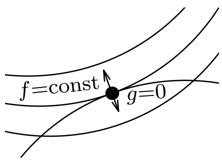
3 Open mappings and constrained optimization

A first order necessary condition ("Lagrange multipliers") for constrained extrema is proved via open mappings and used for optimization.

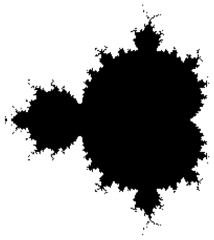
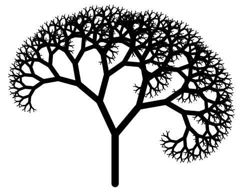
3a What is the problem

As was noted in Sect. 2c, local extrema of a differentiable function f can be found using the necessary condition $(Df)_x = 0$, which is important for optimization. Now we turn to a harder task: to maximize $f(x,y)$ subject to a constraint $g(x,y) = 0$; in other words, to maximize f on the set $Z_g = \{(x,y) : g(x,y) = 0\}$. Here $f, g : \mathbb{R}^2 \to \mathbb{R}$ are given differentiable functions (the *objective function* and the *constraint function*).



It is easy to guess a necessary condition: ∇f and ∇g must be collinear. [Sh:Sect.5.4] It is easy to prove this guess *taking for granted* that Z_g, being a curve, can be parametrized by a differentiable path γ, that is, $g(x,y) = 0 \iff \exists t \; (x,y) = \gamma(t)$. Is it really the general case?

Rather unexpectedly, *every* closed subset of $\mathbb{R}^2$ is Z_g for some $g \in \mathbb{C}^1(\mathbb{R}^2)$. (The proof is beyond this course.)

A simple example: $g(x,y) = x^2 - y^2$; $g \in \mathbb{C}^1(\mathbb{R}^2)$; Z_g is the union of two straight lines intersecting at the origin. Note that $\nabla g = 0$ at the origin.

Another example:

$$g(x,y) = \begin{cases} x^2 + y^2 & \text{for } x \le 0, \\ y^2 & \text{for } x \ge 0. \end{cases}$$

Again, $g \in \mathbb{C}^1(\mathbb{R}^2)$ (think, why); $Z_g = [0,\infty) \times \{0\}$, a ray from the origin. Again, $\nabla g = 0$ at the origin. The function $f : (x,y) \mapsto x$ reaches its minimum on Z_g at the origin. Can we say that ∇f and ∇g are collinear at the origin? Rather, they are linearly dependent.

We assume that $\nabla f(x_0,y_0)$ and $\nabla g(x_0,y_0)$ are linearly independent, $g(x_0,y_0) = 0$, and want to prove that (x_0,y_0) cannot be a local constrained[1] extremum[2] of f on Z_g. Assume for simplicity $x_0 = y_0 = 0$ and $f(0,0) = 0$. Consider the mapping $h : \mathbb{R}^2 \to \mathbb{R}^2$, $h(x,y) = \big(f(x,y), g(x,y) \big)$ near the origin, and its linear approximation $T = (Dh)_{(0,0)} : \mathbb{R}^2 \to \mathbb{R}^2$; $T(x,y) = (ax+by, cx+dy)$ where $a = (D_1 f)_{(0,0)}$, $b = (D_2 f)_{(0,0)}$, $c = (D_1 g)_{(0,0)}$, $d = (D_2 g)_{(0,0)}$. Vectors $\nabla f(0,0) = (a,b)$ and $\nabla g(0,0) = (c,d)$ are linearly independent, thus $\left|\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right| \ne 0$, which means that T is invertible. (Alternatively, use Lemma 2f3.)

It follows that $T(x_1,y_1) = (1,0)$ for some x_1, y_1. We have

$$f(tx_1, ty_1) = t + o(t) \,, \quad g(tx_1, ty_1) = o(t) \,.$$

Does it show that the origin cannot be a local constrained extremum of f on Z_g? No, it does not. We still did not find x_t, y_t such that

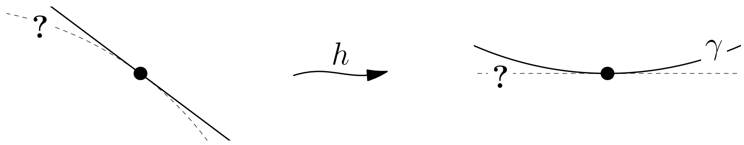
$$f(x_t, y_t) = t + o(t) \,, \quad g(x_t, y_t) = 0 \,.$$

In other words: we know that the image $V = h(U)$ of a neighborhood U of the origin contains a differentiable path $\gamma : (-\varepsilon, \varepsilon) \to \mathbb{R}^2$ such that $\gamma(0) = (0,0)$

[1] In other words, conditional.

