1 Introduction

The field of Mathematical Optimization is concerned with finding minimums and maximums of functions of one or several variables. Often the variables must also satisfy certain equations, called constraints. Of importance also, when possible, is finding the values of the variables that minimize or maximize the function. Sometimes the functions can be described in a geometric way as in the following example.

Example 1. The ancient greek mathematician Heron of Alexandria posed one of the first optimization problems. Let A and B be two points and let ℓ be a line in the plane. Find a point C on the line ℓ so that the sum of the distances $\operatorname{dist}(A, C) + \operatorname{dist}(C, B)$ is as small as possible.

Solution: In class.

Example 2 (The Arithmetic-Geometric Mean Inequality)**.** Let x and y be two non-negative numbers, that is $x \geq 0$ and $y \geq 0$. Then we always have the inequality

$$\sqrt{xy} \leq \frac{x+y}{2}.$$

Equality holds if and only if $x = y$.

Example 3. *Suppose you have enough material to build a fence of length ℓ. What is the maximal rectangular area that you can enclose with the fence?*

Solution: In class.

Most inequalities, can be interpreted as optimization problems. Indeed, the arithmetic geometric mean inequality can be reformulated as

$$\begin{array}{ll} \text{minimize} & x+y \\ \text{subject to} & xy = 1 \\ & x, y \geq 0 \end{array}$$

or equivalently

$$\begin{array}{ll} \text{maximize} & xy \\ \text{subject to} & x+y = 1 \\ & x, y \geq 0 \end{array}$$

These are examples of constrained optimization problems. The letters x and y are called *decision variables*, since have to decide what values of the variables attains the maximum/minimum. The function that we have to minimize or maximize is called the *objective function*, while the conditions that the decision variables must satisfy are called *constraints*.

Example 4. Find the minimum of the function $f(x) = x + 1/x$ on $(0, \infty)$.

Solution: In class.

Recall the following standard theorem from Calculus.

Theorem 5 (First and second derivative tests)**.** *Suppose $f(x)$ is differentiable on an open interval (a,b). If $x^* \in (a,b)$ is a local extremum (minimum or a maximum) then $f'(x^*) = 0$.*

In addition, suppose that $f(x)$ is twice differentiable on (a,b). If $f''(x^*) > 0$ then x^* is a local minimum. Respectively, if $f''(x^*) < 0$ then x^* is a local maximum. If $f''(x^*) = 0$ then nothing can be concluded and alternative ideas must me employed.

Example 6. Find the local maxima and minima of the function $f_n(x) = x^n$, $n \geq 1$.

Solution: To find the candidates for local minima and maxima we need to find the critical points of $f_n(x)$. Differentiating

$$f'(x) = nx^{n-1}$$

and setting $f_n'(x) = 0$ we obtain that when $n = 1$ there is no critical point, while for $n > 1$ there is one critical point $x = 0$. Thus, when $n = 1$, $f_1(x)$ has no extrema. To determine if the critical point is a minima or a maxima, we may try the second derivative test. Since $f_n''(x) = n(n-1)x^{n-2}$ we see that $f_2''(0) = 2 > 0$ when $n = 2$ and $f_n''(0) = 0$ when $n > 2$. Thus, the second derivative test implies that $x = 0$ is a local minimum when $n = 2$ but it fails when $n > 2$.

To get around this problem when $n > 2$, we need to determine the sign of $f_n'(x)$ right before $x = 0$ and right after $x = 0$. This is easy, let $n > 2$, we need to consider two cases. If n is an odd number then $f_n'(x) > 0$ for all $x \neq 0$ thus $f_n(x)$ is strictly increasing for all $x \neq 0$ and $x = 0$ is not an extremum. If n is an even number then $f_n'(x) < 0$ for all $x < 0$ and $f_n'(x) > 0$ for all $x > 0$ thus $f_n(x)$ is strictly decreasing for $x < 0$ and strictly increasing for $x > 0$, thus $x = 0$ is a global minima.

Example 7. Find the rectangle, inscribed in the circle centered at the origin of the coordinate system with radius 1, having the largest area.

Solution: Without loss of generality, we may assume that the sides of the rectangle are parallel to the coordinate axis. Then, let $(x, 0)$ and $(-x, 0)$ be the points where the vertical sides cross the x-axis. Since the rectangle is inscribed in circle with radius one, $x \in [0,1]$. By the Pythagorean theorem, we find that the four vertices of the rectangle are $(x, \sqrt{1-x^2})$, $(x, -\sqrt{1-x^2})$, $(-x, -\sqrt{1-x^2})$, and $(-x, \sqrt{1-x^2})$. The area is

$$f(x) = 4x\sqrt{1-x^2} \quad x \in [0,1].$$

Since $f(0) = f(1) = 0$, and since $f(x) \geq 0$ for $x \in [0,1]$, the maximum of $f(x)$ occurs in the open interval $(0,1)$. We need to find the derivatives of f.

$$f'(x) = 4\sqrt{1-x^2} - \frac{4x^2}{\sqrt{1-x^2}}.$$

In order to find the critical points of f, we need to solve the equation $f(x) = 0$. The solution is $x = 1/\sqrt{2}$. This is the only critical point in $(0,1)$ and hence it must be the point where f attains its maximum. (Why?) The maximal area of an inscribed rectangle is $f(1/\sqrt{2}) = 2$. We can calculate the lengths of the sides of the the maximal rectangle (how?) to see that they are both equal of length $\sqrt{2}$. Thus, the rectangle with maximal area inscribed in a circle is a square. Did it matter that the circle had radius 1, or that its center is the origin of the coordinate system?

We conclude this section with a result that generalizes Theorem 5.

Theorem 8 (n-th derivative test)**.** *Suppose $f(x)$ is n-times continuously differentiable on an open interval (a, b). Suppose $x^* \in (a, b)$ is such that*

$$f'(x^*) = f''(x^*) = \cdots = f^{(n-1)}(x^*) = 0 \quad \textit{but} \quad f^{(n)}(x^*) \neq 0.$$

If n is even and $f^{(n)}(x^*) > 0$ then x^* is a local minimum. If n is even and $f^{(n)}(x^*) < 0$ then x^* is a local maximum. If n is odd then there is neither local minimum nor a local maximum at x^*.

Exercise 9. Find the local maxima and minima of the function $f_n(x) = x^n e^x$, $n \geq 1$.

Exercise 10. Given a point C in the interior of an acute angle, find the points A and B on the sides of the angle such that the perimeter of the triangle ABC is as short as possible.

Exercise 11. Find the minima of the functions

$$\begin{aligned} f(x) &= x\ln(x) \\ g(x) &= x - \ln(x), \end{aligned}$$

on the interval $(0, \infty)$. Show all details of your work.

Exercise 12. Show that the function $f(x) = x^2 - 1/\ln(x)$ has no local or global minima on either $(0, 1)$ or $(1, \infty)$.

Exercise 13. Prove Theorem 8. Hint: Use the n-th order Taylor expansion of f at x^*.

Exercise 14. Find the local extremums of the functions.

(i) $x^3 - 12x$;

(ii) $x^3 - 3x^2 + 6x + 7$;

(iii) $x^3 - 9x^2 + 15x - 3$;

(iv) $x^4 - 8x^3 + 22x^2 - 24x + 12$;

(v) $x^5 - 5x^4 + 5x^3 - 1$;

(vi) $\frac{x}{1+x^2}$;

(vii) $\frac{x^2-7x+6}{x-10}$;

(viii) $x^2\ln(x)$, $x \in (0, \infty)$;

(ix) x^x, $x \in (0, \infty)$;

(x) $x(\ln(x))^2$, $x \in (0, \infty)$;

(xi) $x^n e^{-x}$;

(xii) $x^2 e^{-x^2}$;

(xiii) $e^{-x} - e^{-2x}$;

(xiv) $x^3(x-1)^{2/3}$, $x \in [-2, 2]$;

(xv) $e^x \cos(x)$.

