Geometric Applications of Sturm Comparison Theorem

S. Kumaresan
School of Math. and Stat.
University of Hyderabad
Hyderabad 500046
kumaresa@gmail.com

Abstract

After recalling the Sturm's comparison theorem, we illustrate some of its typical applications in the theory of surfaces. Even though some acquaintance with Gauss theory of surfaces will be advantageous, it is not absolutely necessary, as I have striven to illustrate the uses in the case of a sphere in 3-space.

Introduction

Consider the spheres $S^2(r)$ and $S^2(R)$ of radii $0 < r < R$ in the euclidean 3-space E^3. Take two "similarly situated" curves, say, the equators c_r and c_R on the spheres. We know that $l(c_r) < l(c_R)$, where $l(c)$ denotes the length of a curve c in E^3. What we are going to do is to give a "complicated" proof of this obvious fact (and other such) using Sturm's theorem. If you have a smattering of the theory of surfaces, you can easily adapt the proof which we present for the spheres to the general case of comparing the lengths of 2 "similarly situated" curves in 2 different surfaces. However let me assure you that if you follow diligently, there is no need for a previous exposure to Differential Geometry of surfaces. In fact, I have tried my best to write this in a such a way that this exposition will enhance your understanding if you start learning differential geometry of surfaces.

I hope you have an intellectually stimulating reading.

1 Sturm's Theorem

We recall Sturm's theorem and its proof briefly.

Theorem 1 (Sturm)**.** *Let y_i be nonnegative solutions of $y_i'' + k_i y_i = 0$ on an interval $[0, l]$. Assume that both of them satisfy either Eq. 1 or Eq. 2:*

$$y_1(0) = y_2(0) = a > 0 \qquad \& \qquad y_1'(0) = y_2'(0) \tag{1}$$

$$y_1(0) = y_2(0) = 0 \qquad \& \qquad y_1'(0) = y_2'(0) = b > 0 \tag{2}$$

We also assume that $k_1(s) \leq k_2(s)$. If s_i is the first zero of y_i to the right of 0, then i) $s_2 \leq s_1$ and ii) $y_1 \geq 0$ and $y_2(s) \leq y_1(s)$ for $0 < s \leq s_2$. Furthermore, equality holds at $s_0 \in (0, l)$ iff $k_1(s) = k_2(s)$ for $0 < s < s_0$.

Proof. We shall only give a sketch of a proof. Multiply the first equation by y_2 and the second by y_1 and then subtract the latter from the former. We add and subtract the term $y_1'y_2'$ to this. The resulting equation is

$$y_1''y_2 - y_1'y_2' - (y_1y_2'' - y_1'y_2') + (k_2(s) - k_1(s))y_1y_2 = 0.$$

Integration and the initial conditions lead us to

$$(y_1'y_2 - y_1y_2')(s) = \int_0^s (k_1 - k_2)y_1y_2 ds.$$

We note the identity $(d/ds)(y_1/y_2) = (y_1'y_2 - y_1y_2')/y_2^2$ and proceed to complete the proof. □

2 Surfaces-A Rapid Introduction

Let $\mathbb{R}^n$ be the n-dimensional vector space over $\mathbb{R}$. We have a *natural basis* $\{e_i : 1 \leq i \leq n\}$ of $\mathbb{R}^n$ over $\mathbb{R}$, where $e_i := (0, \ldots, 0, 1, 0, \ldots, 0)$, 1 at the i-th place. We identify any $x \in \mathbb{R}^n$ with the n-tuple $(x_1, \ldots, x_n)$ where $x := \sum x_ie_i$. We then have the euclidean inner product $\langle x, y\rangle := \sum_{i=1}^n x_iy_i$, where $x = \sum x_ie_i$ and $y = \sum y_ie_i$. We let E^n stand for $\mathbb{R}^n$ with the euclidean inner product. E^n is called the n-dimensional euclidean space. We have the norm on E^n induced by the inner product $\|x\| := \sqrt{\langle x, x\rangle}$ for $x \in E^n$. We often denote E^n by E.

