3 Open mappings
and constrained optimization

3a What is the problem
3b Open mappings
3c Linear and nonlinear
3d Curves
3e Surfaces
3f Lagrange multipliers
3g Example: arithmetic, geometric, harmonic, and more general means
3h Example: Three points on a spheroid
3i Example: Singular value decomposition
3j Sensitivity of optimum to parameters

Continuously differentiable mappings behave locally like linear, which is easy to guess but not easy to prove. A first order necessary condition (“Lagrange multipliers”) for constrained extrema is proved and used for optimization.

3a What is the problem

By (2c3), 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: $\nabla f$ and $\nabla g$ 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 $\gamma$, 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 $g \in C^1(\mathbb{R}^2)$. (The proof is beyond this course.)

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

Another example:

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

Again, $g \in C^1(\mathbb{R}^2)$ (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 $\nabla f$ and $\nabla g$ 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 extremum of $f$ on $Z_g$. Assume for simplicity $x_0 = y_0 = 0$ and $f(0, 0) = 0$. Consider the mapping $h : \mathbb{R}^2 \to \mathbb{R}^2$, $h(x, y) = (f(x, y), g(x, y))$ 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 $|c/a - d/b| \neq 0$, which means that $T$ is invertible. (Alternatively, use Lemma 2f2.)

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 \( \gamma(0) = (0, 0) \) 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.** Let \( X, Y \) be metrizable spaces. A mapping \( f : X \to Y \) is \textit{open} if \( f(U) \subset Y \) is open for every open \( U \subset X \).

This is a local notion, due to an equivalent definition \textit{3b2}.

**3b2 Definition.** (equivalent to \textit{3b1})

Let \( X, Y \) be metrizable spaces. A mapping \( f : X \to Y \) is \textit{open} if for every \( x \in X \) and every neighborhood \( U \) of \( x \) the set \( f(U) \) is a neighborhood of \( f(x) \).

Reminder: a neighborhood need not be open.

**3b3 Exercise.** Prove equivalence of these two definitions.

A bijection \( f : X \to Y \) is open if and only if \( f^{-1} : Y \to X \) is continuous.

Thus, a continuous bijection is open if and only if it is a homeomorphism.

By 1a14, every continuous bijection \( \mathbb{R} \to \mathbb{R} \) is open (hence, homeomorphism). But generally (for \( X \to Y \)) it is not; recall 1a15–1a17.

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

The usual projection \( g : \mathbb{R}^{n+1} \to \mathbb{R}^n \) is continuous and open, but not one-to-one.

The usual embedding \( f : \mathbb{R}^n \to \mathbb{R}^{n+1} \) (or \( \mathbb{R}^{n+k} \)) is a homeomorphism \( \mathbb{R}^n \to f(\mathbb{R}^n) \subset \mathbb{R}^{n+1} \), but not an open mapping. If \( U \subset \mathbb{R}^n \) is open then \( f(U) \) is relatively open in \( f(\mathbb{R}^n) \), but not open in \( \mathbb{R}^{n+1} \) (unless \( U = \emptyset \)). In this
case \( f(U) = \overline{f(U)} \), but \( f(\partial U) \neq \partial(f(U)) \) since \( \partial(f(U)) = \overline{f(U)} \setminus f(U)^o = f(\overline{U}) \setminus \emptyset = f(\overline{U}) \). Rather, \( f(\partial U) \) is the relative boundary of \( U \) in \( f(\mathbb{R}^n) \).

Let \( X \) be a metrizable space and \( A \subset X \). Every subset \( U \subset A \) open in \( X \) is relatively open in \( A \) (recall 1c3).

3b5 Exercise. A set \( A \) in a metrizable space \( X \) is open if and only if every relatively open subset of \( A \) is open in \( X \).

Prove it.

3b6 Exercise. Let \( X, Y \) be metrizable spaces, \( U \subset X \), \( V \subset Y \), \( f : U \to V \) a homeomorphism, and \( U \) is open. Than \( f \) is open if and only if \( V \) is open.

Prove it.

Let \( U \subset \mathbb{R}^n \) be relatively open in its closure \( \overline{U} \). As we know, \( U \) need not be open (in \( \mathbb{R}^n \)). We seek a useful sufficient condition for \( U \) to be open.

To this end we introduce two technical notions.\(^1\) We call \( a \in U \) a bad point if there exist \( x_1, x_2, \cdots \in \mathbb{R}^n \setminus U \) such that \( x_n \to a \). We call \( a \in U \) a very bad point if there exists \( x \in \mathbb{R}^n \) such that \( \text{dist}(x, U) = |x - a| > 0 \). (Here \( \text{dist}(x, U) = \inf_{y \in U} |x - y| \), of course.)\(^2\)

Clearly, \( U \) is open if and only if it has no bad points, and every very bad point is a bad point. A bad point need not be very bad, and nevertheless, existence of a bad point implies existence of a very bad point. A wonder!

