtain

$$d\mathsf{H}_{ik} - \sum_{j=1}^m \mathsf{H}_{ij}\omega_{jk} + \sum_{j=1}^m \mathsf{H}_{jk}\mathsf{G}^\star(\theta_{ij}) = \sum_{l=1}^m \mathsf{H}_{ikl}\omega_l, \tag{1.3.44}$$

where $\mathsf{H}_{ikl}$ in symmetric in the indices $i, k, l$. The Laplacian of $\mathsf{G}$ is the set of $m$ functions $(\sum_{k=1}^m \mathsf{H}_{jkk})$ where $j = 1, \cdots, m$. Now assume $M$ has constant mean curvature, then $\sum_i d\mathsf{H}_{ii} = 0$. Setting $k = i$ and summing over $i$, (1.3.44) yields

$$\sum_{i,j} \mathsf{H}_{ij}\omega_{ij} + \sum_{i,j} \mathsf{H}_{ij}\mathsf{G}^\star(\theta_{ij}) = \sum_{i,l} \mathsf{H}_{iil}\omega_l. \tag{1.3.45}$$

Let $e_1, \cdots, e_m$ be along principal directions so that the matrix $(\mathsf{H}_{ij})$ is diagonal. Then $\sum_{i,j} \mathsf{H}_{ij}\omega_{ij}$ and $\sum_{i,j} \mathsf{H}_{ij}\mathsf{G}^\star(\theta_{ij})$ vanish and (1.3.45) simplifies to

$$\sum_{i,l} \mathsf{H}_{iil}\omega_l = 0. \tag{1.3.46}$$

In view of the symmetry of $\mathsf{H}_{ijk}$ it follows that for every $l$ we have $\sum_i \mathsf{H}_{lii} = 0$ or equivalently the Gauss map of a hypersurface of constant mean curvature is harmonic. For minimal surfaces in $\mathbb{R}^3$ we shall prove in example **??** that the map Gauss map is anti-holomorphic in a sense that will be clarified later. One can obtain generalizations for submanifolds of codimension $k > 1$ by looking at the (generalized) Gauss map which takes values in the Grassmann manifold of oriented $k$-planes in $\mathbb{R}^{m+k}$. ♠

### 1.3.5 Congruences of Geodesics and Jacobi's Equation

Let $\Gamma$ be a *congruence* of geodesics on the Riemannian manifold $M$ of dimension $m$. This means that there is an $m-1$ dimensional submanifold $D \subset M$ such that through every point of $D$ there passes (and in a transverse manner) exactly one $\gamma \in \Gamma$. Since this definition is purely local we assume $D$ is a disc and work in one coordinate neighborhood. We want to investigate the conditions under which there is a function $\psi$ such that $\Gamma$ is precisely the orthogonal trajectories (after parametrization by arc-length) of the hypersurfaces $\psi(x) = c$. (Compare with the subsection on Geodesics especially condition •.) Let $U \subset M$ be an open set such that every point $x \in U$ lies on exactly one geodesic in $\Gamma$, and $\sigma : U \rightarrow \mathcal{T}M$ be the section defined by $\sigma(x) = (x, \dot{\gamma}_x)$ where $\dot{\gamma}_x$ is the tangent vector at $x$ to the unique geodesic in $\Gamma$ passing through $x$. Recall that $\tilde{\varepsilon}$ and $\tilde{\omega} = -d\tilde{\varepsilon}$ are the pull-back, by the Riemannian metric, of the canonical 1-form and symplectic 2-form on $\mathcal{T}^\star M$ to $\mathcal{T}M$. Set $\varepsilon_\Gamma = \sigma^\star(\tilde{\varepsilon})$ and $\omega_\Gamma = -d\varepsilon_\Gamma = \sigma^\star(\tilde{\omega})$.

**Lemma 1.3.6** *With the above notation and hypotheses, a necessary and sufficient condition for the existence $\psi$ with the required properties is $\omega_\Gamma = 0$.*

**Proof** - To prove necessity, assume $\Gamma$ is the gradient flow of some function $\psi$. We have the expression

$$\varepsilon_\Gamma = \sum_i (\sum_j g_{ij}\dot{\gamma}_j) dx_i. \tag{1.3.47}$$

Here $\dot{\gamma} = (\dot{\gamma}_1, \cdots, \dot{\gamma}_m)$ is the coordinate expression of the tangent vector field to the geodesic $\gamma$ which by assumption is the gradient vector field of $\psi$. The gradient vector field has coordinate representation $g^{-1}\Psi'$ which when substituted in the expression for $\varepsilon_\Gamma$ yields $\varepsilon_\Gamma = d\psi$. Consequently, $\omega_\Gamma = -d\varepsilon_\Gamma = 0$ as claimed.

To prove sufficiency assume $\omega_\Gamma = 0$ and it is no loss of generality to require the geodesics in $\Gamma$ to be parametrized by arc length. Then $d\varepsilon_\Gamma = 0$ which implies $\varepsilon_\Gamma = d\psi$ for some function $\psi$ on $U$. To apply condition • of the subsection on geodesics, we compute

$$\Psi g^{-1}\Psi' = \sum_{i,j} g_{ij}\dot\gamma_i\dot\gamma_j = 1.$$

Therefore condition • is applicable and the orthogonal trajectories to the hypersurfaces $\psi = c$ are geodesics. In view of (1.3.47) and $d\psi = \varepsilon_\Gamma$, the gradient flow of $\psi$ is represented by

$$g^{-1}(g(\dot\gamma_1, \cdots, \dot\gamma_m)') = (\dot\gamma_1, \cdots, \dot\gamma_m)',$$

where superscript $'$ denotes the transpose. This expression is precisely the tangent vector field to the geodesics in $\Gamma$ thus proving sufficiency. ♣

Lemma 1.3.6, originated from ([C1],§4). By an *orthogonal congruence of geodesics* we mean we mean a function $\psi$ on $M$ such that the orthogonal trajectopries to the hypersurfaces $\psi =$ const. form a congruence of geodesics. Lemma 1.3.6 can be rephrased as local Lagrangian sections $\jmath$ of $M$ into its tangent bundle are equivalent to considering orthogonal congruences of geodesics. More generally, consider Lagrangian submanifolds $L \subset \mathcal{T}M$ such that the differential of the restriction of the projection $\pi : \mathcal{T}M \to M$ to $L$ is generically of maximal rank $m$. The points where the differential of the projection $\pi$ when restricted to $L$ fails to have maximal rank are referred to as *singularities* of $(L, \pi)$ or simply of $L$. These singularities (called *caustics* in physics terminology) are the same as "focusing" of congruences of geodesics. The focusing phenomenon is related to the notion of Jacobi field which we will introduce shortly.

First we examine the two dimensional case. Let $M$ be a surface with a Riemannian metric $ds^2$ and $\psi$ a function (defined on an open subset of $M$) such that the orthogonal trajectories to $\psi =$ const. are geodesics. As noted in the subsection on Geodesics, we can assume $ds^2(\text{grad}\psi, \text{grad}\psi) = 1$ so that $\psi$ is arc length along the geodesics $\gamma \in \Gamma$ (up to a constant specifying the initial point). The curve defined by $\psi = c$ will be denoted by $M_c$. We denote the family of geodesics orthogonal to the curves $M_c$ by $\Gamma$. Now assume completeness of the surface $M$ so that geodesics orthogonal to $M_\circ$ can be continued indefinitely. The family of geodesics $\Gamma$ defines the Lagrangian submanifold $L_\Gamma \subset \mathcal{T}M$ as

$$L_\Gamma = \{(\gamma_\phi(t), \dot\gamma_\phi(t))\}. \tag{1.3.48}$$

Therefore completeness suggests that $L_\Gamma$ can be extended by using the defining relation (1.3.48). However, there is no guarantee that $L_\Gamma$ thus defined is a submanifold and if so the differntial of the projection $\pi$ restricted to $L_\Gamma$ has maximal rank. We will now show how the singularities of $(L_\Gamma, \pi)$ can be explicitly described in terms of the metric $ds^2$. After a change of variable we can assume the metric and the curvature are

$$ds^2 = d\psi^2 + G^2 d\phi^2, \quad K = -\frac{1}{G}\frac{\partial^2 G}{\partial \psi^2}, \tag{1.3.49}$$

where $\psi$ defines the orthogonal congruence of geodesics $\Gamma$. Assume $G(0,\phi) \neq 0$, for $\phi$ in some open interval $(-\epsilon, \epsilon)$ which we identify with $M_\circ$, and let $\gamma = \gamma_\phi$ be the geodesic orthogonal to the curve $M_\circ$ and passing through the point $\phi$. Let $\gamma$ be a fixed geodesic orthogonal to the loci $\psi =$ const, and the coordinate $\phi$ on each locus $M_c : \psi = c$ be such that $(c, \phi)$ is the point of intersection of $\gamma_\phi$ with $M_c$. From (1.3.49) it follows that $G$ satisfies an ordinary differential equation along each geodesic $\gamma = \gamma_\phi$, namely,

$$\frac{d^2 G}{dt^2} + KG = 0. \tag{1.3.50}$$

Here $t$ is the arc length along the geodesics in $\Gamma$. In this form the quantity[11] $G$ is meaningfully defined along each $\gamma$, and we can more easily understand its geometric significance. Initially the function $\psi$ is defined only on an open set $U \subset M$. We can extend $\psi$ by *continuing* it along geodesics. This means that we want to define the value of $\psi$ at the point $\gamma_\phi(t)$ to be $t$. Since a geodesic may re-enter the set $U$, this (at best) gives us a multi-valued function[12]. The quantity $G$ may vanish at some points, but this vanishing is only apparent since $ds^2$ is positive definite, and therefore the expression (1.3.49) for the metric is not valid when $G = 0$. Nevertheless the points where $G$ vanishes have an important geometric significance. In fact, the proof of lemma 1.3.7 shows that points where $G$ vanishes are precisely the points where the differential of $\pi_{|L_\Gamma}$ fails to have maximal rank. The quantity $G$, defined by the differential equation (1.3.50), is no longer dependent on the coordinate system and is meaningfully defined on $\gamma$. In view of the uniqueness of solutions and smooth dependence on initial conditions of second order ordinary differential equations, $L_\Gamma$ is locally a surface with coordinates $(t, \phi)$, however, globally $L_\Gamma$ is an immersed surface and may have self intersections given by isolated closed geodesics. The following lemma clarifies the issue of singularities of $(L_\Gamma, \pi)$:

**Lemma 1.3.7** *$L_\Gamma$ is a submanifold of $\mathcal{T}M$ and the differential of the restriction of $\pi$ to $L_\Gamma$ has maximal rank at all points where $G \neq 0$, i.e., $(t, \phi)$ fails to give on $M$ precisely at points where $G$ vanishes.*

---

[11] $G$ is a tensor component, and so we refer to it as a quantity rather than a function along $\gamma$.

[12] The function $\psi$ is more naturally regarded as a function on $L_\Gamma$. This is reminiscent of the construction of a Riemann surface from a polynomial in two variables. However the nature of singularities in this case is quite different from that of functions of one complex variable.

**Proof** - We have already proven the first assertion. The 1-forms $\omega_1 = d\psi$ and $\omega_2 = Gd\phi$ are unit cotangent vectors to geodesics $\gamma_\phi \in \Gamma$ and the curves $M_c$ respectively. Taking $(t,\phi)$ as coordinates on $L_\Gamma$, then the basic relation

$$dx = \omega_1 e_1 + \omega_2 e_2,$$

where $x$ is a generic point of $M$ expressed relative to the $(t,\phi)$ coordinates on $L_\Gamma$, implies that at points where $G$ does not vanish the differential of the projection $\pi_{|L_\Gamma}$ has maximal rank. ♣

**Exercise 1.3.29** *Let $M \subset \mathbb{R}^3$ be a sphere with the induced metric from $\mathbb{R}^3$. Let $p \in S^2$ and $\psi(q)$ be the distance of $p$ to $q$. Show that $G$ vanishes precisely at the point $p'$ anti-podal to $p$. Prove that $\pi^{-1}(p') \cap L_\Gamma$ is a circle.*

There is an inequality associated with congruences of geodesics which leads to an important insight in Riemannian geometry. To motivate and describe this inequality intuitively for surfaces, we look at a variation of a geodesic $\gamma \in \Gamma$, i.e., we consider a 1-parameter family of curves $\gamma_\delta$ such that $\gamma_\circ = \gamma$ and the curves $\gamma_\delta$ are defined on a short interval $[a,b]$. Let $J(\delta)$ denote the length of the arc of the curve $\gamma_\delta$ on $[a,b]$. If the variation is such that $\gamma_\delta \in \Gamma$, then we expect $J''(0) = 0$ where the second derivative of $J$ is computed relative to the variation parameter $\delta$. On the other hand, if the variation $\gamma_\delta$ is no longer a geodesic, then we expect $J''(0) > 0$. We now work out this inequality rigorously and then discuss an application of it.

It is convenient to introduce a new coordinate system which is better adopted to our problem. Let $v$ curves (i.e., curves $t = \text{const.}$) be the geodesics orthogonal to $\gamma$ with $v$ measuring (signed) arc length from $\gamma$. The orthogonal trajectories to the $v$ curves are the $t$-curves. The curve $v = 0$ is a geodesic but the curves $v = c \neq 0$ may not be geodesics. We have

**Lemma 1.3.8** *Relative to the $(t,v)$ coordinates the metric takes the form*

$$ds^2 = H(t,v)^2 dt^2 + dv^2,$$

*with $H$ satisfying*

$$H(t,0) = 1, \quad \frac{\partial H}{\partial v}(t,0) = 0.$$

**Proof** - The fact the metric has the required form and $H(t,0) = 1$ are immediate. If $\frac{\partial H}{\partial v}(t_\circ, 0) \neq 0$, then we may assume $v \rightarrow H(t,v)^2$ is an increasing function of $v \in (-\epsilon, \epsilon)$ for $t$ in a neighborhood of $t_\circ$. Then it is a simple matter to construct a curve joining two pints

$\gamma(t_1)$ and $\gamma(t_2)$ near $\gamma(t_\circ)$ and of length $< |t_2 - t_1|$ contradicting the local length minimizing property of geodesics. ♣

Let $v = v(t, \delta)$ be a 1-parameter variation of the $t$-curve $v = 0$ for $\delta \in (-\epsilon, \epsilon)$ subject to $v(t, 0) = \gamma(t)$. Let $J(\delta)$ denote the length of the curve $t \to v(t, \delta)$ between 0 and $a$. Then using lemma 1.3.8 we make the substitutions

$$v(t,\delta) = v(t,0) + \delta\frac{\partial v}{\partial \delta}(t,0) + O(\delta^2), \quad H(t,v) = 1 + \frac{\partial^2 H}{\partial v^2}(t,0) + O(\delta^3),$$

in

$$J(\delta) = \int_a^b \sqrt{H^2 + (\frac{\partial v}{\partial t})^2} dt$$

to obtain

$$J(\delta) = a + \frac{\delta^2}{2}\int_a^b \left[(\frac{\partial \xi}{\partial t})^2 - K\xi^2\right] dt + O(\delta^3),$$

where the Gaussian curvature $K = -\frac{\partial^2 H}{\partial v^2}$ along $\gamma$, and $\xi = \frac{\partial v}{\partial \delta}(t, 0)$. Therefore

$$J''(0) = \frac{1}{2}\int_a^b \left[(\frac{\partial \xi}{\partial t})^2 - K\xi^2\right] dt. \tag{1.3.51}$$

Integrating by parts we obtain

$$J''(0) = -\frac{1}{2}\int_a^b \xi\left(\frac{\partial^2 \xi}{\partial t^2} + K\xi\right) dt + \xi\frac{\partial \xi}{\partial t}\Big]_a^b. \tag{1.3.52}$$

Notice that the integrand in (1.3.52) is $\xi$ times the differential equation (1.3.50) with $\xi$ replacing $G$. The form of the extremal property useful for our application is the following lemma:

**Lemma 1.3.9** *Assume $\xi$ satisfies the differential equation (1.3.50) on an interval $[a, b]$ with $\xi(b) = 0$ and $\xi(t) \neq 0$ for $t \in [a, b)$. Then for any other quantity $\eta$ with $\eta(a) = \xi(a)$ and $\eta(b) = 0$ we have*

$$\int_a^b \left[(\frac{d\xi}{dt})^2 - K\xi^2\right] dt \leq \int_a^b \left[(\frac{d\eta}{dt})^2 - K\eta^2\right] dt.$$

**Proof** - Since $\xi$ does not vanish on $[a,b)$, we write $\eta = f\xi$. Substituting and using the equation $\frac{d^2\xi}{dt^2} + K\xi = 0$ we obtain

$$\int_a^b \left[ (\frac{d\eta}{dt})^2 - K\eta^2 \right] dt = \int_a^b \left[ (f\frac{d\xi}{dt} + \frac{df}{dt}\xi)^2 + f^2\xi\frac{d^2\xi}{dt^2} \right] dt.$$

Integrating the term $\int_a^b f^2\xi\frac{d^2\xi}{dt^2}dt$ by parts and simplifying we obtain

$$\int_a^b \left[ (\frac{d\eta}{dt})^2 - K\eta^2 \right] dt = \int_a^b (\frac{df}{dt})^2\xi^2 dt + f^2\xi\frac{d\xi}{dt}\Big]_a^b. \tag{1.3.53}$$

Substituting $f \equiv 1$ in (1.3.53) we obtain

$$\int_a^b \left[ (\frac{d\xi}{dt})^2 - K\xi^2 \right] dt = \xi\frac{d\xi}{dt}\Big]_a^b. \tag{1.3.54}$$

Comparing (1.3.53) and (1.3.54) and using the boundary conditions we obtain the desired result. ♣

Logically, the statement and proof of lemma 1.3.9 are independent of the computation of the second derivative $J''(0)$, however, the computation of the latter (formula 1.3.51) motivates the lemma and will be used in its application. Let $p \in M$ and $\psi$ be the distance function from $p$ which implies that the metric has the required form given in (1.3.50). The function $\psi$ defines an orthogonal congruence of geodesics $\Gamma$ starting at $p$. Since $M$ is assumed to be complete, every $q \in M$ lies on at least one geodesic $\gamma \in \Gamma$. We say $q$ is *conjugate* to $p$ (along a geodesic $\gamma$ joining $p$ to $q$) if there is a non-trivial solution to the differential equation

$$\frac{d^2\xi}{dt^2} + K\xi = 0 \tag{1.3.55}$$

vanishing at $p$ and $q$. Here $t$ denotes the arc length along the geodesic $\gamma$. In other words, if $G$ vanishes at $q$ then $q$ is conjugate to $p$, or $q$ is a point where $(L_\Gamma, \pi)$ is singular. Notice that $G$ is now regarded as a solution to the differential equation $\frac{d^2G}{dt^2} + KG = 0$ extending the metric coefficient $G$ in $ds^2 = d\psi^2 + G^2 d\phi^2$ which is initially defined in a punctured neighborhood $U\backslash p$ of $p$. The notion of conjugate point is related to the question of how far along a geodesic $\gamma$ we can go while minimizing the distance between $p$ and $\gamma(t)$. This is a rather difficult question and the following proposition gives a partial answer:

**Proposition 1.3.6** *Let $M$ be a complete surface, $p \in M$ and $\gamma$ a geodesic with $\gamma(0) = p$. Then after passing through a point conjugate to $p$ along $\gamma$, the geodesic $\gamma$ no longer minimizes the distance between $p$ and $\gamma(t)$.*

**Proof** - Let $\gamma$ be a geodesic and assume a singularity of $(L_\Gamma, \pi)$ occurs at $\tilde{q}$ with $q = \pi(\tilde{q}) = \gamma(b)$. We assume $q$ is the first conjugate to $p$. Let $a < b$ be sufficiently close to $b$ so that all points in the disc $D$ of radius $2|b-a|$ centered at $\gamma(b)$ can be joined by a unique, necessarily length minimizing, geodesic lying entirely inside the disc. Let $b < c < 2b - a$ so that the point $\gamma(c)$ lies beyond the conjugate point $q = \gamma(b)$. Let $\xi$ be a non-trivial solution to the differential equation (1.3.55) vanishing at $p$ and $q$, and $\zeta$ be the solution to (1.3.55) defined on $[a, 2b-a]$ with boundary conditions

$$\zeta(a) = \xi(a), \quad \zeta(2b-a) = 0.$$

By taking $|b-a|$ sufficiently small, we are ensured of the existence and uniqueness of $\zeta$ from elementary theory of linear second order ordinary differential equations. Define

$$\eta(t) = \begin{cases} \xi & \text{if} 0 \leq t \leq a, \\ \zeta & \text{if} t \in [a, 2b-a]. \end{cases}$$

Consider the variation of the geodesic $\gamma$ defined by $\eta$ as discussed above. Since the variation depends on $\eta$ (in $(t, v)$ coordinates $\eta$ is $\frac{\partial v}{\partial \delta}(t, 0)$) and the end-points of the interval under consideration we write $J(\delta; \eta, [a, b])$ rather than $J(\delta)$ in order to specify all the data. Since $J''(0; \xi, [0, b]) = 0$ (see formula (1.3.52)) we have

$$J''(0; \eta, [0, 2b-a]) = J''(0; \eta, [0, 2b-a]) - J''(0; \xi, [0, b]),$$

which gives

$$J''(0; \eta, [0, 2b-a]) = J''(0; \zeta, [b-a, 2b-a]) - J''(0; \xi, [b-a, b]). \tag{1.3.56}$$

Let $\bar{\xi}$ be defined by

$$\bar{\xi} = \begin{cases} \xi & \text{on} [b-a, b], \\ 0 & \text{on} [b, 2b-a]. \end{cases}$$

Then from lemma 1.3.9 it folows that

$$J''(0; \zeta, [b-a, 2b-a]) \leq J''(0; \bar{\xi}, [b-a, b]) = J''(0; \xi, [b-a, b]),$$

and consequently

$$J''(0; \eta, [0, 2b-a]) < 0. \tag{1.3.57}$$

Hence $\gamma$ does not minimize the distance between $p$ and $\gamma(2b-a)$. ♣

We need the following basic proposition, which is a special case of the *Sturm Comparison* theorem, from the theory of second order ordinary differential equations to infer an important geometric corollary from proposition 1.3.6:

**Proposition 1.3.7** *Consider two differential equations*

$$\frac{d^2\xi_j}{dt^2} + K_j(t)\xi_j = 0, \quad j = 1, 2,$$

*on the same interval* $[a, b]$. *Assume* $0 \leq K_1(t) \leq K_2(t)$ *for* $t \in [a, b]$ *and a (non-trivial) solution* $\xi_1$ *with consecutive zeros at* $t_1 < t_2 \in (a, b)$. *If the solution* $\xi_2$ *vanishes at* $t_1$ *then it also vanishes at* $t_3$ *with* $t_1 < t_3 \leq t_2$.

**Proof** - Multiplying the equation for $\xi_1$ by $\xi_2$, and the equation for $\xi_2$ by $\xi_1$, subtracting and integrating we get

$$\int_{t_1}^{t_2} (K_2 - K_1)\xi_1\xi_2 dt + \int_{t_1}^{t_2} \left[\xi_1 \frac{d^2\xi_2}{dt^2} - \xi_2 \frac{d^2\xi_1}{dt^2}\right] dt = 0.$$

Now $\xi_1 \frac{d^2\xi_2}{dt^2} - \xi_2 \frac{d^2\xi_1}{dt^2} = \frac{d}{dt}\left(\xi_1 \frac{d\xi_2}{dt} - \xi_2 \frac{d\xi_1}{dt}\right)$. Therefore using the boundary conditions at $t_i$ we obtain

$$\int_{t_1}^{t_2} (K_2 - K_1)\xi_1\xi_2 dt - \xi_1'(t_2)\xi_2(t_2) = 0. \tag{1.3.58}$$

We may assume $\xi_1 > 0$ on $(t_1, t_2)$, $\xi_2 > 0$ near $t_1$ and $\xi_1'(t_2) < 0$ since $t_1$ and $t_2$ are consecutive zeros. Therefore $\xi_2(t_2) > 0$ which contradicts (1.3.58). ♣

Now we discuss some applications of proposition 1.3.7. For the sphere of constant curvature $K$ the solutions of the differential equation $\frac{d^2\xi_1}{dt^2} + K\xi_1 = 0$ vanishing at 0 are scalar multiples of $\sin\sqrt{K}t$, and therefore conjugate points are a distance $\frac{\pi n}{\sqrt{K}}$ apart. It is convenient to define the *diameter* of a Riemannian manifold $M$ as

$$\mathrm{diam}(M) = \sup_{p,q \in M} d(p, q),$$

where $d$ is the distance function on $M$ induced from the Riemanian metric. If a complete surface $M$ has curvature $K_M$ bounded below by $K > 0$, then Sturm's Comparison theorem implies that conjugate points are at most a distance $\frac{\pi}{\sqrt{K}}$ apart. Therefore

**Corollary 1.3.2** (Bonnet) *Let* $M$ *be a complete surface with curvature* $K_M \geq K$ *for some constant* $K > 0$. *Then the diameter of* $M$ *is bounded above by* $\frac{\pi}{\sqrt{K}}$ *and* $M$ *is compact.*

**Proof** - We have already shown the bound for the diameter and compactness is by a standard elementary argument. ♣

Using the Sturm Comparison Theorem twice we obtain the following corollary:

**Corollary 1.3.3** *Consider the differential equation*

$$\frac{d^2\xi}{dt^2} + K(t)\xi = 0,$$

*on the interval* $[a,b]$. *Assume* $0 < k_1 \leq K(t) \leq k_2$. *If a solution* $\xi$ *has consecutive zeros at* $t_1 < t_2 \in (a,b)$, *then*

$$\frac{\pi}{\sqrt{k_2}} \leq t_2 - t_1 \leq \frac{\pi}{\sqrt{k_1}}.$$

**Proof** - Let $K_2 = k_2$ in proposition 1.3.7 and solve the equation for $j = 2$ to obtain one inequality. Proceed similarly for the second inquality. ♣

As a consequence of corollary 1.3.3 we obtain

**Corollary 1.3.4** *Let* $M$ *be a surface with curvature* $K$ *satisfying the bounds*

$$0 < k_1 \leq K \leq k_2.$$

*Let* $\gamma$ *be a geodesic on* $M$ *parametrized by arc length with* $\gamma(t_1)$ *and* $\gamma(t_2)$, $t_1 < t_2$, *consecutive conjugate points along* $\gamma$. *Then*

$$\frac{\pi}{\sqrt{k_2}} \leq t_2 - t_1 \leq \frac{\pi}{\sqrt{k_1}}.$$

An examination of the proof of proposition 1.3.7 shows that we have in fact proven more. What is important for our applications is to remove the requirement that $K_1 \geq 0$ since we want to understand conjugate points along geodesics on surfaces of non-positive curvature. This case is much simpler since if $-K \geq 0$ then it is elementary that a solution of $\frac{d^2\xi}{dt^2} = -K\xi$ with initial conditions $\xi(a) = 0$, $\frac{d\xi}{dt}(a) \geq 0$, satisfies the inequality

$$\xi(t) > 0, \quad \text{on} \quad (a,b].$$

Therefore we have shown

**Corollary 1.3.5** *There are no conjugate points on a surface of non-positive curvature.*

It should be emphasized that corollary 1.3.5 does not mean that a geodesic $\gamma$ on a surface of non-positive curvature is distance minimizing between all points $\gamma(a)$ and $\gamma(b)$. The flat torus provides such an example.