A curve in U, an open subset of E, is a "smooth" map $c : (-\varepsilon, \varepsilon) \to U$ for some positive ε. By smooth we mean that it has sufficiently many continuous derivatives on $(-\varepsilon, \varepsilon)$. This is same as saying that the component functions c_i have sufficiently many continuous derivatives. Here, of course, $c(t) := (c_1(t), \ldots, c_n(t))$. We say that it passes through the point $p := c(0)$. The *tangent* to this curve at $c(t)$ is by definition the vector

$$c'(t) := \lim_{h\to 0} \frac{c(t + h) - c(t)}{h} = (c_1'(t), \ldots, c_n'(t)).$$

$c'(t)$ is also called the *velocity* vector of c at $c(t)$. The length $l(c)$ of the curve c is defined by setting $l(c) := \int_{-\varepsilon}^{\varepsilon} \|c'(t)\| dt$.

Ex. 2. Let $h : (a, b) \to (-\varepsilon, \varepsilon)$ be a smooth map such that $h'(s) \neq 0$ for any $a < s < b$. Then $c \circ h : (a, b) \to U$ is a curve tracing the same image as c. h is called a reparameterization of c. Show that $l(c \circ h) = l(c)$.

We can put this in a picturesque language: The train travels the same distance whether it is slow or fast!

We restrict ourselves to $n = 3$.

There are geometric objects other than open subsets in E^n on which we can talk of smooth curves. For example, $E^2 := \sum_{i=1}^2 \mathbb{R}e_i$. A more suggestive example is any sphere of radius $r > 0$:

$$S(r) \equiv S^2(r) := \{x \in E^3 : \langle x, x\rangle = r^2\}.$$

It is clear that any point on $S(r)$ has a neighbourhood which is homeomorphic to a disk in E^2. Roughly speaking, a *surface* in E^3 is a subset S of E^3 with this property. More precisely,

Definition 3. A surface in E^3 is a set $S \subset E^3$ such that every point $p \in S$ has an open neighbourhood V in E^3 with the following property:
There exists an open set U in E^2 and a smooth map $\varphi : U \to V$ such that i) $\varphi : U \to V \cap S$ is a homeomorphism and ii) φ is regular, i.e., the Frechet derivative or what is the same the Jacobian $d\varphi(u,v) := \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} & \frac{\partial z}{\partial u} \\ \frac{\partial x}{\partial y} & \frac{\partial y}{\partial v} & \frac{\partial z}{\partial v} \end{pmatrix} (u,v)\colon E^2 \to E^3$ is of φ is of rank 2 at every point $(u,v) \in U$. Here we have written $\varphi(u,v) = (x(u,v), y(u,v), z(u,v))$. (φ, U) is called a *parameterization* of $V \cap S$.

Example 4. $S(r)$ has a parameterization $(u,v) \mapsto (r\cos u \cos v, r \cos u \sin v, r \sin u)$ for $0 < v < 2\pi$ and $0 < u < \pi$.

Example 5. A surface of revolution. Let $c : (a,b) \ni u \mapsto (x(u), 0, z(u))$ be a curve in the xz-plane. We revolve it around the z-axis to get a surface of revolution with the parameterization: $\varphi(u,v) := (x(u)\cos v, x(u) \sin v, z(u))$ for $(u,v) \in (a,b) \times (0, 2\pi)$. (Notice that Example eg:1 is a special case.)

Example 6. Let $f : U \to \mathbb{R}$ be a smooth function on an open set U of E^2. Then $S := \{(x,y,z) : z = f(x,y), x, y, z \in \mathbb{R}\}$ is a surface. The parameterization is given by $(x,y) \mapsto (x, y, f(x,y))$.

Example 7. Let $f : \Omega \subset E^3 \to \mathbb{R}$ be a smooth function. We say that $a \in f(\Omega)$ is a *regular value* if for any $p \in \Omega$ with $f(p) = a$, the gradient of f at p is nonzero: $(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z})(p) \neq 0$. If a is a regular value of f then $S := \{p \in \Omega : f(p) = a\}$ is a surface. I shall leave this as an instructive exercise in the use of inverse function theorem.

We define a (smooth) curve in a surface S as follows: $c : (-\varepsilon, \varepsilon) \to S$ is a curve if for every $s \in (-\varepsilon, \varepsilon)$, we have a neighbourhood V of $c(s)$ in E^3 and a $\delta > 0$ such that the restriction $c : (s - \delta, s + \delta) \to V$ is smooth. We can therefore speak of tangents and the lengths of curves as earlier.