3b7 Lemma. Let \( U \subset \mathbb{R}^n \) be relatively open in its closure. If \( U \) has no very bad points then \( U \) is open.

**Proof.** Let \( a \in U \); we need a neighborhood of \( a \) contained in \( U \). We note that \( \text{dist}(a, \overline{U} \setminus U) > 0 \) (since \( U \) is relatively open in \( \overline{U} \)) and introduce \( \varepsilon = \frac{1}{2} \text{dist}(a, \overline{U} \setminus U) \). It is sufficient to prove that \( U \) contains \( \{ x \in \mathbb{R}^n : |x - a| < \varepsilon \} \).

Assuming the contrary we have \( x \in \mathbb{R}^n \setminus U \) such that \( |x - a| < \varepsilon \), thus \( x \notin \overline{U} \) (since \( |a - x| < \text{dist}(a, \overline{U} \setminus U) \)); taking into account that \( x \notin U \) we get \( x \notin \overline{U} \).

\(^1\)Not a standard terminology; introduced for convenience, to be used within sections 3b–3c only.

\(^2\)It may seem that bad points are well-defined in affine spaces while very bad points are well-defined only in presence of Euclidean metric. In fact, Euclidean metric does not matter. But never mind, we do not need this fact.
By compactness (of the relevant part of $U$), $\text{dist}(x, U) = |x - y| > 0$ for some $y \in U$; we’ll prove that $y$ is a very bad point of $U$.

We introduce $\delta = |x - y|$ and note that $\delta = \text{dist}(x, U) \leq |x - a| < \varepsilon$. Thus $|a - y| \leq |a - x| + |x - y| < \varepsilon + \delta < 2\varepsilon = \text{dist}(a, U \setminus U)$, which gives $y \notin U \setminus U$, that is, $y \in U$. Finally, $y$ is very bad since $|x - y| = \text{dist}(x, U) \leq \text{dist}(x, U)$.

3c Linear and nonlinear

3c1 Definition. A mapping $f : \mathbb{R}^n \to \mathbb{R}^n$ is a (local) homeomorphism near a point $x \in \mathbb{R}^n$ if there exist neighborhoods $U$ of $x$ and $V$ of $f(x)$ such that $f|_U$ is a homeomorphism $U \to V$.

The same applies to mappings from one $n$-dimensional affine space to another.

We know (recall Sect. 1d) that a linear operator $\mathbb{R}^n \to \mathbb{R}^n$ is a homeomorphism if and only if it is bijective. Otherwise it cannot be a homeomorphism near 0 (or any other point).

3c2 Theorem. Let $f : \mathbb{R}^n \to \mathbb{R}^n$ and $x \in \mathbb{R}^n$. If $f$ is continuously differentiable near $x$ and the linear operator $(Df)_x$ is a homeomorphism then $f$ is a homeomorphism near $x$.

The same holds for mappings from one $n$-dimensional affine space to another.

We prove 3c2 in two stages. First, we get a homeomorphism $U \to f(U)$ for some neighborhood $U$ of $x$. Second, we prove that $f(U)$ is a neighborhood of $f(x)$. Here is the exact formulation of the first stage.

3c3 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 $x \in U$.

Spaces treated in Sect. 1b help to prove 3c3.

3c4 Lemma. WLOG we may assume that $x_0 = 0$, $f(x_0) = 0$, and $(Df)_0 = \text{id}$.

Proof. We generalize 3c3 replacing $f : \mathbb{R}^n \to \mathbb{R}^n$ with $f : X \to Y$ where $X,Y$ are $n$-dimensional affine spaces.2 We upgrade $X,Y$ to vector spaces taking $x_0 = 0$ and $f(x_0) = 0$.3 We choose a basis $(e_1, \ldots, e_n)$ in $X$, thus

---

1We could assume that $Df$ is continuous near $x_0$, but this would not simplify the proof.
2Did you know that sometimes a more general claim is easier to prove?
3We could not do it dealing with a single space.
upgrading $X$ to a Cartesian space. We choose in $Y$ the corresponding basis
\[ (Df)_0 e_1, \ldots, (Df)_0 e_n, \]
thus upgrading $Y$ to a Cartesian space and in addition ensuring that the matrix of the operator $(Df)_0$ is the unit matrix.\footnote{Once again, we could not do it dealing with a single space. By the way, an arbitrary matrix is not diagonalizable in the single-space setup, but diagonalizable in the two-spaces setup.}

Now $x_0 = 0$, $f(x_0) = 0$, and $(Df)_0 = \text{id}$.\footnote{Proof of Prop. 3c3} for $x_0 = 0$, $f(x_0) = 0$, and $(Df)_0 = \text{id}$.