So far our analysis in this subsection was limited to the two dimensional case and was based on the differential equation satisfied by the metric coefficient $G$ (1.3.50). Next we derive the higher dimensional analogue of (1.3.50). For this purpose let, as before, $\psi$ be a function on $M$ defining an orthogonal congruence of geodesics $\Gamma$. We may assume grad$\psi$ has norm 1 so that $\psi = t$ is also arc length along the geodesics $\gamma \in \Gamma$. Let $e_1, \cdots, e_m$ be a moving with $e_m$ the unit tangent vector field to the geodesics in $\Gamma$ and the remaining $e_j$'s are obtaining by specifying them on a fixed hypersurface $\psi = 0$ and parallel translation along $\gamma \in \Gamma$. The corresponding coframe is denoted as usual by $\omega_1, \cdots, \omega_{m-1}, \omega_m = d\psi$. The 1-forms $\omega_1, \cdots, \omega_{m-1}$ depend on $\psi$ but do not contain the differential $d\psi$. The submatrix $(\omega_{ij})$, $i, j = 1, \cdots, m-1$, is the Levi-Civita connection of the Riemannian submanifolds defined by $\psi = c$. They may depend on $\psi$, but do not contain the differential $d\psi$ since $\omega_{ij}(e_m) = 0$ by the construction of the moving frame. Regarding $\omega_j$, $j = 1, \cdots, m-1$, as 1-forms depending a parameter $\psi$, we obtain from $d\omega_j + \sum \omega_{jk} \wedge \omega_k = 0$,

$$\frac{d\omega_j}{d\psi} = \omega_{jm}, \qquad j = 1, \cdots, m-1. \tag{1.3.59}$$

From the defining equation of curvature

$$d\omega_{jm} + \sum \omega_{jk} \wedge \omega_{km} = \Omega_{jm} = -\sum_{k<l} R_{jmkl} \omega_k \wedge \omega_l.$$

and the fact that the differential $d\psi$ does not appear in $\omega_{jk}$ for $j, k \neq m$, it follows that

$$\frac{d\omega_{jm}}{d\psi} = -\sum_l R_{jmlm} \omega_l. \tag{1.3.60}$$

Comparing (1.3.59) and (1.3.60) we obtain

$$\frac{d^2\omega_j}{d\psi^2} + \sum_k R_{jmkm} \omega_k = 0.$$

This is the analogue of (1.3.50) and is named after *Jacobi* who studied the two dimensional case. We know from experience that it is convenient to look at this linear ordinary differential equation more abstractly by regarding the quantities $\omega_j$, $j = 1, \cdots, m-1$, as the unknows which is justified since $\omega_1, \cdots, \omega_{m-1}$ span an $(m-1)$-dimensional vector space. Therefore we write Jacobi's equation in the form

$$\frac{d^2\xi_j}{dt^2} + \sum_k R_{jmkm} \xi_k = 0, \quad j = 1, \cdots, m-1. \tag{1.3.61}$$

In deriving equation (1.3.61) we made an arbitrary choice of orthonormal frame $\omega_1, \cdots, \omega_{m-1}$ for the hypersurfaces $M_c$. A change of frame by a gauge transformation $A \in O(m-1)$ (along

$\gamma$) will replace $(\omega_1, \cdots, \omega_{m-1})'$ by $A(\omega_1, \cdots, \omega_{m-1})'$ where $'$ denotes the transpose of the row vector. The symmetric matrix $(R_{imjm})_{i,j=1,\cdots,m-1}$ is transformed into

$$(R_{imjm}) \longrightarrow A(R_{imjm})A'.$$

Since $(R_{imjm})$ is a symmetric matrix, by appropriate choice of gauge transformation we may assume $(R_{imjm})$ is a diagonal matrix. Consequently the coframe $\omega_1, \cdots, \omega_{m-1}$ can be chosen such that the system of equations (1.3.61) is decoupled into $(m-1)$ second order ordinary differential equations:

$$\frac{d^2\xi_j}{dt^2} + R_{jmjm}\xi_j = 0, \quad j = 1, \cdots, m-1. \tag{1.3.62}$$

Note that relative to this (co)frame the components $R_{imjm}$, $i \neq j$, of the curvature tensor vanish. To extend the preceding theory to $m$-dimensional case we make use of the following complement to the Sturm Comparison theorem, proposition 1.3.7:

**Lemma 1.3.10** *Consider two differential equations*

$$\frac{d^2\xi_j}{dt^2} + K_j(t)\xi_j = 0, \qquad j = 1, 2,$$

*on the same interval $[a, b]$. Assume*

1. $0 \leq K_1(t) \leq K_2(t)$ *are continuous functions.*
2. *For some $t_\circ$, $K_1(t_\circ) < K_2(t_\circ)$.*
3. *$\xi_2$ is a non-trivial solution solution of the second equation with consecutive zeros at $t_1, t_2$ and $a < t_1 < t_\circ < t_2 < b$.*

*Then a non-trivial solution $\xi_1$ of the fist equation with $\xi_1(t_1) = 0$ does not vanish in the interval $(t_1, t_2]$.*

**Proof** - Proceeding as in the proof of proposition 1.3.7, we obtain

$$\int_{t_1}^{t_2} (K_2 - K_1)\xi_1\xi_2 dt + \xi_1(t_2)\xi_2'(t_2) = 0. \tag{1.3.63}$$

If $\xi_1(t_2) = 0$ then (1.3.63) implies that $\xi_1$ changes sign on the interval $(t_1, t_2)$, and therefore $\xi_1$ vanishes at some point in $(t_1, t_2)$. Proposition 1.3.7 is applicable to show that $t_1 < t_2$ are not consecutive zeros of $\xi_2$ contrary to hypothesis. ♣

Analogues of corollaries 1.3.2 and 1.3.5 in arbitrary dimensions can now be obtained by reduction to the two dimensional case and making use of example **??** and lemma 1.3.10. Let $M$ be a Riemannian manifold and $\gamma$ a geodesic with $\gamma(0) = p \in M$. Let $e_m$ be the unit tangent vector field to $\gamma$, and $e_1(p), \cdots, e_m(p)$ an orthonormal basis for $\mathcal{T}_pM$. Let $q = \gamma(b)$ be the first conjugate point to $p$ along $\gamma$. Extend $e_1(p), \cdots, e_m(p)$ to a moving frame in a neighborhood of $\gamma(t)$, $0 \leq t \leq b$. Let $\omega_1, \cdots, \omega_m$ be the dual coframe (on $(0, b)$). Without loss of generality we may assume that $\xi_j = \omega_j$, $j = 1, \cdots, m-1$, satisfy the decoupled Jacobi equation (1.3.62) on $[0, b]$ and with $\xi_1$ vanishing at 0 and $b$. By example 1.2.25 the curvature $K_N$ of the surface

$$N = \mathrm{Exp}_{\gamma(t)}(se_1), \qquad -\epsilon < s < \epsilon$$

along $\gamma$ is bounded above by the sectional curvature $R_{1m1m}$ of $M$ along $\gamma$ with equality if and only if the vector field $e_2$ is parallel along $\gamma$ in $M$. If $e_2$ were not parallel along $\gamma$ in $M$ then $K_N < R_{1m1m}$ at some point of $\gamma$. Then lemma 1.3.10 will apply to show that $\xi_1$ cannot vanish at $\gamma(b)$. Therefore

$$K_N = R_{1m1m} \quad \text{on } \gamma. \tag{1.3.64}$$

With this observation the analysis of the two dimensional case becomes applicable to the general case and it is straightforward to deduce the validity of the following corollaries:

**Corollary 1.3.6** *Let $M$ be a complete Riemannian manifold with sectional curvatures bounded below by a constant $K > 0$. Then* $\mathrm{diam}(M) \leq \frac{\pi}{\sqrt{K}}$ *and $M$ is compact.*

The case of Riemannian manifolds of non-positive curvature is simpler, and one easily shows that

**Corollary 1.3.7** *Let $M$ be a Riemannian manifold of non-positive sectional curvature. Then there are no conjugate on $M$.*

**Corollary 1.3.8** *Let $M$ be a Riemannian manifolds with sectional curvatures $K$ satisfying the bounds*

$$0 < k_1 \leq K \leq k_2.$$

*Let $\gamma$ be a geodesic on $M$ parametrized by arc length with $\gamma(t_1)$ and $\gamma(t_2)$, $t_1 < t_2$, consecutive conjugate points along $\gamma$. Then*

$$\frac{\pi}{\sqrt{k_2}} \leq t_2 - t_1 \leq \frac{\pi}{\sqrt{k_1}}.$$

We will return to the discussion of conjugate points in chapter 3.

# 1.4 Geometry of Surfaces

## 1.4.1 Flat Surfaces and Parallel Translation

In this example we study flat surfaces $M \subset \mathbb{R}^3$ (*flat* means all sectional curvatures are zero which in the case of surfaces is vanishing of curvature). Cylinders and cones are simple examples of flat surfaces, and exercise 1.2.1(b) provides a non-obvious class of such surfaces. There is a general procedure for (locally) constructing all generic flat surfaces. Let $\gamma$ be a curve (parametrized by arc length) in $\mathbb{R}^3$ which we assume is generic in the sense that even locally it does not lie in any affine plane. Let $e_1, e_2, e_3$ be a Frenet frame for $\gamma$ as explained in example 1.1.3. Consider the family of *osculating planes* to $\gamma$, i.e., affine planes with origin moved to the point $\gamma(t)$ and spanned by the vectors $e_1(t), e_2(t)$. Thus we have a one parameter family of planes defined by the equations

$$(\mathbf{x} - \gamma(t)).e_3(t) \;=\; 0. \tag{1.4.1}$$

The enveloping surface $M_\gamma$ of this family of planes is obtained eliminating $t$ from (1.4.1) and its $t$-derivative, namely

$$(\mathbf{x} - \gamma(t)).e_2(t) \;=\; 0. \tag{1.4.2}$$

The surface $M_\gamma$ is parametrically given by

$$\mathbf{x} = \mathbf{x}(t, u) = \gamma(t) + ue_1(t). \tag{1.4.3}$$

To compute the curvature of $M_\gamma$ we note that the Riemannian metric on $M_\gamma$ relative to the parametrization (1.4.3) is given by $\begin{pmatrix} 1 + \frac{u^2}{\tau^2} & 1 \\ 1 & 1 \end{pmatrix}$ where $\frac{1}{\tau}$ is the torsion of the curve $\gamma$ (see example 1.1.3) and depends only on $t$. By a linear change of coordinates $t = t',\ u = u' - t'$, the metric takes the form

$$\begin{pmatrix} \frac{(u'-t')^2}{\tau(t')^2} & 0 \\ 0 & 1 \end{pmatrix}$$

Applying exercise 1.2.1 we see that the curvature of $M_\gamma$ is identically zero. Conversely, consider a surface $M \subset \mathbb{R}^3$ with vanishing curvature. Let $\{e_1, e_2, e_3\}$ be an orthonormal frame for $M$ diagonalizing the second fundamental form, and let $e_1$ be along the line of curvature corresponding to principal curvature zero. Let $\gamma$ be a line curvature with tangent vector field $e_1$. From the structure equations we have $de_3 = \omega_{13}e_1 + \omega_{23}e_2$. Since the second fundamental form is already diagonal relative to this basis, we have $\omega_{13} = 0$. It follows that

$$de_1 = \omega_{21}e_2$$

Now under some genericity assumption, we may assume $\omega_{21}(e_1) \neq 0$ relative to this frame. It then follows that the osculating planes are in fact tangent to the surface $M$ so that $M$ is the enveloping surface of the osculating planes to $\gamma$. There are two degenerate cases to consider, namely, (a) $\omega_{21} = 0$, and (b) $\omega_{21} \neq 0$ but $\omega_{21}(e_1) = 0$. In the first case we have $de_1 = 0$ and $de_2 = \omega_{32}e_3$. Therefore the integral curves of the vector field $e_1$ are straight lines. Furthermore the plane spanned by $e_2, e_3$ is independent of the point $x \in M$ ($x$ is translated to the origin) and the integral curves for the vector field $e_2$ lie in this plane. This makes $M$ into a cylinder. By a similar argument one shows that in the second degenerate $M$ is a cone.

It is clear the developable surface $M_\gamma$ contains the one parameter family of lines $u \rightarrow \gamma(t) + ue_1(t)$. A surface generated by the motion of a straight line (such as $M_\gamma$) is called a *ruled surface*. It should be pointed out that, in general, a ruled surface is not flat. The following example explains why a ruled surface may not be flat (see also exercise 1.4.1 in the subsection on quadrics below):

**Example 1.4.1** Let $M \subset \mathbb{R}^3$ be a ruled surface given by

$$(s, v) \longrightarrow \delta(s) + v\xi(s),$$

where $\delta$ is a curve in $\mathbb{R}^3$ and $\xi(s)$ is a vector in $\mathbb{R}^3$ with initial point $\delta(s)$. Then the tangent space to $M$ at the point with coordinates $(s, v)$ is the image of the linear map from $\mathbb{R}^2$ to $\mathbb{R}^3$ given by the matrix

$$(\dot{\delta} + u\dot{\xi}, \xi)$$

where $\dot{\delta}$ etc. denotes derivative of $\delta$ etc. with respect to the variable $s$. Let $L(s)$ denote the line $\delta(s) + v\xi(s)$ as $v$ varies and $s$ remains fixed (called a *ruling* of the ruled surface). If, for fixed $s$, the vectors $\dot{\delta}$, $\dot{\xi}$ and $\xi$ are linearly dependent, then the tangent spaces to $M$ along $L(s)$ have the same normal. Consequently, the second fundamental form is a singular matrix with the unit vector along $L(s)$ an eigenvector for eigenvalue zero. This is the case for a developable surface. However, for a general ruled surface, the vectors $\dot{\delta}$, $\dot{\xi}$ and $\xi$ are linearly independent. Then the normals to the surface along $L(s)$ depend on $v$, the second fundamental form will be a nonsingular matrix and $M$ will have non-zero Gaussian curvature.

The notion of a developable surface and the above example can be used to give a precise geometric interpretation to the concept of parallel translation along a curve on a surface $M \subset \mathbb{R}^3$. Let $\delta$ be a curve on the surface $M \subset \mathbb{R}^3$ and consider the family of tangent planes $\mathcal{T}_{\delta(s)}M$ to $M$ along $\delta$. Denoting an orthonormal moving frame by $f_1, f_2, f_3$ with $f_3$ normal to $M$, the envelope of the tangent plane to $M$ along $\delta$ is given by eliminating $s$ from the equations

$$(\mathbf{x}-\delta(s)).f_3=0, \quad (\mathbf{x}-\delta(s)).(\omega_{13}(\dot{\delta})f_1+\omega_{23}(\dot{\delta})f_2)=0.$$

Therefore it has the parametric representation

$$(s,v)\longrightarrow \delta(s)+v\xi(s),$$

where $\xi(s)=-\omega_{23}(\dot{\delta})f_1+\omega_{13}(\dot{\delta})f_2$. From $d\dot{f}_A=\sum\omega_{BA}f_B$ it follows that the coefficient of $f_3$ in $\dot{\xi}$ vanishes and consequently the vectors $\dot{\delta}$, $\xi$ and $\dot{\xi}$ are linearly dependent. Hence the envelope is a developable surface by example 1.4.1[13]. In view of the general construction of flat surfaces, there is a curve $\gamma$ such that the envelope is of the form $M_\gamma$ as described earlier. The flat surface $M_\gamma$ is tangent to $M$ along the curve $\delta$. On $M_\gamma$ let $(x,y)$ be coordinates so that the metric takes the Euclidean form $ds^2=dx^2+dy^2$ (see exercise 1.2.15). Then the connection form $\omega_{12}$ for $M_\gamma$ vanishes identically relative to the frame $e_1,e_2$ parallel to the coordinate axes in the $(x,y)$-plane. Clearly this frame extends to an orthonormal frame $e_1,e_2,e_3$, with $e_3$ normal to $M_\gamma$. The restriction of $e_1,e_2,e_3$ to the curve $\delta$ extends to an orthonormal frame $e'_1,e'_2,e'_3$ for $M$. Let $\theta_{AB}$ be defined by the relation $de'_A=\sum\theta_{BA}e'_B$, then $\theta_{12}$ is the connection form for $M$. Since $M$ and $M_\gamma$ are tangent along the curve $\delta$ we have

$$\theta_{12}(\dot{\delta}) \;=\; \omega_{12}(\dot{\delta}) \;=\; 0.$$

This means that parallel translation of the frame $e_1,e_2$ along the curve $\delta$ on $M$ is the same as Euclidean parallel translation in the $(x,y)$-coordinates which reduce the metric on the developable surface $M_\gamma$ (which is the envelope of the tangent planes to $M$ along $\delta$) to the Euclidean form $dx^2+dy^2$.

### 1.4.2 Quadrics

Quadric surfaces provide interesting examples of surfaces in $\mathbb{R}^3$. Understanding their geometry is a good demonstration of how a judicious choice of coordinates or frames is essential in unravelling a geometric structure. We consider the quadric surface defined by the equation

$$Q:\ \frac{x_1^2}{a_1}+\frac{x_2^2}{a_2}+\frac{x_3^2}{a_3}=1, \tag{1.4.4}$$

where we assume $a_1>a_2>a_3; a_1a_2a_3\neq 0$ and $a_1>0$. If $a_3>0$ then $Q$ is an *ellipsoid*; $Q$ is a *hyperboloid of one sheet* if $a_2>0>a_3$, and a *hyperboloid of two sheets* if $a_1>0>a_2$.

[13]More generally, the envelope of a one parameter family of planes is a developable by a similar argument.

One can parametrize $Q$ by using trigonometric and hyperbolic functions similar to the use of polar coordinates on the sphere, however, relative to such parametrizations the metric will be in non-diagonal form and computations appear to be intractable or extremely laborious. There is a remarkable way of parametrizing $Q$ which greatly simplifies the computation of many quantities of interest which we now describe. Consider in addition to $Q$ the family of quadrics

$$Q(\lambda):\ \frac{x_1^2}{a_1+\lambda}+\frac{x_2^2}{a_2+\lambda}+\frac{x_3^2}{a_3+\lambda}=1, \tag{1.4.5}$$

where $\lambda$ is a parameter. In analytic geometry one refers to $Q(\lambda)$'s as a family of *confocal* quadrics. For each $(x_1,x_2,x_3)$ we define

$$q(\lambda)\stackrel{\text{def}}{\equiv}(a_1+\lambda)(a_2+\lambda)(a_3+\lambda)-x_1^2(a_2+\lambda)(a_3+\lambda)-x_2^2(a_3+\lambda)(a_1+\lambda)-x_3^2(a_1+\lambda)(a_2+\lambda). \tag{1.4.6}$$

This is a cubic equation in $\lambda$ and for $x=(x_1,x_2,x_3)\in Q$ one of its roots is $\lambda=0$. Note the geometric meaning of $q(\lambda)=0$, namely, for fixed $y=(y_1^\circ,y_2^\circ,y_3^\circ)\in Q$, substituting the solutions $\lambda=u,v$ of $q(\lambda)=0$ (with $y_j$'s replacing $x_j$'s) in (1.4.5) we obtain equations of two other quadrics, confocal with (1.4.4), and passing through the point $y\in Q$. Therefore

$$q(\lambda)=\lambda(\lambda-u)(\lambda-v), \tag{1.4.7}$$

where $u$ and $v$ depend on $x\in Q$. We use $(u,v)$ as coordinates on $Q$. Expressing $x_1,x_2,x_3$ in terms of $(u,v)$ is a simple matter. In fact, substituting $\lambda=-a_i$ in (1.4.6) and using (1.4.7), we obtain

$$x_1^2=\frac{a_1(a_1+u)(a_1+v)}{(a_1-a_2)(a_1-a_3)},\quad x_2^2=\frac{a_2(a_2+u)(a_2+v)}{(a_2-a_1)(a_2-a_3)},\quad x_3^2=\frac{a_3(a_3+u)(a_3+v)}{(a_3-a_1)(a_3-a_2)}. \tag{1.4.8}$$

This parametrization is valid in every connected open subset of the region $x_1x_2x_3\neq 0$. After a simple calculation we see that the metric on $Q$ relative to this parametrization is given the $2\times 2$ matrix

$$ds^2:\ \begin{pmatrix}\frac{u(u-v)}{4(a_1+u)(a_2+u)(a_3+u)} & 0\\ 0 & \frac{v(v-u)}{4(a_1+v)(a_2+v)(a_3+v)}\end{pmatrix}. \tag{1.4.9}$$

Not only the metric is in diagonal form relative to this parametrization, the second fundamental form is also diagonal if we take frames along the curves $v=$ const. and $u=$ const. Since this reflects a more general phenomenon we first make the following observation:

The quadric surfaces defined by (1.4.4) maybe regarded as part of the family (1.4.5). Any two surfaces belonging to (1.4.5) intersect orthogonally in the sense that their normal vectors are orthogonal. This is proven by using (1.4.8) and the fact that the tangent plane to the quadric $Q(\lambda)$ at the point $(c_1, c_2, c_3)$ is given by

$$\frac{c_1 x_1}{a_1+\lambda} + \frac{c_2 x_2}{a_2+\lambda} + \frac{c_3 x_3}{a_3+\lambda} = 1. \tag{1.4.10}$$

Since three quadrics (necessarily orthogonal) of the family (1.4.5) pass through every point of the open subset of $\mathbb{R}^3$ defined by $x_1x_2x_3 \neq 0$, we say we have a *triply orthogonal* family of surfaces. Let $e_1, e_2,$ and $e_3$ be a moving frame with $e_i$'s normals to the three quadrics. Then we have $dx = \sum_{i=1}^3 \omega_i e_i$, and $de_i = \sum_{j=1}^3 \omega_{ij} e_i$. Now set

$$\omega_{23} = b_{11}\omega_1 + b_{12}\omega_2 + b_{13}\omega_3,\ \omega_{31} = b_{21}\omega_1 + b_{22}\omega_2 + b_{23}\omega_3,\ \omega_{12} = b_{31}\omega_1 + b_{32}\omega_2 + b_{33}\omega_3$$

where $b_{ij}$'s are functions on $\mathbb{R}^3$. Now consider the surface with normal $e_3$, i.e. defined by the equation $\omega_3 = 0$. Then recall that the second fundamental form of this surface is obtained by looking at $0 = d\omega_3 = \omega_{13} \wedge \omega_1 + \omega_{23} \wedge \omega_2$ valid on the surface, and using Cartan's lemma to conclude from the structure equations that $\omega_{13} = A_{11}\omega_1 + A_{12}\omega_2$, and $\omega_{23} = A_{21}\omega_1 + A_{22}\omega_2$ with $A_{12} = A_{21}$. It follows that $b_{11} = -b_{22}$. Similarly by looking at the surfaces with normals $e_1$ and $e_2$ we conclude that $b_{22} = -b_{33}$ and $b_{11} = -b_{33}$. Therefore $b_{11} = b_{22} = b_{33} = 0$. Substituting in the matrix of the second fundamental form we see that $A_{12} = 0$. Therefore we have shown