Exercise 15. For any α in the interval

(i) $(0, 4)$ let $m(\alpha)$ be the global minimum of the function $\ln(x^2 + \alpha x + \alpha)$. For what value of α does $m(\alpha)$ attains its maximum?

(ii) $(-\infty, 1]$ let $m(\alpha)$ be the global minimum of the function $(1+x)\arctan(x) - (x+1)\arctan(\alpha) - x\frac{1+\alpha}{1+\alpha^2}$, $x \in (-\infty, 1]$. For what value of α does $m(\alpha)$ attains its maximum?

(iii) $(-\infty, 0]$ let $m(\alpha)$ be the global minimum of the function $x\ln(\frac{1+x^2}{1+\alpha^2}) - \frac{2\alpha^2 x}{1+\alpha^2} + 2\alpha$, $x \in [0, \infty)$. For what value of α does $m(\alpha)$ attains its maximum?

(iv) $(0, \infty)$ let $M(\alpha)$ be the global maximum of the function $\arctan(x) - \frac{\alpha}{2}\ln(1+x^2) + \frac{\alpha}{2}\ln(a)$, $a > 1$. For what value of α does $M(\alpha)$ attains its minimum?

Exercise 16. Find the optimal value and the optimal solution of the optimization problem

$$\begin{array}{ll} \text{minimize} & (x_1 + x_2 + x_3)\left(\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3}\right) \\ \text{subject to} & x_1 > 0, x_2 > 0, x_3 > 0. \end{array}$$

Exercise 17. Find the right circular cylinder of volume V having the least surface area.

Exercise 18. (i) Find the cylinder inscribed in a sphere with radius one with the largest surface area.

(ii) Find the cylinder inscribed in a sphere with radius one with the largest volume.

(iii) Find the cone inscribed in a sphere with radius one with the largest surface area.

(iv) Find the cone inscribed in a sphere with radius one with the largest volume.

Exercise 19 (The Arithmetic-Geometric Mean Inequality)**.** Let $x_1, ..., x_n$ non-negative numbers. Then we always have the inequality

$$\sqrt[n]{x_1 \cdots x_n} \le \frac{x_1 + \cdots + x_n}{n}.$$

Equality holds if and only if $x_1 = \cdots = x_n$.

Exercise 20. Find all pairs of real numbers $x \ge 0$, $y \ge 0$ that minimize the polynomial $P(x, y) = x^4y^2 + x^2y^4 - 3x^2y^2 + 1$.

Exercise 21. If a triangle has sides with lengths a, b, and c, then (according to Heron's formula) its ares is given by $\sqrt{s(s-a)(s-b)(s-c)}$, where $s = (a+b+c)/2$ is half of the perimeter. Given enough material to make a fence with length ℓ, enclose a triangular-shaped land with the largest area and perimeter ℓ.

Exercise 22 (The harmonic-geometric mean inequality)**.** Let $x_1, ..., x_n$ non-negative numbers. Then we always have the inequality

$$\frac{n}{\frac{1}{x_1} + \cdots + \frac{1}{x_1}} \leq \sqrt[n]{x_1 \cdots x_n}.$$

Equality holds if and only if $x_1 = \cdots = x_n$.

Exercise 23 (Power mean inequality)**.** *For all non-negative numbers x_1, x_2,...,x_n and all integer powers $m = 1, 2, ...$ one has the inequality*

$$\frac{x_1 + \cdots + x_n}{n} \leq \sqrt[m]{\frac{x_1^m + \cdots + x_n^m}{n}}.$$

Equality holds if and only if $x_1 = \cdots = x_n$.

1.1 Norm and inner product

Let $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$ be two vectors in $\mathbb{R}^n$. The distance between x and y is given by

$$\operatorname{dist}(x, y) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}.$$

The *norm* or *length* of vector x is

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}.$$

Note that the norm of x (or the length of x) is the distance from x to the origin of the coordinate system $O = (0, ..., 0)$. The *inner product* between x and y is defined by

$$\langle x, y \rangle := x_1 y_1 + \cdots + x_n y_n.$$

Note that $\|x\| = \sqrt{\langle x, x \rangle}$ and that $\operatorname{dist}(x, y) = \|x - y\|$.

Theorem 24 (Cauchy-Schwarz inequality)**.** *For any two vectors x and y in $\mathbb{R}^n$ we have*

$$|\langle x, y \rangle| \leq \|x\| \|y\|. \tag{1}$$

Equality holds if and only if either $y = 0$ or there is a scalar t_0 such that $x = t_0 y$.

Proof. If $y = 0$ the inequality is trivial so suppose that $y \neq 0$, that is $\langle y, y \rangle = \|y\|^2 > 0$. Consider the quadratic function in t:

$$f(t) := \langle x - ty, x - ty \rangle = \langle y, y \rangle t^2 - 2\langle x, y \rangle t + \langle x, x \rangle.$$

Since $\langle x - ty, x - ty \rangle = \|x - ty\|^2 \geq 0$ we see that $f(t) \geq 0$ for all t. This can happen if and only if the discriminant of the quadratic function is non-positive:

$$D = \langle x, y \rangle^2 - \langle y, y \rangle \langle x, x \rangle \leq 0.$$

Moving $\langle y, y \rangle \langle x, x \rangle$ to the right of the inequality and taking square roots from both sides we arrive at (1).

Suppose that (1) holds with equality. Then, the quadratic equation $f(t) = 0$ has a real root t_0. That is, $0 = f(t_0) = \|x - t_0 y\|$ implying that $x = t_0 y$. In the opposite direction, if there is a number t_0 such that $x = t_0 y$ then again one can easily verify the equality. $\square$

Again the Cauchy-Schwartz inequality can be recast as an optimization problem. Indeed, for a fixed vector $y \in \mathbb{R}^n$ solve

$$\begin{array}{ll} \text{maximize} & \langle x, y\rangle \\ \text{subject to} & \|x\| = 1 \\ & x \in \mathbb{R}^n. \end{array}$$

By the Cauchy-Schwarz inequality, the optimal solution of this problem is $x^* = y/\|y\|$ with optimal value $\langle x^*, y\rangle = \|y\|$

We saw that the notion of a norm in $\mathbb{R}^n$ allows us to measure distances, while the notion of an inner product allows us to measure *angles* between vectors in $\mathbb{R}^n$. Indeed, let x and y be two vectors in $\mathbb{R}^n$ that have length 1, that is $\|x\| = \|y\| = 1$, then by the Cauchy-Schwarz inequality we have $|\langle x, y\rangle| \leq 1$ or equivalently

$$-1 \leq \langle x, y\rangle \leq 1.$$

We define, the *angle* between x and y to be the unique number θ in $[0, \pi]$ that satisfies $\cos(\theta) = \langle x, y\rangle$. In general, when the vectors x and y do not necessarily have length one we define the *angle* between x and y to be the unique number θ in $[0, \pi]$ that satisfies

$$\cos(\theta) = \frac{\langle x, y\rangle}{\|x\|\|y\|}.$$

Thus, $\langle x, y\rangle > 0$ holds if and only if the angle between x and y is sharp; $\langle x, y\rangle = 0$ holds if and only if x and y are perpendicular; and $\langle x, y\rangle < 0$ holds if and only if the angle between x and y is obtuse.

Exercise 25. Consider the set

$$C = \left\{x = (x_0, x_1, ..., x_n) \in \mathbb{R} \times \mathbb{R}^n : x_0 \geq \sqrt{x_1^2 + \cdots + x_n^2}\right\}.$$

The set C is called the *Minkowski's light cone* and plays an important role is the Einstein's theory of relativity. For any two vectors $x, y \in C$ define the *Lorentz inner product* between x and y by

$$[x, y] := x_0 y_0 - x_1 y_1 - x_2 y_2 - \cdots - x_n y_n,$$

and the *Lorenz norm*

$$[\![x]\!] := \sqrt{[x, x]} = \sqrt{x_0^2 - x_1^2 - x_2^2 - \cdots x_n^2}.$$

Show that for any two vectors x, y from the Minkowski's light cone C the following *reversed* Cauchy-Schwarz inequality holds

$$[\![x]\!][\![y]\!] \leq [x, y].$$

Equality holds if and only if $y_0 x_i = x_0 y_i$ for all $i = 1, ..., n$.