Example 8. In $S(r)$, consider the equator c in the xy-plane: $c : s \mapsto r\cos s e_1 + r \sin s e_2 \equiv (r\cos s, r \sin s, 0)$ for $0 \leq s \leq 2\pi$. Then $c'(s) = -r\sin s e_1 + r\cos s e_2 \equiv (-r\sin s, r\cos s, 0)$ so that $\| c'(s) \| = r$ for all s and hence $l(c) = 2\pi r$, as is to be expected.

Ex. 9. Let $\varphi : U \to V \cap S$ be a parameterization around $p \in S$. We let $X_1 := (\frac{\partial x}{\partial u}, \frac{\partial y}{\partial u}, \frac{\partial z}{\partial u})$ and $X_2 := (\frac{\partial x}{\partial v}, \frac{\partial y}{\partial v}, \frac{\partial z}{\partial v})$. X_i are tangent vectors at $\varphi(u,v)$ to S and they are linearly independent for all $(u,v) \in U$. (Can you think of some natural curves which have these as their tangent vectors?)

Ex. 10. We let $T_pS := \mathbb{R}X_1 \oplus \mathbb{R}X_2$ for all $p = \varphi(u,v)$ with $(u,v) \in U$. Then T_pS is a 2-dimensional subspace of E^3 independent of the parameterization. It is called the *tangent space* of S at p.

Ex. 11. The vector or the cross product $\mathbf{n}_p := X_1 \times X_2(u,v)$ is nonzero. It is called the *normal field* to the surface at p, since it is perpendicular to T_pS.

Ex. 12. Verify that for $S = S(r)$ the normal at $x \in S$ is given by $\mathbf{n}_x = x/r$. In the case of $S = f^{-1}(a)$ (Eg. 7), we have $\mathbf{n}_p = \nabla f(p)$, the gradient of f at p.

Geodesics

Definition 13. A *geodesic* on a surface S is a curve $c : (-\varepsilon, \varepsilon) \to S$ with zero acceleration as observed from the surface. By this we mean that the *acceleration* $c''(s) \perp T_{c(s)}S$ for all s. Thus, the tangential component $(c'')^\top = 0$.

A more intuitive definition runs as follows: The length of the segment between any two of its sufficiently nearby points is less than or equal to the length of any curve joining these points. That is, given any two points $c(t), c(t+\delta)$ sufficiently nearby, the length of the segment of the curve between these points, viz., $\int_t^{t+\delta} \| c'(s) \| \, ds$ is less than or equal to $l(\sigma)$, the length of any curve σ that joins these points. This definition is equivalent to the above one, but we shall not prove it in our lectures.

The analytical definition $(c'')^\top = 0$ can be translated into a system of second order ordinary differential equations via parameterization. Hence by the existence and uniqueness theorem in the theory of ordinary differential equations, it follows that there exists a unique geodesic with given initial data. That is, if $p \in S$ and $v \in T_pS$ are given there exists a "unique" geodesic $c_{p,v} : (-\varepsilon, \varepsilon) \to S$ such that $c(0) = p$ and $c'(0) = v$. (ε may depend upon v.)

Example 14. The great circles, i.e., the intersection of the planes through the origin and $S(r)$, are the geodesics on a sphere $S(r)$. We can describe these without messy notation. Let $x \in S$ and $v(\neq 0) \in T_pS$. Note that this means $\langle x, v \rangle = 0$. (See Exer. 9.) We then want the description of a circle centered at the origin of radius r in the plane spanned by x and v:

$$c(s) \equiv c_{x,v} := \cos sx + \frac{r}{\| v \|} \sin sv.$$

One easily sees that $c''(s) = -c(s) \perp T_pS$ and that $c_{x,av}(s) = c_{x,v}(as)$ for any $a \in \mathbb{R}$.

Example 15. The curves $c(s) := (r \cos s, r \sin s, rs)$ are geodesics on a cylinder $S := \{(x, y, z) : x^2 + y^2 = r^2\}$.