[2] Not necessarily strict; that is, either $f(x_0,y_0) \le f(x,y)$ for all (x,y) near (x_0,y_0) (minimum), or "$\ge$" (maximum).

and $\gamma'(0) = (1, 0)$, but we still do not know, whether V contains $(-\varepsilon, \varepsilon) \times \{0\}$ or not.



We know that T is onto, but we still do not know, whether h is locally onto. In more technical language: whether h is an open mapping, as defined below.

Of course, we need a multidimensional theory; $\mathbb{R}^2$ is only the simplest case.

3b Open mappings

3b1 Definition. A mapping $f : \mathbb{R}^n \to \mathbb{R}^m$ is *open* if $f(U)$ is open for every open $U \subset \mathbb{R}^n$.

3b2 Remark. For $m < n$, the usual embedding $f : \mathbb{R}^m \to \mathbb{R}^n$ is a homeomorphism $\mathbb{R}^m \to f(\mathbb{R}^m) \subset \mathbb{R}^n$, but not an open mapping. On the other hand, the usual projection $g : \mathbb{R}^n \to \mathbb{R}^m$ is open, but not one-to-one.

Similarly, for an open $U \subset \mathbb{R}^n$ a mapping $f : \mathrm{U} \to \mathbb{R}^m$ is called *open* if $f(U_1)$ is open for every open $U_1 \subset U$.[1]

3b3 Exercise. Let $U, V \subset \mathbb{R}^n$ be open and $f : U \to V$ a homeomorphism; prove that f is open. Does it hold if V is not assumed to be open?

3b4 Exercise. Prove or disprove: a continuous function $\mathbb{R} \to \mathbb{R}$ is open if and only if it is strictly monotone.

3b5 Exercise. Let $U \subset \mathbb{R}^n$ be open, and $f : U \to \mathbb{R}^m$. Consider all open $U_1 \subset U$ such that $f|_{U_1}$ is open. If these U_1 cover U then f is open.

Prove it.

3b6 Exercise. A mapping $f : U \to \mathbb{R}^m$ is open if and only if for every $x \in U$ and every neighborhood $U_1 \subset U$ of x the set $f(U_1)$ is a neighborhood of $f(x)$.

Prove it.

Reminder: a neighborhood need not be open.

3b7 Proposition. Let $U \subset \mathbb{R}^n$ be open, and $f \in C^1(U \to \mathbb{R}^n)$. If the operator $(Df)_x$ is invertible for all $x \in U$ then f is open.

[1] If you know that every subset of $\mathbb{R}^n$ is itself a topological space, you probably know the notion of an open mapping $X \to Y$ for given $X \subset \mathbb{R}^n$, $Y \subset \mathbb{R}^m$. Then you may recall 1b17 and think, whether the continuous bijection $f : \mathbb{R}^2 \to B$ is open, or not.

3b8 Lemma. Let $U \subset \mathbb{R}^n$ be open and bounded, $f : \overline{U} \to \mathbb{R}^n$ a continuous mapping, differentiable on U. If f is a homeomorphism[1] $\overline{U} \to f(\overline{U})$ and the operator $(Df)_x$ is invertible for all $x \in U$ then $f|_U$ is open. (Here $\overline{U}$ is the closure of U.)

3b9 Proposition. Assume that $x_0 \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}^n$ is differentiable near x_0, Df is continuous at x_0, and the operator $(Df)_{x_0}$ is invertible. Then there exists a bounded open neighborhood U of x_0 such that $f|_{\overline{U}}$ is a homeomorphism $\overline{U} \to f(\overline{U})$, and f is differentiable on U, and the operator $(Df)_x$ is invertible for all $x \in U$.

Clearly, Prop. 3b7 follows from 3b8 (to be proved in Sect. 3d) and 3b9(to be proved in Sect. 3e) via 3b5.

In fact, for every open $U \subset \mathbb{R}^n$, every continuous one-to-one mapping $U \to \mathbb{R}^n$ is open (and therefore a homeomorphism $U \to f(U)$). This is a well-known topological result, "the Brouwer invariance of domain theorem". Then, why Lemma 3b8?[2] For two reasons.

First, invariance of domain is proved using *algebraic* topology (the Brouwer fixed point theorem). Lemma 3b8 is much simpler to prove, due to differentiability assumption.

Second, in this course we improve our understanding of *differentiable* mappings. Continuous mappings in general are a different story.