We have $(Df)_x \to (Df)_0 = \text{id}$, that is,
\[ \| (Df)_x - \text{id} \| \to 0 \quad \text{as } x \to 0. \]

For every $\varepsilon > 0$ there exists a neighborhood $U_\varepsilon$ of 0 such that $f$ is continuous on $\overline{U_\varepsilon}$, differentiable on $U_\varepsilon$, and
\[ \| (Df)_x - \text{id} \| \leq \varepsilon \quad \text{for all } x \in U_\varepsilon. \]

We choose $U_\varepsilon$ to be convex (just a ball, if you like) and apply 2d10 to the mapping $f - \text{id}$ (its derivative being $Df - \text{id}$): \[ |(f - \text{id})(x) - (f - \text{id})(y)| \leq \varepsilon |x - y|, \]
that is, \[ |(f(x) - f(y)) - (x - y)| \leq \varepsilon |x - y| \quad \text{for all } x, y \in \overline{U_\varepsilon}. \]

It follows (assuming $\varepsilon < 1$) that $f(x) - f(y) \neq 0$ for $x - y \neq 0$; that is, $f|_{\overline{U_\varepsilon}}$ is one-to-one. Moreover, the triangle inequality gives \[ (1 - \varepsilon)|x - y| \leq |f(x) - f(y)| \leq (1 + \varepsilon)|x - y| \]
for all $x, y \in \overline{U_\varepsilon}$. Thus, $f|_{\overline{U_\varepsilon}}$ is a homeomorphism $\overline{U_\varepsilon} \to f(\overline{U_\varepsilon})$.

Finally, \[ |((Df)_x - \text{id})(h)| \leq \varepsilon |h|, \]
that is, \[ |(Df)_x(h) - h| \leq \varepsilon |h| \quad \text{for all } x \in U_\varepsilon, h \in V; \]
the triangle inequality (again) gives \[ (1 - \varepsilon)|h| \leq |(Df)_x(h)| \leq (1 + \varepsilon)|h|, \]
which shows that the operator $(Df)_x$ is one-to-one, therefore invertible. \footnote{The first stage of the proof of Theorem 3c2 is thus completed. On the second stage we prove that $f(U)$ is a neighborhood of $f(x)$. Here is the exact formulation.}
3c5 Proposition. Assume that $U \subset \mathbb{R}^n$ is a bounded open set, $f : \overline{U} \to \mathbb{R}^n$ a homeomorphism $\overline{U} \to f(\overline{U})$, $f$ is differentiable on $U$, and the operator $(Df)_x$ is invertible for all $x \in U$. Then $f(U)$ is open.

Proof. By Lemma 3b7 it is sufficient to prove that the set $V = f(U)$ is relatively open in its closure and has no very bad points.

Being open in $\mathbb{R}^n$, $U$ is relatively open in $\overline{U}$, therefore $V = f(U)$ is relatively open in the set $f(\overline{U})$ of all $f(\lim_k x_k)$ for $x_k \in U$ such that $(x_k)_k$ converges. On the other hand, $\overline{V} = f(\overline{U})$ is the set of all $\lim_k f(x_k)$ for $x_k \in U$ such that $(f(x_k))_k$ converges. Continuity of $f$ implies $f(\overline{U}) \subset \overline{V}$. Compactness of $\overline{U}$ implies $f(\overline{U}) \supset \overline{V}$. Thus, $V$ is relatively open in its closure $\overline{V} = f(\overline{U})$.

Assuming existence of a very bad point in $V$ we get $V \ni b = f(a), a \in U$, and $x \in \mathbb{R}^n$ such that $\text{dist}(x, V) = |x - b| > 0$. A function $|x - f(\cdot)|$ on $U$ has at $a$ a minimum. However, this function is $\varphi \circ f$ where $\varphi(\cdot) = |x - \cdot|;^3$ thus $D(\varphi \circ f)_a = (D\varphi)_b \circ (Df)_a \neq 0$, since $(Df)_a$ is bijective and $(D\varphi)_b \neq 0$. A contradiction. □

3c6 Remark. 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 3b7? For two reasons.

First, invariance of domain is proved using algebraic topology (the Brouwer fixed point theorem). Lemma 3b7 much simpler to prove, suffices due to differentiability.

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

3c7 Exercise. Prove invariance of domain in dimension one.\(^6\)

3c8 Exercise. Consider the set $U \subset \mathbb{R}^n$ of all $(a_0, \ldots, a_{n-1})$ such that the polynomial

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

has $n$ pairwise distinct real roots.

---

1Recall Sect. 1c.

2True, $x_k \to x \iff f(x_k) \to f(x)$ for $x_k, x \in \overline{U}$, but the question is, what to do if $f(x_k) \to y \in \overline{U} \setminus f(\overline{U})$; the answer is, choose a convergent $(x_k)_i$.

