Exercises
Optmization & Operational Research: Part I

The difficulty of the exercises is denoted with some (*), the more (*) you have the more difficult is the exercise from my opinion. The exercises are classified with respect to each part of the course.

Inner products and norms

Exercise 1 :

Which of the following applications define an inner product :

- (*) $f(x, y) = x_1y_1 + x_2y_2$.
- (*) $f(x, y) = x_1y_1 + x_2y_2 - x_3 + y_3$.
- (**) $f(x, y) = x_1y_1 + 2x_2y_2 + 3x_3y_3$.
- (**) $f(x, y) = x_1^2y_1^2 + x_2^2y_2^2 + x_3^3y_3^3$.

Exercise 2 : Frobenius' Norm

- (*) Show that application $\langle ., . \rangle : \mathcal{M}_{n,m}(\mathbb{R}) \times \mathcal{M}_{n,m}(\mathbb{R}) \to \mathbb{R}$ defined by :

$$\langle A, B \rangle = trace(A^T B),$$

defined an inner product.

- (*) Show that $\|A\|_F = \sqrt{trace(A^T A)} = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{m} (a_{ij}^2)}$, and show that it defines a norm.
- (**) Show that $\|Ax\|_2 \leq \|A\|_F \|x\|_2$ where $A \in \mathcal{M}_{n,m}(\mathbb{R})$ and $x \in \mathbb{R}^m$.
- (**) Show that $\|AB\|_F \leq \|A\|_F \|B\|_F$ where $A \in \mathcal{M}_{n,m}(\mathbb{R})$ and $B \in \mathcal{M}_{m,p}(\mathbb{R})$
- (*) Calculate the Frobenius norm of the following matrices :

$$A = \begin{pmatrix} 1 & -3 \\ -3 & 1 \end{pmatrix} \quad B = \begin{pmatrix} 3 & -2 & 3 \\ -2 & 1 & -2 \\ -3 & 2 & 3 \end{pmatrix}$$

Exercise 3 :

The aim of the exercise is to prove the inequality of Minkowski, i.e the triangular inequality for the L^p norm for $p \in [1, \infty[$. x, y are considered as vectors here.

1) Let $0 < p, q < \infty$ such that $\frac{1}{p} + \frac{1}{q} = 1$

- (*) Show that $\ln(xy) = \frac{\ln(x^p)}{p} + \frac{\ln(y^q)}{q}$ for all $x, y > 0$.
- (*) Use the convexity of the exponential to show *Young's inequality* :

$$|xy| \leq \frac{|x|^p}{p} + \frac{|y|^q}{q}.$$

2) We want to prove now that : $\|xy\|_1 \leq \|x\|_p \|y\|_q$ (*Hölder's inquality*)
We consider $0 < p, q < \infty$ such that $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$ and $x, y \in \mathbb{R}^n$.

- (**) By a good choice of p, q, x and y, show that, applying Young's inequality :

$$|x_i y_i|^r \leq \frac{1}{p'}|x_i|^p + \frac{1}{q'}|y_i|^q,$$

  where you should determine the value of p' and q'.

- (**) Prove Hölder's inequality using the previous result (first you have to take the sum on all i and consider the special case where $r = 1$).
  Hint: set $x_i = \frac{x_i}{\|x\|_p^p}$ and $y_i = \frac{y_i}{\|y\|_q^q}$

3) The triangle inequality for the L^p norm.

- (***) Use successively the triangle inequality and *Hölder's inequality* to show that :

$$\|x + y\|_p^p \leq (\|x\|_p + \|y\|_p) \frac{\|x + y\|_p^p}{\|x + y\|_p}.$$

  This last inequality is called the *inequality of Minkowski.*

- (*) Show that the application $f(x) = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p}$ is a norm.