3b10 Exercise. Prove invariance of domain in dimension one.[3]

3b11 Exercise. Taking Prop. 3b7 for granted, prove the more general claim:

Let $U \subset \mathbb{R}^n$ be open, and $f \in C^1(U \to \mathbb{R}^m)$. If the operator $(Df)_x$ maps $\mathbb{R}^n$ *onto* $\mathbb{R}^m$ for all $x \in U$ then f is open.[4]

If the linear approximation is onto then the nonlinear mapping is locally onto.

3b12 Exercise. Taking Prop. 3b9 for granted, prove the following claim:

Assume that $x_0 \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable near x_0, Df is continuous at x_0, and the operator $(Df)_{x_0}$ is one-to-one. Then there exists a bounded open neighborhood U of x_0 such that $f|_{\overline{U}}$ is a homeomorphism $\overline{U} \to f(\overline{U})$.[5]

[1] That is, f is continuous and one-to-one, and $f^{-1} : f(\overline{U}) \to \overline{U}$ is also continuous.

[2] Still another alternative to Lemma 3b8 will be discussed in Sect. 4d.

[3] Hint: recall 3b4.

[4] Hint: the operator maps some m-dimensional subspace of $\mathbb{R}^n$ onto $\mathbb{R}^m$. That is, $(Df)_{x_0} \circ T$ is onto for some linear $T : \mathbb{R}^m \to \mathbb{R}^n$.

[5] Hint: the operator $T \circ (Df)_{x_0}$ is one-to-one for some linear $T : \mathbb{R}^m \to \mathbb{R}^n$.

If the linear approximation is one-to-one then the nonlinear mapping is locally one-to-one.

If the linear approximation is bijective then the nonlinear mapping is locally bijective.

3c From differentiability to geometry

Assume that U is an open neighborhood of x_0, $f : U \to \mathbb{R}^n$ is differentiable at x_0, $(Df)_{x_0}$ is an invertible operator, but $f(U)$ is not a neighborhood of $y_0 = f(x_0)$.

For every $h \in \mathbb{R}^n$ the path

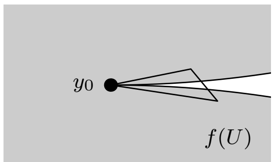
$$\gamma_h(t) = f(x_0 + th) \quad \text{for } t \in (-\varepsilon, \varepsilon)$$

is well-defined (for some ε), differentiable, contained in $f(U)$, and

$$\gamma_h(0) = y_0 \,, \quad \gamma_h'(0) = D_h f(x_0) = T(h)$$

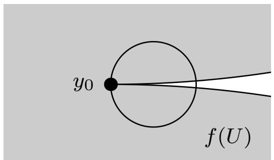
where $T = (Df)_{x_0}$. Due to invertibility of T, every vector is of the form $T(h)$.

3c1 Exercise.
(a) For $n = 2$ prove that $f(U)$ intersects every open triangle with one vertex at y_0.
(b) What about a generalization to $n > 2$?



We really need much less than 3c1(b).

3c2 Exercise.
Prove that $f(U)$ intersects every open ball whose boundary contains y_0.[1]



Let us call $y_0 \in f(U)$ a *regular boundary point*,[2] if $f(U)$ does not intersect some open ball whose boundary contains y_0. We conclude.

3c3 Lemma. If U is an open neighborhood of x_0, $f : U \to \mathbb{R}^n$ is differentiable at x_0, $(Df)_{x_0}$ is an invertible operator, then $y_0 = f(x_0)$ cannot be a regular boundary point.

Irregular boundary points are still a challenge.

[1]Hint: either use the path γ_h for an appropriate h, or alternatively, differentiate a function $|f(\cdot) - a|^2$ where a is the center of the ball.

[2]Not a standard terminology; introduced for convenience, to be used within sections 3c, 3d only.

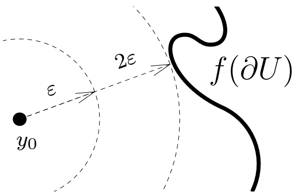
3d From topology to geometry

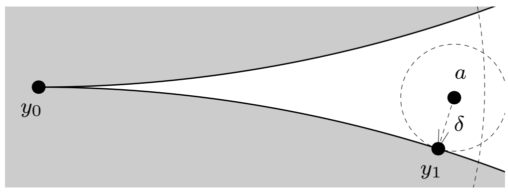
3d1 Lemma. Let $U \subset \mathbb{R}^n$ be open and bounded, $f : \overline{U} \to \mathbb{R}^n$ continuous. If f is a homeomorphism $\overline{U} \to f(\overline{U})$ with no regular boundary points (of $f(U)$), then $f(U)$ is open.

Proof. [Sh:p.198–199] Let $y_0 \in f(U)$; we have to prove that $f(U)$ is a neighborhood of y_0.