Hint: Imitate the proof of the original Cauchy-Schwatz inequality. Show first that the quadratic polynomial in t has two distinct real roots. Thus, the discriminant must be strictly positive.

2 Calculus review

The *open ball* in $\mathbb{R}^n$ centered at $a \in \mathbb{R}$ with radius $r > 0$ is the set defined by

$$B(a, r) := \{x \in \mathbb{R}^n : \|a - x\| < r\}.$$

That is, $B(a, r)$ is the set of all vectors x in $\mathbb{R}^n$ that are at a distance less than r from a. For example, in $\mathbb{R}$ the open ball $B(a, r)$ is the open interval $(a - r, a + r)$. In $\mathbb{R}^2$ the open ball $B(a, r)$ is the open disc centered at a with radius r. In $\mathbb{R}^3$ the open ball $B(a, r)$ is the open ball centered at a with radius r.

A set D in $\mathbb{R}^n$ is called *open* if for every point $a \in D$ there is an open ball centered at a that is entirely contained in D. For example, the whole space $\mathbb{R}^n$ and the empty set are open. The intersection of two open sets is an open set. The union of any number of open sets is an open set. Informally speaking, a set D is open if, whenever x is sufficiently close to a point from D then x is also in D.

Every open interval (a, b) is an open subset of $\mathbb{R}$. The set in $\mathbb{R}^2$ defined by $\{(x, y) \in \mathbb{R}^2 : xy > 1, x > 0\}$ is open. In general, if the definition of a set contains the usual arithmetic operations and inequalities that are *all* strict, then the set is open. For example, let $a \in \mathbb{R}^n$, then the set in $\mathbb{R}^n$ defined by $\{x \in \mathbb{R}^n : a_1x_1 + \cdots + a_nx_n < 0\}$ is open. The set of all vectors with strictly positive coordinates is open: $\mathbb{R}^n_{++} := \{x \in \mathbb{R}^n : x_1 > 0, ..., x_n > 0\}$. This set is called the *strictly positive* orthant.

A set D in $\mathbb{R}^n$ is called *closed* if its compliment $D^c := \mathbb{R}^n \setminus D$ is open. The intersection of any number of closed sets is a closed set. The union of two closed sets is a closed set. Usually, if the definition of a set contains the usual arithmetic operations, and the relationships $\geq, \leq, =$ then the set is closed. The whole space $\mathbb{R}^n$ and the empty set are both closed.

A set D in $\mathbb{R}^n$ is *bounded* if there is a ball $B(a, r)$ in $\mathbb{R}^n$ (with possibly very large radius) that contains D.

Let $f(x_1, ..., x_n)$ be a function defined on a subset D of $\mathbb{R}^n$. The point, $x_0 \in D$ is *global minimizer* of f on D if $f(x_0) \leq f(x)$ for all $x \in D$ and $f(x_0)$ is the *global minimum* of $f(x)$ on D. (Respectively, the point, $x_0 \in D$ is *global maximizer* of f on D if $f(x_0) \geq f(x)$ for all $x \in D$ and $f(x_0)$ is the *global maximum* of $f(x)$ on D.) The point, $x_0 \in D$ is a *local minimizer* of f on D if $f(x_0) \leq f(x)$ for all $x \in D \cap B(x_0, r)$ for some radius $r > 0$ and $f(x_0)$ is a *local minimum* of $f(x)$ on D. (Respectively, the point, $x_0 \in D$ is a *local maximizer* of f on D if $f(x_0) \geq f(x)$ for all $x \in D \cap B(x_0, r)$ for some radius $r > 0$ and $f(x_0)$ is a *local maximum* of $f(x)$ on D.) Note that if x_0 is a global (local) minimizer of f on D if and only if x_0 is a global (local) maximizer of $-f(x)$ on D. Thus, finding minimizers is equivalent to finding maximizers.

The reason why we are interested in closed and bounded sets is given by the next theorem.

Theorem 26 (The fundamental theorem of optimization)**.** *If $f(x)$ is a continuous function defined on a closed and bounded set D in $\mathbb{R}^n$ then there is a point $x^* \in D$ that is a global maximizer of $f(x)$ on D. (Also, there is a, possibly different, point that is a global minimizer of $f(x)$ on D.)*

The theorem fails if D is not closed and/or bounded. For example consider the following sets and functions on the real line. The set $D = [1, \infty)$ is closed but not bounded subset of $\mathbb{R}$ and the function $f(x) = 1/x$ doesn't have a minimum on D. Alternatively, the set $D = (0, 1]$ is bounded but not closed and the function $f(x) = 1/x$ doesn't have a maximum on D. If $D = (0, \infty)$, neither closed nor bounded set, then the function $f(x) = 1/x$ doesn't have any extremum on D. As a final example, the set $D = (-\pi/2, \pi/2)$ is bounded but not closed, and the function $f(x) = \tan(x)$ has no extremums on D.

The reason why we are interested in open subsets of $\mathbb{R}^n$ is given by the next theorem.

Theorem 27 (First order necessary conditions)**.** *Suppose that $f(x)$ is a differentiable function defined on an open subset D of $\mathbb{R}^n$. If $x^* \in D$ is a local minimizer (or a local maximizer) of $f(x)$ on D, then*

$$\frac{\partial f}{\partial x_i}(x^*) = 0 \qquad \text{for all } i = 1, ..., n.$$

Proof. Suppose that x^* is a local minimizer. (The proof when x^* is a local maximizer is analogous.) By definition the i-th partial derivative of f at x^* is

$$\frac{\partial f}{\partial x_i}(x^*) = \lim_{h \to 0} \frac{f(x_1^*, ..., x_i^* + h, ..., x_n^*) - f(x^*)}{h}.$$

Since D is open set, when h is close to 0 the point $(x_1^*, ..., x_i^* + h, ..., x_n^*)^T$ is in D, thus the function value $f(x_1^*, ..., x_i^* + h, ..., x_n^*)$ is defined. Since x^* is local minimizer, if h is close to 0 then $f(x^*) \leq f(x_1^*, ..., x_i^* + h, ..., x_n^*)$. Thus, when h approaches 0 with positive values we get

$$\frac{\partial f}{\partial x_i}(x^*) = \lim_{h \to 0^+} \frac{f(x_1^*, ..., x_i^* + h, ..., x_n^*) - f(x^*)}{h} \geq 0.$$

Alternatively, when h approaches 0 with negative values, we get

$$\frac{\partial f}{\partial x_i}(x^*) = \lim_{h \to 0^-} \frac{f(x_1^*, ..., x_i^* + h, ..., x_n^*) - f(x^*)}{h} \leq 0,$$

because the numerator is positive while the denominator is negative. Hence $\frac{\partial f}{\partial x_i}(x^*) = 0$. □

When $f(x)$ is a differentiable function defined on an open subset D of $\mathbb{R}^n$ the vector

$$\nabla f(x) := \Big(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), ..., \frac{\partial f}{\partial x_n}(x)\Big)^T$$

is called the *gradient* of f at x. A point $x^* \in D$ is called a *critical point* for $f(x)$ on D if $\nabla f(x^*) = 0$. Thus, Theorem 27 says that the local extremums of a differentiable function f are among its critical points. If the partial derivatives $\frac{\partial f}{\partial x_i}(x)$ are continuous in x on D then f is called *continuously differentiable* on D and we write $f \in C^1(D)$.