3Alternatively, consider a path $\gamma : [t_0, t_1] \to U$ such that some $t \in (t_0, t_1)$ satisfies $\gamma(t) = a$ and $\gamma'(t) = ((Df)_a)^{-1}(b - x)$.

4By the way, it follows from the Brouwer invariance of domain theorem that an open set in $\mathbb{R}^{n+1}$ cannot be homeomorphic to any set in $\mathbb{R}^n$ (unless it is empty). Think, why.

5Still another alternative to Lemma 3b7 will be discussed in Sect. 4d, see 4d2.

6Hint: recall 3b4.
(a) Prove that $U$ is open.

(b) Define $\psi : U \to \mathbb{R}^n$ by $\psi(a_0, \ldots, a_{n-1}) = (t_1, \ldots, t_n)$ where $t_1 < \cdots < t_n$ are the roots of the polynomial. Prove that $\psi$ is a homeomorphism $U \to V$ where $V = \{(t_1, \ldots, t_n) : t_1 < \cdots < t_n\}$.\footnote{Hint: use 2c11(b).}

### 3d Curves

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

#### 3d1 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 $h(x, y) = (f(x, y), g(x, y))$ is continuously differentiable near $(x_0, y_0)$, and $(Dh)(x_0, y_0)$ is invertible by 2f2. Theorem 3c2 provides a neighborhood $U$ of $(x_0, y_0)$ such that $V = h(U)$ is a neighborhood of $h(x_0, y_0) = (z_0, 0)$ 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 $\gamma$ by $\gamma(t) = (h|_U)^{-1}(t, 0)$.

Clearly $\gamma$ is continuous, $\gamma(z_0) = (x_0, y_0)$, $\gamma(t) \in U$ and $h(\gamma(t)) = (t, 0)$, that is, $f(\gamma(t)) = t$ and $g(\gamma(t)) = 0$. \hfill $\square$

#### 3d2 Corollary

If $f, g, x_0, y_0$ are as in 3d1 then $(x_0, y_0)$ cannot be a local constrained extremum of $f$ on $Z_g$.

#### 3d3 Remark

(a) Prop. 3d1 does not claim differentiability of the path $\gamma$ (but only its continuity).

(b) Prop. 3d1 does not claim that $\gamma$ 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 4c12).

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 $Z_{g_1, g_2} = Z_{g_1} \cap Z_{g_2}$.

#### 3d4 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 $g_1(x_0, y_0, z_0) = g_2(x_0, y_0, z_0) = 0$. \hfill \footnote{Hint: use 2c11(b).}
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 $t \in (w_0 - \varepsilon, w_0 + \varepsilon)$.

3d5 Exercise. Prove Prop. 3d4.

3d6 Corollary. If $f, g_1, g_2, x_0, y_0, z_0$ are as in 3d4 then $(x_0, y_0, z_0)$ cannot be a local constrained extremum of $f$ on $Z_{g_1, g_2}$.

3d7 Exercise. Generalize 3d4 and 3d6 to $f, g_1, \ldots, g_{n-1} : \mathbb{R}^n \to \mathbb{R}$.

3e 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 $(x, y, z) \mapsto (f(x, y, z), g(x, y, z))$ 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!

3e1 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 $t \in (w_0 - \varepsilon, w_0 + \varepsilon)$.

\textbf{Proof.} We choose a vector $a \in \mathbb{R}^3$ such that the three vectors $\nabla f(x_0, y_0, z_0)$, $\nabla g(x_0, y_0, z_0)$ and $a$ 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 $g_2(x_0, y_0, z_0) = 0$ and $\nabla g_2(x_0, y_0, z_0) = a$ (for example, an affine function $g_2(\cdot) = \langle \cdot, a \rangle + \text{const}$). It remains to apply Prop. 3d4 to $f, g, g_2$. \hfill \Box

3e2 Corollary. If $f, g, x_0, y_0, z_0$ are as in 3e1 then $(x_0, y_0, z_0)$ cannot be a local constrained extremum of $f$ on $Z_g$.

3e3 Exercise. Generalize 3e1 and 3e2 to $f, g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$, $1 \leq m \leq n - 1$.

3f Lagrange multipliers

3f1 Theorem. Assume that $x_0 \in \mathbb{R}^n$, $1 \leq m \leq n - 1$, functions $f, g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable near $x_0$, $g_1(x_0) = \cdots = g_m(x_0) = 0$, and vectors $\nabla g_1(x_0), \ldots, \nabla g_m(x_0)$ are linearly independent. If $x_0$ is a local

\text{Hint: similar to the proof of 3e1} $h(x, y, z) = (f(x, y, z), g_1(x, y, z), g_2(x, y, z))$, \ldots
constrained extremum of $f$ subject to $g_1(\cdot) = \cdots = g_m(\cdot) = 0$ then there exist $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ such that

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

This is a reformulation of the generalization meant in $\text{3e3}$.