Derivatives

Exercise 1 : Calculous

Calculate the first and second order derivatives of the following functions :

- (*) $f(x, y) = 4x^2 + \exp(xy)$.
- (*) $f(x, y) = 7xy + \cos(x) + x^2 + 4y^2$.
- (*) $f(x, y) = 4(x - y)^2 + 5(x^2 - y)^2$.
- (*) $f(x, y) = \exp(x^2 + y^2)$.

Exercise 2 : Schwarz theorem : a counter example

During the lesson we have seen that, given a function f twice continuously differentiable, we always have :

$$\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right) = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right).$$

Let us now consider the function f defined by :

$$f(x, y) = \frac{xy(x^2 - y^2)}{x^2 + y^2}, if\ (x, y) \neq (0, 0)\ and\ f(x, y) = 0\ if\ (x, y) = 0.$$

- (*) Calculate $\frac{\partial f}{\partial x}(0, y)$ and $\frac{\partial f}{\partial y}(x, 0)$.
- (*) Calculate $\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)(0, 0)$ and $\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)(0, 0)$.
- (*) What can we conclude about f ?

Convex set

Exercise 1 : Using definition

- (*) Given two convex sets C_1 and C_2, prove that the intersection $C = C_1 \cap C_2$ is also convex.
- (**) Show that a set C is convex if and only if its intersection with every straight line is convex .
- (***) Show (using induction) that definition of convexity holds for more than two points.

Exercise 2 :

- (*) Let C be a set defined by :

  $$C = \{x \in \mathbb{R} \mid 3x^2 - 6x + 2 \leq 0\}$$

  Show that C is convex.

- (**) In general, show that the set C defined by :

$$C = \{x \in \mathbb{R}^n \mid x^T A x - b^T x + c \leq 0\},$$

where $A \in S^n(\mathbb{R}), b \in \mathbb{R}^n$ and $c \in \mathbb{R}$ is convex if A is a PSD matrix.

Exercise 3 :

- (**) Show that the hyperbolic set $\{x \in \mathbb{R}^n_+ \mid \prod_{i=1}^n x_i \geq 1\}$ is convex.
  Hint : you can first show that, for all $x, y \in \mathbb{R}_{++}$ and $\theta \in [0,1]$ we have : $x^\theta y^{1-\theta} \leq \theta x + (1-\theta) y$.

Convex function

Exercise 1 : Calculous

For the following functions, explain why they are convex :

- (*) $f : \mathbb{R}^n \to \mathbb{R}, f(x) = \sum_{i=1}^n x_i^2$.
- (*) $f(x,y) = 3x^2 + (y-3)^2 + 4x + 6y + 5$.
- (*) $f(x,y) = x^4 + 6y^4 + 2y^2 + 9x^2 + 3$.
- (**) $f(x,y) = 6x^2 + 5y^2 + 6xy$.
- (**) $f(x,y) = \exp(xy)$ for $x > 1$ and $y < -1$.

Exercise 2 : Calculous

Are the following functions convex or not ?

- (*) $f(x,y) = (1-x)^2 + 100(y-x^2)^2$.
- (*) $f(x,y) = (x+2y-7)^2 + (2x+y-5)^2$.
- (**) $f(x,y) = 2x^2 - 1.05x^4 + xy + y^2$.
- (**) $f(x,y) = \sin(x+y) + (x-y)^2 - 1.5x + 2.5y + 1$.
- (***) $f(x,y) = 10 + (x^2 - \cos(2\pi x)) + (y^2 - \cos(2\pi y))$.

Find the local or global minimum of the two first function

Exercise 3 : A PSD matrix

Let $(x_1, x_2, ..., x_n)$ be n vector of $\mathbb{R}^p$, we denote by $X \in \mathbb{R}^{n \times p}$ the matrix where each the i^{th} row is the vector x_i.
We consider the matrix $G \in \mathbb{R}^{n \times n}$ defined by $G = XX^T$. The matrix G is called the *Gram Matrix*.