We have $\overline{U} = U \uplus \partial U$, therefore $f(\overline{U}) = f(U) \uplus f(\partial U)$ (since f is one-to-one). Sets $\overline{U}$ and ∂U are compact, therefore $f(\overline{U})$ and $f(\partial U)$ are compact (since f is continuous).

The distance $\operatorname{dist}(y_0, f(\partial U)) = \inf_{z \in \partial U} |y_0 - z|$ is positive (by compactness); denote it by 2ε. The open ε-neighborhood of y_0 does not intersect $f(\partial U)$. It is sufficient to prove that it is contained in $f(U)$.





Assuming the contrary we take $a \in \mathbb{R}^n$ such that $|a - y_0| < \varepsilon$ but $a \notin f(U)$. Observe that $a \notin f(\partial U)$ and therefore $a \notin f(\overline{U})$. By compactness there exists $y_1 \in f(\overline{U})$ such that $|a - y_1| = \operatorname{dist}(a, f(\overline{U})) > 0$; we denote this distance by δ and note that $\delta < \varepsilon$ (since $y_0 \in f(\overline{U})$). It follows that $y_1 \notin f(\partial U)$ (since $|y_0 - y_1| \le |y_0 - a| + |a - y_1| < \varepsilon + \delta < 2\varepsilon$) and therefore $y_1 \in f(U)$. The open δ-neighborhood of a does not intersect $f(\overline{U})$; thus y_1 is a regular boundary point. Contradiction. $\square$

In combination with 3c3 it proves Lemma 3b8.

For proving 3b7 it remains to prove 3b9.

3e Local homeomorphism

We generalize Prop. 3b9 as follows.

3e1 Proposition. Assume that S_1, S_2 are n-dimensional affine spaces, $x_0 \in S_1$, $f : S_1 \to S_2$ is differentiable near x_0, Df is continuous at x_0, and the operator $(Df)_{x_0} : \vec{S}_1 \to \vec{S}_2$ is invertible. Then there exists a bounded open neighborhood U of x_0 such that $f|_{\overline{U}}$ is a homeomorphism $\overline{U} \to f(\overline{U})$, and f is differentiable on U, and the operator $(Df)_x$ is invertible for all $x \in U$.

Maybe you think that the more general 3e1 is harder to prove than 3b9. No, it is easier to prove! Recall Sect. 1c: irrelevant structure is a nuisance. Dealing with two affine spaces we may upgrade them to vector spaces such that $x_0 = 0$ and $f(x_0) = 0$. (We could not do it in a single space unless $f(x_0) = x_0$.) More importantly, dealing with two vector spaces we can diagonalize an arbitrary linear operator… Recall Sect. 1c (again): two bases give more freedom than one basis.

We have two vector spaces $\vec{S}_1, \vec{S}_2$ and an invertible linear operator $(Df)_{x_0} = T : \vec{S}_1 \to \vec{S}_2$. We choose a basis $(e_1, \dots, e_n)$ of $\vec{S}_1$ and the *corresponding* basis $(Te_1, \dots, Te_n)$ of $\vec{S}_2$. Then the matrix of T becomes the unit matrix! Accordingly, T turns into the identity operator $\mathrm{id} : \mathbb{R}^n \to \mathbb{R}^n$. Having $T = \mathrm{id}$ we may forget the coordinates, downgrading $\mathbb{R}^n$ to vector space, and reformulate Prop. 3e1 as follows.

3e2 Proposition. Assume that V is a vector ^{fd}space, $f : V \to V$ is differentiable near 0, Df is continuous at 0, $f(0) = 0$, and $(Df)_0$ is the identity operator $V \to V$. Then there exists a bounded open neighborhood U of 0 such that $f|_{\overline{U}}$ is a homeomorphism $\overline{U} \to f(\overline{U})$, and f is differentiable on U, and the operator $(Df)_x$ is invertible for all $x \in U$.

3e3 Exercise. Generalize 2f1: $|f(b) - f(a)| \le M|b - a|$ for $f : \mathbb{R}^n \to \mathbb{R}^m$ provided that $\|Df\| \le M$ on $[a, b]$.[1]

Proof. We have $(Df)_x \to (Df)_0 = \mathrm{id}$ as $x \to 0$. We upgrade V to a Euclidean space (arbitrarily) and use the operator norm (recall Sect. 1e):

$$\|(Df)_x - \mathrm{id}\| \to 0 \quad \text{as } x \to 0 \, .$$

For every $\varepsilon > 0$ there exists a neighborhood U_ε of 0 such that f is differentiable on U_ε, and

$$\|(Df)_x - \mathrm{id}\| \le \varepsilon \quad \text{for all } x \in U_\varepsilon \, .$$