The following is an important theorem from Calculus. Given two points $x, y \in \mathbb{R}^n$, the *line segment*, $[x, y]$ joining x and y is defined to be the set

$$[x, y] := \{z \in \mathbb{R}^n \mid z = (1 - t)x + ty,\ t \in [0, 1]\}.$$

Theorem 28 (Mean value theorem)**.** *Suppose that $f(x)$ is continuously differentiable on an open set containing the line segment $[x, y]$. Then, there is a point $z \in [x, y]$ such that*

$$f(x) = f(y) + \langle \nabla f(z), x - y \rangle.$$

Sometimes the mean value theorem is also called first order Taylor expansion of f around y. Notice that when x is close to y, $x-y$ is close to 0, and thus the term $\langle \nabla f(z), x-y\rangle$ is close to zero. Thus, the mean value theorem says that when x is close to y $f(x)$ is close to $f(y)$, with error $f(x)-f(y)$ equal to $\langle \nabla f(z), x-y\rangle$.

When $f(x)$ is a twice differentiable function defined on an open subset D of $\mathbb{R}^n$ the matrix of all second partial derivatives

$$\nabla^2 f(x) := \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) & \vdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(x) \end{pmatrix}$$

is called the *Hessian* of f at x. If all second partial derivatives $\frac{\partial^2 f}{\partial x_i \partial x_j}(x)$ are continuous in x on D then f is called *twice continuously differentiable* on D and we write $f \in C^2(D)$.

Next is another standard result from Calculus.

Theorem 29. *If $f \in C^2(D)$, where D is an open subset of $\mathbb{R}^n$, then*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x) \quad \textit{for all } i, j = 1, ..., n.$$

In other words, when f is twice *continuously* differentiable, it doesn't matter in what order we differentiate to find the second partial derivatives. Note that in that case the Hessian $\nabla^2 f(x)$ is an $n \times n$ *symmetric* matrix. Recall that an $n \times n$ matrix A is called symmetric if $A = A^T$.

Theorem 30 (Second order Taylor theorem)**.** *Suppose that $f(x)$ is twice continuously differentiable on an open set containing the line segment $[x, y]$. Then, there is a point $z \in [x, y]$ such that*

$$\begin{aligned} f(x) &= f(y) + \langle \nabla f(y), (x-y)\rangle + \frac{1}{2}\langle \nabla^2 f(z)(x-y), (x-y)\rangle \\ &= f(y) + \nabla f(y)^T (x-y) + \frac{1}{2}(x-y)^T \nabla^2 f(z)(x-y). \end{aligned}$$

Notice that when x is close to y, $x-y$ is close to 0, and thus the term $\langle \nabla f(y), x-y\rangle$ is close to zero but the term $\frac{1}{2}(x-y)^T \nabla^2 f(z)(x-y)$ is *much closer* to zero since it contains the small factor $(x-y)$ twice. Thus, the second order Taylor theorem says that when x is close to y the value of the function $f(x)$ is close to the value of the affine in x function $f(y) + \nabla f(y)^T(x-y)$, where the error of the approximation is equal to $\frac{1}{2}(x-y)^T \nabla^2 f(z)(x-y)$.

2.1 A digression on symmetric matrices.

Suppose that A is an $n \times n$ symmetric matrix.

- A is called *positive semidefinite* matrix if $x^T A x \geq 0$ for all $x \in \mathbb{R}^n$.

- A is called *positive definite* matrix if $x^T Ax > 0$ for all $x \in \mathbb{R}^n$ s.t. $x \neq 0$.
- A is called *negative semidefinite* matrix if $x^T Ax \leq 0$ for all $x \in \mathbb{R}^n$.
- A is called *negative definite* matrix if $x^T Ax < 0$ for all $x \in \mathbb{R}^n$ s.t. $x \neq 0$.
- A is called *indefinite* matrix if there is a vector $y \in \mathbb{R}^n$ with $y^T Ay > 0$ and a vector $z \in \mathbb{R}^n$ with $z^T Az < 0$.

For example, note that the zero matrix is not indefinite. It is positive semidefinite, and also negative semidefinite. In fact, it is the only matrix that is both positive and negative semidefinite.

It is important to have an easy way to check if a matrix is positive or negative (semi-)definite. One criteria is given by the eigenvalues. The eigenvalues of a matrix A are the roots of

$$\det(A - \lambda I) = 0$$

considered as a polynomial (of degree n) in λ. Since each polynomial of degree n has n roots, the matrix A has n eigenvalues. It is important to keep in mind that since the matrix A is symmetric, all eigenvalues are real numbers. Denote them by $\lambda_1(A)$, $\lambda_2(A)$,...,$\lambda_n(A)$.

Theorem 31. *Suppose A is an $n \times n$ symmetric matrix. Then*

- *A is positive semidefinite if and only if $\lambda_i(A) \geq 0$ for all $i = 1, ..., n$.*
- *A is positive definite if and only if $\lambda_i(A) > 0$ for all $i = 1, ..., n$.*
- *A is negative semidefinite if and only if $\lambda_i(A) \leq 0$ for all $i = 1, ..., n$.*
- *A is negative definite if and only if $\lambda_i(A) < 0$ for all $i = 1, ..., n$.*
- *A is indefinite if and only if $\lambda_i(A) < 0$ and $\lambda_j(A) > 0$ for some distinct i, j.*

Suppose that A is an $n \times n$ symmetric matrix. Define Δ_k to be the determinant of the $k \times k$-submatrix in the upper left-hand corner of A, for $k = 1, 2, ..., n$. The determinant Δ_k is called the *k-th principal minor* of A.

Theorem 32. *If A is an $n \times n$ symmetric matrix and if Δ_k is the k-th principal minor of A for $k = 1, 2, ..., n$, then:*

(i) *A is positive definite if and only if $\Delta_k > 0$ for $k = 1, 2, ..., n$.*

(ii) *A is negative definite if and only if $(-1)^k \Delta_k > 0$ for $k = 1, 2, ..., n$. That is, $\Delta_1 < 0$, $\Delta_2 > 0$, $\Delta_3 < 0$, and so on.*

Similar theorem doesn't hold for positive (negative) semidefinite matrices.

Now, we have three possible ways to check if a matrix is positive (or negative) definite: we can either check if it satisfies the definition, or apply Theorem 31, or apply Theorem 32. One cannot say in advance which criteria will be better in a given situation.

3 Second order conditions

Theorem 33 (Global second order sufficient conditions)**.** *Suppose that $f(x)$ is twice continuously differentiable on $\mathbb{R}^n$. Suppose that x^* is a critical point of $f(x)$. Then*

(i) If the Hessian $\nabla^2 f(x)$ is positive semidefinite for every $x \in \mathbb{R}^n$, then x^* is a global minimizer for $f(x)$.

(ii) If the Hessian $\nabla^2 f(x)$ is negative semidefinite for every $x \in \mathbb{R}^n$, then x^* is a global maximizer for $f(x)$.

Proof. We show only the first part of the theorem, since the second one is analogous. Since x^* is a critical point of $f(x)$, the first partial derivatives of $f(x)$ are zero, that is $\nabla f(x^*) = 0$. Let x be any other point in $\mathbb{R}^n$ other than x^*, then by the second order Taylor expansion, there is a point $z \in [x, x^*]$ such that

$$\begin{aligned} f(x) &= f(x^*) + \nabla f(x^*)^T(x - x^*) + \frac{1}{2}(x - x^*)^T \nabla^2 f(z)(x - x^*) \\ &= f(x^*) + \frac{1}{2}(x - x^*)^T \nabla^2 f(z)(x - x^*). \end{aligned}$$

If the Hessian $\nabla^2 f(z)$ is positive semidefinite for every $z \in \mathbb{R}^n$ then

$$f(x) - f(x^*) = \frac{1}{2}(x - x^*)^T \nabla^2 f(z)(x - x^*) \geq 0.$$

Thus, $f(x) \geq f(x^*)$ for all $x \in \mathbb{R}^n$, showing that x^* is a global minimizer. □

Theorem 34 (Local second order sufficient conditions)**.** *Suppose that $f(x)$ is twice continuously differentiable on an open set D of $\mathbb{R}^n$. Suppose that x^* is a critical point of $f(x)$.*

(i) If the Hessian $\nabla^2 f(x^*)$ is positive definite, then x^* a local minimizer for $f(x)$.