- (**) Show that the Gram Matrix is a PSD matrix using the definition of a PSD matrix.

Optimization and Algorithm

Exercise 1 : A Quadratic function : Matyas function

We consider the function $f : [-10, 10]^2 \to \mathbb{R}$ defined by :

$$f(x, y) = 0.26(x^2 + y^2) - 0.48yx$$

- (*) Is the function f convex or not ?
- (*) Find the solution(s) of the equation $\nabla f(x, y) = 0$.
- (*) What is the global minimum of the function ?
- (*) We set $u_0 = (x, y)^{(0)} = (1, 1)$, the initial point of the gradient descent with the optimal learning rate (or optimal step)
  (a) First recall what the gradient descent with optimal step consists of.
  (b) Calculate u_1 and u_2.

Exercise 2 : The Rosenbrock function

We consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by :

$$f(x, y) = (1 - x)^2 + 10(y - x^2)^2.$$

- (*) Is the function f convex or not ?
- (*) Find the solution(s) of the equation $\nabla f(x, y) = 0$.
- (*) What is the global minimum of the function ?
- (*) We set $u_0 = (x, y)^{(0)} = (2, 2)$, the initial point of the gradient descent with a learning rate $\rho = 0.5$.

(a) First recall what the gradient descent consists of.
(b) Calculate u_1 and u_2.

Exercise 3 : The Rastrigin function

We consider the function $f : [-\pi, \pi]^2 \to \mathbb{R}$ defined by :

$$f(x, y) = 20 + (x^2 - 10\cos(2\pi x)) + (y^2 - 10\cos(2\pi y))$$

- (*) Is the function f convex or not ?
- (***) Find the solution(s) of the equation $\nabla f(x, y) = 0$.
- (*) We assume that this function is positive for all x, y. What is the global minimum of the function ?
- (*) We set $u_0 = (x, y)^{(0)} = (2, 2)$, the initial point of the gradient descent with a learning rate $\rho = 0.5$.
  (a) First recall what the gradient descent consists of.
  (b) Calculate u_1 and u_2.
- Are we sure that the algorithm will reach the global minimum ? Why ?

Exercise 4 : A quadratic function

We consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by :

$$f(x, y) = 7y^2 + 4x^2 - 5xy + 2x - 7y + 32.$$

- (*) Is the function f convex or not ?
- (*) Find the solution(s) of the equation $\nabla f(x, y) = 0$.
- (*) What is the global minimum of the function ?
- (*) We set $u_0 = (x, y)^{(0)} = (1, 1)$, the initial point of the Newton's Method
  (a) First recall what is the Newton's Method.
  (b) Calculate u_1 and u_2.

Exercise 5 : A last function

We consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by :

$$f(x, y) = 2x^2 - 1.05x^4 + \frac{x^6}{6} + xy + y^2.$$

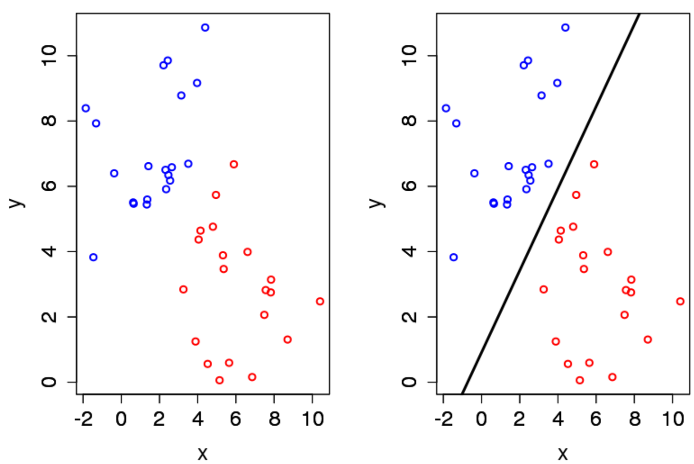
- (*) Calculate the Hessian Matrix.

