M2
Continuous Optimization
P.-A. Zitt

Université Gustave Eiffel
Année 2024–2025

Constrained Optimization

Exercise 1 — KKT. Let $\Omega = [0,4] \times [0,4]$, let $f : \mathbb{R}^2 \to \mathbb{R}$ denote the (Euclidean) distance from $(1,1)$, and let $g = -f$. The goal is to understand the KKT conditions on this simple example.

1. Draw a picture and find the minimum and maximum of f on Ω.
2. Suppose we wish to minimize f on Ω. Write the problem as a constrained optimization problem with linear constraints. Check that the constraints are always qualified.
3. Show that there is no KKT point on the left edge $\{x_1 = 0\}$. Deduce that there is only one KKT point in Ω.
   min $-g$.
4. We now wish to maximize f on Ω. Show that there are now nine different points satisfying the KKT equations (start by looking at the left edge). For each point, say if if it a local minimum, a local maximum, or a point of a different nature, for $f_{|\Omega}$.

Exercise 2 — Constraints qualification. Let $\Omega = \{x \in \mathbb{R}^2, 0 \le x_2 \le x_1^2\}$, and let $f : x \mapsto -x_2$.

1. Are the constraints qualified at x_*? $(0,0)$
2. Compare the tangent space (directions p that may appear as limits of $(x^{(k)} - x_*)/\|x^{(k)} - x_*\|$ for a sequence of feasible points $x^{(k)}$ going to x_*) and the space of directions that are orthogonal to the gradients of all the constraints.
3. Show that the KKT equations are satisfied at x_*.
4. Is x_* a local minimum?

Exercise 3 — LP. Let $c_1 \in \mathbb{R}^{n_1}$, $c_2 \in \mathbb{R}^{n_2}$, l and $u \in \mathbb{R}^{n_2}$, $A_1 \in \mathcal{M}_{p_1,n_1}$, $A_2 \in \mathcal{M}_{p_2,n_1}$, $B_2 \in \mathcal{M}_{p_2,n_2}$, $b_1 \in \mathbb{R}^{p_1}$ and $b_2 \in \mathbb{R}^{p_2}$ be given. Consider the problem of maximizing $c_1^\top x + c_2^\top x_2$, subject to

$$A_1 x_1 = b_1,\ A_2 x_1 + B_2 x_2 \le b_2,\ l \le x_2 \le u,$$

where there are no a priori bounds on x_1.

By adding slack variables and splitting variables as necessary, rewrite the problem to standard form ($Ax = b$, $x \ge 0$).

20 11

M2
Continuous Optimization
P.-A. Zitt

Université Gustave Eiffel
Année 2024–2025

Gradient, convexity

Exercise 1 — Taylor. Show that if f is C^2 on a convex open set Ω, then for all x and y in Ω,

$$\nabla f(y) - \nabla f(x) = \int_0^t \nabla^2 f((1-t)x + ty) \cdot (y - x) dt.$$

Hint : make sure you know where every quantity lives; take the derivative of the scalar function

$$\phi(t) = \langle p, \nabla f((1-t)x + ty) \rangle$$

where p is an arbitrary vector in $\mathbb{R}^n$.

Exercise 2 — Convexity. Using the definition of convexity, check that, for a positive matrix $\mathbb{A}$, the function $q(x) = x^\top \mathbb{A} x$ is convex.

Exercise 3 — Isolated minimizers. Prove that if $f : \Omega \to \mathbb{R}$ has an isolated local minimizer at $x \in \Omega$, then it is a strict local minimum.

Exercise 4 — Graphic visualization. Using Python (libraries `numpy` and `matplotlib.pyplot`; useful functions `numpy.arange`, `matplotlib.pyplot.contour`) visualize the contour lines/level sets of the functions :

$$f(x) = 8x_1 + 12x_2 + x_1^2 - 2x_2^2$$
$$g(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

Show that f has a unique stationary point, which is not a local extremum. Show that g has a unique stationary point, and that it is a global minimum. Do you see these points on the contour plot ?

Exercise 5 — Rates of convergence. Let $x^{(k)}$ converge to x_* in $\mathbb{R}^n$. If the following holds true :

$$\exists M, \exists p, \forall k, \frac{\|x^{(k+1)} - x_*\|}{\|x^{(k)} - x_*\|^p} \leq M,$$

the convergence is called "of order p"; if $p = 1$ the convergence is "Q-linear"; if $p = 2$ it is "Q-quadratic".

