3 Applications

3a Constrained optimization

> One of the most brilliant and well-known achievements of differential calculus is the collection of recipes it provides for finding the extrema of functions. ... Frequently a situation that is more complicated and from the practical point of view even more interesting arises, in which one seeks an extremum of a function under certain constraints ...[1]

Let $Z \subset \mathbb{R}^n$ be a set, $f : Z \to \mathbb{R}$ a function, and $x_0 \in Z$ a point. We say that x_0 is a local maximum point of f on Z, if $f(x) \le f(x_0)$ for all $x \in Z$ near x_0. (A local minimum point is defined similarly.)

In particular, if $Z = g^{-1}\big(\{0\}\big) = \{x : g(x) = 0\}$ for a given $g : \mathbb{R}^n \to \mathbb{R}^m$, a local maximum point of f on Z is called a local maximum point of f subject to the constraint $g(\cdot) = 0$. That is, subject to $g_1(\cdot) = \dots = g_m(\cdot) = 0$ where $\big(g_1(x), \dots, g_m(x)\big) = g(x)$. “Extremum” means either maximum or minimum, of course.

3a1 Theorem. Assume that $x_0 \in \mathbb{R}^n$, $1 \le m \le n-1$, functions $f, g_1, \dots, g_m : \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable near x_0, $g_1(x_0) = \dots = g_m(x_0) = 0$, and the vectors $\nabla g_1(x_0), \dots, \nabla g_m(x_0)$ are linearly independent. If x_0 is a local constrained extremum point of f subject to $g_1(\cdot) = \dots = g_m(\cdot) = 0$, then there exist $\lambda_1, \dots, \lambda_m \in \mathbb{R}$ such that

$$\nabla f(x_0) = \lambda_1 \nabla g_1(x_0) + \dots + \lambda_m \nabla g_m(x_0) \,.$$

[1]Quoted from: Zorich, Sect. 8.7.3a, p. 527.

The numbers $\lambda_1, \dots, \lambda_m$ are called *Lagrange multipliers.*

A physicist could say: in equilibrium, the driving force is neutralized by constraints reaction forces.

In practice, seeking local constrained extrema of f on $Z = g^{-1}\big(\{0\}\big)$ one solves (that is, finds *all* solutions of) a system of $m + n$ equations

$$\begin{aligned} & g_1(x) = \dots = g_m(x) = 0\,, && (m \text{ equations}) \\ & \nabla f(x) = \lambda_1 \nabla g_1(x) + \dots + \lambda_m \nabla g_m(x) && (n \text{ equations}) \end{aligned}$$

for $m + n$ variables

$$\begin{aligned} & \lambda_1, \dots, \lambda_m\,, && (m \text{ variables}) \\ & x\,. && (n \text{ variables}) \end{aligned}$$

For each solution $(\lambda_1, \dots, \lambda_m, x)$ one ignores $\lambda_1, \dots, \lambda_m$ and checks $f(x)$.[1]

In addition, one checks $f(x)$ for all points x that violate the conditions of 3a1; that is, $\nabla g_1(x), \dots, \nabla g_m(x)$ are linearly dependent, or $f, g_1, \dots, g_m$ fail to be continuously differentiable near x.

If the set Z is not compact, one checks *all* relevant limits of f.

If all that is feasible (which is not guaranteed!), one finally obtains the infimum and supremum of f on Z.

More formally: $\sup_{x \in Z} f(x) = \lim_k f(x_k) \in (-\infty, +\infty]$ for some $x_1, x_2, \dots \in Z$. Choosing a subsequence we ensure either $x_k \to x$ for some $x \in \overline{Z}$ or $|x_k| \to \infty$. In the case $x \in Z$ the point x must violate conditions of 3a1. That is enough if Z is compact. Otherwise, if Z is bounded and not closed, the case $x \in \overline{Z} \setminus Z$ must be examined. And if Z is unbounded, the case $|x_k| \to \infty$ must be examined.

In order to prove Th. 3a1 we first generalize Th. 2c3 as follows (recall 2a9).

3a2 Theorem. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable near 0, $f(0) = 0$, and $(Df)_0 = A : \mathbb{R}^n \to \mathbb{R}^m$ be onto. Then f is open at 0.

Proof. We take an m-dimensional subspace $E \subset \mathbb{R}^n$ such that $A|_E$ is an invertible mapping from E onto $\mathbb{R}^m$ (this is possible, as explained in Sect. 2a, Item "linear algebra"). Then $\big(D(f|_E)\big)_0 = A|_E$ is invertible; by Th. 2b1,[2] $f|_E$ is a local diffeomorphism, and therefore,[3] open at 0. It follows that f is open at 0. $\square$

[1] Being ignored in this framework, $(\lambda_1, \dots, \lambda_m)$ are of interest in another framework, see Sect. 3e.

[2] Choosing a basis in E we turn it to a copy of $\mathbb{R}^m$. Or, alternatively, E may be chosen to be spanned by some m out of the n standard basis vectors of $\mathbb{R}^n$.

[3] Use 2a7(a), as in the proof of 2c3.

***Proof** of Theorem 3a1.* WLOG, the extremum is maximum, $x_0 = 0$ and $f(0) = 0$. Assume the contrary: $\nabla f(0)$ is not a linear combination of $\nabla g_1(0), \dots, \nabla g_m(0)$. Then vectors $\nabla g_1(0), \dots, \nabla g_m(0), \nabla f(0)$ are linearly independent. These vectors being the rows of $(D\varphi)_0$, where $\varphi(x) = \big(g_1(x), \dots, g_m(x), f(x)\big)$, we see that $(D\varphi)_0 : \mathbb{R}^n \to \mathbb{R}^{m+1}$ is onto.[1] By Th. 3a2, φ is open at 0.

We take a neighborhood $U \subset \mathbb{R}^n$ of 0 such that $f(x) \le f(x_0)$ for all $x \in U \cap Z$ (where $Z = g^{-1}\big(\{0\}\big)$), note that $\varphi(U)$ is a neighborhood of 0 in $\mathbb{R}^{m+1}$, and therefore $\varphi(U)$ contains $(0, \dots, 0, \varepsilon)$ for $\varepsilon > 0$ small enough. That is, $\varphi(x) = (0, \dots, 0, \varepsilon)$ for some $x \in U$. Then $x \in Z$ and $f(x) > f(0)$, which is a contradiction. □

Theorem 3a1, formulated in terms of gradients, involves a Euclidean metric on $\mathbb{R}^n$. However, it is easy to reformulate it for vector spaces (with no given metric), to be invariant under arbitrary change of basis (not just orthonormal), as follows.