The numbers $\lambda_1, \ldots, \lambda_m$ 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 $Z = Z_{g_1, \ldots, g_m}$ one solves (that is, finds all solutions of) a system of $m + n$ equations

\begin{align*}
g_1(x) = \cdots = g_m(x) &= 0, \quad (m \text{ equations}) \\
\nabla f(x) &= \lambda_1 \nabla g_1(x) + \cdots + \lambda_m \nabla g_m(x) \quad (n \text{ equations})
\end{align*}

for $m + n$ variables

$$\lambda_1, \ldots, \lambda_m, \quad (m \text{ variables})$$

$$x, \quad (n \text{ variables})$$

For each solution $(\lambda_1, \ldots, \lambda_m, x)$ one ignores $\lambda_1, \ldots, \lambda_m$ and checks $f(x)$.

In addition, one checks $f(x)$ for all points $x$ that violate the conditions of $\text{3f1}$ that is, $\nabla g_1(x), \ldots, \nabla g_m(x)$ are linearly dependent, or $f, g_1, \ldots, g_m$ 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: $\sup_Z f = \lim_{k} f(x_k) \in (-\infty, +\infty]$ for some $x_1, x_2, \cdots \in Z$.

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 $\text{[3f1]}$. 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.

Theorem $\text{[3f1]}$ 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 $\nabla f(x_0) = \lambda_1 \nabla g_1(x_0) + \cdots + \lambda_m \nabla g_m(x_0)$ between vectors may be replaced with a relation

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

\footnote{Being ignored in this framework, $(\lambda_1, \ldots, \lambda_m)$ are of interest in another framework, see Sect. $\text{[3]}$.}
between linear functions. And linear independence of vectors $\nabla g_1(x_0), \ldots, \nabla g_m(x_0)$ may be replaced with linear independence of linear functions $(Dg_1)_{x_0}, \ldots, (Dg_m)_{x_0}$; or, due to Lemma 2f2, we may say instead that $(Dg)_{x_0}$ maps $\mathbb{R}^n$ onto $\mathbb{R}^m$. Now it is clear how to generalize Th. 3f1 from $\mathbb{R}^n$ to an $n$-dimensional affine space.

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

Here is an isoperimetric inequality for triangles $\Delta$ on the plane:

$$\text{area}(\Delta) \leq \frac{1}{12\sqrt{3}} (\text{perimeter}(\Delta))^2,$$

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.\(^1\)

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\(^2\) 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 $A^2 = \frac{L}{2} \left( \frac{L}{6} \right)^3 = \frac{L^4}{24\sqrt{3}}; A = \frac{L^2}{2\sqrt{3}\sqrt[3]{3}}$.

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

$$(3g1) \quad (x_1 \ldots x_n)^{1/n} \leq \frac{1}{n} (x_1 + \cdots + x_n) \quad \text{for} \quad n = 1, 2, \ldots \quad \text{and} \quad x_1, \ldots, x_n \geq 0.$$

It follows easily from concavity of the logarithm: the set $A = \{(x, y) : x \in (0, \infty), y \leq \ln x\}$ is convex, therefore the convex combination $\left( \frac{1}{n} (x_1 + \cdots + x_n), \frac{1}{n} (\ln x_1 + \cdots + \ln x_n) \right)$ of points $(x_1, \ln x_1), \ldots, (x_n, \ln x_n) \in A$ belongs to $A$, which gives (3g1). And still, it is worth to exercise Lagrange multipliers.

\(^1\)Generally, $\text{area}(G) \leq \frac{1}{4\pi} (\text{perimeter}(G))^2$ 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).

\(^2\)\(\frac{L}{2} - x = \frac{x+y+z}{2} - x = \frac{y+z-x}{2} > 0\) by the triangle inequality.
3g2 Exercise. Prove [3g1] via Lagrange multipliers.

By the way, the harmonic mean $h$ defined by $\frac{1}{h} = \frac{1}{n} (\frac{1}{x_1} + \cdots + \frac{1}{x_n})$ satisfies
\[ h \leq (x_1 \ldots x_n)^{1/n}; \] just apply [3g1] to $\frac{1}{x_1}, \ldots, \frac{1}{x_n}$.

More generally, the Hölder mean (called also power mean) with exponent $p \in (-\infty, 0) \cup (0, \infty)$ is
\[ M_p(x_1, \ldots, x_n) = \left( \frac{x_1^p + \cdots + x_n^p}{n} \right)^{1/p} \] for $x_1, \ldots, x_n > 0$.