**Lemma 1.4.1** *The moving frame $e_1, e_2, e_3$ consisting of normals to a triply orthogonal family simultaneously diagonalizes the second fundamental forms of the surfaces.*

Since the the quadrics (1.4.5) define a triply orthogonal family, we conclude from lemma 1.4.1 that by taking $e_1$ and $e_2$ be along the $v =$ const. and $u =$ const. curves we diagonalize the second fundamental form. Having made this general observation we proceed to compute the second fundamental form and the principal curvatures of a quadric surface. It is convenient to introduce the quantity $l$ which is the length of the perpendicular from the origin to the tangent plane to $Q$ at $(c_1, c_2, c_3) \in Q$. From (1.4.10) and (1.4.8) it follows easily that

$$\frac{1}{l^2} = \frac{c_1^2}{a_1} + \frac{c_2^2}{a_2} + \frac{c_3^2}{a_3} = \frac{uv}{a_1a_2a_3}.$$

Therefore the unit normal $e_3$ to $Q$ is $e_3 = (\frac{lc_1}{a_1}, \frac{lc_2}{a_2}, \frac{lc_3}{a_3})$. It is a straightforward calculation that relative to the coframe $\omega_1 = \sqrt{\frac{u(u-v)}{4(a_1+u)(a_2+u)(a_3+u)}}du$ and $\omega_2 = \sqrt{\frac{v(v-u)}{4(a_1+v)(a_2+v)(a_3+v)}}dv$, the matrix of the second fundamental form is

$$\begin{pmatrix} \frac{1}{u}\sqrt{\frac{a_1a_2a_3}{uv}} & 0 \\ 0 & \frac{1}{v}\sqrt{\frac{a_1a_2a_3}{uv}} \end{pmatrix}. \tag{1.4.11}$$

It follows that the mean and the Gaussian curvatures of the quadric surface are

$$H = (\frac{1}{u} + \frac{1}{v})\sqrt{\frac{a_1a_2a_3}{uv}}, \quad K = \frac{a_1a_2a_3}{u^2v^2} = \frac{l^4}{a_1a_2a_3}. \tag{1.4.12}$$

**Exercise 1.4.1** *Let $Q_\circ$ be the quadric $x_1^2 + x_2^2 - x_3^2 = 1$. Show that for every point $P \in Q_\circ$ with coordinates $(a, b, c)$, $c \neq 0$, there is point $P' \in Q_\circ$ with coordinates $(\cos\beta, \sin\beta, 0)$ such that the line joining $P$ to $P'$ lies entirely on $Q_\circ$. Therefore $Q_\circ$ is a ruled surface of non-zero curvature. Show that the same is true for all hyperboloids of one sheet.*

**Exercise 1.4.2** *Consider the paraboloid defined by the equation $Q : \frac{x_1^2}{a_1} + \frac{x_2^2}{a_2} = 4x_3$ where $a_2 \geq a_1$ and $a_2 > 0$, and the family of quadric surfaces defined by*

$$Q(\lambda) : \frac{x_1^2}{a_1 - \lambda} + \frac{x_2^2}{a_2 - \lambda} = 4(x_3 - \lambda).$$

*For fixed $(x_1, x_2, x_3) \in Q(\lambda)$, the equation*

$$q(\lambda) \stackrel{\text{def}}{\equiv} x_1^2(a_2 - \lambda) + x_2^2(a_1 - \lambda) - 4(x_3 - \lambda)(a_1 - \lambda)(a_2 - \lambda) = 0$$

*has three roots, one of which is 0. Denote the other two roots by $u$ and $v$. Show that $(u, v)$ can be used to parametrize $Q$ as*

$$x_1^2 = \frac{4a_1(a_1 - u)(a_1 - v)}{a_2 - a_1}, \; x_2^2 = \frac{4a_2(a_2 - u)(a_2 - v)}{a_1 - a_2}, \; x_3 = u + v - a_1 - a_2.$$

*Show that the matrix of $ds^2$ relative to this parametrization is given by*

$$\begin{pmatrix} \frac{u(u-v)}{(a_1-u)(a_2-u)} & 0 \\ 0 & \frac{v(v-u)}{(a_1-v)(a_2-v)} \end{pmatrix}$$

*Prove that relative to the coframe $\omega_1 = \sqrt{\frac{u(u-v)}{(a_1-u)(a_2-u)}}du$, $\omega_2 = \sqrt{\frac{v(v-u)}{(a_1-v)(a_2-v)}}dv$ the matrix of the second fundamental form is diagonal with eigenvalues $\frac{1}{2u}\sqrt{\frac{a_1a_2}{uv}}$ and $\frac{1}{2v}\sqrt{\frac{a_1a_2}{uv}}$ whence calculate the mean and Gaussian curvatures of $Q$. Show also that the family $Q(\lambda)$ is a triply orthogonal family of surfaces in $\mathbb{R}^3$.*

Geodesics on a quadric surface, especially an ellipsoid, have particularly interesting features. To understand their behavior, it is convenient to introduce the notion of Liouville-Stäckel metric. Let $\phi_{ij}$ be functions on $\mathbb{R}^m$ (with coordinates $u_1, \cdots, u_m$) and the properties

$$\phi_{ij} \text{ is a function of } u_i \text{ only}, \quad \det(\phi_{ij}) = \Phi \neq 0.$$

Let $\Phi_{ij}$ denote the $(i,j)$ cofactor of the matrix $\phi = (\phi_{ij})$ so that $\Phi = \sum_j \phi_{ij}\Phi_{ij}$. A Riemannian metric of the form

$$ds^2 = \Phi \sum_i \frac{1}{\Phi_{i1}} du_i^2 \tag{1.4.13}$$

is called a *Liouville-Stäckel* metric. (We have changed notation from $(u,v)$ to $(u_1,\cdots,u_m)$ to emphasize the greater generality of Liouville-Stäckel metric.) The case of immediate interest to us is

**Exercise 1.4.3** *Show that the metric on a quadric surface is a Liouville-Stäckel metric by writing it in the form*

$$ds^2 = g_{11}du_1^2 + g_{22}du_2^2 = \phi_{12}(\frac{\phi_{11}}{\phi_{12}} - \frac{\phi_{21}}{\phi_{22}})du_1^2 + \phi_{22}(\frac{\phi_{11}}{\phi_{12}} - \frac{\phi_{21}}{\phi_{22}})du_2^2.$$

**Exercise 1.4.4** *Let $M$ be a surface with a Riemannian metric $g$, and assume that $M$ admits of a nontrivial one parameter group of isometries. Show that one can choose coordinates such that the metric becomes of the Liouville-Stäckel type.*

The importance of this metric is exemplified by the following exercise:

**Exercise 1.4.5** *For a surface with a Liouville-Stäckel metric, show that the function*

$$\mathcal{L} = g_{11}\frac{\phi_{21}}{\phi_{22}}(\frac{du_1}{ds})^2 + g_{22}\frac{\phi_{11}}{\phi_{12}}(\frac{du_2}{ds})^2$$

*is invariant under the geodesic flow, where $s$ denotes arc length along a geodesic $(u_1(s), u_2(s))$.* (Calculate $\frac{d\mathcal{L}}{ds}$ using the symplectic form of the equations of geodesics.

**Exercise 1.4.6** *The Poincaré metric on the upper half plane is of Liouville-Stäckel type. Identifying the unit tangent bundle of the upper half plane with $SO(1,2) = SL(2,\mathbb{R})/\pm I$ (see example 1.3.2), show that*

$$\mathcal{L}(\begin{pmatrix} a & b \\ c & d \end{pmatrix}) = c^2d^2, \quad \text{for} \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{R}).$$

*Prove also that if a geodesic $\gamma$ on the upper half plane intersects the real axis at $A$ and $B$, then $\mathcal{L}(\gamma) = (A-B)^{-2}$.* (While the geodesic flow on the entire upper half plane is completely integrable, on compact or finite volume quotients $\Gamma \setminus \mathcal{H}_2$, it is not integrable. This fact is reflected in the high degree of non-invariance of the function $\mathcal{L}$ under any sufficiently large discrete subgroup of $SL(2,\mathbb{R})$.)

While the preceding exercise shows that the function $\mathcal{L}$ is invariant under the geodesic flow, there is a conceptual way of understanding this invariance property which sheds light on the structure of the geodesic flow for a Liouville-Stäckel metric. For simplicity of notation set

$$U_1 = \frac{\phi_{11}}{\phi_{12}}, \quad U_2 = \frac{\phi_{21}}{\phi_{22}}, \quad h_1 = \phi_{12}, \quad h_2 = \phi_{22}.$$

Then the metric (1.4.13), in the case of a surface, takes the form

$$ds^2 = (U_1 - U_2)(h_1 du_1^2 + h_2 du_2^2),$$

where $U_i, h_i$ is a function of $u_i$ only. To understand the structure of geodesics for this metric we use condition • of the subsection "Geodesics". It follows from this condition that the orthogonal trajectories to the curves $\psi = c$ for a function $\psi$ satisfying the differential equation

$$\frac{1}{U_1 - U_2}\left[\frac{1}{h_1}\left(\frac{\partial\psi}{\partial u_1}\right)^2 + \frac{1}{h_2}\left(\frac{\partial\psi}{\partial u_2}\right)^2\right] = 1, \tag{1.4.14}$$

are geodesics. The differential equation for $\psi$ can be written in the more convenient form

$$U_1 - \frac{1}{h_1}\left(\frac{\partial\psi}{\partial u_1}\right)^2 = U_2 + \frac{1}{h_2}\left(\frac{\partial\psi}{\partial u_2}\right)^2. \tag{1.4.15}$$

This equation can be integrated since its left (resp. right) hand side is a function of $u_1$ (resp. $u_2$) only. Therefore for every fixed number $L$ the equations

$$U_1 - \frac{1}{h_1}\left(\frac{\partial\psi}{\partial u_1}\right)^2 = L = U_2 + \frac{1}{h_2}\left(\frac{\partial\psi}{\partial u_2}\right)^2, \tag{1.4.16}$$

define a function $\psi = \psi_L = \Psi_1(u_1) + \Psi_2(u_2)$ such that the integral curves for its gradient vector field are geodesics. Explicitly we can write

$$\psi(u_1, u_2) = \int \sqrt{h_1(u_1)(U_1(u_1) - L)}du_1 + \int \sqrt{h_2(u_2)(L - U_2(u_2))}du_2, \tag{1.4.17}$$

where the integral sign means an indefinite integral. Clearly the constant $L$ is invariant under the geodesic flow. To better understand the meaning of $L$, we first prove the following simple lemma:

**Lemma 1.4.2** *The geodesics of the metric $ds^2$, orthogonal to the curves $\psi_L = c$, are given by the differential equation*

$$\sqrt{h_2(U_1 - L)}du_2 - \sqrt{h_1(L - U_2)}du_1 = 0.$$

**Proof** - We write the metric in the form

$$ds^2 = (\sqrt{h_1(U_1 - L)}du_1 + \sqrt{h_2(L - U_2)}du_2)^2 + (\sqrt{h_2(U_1 - L)}du_2 - \sqrt{h_1(L - U_2)}du_1)^2.$$

The first square is $d\psi^2$, and naturally we seek functions $G$ and $\varphi$ such that $ds^2 = d\psi^2 + Gd\varphi^2$. It is easily seen that we can set

$$G(u_1, u_2) = (U_1(u_1) - L)(L - U_2(u_2)), \quad d\varphi = \frac{\sqrt{h_1}}{\sqrt{U_1 - L}}du_1 - \frac{\sqrt{h_2}}{\sqrt{L - U_2}}du_2,$$

to obtain the required form for the metric (note that $d\varphi$ is closed so that $\varphi$ exists). The orthogonal trajectories to $\psi = c$ being geodesics (see proposition 1.2.3), we obtain $\sqrt{G}d\varphi = 0$ as the differential equations of geodesics which is the desired result.

Now let $\theta$ be the angle between a geodesic (with arc-length $s$) and the curve $u_2 = c$. Clearly

$$\cos\theta = \sqrt{h_1(U_1 - U_2)}\frac{du_1}{ds}, \quad \sin\theta = \sqrt{h_2(U_1 - U_2)}\frac{du_2}{ds}. \tag{1.4.18}$$

Combining lemma 1.4.2 and (1.4.18) we obtain

$$\frac{\cos\theta}{\sqrt{U_1 - L}} - \frac{\sin\theta}{\sqrt{L - U_2}} = 0,$$

which implies

$$L = U_1 \sin^2\theta + U_2 \cos^2\theta = \mathcal{L}. \tag{1.4.19}$$

This gives the important interpretation of the function $\mathcal{L}$ and that its invariance under the geodesic flow is immediate since it is equal to $L$. We can now give an explicit description of the the tori $N_c$ of proposition **??**. In fact, for every regular value $L$ of $\mathcal{L}$, the integral curves of grad$\psi_L$ lie on the torus $N_L$, and the torus $N_L$ consists of the orthogonal trajectories to the curve(s) $\psi_L = c$. It remains to determine the regular values or critical points of the function $\mathcal{L}$.

It makes more sense to determine the critical points of $\mathcal{L}$ in a global setting when the manifold or surface $M$ is given rather than when only local information about the metric is available. Thus we restrict ourselves to the case of the ellipsoid $Q$ (1.4.4). The coordinates $(u, v)$ are valid in each connected open subset of $x_1x_2x_3 \neq 0$, and in each such open set we have $U_1 = u$ and $U_2 = v$. Therefore

$$d\mathcal{L} = \sin^2\theta du + \cos^2\theta dv + (u - v)\sin 2\theta d\theta,$$

and $d\mathcal{L} \neq 0$ in the region $x_1x_2x_3 \neq 0$. To understand the behavior of the function $\mathcal{L}$ in a neigborhood of $x_3 = 0$, we note that $\lambda = -a_3$ is a root of the equation $q(\lambda) = 0$ when $x_3 = 0$. In view of (1.4.8) we make the substitution $v = -a_3 + y^2$ which yields the expression

$$\mathcal{L} = u \sin^2\theta + (-a_3 + y^2)\cos^2\theta. \tag{1.4.20}$$

It follows that $d\mathcal{L} = 0$ for $y = 0$ and $\theta = 0$. This means that for $x_3 = 0 \neq x_1x_2$, the arc of the ellipse $\frac{x_1^2}{a_1} + \frac{x_2^2}{a_2} = 1$ together with the unit tangent vector field to it, is a curve in the unit tangent bundle which is critical for the function $\mathcal{L}$. Similar considerations apply to the arcs $x_2 = 0 \neq x_1x_3$ and $x_1 = 0 \neq x_2x_3$. There still remain six points which are the intersections of the ellipsoid with the coordinate axis. We make make the substitution $u = -a_2 + z^2$ and $v = -a_3 + y^2$ in a neighborhood of the point $x_1 = \pm\sqrt{a_1}$ to obtain

$$d\mathcal{L} = 2z\sin^2\theta dz + 2y\cos^2\theta dy + (a_3 - a_2 - y^2 + z^2)\sin 2\theta d\theta.$$

Therefore the critical points of $\mathcal{L}$ in the fibres of the unit tangent bundle over the six points are the directions $\theta = 0, \pm\frac{\pi}{2}, \pi$. Thus the critical points (or manifolds) of $\mathcal{L}$ is exactly six disjoint circles in the unit tangent the ellipsoid. This completes the description of complete integrability of the geodesic flow on the unit tangent bundle of the general ellipsoid $Q$ with $a_1 > a_2 > a_3 > 0$.

Complete integrability of the geodesic flow is related to the question of the existence of closed geodesics. To demonstrate this relationship, we consider the ellipsoid $Q$ for which we have a reasonably clear picture. The geodesic flow leaves each torus $N_c$ of proposition **??** invariant and is linear. Therefore if for some $N_c$ the geodesic flow on $N_c$ has a periodic orbit, all the orbits on $N_c$ are closed and we have a continuum of closed geodesics with the same period. The three distinguished closed geodesics which are the intersections of the hyperplanes $x_i = 0$ with $Q$, all lie on degenerate tori $N_c$ since $\mathcal{L}$ is critical on these geodesics. Therefore we cannot yet conclude the existence of closed geodesics, other than the three distinguished ones, from the above analysis. The conclusion that there are in fact continuums of closed geodesics on an ellipsoid requires further analysis to ensure that the geodesic flow is "rational" relative to the period matrix of the given torus. This is achieved by showing that if the period matrix of the tori $N_c$ are normalized so that $N_c$ becomes isometric to the standard torus represented by the unit square, then the angle that the geodesic flow makes with the $x$-axis changes continuously with $c$ and is not constant, and in particular there are many $c$'s for which the flow is rational.

The geodesic flow is in general not completely integrable and the existence of closed geodesics on Riemannian manifolds is best treated by other techniques.

**Remark 1.4.1** The Liouville-Stäckel metric can be used to establish complete integrability of the geodesic flow on an ellipsoid in $\mathbb{R}^{m+1}$. The metric on a quadric in $\mathbb{R}^{m+1}$ can be put in the form (1.4.13) by using $m+1$-tuples of orthogonal quadrics. One then exhibits $m-1$ functions which together with $E$ (which is the Riemannian metric) give complete integrability. These functions are

$$\mathcal{L}_j = \Phi \sum_{i=1}^{m} \frac{\Phi_{ij}}{(\Phi_{i1})^2} (\frac{du_i}{dt})^2, \qquad \text{for } \ j = 2, \cdots, m.$$

The detailed verification of complete integrability and the structure of the critical manifolds will not be discussed here. The ingenious ideas in unravelling the geometry of quadric surfaces are substantially due to Jacobi.

### 1.4.3 Isothermal Coordinates

We show that any Riemannian metric on a surface is (locally) conformally flat, which means that it has local expression of the form

$$ds^2 \ = \ e^{2\sigma(u,v)}(du^2 + dv^2), \tag{1.4.21}$$

where $\sigma(u, v)$ is a real-valued function. Coordinate systems where the Riemannian metric has expression of the form (1.4.21) are called *isothermal* coordinates. To prove the existence of isothermal coordinates, we start with the metric in the form $ds^2 = \omega_1^2 + \omega_2^2$ where $\omega_1, \omega_2$ is an orthonormal coframe. Set $\varphi = \omega_1 + i\omega_2$. We have

**Lemma 1.4.3** *There is locally a complex valued function $\alpha(u, v)$ of two real variables $u, v$ such that $\varphi = \alpha dw$ where $w = u + iv$. Consequently,*

$$ds^2 = \varphi\overline{\varphi} = |\alpha|^2(du^2 + dv^2)$$

*exhibiting the metric in isothermal coordinates.*

**Proof of Lemma 1.4.3** - By elementary theory of ordinary differential equations in the plane, there is an integrating factor $\mu \neq 0$ such that $\mu(\omega_1 + i\omega_2)$ is an exact differential $dw$. Since

$$dw \wedge \overline{dw} = (-2i)|\mu|^2\omega_1 \wedge \omega_2,$$

the change of variables to $(u, v)$ coordinates, where $w = u + iv$, has nonvanishing Jacobian and is permissible. The second assertion follows from the first. ♣

**Exercise 1.4.7** *Compute isothermal coordinates for $S^2 \subset \mathbb{R}^3$ by stereographic projection and implementing the proof of lemma 1.4.3.*

**Remark 1.4.2** The above proof of the existence of isothermal coordinates (or the integrating factor) requires the metric to be of class $C^2$. While the assumption of $C^2$ can be weakened, the proof becomes considerably more elaborate, and the result is false if we assume mere continuity of the metric (see [Ch1]). ♡

One important implication of the existence of isothermal coordinates is that a Riemannian metric on an orientable surface $M$ gives $M$ the structure of a complex manifold of dimension 1. In fact, we set $w = u + iv$ where $u, v$ are isothermal positively oriented coordinates on $M$. Any positively oriented change of coordinates which preserves the conformally flat form of the expression of the metric (i.e., positively oriented isothermal change of coordinates) is a conformal orientation preserving map of a domain in $\mathbb{R}^2 = \mathbb{C}$ into $\mathbb{C}$ and is therefore complex analytic by elementary complex function theory. This gives $M$ the structure of a complex manifold of dimension 1. It is useful to express the complex structure on an orientable surface defined by a Riemannian metric in terms of a moving (co)frames.

**Lemma 1.4.4** *Let $ds^2 = \omega_1^2 + \omega_2^2$. Then a function $f$ defined on an open set $U \subseteq M$ is holomorphic relative to the complex structure defined by $ds^2$ if and only if $df = g(\omega_1 + i\omega_2)$ for some function $g$ on $U$, i.e., the $\bar\partial$-component of $df$ vanishes. Similarly $df = g(\omega_1 - i\omega_2)$ for some function $g$ on $U$ if and only if $f$ is antiholomorphic.*

**Proof** - Since the lemma is obviously valid for the coframe $\omega_1 = e^\sigma du, \omega_2 = e^\sigma dv$ where $ds^2 = e^{2\sigma}(du^2 + dv^2)$ (isothermal coordinates), it suffices to to establish its invariance under positively oriented orthonormal change of frames. For the coframe

$$\theta_1 = \cos\beta\ \omega_1 - \sin\beta\ \omega_2, \quad \theta_2 = \sin\beta\ \omega_1 + \cos\beta\ \omega_2,$$

we have $\theta_1 + i\theta_2 = e^{i\beta}(\omega_1 + i\omega_2)$ which is the required invariance. Q E D

In particular, consider the sphere $S^2 \subset \mathbb{R}^3$, then the complex structure on $S^2$ from the induced metric is defined by the 1-form $\omega_{13} + i\omega_{23}$ in the notation of example 1.2.1. An important but simple consequence of this observation is the following:

**Corollary 1.4.1** *The Gauss map $g : M \to S^2$, where $g(x) = e_3$, of a minimal surface is antiholomorphic.*

**Proof** - To obtain the coefficients of the second fundamental form we may write $\omega_{j3}$ instead of $g^\star(\omega_{j3})$. Then, by (1.2.5) and the minimality condition $A_{11} + A_{22} = 0$, we have

$$\omega_{13} + i\omega_{23} = (A_{11} + iA_{12})\omega_1 + i(A_{22} - iA_{12})\omega_2 = (A_{11} + iA_{12})(\omega_1 - i\omega_2),$$

which proves the antiholomorphy of the Gauss map. ♣

**Example 1.4.2** One should note that in general it is not true that $dz = \omega_1 + i\omega_2$ where $z$ denotes a coordinate function on the surface $M$. In fact, for $M = S^2$, we have $\omega_1 = d\varphi$ and $\omega_2 = \sin\varphi d\theta$. Now if $dz = \omega_1 + i\omega_2$ then $ddz \neq 0$ which shows that $dz \neq \omega_1 + i\omega_2$ for any coordinate function $z$. ♠

**Exercise 1.4.8** *Show that the antipodal map of $S^2$ is anti-holomorphic.*

**Exercise 1.4.9** *Show that the standard cylindrical coordinates are isothermal for the surface of revolution $f(x_3) = \sqrt{x_1^2 + x_2^2}$. Compute the metric in isothermal coordinates for the catenoid defined by $\cosh x_3 = \sqrt{x_1^2 + x_2^2}$. Find isothermal coordinates for the helicoid defined by the equation $x_2 \tan x_3 = x_1$, and compare it to the metric for the catenoid.* (One such parametrization is $x_1 = \sinh u \sin v,\ x_2 = \sinh u \cos v,\ x_3 = v$.)

**Exercise 1.4.10** *Let $M \subset \mathbb{R}^3$ be the graph of a function $z = z(x, y)$. With the notation of exercise 1.2.2, define*

$$\eta = \frac{1 + p^2}{\sqrt{1 + p^2 + q^2}}dx + \frac{pq}{\sqrt{1 + p^2 + q^2}}dy, \quad \zeta = \frac{pq}{\sqrt{1 + p^2 + q^2}}dx + \frac{1 + q^2}{\sqrt{1 + p^2 + q^2}}dy.$$

*Show that $d\eta = qH$, and $d\zeta = -pH$ where $H$ is the mean curvature of $M$ (see exercise 1.2.2). Therefore $\eta$ and $\zeta$ are closed if $M$ is a minimal surface. Now assume that $M$ is a minimal surface and let $f$ and $h$ be functions such that $\eta = df$ and $\zeta = dh$. Show that*

$$u = x + f(x, y), \quad v = y + h(x, y)$$

*are isothermal coordinates for $M$, and with respect to the $(u, v)$ coordinates, and the metric takes the form*

$$ds^2 = \frac{1 + p^2 + q^2}{2 + p^2 + q^2 + 2\sqrt{1 + p^2 + q^2}}(du^2 + dv^2),$$

*and the Jacobian of change of variables is*

$$\frac{\partial(u, v)}{\partial(x, y)} = 2 + \frac{2 + p^2 + q^2}{\sqrt{1 + p^2 + q^2}}.$$

Just as harmonic functions of two real variables are closely connected to complex analytic functions, harmonic maps of surfaces are related to holomorphic maps. To understand this connection, we let $f : M \to N$ be a mapping of surfaces with Riemannian metrics in isothermal coordinates given by

$$ds_M^2 = e^{2\sigma(x,y)}(dx^2 + dy^2), \quad ds_N^2 = e^{2\rho(u,v)}(du^2 + dv^2). \tag{1.4.22}$$

Then as noted earlier $z = x + iy$ and $w = u + iv$ are complex analytic coordinates on $M$ and $N$ respectively.