We choose U_ε to be convex (just a ball, if you like) and apply 2f1 to the mapping $f - \mathrm{id}$ (its derivative being $Df - \mathrm{id}$): $|(f - \mathrm{id})(x) - (f - \mathrm{id})(y)| \le \varepsilon|x - y|$, that is,

$$|(f(x) - f(y)) - (x - y)| \le \varepsilon|x - y| \quad \text{for all } x, y \in U_\varepsilon \, .$$

It follows (assuming $\varepsilon < 1$) that $f(x) - f(y) \ne 0$ for $x - y \ne 0$; that is, $f|_{U_\varepsilon}$ is one-to-one. Moreover, the triangle inequality gives

$$(1 - \varepsilon)|x - y| \le |f(x) - f(y)| \le (1 + \varepsilon)|x - y|$$

[1] Hint: $\mathbb{R}^n \xrightarrow{f} \mathbb{R}^m \xrightarrow{g} \mathbb{R}$, $g(y) = \langle u, y \rangle$, $|u| = 1$; apply 2f1 to $g \circ f$; optimize in u.

for all $x, y \in U_\varepsilon$. Thus, $f|_{U_\varepsilon}$ is a homeomorphism $U_\varepsilon \to f(U_\varepsilon)$.

Finally, $\big|\big((Df)_x - \mathrm{id}\big)(h)\big| \le \varepsilon|h|$, that is,

$$|(Df)_x(h) - h| \le \varepsilon|h| \quad \text{for all } x \in U_\varepsilon\,, h \in V\,;$$

the triangle inequality (again) gives

$$(1-\varepsilon)|h| \le |(Df)_x(h)| \le (1+\varepsilon)|h|\,,$$

which shows that the operator $(Df)_x$ is one-to-one, therefore invertible. □

Thus, 3e1, 3b9, and finally 3b7 are proved.

3e4 Exercise. Consider the set $U \subset \mathbb{R}^n$ of all $(a_0, \dots, a_{n-1})$ such that the polynomial

$$t \mapsto t^n + a_{n-1}t^{n-1} + \dots + a_0$$

has n pairwise distinct real roots.

(a) Prove that U is open.

(b) Define $\psi : U \to \mathbb{R}^n$ by $\psi(a_0, \dots, a_{n-1}) = (t_1, \dots, t_n)$ where $t_1 < \dots < t_n$ are the roots of the polynomial. Prove that ψ is a homeomorphism $U \to V$ where $V = \big\{(t_1, \dots, t_n) : t_1 < \dots < t_n\big\}$.[1]

3f Curves

We return to the problem discussed in Sect. 3a.

3f1 Proposition. Assume that $f, g : \mathbb{R}^2 \to \mathbb{R}$ are continuously differentiable near a given point (x_0, y_0); vectors $\nabla f(x_0, y_0)$ and $\nabla g(x_0, y_0)$ are linearly independent; and $g(x_0, y_0) = 0$. Denote $z_0 = f(x_0, y_0)$. Then there exist $\varepsilon > 0$ and a path $\gamma : (z_0 - \varepsilon, z_0 + \varepsilon) \to \mathbb{R}^2$ such that $\gamma(z_0) = (x_0, y_0)$, $f(\gamma(t)) = t$ and $g(\gamma(t)) = 0$ for all $t \in (z_0 - \varepsilon, z_0 + \varepsilon)$.

Proof. The mapping $h : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $h(x, y) = \big(f(x, y), g(x, y)\big)$ is continuously differentiable near (x_0, y_0), and $(Dh)_{(x_0,y_0)}$ is invertible by 2f3. Lemma 3b8 and Prop. 3b9 provide a neighborhood U of (x_0, y_0) such that $V = f(U)$ is a neighborhood of $(z_0, 0)$ and $h|_U$ is a homeomorphism $U \to V$. We take $\varepsilon > 0$ such that $(t, 0) \in V$ for all $t \in (z_0 - \varepsilon, z_0 + \varepsilon)$ and define γ by

$$\gamma(t) = (h|_U)^{-1}(t, 0)\,.$$

Clearly γ is continuous, $\gamma(0) = (x_0, y_0)$, $\gamma(t) \in U$ and $h(\gamma(t)) = (t, 0)$, that is, $f(\gamma(t)) = t$ and $g(\gamma(t)) = 0$. □

[1] Hint: use 2e9(b).

3f2 Corollary. If f, g, x_0, y_0 are as in 3f1 then (x_0, y_0) cannot be a local constrained extremum of f on Z_g.

3f3 Remark. (a) Prop. 3f1 does not claim differentiability of the path γ (but only its continuity).