(ii) If the Hessian $\nabla^2 f(x^*)$ is negative definite, then x^* a local maximizer for $f(x)$.

Proof. We prove the first part of the theorem since the second part is analogous. The key is the following

observation. If $g(x)$ is a function on $\mathbb{R}^n$ that is continuous at the point x^* and $g(x^*) > 0$ then there exists a number $r > 0$ such that $g(x) > 0$ for all $x \in B(x^*, r)$.

Define $\Delta_k(x)$ to be the k-th principal minor of the Hessian $\nabla^2 f(x)$. By hypothesis, and by Theorem 32, we know that $\Delta_k(x^*) > 0$ for $k = 1, 2, ..., n$. Now because the second partial derivatives of $f(x)$ are continuous, each $\Delta_k(x)$ is a continuous function of x. Since $\Delta_k(x^*) > 0$ from continuity it follows that there exists a number $r_k > 0$ (one for each $k = 1, 2, ..., n$) such that $\Delta_k(x) > 0$ if $x \in B(x^*, r_k)$. Set $r = \min\{r_1, r_2, ..., r_k\}$ and observe that for all $k = 1, 2, ..., n$ we have $\Delta_k(x) > 0$ if $x \in B(x^*, r)$. Therefore by Theorem 32 the Hessian $\nabla^2 f(x)$ is positive definite for all $x \in B(x^*, r)$.

Choose the radius r even smaller if necessary, so that $B(x^*, r) \subset D$. This can be done because D is an open set.

Now take any $x \in B(x^*, r)$. By the second order Taylor expansion there is a $z \in [x, x^*]$ such that

$$\begin{aligned} f(x) &= f(x^*) + \nabla f(x^*)^T(x - x^*) + \frac{1}{2}(x - x^*)^T \nabla^2 f(z)(x - x^*) \\ &= f(x^*) + \frac{1}{2}(x - x^*)^T \nabla^2 f(z)(x - x^*), \end{aligned}$$

where we used that x^* is a critical point. Observe that $z \in B(x^*, r)$. Hence, $\nabla^2 f(z)$ is positive definite, that is $\frac{1}{2}(x - x^*)^T \nabla^2 f(z)(x - x^*) \geq 0$. Therefore $f(x) \geq f(x^*)$ for every $x \in B(x^*, r)$. Since $B(x^*, r) \subset D$ this shows that x^* is a local minimizer. $\square$

Example 35. Find all critical points of the function. Explain which critical point is global/local minimizer/maximizer. If a critical point is none of the above explain why.

a) $f(x_1, x_2) = x_1^3 - 12x_1x_2 + 8x_2^3$,

b) $f(x_1, x_2) = x_1^5 - x_1x_2^6$

Solution: a) We compute the gradient and the Hessian of f:

$$\nabla f(x_1, x_2) = \begin{pmatrix} 3x_1^2 - 12x_2 \\ -12x_1 + 24x_2^2 \end{pmatrix}, \quad \nabla^2 f(x_1, x_2) = \begin{pmatrix} 6x_1 & -12 \\ -12 & 48x_2 \end{pmatrix}.$$

The critical points are $(2, 1)$ and $(0, 0)$. The Hessian at the point $(2, 1)$ is

$$\nabla^2 f(2, 1) = \begin{pmatrix} 12 & -12 \\ -12 & 48 \end{pmatrix},$$

with $\Delta_1 = 12$, $\Delta_2 = 432$. It follows that the critical point $(2, 1)$ is a local minimizer.

The Hessian at the point $(0, 0)$ is

$$\nabla^2 f(0, 0) = \begin{pmatrix} 0 & -12 \\ -12 & 0 \end{pmatrix},$$

which is indefinite. Therefore $(0, 0)$ is a saddle point as it will be explained shortly in Subsection 3.1.

b) We compute the gradient and the Hessian of f:

$$\nabla f(x_1, x_2) = \begin{pmatrix} 5x_1^4 - x_2^6 \\ -6x_1x_2^5 \end{pmatrix}, \quad \nabla^2 f(x_1, x_2) = \begin{pmatrix} 20x_1^3 & -6x_2^5 \\ -6x_2^5 & 30x_1x_2^4 \end{pmatrix}.$$

The only critical point is $(0, 0)$. At it, the Hessian is

$$\nabla^2 f(0, 0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

which is positive semidefinite. We do not have a criteria that tells us if $(0, 0)$ is local extremum in this case. That is why we have to be more inventive. Consider the points $\{(t, 0) \,|\, t \in \mathbb{R}\}$. The function values at these points are

$$f(t, 0) = t^5.$$

For this function $t = 0$ is neither local min nor local max, therefore the same is true for $(0, 0)$.

3.1 Indefinite Hessian

Suppose $f \in C^2(D)$ and $x^* \in D$ is a critical point of f, that is, $\nabla f(x^*) = 0$. Suppose that $\nabla^2 f(x^*)$ is indefinite matrix, this means that there are nonzero vectors y and z in $\mathbb{R}^n$ such that

$$y^T \nabla^2 f(x^*) y > 0 \quad \text{and} \quad z^T \nabla^2 f(x^*) z < 0.$$

Define the functions

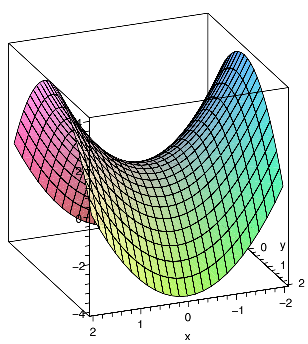
$$Y(t) := f(x^* + ty) \quad \text{and} \quad Z(t) := f(x^* + tz) \quad \text{for } t \text{ close to } 0.$$

Using the chain rule we can compute the first two derivatives of $Y(t)$ and $Z(t)$ at $t = 0$.

$$\begin{aligned} Y'(t) &= y^T \nabla f(x^* + ty), & Z'(t) &= z^T \nabla f(x^* + tz), \\ Y''(t) &= y^T \nabla^2 f(x^* + ty) y, & Z''(t) &= z^T \nabla^2 f(x^* + tz) z. \end{aligned}$$

Thus, $Y'(0) = Z'(0) = 0$ and $Y''(0) = y^T \nabla^2 f(x^*) y > 0$, $Z''(0) = z^T \nabla^2 f(x^*) z < 0$. Therefore, $t = 0$ is a local minimizer for $Y(t)$ and a local maximizer for $Z(t)$. Thus, if we move from x^* in the direction of y or $-y$ the value of $f(x)$ increases, while if we move from x^* in the direction of z or $-z$, the value of $f(x)$ decreases. This is the reason for calling the critical point x^* a *saddle point* for $f(x)$. In other words, x^* is a saddle point for $f(x)$ if $\nabla f(x^*) = 0$ and the Hessian $\nabla^2 f(x^*)$ is indefinite.

For example, on the figure below, you see the graph of the function $f(x, y) = x^2 - y^2$ for which the point $(0, 0)$ is a saddle point. In this example, $y = (1, 0)$ and $z = (0, 1)$.



Exercise 36. Find and classify the extremums, if any, of the following functions:
a) $f(x, y) = y^2 + x^2 y + x^4$;
b) $f(x, y) = x^2 + y^2 + x + y + xy$;
c) $f(x, y) = (x - 1)^4 + (x - y)^4$;
d) $f(x, y) = y^2 - x^3$.

Exercise 37. Find the point(s) where f achieves its global minimum. If f has no global minimum explain why.

a) $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - x_1 x_2 + x_2 x_3 - x_1 x_3$;

b) $f(x, y) = e^{x-y} + e^{y-x}$;

c) $f(x, y, z) = e^{x-y} + e^{y-x} + e^{x^2} + z^2$.