Assume that V is an n-dimensional vector space, $x_0 \in V$, $1 \le m \le n-1$, functions $f, g_1, \dots, g_m : V \to \mathbb{R}$ are continuously differentiable near x_0, $g_1(x_0) = \dots = g_m(x_0) = 0$, and the linear functions $(Dg_1)_{x_0}, \dots, (Dg_m)_{x_0} : V \to \mathbb{R}$ are linearly independent. If x_0 is a local constrained extremum point of f subject to $g_1(\cdot) = \dots = g_m(\cdot) = 0$, then there exist $\lambda_1, \dots, \lambda_m \in \mathbb{R}$ such that

$$(Df)_{x_0} = \lambda_1 (Dg_1)_{x_0} + \dots + \lambda_m (Dg_m)_{x_0} \,.$$

3b Example: arithmetic, geometric, harmonic, and more general means

Here is an isoperimetric inequality for triangles Δ on the plane:

$$\operatorname{area}(\Delta) \le \frac{1}{12\sqrt{3}} \big(\operatorname{perimeter}(\Delta)\big)^2 \,,$$

and equality is attained for equilateral triangles and only for them. In other words, among all triangles with the given perimeter, the equilateral one has the largest area.[2]

---

[1] Recall Sect. 2a, Item "linear algebra".

[2] Generally, $\operatorname{area}(G) \le \frac{1}{4\pi}\big(\operatorname{perimeter}(G)\big)^2$ for any G on the plane, and equality is attained for disks only. This is a famous deep fact. But I do not give an exact formulation (nor a proof, of course).

The proof is based on Heron's formula for the area A of a triangle whose side lengths are x, y, z (and perimeter $L = x + y + z$):

$$A^2 = \frac{L}{2}\left(\frac{L}{2} - x\right)\left(\frac{L}{2} - y\right)\left(\frac{L}{2} - z\right).$$

The sum of the three positive[1] numbers $\frac{L}{2} - x$, $\frac{L}{2} - y$, $\frac{L}{2} - z$ is fixed (equal to $\frac{3L}{2} - L = \frac{L}{2}$); their product is claimed to be maximal when these numbers are equal (to $L/6$), and then $A^2 = \frac{L}{2}\left(\frac{L}{6}\right)^3 = \frac{L^4}{2^4 \cdot 3^3}$; $A = \frac{L^2}{2^2 \cdot 3\sqrt{3}}$.

More generally, $\max\{x_1 \dots x_n : x_1, \dots, x_n \ge 0, x_1 + \dots + x_n = c\}$ is reached for $x_1 = \dots = x_n = c/n$ and is equal to $(c/n)^n$. Equivalently, $\max\{(x_1 \dots x_n)^{1/n} : x_1, \dots, x_n \ge 0, (x_1 + \dots + x_n)/n = c\}$ is reached for $x_1 = \dots = x_n = c$ and is equal to c, which is the well-known inequality for geometric mean and arithmetic mean,

$$\text{(3b1)} \quad (x_1 \dots x_n)^{1/n} \le \frac{1}{n}(x_1 + \dots + x_n) \quad \text{for } n = 1, 2, \dots \text{ and } x_1, \dots, x_n \ge 0.$$

It follows easily from concavity of the logarithm: the set $A = \{(x, y) : x \in (0, \infty), y \le \ln x\}$ is convex, therefore the convex combination $\left(\frac{1}{n}(x_1 + \dots + x_n), \frac{1}{n}(\ln x_1 + \dots + \ln x_n)\right)$ of points $(x_1, \ln x_1), \dots, (x_n, \ln x_n) \in A$ belongs to A, which gives (3b1). And still, it is worth to exercise Lagrange multipliers.

3b2 Exercise. Prove (3b1) via Lagrange multipliers.

By the way, the harmonic mean h defined by $\frac{1}{h} = \frac{1}{n}\left(\frac{1}{x_1} + \dots + \frac{1}{x_n}\right)$ satisfies $h \le (x_1 \dots x_n)^{1/n}$; just apply (3b1) to $\frac{1}{x_1}, \dots, \frac{1}{x_n}$.

More generally, the Hölder mean (called also power mean) with exponent $p \in (-\infty, 0) \cup (0, \infty)$ is

$$M_p(x_1, \dots, x_n) = \left(\frac{x_1^p + \dots + x_n^p}{n}\right)^{1/p} \quad \text{for } x_1, \dots, x_n > 0.$$

In particular, M_1 is the arithmetic mean and M_{-1} is the harmonic mean. For $p \to 0$ L'Hôpital's rule gives

$$\ln \lim_{p \to 0} M_p((x_1, \dots, x_n) = \lim_{p \to 0} \frac{1}{p} \ln \frac{x_1^p + \dots + x_n^p}{n} =$$
$$= \lim_{p \to 0} \frac{x_1^p \ln x_1 + \dots + x_n^p \ln x_n}{x_1^p + \dots + x_n^p} = \frac{\ln x_1 + \dots + \ln x_n}{n} = \ln(x_1 \dots x_n)^{1/n};$$

[1] $\frac{L}{2} - x = \frac{x+y+z}{2} - x = \frac{y+z-x}{2} > 0$ by the triangle inequality.

accordingly, one defines

$$M_0(x_1,\dots,x_n) = (x_1 \dots x_n)^{1/n}\,,$$

and observes that $M_{-1}(x_1,\dots,x_n) \le M_0(x_1,\dots,x_n) \le M_1(x_1,\dots,x_n)$. For $p \to +\infty$ we have

$$\frac1n \max(x_1^p,\dots,x_n^p) \le \frac{x_1^p+\dots+x_n^p}{n} \le \max(x_1^p,\dots,x_n^p)\,,$$

therefore $M_p(x_1,\dots,x_n) \to \max(x_1,\dots,x_n)$; one writes

$$M_{+\infty}(x_1,\dots,x_n) = \max(x_1,\dots,x_n)\,; \quad M_{-\infty}(x_1,\dots,x_n) = \min(x_1,\dots,x_n)$$

(the latter being similar to the former) and observes that $M_{-\infty}(x_1,\dots,x_n) \le M_{-1}(x_1,\dots,x_n) \le M_0(x_1,\dots,x_n) \le M_1(x_1,\dots,x_n) \le M_{+\infty}(x_1,\dots,x_n)$. That is interesting! Maybe $M_p \le M_q$ whenever $p \le q$?