(b) Prop. 3f1 does not claim that γ covers *all* points of Z_g near (x_0, y_0). Moreover, the set $U \cap Z_g$ need not be connected.

We'll return to these points later (in 4c10).

The next case is, dimension three. We guess that a single constraint $g(x, y, z) = 0$ leads to a surface Z_g, not a curve; a curve is rather $Z_{g_1,g_2} = Z_{g_1} \cap Z_{g_2}$.

3f4 Proposition. Assume that $f, g_1, g_2 : \mathbb{R}^3 \to \mathbb{R}$ are continuously differentiable near a given point (x_0, y_0, z_0); vectors $\nabla f(x_0, y_0, z_0)$, $\nabla g_1(x_0, y_0, z_0)$ and $\nabla g_2(x_0, y_0, z_0)$ are linearly independent; and $g_1(x_0, y_0, z_0) = g_2(x_0, y_0, z_0) = 0$. Denote $w_0 = f(x_0, y_0, z_0)$. Then there exist $\varepsilon > 0$ and a path $\gamma : (w_0 - \varepsilon, w_0 + \varepsilon) \to \mathbb{R}^3$ such that $\gamma(w_0) = (x_0, y_0, z_0)$, $f(\gamma(t)) = t$ and $g_1(\gamma(t)) = g_2(\gamma(t)) = 0$ for all $t \in (w_0 - \varepsilon, w_0 + \varepsilon)$.

3f5 Exercise. Prove Prop. 3f4.[1]

3f6 Corollary. If $f, g_1, g_2, x_0, y_0, z_0$ are as in 3f4 then (x_0, y_0, z_0) cannot be a local constrained extremum of f on Z_{g_1,g_2}.

3f7 Exercise. Generalize 3f4 and 3f6 to $f, g_1, \dots, g_{n-1} : \mathbb{R}^n \to \mathbb{R}$.

3g Surfaces

We turn to a single constraint $g(x, y, z) = 0$ in $\mathbb{R}^3$, and a function $f : \mathbb{R}^3 \to \mathbb{R}$. How to proceed? The mapping $(x, y, z) \mapsto \big(f(x, y, z), g(x, y, z)\big)$ from $\mathbb{R}^3$ to $\mathbb{R}^2$ surely is not expected to be a local homeomorphism. However, we may add another constraint, getting a curve on the surface!

3g1 Proposition. Assume that $f, g : \mathbb{R}^3 \to \mathbb{R}$ are continuously differentiable near a given point (x_0, y_0, z_0); vectors $\nabla f(x_0, y_0, z_0)$ and $\nabla g(x_0, y_0, z_0)$ are linearly independent; and $g(x_0, y_0, z_0) = 0$. Denote $w_0 = f(x_0, y_0, z_0)$. Then there exist $\varepsilon > 0$ and a path $\gamma : (w_0 - \varepsilon, w_0 + \varepsilon) \to \mathbb{R}^3$ such that $\gamma(w_0) = (x_0, y_0, z_0)$, $f(\gamma(t)) = t$ and $g(\gamma(t)) = 0$ for all $t \in (w_0 - \varepsilon, w_0 + \varepsilon)$.

[1] Hint: similar to the proof of 3f1; $h(x, y, z) = ((f(x, y, z), g_1(x, y, z), g_2(x, y, z)), \dots$

Proof. We choose a vector $a \in \mathbb{R}^3$ such that the three vectors a, $\nabla f(x_0, y_0, z_0)$ and $\nabla g(x_0, y_0, z_0)$ are linearly independent. We choose a function $g_2 : \mathbb{R}^3 \to \mathbb{R}$, continuously differentiable near (x_0, y_0, z_0), such that $\nabla g_2(x_0, y_0, z_0) = a$ (for example, the linear function $g_2(\cdot) = \langle \cdot, a \rangle$). It remains to apply Prop. 3f4 to f, g, g_2. □

3g2 Corollary. If f, g, x_0, y_0, z_0 are as in 3g1 then (x_0, y_0, z_0) cannot be a local constrained extremum of f on Z_g.

3g3 Exercise. Generalize 3g1 and 3g2 to $f, g_1, \dots, g_m : \mathbb{R}^n \to \mathbb{R}$, $1 \le m \le n - 1$.

3h Lagrange multipliers

3h1 Theorem. Assume that $x_0 \in \mathbb{R}^n$, 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 vectors $\nabla g_1(x_0), \dots, \nabla g_m(x_0)$ are linearly independent. If x_0 is a local constrained extremum 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

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

This is a reformulation of the generalization meant in 3g3.

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 $Z = Z_{g_1,\dots,g_m}$ 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 3h1; 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 at x.