- (*) What are the quantities we have to calculate to prove that a 2×2 matrix is PSD ? Calculate them.
- (*) We assume that the function f is non-negative and non-convex, i.e $f(x, y) \geq 0$. Show that $(0, 0)$ is a solution of $\nabla f(x, y) = 0$.
- (*) What is the global minimum of the function ?
- (*) We set $u_0 = (x, y)^{(0)} = (1, 1)$,the initial point of the gradient descent with a learning rate $\rho = 0.5$.
  (a) First recall what is the gradient descent.
  (b) Calculate u_1 and u_2.
- Are we sure that the algorithm will reach the global minimum ? Why ?

Exercise 6 : An application of Newton's Method: The logistic regression

Let us consider $\mathbf{X} = (x_1, x_2, ..., x_n) \in \mathbb{R}^d$ and $\mathbf{Y} = (y_1, y_2, ..., y_n) \in \{0, 1\}^n$ be respectively the matrix of the feature vector of n instances and their label.

The logistic regression is used as a binary classifier (it can be extended to multiclass classification problem) where the classifier returns the probability of an example to belong to class of reference (let say the class 1). An example of classifier trained using a logistic regression model is shown below.



The logistic regression is based on the following model:

$$\ln \left(\frac{p(1 \mid X)}{p(0 \mid X)} \right) = w_0 + w_1 x_1 + w_2 x_2 + ... + w_d x_d.$$

In other words we estimate the log of the ratio of the probabilities of being in the class 1 with the one being in the class 0. This model is called a **LOGIT** model. The quantity $p(1 \mid X = x)$ is called the posterior probability of being in the class 1, i.e. the probability for the example to be in the class 1.

- Using the above equation, give an expression of $p(1 \mid X)$ which depends on the vector of parameters $\mathbf{w} \in \mathbb{R}^{d+1}$. We will note g the obtained function, this function is called the **logistic** function.

- Show that, for any $w \in \mathbb{R}$, we have $\nabla_w g(w) = g(w)(1 - g(w))$.
- Study the convexity of function g.
- What about $\ln(g(w))$?

A classical method to estimate the parameter of of a logistic regression model is to find the parameter that maximize the likelihood of you data. The likelihood of an instance x_i under this model is given by:

$$P(y_i \mid x_i, w) = g(w, x_i)^{y_i}(1 - g(w, x_i))^{1-y_i}.$$

The law is the same as the Bernoulli law $\mathcal{B}(p)$ with probability $p = g(w, x_i)$ where p is probability of being in the class 1.

- We denote by L the likelihood of our date and ℓ the log-likelihood of the data. Determine the expression of $-\ell$, the opposite of the log-likelihood.
- Study the convexity of such problem.
- Write the Newton's method to solve the problem of minimization:

$$\min_{w \in \mathbb{R}^{d+1}} - \ell(w, \mathbf{X}, \mathbf{Y})$$

Exercise 7: The Backtracking Line Search

In class, we have seen that the efficiency of the gradient descent algorithm depends on the choice of the learning rate ρ. It its simple version the learning rate is fixed. In practice, the value of the learning rate is decreasing with respect to the number iteration, it can be $\rho_k = \left(\frac{1}{2}\right)^{k-1}$.
Such way to choose the learning rate is called an **inexact line search method**, indeed, we are not sure to reach the minimum of the function along the choosen direction of descent.

An other inexact but more reliable method to choose the value of the step ρ_k at iteration k is the **Backtracking Line Search**:

Given a direction of descent d_x, two real numbers $\alpha \in [0, 0.5]$ and $\beta \in (0, 1)$, we set $\rho = \beta\rho$ while:

$$f(x + \rho d_x) \geq f(x) + \alpha\rho\langle\nabla f(x), d_x\rangle.$$

The name **backtracking** comes from the fact that the value ρ is updated till the stopping condition holds.