A more natural way of parameterizing a geodesic through the arc length, i.e., a parameterization with $\| c' \| = 1$. If $c : [0, T] \to S$ is a geodesic, we first of all note that, for any t, $c'(t) \neq 0$ since otherwise it has to be the constant curve due to uniqueness. (Do you understand this reasoning completely?) If we introduce $s(t) := \int_0^t \| c'(t) \| \, dt$, s is then a strictly increasing function and hence we can use it to reparametrize c. In the new s-parameterization, we have $l(c|_{[0,a]}) = a$ for any $0 \leq a \leq T$.

The geodesics of the sphere with respect to the arc length are given by $c(s) = \cos \frac{s}{r} x + r \sin \frac{s}{r} v$ for $v \in T_pS$ with $\| v \| = 1$. (This is enough; see the last line of Eg. 14.) The reader is urged to check the details of this.

Gaussian Curvature

The most important concept in the theory of surfaces is Gaussian curvature. If S is a surface in E^3 and (φ, U) is a parameterization of a piece of S, the vector field $\mathbf{n}_p := X_1 \times X_2$ is a nonzero normal vector field on $\varphi(U) \subset S$. Hence we have a normal field $\mathbf{n}$ of unit norm on $\varphi(U)$, unique up to sign. Thus we have a map $\mathbf{n} : \varphi(U) \to S^2(1)$, called the *Gauss map.*

Example 16. 1) For $S = S(r)$, the Gauss map is $x \mapsto x/r$.
2) For $S = E^2$, the Gauss map is the constant $x \mapsto e_3$.
3) For the cylinder in Eg. 15, the Gauss map is $(x, y, z) \mapsto (1/r)(x, y, 0)$.

Definition 17. As x varies over a small *oriented* area around p in S, the Gauss map $\mathbf{n}_x$ sweeps out an *oriented* area in $S^2(1)$. The rate of change of this *oriented* area is called the *Gaussian curvature* of S at p.

A more precise definition is analytical. We define $K(p)$, the Gaussian curvature of S at p to be the determinant of the Jacobian $d\mathbf{n}_p$ of the Gauss map. This is what we should expect if we remember the change of variable formula from calculus of several variables. It is important to realize that $K(p)$ does not depend on the choice of the sign of $\mathbf{n}$.

Example 18. For the sphere $S(r)$ the Gauss map can be thought of the restriction of the linear map $x \mapsto x/r$ from E^2 to itself and hence the curvature is given by $K(p) = 1/r^2$. Thus a sphere of smaller radius is more 'curved' than a sphere of larger radius. This is certainly clear if you look at the way the Gauss map sweeps out areas.

Example 19. The plane has curvature 0. This is clear either from the geometric definition or from the analytical one.

Example 20. The cylinder springs a surprise on almost everybody. Contrary to general expectations, it is not 'curved' at all, i.e., its Gaussian curvature is zero. For as x varies over a region of the surface, $\mathbf{n}_x$ varies over a part of the equator only, whose *area* is 0!

Complete Surfaces & Hopf-Rinow Theorem

In general, for a given surface S and $p \in S$ and $v \in T_pS$, the geodesic may not be defined on all of $\mathbb{R}$. See, for example, the surface $E^2 \setminus \{(0, 0, 0)\}$ in E^3. We usually deal with surfaces on which all geodesics are defined on the entire line $\mathbb{R}$. Such surfaces are said to be *geodesically complete*. The spheres $S(r)$, the plane $E^2 \subset E^3$, the cylinder $S := \{(x, y, z) : x^2 + y^2 = r^2\}$ are geodesically complete.

On any surface M, we have a pseudo-metric: $d(p, q) := \inf l(\sigma)$, where the infimum is taken over all (piece-wise) smooth curves σ joining p and q. One can show easily that d is, in fact, a metric and that the metric topology on M coincides with the subspace topology.

We now quote the basic result about the geodesically complete surfaces without a proof:

Theorem 21 (Hopf-Rinow)**.** *hopf Let M be a surface in E^3. The following are equivalent:*
M is geodesically complete.
(M, d) is complete.
Any closed and d-bounded subset of M is compact.