We treat M_p as a function on $(0,\infty)^n \subset \mathbb{R}^n$ and calculate its gradient ∇M_p, or rather, the direction of the vector ∇M_p; indeed, we only need to know when two vectors ∇M_p, ∇M_q are linearly dependent, that is, collinear (denote it $\shortparallel$). We have $\nabla M_p \shortparallel \nabla M_p^p \shortparallel \nabla(nM_p^p) \shortparallel (x_1^{p-1},\dots,x_n^{p-1})$ for $p \ne 0$; however, this result holds for $p=0$ as well, since $\nabla M_0 \shortparallel \nabla \ln M_0 \shortparallel (x_1^{-1},\dots,x_n^{-1})$. Thus, ∇M_p, ∇M_q are collinear if and only if $\frac{x_1^{q-1}}{x_1^{p-1}} = \dots = \frac{x_n^{q-1}}{x_n^{p-1}}$, that is, $x_1^{q-p} = \dots = x_n^{q-p}$, or just $x_1 = \dots = x_n$. In this case, evidently, $M_p = M_q$. Does it prove that $M_p \le M_q$ always? Not yet. Functions M_p, M_q are continuously differentiable on the open set $G = (0,\infty)^n$, and on the set $Z_p = \{x \in G : M_p(x) = 1\}$[1] the conditions of 3a1 are violated at one point $(1,\dots,1)$ only. This could not happen on a compact Z_p! Surely Z_p is not compact, and we must examine $\overline{Z}_p \setminus Z_p$ and/or ∞.

Case 1: $0 < p < q < \infty$. The set Z_p is bounded, since $\max(x_1,\dots,x_n) \le (x_1^p+\dots+x_n^p)^{1/p} = n^{1/p} M_p(x_1,\dots,x_n) = n^{1/p}$, but not closed.[2] Functions M_p, M_q are continuous on $\overline{G} = [0,\infty)^n$. Maybe the (global) minimum of M_q on $\overline{Z_p} = \{x \in \overline{G} : M_p(x) = 1\}$ is reached at some $x \in \overline{Z}_p \setminus Z_p$? In this case at least one coordinate of x vanishes. We use induction in n. For $n=1$, $M_p(x) = x = M_q(x)$. Having $M_p \le M_q$ in dimension $n-1$ we get (assuming

[1]No need to consider $M_p(x) = c$, since $M_p(\lambda x) = \lambda M_p(x)$ for all $\lambda \in (0,\infty)$ and all p, thus $\frac{M_q(\lambda x)}{M_p(\lambda x)}$ does not depend on λ.

[2]For example, the point $(n^{1/p},0,\dots,0)$ belongs to $\overline{Z}_p \setminus Z_p$.

$x_n = 0$)

$$\frac{M_q(x)}{M_p(x)} = \frac{\Big(\frac1n(x_1^q + \dots + x_{n-1}^q + 0^q)\Big)^{1/q}}{\Big(\frac1n(x_1^p + \dots + x_{n-1}^p + 0^p)\Big)^{1/p}} =$$

$$= \Big(\frac{n}{n-1}\Big)^{\frac1p - \frac1q} \frac{\Big(\frac1{n-1}(x_1^q + \dots + x_{n-1}^q)\Big)^{1/q}}{\Big(\frac1{n-1}(x_1^p + \dots + x_{n-1}^p)\Big)^{1/p}} \ge \Big(\frac{n}{n-1}\Big)^{\frac1p - \frac1q} > 1\,,$$

therefore $M_q > M_p$ on $\overline{Z}_p \setminus Z_p$.

CASE 2: $0 = p < q < \infty$. Follows from Case 1 via the limiting procedure $p \to 0+$.

CASE 3: $-\infty < p < q < 0$. Follows from Case 1 applied to $1/x_1, \dots, 1//x_n$, since

$$1/M_{-p}(x_1^{-1}, \dots, x_n^{-1}) = \Big(\frac{x_1^p + \dots + x_n^p}{n}\Big)^{1/p} = M_p(x_1, \dots, x_n)\,;$$
$$M_p(x_1, \dots, x_n) = 1/M_{-p}(x_1^{-1}, \dots, x_n^{-1}) \le 1/M_{-q}(x_1^{-1}, \dots, x_n^{-1}) = M_q(x_1, \dots, x_n)\,.$$

CASE 4: $-\infty < p < q = 0$. Follows from Case 3 via the limiting procedure $q \to 0-$.

CASE 5: $-\infty < p < 0 < q < \infty$. Follows from Cases 2 and 4: $M_p \le M_0 \le M_q$.

So, $M_p \le M_q$ whenever $p \le q$.

Some practical advice.

The system of $m+n$ equations proposed in Sect. 3a is only one way of finding local constrained extrema. Not necessarily the simplest way.

No need to find ∇f when $f(\cdot) = \varphi(g(\cdot))$; just find ∇g and note that ∇f is collinear to ∇g.

In many cases there are alternatives to the Lagrange method. For example, we could replace $\inf\{M_q(x) : M_p(x) = 1\}$ with $\inf\big\{\frac{M_q(x)}{M_p(x)} : M_1(x) = 1\big\}$, substitute $x_n = n - (x_1 + \dots + x_{n-1})$ and optimize in $x_1, \dots, x_{n-1}$ without constraints. Alternatively we could use convexity of the function $t \mapsto t^{q/p}$, that is, convexity of the set $A = \big\{(t,u) : t \in (0,\infty), u \ge t^{q/p}\big\}$. The convex combination $\big(\frac1n(x_1^p + \dots + x_n^p), \frac1n(x_1^q + \dots + x_n^q)\big)$ of points $(x_1^p, x_1^q), \dots, (x_n^p, x_n^q) \in A$ belongs to A, which gives $\big(\frac1n(x_1^p + \dots + x_n^p)\big)^{q/p} \le \frac1n(x_1^q + \dots + x_n^q)$, that is, $M_p \le M_q$. Moreover, the same applies to *weighted* mean