In particular, $M_1$ is the arithmetic mean and $M_{-1}$ is the harmonic mean. For $p \to 0$ L’Hôpital’s rule gives
\[ \ln \lim_{p \to 0} M_p((x_1, \ldots, x_n)) = \lim_{p \to 0} \frac{1}{p} \ln \frac{x_1^p + \cdots + x_n^p}{n} = \lim_{p \to 0} \frac{x_1^p \ln x_1 + \cdots + x_n^p \ln x_n}{x_1^p + \cdots + x_n^p} = \frac{\ln x_1 + \cdots + \ln x_n}{n} = \ln(x_1 \ldots x_n)^{1/n}; \]
accordingly, one defines
\[ M_0(x_1, \ldots, x_n) = (x_1 \ldots x_n)^{1/n}, \] and observes that $M_{-1}(x_1, \ldots, x_n) \leq M_0(x_1, \ldots, x_n) \leq M_1(x_1, \ldots, x_n)$. For $p \to +\infty$ we have
\[ \frac{1}{n} \max(x_1^p, \ldots, x_n^p) \leq \frac{x_1^p + \cdots + x_n^p}{n} \leq \max(x_1^p, \ldots, x_n^p), \]
therefore $M_p(x_1, \ldots, x_n) \to \max(x_1, \ldots, x_n)$; one writes
\[ M_{+\infty}(x_1, \ldots, x_n) = \max(x_1, \ldots, x_n); \quad M_{-\infty}(x_1, \ldots, x_n) = \min(x_1, \ldots, x_n) \]
(the latter being similar to the former) and observes that $M_{-\infty}(x_1, \ldots, x_n) \leq M_{1}(x_1, \ldots, x_n) \leq M_0(x_1, \ldots, x_n) \leq M_{1}(x_1, \ldots, x_n) \leq M_{+\infty}(x_1, \ldots, x_n)$. That is interesting! Maybe $M_p \leq M_q$ whenever $p \leq q$?

We treat $M_p$ as a function on $(0, \infty)^n \subset \mathbb{R}^n$ and calculate its gradient $\nabla M_p$, or rather, the direction of the vector $\nabla M_p$; indeed, we only need to know when two vectors $\nabla M_p, \nabla M_q$ are linearly dependent, that is, collinear (denote it $\parallel$). We have $\nabla M_p \parallel \nabla M_p \parallel \nabla (nM_p) \parallel (x_1^{p-1}, \ldots, x_n^{p-1})$ for $p \neq 0$; however, this result holds for $p = 0$ as well, since $\nabla M_0 \parallel \nabla \ln M_0 \parallel (x_1^{-1}, \ldots, x_n^{-1})$. Thus, $\nabla M_p, \nabla M_q$ are collinear if and only if $\frac{x_1^{q-1}}{x_1} = \cdots = \frac{x_n^{q-1}}{x_n}$, that is, $x_1^{q-p} = \cdots = x_n^{q-p}$, or just $x_1 = \cdots = x_n$. In this case, evidently,
\( M_p = M_q \). Does it prove that \( M_p \leq M_q \) always? Not yet. Functions \( M_p, M_q \) are continuously differentiable on the open set \( G = (0, \infty)^n \), and on the set \( Z_p = \{ x \in G : M_p(x) = 1 \} \) the conditions of [3ff] are violated at one point \((1, \ldots, 1)\) 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 \( \infty \).

**Case 1:** \( 0 < p < q < \infty \). The set \( Z_p \) is bounded, since \( \max(x_1, \ldots, x_n) \leq (x_1^p + \cdots + x_n^p)^{1/p} = n^{1/p}M_p(x_1, \ldots, x_n) = n^{1/p} \), but not closed. 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 \( M_p \leq M_q \) in dimension \( n - 1 \) we get (assuming \( x_n = 0 \))

\[
\frac{M_q(x)}{M_p(x)} = \left( \frac{1}{n} (x_1^q + \cdots + x_{n-1}^q + 0^q) \right)^{1/q} / \left( \frac{1}{n} (x_1^p + \cdots + x_{n-1}^p + 0^p) \right)^{1/p} = \\
\left( \frac{n}{n-1} \right)^{1/q} \left( \frac{1}{n-1} (x_1^q + \cdots + x_{n-1}^q) \right)^{1/q} / \left( \frac{1}{n-1} (x_1^p + \cdots + x_{n-1}^p) \right)^{1/p} \geq \left( \frac{n}{n-1} \right)^{1/q} > 1,
\]

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 \( 1/x_1, \ldots, 1/x_n \), since

\[
1/M_{-p}(x_1^{-1}, \ldots, x_n^{-1}) = \left( \frac{x_1^p + \cdots + x_n^p}{n} \right)^{1/p} = M_p(x_1, \ldots, x_n);
\]

\[
M_p(x_1, \ldots, x_n) = 1/M_{-p}(x_1^{-1}, \ldots, x_n^{-1}) \leq 1/M_{-q}(x_1^{-1}, \ldots, x_n^{-1}) = M_q(x_1, \ldots, x_n).
\]

**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: \( M_p \leq M_0 \leq M_q \).

So, \( M_p \leq M_q \) whenever \( p \leq q \).