Exercise 38. Locate all local/global maximizers, minimizers and saddle points of the functions
(a) $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - 4x_1x_2$;
(b) $f(x, y, z) = e^x + e^y + e^z + 2e^{-x-y-z}$.

Exercise 39. Let A be an $n \times n$ symmetric positive semidefinite matrix, let $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Find the optimal value and the minimizer and/or maximizer of $f(x) = x^T Ax + b^T x + c$.

4 Lagrange multipliers

It turns out that some constrained optimization problems can be easily turned into unconstrained ones. We explain how that is done now.

Suppose that $f(x)$, $h_1(x)$,...,$h_p(x)$ are continuously differentiable functions defined on an open set D of $\mathbb{R}^n$. We are interested in solving the optimization *program*.

$$(P): \begin{cases} \text{min/max} & f(x) \\ \text{subject to} & h_1(x) = 0 \\ & h_2(x) = 0 \\ & \vdots \\ & h_p(x) = 0 \\ & x \in D. \end{cases}$$

This problem has n variables $x = (x_1, ..., x_n)$, called *decision variables*. The function $f(x)$ is called the *objective function*. The equations $h_1(x) = 0$,..., $h_p(x) = 0$ are called *constraints*. That is, there are p equality constraints. The set

$$FR := \{x \in D : h_1(x) = 0, ..., h_p(x) = 0\}$$

is called the *feasible region*. The goal of the optimization program is two-fold: 1) to find the points $x^* \in FR$ yielding the smallest (resp. largest) value of $f(x)$ on FR; and 2) to calculate $f(x^*)$. In other words, we want to find the global minimum (resp. maximum) of $f(x)$ on FR and the points in FR that achieve it. The points (possibly more than one) that minimize (resp. maximize) $f(x)$ on FR are called *minimizers* (resp. *mzaximizers*) or just *optimal solutions*. If x^* is a maximizer (or minimizer) then $f(x^*)$ is called *optimal value*. Note that even though there may be more than one maximizers (or minimizers) there is always only one optimal value.

It turns out that solving the optimization problem above reduces to finding the critical points of the *Lagrangian function*:

$$L(x, \lambda) := f(x) - \sum_{i=1}^{p} \lambda_i h_i(x),$$

where $\lambda = (\lambda_1, ..., \lambda_p)$ is a vector of new variables called *Lagrange multipliers*. Note that the Lagrangian is a function on $n + p$ variables. Its gradient is given by

$$\nabla L(x, \lambda) = \begin{pmatrix} \nabla f(x) - \sum_{i=1}^{p} \lambda_i \nabla h_i(x) \\ -h_1(x) \\ -h_2(x) \\ \cdots \\ -h_p(x) \end{pmatrix},$$

where the column vector $\nabla f(x) - \sum_{i=1}^{p} \lambda_i \nabla h_i(x)$ contains the partial derivatives with respect to $x_1, ..., x_n$, while $-h_1(x)$,...,$-h_p(x)$ are the partial derivatives with respect to λ_1,...,λ_p.

Theorem 40 (Lagrange multiplier theorem - necessary conditions)**.** *If $x^* \in FR$ is an optimal solution of (P) and the vectors $\nabla h_1(x^*)$,...,$\nabla h_p(x^*)$ are linearly independent, then there are (unique) numbers $\lambda^* = (\lambda_1^*, ..., \lambda_p^*)$ such that*

$$\nabla L(x^*, \lambda^*) = 0. \tag{2}$$

Notice that if $p = 0$, that is, if there are no constraints, then $L(x, \lambda) = f(x)$, $FR = D$, and the theorem says that if $x^* \in D$ is a maximizer (or minimizer) of $f(x)$ then $\nabla L(x^*, \lambda^*) = \nabla f(x^*) = 0$. That is, if there are no constraints, Theorem 40 reduces to Theorem 27.

Equation (2) in fact defines a system of $n+p$ equations in $n+p$ unknowns. Solving this system gives all vectors x^* some of which are the optimal solutions of (P). Usually, the difficulty is that the system of equations is non-linear and thus, often difficult to solve.

Note that if $p > n$ then the vectors $\nabla h_1(x^*)$,...,$\nabla h_p(x^*)$ are always linearly dependent (since each gradient is an n-dimensional vector and there are p of them), thus the theorem can not be applied. Remember, the theorem only finds those candidates x^* for optimal solutions at which $\nabla h_1(x^*)$,...,$\nabla h_p(x^*)$ are linearly independent. Sometimes there are optimal solutions x^* at which $\nabla h_1(x^*)$,...,$\nabla h_p(x^*)$ are linearly dependent, those have to be found separately, by other means.

Example 41. Find the maximum of the function $f(x) = x_1 + 3x_2 - 2x_3$ on the sphere $x_1^2 + x_2^2 + x_3^2 = 14$.

Solution: Here $D = \mathbb{R}^3$ the whole space, and there is one constraint: $h_1(x) = 0$, where $h_1(x) = x_1^2 + x_2^2 + x_3^2 - 14$. The feasible region is the sphere centered at the origin $(0, 0, 0)$ with radius $\sqrt{14}$.

First, let us see if there are points x^* in the feasible region for which the vector $\nabla h_1(x^*) = (2x_1^*, 2x_2^*, 2x_3^*)$ is linearly dependent. Recall that one vector is linearly dependent if and only if it is the zero vector. Note that $\nabla h_1(x^*) = 0$ when $x^* = 0$. But $x^* = 0$ doesn't belong to the feasible region so we conclude that for every $x^* \in FR$ the vector $\nabla h_1(x^*)$ is linearly independent.

The Lagrangian is

$$L(x_1, x_2, x_3, \lambda) = x_1 + 3x_2 - 2x_3 - \lambda(x_1^2 + x_2^2 + x_3^2 - 14).$$

If $x^* \in FR$ is an optimal solution (necessarily $\nabla h_1(x^*)$ is linearly independent), then there is a number λ^* such that $\nabla L(x^*, \lambda^*) = 0$. Written out the system of equations is

$$\begin{aligned} 1 - 2\lambda^* x_1^* &= 0 \\ 3 - 2\lambda^* x_2^* &= 0 \\ -2 - 2\lambda^* x_3^* &= 0 \\ (x_1^*)^2 + (x_2^*)^2 + (x_3^*)^2 &= 14. \end{aligned}$$

First we solve this system. From the first three equations we find

$$x_1^* = \frac{1}{2\lambda^*}, \quad x_2^* = \frac{3}{2\lambda^*}, \quad x_3^* = -\frac{2}{2\lambda^*}.$$

Substituting into the last equation we find two possible values for $\lambda^* = \pm 1/2$ and thus two possible solutions

$$(x_1^*, x_2^*, x_3^*, \lambda^*) = (1, 3, -2, -\frac{1}{2}), \quad (x_1^*, x_2^*, x_3^*, \lambda^*) = (-1, -3, 2, \frac{1}{2}).$$

Thus, by Theorem 40 both $(x_1^*, x_2^*, x_3^*) = (1, 3, -2)$ and $(x_1^*, x_2^*, x_3^*) = (-1, -3, 2)$ are *candidates* for optimal solution.

Third, argue from scratch that the optimization problem must have a maximizer. In this case the feasible region is a sphere, that is a closed and bounded set. The objective function is continuous. By the fundamental theorem of optimization, Theorem 26, every continuous function has a maximizer (resp. minimizer) over a closed and bounded set.

Finally, the fourth step is to compute the objective value at each candidate solution. The objective value of the first solution is 14, while that of the second solution is -14. Since the problem must have a maximizer (and a minimizer) $(x_1^*, x_2^*, x_3^*) = (1, 3, -2)$ must be the maximizer with maximal value 14, while $(x_1^*, x_2^*, x_3^*) = (-1, -3, 2)$ must be the minimizer with minimal value -14.