1. Show that Q-quadratic convergence implies Q-linear convergence.
2. What is the order of convergence of the sequence $1 + (1/3)^k$? Of $(1 + e^{-e^k})$? Of $1 + 1/k^2$?
3. Assume that $x^{(k)}$ has P "correct" digits. How many correct digits does $x^{(k+1)}$ have, if the convergence is linear ? It it is quadratic ?

Exercise 6 — Descent direction. Let $f : x \mapsto (x_1 + x_2^2)^2$. At $x = \binom{1}{0}$, check that $p = \binom{-1}{1}$ is a descent direction. Analyze f along the half line $\{x + tp, t \geq 0\}$ and find all minimizers in this direction.

before christmas

b 2

M2
Continuous Optimization
P.-A. Zitt

$\lambda_n = \frac{1}{\lambda_1}$?

Université Gustave Eiffel
Année 2024–2025

$2A^{-1}b$

Homework assignment

0965729998

(a, b)

Exercise 1 — An inequality by Kantorovich. The goal is to prove that, for any SPD matrix $\mathbb{A}$, and any vector x,

$$\|x\|_2^2 \le \langle x, \mathbb{A}x\rangle^{1/2}\langle x, \mathbb{A}^{-1}x\rangle^{1/2} \le \frac{\lambda_1+\lambda_n}{2\sqrt{\lambda_1\lambda_n}}\|x\|_2^2,$$

$\sqrt{\frac{\lambda_1}{\lambda_n}} + \sqrt{\frac{\lambda_n}{\lambda_1}}$

where λ_1 (resp. λ_n) is the smallest (resp. largest) eigenvalue of $\mathbb{A}$.

1. Show that it is enough to prove the result for $\mathbb{A}$ a diagonal matrix, with entries $\lambda_1 \le \cdots \le \lambda_n$, and that one can assume that $\lambda_n = 1/\lambda_1$.
2. Show that if $\lambda_1 \le \lambda_i \le 1/\lambda_1$, then $\lambda_i + 1/\lambda_i \le \lambda_1 + 1/\lambda_1$.
3. Using the bound $ab \le \frac{1}{2}(a^2+b^2)$, show that

$$\Big(\sum \lambda_i x_i^2\Big)^{1/2}\Big(\sum \lambda_i^{-1}x_i^2\Big)^{1/2} \le \frac{1}{2}\sum_i(\lambda_i+\lambda_i^{-1})x_i^2.$$

$\lambda_1 \le \frac{1}{\lambda_1}$

4. Conclude.

$x - \frac{1}{x^2} \quad x^2.$

Exercise 2 — Rate of convergence and condition number. Assume that $\mathbb{A}$ is SPD and denote by $0 < \lambda_1 \le \cdots \le \lambda_n$ its eigenvalues. Consider the minimization of $q(x) = \frac{1}{2}\langle x, \mathbb{A}x\rangle - \langle \mathbb{b}, x\rangle$ with the descent algorithm that :

1. follows the gradient : $p^{(k)} = -\nabla q(x^{(k)})$,
2. selects the optimal step size at each step.

$a = 2$

$\frac{\lambda_1}{\lambda_n} + \frac{\lambda_n}{\lambda_1}$

Let $e^{(k)} = x^{(k)} - x_*$ (the "error" at step k) and $r^{(k)} = -\mathbb{A}e^{(k)}$ (the "residue").

1. Recall why $r^{(k)} = \mathbb{b} - \mathbb{A}x^{(k)}$.
2. Show that

$$\|e^{(k+1)}\|_{\mathbb{A}}^2 = \|e^{(k)}\|_{\mathbb{A}}^2 - \frac{\|r^{(k)}\|_2^4}{\|r^{(k)}\|_{\mathbb{A}}^2}.$$

$ab = 1$

Rose depend and

3. Using Kantorovich's inequality, deduce that

$$\|e^{(k+1)}\|_{\mathbb{A}} \le \|e^{(k)}\|_{\mathbb{A}}\frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1}.$$

$ab = 1$

$\lambda_1\lambda_n = k$

4. Conclude that

$\lambda_1\lambda_2$

$$\|e^{(k)}\|_{\mathbb{A}} \le \left(\frac{\kappa_2(\mathbb{A})-1}{\kappa_2(\mathbb{A})+1}\right)^k \|e^{(0)}\|_{\mathbb{A}},$$

$\lambda_n = \frac{k}{\lambda_1}$

where $\kappa_2(\mathbb{A}) = |||\mathbb{A}|||_2|||\mathbb{A}^{-1}|||_2$ is the condition number of the matrix $\mathbb{A}$ relatively to the Euclidean norm in $\mathbb{R}^n$.