$$M_{p,w}(x) = (x_1^p w_1 + \dots + x_n^p w_n)^{1/p}$$

for given $w_1, \dots, w_n \ge 0$ satisfying $w_1 + \dots + w_n = 1$. In particular, $M_{1,w}(x) \le M_{p,w}(x)$ for $p \ge 1$, that is, $x_1 w_1 + \dots + x_n w_n \le (x_1^p w_1 + \dots + x_n^p w_n)^{1/p}$. Substituting $x_i = a_i b_i^{-q/p}$ and $w_i = b_i^q$ where q is such that $\frac{1}{p} + \frac{1}{q} = 1$ we have $\sum_i a_i b_i^{-q/p} b_i^q \le \bigl(\sum_i a_i^p b_i^{-q} b_i^q\bigr)^{1/p}$, that is, $\sum_i a_i b_i \le (\sum_i a_i^p)^{1/p}$ provided that $\sum_i b_i^q = 1$. This leads easily to the *Hölder's inequality*

$$\Bigl| \sum_i x_i y_i \Bigr| \le \Bigl(\sum_i |x_i|^p \Bigr)^{1/p} \Bigl(\sum_i |y_i|^q \Bigr)^{1/q}$$

for $p, q \in (1, \infty)$, $\frac{1}{p} + \frac{1}{q} = 1$, and arbitrary $x_i, y_i \in \mathbb{R}$. The right-hand side may be rewritten as $n M_p(|x|) M_q(|y|)$, admitting $p, q \in [1, \infty]$. Note the special cases $p = q = 2$ and $p = 1, q = \infty$.

However, the shown way to this inequality is rather tricky.

3b3 Exercise. Given $a_1, \dots, a_n > 0$, maximize $a_1 x_1 + \dots + a_n x_n$ on $\{x \in [0, \infty)^n : x_1^p + \dots + x_n^p = 1\}$ using the Lagrange method.[1] Deduce Hölder's inequality.

Hölder's inequality persists in the case of countably many variables x_i and y_i. If two series $\sum |x_i|^p$ and $\sum |y_i|^q$ converge (and $\frac{1}{p} + \frac{1}{q} = 1$), then the series $\sum x_i y_i$ also converges (and the inequality holds).

3b4 Exercise. Given $a, b, c, k > 0$, find the maximum of the function $f(x, y, z) = x^a y^b z^c$ where $x, y, z \in [0, \infty)$ and $x^k + y^k + z^k = 1$.

3b5 Exercise. Find the maximum of y over all points $(x, y) \in \mathbb{R}^2$ that satisfy the equation $x^2 + xy + y^2 = 27$.

3c Example: Three points on a spheroid

We consider an ellipsoid of revolution (in other words, spheroid)

$$x^2 + y^2 + \alpha z^2 = 1$$

for some $\alpha \in (0, 1) \cup (1, \infty)$, and three points P, Q, R on this surface. We want to maximize $|PQ|^2 + |QR|^2 + |RP|^2$.

We'll see that the maximum is reached when P, Q, R are situated either in the horizontal plane $z = 0$ or the vertical plane $y = 0$ (or another vertical plane through the origin; they all are equivalent due to symmetry). Thus, the three-dimensional problem boils down to a pair of two-dimensional problems (not to be solved here).

[1] Hint: induction in n is needed again.

We introduce 9 coordinates,

$$P = (x_1, y_1, z_1)\,, \quad Q = (x_2, y_2, z_2)\,, \quad R = (x_3, y_3, z_3)$$

and 4 functions $f, g_1, g_2, g_3 : \mathbb{R}^9 \to \mathbb{R}$ of these coordinates,

$$\begin{aligned} f(x_1, \dots, z_3) =& (x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 \\ & + (x_2 - x_3)^2 + (y_2 - y_3)^2 + (z_2 - z_3)^2 \\ & + (x_3 - x_1)^2 + (y_3 - y_1)^2 + (z_3 - z_1)^2\,; \\ g_1(x_1, \dots, z_3) =& x_1^2 + y_1^2 + \alpha z_1^2 - 1\,, \\ g_2(x_1, \dots, z_3) =& x_2^2 + y_2^2 + \alpha z_2^2 - 1\,, \\ g_3(x_1, \dots, z_3) =& x_3^2 + y_3^2 + \alpha z_3^2 - 1\,. \end{aligned}$$

We use the approach of Sect. 3a with $n = 9$, $m = 3$. The functions f, g_1, g_2, g_3 are continuously differentiable on $\mathbb{R}^9$. The set $Z = Z_{g_1, g_2, g_3} \subset \mathbb{R}^9$ is compact. The gradients of g_1, g_2, g_3 do not vanish on Z (check it) and are linearly independent (and moreover, orthogonal).

We introduce Lagrange multipliers $\lambda_1, \lambda_2, \lambda_3$ corresponding to g_1, g_2, g_3 and consider a system of $m + n = 12$ equations for 12 unknowns. The first three equations are

$$x_1^2 + y_1^2 + \alpha z_1^2 = 1\,, \quad x_2^2 + y_2^2 + \alpha z_2^2 = 1\,, \quad x_3^2 + y_3^2 + \alpha z_3^2 = 1\,.$$

Now, the partial derivatives. We have

$$\frac{\partial f}{\partial x_1} = 2(x_1 - x_2) - 2(x_3 - x_1) = 4x_1 - 2x_2 - 2x_3\,,$$

which is convenient to write as $6x_1 - 2(x_1 + x_2 + x_3)$; similarly,

$$\begin{aligned} \frac{\partial f}{\partial x_k} &= 6x_k - 2(x_1 + x_2 + x_3)\,, \\ \frac{\partial f}{\partial y_k} &= 6y_k - 2(y_1 + y_2 + y_3)\,, \\ \frac{\partial f}{\partial z_k} &= 6z_k - 2(z_1 + z_2 + z_3) \end{aligned}$$

for $k = 1, 2, 3$. Also,

$$\frac{\partial g_k}{\partial x_k} = 2x_k\,, \quad \frac{\partial g_k}{\partial y_k} = 2y_k\,, \quad \frac{\partial g_k}{\partial z_k} = 2\alpha z_k\,;$$

other partial derivatives vanish. We get 9 more equations:

$$
\begin{aligned}
6x_k - 2(x_1 + x_2 + x_3) &= \lambda_k \cdot 2x_k \, , \\
6y_k - 2(y_1 + y_2 + y_3) &= \lambda_k \cdot 2y_k \, , \\
6z_k - 2(z_1 + z_2 + z_3) &= \lambda_k \cdot 2\alpha z_k
\end{aligned}
$$

for $k = 1, 2, 3$. That is,

$$
\begin{aligned}
(3 - \lambda_k) x_k &= x_1 + x_2 + x_3 \, , \\
(3 - \lambda_k) y_k &= y_1 + y_2 + y_3 \, , \\
(3 - \alpha\lambda_k) z_k &= z_1 + z_2 + z_3 \, .
\end{aligned}
$$

We note that

$$(x_1 + x_2 + x_3) y_k = (3 - \lambda_k) x_k y_k = (y_1 + y_2 + y_3) x_k$$

for $k = 1, 2, 3$.