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

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.

Theorem 3h1 generalizes readily from $\mathbb{R}^n$ to an n-dimensional Euclidean affine space. But if no Euclidean norm is given on the affine space then the gradient is not defined. However, the gradient vector $\nabla f(x_0)$ is rather a substitute of the linear function $(Df)_{x_0}$, namely, $(Df)_{x_0} : h \mapsto \langle \nabla f(x_0), h \rangle$ (recall Sect. 2f). Thus, the relation $\nabla f(x_0) = \lambda_1 \nabla g_1(x_0) + \dots + \lambda_m \nabla g_m(x_0)$ between vectors may be replaced with a relation

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

between linear functions. And linear independence of vectors $\nabla g_1(x_0), \dots, \nabla g_m(x_0)$ may be replaced with linear independence of linear functions $(Dg_1)_{x_0}, \dots, (Dg_m)_{x_0}$; or, due to Lemma 2f3, we may say instead that $(Dg)_{x_0}$ maps $\mathbb{R}^n$ onto $\mathbb{R}^m$. Now it is clear how to generalize Th. 3h1 from $\mathbb{R}^n$ to an affine ^{fd}space.

3i Examples

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).

We introduce nine 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 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. 3h 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 zum!) 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 $z_1 = z_2 = z_3 = \frac{z_1+z_2+z_3}{3-3\alpha}$, therefore $z_1 = z_2 = z_3 = 0$ (since $\alpha \neq 1$); P, Q, R are situated on the horizontal plane $\{(x, y, z) : z = 0\}$.

Isoperimetry for triangles

Denote by A the area, and by L the length. The Dido isoperimetric inequality says that, for any plane figure G,

$$\mathrm{A}(G) \leq \frac{\mathrm{L}(\partial G)^2}{4\pi}\,,$$

and equality is attained for discs only. For triangles, the estimate can be improved:

$$\mathrm{A}(\Delta) \leq \frac{\mathrm{L}(\partial\Delta)^2}{12\sqrt{3}},$$

for any plane triangle Δ, and the equality sign attains for the equilateral triangles and only for them.

In other words, among all triangles with the given perimeter, the equilateral one has the largest area.

Proof. We use the Heron formula that relates the area A and the perimeter $L = x + y + z$:

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

Set $L = 2s$. Then we need to maximize the function

$$f(x, y, z) = s(s - x)(s - y)(s - z)$$

under condition

$$g(x, y, z) = x + y + z - 2s = 0\,.$$

Of course, we have additional restrictions

$$x, y, z > 0, \qquad x + y > z, \quad x + z > y, \quad y + z > x\,,$$

which define the domain U in the space (x, y, z). (Draw this domain!) On the boundary of this domain (when the inequalities turn to the equations), the function f identically vanishes. Thus, f attains its maximal value inside U and we can use the Lagrange multipliers.

The Lagrange equations are

$$\begin{cases} -s(s-y)(s-z) = \lambda \\ -s(s-x)(s-z) = \lambda \\ -s(s-x)(s-y) = \lambda \\ x+y+z = 2s \end{cases}$$

The first three equations give us

$$(s-y)(s-z) = (s-x)(s-z) = (s-x)(s-y)\,,$$

whence

$$x = y = z = \frac{2}{3}s\,,$$

and $A^2 = s \cdot (s/3)^3$. The result follows. $\square$

Extrema of quadratic forms

We are looking for the maximal and minimal values of the quadratic form

$$f(x) = \sum_{i,j=1}^{n} a_{ij} x_i x_j \qquad a_{ij} = a_{ji}\,,$$

on the unit sphere

$$\sum_{i=1}^{n} x_i^2 = 1\,.$$

In this case, $f(x) = \langle Ax, x \rangle$, where A is a symmetric linear operator with the matrix (a_{ij}). Thus $\nabla f(x) = 2Ax$ (recall 2b11(a)). Furthermore, $g(x) = |x|^2 - 1$, and $\nabla g(x) = 2x$. Therefore, the Lagrange equations take the form

$$\begin{cases} 2Ax = 2\lambda x \\ (x, x) = 1\,. \end{cases}$$

Hence, λ is the eigenvalue of A, and the maximum of the form is the largest eigenvalue, the minimum of the form is the smallest eigenvalue.

The operator norm.

As a corollary, we compute the (operator) norm of a linear operator $A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$. By definition,

$$\|A\| = \max_{|x|=1} |Ax| .$$

Thus, we need to maximize the function $f(x) = |Ax|^2 = (Ax, Ax)$ under additional condition $|x|^2 = 1$.