Example 42. Solve the optimization problem.

$$(P): \begin{cases} \text{minimize} & x_1 + x_2 \\ \text{subject to} & x_1^2 + x_2^2 - 2x_1 = 0 \\ & x_1^2 + x_2^2 - 4x_1 = 0 \\ & x \in \mathbb{R}^2. \end{cases}$$

The Lagrangian is

$$L(x_1, x_2, \lambda_1, \lambda_2) = x_1 + x_2 - \lambda_1(x_1^2 + x_2^2 - 2x_1) - \lambda_2(x_1^2 + x_2^2 - 4x_1),$$

where $f(x) = x_1 + x_2$, $h_1(x) = x_1^2 + x_2^2 - 2x_1$, and $h_2(x) = x_1^2 + x_2^2 - 4x_1$. If x^* is an optimal solution, for which $\nabla h_1(x^*)$, $\nabla h_2(x^*)$ are linearly independent, then there are numbers λ_1^*, λ_2^* such that $\nabla L(x^*, \lambda^*) = 0$. Written out the system of equations is

$$\begin{aligned} 1 - \lambda_1^*(2x_1^* - 2) - \lambda_2^*(2x_1^* - 4) &= 0 \\ 1 - \lambda_1(2x_2^*) - \lambda_2^*(2x_2^*) &= 0 \\ (x_1^*)^2 + (x_2^*)^2 - 2x_1^* &= 0 \\ (x_1^*)^2 + (x_2^*)^2 - 4x_1^* &= 0. \end{aligned}$$

Subtracting the fourth equation from the third we get $x_1^* = 0$ and then from any of the last two equations that $x_2^* = 0$. Substituting into the first two equations we see that there are no values of λ_1^* and λ_2^* that satisfy the first two equations. Thus, the system $\nabla L(x^*, \lambda^*) = 0$ has no solution.

This doesn't yet mean that the optimization problem has no optimal solutions! It means that the optimization problem doesn't have optimal solutions x^* for which the gradients $\nabla h_1(x^*)$ and $\nabla h_2(x^*)$ are linearly independent!

Let us see now if there are points x^* in the feasible region for which the vectors $\nabla h_1(x^*)$ and $\nabla h_2(x^*)$ are linearly dependent. (Calculate that $\nabla h_1(x^*) = (2x_1^* - 2, 2x_2^*)$ and $\nabla h_2(x^*) =$

$(2x_1^* - 4, 2x_2^*)$.) To find such points x^* ask if there is a *nonzero* vector $\beta^* = (\beta_1^*, \beta_2^*)$ such that (x^*, β^*) solves the system

$$\begin{aligned} \beta_1^* \nabla h_1(x^*) + \beta_2^* \nabla h_2(x^*) &= 0 \\ h_1(x^*) &= 0 \\ h_2(x^*) &= 0. \end{aligned}$$

Written out the system is

$$\begin{aligned} \beta_1^*(2x_1^* - 2) + \beta_2^*(2x_1^* - 4) &= 0 \\ \beta_1^*(2x_2^*) + \beta_2^*(2x_1^*) &= 0 \\ (x_1^*)^2 + (x_2^*)^2 - 2x_1^* &= 0 \\ (x_1^*)^2 + (x_2^*)^2 - 4x_1^* &= 0. \end{aligned}$$

This system looks very much like the one we already solved (this is always the case, why?). It is easy to see that the only x^* for which there is a nonzero $\beta^* = (\beta_1^*, \beta_2^*)$ so that (x^*, β^*) is a solution to the system is $x^* = (0, 0)$.

By Theorem 26, the optimization problem must have an optimal solution because we are minimizing a continuous function over a closed bounded feasible region. (Why?)

We conclude that $x^* = (0, 0)$ is an optimal solution with optimal value $f(x^*) = 0$!

Let us summarize the procedure for solving optimization problems of the form (P).

(i) Find all points x^* in the feasible region FR such that the vectors $\nabla h_1(x^*)$,...,$\nabla h_p(x^*)$ are linearly dependent. Denote this set by S_1. This is the set of points to which Theorem 40 cannot be applied. They must be investigated separately later.

(ii) Find all solutions x^* of the system $\nabla L(x^*, \lambda^*) = 0$. Check if the vectors $\nabla h_1(x^*)$,...,$\nabla h_p(x^*)$ are linearly independent for each solution. Discard those x^*'s for which this condition fails. Denote by S_2 the set of remaining solutions.

(iii) Argue from general principles that the problem has to have an optimal solution. Most of the times the objective function is continuous and the feasible region is closed and bounded set.

(iv) Calculate the objective value $f(x^*)$ for all x^*'s in the set $S_1 \cup S_2$. The x^* that gives the largest value $f(x^*)$ is the maximizer, while the x^* that gives the smallest value $f(x^*)$ is the minimizer. (That is, the Lagrange multimplier procedure simultaneously finds both the maximizers and the minimizers.)

Note that it doesn't matter in what order you perform the first three steps in the above procedure as long as they are all completed before the last step.

4.1 Related results

We conclude this section with two more related results. A function $h(x_1, x_2, ..., x_n)$ is called *linear*[1] if it has the form

$$h(x_1, x_2, ..., x_n) = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n - \alpha,$$

[1] Technically, this function is affine, but let's not complicate matters.

for some vector $a = (a_1, a_2, ..., a_n)$ and a scalar α. For example, $x_1 - 3.4x_2 + 2x_3 - 0.333$ is a linear function. Note also that every linear function can be written concisely as $\langle a, x\rangle - \alpha$.

Theorem 43 (Lagrange multiplier theorem - linear constraints)**.** *If $x^* \in FR$ is a solution of (P) and the functions $h_1(x)$,...,$h_p(x)$ are all linear, then there are (possibly not unique) numbers $\lambda^* = (\lambda_1^*, ..., \lambda_p^*)$ such that*

$$\nabla L(x^*, \lambda^*) = 0. \tag{3}$$

This result is a significant simplification of Theorem 40. In order to solve an optimization problem with linear constraints all we have to do is

(i) Find all solutions x^* of the system $\nabla L(x^*, \lambda^*) = 0$.

(ii) Argue from general principles that the problem has to have an optimal solution.

(iii) Calculate the objective value $f(x^*)$ for all solutions x^* found in step (i). The x^* that gives the largest value $f(x^*)$ is the maximizer, while the x^* that gives the smallest value $f(x^*)$ is the minimizer.

In the case when the functions $f(x)$, $h_1(x)$,...,$h_p(x)$ are twice continuously differentiable, we may have a criteria distinguishing between minimizers and maximizers.

Theorem 44 (Lagrange multiplier theorem - sufficient conditions)**.** *If $x^* \in \mathbb{R}^n$ and $\lambda^* \in \mathbb{R}^p$ satisfy*

$$\nabla L(x^*, \lambda^*) = 0$$

and

$$y^T \Big(\nabla^2 f(x^*) - \sum_{i=1}^{p} \lambda_i^* \nabla^2 h_i(x^*)\Big) y > 0, \tag{4}$$

for all nonzero vectors y that are perpendicular to $\nabla h_1(x^*)$,...,$\nabla h_p(x^*)$, then x^* is a local minimizer of $f(x)$ on FR.

If the inequality in (4) is $<$ then x^* is a local maximizer. If (4) holds with equality for some nonzero vector y perpendicular to $\nabla h_1(x^*)$,...,$\nabla h_p(x^*)$ then the criteria fails and we have to find alternative ways to deal with the problem.

Note that when there are no constraints, that is $p = 0$ then this theorem becomes identical to Theorem 34.

Armed with this theorem, we do not have to justify from general principles any more that the optimization problem has an optimal solution. One drawback is that it cannot tell when a local extremum is global. Another drawback is that the criteria sometimes fails to give any information.

4.2 Exercises

Exercise 45. Find the minimum value of $x_1^2+3x_2^2+2x_3^2$ subject to the condition $2x_1+3x_2+4x_3 = 15$.

Exercise 46. Find the minimum value of $2x_1^2+x_2^2+2x_3^2$ subject to the condition $2x_1+3x_2-2x_3 = 13$.