Any of the above statements implies the following: Given any $p, q \in M$, there exists a geodesic $c : [0, 1] \to M$ such that $c(0) = p$ and $c(1) = q$ and $l(c) = d(p, q)$. □

If M is geodesically complete, we can define the *exponential map.* If $v \in T_pM$, then we set $\exp_p v := c_{p,v}(1)$ since 1 is in the domain of any geodesic.

Example 22. Let $x \in S := S^2(1)$. Let $u \in T_xS$ be of norm 1, i.e., $\langle u, x\rangle = 0$ and $\|u\| = 1$. Take $v := \pi/2v$. Then

$$\exp_x v := c_{x,v}(1) = c_{x,u}(\pi/2) = \cos\pi/2x + \sin\pi/2u = u.$$

3 Rauch Comparison for Spheres

It is clear that on any geodesic c in $S(r)$ if we start from $c(0)$, after πr, the geodesic ceases to be minimizing. That is, if we consider any point $c(s)$ with $s > \pi r$, then there is a shorter way of reaching it, viz., along the opposite direction. Jacobi attempted to find conditions on the parameter value s_0 from which onwards the geodesic ceases to be minimizing. He approached the problem through the calculus of variations. We shall adopt somewhat simpler and more geometric approach.

Let c be a given geodesic. Following Jacobi we start with a 1-parameter variation c_t $(-\varepsilon < t < \varepsilon)$ of geodesics such that $c_0 = c$ and $c_t(0) = c(0)$. We wish to write such a family in the case the geodesic $c(s) := r(\cos(s/r)e_1 + \sin(s/r)e_2)$ in $S(r)$. It is easy if you draw some picture or if you realize that you have to vary the tangent e_2 of c: If we write $H(s,t) := c_t(s)$, then

$$H(s,t) := r(\cos(s/r)e_1 + \sin(s/r)(\cos t\, e_2 + \sin t\, e_3)).$$

The *variational field* of a geodesic variation H is by definition $\frac{\partial H}{\partial t}|_{t=0}$. Hence in our case it is

$$J(s) := \frac{\partial H}{\partial t}|_{t=0} = r\sin(s/r)(-\sin t\, e_2 + \cos t\, e_3)|_{t=0} = r\sin(s/r)e_3.$$

J is called the *Jacobi field* of the geodesic variation.

Remark: All of the above can easily be done in a more general set up. The key to this is the following observations:
i) e_3 is the tangent to the curve $t \mapsto \sigma(t) := \cos t\, e_2 + \sin t\, e_3$ *through* $e_2 \equiv c'(0)$ *in* $T_{e_1}S$ ii) $J(0) = 0$ and $J'(0) = e_3$ and iii) $\exp_{e_1}(s\sigma(t)) = H(s,t)$.

Now we are ready to exhibit the link that connects geometry and analysis. It is the fact that J satisfies the following second order equation: $J'' + (1/r^2)J = 0$ with $J(0) = 0$ and $J'(0) = e_3$. This can be reduced to an equivalent scalar equation by writing $J(s) = j(s)e_3$ where $j(s) := r\sin(s/r)$. Hence we get $j'' + (1/r^2)j = 0$ with the initial conditions $j(0) = 0$ and $j'(0) = 1$. The important thing to note here is that the coefficient is the Gaussian curvature.

In the general case also we get an exact analogue, if we start with a 1-parameter family of geodesic variation of a given geodesic c. That is, the jacobi field, viz., the variational field will satisfy the *Jacobi equation* $J''(s) + K(s)J(s) = 0$ along c where $K(s)$ is the Gaussian curvature function along c. Again this can be reduced to a scalar equation.

We can now state the result of Jacobi without proof :

Theorem 23 (Jacobi)**.** *If H is a 1-parameter variation of geodesics and $J(s) := \frac{\partial H}{\partial t}|_{t=0}$ is the Jacobi field along $c := H(s,0)$ such that $c_t(0) := H(0,t) = c(0)$ for all t and such that there exists s_0 with $J(s_0) = 0$ for some $s_0 > 0$, then c is not minimizing after s_0.*

Using this we give our first application of Sturm's theorem to the geometry of surfaces.