Observe that $f(x) = (Ax, Ax) = (A^*Ax, x)$. Hence, by the previous paragraph, $\|A\|^2$ equals the maximal eigenvalue of the symmetric matrix A^*A.

3i1 Exercise. Let A be an invertible linear operator. Find $\|A^{-1}\|$.

The Hölder inequality

Let $1 < p < \infty$. Then

$$\Big| \sum x_i y_i \Big| \le \Big(\sum |x_i|^p \Big)^{1/p} \Big(\sum |y_i|^q \Big)^{1/q} , \tag{3i2}$$

where q is 'the dual exponent' to p: $\frac{1}{p} + \frac{1}{q} = 1$.

Proof. We assume that all x_i's and y_i's are non-negative. Since Hölder's inequality is homogeneous with respect to multiplication of all x_i by the same positive number, we assume that $\sum x_i^p = 1$. Given $y \in \mathbb{R}^n$ with non-negative coordinates, define the function $f(x) = \sum x_i y_i$. That is, for a compact set $K = \{x \in \mathbb{R}^n : x_i \ge 0, \ \sum x_i^p = 1\}$, we want to prove that $\max_K f \le (\sum y_i^q)^{1/q}$. We use induction with respect to the number n of variables. For $n = 1$, we have $K = \{1\}$, and there is nothing to prove.

For an arbitrary $n \ge 2$, we look at the extremum of f on K. The Lagrange multipliers technique can be applied only on the set $K_0 = \{x \in \mathbb{R}^n : x_i > 0, \ \sum x_i^p = 1\}$ (why?). However, the rest $K \setminus K_0$ consists of x such that at least one of the coordinates x_i vanishes. Hence, by the assumption of the induction $\max_{K \setminus K_0} f \le (\sum y_i^q)^{1/q}$. Now, using the Lagrange method, we shall find that the conditional extremum of f under assumptions $g(x) = \sum x_i^p - 1 = 0$ *equals* $(\sum y_i^q)^{1/q}$. This will prove Hölder's inequality (and also will show that it cannot be improved).

The Lagrange equations have the form

$$y_i = \lambda p x_i^{p-1} , \qquad 1 \le i \le n ,$$
$$\sum x_i^p = 1 .$$

We have: $x_i = cy_i^{\frac{1}{p-1}}$; $1 = \sum x_i^p = c^p \sum y_i^{\frac{p}{p-1}} = c^p \sum y_i^q$; $c = \frac{1}{(\sum y_i^q)^{1/p}}$; $\sum x_i y_i = c \sum y_i^{\frac{1}{p-1}+1} = c \sum y_i^{\frac{p}{p-1}} = c \sum y_i^q = (\sum y_i^q)^{1-\frac{1}{p}} = (\sum y_i^q)^{1/q}$ (recall that $\frac{p}{p-1} = q$). $\Box$

We proved Hölder's inequality in the case of finitely many variables x_i and y_i. It persists in the case of countable many variables x_i and y_i. In this case, it means that if two series $\sum |x_i|^p$ and $\sum |y_i|^q$ converge (and q is dual to p), then the series $\sum x_i y_i$ also converges and inequality (3i2) holds.

3i3 Exercise. Prove that, for $x_i > 0$,

$$\frac{n}{\frac{1}{x_1} + \dots + \frac{1}{x_n}} \le \sqrt[n]{x_1 x_2 \dots x_n} \le \frac{x_1 + x_2 + \dots + x_n}{n} \, .$$

The equality sign attains only in the case when all x_i's are equal.[1]

3i4 Exercise. Find the maximum of the function $f(x, y, z) = x^a y^b z^c$ $(a, b, c > 0)$, where x, y and z are positive, and $x^k + y^k + z^k = 1$ $(k > 0)$.

3i5 Exercise. Find the maximum of y over all points $(x, y) \in \mathbb{R}^2$ that satisfy the equation $x^2 + xy + y^2 = 27$.

[Sh:Sect.5.4]

3j Sensitivity of optimum to parameters

When using a mathematical model one often bothers about sensitivity[2] of the result (the output of the model) to the assumptions (the input). Here is one of such questions.[3]

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

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

[1]Hint: to get the first inequality, minimize $x_1 \cdot x_2 \cdot \ldots \cdot x_n$ under assumption that all x_i's are positive and $\sum_i x_i^{-1} = 1$. To get the second inequality, maximize $x_1 x_2 \dots x_n$ under assumption that all x_i are positive and $\sum_i x_i = 1$.

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

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

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,

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

and

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

(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

$$\sup_{Z_c} f = \sup_{Z_0} f + \lambda_1(0)c_1 + \dots + \lambda_m(0)c_m + o(|c|) \tag{3j1}$$

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 (3j1) holds.[1]

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

Index