**Exercise 1.4.11** *Let the metrics on $M$ and $N$ be given by (1.4.22). Set $\omega = \omega_1 + i\omega_2 = e^{\sigma(x,y)}dx + ie^{\sigma(x,y)}dy$ and a similar expression for $\theta$. Show that $d\omega + i\omega_{21} \wedge \omega = 0$ and a similar equation for $d\theta$. Define $f_\omega$ and $f_{\bar{\omega}}$ by $f^\star(\theta) = f_\omega \omega + f_{\bar{\omega}}\bar{\omega}$. Proceeding as in the definition of the coefficients $f^a_{ij}$, define $f_{\omega\omega}$, $f_{\omega\bar{\omega}}$, $f_{\bar{\omega}\omega}$ and $f_{\bar{\omega}\bar{\omega}}$, and show that $f_{\omega\bar{\omega}} = f_{\bar{\omega}\omega}$. Prove that $f$ is harmonic if and only if $f_{\omega\bar{\omega}} = 0$.* (This maybe regarded as the analogue for the expression $4\frac{\partial^2}{\partial z \partial \bar{z}} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$.)

We set $\omega = \omega_1 + i\omega_2$, $\theta = \theta_1 + i\theta_2$ and define $f_\omega$, $f_{\bar{\omega}}$ etc. as in exercise 1.4.11. Consider the quadratic differential

$$\Psi \;=\; f_\omega \overline{f_{\bar{\omega}}} \omega^2. \tag{1.4.23}$$

It is a simple matter to see that $\Psi$ is invariant under gauge transformations $\lambda : N \to U(1)$ and $\mu : M \to U(1)$. Therefore $\Psi$ does not depend on the choice of frames on $M$ and $N$. Taking exterior derivative of $f^\star(\theta)$ and using Cartan's lemma in the familiar fashion we obtain

$$df_\omega + if_\omega\omega_{12} - if_\omega f^\star(\theta_{12}) = f_{\omega\omega}\omega + f_{\omega\bar{\omega}}\bar{\omega}, \quad df_{\bar{\omega}} + if_{\bar{\omega}}\omega_{12} - if_{\bar{\omega}} f^\star(\theta_{12}) = f_{\bar{\omega}\omega}\omega + f_{\bar{\omega}\bar{\omega}}\bar{\omega}, \tag{1.4.24}$$

with $f_{\bar{\omega}\omega} = f_{\omega\bar{\omega}}$. As shown in exercise 1.4.11 harmonicity is equivalent to $f_{\omega\bar{\omega}} = 0$. We can choose gauge transformations such that $\omega_{12}$ and $\theta_{12}$ vanish at given points in $M$ and $N$. It then follows that vanishing of $f_{\omega\bar{\omega}}$ at $x \in M$ is equivalent

$$df_\omega = f_{\omega\omega}\omega, \text{ and } d\overline{f_{\bar{\omega}}} = f_{\bar{\omega}\bar{\omega}}\omega \text{ at } x \in M.$$

In view of the independence of $\Psi$ from the choice of frame (gauge transformation), this means

**Proposition 1.4.1** *$\Psi$ is holomorphic if and only if $f$ is harmonic.*

**Remark 1.4.3** Exercise 1.4.12 below gives the coordinate version of the calculation leading to proposition 1.4.1. The reason for including it is to demonstrate the greater simplicity and transparency that one often achieves by using moving frames rather than coordinates. ♡

**Exercise 1.4.12** *Show that harmonicity of $f$ can be expressed by the equation*

$$\frac{\partial^2 f}{\partial z \partial \bar{z}} + 2\frac{\partial \rho}{\partial w}\frac{\partial f}{\partial z}\frac{\partial f}{\partial \bar{z}} = 0,$$

*where $z = x + iy$, $w = u + iv$, $\frac{\partial}{\partial \bar{z}} = \frac{1}{2}(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y})$ and $\frac{\partial}{\partial z} = \frac{1}{2}(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y})$ relative to isothermal coordinates given by (1.4.22). Show that for a harmonic mapping the quadratic differential*

$$\Psi = e^{2\rho(f(z))}\frac{\partial f}{\partial z}\frac{\partial \bar{f}}{\partial z}dz^2$$

*is holomorphic.*

Since we have not developed the basic facts regarding complex manifolds even in dimension one, we cannot yet exploit the implications of holomorphy of $\Psi$. For example, anticipating the elementary fact that that there are no holomorphic quadratic differentials on $\mathbb{C}P(1) \simeq S^2$ (which will be discussed in volume 2) we see that for every harmonic map $f : S^2 \to S^2$ we have $\Psi = 0$. Consequently $f$ is either holomorphic or antiholomorphic. This maybe regarded as an analogue of the fact that harmonic functions in the plane are representable as a sum of a holomorphic and an antiholomorphic function.

**Example 1.4.3** The existence of isothermal coordinates allows us to extend the isoperimetric inequality to negatively curved surfaces (at least locally). We assume the metric is in the form $ds^2 = e^{-2\rho}(dx^2 + dy^2)$ and note that the curvature of the surface is given by $e^{2\rho}\Delta\rho < 0$ by assumption. Let $\Gamma$ be a simple closed curve on $M$ enclosing a relatively compact open set $D$. Since our considerations are local at this point we may assume $D \subset \mathbb{C}$ with piece-wise smooth boundary $\Gamma$. The area of $D$ and length of $\gamma$ are given by

$$S = \int_D e^{-2\rho(x,y)}dx \wedge dy, \quad L = \int e^{-\rho}\sqrt{dx^2 + dy^2}. \tag{1.4.25}$$

Let $\varphi_1$ be the harmonic function on $D$ with boundary values given by *rho*. Let $\varphi_2$ be the conjugate harmonic function so that $\varphi(z) = \varphi_1(z) + i\varphi_2(z)$ is holomorphic. Consider the mapping $F : D \to \mathbb{C}$ defined by

$$F(z) = \int_{z_\circ}^{z} e^{-\varphi(z)}dz,$$

where $z_\circ \in D$ is any fixed point. Then $F'$ is nowhere vanishing and we obtain a diffeomorphism of $D$ onto a region $D'$ bounded by a curve $\Gamma'$. The area of $D'$ and length of $\Gamma'$ are given by

$$S' = \int_D e^{-2\varphi_1(x,y)}dx \wedge dy, \quad L = \int e^{-\varphi_1}\sqrt{dx^2 + dy^2}. \tag{1.4.26}$$

Since $\varphi_1$ is harmonic with boundary values $\rho$, we have

$$\varphi_1 - \rho = 0 \text{ on } \Gamma, \quad \Delta(\rho - \varphi_1) \leq 0 \text{ in } D,$$

which implies $\rho \geq \varphi_1$ in $D$. It therefore follows from (1.4.25) and (1.4.26) that

$$S \leq S', \quad L' = L.$$

Thus isoperimetric inequality for the Euclidean plane implies the analogous result for negatively curve surface, viz.,

$$S \leq \frac{L}{4\pi}.$$

Of course our considerations were purely local. The same inequality is valid in general if we require the surface to be simply connected. This notion is introduced in chapter 4. There are many generaliztions of the isoperimetric inequality to Riemannian manifolds (see e.g. [Cha] and references thereof). ♠

### 1.4.4 Mean Curvature

In example 1.2.6 we showed how one can construct surfaces of revolution with prescribed mean curvature. The construction was local in nature. In the constant curvature case, the differential equation has an remarkable geometric interpretation which was discovered by Delaunay. We present below Delaunay's ingenius geometric construction of a complete surface of constant mean curvature other than the sphere.

Let $\Gamma$ be a simple closed convex curve in the plane. We fix a point $Q$ in the interior of $\Gamma$ and let $L$ be a line tangent to $\Gamma$ at some point $P \in \Gamma$. Since $\Gamma$ is convex we can imagine the rolling motion of $\Gamma$ on the line $L$. Let $s$ denote the arc length along $\Gamma$ measured from $P$ and $P_s \in \Gamma$ denote the point whose distance from $P$ along $\Gamma$, moving in the counterclockwise direction, is $s$. We assume the rolling motion of $\Gamma$ on $L$ is such that the point of contact of $\Gamma$ with $L$ moves counterclockwise on $\Gamma$. The curve $\Gamma_s$ differs from $\Gamma$ by a proper Euclidean motion $g_s \in SE(2)$ and we can assume $g_s$ depends smoothly on $s$ with $g_\circ = \text{id}$. Then $s \rightarrow g_s(Q)$ describes a curve $\Gamma'$ in the plane which we refer to as the *roulette* of $\Gamma$ (relative to $Q$). We want to study the mean curvature of the surface obtained by the rotation of $\Gamma'$ around the line $L$. It is convenient to introduce a new Cartesian coordinate system with the positive $x$-axis being the line $L$ pointing in the direction of the rolling motion of $\Gamma$, and choose $y$ accordingly. Let $\Gamma$ be described by a parameter $\phi$ and $s(\phi)$ denote the distance along $\Gamma$ of the point corresponding to parameter $\phi$ to $P$. We let $(x(\phi), y(\phi))$ denote the coordinates of $g_{s(\phi)}(Q)$ in the $(x, y)$ system. Denoting the point on $\Gamma$ by $p(\phi)$, we obtain

$$x(\phi) = s(\phi) - \frac{p(\phi) \cdot p'(\phi)}{|p'(\phi)|}, \quad y(\phi) = \sqrt{|p(\phi)|^2 - \frac{(p(\phi) \cdot p'(\phi))^2}{|p'(\phi)|^2}} \tag{1.4.27}$$

where $\cdot$ denotes the standard inner product on $\mathbb{R}^2$ with $g_{s(\phi)}(Q)$ as the origin (see figure (XXXX)). Notice the appearance of the term $s(\phi)$ in the expression for $x(\phi)$ reflects the rolling motion of $\Gamma$, and the validity the expression for $x(\phi)$ is immediate. To prove the formula for $y(\phi)$, let $e_1(\phi), e_2(\phi)$ be a moving frame with $e_1(\phi)$ tangent to $\Gamma_{s(\phi)}$. Then the expression for $y(\phi)$ follows easily from

$$p(\phi) \cdot p(\phi) = (p(\phi) \cdot e_1(\phi))^2 + (p(\phi) \cdot e_2(\phi))^2. \tag{1.4.28}$$

$\Gamma_{s(\phi)}$. We now calculate the differential equation satisfied by the roulette $\Gamma'$. In fact differentiating (1.4.27) we obtain

$$\frac{dx}{ds} = -\kappa p(\phi) \cdot e_2(\phi), \quad \frac{dy}{ds} = \kappa p(\phi) \cdot e_1(\phi), \tag{1.4.29}$$

where $\kappa$ denotes the curvature of the curve $\Gamma$. Therefore

$$\frac{dy}{dx} = -\frac{p \cdot e_1}{p \cdot e_2}. \tag{1.4.30}$$

From (1.4.28) and (1.4.30) we obtain

$$y^2\left(1 + (\frac{dy}{dx})^2\right) = p \cdot p. \tag{1.4.31}$$

In cases when one can obtain a reasonable expression for $p \cdot p$, this differential equation gives decisive information about $\Gamma'$ and the surface obtained by rotating it around the $x$-axis. This point is demonstrated by the following example and exercise 1.4.13 below:

**Example 1.4.4** Now specialize to the case where $\Gamma$ is the ellipse

$$p(\phi) = (-c + a\cos\phi, b\sin\phi),$$

that is, $\Gamma$ is the ellipse with major and minor axes $2a$ and $2b$ and foci $2c$ apart where $c^2 = a^2 - b^2$. This parametrization refers to the Cartesian coordinates with the origin at one of the foci and we let $Q$ be the other focus. We can easily calculate $y$ in terms of $\phi$ from (1.4.27) and obtain

$$\cos\phi = \frac{a(b^2 - y^2)}{c(b^2 + y^2)}. \tag{1.4.32}$$

Substituting in $p \cdot p = (a - c\cos\phi)^2$ to eliminate $\phi$, (1.4.31) yields

$$\sqrt{1 + (\frac{dy}{dx})^2} = \frac{2ay}{b^2 + y^2}. \tag{1.4.33}$$

as the differential equation for the roulette. We would like to compare this differential equation with (1.2.14), however, the latter is a second order equation. When $H$ is constant (1.2.14) can be simplified. In fact, multiplying both sides of (1.2.14) by $f'$ we obtain

$$2Hff' = f'\left(1 + f'^2\right)^{-\frac{1}{2}} - ff'f''\left(1 + f'^2\right)^{-\frac{3}{2}}.$$

The right hand side is the derivative of $f(1 + f'^2)^{-\frac{1}{2}}$, and therefore we can carry out an integration to obtain

$$\sqrt{1 + \left(\frac{df}{dx}\right)^2} = \frac{2f}{2Hf^2 + \gamma}, \tag{1.4.34}$$

where $\gamma$ is a constant of integration. Comparing (1.4.33) with (1.4.34) we conclude that $M_\Gamma$ has constant mean curvature. The solution obtained in example 1.2.6 was local, however, since we can roll the ellipse in either direction indefinitely, Delaunay's construction, in the constant mean curvature case, gives a complete surface. ♠

**Exercise 1.4.13** *By adopting the argument of example 1.4.4, show that when $\Gamma$ is a parabola, and $Q$ is its focus, then $\Gamma'$ is the catenary*

$$y = c\cosh\frac{x}{c},$$

*and the surface obtained by rotating $\Gamma'$ around the $x$-axis is a minimal surface.*

To discuss some integral inequalities involving the mean curvature we begin with the following lemma:

**Lemma 1.4.5** *Let $U \subset \mathbb{R}^3$ be an open relatively compact subset with smooth boundary $\partial U = M$, and $\mathsf{G} : M \to S^2$ the Gauss map assigning to each point of $M$ the unit outward normal. Let $M_+$ denote the subset of $M$ with non-negative Gaussian curvature. Then the restriction of $\mathsf{G}$ to points with Gaussian curvature $\kappa \geq 0$ is onto $S^2$.*

**Proof** - We give an intuitive proof. It is no loss of generality to assume that the origin lies in $U$. Let $e \in S^m$, and $P_r$ denote the hyperplane with normal $e$ and distance $r$ from the origin.

For $R$ sufficiently large $P_R \cap \bar{U} = \emptyset$. Let $r_\circ$ be the infimum of all $r$ such that $P_r \cap \bar{U} = \emptyset$. Then $P_{r_\circ}$ is a tangent space to $M$, and $M$ lies on one side of $P_{r_\circ}$ so that $\kappa$ is non-negative on $P_{r_\circ} \cap M$. In fact if $P_{r_\circ} \cap \bar{U} = \emptyset$ then for $r$ in a small neighborhood of $r_\circ$ we would have $P_r \cap \bar{U} = \emptyset$. The infimum requirement on $r_\circ$ implies that $M$ lies on one side of $P_{r_\circ}$ and $P_{r_\circ} \cap \bar{U} \subset P_{r_\circ}$ so that $P_{r_\circ}$ is a tangent space. Therefore the restriction of $\mathsf{G}$ to $M_+$ is onto. ♣

Let $M \subset \mathbb{R}^3$ be a compact (embedded) surface so that it bounds an open relatively compact subset $U$, and $\kappa_+ = \max(0, \kappa)$ where $\kappa$ denotes the Gaussian curvature of $M$. Since $\mathsf{G}^\star(dv_{S^2}) = \kappa\omega_1 \wedge \omega_2$, and by lemma 1.4.5 the restriction of $\mathsf{G}$ to $M_+$ is onto $S^2$ we obtain

$$\int_M \kappa_+ dv_M \geq 4\pi. \tag{1.4.35}$$

We can now prove

**Corollary 1.4.2** (Willmore) *For a compact embedded surface $M \subset \mathbb{R}^3$ we have*

$$\int_M H^2 dv_M \geq 4\pi,$$

*where $H$ denotes the mean curvature of $M$.*

**Proof** - We have

$$H^2 = \frac{(\kappa_1 + \kappa_2)^2}{4} = \kappa_1\kappa_2 + \frac{(\kappa_1 - \kappa_2)^2}{4}.$$

Therefore $H^2 \geq \kappa_+$ and the required result follows from (1.4.35). ♣

Related to corollary 1.4.2 is the *Willmore conjecture* that for an embedded torus $M \subset \mathbb{R}^3$ we have

$$\int_M H^2 dv_M \geq 2\pi^2. \tag{1.4.36}$$

Many special cases of this conjecture have been verified.

There are inequalities analogous to the isoperimetric inequality but involving the mean curvature of a surface or hypersurface. A particularly useful one is due to Ros which we now describe. Since the proofs of these facts for hypersurfaces are almost identical to those for surfaces we work in the more general framework of hypersurfaces. Let $M \subset \mathbb{R}^{m+1}$ be a compact connected hypersurface bounding a region $U$ ($\partial U = M$). For a point $x \in U$ let $d_U(x)$ denote the distance of $x$ to $M$, i.e.,

$$d_U(x) = \inf_{y \in M} d(x, y),$$

where $d(x, y)$ denotes the Euclidean distance between $x$ and $y$. The following geometric lemma is an important observation:

**Lemma 1.4.6** *Let $U \subset \mathbb{R}^{m+1}$ be an open relatively compact subset with smooth boundary $\partial U = M$. There is a compact set $C \subset U$ of measure zero such that for every $x$ in the complement of $C$ in $U$, $d_U(x)$ is realized by a unique point $y(x) \in M$.*

We shall not give a proof of this intuitive lemma. It is clear that if $d_U(x)$ is realized by more than point $y \in M$, then for every $z \neq x$ on the ray joining $x$ to any such $y$, $d_U(z)$ is uniquely realized by $y$. This observation may be used to give a formal proof of lemma 1.4.6, but the details are not relevant to our context and will not be discussed here. The structure of the set $C$ can be quite complex and reflects the topology of $M$, but this is not the issue at this point.

We denote a general point of $M$ by $p$ and let $e_1 \cdots, e_{m+1}$ be a moving frame with $e_{m+1}$ the unit normal to $M$ pointing to the interior $U$. The open set $V = U \backslash C$ has parametric representation

$$x = p + te_3(p),$$

where the domain of $t$ is an interval $(0, c_p)$ which depends on $p$. This parametrization means that for every $x \in V$ we let $p \in M$ be the unique point realizing $d_U(x)$. Taking exterior derivative we obtain

$$dx = \omega_1 e_1 + \cdots + \omega_m e_m + tde_{m+1} + dte_{m+1}.$$

Writing $de_{m+1} = \omega_{1\ m+1} e_1 + \cdots \omega_{m\ m+1} e_m$ and expressing $\omega_{im+1}$ in terms of the second fundamental form of the hypersurface $M$, we obtain the following expression for the volume element on $V$:

$$dv = (1 - mtH_{(1)} + \binom{m}{2} t^2 H_{(2)} + \cdots)\omega_1 \wedge \cdots \wedge \omega_m \wedge dt,$$

where $H_{(1)} = H$ (mean curvature), $H_{(2)}, \cdots$ are the normalized elementary symmetric functions of principal curvatures of the hypersurface $M$. Therefore setting

$$A = \int_\circ^{c_p} (1 - mtH + \binom{m}{2} t^2 H_{(2)} + \cdots)dt,$$

we obtain

$$\mathrm{vol}(U) = \int_M A dv_M. \tag{1.4.37}$$

We have the factorization

$$1 - mtH + \binom{m}{2} t^2 H_{(2)} + \cdots = (1 - \kappa_1 t) \cdots (1 - \kappa_m t),$$

and in view examples 1.1.1 and 1.2.20 we have

$$c_p \leq \min(\frac{1}{|\kappa_1|}, \cdots, \frac{1}{|\kappa_m|}). \tag{1.4.38}$$

It follows that each term $(1 - \kappa_i t)$ in the integral defining $A$ is positive. By the arithmetic/geometric mean inequality we have

$$(1 - \kappa_1 t) \cdots (1 - \kappa_m t) \leq (1 - H)^m.$$

It also follows from (1.4.38) that $c_p \leq \frac{1}{H}$, and consequently

$$A \leq \int_\circ^{\frac{1}{H}} (1 - tH)^m dt = \frac{1}{(m+1)H_{(1)}}.$$

Therefore (1.4.37) implies

$$\mathrm{vol}(U) \leq \frac{1}{m+1} \int_M \frac{1}{H} \omega_1 \wedge \cdots \wedge \omega_m. \tag{1.4.39}$$

We can strengthen the above conclusion as

**Proposition 1.4.2** *Let $M \subset \mathbb{R}^{m+1}$ be a compact surface bounding a region $U$. Then (1.4.39) holds with equality if and only if $M$ is a sphere.*

**Proof** - It only remains to prove the assertion about equality. Clearly equality holds for a sphere. In view of the application of the arithmetic/geometric mean inequality in the proof of (1.4.39), for $=$ to hold it is necessary for every point of $M$ to be umbilical which implies that $M$ is a sphere (see example 1.2.16). ♣.

**Corollary 1.4.3** (Alexandrov) - *A compact hypersurface $M$ of constant mean curvature embedded in $\mathbb{R}^{m+1}$ is necessarily a sphere.*

**Proof** - A compact hypersurface embedded in $\mathbb{R}^{m+1}$ necessarily bounds an open relatively compact set[14] $U$ so that $\partial U = M$. We make use of the identity (see corollary **??**)

$$\mathrm{vol}(M) = -\int_M H < e_{m+1}, x > \omega_1 \wedge \omega_m \tag{1.4.40}$$

[14]The fact that a compact embedded hypersurface decomposes $\mathbb{R}^{m+1}$ into two components, namely the interior and exterior of $M$ requires proof, but is geometrically so plausible that we will assume it without a formal argument. This issue will be discussed in a more general framework in the context of cohomology in the next volume. One may call a point $x \notin M$ an exterior point (resp. interior point) of $M$ if generically any ray emanating from $x$ intersects $M$ in an even (resp. odd) number of points.

where $x$ represents a general point of $M \subset \mathbb{R}^{m+1}$, $e_{m+1}$ is interior normal to the hypersurface etc. For $H_1$ constant, from (1.4.40) and Stokes' theorem (see example **??** and in particular formula (**??**) of chapter 1) we obtain

$$\begin{aligned} \mathrm{vol}(M) &= -H \int_M < e_{m+1}, \mathbf{x} > \omega_1 \wedge \cdots \wedge \omega_m \\ &= (m+1)H \int_U \omega_1 \wedge \cdots \wedge \omega_{m+1} \\ &= (m+1)H\mathrm{vol}(U). \end{aligned}$$

Therefore equality holds in (1.4.39) and $M$ is a sphere. ♣

The situation regrding immerions of surfaces of constant mean curvature is quite different. In fact it is possible to immerse a torus in $\mathbb{R}^3$ such that it has constant mean curvature, and glue constant mean curvature tori together in such a way that the resulting surface will have the same property. The reader is referred to [GB] and referencs thereof for for this matter.

# 1.5 Convexity

## 1.5.1 Support Function

In order to investigate the geometric notion of convexity we introduce the analytical concepts of sublinear and convex functions. A function $f : \mathbb{R}^{m+1} \to \mathbb{R}$ is called *sublinear* if

$$f(x+y) \leq f(x) + f(y), \quad f(\alpha x) = \alpha f(x) \quad \text{for} \quad \alpha \geq 0, \; x, y \in \mathbb{R}^{m+1}.$$

A sublinear function is a *convex function* in the sense that

$$f(\alpha_1 x_1 + \cdots + \alpha_k x_k) \leq \alpha_1 f(x_1) + \cdots + \alpha_k f(x_k),$$

for $x_j \in \mathbb{R}^{m+1}$ and $\alpha_j \in [0,1]$ with $\sum \alpha_j = 1$. Implicit in the definition of a convex function is that its domain is a convex subset of $\mathbb{R}^{m+1}$. If for a sublinear function $f$, $0 \neq x \in \mathbb{R}^{m+1}$ is such that $f(-x) = -f(x)$ then we call $x$ a *linearity direction* for $f$.

**Exercise 1.5.1** *Let $Q : \mathbb{R}^{m+1} \to \mathbb{R}_+$ be a convex function, and $g : \mathbb{R}_+ \to \mathbb{R}$ a monotone increasing convex function. Show that $g(Q(x))$ is a convex function.*

**Exercise 1.5.2** *For a sublinear function $f$, the set of linearity directions together with $\mathbf{0}$ is a linear subspace $L_f$ on which $f$ is linear. Every subspace of $\mathbb{R}^{m+1}$ on which $f$ is linear is contained in $L_f$.*

**Exercise 1.5.3** *Let $U \subset \mathbb{R}$ be an interval and $f : U \to \mathbb{R}$ a differentiable function. Then $f$ is convex if and only if $f'$ is an increasing function.*

The basic analytical properties of convex functions are given in the proposition 1.5.1 below. Their relevance to geometry will become clear later in this subsection.