Case 1: $x_1 + x_2 + x_3 \neq 0$ or $y_1 + y_2 + y_3 \neq 0$.

Then P, Q, R are situated on the vertical plane $\{(x, y, z) : (x_1+x_2+x_3)y = (y_1 + y_2 + y_3)x\}$.

Case 2: $x_1 + x_2 + x_3 = y_1 + y_2 + y_3 = 0$ and $(\lambda_1, \lambda_2, \lambda_3) \neq (3, 3, 3)$.

If $\lambda_1 \neq 3$ then $x_1 = y_1 = 0$; the three vectors $(x_1, y_1), (x_2, y_2), (x_3, y_3) \in \mathbb{R}^2$ (of zero sum!) are collinear; therefore P, Q, R are situated on a vertical plane (again). The same holds if $\lambda_2 \neq 3$ or $\lambda_3 \neq 3$.

Case 3: $x_1 + x_2 + x_3 = y_1 + y_2 + y_3 = 0$ and $\lambda_1 = \lambda_2 = \lambda_3 = 3$.

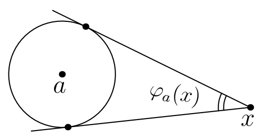
Then $z_1 = z_2 = z_3 = \frac{z_1+z_2+z_3}{3-3\alpha}$ (since $\alpha \neq 0$), therefore $z_1 = z_2 = z_3 = 0$; P, Q, R are situated on the horizontal plane $\{(x, y, z) : z = 0\}$.

Another practical advice.

If Lagrange method does not solve a problem to the end, it may still give a useful information. Combine it with other methods as needed.

3c1 Exercise. [1]

Let $a, b \in \mathbb{R}^n$ be linearly independent, $|a| = 5$, $|b| = 10$. Functions φ_a, φ_b on the sphere $S_1(0) = \{x : |x| = 1\} \subset \mathbb{R}^n$ are defined as follows: $\varphi_a(x)$ is the angular diameter of the sphere $S_1(a) = \{y : |y - a| = 1\}$ viewed from x; similarly, $\varphi_b(x)$ is the angular diameter of $S_1(b)$ from x.



Prove that every point of local extremum of the function $\varphi_a + \varphi_b$ on $S_1(0)$ is some linear combination of a, b.[2]

---

[1] Exam of 26.01.14, Question 2.

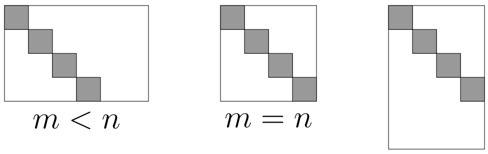
[2] Hint: show that $\sin \frac{1}{2}\varphi_a(x) = 1/|x - a|$; use the gradient.

3d Example: Singular value decomposition

3d1 Proposition. Every linear operator from one finite-dimensional Euclidean vector space to another sends some orthonormal basis of the first space into an orthogonal system in the second space.

This is called the Singular Value Decomposition.[1] It may be reformulated as follows.

3d2 Proposition. Every linear operator from an n-dimensional Euclidean vector space to an m-dimensional Euclidean vector space has a diagonal $m \times n$ matrix in some pair of orthonormal bases.



$m < n$ $m = n$ $m > n$

In particular, this holds for every linear operator $\mathbb{R}^n \to \mathbb{R}^n$. It does not mean that every matrix is diagonalizable! Two bases give much more freedom than one basis.

Do you think this is unrelated to constrained optimization? Wait a little.

Prop. 3d1 will be derived from Prop. 3d3 below.

3d3 Proposition. Every finite-dimensional vector space endowed with two Euclidean metrics contains a basis orthonormal in the first metric and orthogonal in the second metric.

Proof. Let an n-dimensional vector space V be endowed with two Euclidean metrics. It means, two norms $|\cdot|$ and $|\cdot|_1$ corresponding to two inner products $\langle\cdot,\cdot\rangle$ and $\langle\cdot,\cdot\rangle_1$ by $|x|^2 = \langle x, x\rangle$ and $|x|_1^2 = \langle x, x\rangle_1$. We denote by E the Euclidean space $(V, |\cdot|)$ and define a mapping $A : E \to E$ by

$$\forall x, y \in E \quad \langle x, y\rangle_1 = \langle Ax, y\rangle\,;$$

it is well-defined, since the linear form $\langle x, \cdot\rangle_1$, as every linear form, is $\langle a, \cdot\rangle$ for some $a \in E$. It is easy to see that A is a linear operator, symmetric in the sense that

$$\forall x, y \in E \quad \langle Ax, y\rangle = \langle x, Ay\rangle\,.$$

[1]See: Todd Will, "Introduction to the Singular Value Decomposition", `http://websites.uwlax.edu/twill/svd/` *Quote:*

The Singular Value Decomposition (SVD) is a topic rarely reached in undergraduate linear algebra courses and often skipped over in graduate courses.

Consequently relatively few mathematicians are familiar with what M.I.T. Professor Gilbert Strang calls "absolutely a high point of linear algebra."

We want to maximize $|\cdot|_1^2$ on the sphere $S = \{x \in E : |x| = 1\}$. We have[1]

$$\nabla |x|^2 = 2x\,, \quad \nabla |x|_1^2 = 2Ax$$

by 1d1(a), or just by a very simple calculation:

$$|x+h|^2 = |x|^2 + \langle x, h\rangle + \langle h, x\rangle + |h|^2 = |x|^2 + 2\langle x, h\rangle + o(|h|)\,,$$
$$|x+h|_1^2 = |x|_1^2 + \langle x, h\rangle_1 + \langle h, x\rangle_1 + |h|_1^2 = |x|_1^2 + 2\langle Ax, h\rangle + o(|h|)\,.$$

These two gradients are collinear if and only if $\exists \lambda \;\; Ax = \lambda x$; it means, x is an eigenvector of A, and λ is the eigenvalue. Now we could use well-known results of linear algebra, but here is the analytic way.

By compactness, $|\cdot|_1^2$ reaches its maximum on S; by Theorem 3a1, a maximizer is an eigenvector. Existence of an eigenvector is thus proved. Denote it by e_n, and the eigenvalue by λ_n.