Theorem 24 (Bonnet)**.** *Let M be a complete surface (in $\mathbb{R}^3$). Assume that the curvature function $K(p) \geq r$ for some $r > 0$ for all $p \in M$. Then M has diameter $\leq \pi/\sqrt{r}$. Hence M is compact.*

Proof. The proof uses the Hopf-Rinow theorem. Let p and q be points of M. Since M is complete, there exists a geodesic c joining p and q such that $l(c) = d(p,q)$. We may assume that it is parameterized by arc length. By Sturm's theorem it follows that any Jacobi field along c has a zero at or before $\pi/\sqrt{r}$. Hence c is not minimizing if $d(p,q) > \pi/\sqrt{r}$. Hence $d(p,q) \leq \pi/\sqrt{r}$. □

The reader should note that we used only a weaker version of Sturm's theorem. We now discuss another consequence of Sturm's theorem which compares the solutions themselves. Roughly speaking, the result we wish to illustrate says that the length of a curve on a surface of lesser curvature is greater than or equal to that of a "similar" curve on a surface of larger curvature. This result is known as Rauch's theorem.

To illustrate Rauch's theorem in our case, let me first show how to relate the length of a curve with Jacobi fields.

Let $c(s) := r\cos s\, e_1 + \sin s\, e_2$ be the equator in $S(r)$. To compute the length of c, say, from 0 to π, we need to evaluate $\int_0^\pi \|c'(s)\|\, ds$. In particular, we need to compute $\|c'(s)\|$. As we have a concrete parameterization we could forge ahead straight-away. In stead, we indicate the strategy of Rauch which involves jacobi fields and Sturm's theorem.

From $p = e_1$, we have the unique minimizing geodesic $\gamma_s : t \mapsto H(s,t) := \gamma_s(t) := \cos t\, e_1 + \sin t(\cos s\, e_2 + \sin s\, e_3)$ to the point $c(s)$. $\gamma_s(\pi/2) = c(s)$. Now if we fix an s_0, then $s \mapsto \gamma_s$ is a geodesic variation of γ_{s_0} and hence we have a corresponding Jacobi field J_{s_0} along γ_{s_0} given by

$$J_{s_0}(t) := \frac{\partial H}{\partial s}|_{s=s_0}(t) = \sin t(-\sin s_0\, e_2 + \cos s_0\, e_3).$$

Now the important observation is that the value of $J_{s_0}(t)$ at $t = \pi/2$ is $c'(s_0)$!

Thus to compare the lengths of the equators c_r and c_R we proceed as above to get jacobi fields J_s^r and J_s^R with the property that $J_s^r(\pi/2) = c_r'(s)$ and $J_s^R(\pi/2) = c_R'(s)$. Now J_s^r is a solution of the Jacobi equation $J_s^{r\prime\prime} + r^{-2}J_s^r = 0$. Analogous result holds for r replaced by R. It is easy to see that we can invoke Sturm to conclude that $J_s^R(t) \geq J_s^r(t)$ for all $t \in [0,\pi/2]$ and for any fixed s. In particular, at $t = \pi/2$, we have

$$\|c_r'(s)\| = J_s^r(\pi/2) \leq J_s^R(\pi/2) = \|c_R'(s)\|.$$

We thus have a point-wise estimation of the tangent vectors of the curves under question and hence $l(c_r) \leq l(c_R)$.

How far this proof goes over to the general situations? Almost verbatim, if one knows the correct jargon! Let M_i be surfaces and $p_i \in M_i$. We fix a linear isometry of $T_{p_1}M_1$ onto

$T_{p_2}M_2$. Let $\varepsilon > 0$ be such that the exponential map $\exp_{p_i}$ is a diffeomorphism on the ball $B_{p_i}(0, \varepsilon) \subset T_{p_i}M_i$. Let $\sigma_i : [0, 1] \to B_{p_i}(0, \varepsilon)$ be a curve. Let $c_i(s) := \exp_{p_i}(\sigma_i(s))$. Then we can compare the lengths of the tangent vectors of the curves $c_i(s)$ (and hence the lengths of the curves themselves) in a way entirely analogous to the above.

We may also use the above to deduce an area comparison theorem (due to Bishop) and an angle comparison theorem (due to Alexandrov).

For a nice introduction to Differential Geometry of Surfaces, we refer the readers to do Carmo, M., *Differential geometry of curves and surfaces*, Prentice-Hall, New Jersey, 1976.