**Proposition 1.5.1** *Let $U$ be an open convex subset of $\mathbb{R}^{m+1}$, and $f : U \to \mathbb{R}$ be a convex function.*

1. *$f$ is continuous.*
2. *$f$ is Lipschitz continuous on any compact subset $K$ of $U$.*
3. *If the partial derivatives $D_i f$ exist at a point $x \in U$, then $f$ is differentiable at $x$.*
4. *For $m = 0$, (i.e., $U \subset \mathbb{R}$ and $f : U \to \mathbb{R}$) the right and left derivatives of $f$ exist at every point in the interior in $U$.*

5. *For $U \subset \mathbb{R}$ an open interval, and $f : U \to \mathbb{R}$ a convex function, the right and left derivatives of $f$ satisfy the inequality $f'_l \leq f'_r$. If $y > x$ then $f'_l(y) \geq f'_r(x)$, and consequently, $f$ fails to be differentiable at most at a countable set of points.*

To maintain the continuity of the presentation, the proof of the above proposition, which belongs to real variable theory, is postponed to the end of this subsection.

The most important tool in the study of convex bodies is the support function. Let $K$ be a convex region in $\mathbb{R}^{m+1}$ (i.e., closure of a convex open set) with smooth boundary. Then the *support function* of $K$ is the real valued function $h_K$ on $S^m$ defined by

$$h_K(e) =< \mathsf{G}^{-1}(e), e >,$$

where $\mathsf{G}$ denotes, as usual the Gauss mapping of the boundary $\partial K$. Several issues should be clarified regarding this definition. The function $h_K$, if defined, measures the distance of the origin $\mathbf{0}$ to the tangent plane to $\partial K$ at the point whose normal is $e$. If we assume $\partial K$ is smooth and has positive Gauss-Kronecker curvature everywhere, then $\mathsf{G}$ is a diffeomorphism and therefore $h_K$ is defined. However, a general convex body need not have smooth boundary and even in the smooth case, $\mathsf{G}^{-1}$ may be many-valued. Although primarily our concern here is with bodies with smooth boundary, it is useful for our study of convexity to allow a greater generality on the convex bodies. To this end we modify the definition of $h_K$ by setting

$$h_K(e) = \sup_{x \in K} < x, e > . \tag{1.5.1}$$

If the boundary of the convex body $K$ is smooth, and $y \in \partial K$ is such that $\mathsf{G}(y) = e$, then from the fact that $K$ lies in a half space with boundary $\mathcal{T}_x \partial K$, it follows that the supremum in (1.5.1) is achieved at $y = x$. Therefore the two definitions are compatible. Note that this definition is applicable to compact sets contained in lower dimensional subspaces as well. It is convenient to extend the definition of $h_K$ to a general non-zero vector $u \in \mathbb{R}^{m+1}$. In this case we set

$$h_K(u) = \sup_{x \in K} < x, u > .$$

**Example 1.5.1** For simple convex sets, the calculation of the the support function is routine. For instance, if $K = v$ is a single point then $h_K$ is the linear function

$$h_K(u) =< u, v > . \tag{1.5.2}$$

If $K = B_R^{m+1}$ is the closed ball of radius $R > 0$ centered at the origin in $\mathbb{R}^{m+1}$, then

$$h_K(u) = ||u||R. \tag{1.5.3}$$

If the convex set $K$ is presented as the convex closure of set of its extreme points $\mathcal{E}(K)$, then it is immediate that

$$h_K(u) = \sup_{x\in\mathcal{E}(K)} < x, u > . \tag{1.5.4}$$

In particular, for the cube or more generally for any convex polytope[15] $h_K$ is obtained by taking "sup" over a finite set of points. For any convex set $K \subset \mathbb{R}^{m+1}$ and $N \geq m+1$, we may regard $K$ as a convex subset of $\mathbb{R}^N$. Then the support function $h_K^{(N)}$ of $K$ regarded as a subset of $\mathbb{R}^N$ is related to $h_K = h_K^{(m+1)}$ by

$$h_K^{(N)}(u) = h_K(P(u)). \tag{1.5.5}$$

where $P$ denotes the operator of orthogonal projection from $\mathbb{R}^N$ onto $\mathbb{R}^{m+1}$. If $K$ and $K'$ are compact convex subsets of $\mathbb{R}^{m+1}$, then it is clear that

$$h_{K+K'} = h_K + h_{K'}. \tag{1.5.6}$$

Formulae (1.5.5) and (1.5.6) are useful in computing support functions. For example, they immedately imply that the support function of the cube $K$ with vertices at $(\pm1, \cdots, \pm1) \subset \mathbb{R}^{m+1}$ is

$$h_K(u_1, \cdots, u_{m+1}) = |u_1| + \cdots + |u_{m+1}|.$$

Note that this function is differentiable outside the hyperplanes $u_j = 0$. (See also exercise 1.5.5.) ♠

**Exercise 1.5.4** *Let $\theta$ denote the parameter on the circle $S^1 = \{e^{i\theta}\}$, and $h$ be a real valued $C^2$ function of $\theta$ satisfying the differential inequality*

$$\frac{d^2h}{d\theta^2} + h > 0.$$

*Show that there is a convex set $K$ in $\mathbb{R}^2$ with support function $h_K(e^{i\theta}) = h(\theta)$.* ($h$ is extended to $\mathbb{R}^2$ via the homogeneity condition $h_K(\rho e^{i\theta}) = \rho h_K(e^{i\theta})$, $\rho \geq 0$, satisfied by $h_K$. Consider the curve $p(\theta) = h'(\theta)(-\sin\theta, \cos\theta) + h(\theta)(\cos\theta, \sin\theta)$.)

The half space $\mathsf{H}^-_{(u,\gamma)}$, $\mathbf{0} \neq u \in \mathbb{R}^{m+1}$, is defined as

$$\mathsf{H}^-_{(u,\gamma)} = \{x \in \mathbb{R}^{m+1} \mid \ < x, u > \leq \gamma\}$$

[15] A *convex polytope* is the convex closure of a finite set of points in $\mathbb{R}^{m+1}$.

Let $K \subset \mathbb{R}^{m+1}$ be a closed convex body and $0 \neq u \in \mathbb{R}^{m+1}$. We define the *support hyperplane* and *support half space* as

$$\mathsf{H}_K(u) = \{x \in \mathbb{R}^{m+1} | < x, u >= h_K(u)\}, \quad \mathsf{H}_K^-(u) = \{x \in \mathbb{R}^{m+1} | < x, u > \leq h_K(u)\},$$

and set $F_K(u) = K \cap H_K(u)$.

**Exercise 1.5.5** *Let $K$ be a cube centered at the origin. Describe geometrically the hyperplanes $\mathsf{H}_K(u)$ and the sets $F_K(u)$.*

There is a useful notion of duality in convex geometry. For a convex function $f : \mathbb{R}^{m+1} \rightarrow \mathbb{R}$, we define the *conjugate* function $f^\star$ as

$$f^\star(x) = \sup_{y \in \mathbb{R}^{m+1}} \big( < x, y > -f(y) \big).$$

**Exercise 1.5.6** *Let $f$ be lower semi-continuous*[16] *convex function on $\mathbb{R}^{m+1}$ with values in $\mathbb{R} \cup \infty$. Show that $f^\star$ is a lower semi-continuous convex function with values in $\mathbb{R} \cup \infty$ and $f^{\star\star} = f$.*

**Example 1.5.2** Let $K$ be a compact convex set and $h_K$ be its support function. The conjugate function $h_K^\star$ has a simple description. Since $K$ is compact

$$h_K(u) = \sup_{x \in K} < x, u >< \infty$$

and $h_K$ is in fact a sublinear function when extended by 0 to the origin. Now

$$h_K^\star(v) = \sup_{x \in \mathbb{R}^{m+1}} \Big[ < x, v > - \sup_{y \in K} < x, y > \Big].$$

Setting $x = \mathbf{0}$ we deduce that $h_K^\star \geq 0$. Since $K$ is compact, $\sup_{y \in K} < x, y >$ is achieved at some $z(x) \in K$. Therefore for $v \in K$ we have

$$h_K^\star(v) = \sup_{x \in \mathbb{R}^{m+1}} \Big[ < x, v > - < x, z(x) > \Big] = 0$$

[16] A function $f$ is *lower semi-continuous* at $x$ if for every $\epsilon > 0$ there is a neighborhood $U_x$ such that $f(y) > f(x) - \epsilon$ for all $y \in U_x$. If $f$ is lower semi-continuous at all $x$, then it is called a *lower semi-continuous* function. If $K$ is the closure of open set and $f : K \rightarrow \mathbb{R}$ is a continuous function, then extending $f$ by $\infty$ outside of $K$ gives a lower semi-continuous function.

On the other hand, if $v \notin K$ then clearly

$$h_K^\star(v) = \sup_{x \in \mathbb{R}^{m+1}} \Big[ < x, v > - < x, z(x) > \Big] = \infty.$$

Therefore

$$h_K^\star(v) = \begin{cases} 0 & \text{if} v \in K, \\ \infty & \text{otherwise.} \end{cases} \tag{1.5.7}$$

We refer to a function of the form (1.5.7) as the *associated function* of the compact set $K$. ♠

What we accomplished in the above example is more than the calculation of the conjugate of the support function of a compact set. In fact, if $f$ is a sublinear function, not identically zero, then for every $\lambda > 0$,

$$f^\star(v) = \sup_{x \in \mathbb{R}^{m+1}} \Big[ < \lambda x, v > - f(\lambda x) \Big] = \lambda f^\star(v)$$

which implies that $f^\star$ takes only values 0 and $\infty$. If we set $K = \{v \mid f^\star(v) = 0\}$, it follows easily that $K$ is compact and $f^{\star\star} = f$ is its support function. Therefore we have already shown most of:

**Proposition 1.5.2** *There is a one to one correspondence between non-trivial* $\mathbb{R}$*-valued sublinear functions on* $\mathbb{R}^{m+1}$ *and compact convex subsets of* $\mathbb{R}^{m+1}$*. The correspondence is given by assigning to each* $K$ *its support function, and the inversion is achieved by the zero set of the conjugate of a sublinear function. Under the correspondence linear functions correspond to points, and for compact convex sets* $K, K'$

$$K' \subset K \quad \Longleftrightarrow \quad h_{K'} \leq h_K.$$

**Proof** - The fact that linear functions correspond to points was noted in example 1.5.2 and is almost immediate. From the definition of $h_K$ it follows that $K' \subset K$ implies $h_K \leq h_{K'}$. The reverse implication follows from the description of the correspondence and the definition of conjugate function. ♣

**Remark 1.5.1** Let $K = \{\mathbf{0}\}$, then $h_K$ is the zero function, and $h_K^\star = \infty$. With the convention $0.\infty = 0$, the zero set of $h_K^\star$ becomes the origin which is compatible with proposition 1.5.2. ♡

If we define directional derivative as a one-sided derivative, namely,

$$f'(x;v) = \lim_{t \downarrow 0} \frac{f(x+tv) - f(x)}{t},$$

then proposition 1.5.1 and sublinearity of $h_K$ imply that directional derivatives of $h_K$ exist everywhere, and $h'_K(u;v)$ is a sublinear function of $v$. *For this reason by* directional derivative *we mean it in the one-sided sense.* The following lemma plays a key role in relating the support function to differential geometric techniques:

**Lemma 1.5.1** *For a compact convex subset $K \subset \mathbb{R}^{m+1}$, the directional derivative of $h_K$ at $u$ is given by*

$$h'_K(u;v) = h_{F_K(u)}(v).$$

**Proof** - Let $K'$ be the convex compact subset of of $\mathbb{R}^{m+1}$ corresponding to $h'_K(u;.)$ (for fixed $u$) under the correspondence of proposition 1.5.2. It follows from sublinearity of $h_K$ that

$$h'_K(u;v) \leq h_K(v).$$

Consequently $K' \subset K$ by proposition 1.5.2. Let $y \in K'$, then

$$< y, u > \leq h_K(u). \tag{1.5.8}$$

On the other hand,

$$h'_K(u;-u) = \lim_{t \downarrow 0} \frac{h_K((1-t)u) - h_K(u)}{t} = -h_K(u).$$

Therefore

$$< y, -u > \leq h_{K'}(-u) = h'_K(u, -u) = -h_K(u). \tag{1.5.9}$$

Comparing (1.5.8) and (1.5.9) we obtain $< y, u >= h_{F_K}(u)$ and consequently $K' \subset F_K(u)$. Conversely, assume $y \in F_K$, then for $\mathbf{0} \neq v \in \mathbb{R}^{m+1}$ we have $< y, v > \leq h_K(v)$. Let $v = u + tx$, then

$$< y, x >= \frac{< y, v > - < y, u >}{t} \leq \frac{h_K(u+tx) - h_K(u)}{t}$$

which implies

$$< y, x > \leq h'_K(u;x). \tag{1.5.10}$$

From the description of the correspondence between sublinear function and compact convex sets, the definition of conjugate function and (1.5.10) it follows that $y \in K'$ or $F_K(u) \subset K'$. Therefore $K' = F_K(u)$. ♣

An important consequence of lemma 1.5.1 is

**Corollary 1.5.1** *Let $K \subset \mathbb{R}^{m+1}$ be a compact convex set. Then $h_K$ is differentiable at $u \neq \mathbf{0}$ if and only if $F_K(u)$ consists of a single point. In such a case,*

$$\mathrm{grad} h_K(u) = F_K(u).$$

**Proof** - Differentiability of $h_K$ at $u$ is equivalent to linearity of $h'_K(u;.)$ which is equivalent to $K'$ consisting of a single point by proposition 1.5.2. ♣

Note that at points where $h_K$ fails to be differentiable, lemma 1.5.1 still gives an elegant description of the failure of differentiability. As a simple application of the concept of support function we prove

**Corollary 1.5.2** *A compact convex set $K \subset \mathbb{R}^{m+1}$ is uniquely determined by its projections on the lines through the origin, and consequently by its projection on two (or higher) dimensional linear spaces containing a fixed line.*

**Proof** - It follows from the definition of support function $h_K$ that it is determined by the one dimensional projections of $K$. The required result follows from example 1.5.2 or proposition 1.5.2. ♣

**Example 1.5.3** In this example we compute the Laplacian of the support function $h_K$ when the boundary $\partial K$ is a smooth hypersurface of strictly positive Gauss-Kronecker curvature. The Gauss map in this case is a diffeomorphism. The function $u \to F_K(u)$ is the inverse of the Gauss mapping and consequently

$$\mathrm{grad} h_K(u) = \mathsf{G}^{-1}(\frac{u}{||u||}), \tag{1.5.11}$$

for a non-zero vector $u \in \mathbb{R}^{m+1} \backslash \mathbf{0}$. Therefore $\Delta h_K$, where $\Delta$ denotes the Euclidean Laplacian, is the trace of the derivative of the map $e \to \mathsf{G}^{-1}(e)$. Since the derivative of $\mathsf{G}$ is the second fundamental form, $\Delta h_K$ is the sum of the eigenvalues of the inverse of the second fundamental form, i.e.,

$$\Delta h_K = \frac{1}{\kappa_1} + \cdots + \frac{1}{\kappa_m}, \tag{1.5.12}$$

where $\kappa_j$'s are the principal curvatures of $M = \partial K$. Equation (1.5.12) plays an important in the existence part of Christoffel's problem. However, since this aspect of Christoffel's problem involves a considerable amount of analysis, we will not discuss it in this volume. ♠

Finally we give the proof of proposition 1.5.1:

**Proof of proposition 1.5.1** - The proofs of the five parts of the proposition are given separately:

**1** - Since replacing $f(x)$ by $f(x+y)$ does not affect the assertion of the lemma, we may assume $x = \mathbf{0}$ is the origin. We have, for $0 \leq \alpha \leq 1$, $\alpha_i \geq 0$ with $\sum \alpha_i = 1-\alpha$ and $\mathbf{0}, y_1, \cdots y_n \in U$

$$f(\alpha\mathbf{0} + \sum \alpha_i y_i) \leq \alpha f(\mathbf{0}) + \sum \alpha_i f(y_i). \tag{1.5.13}$$

Now let $y_i$ be such that $\mathbf{0}$ lies in the interior of the convex closure of $y_i$'s and let $\alpha \to 1$ to obtain

$$\lim_{y\to\mathbf{0}} f(y) \leq f(\mathbf{0}). \tag{1.5.14}$$

Let $u \in \mathbb{R}^{m+1}$ with $|u| > 0$ small so that $\mathbf{0} \pm u$ lies in $U$. Then

$$f(\mathbf{0}) \leq \frac{1}{2} f(\mathbf{0}+u) + \frac{1}{2} f(\mathbf{0}-u). \tag{1.5.15}$$

Comparing (1.5.14) and (1.5.15) we obtain the desired continuity. ♣
**(2)** - It suffices to show that for some $\delta > 0$, $c > 0$, and all $x, y \in K$ with $|x-y| < \delta$ we have

$$|f(x) - f(y)| \leq c|x-y|.$$

Let $x_1, \cdots, x_N$ be the vertices of a small tetrahedron $\mathcal{T}$ containing $x, y$ in its interior. Assume $f(y) \geq f(x)$. We have

$$y = \alpha x + \sum \alpha_i x_i, \quad \alpha, \alpha_i \geq 0, \quad \sum \alpha_i = 1 - \alpha.$$

We can furthermore assume $\alpha_i = 0$ for some $i$ so that $y$ lies in the convex closure of $x$ and one face of the tetrahedron $\mathcal{T}$. By appropriate choice of $\mathcal{T}$ we can furthermore assume $\alpha \geq \frac{1}{2}$. By convexity of $f$ we have

$$0 \leq f(y) - f(x) \leq (1-\alpha) f(x) + \sum_i \alpha_i x_i = (1-\alpha) f(x) + \beta \sum_i \alpha_i' x_i,$$

where $\beta = \max\{\alpha_i\}$ and $\alpha_i' = \frac{\alpha_i}{\beta} \leq 1$. Clearly $1-\alpha = c_1|x-y|$ and since $\alpha \geq \frac{1}{2}$, $\beta \leq c_2|x-y|$. Therefore

$$0 \leq f(y) - f(x) \leq \left[ c_1 |f(x)| + c_2 \left| \sum \alpha_i' f(x_i) \right| \right] |x-y|. \tag{1.5.16}$$

By continuity of $f$ (item 1), there is a uniform bound

$$|f(x)|, |f(y)|, |f(x_i)| \leq c_3,$$

which together with (1.5.16) implies the required result. ♣

**(3)** - Let $e_1, \cdots, e_{m+1}$ be the standard basis for $\mathbb{R}^{m+1}$. Differentiability of $f$ means for $h = \sum h_i e_i \in \mathbb{R}^{m+1}$ we have

$$|f(x+h) - f(x) - \sum_{i=1}^{m+1} D_i f(x) h_i| = o(||h||) \quad \text{as} \quad ||h|| \to 0. \tag{1.5.17}$$

Denote the quantity inside the absolute value sign $|.|$ on the left hand side of (1.5.17) by $g(h)$. Convexity of $f$ implies that $g$ is a convex function. Let $\epsilon_i = \pm 1$, $h_i \geq 0$ and set $[h] = \sum h_i > 0$. Then

$$|g(h)| \leq \frac{1}{m+1} \sum_{i=1}^{m+1} \frac{h_i}{[h]} g(h_i \epsilon_i e_i). \tag{1.5.18}$$

Existence of partial derivatives $D_i f(x)$ implies $g(h_i \epsilon_i e_i) = o(h_i)$, which together with (1.5.18) implies $g(h) = o(||h||)$ as required. ♣

**(4)** - Let $u_j$ and $v_j$ be strictly decreasing sequences converging to $\mathbf{0}$ with $x+u_j, x+v_j, x \in U$. Let

$$\lim_{j\to\infty} \frac{f(x+u_j) - f(x)}{u_j} = A \qquad \lim_{j\to\infty} \frac{f(x+v_j) - f(x)}{v_j} = B.$$

Existence of right derivative follows once we show $A = B$. The above limits can be written as

$$f(x+u_j) = f(x) + Au_j + o(u_j), \quad f(x+v_j) = f(x) + Bv_j + o(v_j), \tag{1.5.19}$$

as $j \to \infty$ (or $u_j, v_j \downarrow 0$). After passing to a subsequence we may assume $u_j$'s and $v_j$'s intertwine, i.e.,

$$u_1 > v_1 > \cdots > u_j > v_j > u_{j+1} > v_{j+1} > \cdots.$$

Now if $B > A$, then (1.5.19) implies that for $j$ sufficiently large and any $\alpha \in [0,1]$

$$f(v_j) > \alpha f(u_j) + (1-\alpha) f(u_{j+1})$$

contradicting convexity of $f$. Therefore $A = B$. ♣

**(5)** - Let $u > 0$, then from convexity of $f$ it follows that

$$2f(x) \leq f(x+u) + f(x-u), \tag{1.5.20}$$

which implies $f'_l \leq f'_r$. Setting $y = x+u$ and $z = x-u$, (1.5.20) also implies $f'_r(z) \leq f'_l(y)$ from which remaining assertions follow. ♣

### 1.5.2 Problems of Christoffel, Minkowski and Weyl

In this subsection we investigate some uniqueness results in differential geometry generally known as rigidity with special reference to convex surfaces. An important observation in the application of the method of moving frames to deduce rigidity in this and other frameworks is the following simple lemma:

**Lemma 1.5.2** *Let $f, h : M \rightarrow G$ be smooth maps of a manifold $M$ into the analytic (linear) group $G$. Then there is $k \in G$ such that $f(x) = kh(x)$ for all $x \in M$ if and only if $f^\star(\omega) = h^\star(\omega)$ where $\omega = g^{-1}dg$.*

**Proof** - There is a function $k : M \rightarrow G$ be such that $f(x) = k(x)h(x)$, and the question is when we can make $k$ a constant. We have

$$f^{-1}df = f^{-1}(dk)h + h^{-1}dh.$$

Since $f^\star(\omega) = f^{-1}df$ and $h^\star(\omega) = h^{-1}dh$, we obtain

$$f^{-1}(dk)h = f^\star(\omega) - h^\star(\omega).$$

Therefore $dk = 0$ or $k$ is a constant if and only if $f^\star(\omega) = h^\star(\omega)$. ♣

**Corollary 1.5.3** *Let $f, h : M \rightarrow \mathbb{R}^{m+1}$ be two embeddings of an $m$-dimensional manifold in $\mathbb{R}^{m+1}$. Denote the induced metrics on $M$ by the embeddings by $ds_f^2$ and $ds_h^2$ respectively, and let $\mathsf{H}_f$ and $\mathsf{H}_h$ be the corresponding second fundamental forms. Then $f$ and $h$ differ by a Eulidean motion if and only if*

$$ds_f^2 = ds_h^2, \quad \text{and} \quad \mathsf{H}_f = \mathsf{H}_h.$$

**Proof** - We regard the embeddings $f$ and $h$ as mappings into the group $SE(m+1)$ by choosing moving frames with $e_{m+1}$ orthogonal to the images of $f$ and $h$. We write the induced Riemannian metrics in the form $ds_f^2 = \sum \omega_{i,f}^2$ and $ds_h^2 = \sum \omega_{i,h}^2$. Their equality implies that we can assume

$$\omega_{i,f} = \omega_{i,h} \tag{1.5.21}$$

and therefore we omit the subscripts referring to $f$ or $h$. Clearly

$$f^\star(\omega_{m+1}) = 0 = h^\star(\omega_{m+1}).$$

Therefore

$$d\omega_i + \sum \omega_i \wedge f^\star(\omega_{ji}) = 0 = d\omega_i + \sum \omega_i \wedge h^\star(\omega_{ji}),$$

and

$$\sum_j (f^\star(\omega_{ij}) - h^\star(\omega_{ij})) \wedge \omega_j = 0.$$

By Cartan's lemma

$$f^\star(\omega_{ij}) - h^\star(\omega_{ij}) = \sum_k b_{ijk}\omega_k, \qquad \text{with} \quad b_{ijk} = b_{ikj}.$$

On the other hand, $b_{ijk} = -b_{jik}$ which implies (see proof of proposition **??**) that $b_{ijk} = 0$. Therefore

$$f^\star(\omega_{ij}) = h^\star(\omega_{ij}) \tag{1.5.22}$$

Finally the equality of second fundamental forms implies

$$f^\star(\omega_{im+1}) = f^\star(\omega_{im+1}) \tag{1.5.23}$$

Equations (1.5.21), (1.5.22), (1.5.23) and lemma 1.5.2 imply the desired result. ♣

Let $M \subset \mathbb{R}^3$ be a compact convex submanifold of dimension 2, and $\mathsf{G} : M \to S^2$ be the Gauss map. The first observation is

**Lemma 1.5.3** *For a compact convex surface $M$, the Gaussian curvature $K_M = K$ is non-negative.*

**Proof** - The convexity condition at $x \in M$ means that the surface lies entirely on one side of the tangent plane $\mathcal{T}_x M \subset \mathbb{R}^3$. Therefore after a rotation and a translation of the surface (or coordinates) we may assume $\mathcal{T}_x M$ is the plane of $(x_1, x_2)$ of coordinates, $x$ is the origin $\mathbf{0}$ and near $\mathbf{0}$, $M$ is the graph of a non-negative function of two variables $x_1, x_2$. Therefore the Hessian of $f$ at $\mathbf{0}$ is positive semi-definite and in view of example **??** Gaussian curvature of $M$ is non-negative. ♣

The convex surface $M$ is called *strictly convex* if its Gaussian curvature is everywhere positive. We have

**Lemma 1.5.4** *Let $M \subset \mathbb{R}^3$ be a compact strictly convex surface. Then the Gauss map $\mathsf{G} : M \to S^2$ is a diffeomorphism.*

**Proof** - In view of the expression (**??**) for Gauss-Kronecker curvature $K$ of a hypersurface $M \subset \mathbb{R}^{m+1}$, positivity of $K$ implies that the Gauss map $\mathsf{G} : M \to S^m$ is a local diffeomorphism. The fact that this implies $\mathsf{G}$ is a diffeomorphism follows from the basic theory of covering spaces which is discussed in chapter 4, and therefore we assume its validity for now. ♣

Let $F = F_M : S^2 \to M$ be the inverse to the Gauss map $\mathsf{G}_M$ for the strictly convex compact surface $M$. Then $S^2$ defines a parametrization of $M$ which allows us to regard the Gaussian curvature $K_M$ as a function on $S^2$ by composing it with $F_M$. We set $\tilde{K}_M = K_M \circ F_M$. The question arises whether every positive function $H$ on $S^2$ is of the form $\tilde{K}_M$ for some compact strictly convex surface $M$. Our immediate goal is to show $K_M$ satisfies an identity which in particular proves that the answer to the question is negative.