If $x \perp e_n$ then $Ax \perp e_n$ due to symmetry of A: $\langle Ax, e_n\rangle = \langle x, Ae_n\rangle = \langle x, \lambda_n e_n\rangle = \lambda_n \langle x, e_n\rangle = 0$. We consider a hyperplane (that is, $(n-1)$-dimensional subspace)

$$E_{n-1} = \{x \in E : x \perp e_n\}$$

and the restricted operator

$$A_{n-1} : E_{n-1} \to E_{n-1}\,, \quad A_{n-1}x = Ax \text{ for } x \in E_{n-1}\,.$$

The Euclidean space E_{n-1} is endowed with two Euclidean metrics $|\cdot|$ and $|\cdot|_1$ (restricted to E_{n-1}), and $\langle x, y\rangle_1 = \langle A_{n-1}x, y\rangle$ for $x, y \in E_{n-1}$.

Now we use induction in n. The case $n = 1$ is trivial. The claim for $n-1$ applied to E_{n-1} gives a basis $(e_1, \dots, e_{n-1})$ of E_{n-1} orthonormal in $|\cdot|$ and orthogonal in $|\cdot|_1$. Thus, $(e_1, \dots, e_{n-1}, e_n)$ is a basis of E. We normalize e_n to $|e_n| = 1$; now this basis is orthonormal in $|\cdot|$. It is also orthogonal in $|\cdot|_1$, since $\langle e_k, e_n\rangle_1 = \langle Ae_k, e_n\rangle = 0$ for $k = 1, \dots, n-1$. □

3d4 Remark. Positivity of the quadratic form $x \mapsto |x|_1^2 = \langle x, x\rangle_1$ was not used. The same holds for arbitrary quadratic form on a Euclidean space. (In contrast, positivity of $|\cdot|^2$ was used.)

Proof *of Prop. 3d1.* We have two Euclidean spaces E, E_2 and a linear operator $T : E \to E_2$. First, assume in addition that T is one-to-one. Then T induces a second Euclidean metric on E:

$$|x|_1 = |Tx|\,; \quad \langle x, y\rangle_1 = \langle Tx, Ty\rangle$$

[1]All gradients are taken in $E = (V, |\cdot|)$, not $(V, |\cdot|_1)$!

(of course, $|Tx|$ is the norm in E_2). Prop. 3d3 gives an orthonormal basis $(e_1, \dots, e_n)$ of E, orthogonal in the second metric: $\langle e_k, e_l \rangle_1 = 0$ for $k \ne l$. That is, $\langle Te_k, Te_l \rangle = 0$, which shows that $(Te_1, \dots, Te_n)$ is an orthogonal system in E_2.

If T is not one-to-one, the same argument applies due to Remark 3d4.[1] □

Prop. 3d2 follows immediately, and gives a diagonal matrix. Its diagonal elements can be made ≥ 0 (changing signs of basis vectors as needed) and decreasing (renumbering basis vectors as needed); this way one gets the so-called *singular values* of the given operator T. They depend on T only, not on the choice of the pair of bases,[2,3] and are the square roots of the eigenvalues of the operator $A = T^*T$. The highest singular value is the operator norm $\|T\|$ of T (think, why). The lowest singular value (if not 0) is $1/\|T^{-1}\|$.

3e Sensitivity of optimum to parameters

When using a mathematical model one often bothers about sensitivity[4] of the result (the output of the model) to the assumptions (the input). Here is one of such questions.[5]

What happens if the restrictions $g_1(x) = \dots = g_m(x) = 0$ are replaced with $g_1(x) = c_1, \dots, g_m(x) = c_m$?

Assume that the system of $m + n$ equations

$$\begin{aligned} &g_1(x) = c_1, \dots, g_m(x) = c_m \,, && (m \text{ equations}) \\ &\nabla f(x) = \lambda_1 \nabla g_1(x) + \dots + \lambda_m \nabla g_m(x) && (n \text{ equations}) \end{aligned}$$

for $(\lambda, x) \in \mathbb{R}^m \times \mathbb{R}^n$ has a solution $(\lambda(c), x(c))$ for all $c \in \mathbb{R}^m$ near 0, *and* the mapping $c \mapsto x(c)$ is differentiable at 0. Then, by the chain rule,

$$\frac{\partial}{\partial c_k}\Big|_{c=0} f(x(c)) = \Big\langle \nabla f(x(0)), \frac{\partial}{\partial c_k}\Big|_{c=0} x(c) \Big\rangle \quad \text{for } k = 1, \dots, m \,.$$

On the other hand,

$$\nabla f(x(0)) = \lambda_1(0) \nabla g_1(x(0)) + \dots + \lambda_m(0) \nabla g_m(x(0))$$

[1] Alternatively, define $|x|_1^2 = |Tx|^2 + |x|^2$, $\langle x, y \rangle_1 = \langle Tx, Ty \rangle + \langle x, y \rangle$.

[2] The only freedom in this choice (in addition to sign change and renumbering) is, rotation within each eigenspace of dimension > 1 (if any).

[3] On the space of operators, the *Schatten norm* is $\|T\|_p = \big(|s_1|^p + \dots + |s_n|^p \big)^{1/p}$ where $s_1, \dots, s_n$ are the singular values of T (and $1 \le p \le \infty$).

[4] Closely related ideas: stability, robustness; uncertainty; elasticity, ...

[5] A more general one: $g_1(x, c_1) = 0, \dots, g_m(x, c_m) = 0$.

and

$$\left\langle \nabla g_1(x(0)), \frac{\partial}{\partial c_k}\Big|_{c=0} x(c) \right\rangle = \frac{\partial}{\partial c_k}\Big|_{c=0} g_1(x(c)) = \begin{cases} 1, & \text{if } k=1, \\ 0, & \text{otherwise} \end{cases}$$

(since $g_1(x(c)) = c_1$). The same holds for $g_2, \dots, g_m$. Therefore

$$\frac{\partial}{\partial c_k}\Big|_{c=0} f(x(c)) = \lambda_k(0) \, .$$

It means that $\lambda_k = \lambda_k(0)$ is the sensitivity of the critical value to the level c_k of the constraint $g_k(x) = c_k$. That is,

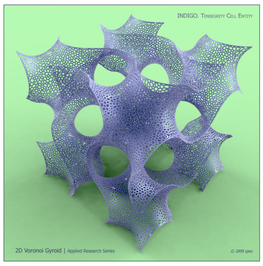
$$f(x(c)) = f(x(0)) + \lambda_1(0)c_1 + \dots + \lambda_m(0)c_m + o(|c|) \, .$$