Some practical advice.

\(^1\)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 \( M_p(\lambda x) \) does not depend on \( \lambda \).

\(^2\)For example, the point \((n^{1/p}, 0, \ldots, 0)\) belongs to \( \partial Z_p \).
The system of \( m + n \) equations proposed in Sect. 3H is only one way of finding local constrained extrema. Not necessarily the simplest way.

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

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

\[
M_{p,w}(x) = (x_1^p w_1 + \cdots + x_n^p w_n)^{1/p}
\]

for given \( w_1, \ldots, w_n \geq 0 \) satisfying \( w_1 + \cdots + w_n = 1 \). In particular, \( M_{1,w}(x) \leq M_{p,w}(x) \) for \( p \geq 1 \), that is, \( x_1 w_1 + \cdots + x_n w_n \leq (x_1^p w_1 + \cdots + x_n^p w_n)^{1/p} \).

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

\[
\sum_i a_i b_i^{-q/p} b_i^q \leq \left(\sum_i a_i^p b_i^{-q/p} b_i^q\right)^{1/p},
\]

that is, \( \sum_i a_i b_i \leq \left(\sum_i a_i^p\right)^{1/p} \) provided that \( \sum_i b_i^q = 1 \). This leads easily to the Hölder’s inequality

\[
|\sum_i x_i y_i| \leq \left(\sum_i |x_i|^p\right)^{1/p} \left(\sum_i |y_i|^q\right)^{1/q}
\]

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.

**3g3 Exercise.** Given \( a_1, \ldots, a_n > 0 \), maximize \( a_1 x_1 + \cdots + a_n x_n \) on \( \{x \in [0,\infty)^n : x_1^p + \cdots + x_n^p = 1\} \) 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).

**3g4 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 \( x^k + y^k + z^k = 1 \).

\(^1\)Hint: induction in \( n \) is needed again.
3g5 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]

3h 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).

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,

$$f(x_1, \ldots, z_3) = (x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2$$

$$+ (x_2 - x_3)^2 + (y_2 - y_3)^2 + (z_2 - z_3)^2$$

$$+ (x_3 - x_1)^2 + (y_3 - y_1)^2 + (z_3 - z_1)^2;$$

$$g_1(x_1, \ldots, z_3) = x_1^2 + y_1^2 + \alpha z_1^2 - 1,$$

$$g_2(x_1, \ldots, z_3) = x_2^2 + y_2^2 + \alpha z_2^2 - 1,$$

$$g_3(x_1, \ldots, z_3) = x_3^2 + y_3^2 + \alpha z_3^2 - 1.$$ We use the approach of Sect. 3f 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,

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

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:

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

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

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

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 $z_1 = z_2 = z_3 = z_1 + z_2 + z_3 = 0$ (since $\alpha \neq 1$);

$P, Q, R$ are situated on the horizontal plane $\{ (x, y, z) : z = 0 \}$.

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.
3h1 Exercise.  
Let \(a, b \in \mathbb{R}^n\) be linearly independent, \(|a| = 5, |b| = 10\). Functions \(\varphi_a, \varphi_b\) 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

3i Example: Singular value decomposition

3i1 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.³ It may be reformulated as follows.

3i2 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.

\[
\begin{array}{ccc}
\text{m < n} & \text{m = n} & \text{m > n} \\
\text{\includegraphics{matrix.png}} & \text{\includegraphics{matrix.png}} & \text{\includegraphics{matrix.png}}
\end{array}
\]

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. 3i1 will be derived from Prop. 3i3 below.

3i3 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.

---

¹Exam of 26.01.14, Question 2.
²Hint: show that \(\sin \frac{1}{2} \varphi_a(x) = 1/|x - a|\); use the gradient.
³See Todd Will, "Introduction to the Singular Value Decomposition", [http://www.uwlax.edu/faculty/will/svd/index.html](http://www.uwlax.edu/faculty/will/svd/index.html)

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

$$\forall x, y \in E \quad ⟨x,y⟩_1 = ⟨A(x),y⟩;$$

it is well-defined, since the linear form $⟨x,·⟩_1$, as every linear form, is $⟨a,·⟩$ for some $a \in E$. It is easy to see that $A$ is a linear operator, symmetric in the sense that

$$\forall x, y \in E \quad ⟨Ax,y⟩ = ⟨x, Ay⟩.$$

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

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

by 2b11, or just by a very simple calculation:

$$|x + h|^2 = |x|^2 + ⟨x,h⟩ + |h|^2 = |x|^2 + 2⟨x,h⟩ + o(|h|),$$

$$|x + h|^2_1 = |x|^2_1 + ⟨x,h⟩_1 + |h|^2_1 = |x|^2_1 + 2⟨Ax,h⟩ + o(|h|).$$

These two gradients are collinear if and only if $\exists \lambda$ $Ax = λx$; 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, $|·|^2$ reaches its maximum on $S$; by Theorem 3f1, 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 $A$: $⟨Ax,e_n⟩ = ⟨x,Ae_n⟩ = ⟨x,λ_ne_n⟩ = λ_n⟨x,e_n⟩ = 0$. 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 $|·|$ and $|·|_1$ (restricted to $E_{n-1}$), and $⟨x,y⟩_1 = ⟨A_{n-1}x,y⟩$ 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 $(e_1, \ldots, e_{n-1})$ of $E_{n-1}$ orthonormal in $|·|$ and orthogonal in $|·|_1$. Thus, $(e_1, \ldots, e_{n-1}, e_n)$ is a basis of $E$. We normalize $e_n$ to $|e_n| = 1$; now this basis is orthonormal in $|·|$. It is also orthogonal in $|·|_1$, since $⟨e_k,e_n⟩_1 = ⟨Ae_k,e_n⟩ = 0$ for $k = 1, \ldots, n-1$.

\[\square\]

1 All gradients are taken in $E = (V, |·|)$, not $(V, |·|_1)$!
3i4 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. 3i1. 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 \( E \):

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

(of course, \( |Tx| \) is the norm in \( E_2 \)). Prop. 3i3 gives an orthonormal basis \( (e_1, \ldots, e_n) \) of \( E \), orthogonal in the second metric: \( \langle e_k, e_l \rangle = 0 \) for \( k \neq l \).

That is, \( \langle Te_k, Te_l \rangle = 0 \), which shows that \( (Te_1, \ldots, Te_n) \) is an orthogonal system in \( E_2 \).

If \( T \) is not one-to-one, the same argument applies due to Remark 3i4. □

Prop. 3i2 follows immediately, and gives a diagonal matrix. Its diagonal elements can be made \( \geq 0 \) (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, 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 \( 1/\|T^{-1}\| \).

3j Sensitivity of optimum to parameters

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

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

Assume that the system of \( m + n \) equations

\[
\begin{align*}
g_1(x) &= c_1, \ldots, g_m(x) = c_m, \\
\nabla f(x) &= \lambda_1 \nabla g_1(x) + \cdots + \lambda_m \nabla g_m(x)
\end{align*}
\]

1Alternatively, define \( |x|_1^2 = |Tx|^2 + |x|^2, \langle x, y \rangle_1 = \langle Tx, Ty \rangle + \langle x, y \rangle \).

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

3On the space of operators, the Schatten norm is \( \|T\|_p = (|s_1|^p + \cdots + |s_n|^p)^{1/p} \) where \( s_1, \ldots, s_n \) are the singular values of \( T \) (and \( 1 \leq p \leq \infty \)).

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

5A more general one: \( g_1(x, c_1) = 0, \ldots, g_m(x, c_m) = 0 \).

...
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,
\[
\frac{\partial}{\partial c_k}\bigg|_{c=0} f(x(c)) = \left\langle \nabla f(x(0)), \frac{\partial}{\partial c_k}\bigg|_{c=0} x(c) \right\rangle \quad \text{for} \quad k = 1, \ldots, m.
\]
On the other hand,
\[
\nabla f(x(0)) = \lambda_1(0) \nabla g_1(x(0)) + \cdots + \lambda_m(0) \nabla g_m(x(0))
\]
and
\[
\left\langle \nabla g_1(x(0)), \frac{\partial}{\partial c_k}\bigg|_{c=0} x(c) \right\rangle = \frac{\partial}{\partial c_k}\bigg|_{c=0} g_1(x(c)) = \begin{cases} 1, & \text{if } k = 1, \\ 0, & \text{otherwise} \end{cases}
\]
(since \(g_1(x(c)) = c_1\)). The same holds for \(g_2, \ldots, g_m\). Therefore
\[
\frac{\partial}{\partial c_k}\bigg|_{c=0} f(x(c)) = \lambda_k(0).
\]
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,
\[
f(x(c)) = f(x(0)) + \lambda_1(0)c_1 + \cdots + \lambda_m(0)c_m + o(|c|).
\]
Does it mean that
\[
(3j1) \quad \sup_{Z_c} f = \sup_{Z_0} f + \lambda_1(0)c_1 + \cdots + \lambda_m(0)c_m + o(|c|)
\]
where \(Z_c = \{x : g_1(x) = c_1, \ldots, g_m(x) = c_m\}\)? 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.\(^1\)

Index

- constraint function, 51
- Lagrange multipliers, 60
- Hölder mean, 62
- objective function, 51
- Hölder’s inequality, 64
- open mapping, 53
- homeomorphism near a point, 55
- invariance of domain, 57
- \(Z_g, 51\)
- \(Z_{g_1, g_2}, 58\)