Recall that given differential forms $\eta$ and $\theta$ with values in vector spaces $V_1$ and $V_2$, then their wedge product $\eta \wedge \theta$ is defined and is a $(p+q)$-form with values in a vector space $W$ ($\eta$ is a $p$-form and $\theta$ a $q$-form) provided we have a bilinear pairing $B : V_1 \times V_2 \to W$. In fact, if $\eta = f dx_{i_1} \wedge \cdots \wedge dx_{i_p}$ and $\theta = h dx_{j_1} \wedge \cdots \wedge dx_{j_q}$, then

$$\eta \wedge \theta = B(f,h) dx_{i_1} \wedge \cdots \wedge dx_{i_p} \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_q}.$$

Bilinearity allows us to extend this definition to general $p$ and $q$ forms. It is straightforward to see that this definition makes sense on manifolds and has the correct transformation properties, but we shall not dwell on straightforward foundations.

To derive a necessary condition for a function on $S^2$ to be the curvature of a compact strictly convex surface, we consider $\mathbb{R}^3$ with vector product $\times$ as a bilinear pairing. Then $\eta = x \times dx$ is a vector valued 1-form on $S^2$ where $x = F_M(z)$ represents a point on $M$. We choose positively oriented frame $e_1, e_2, e_3$ with $e_3$ normal to the surface $M$. Then, using $dx = \omega_1 e_1 + \omega_2 e_2$, we obtain

$$d\eta = dx \times dx = 2(\omega_1 \wedge \omega_2)(e_1 \times e_2) = 2(\omega_1 \wedge \omega_2) e_3. \tag{1.5.24}$$

Now

$$\omega_1 \wedge \omega_2 = \frac{1}{K_M} \omega_{13} \wedge \omega_{23}, \tag{1.5.25}$$

and $\omega_{13} \wedge \omega_{23} = \mathsf{G}_M^\star(dv_{S^2})$. Since by Stokes' theorem $\int_M d\eta = 0$ we obtain the equation

$$\int_{S^2} \frac{1}{\tilde{K}_M} e_3 dv_{S^2} = 0 \tag{1.5.26}$$

which is therefore a necessary condition for a positive function on $S^2$ to be the curvature of compact strictly convex surface. The analogue of this condition for polyhedra is discussed in §5.4.

Let $M, N \subset \mathbb{R}^3$ be compact strictly convex surfaces, i.e., Gaussian curvatures are positive everywhere. We use super(sub)scripts $M$ and $N$ to denote various geometric quantities associated with $M$ and $N$. Perhaps the best known examples of rigidity problems in the theory of (strictly) convex surfaces are

1. (*Christoffel*) - Assume
$$\frac{1}{\kappa_1^M}+\frac{1}{\kappa_2^M}=\frac{1}{\kappa_1^N}+\frac{1}{\kappa_2^N}$$
at all points of $M$ and $N$ where the outer unit normals $e_3^M$ and $e_3^N$ coincide. Then $M$ and $N$ differ by a translation.

2. (*Minkowski*) - Assume the Gaussian curvatures $\kappa^M$ and $\kappa^N$ coincide at points where the outer unit normals $e_3^M$ and $e_3^N$ coincide. Then $M$ and $N$ differ by a translation.

3. (*Weyl*) - Assume there is a diffeomorphism $\phi : M \rightarrow N$ such that $\phi^\star(ds_N^2) = ds_M^2$. Then $M$ and $N$ differ by a Euclidean motion.

The solutions to these problems rely on vector valued differential forms and certain integral identities. We make use of the fact that surfaces with identitical first and second fundamental forms differ by a Euclidean motion. The first two problems can be treated in a unified framework since the hypotheses are based on a common diffeomorphism with $S^2$. Weyl's problem is of a somewhat different nature since the diffeomorphism $\phi$ is not specified in terms of the Gauss map. Nevertheless, the proof is to a large extent of the same spirit as those of Christoffel and Minkowski. Since $M$ and $N$ have positive Gaussian curvatures everywhere, their Gauss maps give diffeomorphisms onto the unit sphere $S^2$. Generic points of $M$ and $N$ are denoted by $x^M$ and $x^N$ respectively. Implicit in this notation is that the parameter space is the unit sphere and the parametrization is effected by the inverse of the Gauss map. Thus for the unit normals $e_3$ it is redundant to use the superscript $M$ or $N$ and consequently $e_1, e_2$ may denote a moving frame for both $M$ and $N$. We introduce a number of differential forms some of which will be used in the solutions to the problems of Christoffel and Minkowski.
$$\begin{gathered}
A_{\circ\circ}^M =< x^M, e_3\times de_3 >,\;\; A_{\circ 1}^M =< x^M, e_3\times dx^N >,\;\; A_{1\circ}^M =< x^M, e_3\times dx^M >;\\
A_{\circ\circ}^N =< x^N, e_3\times de_3 >,\;\; A_{\circ 1}^N =< x^N, e_3\times dx^N >,\;\; A_{1\circ}^N =< x^N, e_3\times dx^M >,\\
B_{\circ\circ}^N =< x^M, x^N\times de_3 >,\;\; B_{\circ 1}^N =< x^M, x^N\times dx^N >,\;\; B_{1\circ}^N =< x^M, x^N\times dx^M > .
\end{gathered}$$
Similarly,
$$\begin{gathered}
C_{\circ\circ}^M =< x^M, de_3\times de_3 >,\;\; C_{\circ 1}^M =< x^M, de_3\times dx^N >,\;\; C_{\circ 2}^M =< x^M, dx^N\times dx^N >,\\
C_{1\circ}^M =< x^M, de_3\times dx^M >,\;\; C_{11}^M =< x^M, dx^M\times dx^N >,\;\; C_{2\circ}^M =< x^M, dx^M\times dx^M > .
\end{gathered}$$
Replacing only the first $x^M$ by $x^N$ in the defintition of $C_{rs}^M$ we obtain $C_{rs}^N$. Similarly replacing the first $x^M$ in the definition of $C_{rs}^M$ by $e_3$ we obtain a differential form which we denote by $D_{rs}$. There is no need for the superscripts $M$ or $N$ in the definition of $D_{rs}$. Denoting by $h^M$ etc. the support function for the region bounded by $M$ etc. we obtain

**Lemma 1.5.5** *With the above notation and $0 \leq r+s \leq 1$ we have*

$$dA^M_{rs} = C^M_{rs} - D_{r+1\ s} = h^M D_{rs} - D_{r+1\ s}, \quad dA^N_{rs} = C^N_{rs} - D_{r+1\ s} = h^N D_{rs} - D_{r+1\ s},$$

*and*

$$dB_{rs} = C^M_{r\ s+1} - C^N_{r+1\ s} = h^M D_{r\ s+1} - h^N D_{r+1\ s}.$$

The proof of lemma 1.5.5 is by straightforward calculations. We want to relate these quantities to the second fundamental forms and curvature functions of $M$ and $N$. The matrix of the second fundamental are denoted by

$$\mathsf{H}_M = \begin{pmatrix} a^M & b^M \\ b^M & c^M \end{pmatrix}, \quad \mathsf{H}_N = \begin{pmatrix} a^N & b^N \\ b^N & c^N \end{pmatrix}$$

The inverse to the second fundamental forms will be denoted by

$$\mathsf{H}_M^{-1} = \begin{pmatrix} \tilde{a}_M & \tilde{b}_M \\ \tilde{b}_M & \tilde{c}_M \end{pmatrix}, \quad \mathsf{H}_N^{-1} = \begin{pmatrix} \tilde{a}_N & \tilde{b}_N \\ \tilde{b}_N & \tilde{c}_N \end{pmatrix}$$

Clearly

$$\tilde{a}_M = \frac{c^M}{\kappa^M}, \quad \tilde{b}^M = \frac{-b^M}{\kappa^M}, \quad \tilde{c}_M = \frac{a^M}{\kappa^M},$$

where $\kappa^M$ denotes the Gaussian curvature of $M$. For $S^2$ we have

$$de_3 = \theta_1 e_1 + \theta_2 e_2,$$

and the 1-forms $\omega^M_{13}$ and $\omega^M_{23}$ are the pull-backs, via the Gauss map of $M$, of $\theta_1$ and $\theta_2$ respectively. For indeterminates $\xi$ and $\eta$ we set

$$\det\Big(I + \xi\mathsf{H}_M^{-1} + \eta\mathsf{H}_N^{-1}\Big) = \sum_{\circ \leq r+s \leq 2} \frac{2}{r!s!(2-r-s)!}\xi^r\eta^s P_{rs},$$

where $P_{rs}$ is a polynomial of degrees $r$ and $s$ in the entries $\mathsf{H}_M^{-1}$ and $\mathsf{H}_N^{-1}$ respectively. It is straightforward to verify that

$$2P_{rs}\theta_1 \wedge \theta_2 = D_{rs}, \tag{1.5.27}$$

so that $\det\big(I + \xi\mathsf{H}_M^{-1} + \eta\mathsf{H}_N^{-1}\big)\theta_1 \wedge \theta_2$ is like a generating function for $D_{rs}$'s. Since $\theta_1 \wedge \theta_2$ is the volume element $dv_{S^2}$ of $S^2$, (1.5.27) and lemma 1.5.5 imply

**Lemma 1.5.6** *With above notation we have, for* $0 \leq r+s \leq 1$,

$$\int \big(h^M P_{rs} - P_{r+1\ s}\big) dv_{S^2} = 0, \quad \int \big(h^N P_{rs} - P_{r\ s+1}\big) dv_{S^2} = 0, \quad \int \big(h^M P_{r\ s+1} - h^N P_{r+1\ s}\big) dv_{S^2} = 0.$$

With these preliminaries out of the way, we proceed with the solutions to the problems of Christoffel, Minkowski and Weyl.

*Christoffel's Problem* - It is an elementary calculation that

$$\det\big(\mathsf{H}_N^{-1} - \mathsf{H}_M^{-1}\big) = P_{2\circ} + P_{\circ 2} - 2P_{11} \tag{1.5.28}$$

It follows from lemma 1.5.6 that

$$\int h^M \big(P_{\circ 1} - P_{1\circ}\big) dv_{S^2} = \int \big(P_{11} - P_{2\circ}\big) dv_{S^2}, \quad \int h^N \big(P_{\circ 1} - P_{1\circ}\big) dv_{S^2} = \int \big(P_{\circ 2} - P_{11}\big) dv_{S^2}. \tag{1.5.29}$$

According to the hypotheses of Christoffel's problem $P_{\circ 1} = P_{1\circ}$. Therefore (1.5.28) and (1.5.29) imply

$$\int \Lambda dv_{S^2} = 0, \tag{1.5.30}$$

where $\Lambda$ denotes the left hand side of (1.5.28). Expanding $\Lambda$ we obtain

$$-\Lambda = (\tilde{a}_M - \tilde{a}_N)^2 + (\tilde{c}_M - \tilde{c}_N)^2 + 2(\tilde{b}_M - \tilde{b}_N)^2.$$

Thus (1.5.30) implies

$$\tilde{a}_M = \tilde{a}_N, \quad \tilde{b}_M = \tilde{b}_N, \quad \tilde{c}_M = \tilde{c}_N.$$

In particular, $M$ and $N$ have the same Gaussian curvatures at points with common normals. Since the metric form is determined by curvature, both first and second fundamental forms are equal at points with common normals. Therefore the surfaces differ by a Euclidean motion which is necessarily a translation. ♣

*Minkowski's Problem* - From lemma 1.5.6 we obtain

$$2\int h^M \big(P_{\circ 2} - P_{11}\big) dv_{S^2} = \int \Big[h^N \big(P_{11} - P_{2\circ}\big) - h^M \big(P_{11} - P_{\circ 2}\big)\Big] dv_{S^2}. \tag{1.5.31}$$

Let $\tilde{P}$ denote the element of degree two of the symmetric algebra on $\mathbb{R}^3$ such that

$$\tilde{P}(\mathsf{H}_M, \mathsf{H}_M) = P_{2\circ}$$

We need the following simple lemma whose proof is given below:

**Lemma 1.5.7** *With above notation*

$$\tilde{P}(\mathsf{H}_M^{-1}, \mathsf{H}_N^{-1}) \geq \sqrt{P_{2\circ}P_{\circ 2}},$$

*with equality if and only if* $\mathsf{H}_M^{-1} = \rho \mathsf{H}_N^{-1}$ *for some scalar* $\rho > 0$.

The equality of Gaussian curvatures as required by the hypotheses of Minkowski's problem imply $P_{\circ 2} = P_{2\circ}$. Lemma 1.5.7 implies

$$P_{11} = \tilde{P}(\mathsf{H}_M^{-1}, \mathsf{H}_N^{-1}) \geq P_{\circ 2} = P_{2\circ}. \tag{1.5.32}$$

Therefore the left hand side of (1.5.31) is non-positive. On the other hand, the right hand side of (1.5.31) is anti-symmetric in $M$ and $N$. Therefore both sides vanish identically and by lemma 1.5.7

$$\mathsf{H}_M^{-1} = \rho \mathsf{H}_N^{-1}$$

for some $\rho > 0$. The hypothesis $\kappa_M = \kappa_N$ implies $\rho = 1$ and therefore the second fundamental forms of $M$ and $N$ are identical. Therefore as in the case of Christoffel's problem, the surfaces differ by a Euclidean motion which is necessarily a translation. ♣

**Proof of lemma 1.5.7** - We write $a, a'$ etc. instead of $\tilde{a}_M, \tilde{a}_N$ etc., and note that the quantities $ac - b^2$ and $a'c' - b'^2$ are positive. Squaring and expanding, the inequality in contention becomes

$$[ac' + a'c - 2bb']^2 \geq 4(ac - b^2)(a'c' - b'^2). \tag{1.5.33}$$

The inequality is invariant under the action of $SO(2)$ on the matrices $\mathsf{H}_M^{-1}$ and $\mathsf{H}_N^{-1}$. Therefore we can assume $b = 0$ and (1.5.33) becomes

$$a^2c'^2 + a'^2c^2 \geq 2aa'cc' - acb'^2,$$

from which the required result follows. ♣

*Weyl's Problem* - As suggested by figure XXXX the assumption of convexity is necessary for a positive solution to the problem. Naturally the affirmative solution presented below is based on the establishment of equality of the second fundamental forms of $M$ and $N$ at corresponding points (the first fundamental forms are identical by the hypotheses) and then invoking proposition **??**. In other words, we have equality of $\omega_1, \omega_2$ and $\omega_{12}$ and we want to eatablish equality of $\omega_{i3}$ for $M$ and $N$.

We let $e_1^M, e_2^M, e_3^M$ be a positively oriented moving frame with $e_3^M$ normal to $M$. Consider the vector valued differential forms

$$\gamma = \omega_{31}^N e_1^M + \omega_{32}^N e_2^M, \quad \psi = x^M \cdot (e_3^M \times \gamma),$$

where $x^M$ denotes a generic point on $M$[17]. It is more accurate to write $\phi^\star(\omega^N_{3j})$ instead of $\omega^N_{3j}$, but to avoid cumbersome notation we suppress $\phi$. This should cause no confusion.

**Lemma 1.5.8** *With the above notation and the hypothesis that $\phi$ is an isometry, we have*

$$d\psi = dx^M.(e_3^M \times \gamma) + x^M.(de_3^M \times \gamma).$$

**Proof** - It is a simple calculation that

$$\begin{aligned} d\gamma &= (\omega^N_{32} \wedge \omega^N_{21})e_1^M + (\omega^M_{12} \wedge \omega^N_{31})e_2^M + (\omega^M_{13} \wedge \omega^N_{31})e_3^M + \\ &\quad (\omega^M_{21} \wedge \omega^N_{32})e_1^M - (\omega^N_{21} \wedge \omega^N_{13})e_2^M + (\omega^M_{23} \wedge \omega^N_{32})e_3^M. \end{aligned}$$

Since $\phi$ is an isometry, $\omega^M_{12} = \omega^N_{12}$. Substituting we obtain

$$d\gamma = (\omega^M_{13} \wedge \omega^N_{31} + \omega^M_{23} \wedge \omega^N_{32})e_3^M. \tag{1.5.34}$$

In particular, $e_3^M \times d\gamma = 0$. Computing the exterior derivative of $\psi$ and using $e_3^M \times d\gamma = 0$, we obtain the desired expression. ♣

Since $\int_M d\psi = 0$, it follows from lemma 1.5.8 that

$$\int_M [dx^M \cdot (e_3^M \times \gamma) + x^M \cdot (de_3^M \times \gamma)] = 0.$$

Substituting $de_3^M = \omega^M_{13}e_1^M + \omega^M_{23}e_2^M$, $dx^M = \omega_1^M e_1^M + \omega_2^M e_2^M$, and recalling the notation $h^M = x^M.e_3^M$, we obtain after a simple calculation

$$\int_M (\omega_1^M \wedge \omega^N_{32} - \omega_2^M \wedge \omega^N_{31}) = \int_M h^M(\omega^M_{31} \wedge \omega^N_{32} - \omega^M_{32} \wedge \omega^N_{31}). \tag{1.5.35}$$

The same equation (1.5.35) remains valid if we replace the superscripts $N$ by $M$. That is,

$$\int_M (\omega_1^M \wedge \omega^M_{32} - \omega_2^M \wedge \omega^M_{31}) = \int_M h^M(\omega^M_{31} \wedge \omega^M_{32} - \omega^M_{32} \wedge \omega^M_{31}). \tag{1.5.36}$$

Since $\omega^M_{12} = \omega^N_{12}$ by the isometry assumption, we have

$$\omega^M_{13} \wedge \omega^M_{32} = \omega^N_{13} \wedge \omega^N_{32}.$$

[17]The differential form $\psi$ is similar to $-A^M_{\circ\circ}$, however, since we are not using the Gauss map to identify $M$ and $N$ with $S^2$ it would be incorrect to write $-A^M_{\circ\circ}$ instead of $\psi$. The calculations that follow are in the same spirit as those for the Christoffel and Minkowski problems. For example, lemma 1.5.8 is the analogue of $dA^M_{\circ\circ} = C^M_{\circ\circ} - D_{r+1\ s}$.

Substituting in (1.5.36) and comparing with (1.5.35) we obtain

$$\int_M [\omega_1^M \wedge (\omega_{32}^N - \omega_{32}^M) + \omega_2^M \wedge (\omega_{31}^N - \omega_{31}^M)] = \int_M h^M (\omega_{31}^M - \omega_{31}^N) \wedge (\omega_{32}^M - \omega_{32}^N). \tag{1.5.37}$$

The following simple algebraic observation plays an important role:

**Lemma 1.5.9** *Let $S_1$ and $S_2$ be $2 \times 2$ positive definite symmetric matrices with $\det S_1 = \det S_2$. Then $\det(S_1 - S_2) \leq 0$ with equality if and only if $S_1 = S_2$.*

**Proof** - Replacing $S_j$ by $AS_jA^{-1}$ for some $A \in SO(2)$ we may assume $S_1$ is diagonal. The required result follows easily by an elementary calculation. ♣

Lemma 1.5.9 implies

**Lemma 1.5.10** *With the above notation*

$$(\omega_{31}^M - \omega_{31}^N) \wedge (\omega_{32}^M - \omega_{32}^N) = f\omega_1^M \wedge \omega_2^M,$$

*where $f$ is non-positive function and is zero if and only if $\omega_{3j}^M = \omega_{3j}^N$.*

**Proof** - In terms of the second fundamental forms $\mathsf{H}_M$ and $\mathsf{H}_N$, we have

$$(\omega_{31}^M - \omega_{31}^N) \wedge (\omega_{32}^M - \omega_{32}^N) = \det(\mathsf{H}_M - \mathsf{H}_N)\omega_1^M \wedge \omega_2^M.$$

Now the isometry condition implies that the second fundamental forms of $M$ and $N$ have the same determinant (Gaussian curvature). Therefore by lemma 1.5.9, $f = \det(\mathsf{H}_M - \mathsf{H}_N)$ is non-positive and is zero only if $\mathsf{H}_M = \mathsf{H}_N$. ♣

Now we can complete the solution to Weyl's problem with an argument similar to that given for Minkowski's problem. We can assume $h^M > 0$ and $h^N > 0$, after possibly transforming them by Euclidean motions so that the origin lies in the interior of both $M$ and $N$. In view of lemma 1.5.10, the right hand side of (1.5.37) remains unchanged if we interchange the roles of $M$ and $N$. In view of the isometry hypothesis we have $\omega_j^M = \omega_j^N$ after possibly replacing $N$ by $g(N)$ for any orthogonal matrix $g$ with $\det(g) = -1$. Therefore the left hand side of (1.5.37) is multiplied by $(-1)$ if we interchange the roles of $M$ and $N$. It follows that both sides of (1.5.37) vanish and by lemma 1.5.10 the integrand on right hand side vanishes. The same lemma implies $\mathsf{H}_M = \mathsf{H}_N$ and the required result follows. ♣

The uniqueness or rigidity results proven in this subsection can be generalized to hypersurfaces. The first difficulty one encounters is the definition of analogues of the differential forms $A_{rs}^M$ etc. This is achieved by replacing $< \mathbf{a}_1, \mathbf{a}_2 \times \mathbf{a}_3 >$ with $\det(\mathbf{a}_1, \cdots, \mathbf{a}_m)$. One should be cognizant of the fact that since $\mathbf{a}_j$'s are forms, it is necessary to verify well-definedness of the determinant. Relations similar to $dA_{rs}^M = C_{rs}^M - D_{r+1\ s}$ are easily verified by exterior differentiation. On the other hand, the relevant analogue of the inequality in lemma 1.5.7 is more subtle. For an account of these issues see [Ch4].

### 1.5.3 Mixed Volumes

In order to understand the concepts of mixed volumes and related measures it is useful to examine a very simple special case first.

**Example 1.5.4** Let $K \subset \mathbb{R}^3$ be a rectangular cube and $B$ be the unit ball in $\mathbb{R}^3$. Then $K + \rho B$, $\rho > 0$, is shown in figure XXXX. Its volume can be decomposed into a sum of four terms:

$$v_3(K + \rho B) = v_{111} + 3v_{112}\rho + 3v_{122}\rho^2 + v_{222}\rho^3. \tag{1.5.38}$$

Here $v_{111}$ is the volume of $K$, $v_{222}$ is the volume of the unit ball, $v_{112}$ is one third the area of $\partial K$ and $v_{122}$ is $\frac{\pi}{12}$ the sum of the lengths of the edges of $K$. It is this kind of formula which one would like to generalize to sums of compact convex sets. ♠

For a subset $X \subset \mathbb{R}^{m+1}$ we let $v_n(X)$ denote its volume as an $n$-dimensional object. One way of generalizing example 1.5.4 is to prove a formula such as

$$v_{m+1}(K + \rho B^{m+1}) = \sum_{j=0}^{m+1} \rho^j \binom{m+1}{j} W_j(K), \tag{1.5.39}$$

where $K \subset \mathbb{R}^{m+1}$ is a compact convex set, $B^{m+1} \subset \mathbb{R}^{m+1}$ is the unit ball, and geometrically interpret the quantities $W_j(K)$. Still a more general version is to express the volume of a sum $m+1$ compact convex subsets $K_1, \cdots, K_{m+1} \subset \mathbb{R}^{m+1}$ as

$$v_{m+1}(\alpha_1 K_1 + \cdots + \alpha_{m+1} K_{m+1}) = \sum_{i_1 \cdots i_{m+1}} \alpha_{i_1} \cdots \alpha_{i_{m+1}} v_{m+1}(K_{i_1}, \cdots, K_{i_{m+1}}), \tag{1.5.40}$$

where the indices $i_1, \cdots, i_{m+1}$ run over all possible choices (possibly with repetitions) from $\{1, \cdots, m+1\}$, $\alpha_j \geq 0$, and give a geometric interpretation to the quantity $v_{m+1}(K_{i_1}, \cdots, K_{i_{m+1}})$. Formula (1.5.39) is a special case of (1.5.40). In fact, for $0 \leq j \leq m+1$ and compact convex sets $K_1$ and $K_2$ define

$$V(K_1, K_2; j) = v_{m+1}(\underbrace{K_1, \cdots, K_1}_{m+1-j}, \underbrace{K_2, \cdots, K_2}_{j}).$$

Now setting $K_1 = K, K_2 = B^{m+1}$, $\alpha_1 = 1, \alpha_2 = \rho$, $\alpha_j = 0$ for $j \geq 2$ and $W_j(K) = V(K_1, K_2; j)$ we obtain (1.5.39) from (1.5.40) thereby defining *projection or cross-section measure, or quermaßintegral* $W_j(K)$ as a mixed volume. Formula (1.5.39) is often called *Steiner's formula.*

The validity of (1.5.40) determines the the quantities $v_{m+1}(K_{i_1},\cdots,K_{i_{m+1}})$. To see this note that by setting $\alpha_j=0$ for some $j$'s, we obtain the formulae

$$v_{m+1}(\alpha_1K_1+\cdots+\alpha_kK_k)=\sum\alpha_{i_1}\cdots\alpha_{i_{m+1}}v_{m+1}(K_{i_1},\cdots,K_{i_{m+1}}), \tag{1.5.41}$$

where the summation is over all indices $i_1,\cdots,i_{m+1}$ from $\{1,\cdots,k\}$. The following lemma shows that these equations can be inverted algebraically:

**Lemma 1.5.11** *With the above notation, the equation (1.5.40) (or (1.5.41)) is inverted as*

$$v_{m+1}(K_1,\cdots,K_{m+1})=\frac{1}{(m+1)!}\sum_{l=1}^{m+1}(-1)^{m+1+l}\sum_{i_1<i_2<\cdots<i_l}v_{m+1}(K_{i_1}+K_{i_2}+\cdots+K_{i_l}).$$

(This equation is often called the *polarization formula*.)