Does it mean that

$$\sup_{Z_c} f = \sup_{Z_0} f + \lambda_1(0)c_1 + \dots + \lambda_m(0)c_m + o(|c|) \qquad \text{(3e1)}$$

where $Z_c = \{x : g_1(x) = c_1, \dots, g_m(x) = c_m\}$? Not necessarily, for several reasons (possible non-compactness, non-differentiability, greater or equal value at another critical point when $c = 0$). But if $\sup_{Z_c} f = f(x(c))$ for all c near 0 then (3e1) holds.[1]

3f Manifolds in $\mathbb{R}^n$



Everyone knows what a curve is, until he has studied enough mathematics... Felix Klein[2]

Image: (CC) Jonathan Johanson, `http://cliptic.wordpress.com`

By a manifold (to be defined soon) we mean a differential k-dimensional submanifold of $\mathbb{R}^n$, of class C^1, without boundary.[3] It is also called "k-dimensional smooth surface in $\mathbb{R}^n$" or "k-dimensional submanifold on $\mathbb{R}^n$",[4] or "smooth manifold in $\mathbb{R}^n$"[5] etc.

[1] See also Sect. 13.2 in book: J. Cooper, "Working analysis", Elsevier 2005.

[2] Quoted from: Hubbard, Sect. 3.1 "Manifolds".

[3] 'Generally, "smooth" means "as many times differentiable as is relevant to the problem at hand. ... (Some authors use "smooth" to mean C^∞: "infinitely many times differentiable". For our purposes this is overkill.)' Hubbard, Sect. 3.1, p. 293–294.

[4] Zorich Sect. 8.7.1.

[5] Hubbard Sect. 3.1.

Several equivalent definitions of a manifold are used: via equations;[1] via diffeomorphisms;[2] via graphs of mappings;[3] and via parametrizations (so-called charts, to be treated in Analysis-4).

3f1 Theorem. The following conditions on a set $M \subset \mathbb{R}^n$, a point $x_0 \in M$ and a number $k \in \{1, 2, \dots, n-1\}$ are equivalent:

(a) there exists a mapping $f : \mathbb{R}^n \to \mathbb{R}^{n-k}$, continuously differentiable near x_0, such that $(Df)_{x_0} = A : \mathbb{R}^n \to \mathbb{R}^{n-k}$ is onto, and

$$x \in M \iff f(x) = f(x_0) \quad \text{for all } x \text{ near } x_0\,;$$

(b) there exists a local diffeomorphism φ near x_0 such that

$$x \in M \iff \varphi(x) \in \mathbb{R}^k \times \{0_{n-k}\} \quad \text{for all } x \text{ near } x_0\,;$$

(c) there exists a permutation $(i_1, \dots, i_n)$ of $\{1, \dots, n\}$ and a mapping $g : \mathbb{R}^k \to \mathbb{R}^{n-k}$, continuously differentiable near $(x_{0,i_1}, \dots, x_{0,i_k})$, such that

$$x \in M \iff g(x_{i_1}, \dots, x_{i_k}) = (x_{i_{k+1}}, \dots, x_{i_n}) \quad \text{for all } x \text{ near } x_0\,.$$

Proof. First, WLOG, $x_0 = 0$ (as usual).

Second, the three conditions are insensitive to permutations of the n coordinates of x.[4] Indeed, in (a) we may change the order of arguments of f as needed; in (b) we may change the order of arguments of φ as needed; and in (c) we may change the permutation $(i_1, \dots, i_n)$ as needed.

(a)$\Longrightarrow$(c): WLOG, $f(0) = 0$ and $A = (B\,|\,C)$ with $B = \mathbb{R}^k \to \mathbb{R}^{n-k}$, $C : \mathbb{R}^{n-k} \to \mathbb{R}^{n-k}$, C invertible (using the fact that $\operatorname{rank} A = n-k$). Theorem 2b3 (for n and $n-k$ in place of n and m) gives $g : \mathbb{R}^k \to \mathbb{R}^{n-k}$ such that $g(x_1, \dots, x_k) = (x_{k+1}, \dots, x_n) \iff f(x_1, \dots, x_n) = 0 \iff x \in M$, which gives (c) for $(i_1, \dots, i_n) = (1, \dots, n)$.

(c)$\Longrightarrow$(b): WLOG, $(i_1, \dots, i_n) = (1, \dots, n)$. Similarly to the proof of 2b3$\Longrightarrow$2b1 (in Sect. 2a) we define φ by $\varphi(u, v) = \big(u, g(u) - v\big)$ for $u \in \mathbb{R}^k$ and $v \in \mathbb{R}^{n-k}$; then $\varphi(u, v) \in \mathbb{R}^k \times \{0_{n-k}\} \iff \varphi(u, v) = (u, 0) \iff g(u) = v \iff x \in M$.

(b)$\Longrightarrow$(a): we define $f(x) = (y_{k+1}, \dots, y_n)$ whenever $\varphi(x) = (y_1, \dots, y_n)$; then $f(0) = 0$ and $f(x) = 0 \iff \varphi(x) \in \mathbb{R}^k \times \{0_{n-k}\} \iff x \in M$. $\square$

3f2 Definition. A nonempty set $M \subset \mathbb{R}^n$ is a k-dimensional *manifold,* if the equivalent conditions 3f1(a,b,c) hold for every $x_0 \in M$.

[1] Fleming; also Hubbard, Th. 3.1.10.

[2] Lang, Zorich.

[3] Hubbard.

[4] I mean, coordinates of x, not of $f(x)$ or $\varphi(x)$.

We may say that M is a k-manifold near x_0 when 3f1(a,b,c) hold for M, x_0 and k. Accordingly, M is a k-manifold when it is a k-manifold near every point (of M).

3f3 Exercise. Let $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ be a diffeomorphism, and $M \subset \mathbb{R}^n$.

(a) If M is a k-manifold near x_0, then its image $\varphi(M)$ is a k-manifold near $\varphi(x_0)$;

(b) M is a k-manifold if and only if $\varphi(M)$ is a k-manifold.
Prove it.

This applies, in particular, to shifts, rotations, and all invertible affine transformations of $\mathbb{R}^n$.

3f4 Exercise. Let $M_1, M_2 \subset \mathbb{R}^n$ be k-dimensional manifolds, and $M = M_1 \cup M_2$.

(a) If $\overline{M}_1 \cap M_2 = \emptyset$ and $M_1 \cap \overline{M}_2 = \emptyset$, then M is a k-dimensional manifold. Prove it.