5. Is the convergence linear ? Superlinear ? Comment on the dependence on $\kappa_2(\mathbb{A})$, keeping in mind the geometry of the problem (say, in $\mathbb{R}^2$).

Does depend on .

Exercise 3 — Code. Using your python code for steepest descent and the Newton method, how many iterations are needed to get $\left|x^{(k)} - x_*\right|$ smaller than 10^{-3} ? Smaller than 10^{-6} ? Start from $x^{(0)} = (1.2, 1.2)$, and then from $x^{(0)} = (-1.2, 1)$. (Send your code and the results, e.g. in notebook form).

K_x, K_y $\sqrt{K}$ $\log$

K^2 $K_x = K_y$

M2
Continuous Optimization
P.-A. Zitt

Université Gustave Eiffel
Année 2024–2025

Final Exam

In all the exercises, $\|x\|$ denotes the Euclidean norm and $\langle x, y\rangle$ the usual scalar product in $\mathbb{R}^d$. In exercises 1 and 2, B is a symmetric square matrix $B \in \mathcal{M}_d(\mathbb{R})$. Recall that B has d real eigenvalues $\lambda_1 \leq \dots \leq \lambda_d$, and that there exists an orthogonal basis $(v_1, \dots, v_d)$ such that $Bv_i = \lambda_i v_i$, for all $1 \leq i \leq d$. Finally, recall that B is called "positive semi-definite" if

$$\forall v \in \mathbb{R}^d, \quad \langle v, Bv\rangle \geq 0.$$

Exercise 1. Let $B \in \mathcal{M}_d(\mathbb{R})$ be a symmetric matrix.

1. Show that B is positive semi-definite if and only if $\lambda_1 \geq 0$.
2. Let $v_0 \in \mathbb{R}^d \setminus \{0\}$. Show that :

$$B \text{ is positive semi-definite} \iff \forall v \in \mathbb{R}^d, \langle v, v_0\rangle > 0 \implies \langle v, Bv\rangle \geq 0.$$

3. Let $f \in \mathcal{C}^2(\mathbb{R}^d)$. Let x_0 be a critical point of f. Let $v_0 \in \mathbb{R}^d$. Let $H = \{x, \langle x, v_0\rangle \geq \langle x_0, v_0\rangle\}$. Assume that

$$\forall x \in H, \quad f(x) \geq f(x_0).$$

Show that the Hessian of f at x_0 is positive semi-definite.

$H. \quad f(x) \geq f(x_0)$

Exercise 2 — Trust region.

1. Let B be a symmetric matrix, let $b \in \mathbb{R}^d$, and let

$$f(x) = \frac{1}{2}\langle x, Bx\rangle - \langle b, x\rangle.$$

$\frac{1}{2} x^T B x - b^T x$

Compute $\nabla f(x)$. What is the Hessian of f?

2. In the following we write $\lambda_1 \leq \dots \leq \lambda_d$ the eigenvalues of B. Discuss the existence and uniqueness of a global mininum of f (on $\mathbb{R}^d$) in the three cases $\lambda_1 > 0$, $\lambda_1 = 0$ and $\lambda_1 < 0$.

In the rest of the problem, we look for minimizers of f in the Euclidean ball $B_\delta = \{x : \|x\| \leq \delta\}$:

$$\begin{aligned} &\min f(x) \\ &\text{subject to } x \in B_\delta. \end{aligned} \tag{1}$$

convex

3. Show that f is bounded on B_δ and reaches its lower bound at at least one point.
4. Find a function g_δ such that g_δ *is* C^∞ and $x \in B_\delta \iff g_\delta(x) \geq 0$. Use g_δ to write Problem (1) as a constrained minimization problem with smooth functional constraints.

From now on, x_* denotes a solution of the optimization problem (1).