**Proof** - To prove the assertion replace $K_i$ by $\alpha_iK_i$, $\alpha_i\geq 0$, in the formula in question. It follows from the (1.5.41) that the right hand side of the second equation is homogeneous of degree $m+1$ in $\alpha_1,\cdots,\alpha_{m+1}$. Now observe that if we set $\alpha_1=0$ and $\alpha_j=1$ for $j\neq 1$, the right hand side of the equation vanishes identically. Since the index 1 can be replaced with any other, the only term on the right hand side will be the coefficient of $\alpha_1\cdots\alpha_{m+1}$. Substituting from (1.5.41) we see that the right hand side is the mixed volume $v_{m+1}(K_1,\cdots,K_{m+1})$. ♣

In view of lemma 1.5.11 we define the (*Minkowski*) *mixed volume* as

$$v_{m+1}(K_1,\cdots,K_{m+1})=\frac{1}{(m+1)!}\sum_{l=1}^{m+1}(-1)^{m+1+l}\sum_{i_1<i_2<\cdots<i_l}v_{m+1}(K_{i_1}+K_{i_2}+\cdots+K_{i_l}). \tag{1.5.42}$$

While (1.5.42) defines mixed volumes, it is difficult to to deduce the basic properties of mixed volumes (e.g., the homogeneity property (1.5.40) directly and immediately from this expression. In fact, the expression in (1.5.42) was a purely formal derivation but does not extend to non-convex sets[18]. Therefore the validity of (1.5.40) and other basic properties of mixed volumes given below, which are limited to convex sets, depend on some remarkable and non-trivial cancellations in (1.5.42). Complete proofs of (1.5.40) and its ramifications require a detailed study of the structure of polytopes and the approximation, relative to the Hausdorf metric, of compact convex sets by polytopes. Since this would be quite lengthy we

[18]A modification of the theory of mixed volumes where sets are replaced by support functions admits of extensions to non-convex sets, but this is not the issue here.

concentrate on explaining the geometric content and ideas underlying the theory, without a formal development of the theory of convex polytopes, and refer the reader to, e.g., [Sch], for such a treatment and extensive references to the literature.

Let $P$ be a convex polytope in $\mathbb{R}^{m+1}$ with non-empty interior. Then

$$\partial P = F_1 \cup \cdots \cup F_N,$$

where each $F_j$ is a convex polytope of dimension $m$. We refer to $F_j$'s as *facets* of $P$.

**Lemma 1.5.12** *Let $P$ be a polytope in $\mathbb{R}^{m+1}$ and assume that the origin $\mathbf{0}$ lies in the interior of $P$. Let $e_j$ denote the unit outward normal to the facet $F_j$. Then*

$$\sum_j v_m(F_j)e_j = 0, \quad v_{m+1}(K) = \frac{1}{m+1}\sum_j h_K(e_j)v_m(F_j).$$

**Proof** - Both sides of the first identity are invariant under translations. Therefore we can assume that the origin $\mathbf{0}$ is in the exterior of $P$. Let $f \in S^m$ be a vector such that $P \cap (\mathbb{R}f)^\perp = \emptyset$. Then the orthogonal projection $P'$ of $P$ to $(\mathbb{R}f)^\perp$ is a polytope whose set of extreme points is the projection of a subset $\mathcal{E}'$ of the extreme points $\mathcal{E}(P)$ of $P$. It follows that

$$v_m(P') = \sum_{j:<f,e_j>>0} < f, e_j > v_m(F_j) = -\sum_{j:<f,e_j><0} < f, e_j > v_m(F_j). \tag{1.5.43}$$

Since $f$ is arbitrary from an open set of vectors in $S^m$, the first identity follows from (1.5.43). After possibly a subdivision of the faces of $P$ we may assume each facet $F_j$ is a tetrahedron of dimension $m$. Then the $(m+1)$-dimensional volume of the convex closure of $F_j$ and $\mathbf{0}$ is

$$\frac{1}{m+1}h_{F_j}(e_j)v_m(F_j),$$

from which the second identity follows. ♣

**Remark 1.5.2** The assumption that $\mathbf{0}$ lies in the interior of the polytope $P$ in unnecessary. In fact, if $\mathbf{0}$ lies in the exterior of the tetrahedron $K$ then the quantities $h_{F_j}(e_j)$ will have both positive and negative signs and the $\frac{1}{m+1}h_{F_j}(e_j)v_m(F_j)$'s add up correctly to $v_{m+1}(K)$. We have already made use of a special case of this phenomenon in connection with signed areas and example 1.1.2. ♡

The key concept in understanding the geometry of polytopes and the necessary approximation theory is that of the normal cone which we now introduce. For a convex set $K$ and $x \in K$ we set

$$\mathsf{N}_K(x) = \{u \in \mathbb{R}^{m+1} \mid x \in \mathsf{H}_K(u)\} \cup \{\mathbf{0}\}.$$

$\mathsf{N}_K(x)$ is called the *normal cone* of $K$ at $x \in K$. Two polytopes $P$ and $P'$ are called *similar* if the sets

$$\{\mathsf{N}_P(x) \mid x \in \mathcal{E}(P)\} \quad \text{and} \quad \{\mathsf{N}_{P'}(x) \mid x \in \mathcal{E}(P')\}$$

are identical. It is clear that similarity is an equivalence relation, and $P$, $\alpha P$ and $v+P$, where $\alpha > 0$ and $v \in \mathbb{R}^{m+1}$, are similar. We will see shortly that the class of similar polytopes is sufficiently large to allow certain approximations to compact convex sets.

**Exercise 1.5.7** *Determine when two compact convex n-gons in the plane are similar.*

**Exercise 1.5.8** *Consider the rectangular cube $P$ with vertices at $(\pm 1, \pm 1, \pm 1)$. Show that the normal cones $\mathsf{N}_P(x)$ as $x$ runs over the eight extreme point of the cube are precisely the coordinate octants.*

Exercise 1.5.8 reflects a general geometric phenomenon which is described in the following exercise (see also lemma 1.5.13-(4) below):

**Exercise 1.5.9** *For $x \in P$ let $\mathsf{S}_P(x)$ denote the intersection of all half spaces $\mathsf{H}^-_{(u,\gamma)}$ containing $P$ and such that $x \in \partial\mathsf{H}^-_{(u,\gamma)}$. If $x \in \mathcal{E}(P)$ then $\mathsf{S}_P(x)$ is the smallest cone with its vertex at $x$ and containing $P$, and $\mathsf{N}_P(x)$ is dual to $\mathsf{S}_P(x)$ in the sense*

$$\mathsf{N}_P(x) = \{u \in \mathbb{R}^{m+1} \mid \ < u, y > \geq 0 \text{ for all } y \in x - \mathsf{S}_P(x)\}.$$

More generally for a face $F$ of the polytope $P$ we define the *normal cone on $F$* as

$$\mathsf{N}_P(F) = \{u \in \mathbb{R}^{m+1} \mid F \subset \mathsf{H}_P(u)\}.$$

The *cone closure* of a set of vectors $e_1, \cdots, e_N$ is

$$\mathsf{Cone}(e_1, \cdots, e_n) = \{\alpha_1 e_1 + \cdots + \alpha_N e_N \mid \alpha_j \geq 0\}$$

The following elementary observations are the essential technical tools:

**Lemma 1.5.13** *Let $P$ and $P'$ be compact convex polytopes. Denote the facets of $P$ by $F_1, \cdots, F_N$ and outward normal to $F_i$ by $e_i$. Then*

1. *For $x \in \mathcal{E}(P)$, the normal cone $\mathsf{N}_P(x)$ has non-empty interior.*
2. *If $x \notin \mathcal{E}(P)$, then $\mathsf{N}_P(x)$ has empty interior.*
3. *For distinct vertices $x, x'$ of $P$, the intersection $\mathsf{N}_P(x) \cap \mathsf{N}_P(y)$ has empty interior.*
4. *If $P$ has non-empty interior and $F$ is a face of $P$ then*
$$\mathsf{N}_P(F) = \mathsf{Cone}(e_{i_1}, \cdots e_{i_r}),$$
*where $F_{i_1}, \cdots, F_{i_r}$ are the facets of $P$ containing $F$.*
5. *For $x \in K$, $y \in P'$ we have*
$$\mathsf{N}_{P+P'}(x+y) = \mathsf{N}_P(x) \cap \mathsf{N}_{P'}(y).$$

The proof of the lemma is straightforward and is omitted. An immediate consequence is

**Corollary 1.5.4** *Let $P$ and $P'$ be similar polytopes. Then $P + P'$ is similar to $P$ and $P'$.*

**Proof** - Since $\mathcal{E}(P + P') \subset \mathcal{E}(P) + \mathcal{E}(P')$ the required result follows from lemma 1.5.13. ♣

For a polytope $P$ there is a connection between the subsets $F_P(u)$ and the normal cones $\mathsf{N}_P(x)$. It is clear that every non-zero vector $u$ lies in a normal cone $\mathsf{N}_P(x)$ for some vertex $x$ of $P$. Generically $u$ lies in the interior of a cone $\mathsf{N}_P(x)$ and consequently $\mathsf{H}_P(u) \cap P = \{x\}$. If $u$ lies on the boundary of $\mathsf{N}_P(x)$ for a vertex $x$, then there is a unique face $F$ of $\mathsf{N}_P(x)$ such that $u$ in the (relative) interior of $F$. Let $x_1, \cdots, x_r$ be all the vertices of $P$ such that $u \in \mathsf{N}_P(x_i)$. It follows that $x_1, \cdots, x_r$ are the vertices of $F_P(u)$ of $P$ and

$$\dim F_P(u) = m + 1 - \dim F. \tag{1.5.44}$$

The above geometric picture allows us to establish a correspondence between the subsets $F_P(u)$ and $F_{P'}(u)$ of similar convex polytopes $P$ and $P'$. For a non-zero vector $u$ let $x_1, \cdots, x_r$ be as above, and let $x'_j \in \mathcal{E}(P')$ be such that $\mathsf{N}_P(x_j) = \mathsf{N}_{P'}(x'_j)$. Then $x'_1, \cdots, x'_r$ are the vertices of $F_{P'}(u)$ and in view of (1.5.44)

$$\dim F_P(u) = \dim F_{P'}(u). \tag{1.5.45}$$

We also have

**Lemma 1.5.14** *The faces $F_P(u)$ and $F_{P'}(u)$ of similar polytopes $P$ and $P'$ are similar.*

**Proof** - The proof is a simple application of the above ideas. ♣

Now we are in a position to derive another expression for Minkowski mixed volumes for similar polytopes. For each $i$ let $F_{ij}$, run over the $(m-1)$ faces of the polytope $F_i$. Let $S_i^{m-1}$ denote the unit sphere in the hyperplane orthogonal to $e_i$ and for fixed $i$, let $e_{ij} \in S_i^{m-1}$, run over the normals to the $(m-1)$-dimensional faces of $F_i$. We also set $h_i = h_P(e_i)$ and $h_{ij} = h_{F_i}(e_{ij})$. The number $h_i, h_{ij}, \cdots$ is called the *support numbers* of $P$. The $(m-1)$ dimensional polytope $F_{ij}$ is also the intersection of $F_i$ with another facet of $P$ which we naturally denote by $F_j$ so that $F_{ij} = F_i \cap F_j$. Denoting the angle between the vectors $e_i, e_j \in S^m$ by $\theta_{ij}$, we deduce that the unit vector parallel to the facet $F_i$ and orthogonal to $F_{ij}$ is

$$e_{ij} = \frac{\pm 1}{\sqrt{1+\cos^2\theta_{ij}}}[(\cos\theta_{ij})e_i - e_j].$$

Since

$$\sup_{x\in F_i} < e_i, x >=< e_i, y >, \quad \text{for all} \quad y \in F_i,$$

we obtain

$$h_{ij} = h_{F_i}(e_{ij}) = \frac{\pm 1}{\sqrt{1+\cos^2\theta_{ij}}}[(\cos\theta_{ij})h_i - h_j]. \tag{1.5.46}$$

Therefore by substituting from (1.5.46) in lemma 1.5.12 for the volume of $F_i$, we obtain an expression of the form

$$v_m(F_i) = \sum_j a_j^{(i)} h_j \tag{1.5.47}$$

where the coefficients $a_j^{(i)}$ depend only on the angles $\theta_{ij}$ and the summation may be limited to only those $j$ such that $\dim(F_i \cap F_j) = m-1$. The fact that the coefficients $a_j^{(i)}$ depend only on the angles $\theta_{ij}$ implies that they depend only on the similarity class of the polytope $P$. This fact plays an important role in the development of theory of mixed volumes. We can now prove

**Lemma 1.5.15** *Let $P \subset \mathbb{R}^{m+1}$ be a polytope with facets $F_1, \cdots, F_N$ and support numbers $h_1, \cdots, h_N$. Then*

$$v_{m+1}(P) = \sum_{i_1,\cdots,i_{m+1}} a_{i_1,\cdots,i_{m+1}} h_{i_1}\cdots h_{i_{m+1}},$$

*where the coefficients $a_{i_1,\cdots,i_{m+1}}$ depend only on the similarity class of $P$.*

**Proof** - The proof is by induction on $m$. For $m = 1$ this is a simple exercise. Using lemma 1.5.12 and (1.5.47), the induction is easily completed. ♣

Now assume $P_1, \cdots P_{m+1}$ are similar polytopes. Denote the support numbers of the polytope $P_j$ by $h_1^{(j)}, h_2^{(j)}, \cdots$. We set

$$v'_{m+1}(P_1, \cdots, P_{m+1}) = \sum_{i_1,\cdots,i_{m+1}} a_{i_1,\cdots,i_{m+1}} h_{i_1}^{(1)} \cdots h_{i_{m+1}}^{(m+1)}. \tag{1.5.48}$$

With this definition we have:

**Lemma 1.5.16** *Let $P_1, \cdots, P_{m+1}$ be similar polytopes. Then for $\alpha_i \geq 0$ we have*

$$v_{m+1}(\alpha_1 P_1 + \cdots + \alpha_{m+1} P_{m+1}) = \sum \alpha_{i_1} \cdots \alpha_{i_{m+1}} v'_{m+1}(P_{i_1}, \cdots, P_{i_{m+1}}),$$

*where the indices $i_1, \cdots, i_{m+1}$ range over $1, \cdots, m+1$ independently.*

**Proof** - Since $h_{K+K'} = h_K + h_{K'}$ the required result follows from lemma 1.5.15 and the definition of $v'_{m+1}$. ♣

Another implication of the definition of $v'_{m+1}$ and the additivity of $h_K$ is

**Lemma 1.5.17** *Let $P_1, \cdots, P_{m+1}, Q$ be similar polytopes. Then $v'_{m+1}(P_1, \cdots, P_{m+1})$ is symmetric in the arguments $P_1, \cdots, P_{m+1}$ and*

$$v'_{m+1}(Q + P_1, P_2, \cdots, P_{m+1}) = v'_{m+1}(P_1, \cdots, P_{m+1}) + v'_{m+1}(Q, P_2, \cdots, P_{m+1}).$$

Since the inversion in lemma 1.5.11 was a purely formal derivation, lemma 1.5.16 implies

**Lemma 1.5.18** *With the above notation, for similar polytopes $P_1, \cdots, P_{m+1}$, we have*

$$v'_{m+1}(P_1, \cdots, P_{m+1}) = v_{m+1}(P_1, \cdots, P_{m+1}).$$

Lemmas 1.5.16, 1.5.17 and 1.5.18 imply the the mixed volume $v_{m+1}$ has the desired properties of homogeneity, additivity and symmetry on similar polytopes. To deduce the same for all compact convex subsets of $\mathbb{R}^{m+1}$ we make use of an approximation lemma (1.5.19 below). The approximations are relative to the *Hausdorf distance* of compact subsets of $\mathbb{R}^{m+1}$ which is defined as

$$d(K, K') = \max(\sup_{x\in K} \inf_{y\in K'} ||x - y||, \sup_{y\in K'} \inf_{x\in K} ||x - y||)$$

A useful property of Hausdorf metric is that if $\mathcal{K}_R$ denotes the family of compact convex subsets of $\mathbb{R}^{m+1}$ contained in the ball of radius $R > 0$, then $\mathcal{K}_R$ is compact. This fact is known as *Blaschke's Selection* lemma and its proof is straightforward real analysis.

**Remark 1.5.3** Let $P_1$ and $P_2$ be similar polytopes. It follows from lemma 1.5.18 and the definition of $v'_{m+1}(P_1, \cdots, P_m, P_2)$ that

$$v_{m+1}(P_1, \cdots, P_1, P_2) = \frac{1}{m+1} \sum_j h_{P_2}(e_j) a_j, \tag{1.5.49}$$

where the summation is over all outward unit normals $e_j$ to the faces of $P_1$ (or $P_2$ since they are similar) and $a_j$ is the area (volume) of the facet of $P_1$ with outward unit normal $e_j$. One can give similar interpretations to the quantities $v_{m+1}(P_1, \cdots, P_1, P_2, \cdots, P_2)$, where the areas (volumes) $a_j$ are replaced by those of lower dimensional faces and $h_{P_2}(e_j)$ by products of support numbers. ♡

Since the set of extreme points of a compact convex set $K$ is contained in $\partial K$, we can approximate $K$ with polytopes arbitrarily closely. The following lemma is the key approximation tool:

**Lemma 1.5.19** *Compact convex sets $K_1, \cdots, K_N \subset \mathbb{R}^{m+1}$ can be arbitrarily closely approximated by similar polytopes.*

**Proof** - Let $Q_i$ be a polytope approximating $K_i$ by $\frac{\epsilon}{2}$, $P = Q_1 + \cdots + Q_N$. Then $P_i = Q_i + \alpha P$, for $\alpha > 0$ sufficiently small is an $\epsilon$ approximation to $K_i$. The normal cones of the vertices of $P$ and therefore those of $\alpha P$ are contained in those of $Q_i$ for all $i$ in view of lemma 1.5.13-(5). Another application of the same lemma shows that the normal cones of $P_i$ are identical with those of $P$ and consequently the polytopes $P_i$ are similar. ♣

The proof of lemma 1.5.19 shows that it is quite easy to construct similar polytopes. The following exercise, which is independent of the above lemma, provides a method for generating similar polytopes through appropriate perturbations:

**Exercise 1.5.10** *Let $P$ be a polytope with non-empty interior and facets $F_1, \cdots, F_N$ and corresponding outward unit normals $e_1, \cdots, e_N$. Show that for $|\epsilon_1|, \cdots, |\epsilon_N|$ sufficiently small the polytope*

$$P_{\epsilon_1, \cdots, \epsilon_N} = \bigcap_{i=1}^{N} \mathsf{H}^-_{(e_i, h_P(e_i)+\epsilon_i)}$$

*is similar to $P$.*

A very special case of lemma 1.5.19 is the following exercise which can be done more or less explicitly without any reference to the lemma:

**Exercise 1.5.11** *Consider the convex sets in the plane defined as*

$$K: \ \frac{x^2}{a_1^2} + \frac{y^2}{b_1^2} \le 1; \qquad K': \frac{x^2}{a_2^2} + \frac{y^2}{b_2^2} \le 1.$$

*Show that $K$ and $K'$ can be approximated arbitrarily closely by similar convex polytopes $P$ and $P'$.*

Lemma 1.5.19 and the continuity of volumes relative to the Hausdorf distance shows that the basic properties of mixed volumes, which were established for similar polytopes, remain valid for arbitrary compact convex subsets of $\mathbb{R}^{m+1}$. We summarize the above conclusions and other properties of mixed volumes in the form of a proposition for easy reference:

**Proposition 1.5.3** *Let $K, K_1, \cdots, K_{m+1} \subset \mathbb{R}^{m+1}$ be compact convex sets, and define the mixed volume $v_{m+1}(K_1, \cdots, K_{m+1})$ as in lemma 1.5.11. Then $v_{m+1}$ has the following properties:*

1. *$v_{m+1}$ is continuous relative to the Hausdorf distance and symmetric in the arguments.*
2. *For $\alpha_1 \ge 0, \cdots, \alpha_{m+1} \ge 0$ formula (1.5.40) is valid.*
3. *$v_{m+1}$ is additive in each argument, i.e.,*
$$v_{m+1}(K + K_1, K_2 \cdots, K_{m+1}) = v_{m+1}(K_1, \cdots, K_{m+1}) + v_{m+1}(K, K_2, \cdots, K_{m+1}).$$
4. *$v_{m+1}(K, K, \cdots, K) = v_{m+1}(K)$.*
5. *$v_{m+1}$ is argument wise invariant under translations $T$ in the sense that*
$$v_{m+1}(T(K_1), K_2, \cdots, K_{m+1}) = v_{m+1}(K_1, K_2, \cdots, K_{m+1}).$$
6. *$v_{m+1}$ is invariant under $GL(m+1, \mathbb{R})$ acting diagonally, i.e., for $A \in GL(m+1, \mathbb{R})$ we have*
$$v_{m+1}(A(K_1), A(K_2), \cdots, A(K_{m+1})) = |\det(A)| v_{m+1}(K_1, K_2, \cdots, K_{m+1}).$$
7. *$v_{m+1}(K_1, \cdots, K_{m+1}) \ge 0$.*
8. *If $K \subset K_1$ then*
$$v_{m+1}(K, K_2, \cdots, K_{m+1}) \le v_{m+1}(K_1, K_2, \cdots, K_{m+1}).$$

**Proof** - We have already proven most of these statements and will only comment on the remaining ones. Item (4) follows from lemma 1.5.15 and the definition of $v'_{m+1} = v_{m+1}$. Items (5) and (6) follow from lemma 1.5.11. After applying translations we may assume the origin lies in the interior of $K_1 \cap \cdots \cap K_{m+1}$. Then an examination of the construction of $v'_{m+1}$ shows that we can assume the support numbers $h_j > 0$ and the coefficients $a_{i_1,\cdots,i_{m+1}} \geq 0$. This implies (7). The additional observation that $h_K \leq h_{K_1}$ implies (8). ♣

There are a number of inequalities involving mixed volumes some of which are quite subtle. Here we use the Brunn-Minkowski inequality (see Chapter 1, proposition **??**) to derive a simple inequality and use it for the existence result in proposition 1.5.5 and rigidity in corollary 1.5.5.

**Proposition 1.5.4** (Minkowski Inequality) - *Let $K_1, K_2 \subset \mathbb{R}^{m+1}$ be compact convex sets with non-empty interior. Then*

$$v_{m+1}(K_1, \cdots, K_1, K_2)^{m+1} \geq v_{m+1}(K_1)^m v_{m+1}(K_2).$$

*Equality holds if and only if $K_2 = \mathbf{v} + \beta K_1$, i.e., $K_i$ differ by a homothety and a translation.*

**Proof** - Let $K_\alpha = (1-\alpha)K_1 + \alpha K_2$ and set

$$\phi(\alpha) = v_{m+1}(K_\alpha)^{\frac{1}{m+1}} - (1-\alpha)v_{m+1}(K_1)^{\frac{1}{m+1}} - \alpha v_{m+1}(K_2)^{\frac{1}{m+1}}.$$

Clearly $\phi(0) = \phi(1) = 0$. The homogeneity property of mixed volumes yields

$$v_{m+1}(K_\alpha) = \sum_{j=0}^{m+1} \binom{m+1}{j} (1-\alpha)^{m+1-j} \alpha^j v_{m+1}(K_1, K_2; j). \tag{1.5.50}$$

Substituting from (1.5.50), differentiating $\phi(\alpha)$ and setting $\alpha = 0$ we obtain

$$\phi'(0) = v_{m+1}(K_1)^{-\frac{m}{m+1}} \left[ v_{m+1}(K_1, K_2, ; 1) - v_{m+1}(K_1)^{\frac{m}{m+1}} v_{m+1}(K_2)^{\frac{1}{m+1}} \right] \tag{1.5.51}$$

On the other hand, the Brunn-Minkowski inequality implies concavity of the function $\phi(\alpha)$ and consequently

$$\phi(\alpha) \geq 0. \tag{1.5.52}$$

(1.5.51) and (1.5.52) imply the required inequality. In view of concavity of $\phi$ equality holds only if $\phi$ is identically 0. Proposition **??** of chapter 1 implies the second assertion. ♣

**Exercise 1.5.12** *Let $K_1, K_2 \subset \mathbb{R}^{m+1}$ be compact convex subsets with non-empty interior. Show that*

$$v_{m+1}(K_1, K_2; 1)^2 \geq v_{m+1}(K_1)v_{m+1}(K_1, K_2; 2).$$

(Differentiate the function $\phi$ in the proof of proposition 1.5.4 twice at 0.)