(b) It can happen that $M_1 \cap M_2 = \emptyset$ but M is not a k-dimensional manifold. Give a counterexample.

3f5 Exercise. Let $0 < m < n$, and $g_1, \dots, g_m \in C^1(\mathbb{R}^n \to \mathbb{R})$ be such that the vectors $\nabla g_1(x), \dots, \nabla g_m(x)$ are linearly independent for every $x \in M$ where $M = \{x : g_1(x) = \dots = g_m(x) = 0\}$. Then M is a $(n-m)$-dimensional manifold.

Prove it.

3f6 Exercise. Which of the following subsets of $\mathbb{R}^2$ are 1-dimensional manifolds? Prove your answers, both affirmative and negative.

* $M_1 = \mathbb{R} \times \{0\}$;
* $M_2 = [0, 1] \times \{0\}$;
* $M_3 = (0, 1) \times \{0\}$;
* $M_4 = \{(0, 0)\}$;
* $M_5 = \mathbb{R} \times \{0, 1\}$;
* $M_6 = \mathbb{R} \times \mathbb{Z}$;
* $M_7 = \mathbb{R} \times \{1, \frac{1}{2}, \frac{1}{3}, \dots\}$;
* $M_8 = M_7 \cup M_1$.

3f7 Example. The sphere $S = \{x \in \mathbb{R}^n : |x| = 1\}$ is a $(n-1)$-dimensional manifold (by 3f5 for $m = 1$ and $g(x) = |x|^2 - 1$).

Alternatively, we may prove that S is a manifold around just one point, say, $e_1 = (1, 0, \dots, 0)$, and then use rotation invariance: $U(S) = S$ for every linear isometry $U : \mathbb{R}^n \to \mathbb{R}^n$, and each $x \in S$ is Ue_1 for some U;[1] use 3f3(a). Near e_1 the equality $x_1 = \sqrt{1 - x_2^2 - \dots - x_n^2}$ gives 3f1(c).

3f8 Example. [2] Consider the set M of all 3×3 matrices A of the form

$$A = \begin{pmatrix} a^2 & ab & ac \\ ba & b^2 & bc \\ ca & cb & c^2 \end{pmatrix} \quad \text{for } a, b, c \in \mathbb{R}\,, \; a^2 + b^2 + c^2 = 1\,.$$

These are orthogonal projections to one-dimensional subspaces of $\mathbb{R}^3$, that is, straight lines through the origin. Note that each line contains two points of the sphere $S = \{(a, b, c) \in \mathbb{R} : a^2 + b^2 + c^2 = 1\}$, which gives a 2-to-1 mapping $S \to M$. We treat M as a subset of the six-dimensional space of all symmetric 3×3 matrices.

The set M is invariant under transformations $A \mapsto UAU^{-1}$ where U runs over all orthogonal matrices (linear isometries); these are linear transformations of the six-dimensional space of matrices. If A corresponds to $x = (a, b, c)$ then UAU^{-1} corresponds to Ux. For arbitrary $A, B \in M$ there exists U such that $UAU^{-1} = B$ ("transitive action").

Thus, M looks the same around all its points ("homogeneous space"). In order to prove that M is a 2-manifold (in $\mathbb{R}^6$) it is sufficient to prove this near a single point of M, say,

$$A_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \in M\,,$$

that corresponds to $(a, b, c) = (1, 0, 0)$ (but also $(-1, 0, 0)$, of course). For $(a, b, c) \to (1, 0, 0)$ we have in the linear approximation

$$\begin{pmatrix} a^2 & ab & ac \\ ba & b^2 & bc \\ ca & cb & c^2 \end{pmatrix} \approx \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & b & c \\ b & 0 & 0 \\ c & 0 & 0 \end{pmatrix}$$

(think, why). Thus, in the linear approximation all elements of A are functions of two of them. Returning to the nonlinear situation we want to express a^2, b^2, c^2 and bc in terms of ab and ac (locally, for (a, b, c) near $(1, 0, 0)$). We

[1] Since x is the first vector of some orthogonal basis.

[2] The projective plane in disguise.

have

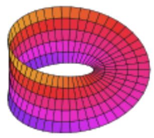
$$(ab)^2 + (ac)^2 = a^2(b^2 + c^2) = a^2(1 - a^2);$$
$$a^2 = \tfrac{1}{2} + \sqrt{\tfrac{1}{4} - (ab)^2 - (ac)^2};$$
$$b^2 = \frac{(ab)^2}{\frac{1}{2} + \sqrt{\dots}}; \quad c^2 = \frac{(ac)^2}{\frac{1}{2} + \sqrt{\dots}}; \quad bc = \frac{(ab)(ac)}{\frac{1}{2} + \sqrt{\dots}};$$

thus, M is a 2-manifold near A_1 according to 3f1(c).[1]

Interestingly, the part of M that corresponds to a spherical zone (symmetrical, around the equator), say $a^2+b^2+c^2 = 1$, $|c| < 1/2$, is homeomorphic to the Möbius strip[2] (without the edge),

$$M = \{h(s,\theta) : s \in (-1,1),\ \theta \in [0,2\pi]\},$$

$$h(s,\theta) = \begin{pmatrix} (R+rs\cos\frac{\theta}{2})\cos\theta \\ (R+rs\cos\frac{\theta}{2})\sin\theta \\ rs\sin\frac{\theta}{2} \end{pmatrix},$$



for given $R > r > 0$. You see, a straight segment on the x, z plane rotates by $\theta/2$ (around the y axis) and at the same time it rotates (in the three dimensions) by θ around the z axis.

A point $h(s,\theta)$ of the Möbius strip corresponds to the point

$$\Big(\sqrt{1 - \tfrac{1}{4}s^2}\,\cos\tfrac{1}{2}\theta,\ \sqrt{1 - \tfrac{1}{4}s^2}\,\sin\tfrac{1}{2}\theta,\ \tfrac{1}{2}s\Big)$$

on the sphere S, and the corresponding point of M. (Think, what happens for $\theta = 2\pi$.)

The rest of M is homeomorphic to a disk (not two disks), and this disk is glued to the Möbius strip in a way unthinkable in three dimensions.[3]

[1] It is easy to check that, locally, every matrix that satisfies these equations belongs to M.

[2] Images from Wikipedia, “Möbius strip”.

[3] Dimension 6 can be reduced to dimension 4 by taking only $(a^2 - b^2, ab, ac, bc)$, see “Real projective plane” in Wikipedia.

Index