5. Show that the constraints are necessarily qualified at x_*.
6. Show that there exists a $\lambda_* \geq 0$ such that :

$\|x\| \leq \delta$

$$\begin{cases} \|x_*\| \leq \delta, \\ \boxed{Bx_* - b = -\lambda x_*,} \\ \lambda_*(\|x_*\|^2 - \delta^2) = 0. \end{cases}$$

x, B

$x B x.$

$(x-x_*)$ $\lambda > -\lambda_1$

We now wish to show that $B+\lambda_* I$ is positive semi-definite.

7. Prove that

$$\forall x \in \mathbb{R}^d, \quad \frac{1}{2}\langle x - x_*, (B+\lambda_*)(x-x_*)\rangle = f(x) - f(x_*) + \frac{\lambda_*}{2}(\|x\|^2 - \|x_*\|^2). \qquad (2)$$

8. Assuming that λ_* is zero, prove that B is positive semi-definite (hint : if v is a non-zero vector, can you write v as a multiple of $x - x_*$ for an $x \in B_\delta$?)

9. Assume that λ_* is non zero. Compute $\|x_*\|$. Show that $B+\lambda_* I$ is positive semi-definite (hint : apply (2) to well-chosen vectors x on the sphere $\{x : \|x\| = \delta\}$).

We now use this characterization to devise an algorithm to solve Problem (1). For simplicity we assume that $\lambda_1 < 0$. For $\lambda > -\lambda_1$, let $x(\lambda) = (B+\lambda)^{-1}b$.

10. Why is $x(\lambda)$ well-defined ?
11. Show that $\psi : \lambda \mapsto \|x(\lambda)\|^2$ is decreasing on $(-\lambda_1, \infty)$. Hint : decompose x and b on a basis of eigenvectors $(v_1, ..., v_d)$ of B.
12. We assume from now on that $\langle b, v_1\rangle \neq 0$, where v_1 is an eigenvector of B associated to the eigenvalue λ_1. Show that the equation $\psi(\lambda) = \delta^2$ has a unique solution $\tilde{\lambda}$ on (λ_1, ∞).
13. Prove that $x(\tilde{\lambda})$ solves Problem (1).
14. What algorithm could be used to find $\tilde{\lambda}$, and therefore a solution to Problem (1) ?

Exercise 3. Let $a_1, ..., a_n, b_1, ..., b_n$ be points in $\mathbb{R}^d$. Assume that there is a mass 1 of water in each of the a_i, and that we want to tranport it on the b_j, so that each b_j receives a total mass 1.

A "transport plan" is a family $(x_{ij})_{1\le i\le n, 1\le j\le n} \in \mathbb{R}^{n^2}$ such that

$$\begin{aligned}
&\forall i, j, && x_{ij} \geq 0,\\
&\forall 1 \leq i \leq d, && \sum_j x_{ij} = 1,\\
&\forall 1 \leq j \leq d, && \sum_i x_{ij} = 1.
\end{aligned}$$

B

The quantity x_{ij} denotes the mass of water that is transported from a_i to b_j.

1. Give the meaning/interpretation of the three families of equations above.
2. Given $p \geq 1$, the optimal (L^p) transport problem consists in finding the plan $x = (x_{ij})$ such that the cost

$$f(x) = \sum_{i,j} x_{ij} |a_i - b_j|^p$$

is as small as possible. Show that this is a linear problem.

3. Show that there exists an optimal transport plan.
4. Let x be an optimal transport plan. Show that there exist $\gamma_{ij} \geq 0$, α_i and β_j such that

$$\forall i, j, \qquad \|a_i - b_j\|^p = \gamma_{ij} + \alpha_i + \beta_j,$$

where the γ_{ij} satisfy an additional condition that you will specify.

5. In this question we fix $p = 2$. Show that, if $\langle b_0 - a_0, b_1 - a_1\rangle \neq 0$, at least one of the four quantities x_{01}, x_{01}, x_{10} and x_{11} must be zero.
6. In this question we fix $p = 1$ and $d = 2$. Using a picture, show that, if the four points (a_1, a_2, b_1, b_2) are not aligned, and if the two planar segments $[a_1, b_2]$ and $[a_2, b_1]$ intersect in a point $c \notin \{a_1, a_2, b_1, b_2\}$, then

$$\|b_2 - a_2\| + \|b_1 - a_1\| < \|b_1 - a_2\| + \|b_2 - a_1\|.$$

Deduce that in this case, an optimal transport plan cannot contain *crossings* : the product $x_{10}x_{01}$ must be zero.