A concept related to mixed volumes is that of mixed area measures which are Borel measures on the sphere $S^m$. If the compact set $K$ has smooth boundary $\partial K$ with the standard volume element $dv_{\partial K}$ induced from the Euclidean metric of $\mathbb{R}^{m+1}$, then we set

$$ds_m(K, .) = \mathsf{G}_\star(dv_{\partial K}),$$

where $\mathsf{G}$ is the Gausss map of the boundary $\partial K$. This means that for a Borel measurable function $\phi : S^m \to \mathbb{R}$ we define its integral as

$$\int_{S^m} \phi(u) ds_m(K, u) = \int_{\partial K} \phi(\mathsf{G}(x)) dv_{\partial K}(x).$$

In particular, for a subset $U \subset S^m$ its $ds_m(K, .)$ measure is the volume, relative to $dv_{\partial K}$, of the set $\mathsf{G}^{-1}(U)$. Since $\partial K$ is generally not a manifold, in order to extend this definition to general compact convex sets we make use use of approximation by convex polytopes. The natural extension of the definition of $ds_m(K, .)$ to polytopes is as follows: Let $\partial P = F_1 \cup \cdots \cup F_N$ be the decomposition of the boundary of the polytope $P$ into facets and $e_j$ be the unit outward normal to $F_j$. Then $ds_m(P, .)$ is the atomic measure, supported on $\cup_j \{e_j\}$, which assigns $v_m(F_j)$ to $e_j$. It is clear that if the sequence of polytopes $P_j$ converges to $K$ relative to the Hausdorf distance, and $K$ has smooth boundary, then $ds_m(P_j, .)$ converges weakly[19]to $ds_m(K, .)$ as defined earlier. Therefore for a general compact convex set we define $ds_m(K, .)$ as the weak limit of the measures $ds_m(P_j, .)$ for any sequence of polytopes converging to $K$ in the Hausdorf metric. The proof that this is well-defined is straightforward and is omitted.

Having defined the area measures $ds_m(K, .)$ we proceed to define mixed area measures much in the same the same way as mixed volumes are related to volumes. The analogue of the polarization formula (lemma 1.5.11) may be used as the definition of *mixed area measures*:

$$ds_m(K_1, \cdots, K_m, .) = \frac{1}{m!} \sum_{k=1}^{m} (-1)^{m+k} ds_m(K_{i_1} + \cdots + K_{i_k}, .), \qquad (1.5.53)$$

[19]A sequence of (Borel) measures $\mu_j$ converges weakly to a (Borel) measure $\mu$ if for every (Borel) meaurable function $\int \phi d\mu_j$ converges to $\int \phi d\mu$.

which can be derived formally from the inversion of the homogeneity requirement

$$ds_m(\alpha_1 K_1 + \cdots + \alpha_m K_m, .) = \sum \alpha_{i_1} \cdots \alpha_{i_m} ds_m(K_{i_1}, \cdots, K_{i_m}, .), \tag{1.5.54}$$

where the summation is over all indices $i_1, \cdots, i_{m+1}$ from $\{1, \cdots, k\}$. Just as in the case of mixed volumes it is more useful to first define mixed area measures for similar polytopes and then extend it to all compact convex sets via the approximation lemma 1.5.19. More precisely let $P_1, \cdots, P_m$ be similar polytopes in $\mathbb{R}^{m+1}$. To define the area measure $ds_m(P_1, \cdots, P_m, u)$ it is necessary to give values to

$$\int_C ds_m(P_1, \cdots, P_m, u),$$

for Borel sets $C \subset S^m$. For a given direction $u \in S^m$ let $F_u^{(j)}$ be the face (if exists) of $P_j$ whose outward unit normal is u. Since the polytopes $P_j$ are similar the faces $F_u^{(j)}$ either exist for all $j$ or for none. In the former case they are all parallel, and by a translation we regard $F_u^{(j)}$ as $m$ dimensional polytopes in $\mathbb{R}^m$. Then we set

$$\int_C ds_m(P_1, \cdots, P_m, u) = \sum_{u \in C} v_m(F_u^{(1)}, \cdots, F_u^{(m)}). \tag{1.5.55}$$

The sum on the right hand side is finite since for only finitely many directions $F_u^{(j)}$'s are non-empty. Starting with this definition of mixed area measure one can develop the theory as in the case mixed volumes by making use of the approximation lemma 1.5.19. We will not go through a formal verification of the fact that we obtain a Borel meassure in this fashion. For similar polytopes it is immediate that

$$v_{m+1}(P_1, \cdots, P_m, P) = \frac{1}{m+1} \int_{S^m} h_P(u) ds_m(P_1, \cdots, P_m, u), \tag{1.5.56}$$

where the integral reduces to a finite sum. Taking limit of $P_j \to K_j$ and $P \to K$ through similar polytopes we obtain the following result:

**Lemma 1.5.20** *With the above notation we have*

$$v_{m+1}(K_1, \cdots, K_m, K) = \frac{1}{m+1} \int_{S^m} h_K(u) ds_m(K_1, \cdots, K_m, u).$$

*for compact convex sets,* $K, K_1, \cdots, K_m \subset \mathbb{R}^{m+1}$.

As an application of the concept of mixed volumes and the Minkowski inequality (proposition 1.5.4) one establish the following rigidity result based on the area measures $ds_m(K,.)$:

**Corollary 1.5.5** *Let $K_1, K_2$ be compact convex sets with non-empty interior and assume $ds_m(K_1,.) = ds_m(K_2,.)$. Then $K_1$ and $K_2$ are translates of each other.*

**Proof** - It follows from the hypothesis that the area measures $ds_m(K_i, \cdots, K_i, .)$, $i = 1, 2$, are identical and consequently

$$v_{m+1}(K_1, K_2, \cdots, K_2) = v_{m+1}(K_1).$$

Applying Minkowski inequality (proposition 1.5.4) we obtain

$$v_{m+1}(K_1)^{m+1} = v_{m+1}(K_1, K_2, \cdots, K_2)^{m+1} \geq v_{m+1}(K_1)v_{m+1}(K_2)^m.$$

Therefore $v_{m+1}(K_1) \geq v_{m+1}(K_2)$. Similarly, $v_{m+1}(K_1) \geq v_{m+1}(K_2)$. It follows that

$$v_{m+1}(K_1, K_2, \cdots, K_2)^{m+1} \geq v_{m+1}(K_1)v_{m+1}(K_2)^m.$$

By proposition 1.5.4 $K_1$ and $K_2$ by a homothety and a translation and therefore differ by a translation since they have the same volume. ♣

A special case of corollary 1.5.5 is

**Corollary 1.5.6** *If two polytopes have the same set of outward unit normals (to facets) and corresponding facets have the same volume (area), then they differ by a translation.*

## 1.5.4 Existence Theorems

In lemma 1.5.12 we showed that for a polytope with facets $F_j$ and corresponding outward normals $e_j$, we have $\sum_j v_m(F_j)e_j = 0$. The question is whether given finite set of distinct vectors $\{e_1, \cdots, e_N\} \subset S^m$, which contains a basis for $\mathbb{R}^{m+1}$, and positive numbers $a_j$ such that

$$\sum_{j=1}^{N} a_j e_j = 0, \tag{1.5.57}$$

there is a polytope $P$ with non-empty interior, facets $F_j$ with outward unit normals $e_j$ and $v_m(F_j) = a_j$. The case of $m = 1$ is particularly simple:

**Lemma 1.5.21** *With the above notation, let $m = 1$. Then the necessary condition (1.5.57) is also sufficient.*

**Proof** - Let $e'_j$ be the unit vector orthogonal to $e_j$ such that $e_j, e'_j$ is a positively oriented basis for $\mathbb{R}^2$. Then (1.5.57) is equivalent to

$$\sum_{j=1}^{N} a_j e'_j = 0. \tag{1.5.58}$$

We may assume that the vectors $e_1, e_2, \cdots, e_N$ are ordered in the counterclockwise direction. Now consider the polygon with side $F_1$ parallel to and directed as the vector $e'_1$. From the end point of $F_1$ draw the face $F_2$ parallel to and directed as the vector $e'_2$. Continuing the process we see that $F_1, F_2, \cdots, F_N$ close up to form a polygon if (1.5.57) is fulfilled. ♣

As an application of the ideas of the preceding subsection we show that the answer is in the affirmative. The first observation is

**Lemma 1.5.22** *Let $\mathbb{R}^N_+$ be the subset of $\mathbb{R}^N$ consisting of vectors $\mathbb{Z} = (z_1, \cdots, z_N)$ with $z_j \geq 0$. Then the set*

$$P_{\mathbb{Z}} = \bigcap_{j=1}^{N} \mathsf{H}^-_{(e_j, z_j)}$$

*is a compact convex set, and has non-empty interior if all $z_j > 0$.*

**Proof** - Convexity of $P_{\mathbb{Z}}$ and non-emptiness of the interior of $P_{\mathbb{Z}}$ are clear. If $P_{\mathbb{Z}}$ were not compact, there would exist non-zero $f$ such that $\mathbb{R}_+ f \subset P_{\mathbb{Z}}$. The hypotheses $\sum a_i e_i = 0$ and $\sum \mathbb{R} e_j = \mathbb{R}^{m+1}$ imply $< f, e_j >> 0$ for some $j$ and therefore $af \notin P_{\mathbb{Z}}$ for $a > 0$ sufficiently large. ♣

The desired polytope can be obtained by a finite dimensional variational argument. Let

$$U = \{\mathbb{Z} \in \mathbb{R}^N_+ \mid v_{m+1}(P_{\mathbb{Z}}) \geq 1\}.$$

The boundary $\partial U$ is defined by the requirement $v_{m+1}(P_{\mathbb{Z}}) = 1$. Let us note that if $m > 1$ and the vectors $e_1, \cdots, e_N$ are in general position in the sense that every subset $e_{i_1}, \cdots, e_{i_{m+1}}$ of distinct vectors is a basis for $\mathbb{R}^{m+1}$, then $\partial U$ is $C^1$ manifold and tangent spaces to points of $\partial U$ are well-defined. In fact, it is clear that if $\mathbb{Z}_\circ = (z^\circ_1, \cdots, z^\circ_N)$ is such that each subset $P_{\mathbb{Z}_\circ} \cap \partial \mathsf{H}^-_{(e_j, z^\circ_j)}$ is

1. Either of positive measure, $v_m(P_{\mathbb{Z}_\circ} \cap \partial \mathsf{H}^-_{(e_j, z^\circ_j)}) > 0$,

2. Or is empty, $P_{\mathbb{Z}_\circ} \cap \partial \mathsf{H}^-_{(e_j, z^\circ_j)} = \emptyset$,

then a neighborhood of $\mathbb{Z}_\circ$ in $\partial U$ is the image of a diffeomorphism of an open subset of $\mathbb{R}^{N-1}$. Now assume $\mathbb{Z}_\circ = (z_1^\circ, \cdots, z_N^\circ)$ is such that $P_{\mathbb{Z}_\circ} \cap \partial \mathsf{H}^-_{(e_j, z_j^\circ)} \neq \emptyset$ but has measure 0. Then for $z_j^\circ - \epsilon < z_j < z_j^\circ$ and $\epsilon > 0$ small, $v_m(P_{\mathbb{Z}_\circ} \cap \partial \mathsf{H}^-_{(e_j, z_j^\circ)}) > 0$ and the general position assumption implies that $v_m(P_{\mathbb{Z}_\circ} \cap \partial \mathsf{H}^-_{(e_j, z_j^\circ)})$ goes to zero as $\epsilon^m$ as $\epsilon \to 0$. For $m > 1$ this will suffice to give $\partial U$ the structure of a $C^1$ manifold.

Consider the function

$$\psi : U \to \mathbb{R}, \quad \psi(\mathbb{Z}) = \frac{1}{m+1} \sum_{j=1}^{N} a_j z_j.$$

Since the coefficients $a_j > 0$, the function $\psi$ attains a minimum on $U$. Assume this minimum is attained at the point $\mathbf{b} = (b_1, \cdots, b_N)$ and $\psi(\mathbf{b}) = \mu^m$, i.e.,

$$\frac{1}{m+1} \sum_{j=1}^{N} a_j b_j = \mu^m. \tag{1.5.59}$$

Since the function $\psi$ attains a minimum at $\mathbb{Z} = \mathbf{b}$

$$v_{m+1}(P_{\mathbf{b}}) = 1. \tag{1.5.60}$$

We will show that the polytope $\mu P_{\mathbf{b}}$ is the solution to our problem.

Replacing $P_{\mathbf{b}}$ by a translate of it we may assume $\mathbf{0}$ is an interior point and consequently $b_j > 0$ for all $j$. Set

$$a_j^\star = v_m(F_{P_{\mathbf{b}}}(e_j)).$$

By lemma 1.5.12 $v_{m+1}(P_{\mathbf{b}}) = \frac{1}{m+1} \sum h_{P_{\mathbf{b}}}(e_j) a_j^\star$ and since $h_{P_{\mathbf{b}}}(e_j) = b_j$ if $a_j^\star \neq 0$, we have by (1.5.60)

$$\frac{1}{m+1} \sum_{j=1}^{N} a_j^\star b_j = 1. \tag{1.5.61}$$

Define the affine hyperplanes $\mathcal{L}_i$ as

$$\mathcal{L}_1 = \{\mathbb{Z} \mid \frac{1}{m+1} \sum_{j=1}^{N} a_j z_j = \mu^m\}, \quad \mathcal{L}_2 = \{\mathbb{Z} \mid \frac{1}{m+1} \sum_{j=1}^{N} a_j^\star z_j = 1\}$$

The key point in understanding the structure of $P_{\mathbf{b}}$ is

**Lemma 1.5.23** *Assume $e_1, \cdots, e_N$ are in general position. Then with the above notation, we have $\mathcal{L}_1 = \mathcal{L}_2$.*

First we show how lemma 1.5.23 implies the desired existence result:

**Proposition 1.5.5** *Let $e_j \in S^m$, $j = 1, \cdots, N$, be a set of distinct vectors containing a basis for $\mathbb{R}^{m+1}$, $a_j > 0$ real numbers such that (1.5.57) is satisfied. Then there is a unique, up to translation, polytope with outward unit normals $e_j$ and corresponding areas (volumes) of facets $a_j$.*

**Proof** - Uniqueness follows from corollary 1.5.5. In view of lemma 1.5.21 we may assume $m > 1$ Under the additional hypothesis that the vectors $e_1, \cdots, e_N$ are in general position, lemma 1.5.23 implies

$$a_i = \mu^m a_i^\star = v_m(F_{\mu P_{\mathbf{b}}}(e_i)),$$

that is, the volume (area) of the facet with normal $e_i$ is $a_i$ as desired, proving the proposition. The general case follows by approximating the set of vectors $\{e_1, \cdots, e_N\}$ by one in general position. To be more precise, let $\{e_1(r), \cdots, e_N(r)\}$, $0 \leq r < 1$ be a smooth one parameter family of unit vectors which for $r > 0$ are in general position and $e_j(0) = e_j$. Let $a_j(r) > 0$ be smooth functions with $a_j(0) = a_j$ and

$$\sum_{j=1}^{N} a_j(r) e_j(r) = 0.$$

Let $P_r$, $r > 0$, be the corresponding polytope. The volumes $v_m(\partial P_r)$ are uniformly bounded and by the isoperimetric inequality $v_{m+1}(P_r)$ is also uniformly bounded. Since the quantities $a_j(r)$, $r \in [0, 1]$ are bounded away from zero it follows from second formula of lemma 1.5.12 that the quantities $h_{P_r}(e_j(r))$ are uniformly bounded. Therefore the polytopes $P_r$ remain in a bounded subset of $\mathbb{R}^{m+1}$. By the Blaschke Selection lemma we can choose a convergent sequence of polytopes $P_{r_j}$. It is clear that the limiting convex set is still a polytope with outward unit normals $e_j$ and the area (volume) of the corresponding facet equal to $a_j$. ♣

**Lemma 1.5.24** *With the above notation for $\mathbb{Z}$ in a neighborhood $V$ of $\mathbf{b}$ in $\mathbb{R}^N$ we have*

$$v_{m+1}(P_{\mathbf{b}}, \cdots, P_{\mathbf{b}}, P_{\mathbb{Z}}) = \frac{1}{m+1} \sum_{j=1}^{N} h_{P_{\mathbb{Z}}}(e_j) a_j^\star.$$

**Proof** - For $V$ sufficiently small the polytopes $P_{\mathbf{b}}$ and $P_{\mathbb{Z}}$ are similar if either all $a_j^\star > 0$ or if $a_j^\star = 0$ then $P_{\mathbf{b}} \cap \partial\mathsf{H}^-_{(e_j,b_j)} = \emptyset$ in which case the validity of the assertion follows from remark 1.5.3. If $a_j^\star = 0$ but $P_{\mathbf{b}} \cap \partial\mathsf{H}^-_{(e_j,b_j)} \neq \emptyset$, then for $z_j > b_j$ the polytopes $P_{\mathbf{b}}$ and $P_{\mathbb{Z}}$ are similar, but for $z_j < b_j$ they are no longer similar since the set of normals to $P_{\mathbb{Z}}$ and $P_{\mathbf{b}}$ will be different. However for $\epsilon > b_j - z_j > 0$ we can approximate $P_{\mathbf{b}}$ arbitrarily closely with polytopes $P_{\mathbf{b}'}$ similar to $P_{\mathbb{Z}}$ by taking $b_j > b'_j > z_j$. Then the assertion remains valid for $P_{\mathbf{b}'}$ replacing $P_{\mathbf{b}}$ and by taking $b'_j \to b_j$ the required result follows. ♣

**Proof of lemma 1.5.23** - Let $\mathbb{Z}$ be in the neighborhood $V$ of $\mathbf{b}$ given in lemma 1.5.24. Substituting $h_{P_{\mathbb{Z}}}(e_j) = z_j$ in formula in lemma 1.5.24 we obtain

$$v_{m+1}(P_{\mathbf{b}}, \cdots, P_{\mathbf{b}}, P_{\mathbb{Z}}) = \frac{1}{m+1}\sum_{j=1}^{N} a_j^\star z_j.$$

By proposition 1.5.4

$$v_{m+1}(P_{\mathbb{Z}}) \leq v_{m+1}(P_{\mathbf{b}}, \cdots, P_{\mathbf{b}}, P_{\mathbb{Z}})^{m+1},$$

which implies that the only point of intersection of $U$ and $\mathcal{L}_2$ is the point $\mathbf{b}$. The affine space $\mathcal{L}_1 \cap V$ passes through $\mathbf{b}$ but does not contain any point from the interior of $U$ since that would contradict minimality of $\psi(\mathbf{b}) = \mu^m$. By the general position assumption of the vectors $e_1, \cdots, e_N$, the boundary $\partial U$ is a manifold and consequently the affine subspaces are the tangent spaces to $\partial U$ at $\mathbf{b}$ and are identical. ♣

It is possible to use proposition 1.5.5 and an approximation argument to prove an existence result for the Minkowski problem discussed in the preceding subsection, yet such an approach is not satisfactory since it does not appear that one can prove smoothness of the resulting manifold in this manner. A satisfactory approach is based on the study of the Monge-Ampère equation which involves analytical techniques which are postponed to another volume (see [CY]). However, we can prove the following existence result for area measures via approximation:

**Proposition 1.5.6** *Let $f \geq 0$ be a non-negative continuous function*[20] *on $S^m$. If*

$$\int_{S^m} ef(e)dv = 0,$$

[20]The same result is valid for a general finite Borel measure $d\mu$ provided it is not concentrated on a great circle. In fact proposition 1.5.5 is the case of a finitely supported measure and the necessary conditions of not being supported on a great circle and $\int e d\mu(e) = 0$ imply that the support set of the measure contains a basis for $\mathbb{R}^{m+1}$. It is for the purpose of avoiding some minor measure theoretic technicalities that we are assuming that the measure is given by a continuous density $f$.

*where $dv$ is the standard volume element on $S^m$, then there is a compact convex set $K$ with $ds_m(K,.) = fdv$.*

**Proof** - We decompose $\mathbb{R}^{m+1}$ into finitely many convex cones $C_1, \cdots, C_N$ bounded by hyperplanes and with common vertex $\mathbf{0}$. We assume that the cones have pairwise disjoint interiors and each set $D_j = C_j \cap S^m$ of small volume (area) as may be necessary. Let $D_1, \cdots, D_l$ be those among $D_j$'s with positive measure relative to $d\mu = fdv$, and set for $j = 1, \cdots, l$,

$$\varepsilon_j = \frac{1}{\mu(D_j)} \int_{D_j} ef(e)dv,$$

where $\mu(D_j)$ is the measure of $D_j$. Each $\varepsilon_j$ is a vector of the form

$$\varepsilon_j = \gamma_j e_j,$$

where $e_j \in S^m$ and $\gamma_j > 0$ by taking the decomposition of $\mathbb{R}^{m+1}$ to be sufficiently fine. Now let $a_j = \gamma_j \mu_j$, then the set of vectors $\{e_1, \cdots, e_j\}$ and positive numbers $a_j$ satisfy the hypothesis of proposition 1.5.5 and therefore we get a polytope $P$ which depends on the decomposition of $\mathbb{R}^{m+1}$ into convex cones with $\mathbf{0}$ as their common vertex. It is clear from the construction that the quantities $\sum a_j$ remain uniformly bounded (in fact, bounded by $\int fdv$) and by isoperimetric inequality the volumes of the polytopes $P$, as we refine the decomposition of $\mathbb{R}^{m+1}$, also remain bounded. Let $b \geq v_{m+1}(P)$ for all such $P$, and $y \in P$. Set $y = \eta e$ with $e \in S^m$ and $\eta > 0$ and $C_y$ be the convex closure of $\mathbf{0}$ and $y$. Then

$$h_P(u) \geq h_{C_y}(u) = \eta < u, e >_+,$$

where $z_+$ means maximum of 0 and $z$. Therefore

$$b \geq v_{m+1}(P) \geq \frac{\eta}{m+1} \int_{S^m} < u, e >_+ f(e)dv \geq c\eta$$

for some positive constant $c$ bounded away from 0. Therefore $\eta$ is bounded and the polytopes $P$ remain in a bounded subset of $\mathbb{R}^{m+1}$. Blaschke's Selection lemma is now applicable to give the desired compact convex set. ♣

## 1.6 Minimal Surface

### 1.6.1 Weierstrass Representation

THIS SECTION IS NOT INCLUDED

# Bibliography


[A] Arnold, V. I. - *Singularities of Caustics and Wave Fronts*, Kluwer Academic Press, Norwell, Ma., 1990.

[C1] Cartan, E. - Le Principe de Dualité et Certaines Integrales Multiples de l'Espace Reglé; *Bull. Soc. Math. France*, vol. **24**, (1896), pp.140-177; *Ouvres Complètes*, Partie II, Centre National de la Recherche Scientifique, Paris, 1984, pp.265-302.

[C2] Cartan, E. - *Leçons sur la Géométrie des Espaces de Riemann*, Gauthier-Villars, Paris 1963.

[Ch] Chern, S. S. - An elementary proof of the existence of isothermal parameters on a surface; *Proc. Am. Math. Soc.*, **6**, (1955), pp. 771-782; *Selected Papers*, volume 2, Springer-Verlag, Berlin-Heidelberg-New York, 1978.

[G] Gray, A. - *Tubes*, Addison-Wesley, Reading, Ma., 1990.

[GB] Große-Braukmann, K. - Gyroids of Constant Mean Curvature; *Experimental Mathematics*, vol. **6**, no.1, (1997), pp. 33-50.

[H] Helgason, S. - *Differential Geometry, Lie Groups and Symmetric Spaces*, Academic Press, New York, 1978.

[KN] Kobayashi, S. and K. Nomizu - *Foundations of Differential Geometry*, Two Volumes, Wiley (Interscience), New York, 1963, 1969.

[N] Nitsche, J. C. C. - *Lectures on Minimal Surfaces*, Volume 1, Cambridge University Press, New York, 1989. (English translation of the first five chapter of the original German text.)

[S] Santaló, L. A. - *Integral Geometry and Geometric Probability*, Addison Wesley Publishing Co., Reading, Ma., 1976.

[SW] Stein, E. and G. Weiss - *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton University Press, 1975.

[W] Weyl, H. - *Classical Groups*, Princeton University Press 1966.

[Wh] Whitehead, J. H. C. - Convex Regions in the Geometry of Paths; *Quarterly Journal of Mathematics* (Oxford), vol. **3**, (1932), pp.33-42.