Exercise 47. Minimize the function $-x_1x_2 - x_1x_3 - x_2x_3$ subject to the condition $x_1+x_2+x_3 = 3$.

Exercise 48. Find the minimum value of $x_1^2+x_2^2+x_3^2$ subject to the conditions $2x_1+2x_2+x_3 = -9$ and $2x_1 - x_2 - 2x_3 = 18$.

Exercise 49. Find the minimum value of $4x_1^2+2x_2^2+3x_3^2$ subject to the conditions $x_1+2x_2+3x_3 = 9$ and $4x_1 - 2x_2 + x_3 = -19$.

Exercise 50. Find the minimum value of $x_1^2 + x_2^2 + x_3^2 + x_4^2$ subject to the condition $2x_1 + x_2 - x_3 - 2x_4 = 5$.

Exercise 51. Find the minimum value of $x_1^2+x_2^2+x_3^2+x_4^2$ subject to the conditions $x_1-x_2+x_3+x_4 = 4$ and $x_1 + x_2 - x_3 + x_4 = -6$.

Exercise 52. Find the points on the curve $4x_1^2 + 4x_1x_2 + x_2^2 = 25$ that are nearest to the origin.

Exercise 53. Find the points on the curve $7x_1^2 + 6x_1x_2 + 2x_2^2 = 25$ that are nearest to the origin.

Exercise 54. Find the points on the curve $x_1^4 + 3x_1x_2 + x_2^4 = 2$ that are farthest from the origin.

Exercise 55. Let $b_1, ..., b_k$ be positive numbers. Find the maximum value of $\sum_{i=1}^k b_i x_i$ subject to the condition $\sum_{i=1}^k x_i^2 = 1$.

Exercise 56. Let A be an $n \times n$ positive definite symmetric matrix. Find the maximum value of $\|x\|$ subject to the condition $x^T Ax = 1$.

Exercise 57. Let A be an $n \times n$ positive definite symmetric matrix. Find the maximum value of $x^T Ax$ subject to the condition $\|x\| = 1$.

Exercise 58. Minimize the function $x_1 + x_2 + \cdots + x_n$ subject to the conditions $x_1x_2 \cdots x_n = 1$, $x_i > 0$ for $i = 1, ..., n$.

Exercise 59. Minimize the function $x_1^2 + x_2^2 + x_3^2$ subject to $x_1^2 + x_2^2/4 + x_3^2/9 = 1$.

Exercise 60. Determine all maximizers and minimizers of the function

$$f(x_1, x_2, x_2) = \frac{1}{x_1^2 + x_2^2 + x_3^2}$$

subject to

$$h_1(x_1, x_2, x_2) = 1 - x_1^2 - 2x_2^2 - 3x_3^2 = 0,$$
$$h_2(x_1, x_2, x_2) = x_1 + 2x_2 + x_3 = 0.$$

Exercise 61. Find the extreme values of $x^2 + 2y^2$ subject to $x^2 + y^2 \leq 1$.

Exercise 62. Find the points on the sphere $x^2 + y^2 + z^2 = 4$ that are closest to and farthest from the point $(3, 1, -1)$.

Exercise 63. A rectnagular box without a lid is made from 12 square meters of cardboard. Find the maximum volume of such a box.

Exercise 64. Find the maximizers and the minimizers of the function subject to the given constraints.
a) $x^2 - y^2$ s.t. $x^2 + y^2 = 1$
b) $4x + 6y$ s.t. $x^2 + y^2 = 13$
c) x^2y s.t. $x^2 + 2y^2 = 6$
d) $x^2 + y^2$ s.t. $x^4 + y^4 = 1$
e) $2x + 6y + 10z$ s.t. $x^2 + y^2 + z^2 = 35$
f) $8x - 4z$ s.t. $x^2 + 10y^2 + z^2 = 5$
g) xyz s.t. $x^2 + 2y^2 + 3z^2 = 6$
h) $x^2y^2z^2$ s.t. $x^2 + y^2 + z^2 = 1$
i) $x + 2y$ s.t. $x + y + z = 1$ and $y^2 + z^2 = 4$
j) $3x - y - 3z$ s.t. $x + y - z = 0$ and $x^2 + 2z^2 = 1$
k) $yz + xy$ s.t. $xy = 1$ and $y^2 + z^2 = 1$.

Exercise 65. Find the maximizers and the minimizers of the function subject to the given constraints.
a) $2x^2 + 3y^2 - 4x - 5$ s.t. $x^2 + y^2 \leq 16$
b) e^{-xy} s.t. $x^2 + 4y^2 \leq 1$.

Exercise 66. Find the maximizers and the minimizers of the function subject to the given constraints.
a) ye^{x-z} s.t $9x^2 + 4y^2 + 36z^2 = 36$ and $xy + yz = 1$
b) $x + y + z$ s.t. $x^2 - y^2 = z$ and $x^2 + y^2 = 4$.

Exercise 67. Find the shortest distance from the point $(0, b)$ on the y-axis to the parabola $x^2 - 4y = 0$.

Exercise 68. a) Find the shortest distance from the point $(a_1, a_2, a_3) \in \mathbb{R}^3$ to the plane $b_1x_2 + b_2x_2 + b_3x_3 + b_0 = 0$.
b) Find the point on the line of intersection of the two planes

$$a_1x_2 + a_2x_2 + a_3x_3 + a_0 = 0$$

and

$$b_1x_2 + b_2x_2 + b_3x_3 + b_0 = 0$$

which is nearest to the origin $O = (0, 0, 0)$.

Exercise 69. Find the local extremum of $x_1^k + \cdots + x_n^k$ subject to the condition $x_1 + \cdots + x_n = a$.

Exercise 70. Find the extremums of $x_1^2 + x_2^2$ subject to $x_1^2 + x_3^2 + x_4^2 = 4$ and $x_2^2 + 2x_3^2 + 3x_4^2 = 9$.

Exercise 71. Find the extremums of $x_1^2 + x_2^2$ subject to $x_1^2 + x_3^2 + x_4^2 = 4$ and $x_2^2 + 2x_3^2 + 3x_4^2 = 9$.

Exercise 72.* Find the extremums of $x^3 + 2xyz - z^2$ subject to $x^2 + y^2 + z^2 = 1$.

Exercise 73.* Show that the extreme values of $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^3$ subject to

$$\sum_{j=1}^{3} \sum_{i=1}^{3} a_{ij} x_i x_j = 1 \quad (a_{ij} = a_{ji})$$

and

$$b_1 x_2 + b_2 x_2 + b_3 x_3 = 0 \quad (b_1, b_2, b_3) \neq (0, 0, 0)$$

are t_1^{-1}, t_2^{-1}, where t_1 and t_2 are the roots of the equation

$$\begin{vmatrix} b_1 & b_2 & b_3 & 0 \\ a_{11} - t & a_{12} & a_{13} & b_1 \\ a_{21} & a_{22} - t & a_{23} & b_2 \\ a_{31} & a_{32} & a_{33} - t & b_3 \end{vmatrix} = 0.$$

Show that this is a quadratic equation in t and give a geometric argument to explain why the roots t_1, t_2 are real and positive.

Exercise 74.*[Hadamard inequality] Let X be an $n \times n$ real matrix and let $X_i = (x_{i1}, x_{i2}, ..., x_{in})$ be its i-th row. Show the inequality

$$\det X \leq \|X_1\| \cdots \|X_n\|.$$

Equality holds if and only if the matrix X is positive definite diagonal matrix. Hint: Prove this by treating $\det X$ as a function on n^2 variables subject to n constraints.

Exercise 75.* Calculate the function and plot its graph

$$\Phi(\lambda) = \min_{(x_1, x_2) \in \mathbb{R}^2} \left\{ e^{x_2} + \lambda \Big(\sqrt{x_1^2 + x_2^2} - x_1 \Big) \right\},$$

where $\lambda \in \mathbb{R}$. Hint: no theory applies here. You have to be clever and argue from general principles.