- Suppose that the function f is strongly convex with $mI \leq \nabla^2 f(x) \leq MI$. Show that for all x and d_x we have:
$$f(x + \rho d_x) \leq f(x) + \rho\nabla\langle f(x), d_x\rangle + \rho^2(M/2)\langle d_x, d_x\rangle.$$
- Using the previous question, show that the backtracking stopping condition holds for:
$$0 < \rho \leq -\frac{\langle\nabla f(x), d_x\rangle}{M\|d_x\|_2^2}.$$

Exercise 8: Minimizing a quadratic-linear fractional function:

We consider the function $f : \mathbb{R}^n \to \mathbb{R}$ defined by:

$$f(x) = \frac{\|Ax - b\|_2^2}{c^T x - d},$$

where x is such that $c^T x + d - 0$. We will assume that the function is bounded below and that it admits only one minimum. We further assume that A is full rank.

- Show that the minimizer x^* of the function f has the form:

$$x^* = x_1 + tx_2,$$

where $x_1 = (A^T A)^{-1} A^T b,\ x_2 = (A^T A)^{-1} c$ and t where t is such that:

$$t = \frac{\|Ax^* - b\|_2^2}{2(c^T x^* - d)}.$$

- Show that the value of t is given by solving a second order equation and find this value. We assume that the obtained polynom has two roots.

Exercise 9: The optimal step algorithm: illustration of convergence

Let $\gamma \in \mathbb{R}_+^*$ consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by:

$$f_\gamma(x) = \frac{1}{2}(x_1^2 + \gamma x_2^2).$$

We want to apply the algorithm of gradient descent with optimal step to find the minimum of the function f and show that is algorithm converge.

- Say where the function reaches its minimum.
- Recall what is the the gradient descent algorithm with optimal step.

We initialize the algorithm at the point $x^{(0)} = (x_1^{(0)}, x_2^{(0)}) = (\gamma, 1)$.

- Compute the value of $x^{(1)}$ and $x^{(2)}$ using the above algorithm.
- Show that, for all $k \in \mathbb{N}$, the value of $x^{(k)}$ is given by:

$$x^{(k)} = \left(\left(\frac{\gamma - 1}{\gamma + 1}\right)^k \gamma, \left(\frac{\gamma - 1}{\gamma + 1}\right)^k (-1)^k \right).$$

- Prove that:

$$f(x^{(k)}) = \left(\frac{\gamma - 1}{\gamma + 1}\right)^{2k} f(x^{(0)}).$$

- Conclude about the convergence.

Technical Proofs and convergence

Exercise 1: An $\alpha-$ elliptical function

Definition 1 *Let V be a $\mathbb{R}^n-$ vectorial space. A function $f : V \mapsto \mathbb{R}$ is said to be $\alpha-$elliptical if f is continuously differentiable on V and if it exists $\alpha > 0$ such that:*

$$\langle \nabla f(v) - \nabla f(u), v - u \rangle \geq \alpha \|v - u\|_2^2, \quad \forall u, v \in V.$$

- Using the above definition prove the following inequality:

  $$f(v) \geq f(u) + \langle \nabla f(u), v - u \rangle + \frac{\alpha}{2} \|v - u\|_2^2, \quad \forall v, u \in V.$$

  Hint: Introduce the function ϕ defined by $\phi(t) = f(u + t(v - u)),\ t \in [0, 1]$ and compute $f(v) - f(u)$. (Recall that $\phi(x) - \phi(y) = \int_x^y \nabla \phi(z)\ dz$)

Definition 2 *Let U an unbounded part of the space $\mathbb{R}^n$. A function $f : U \mapsto \mathbb{R}$ is said to be coercive if:*

$$\lim_{\|u_k\| \to \infty} f(u_k) = \infty.$$

- Using the previous result, show that an $\alpha-$elliptical function is coercive.

Remark: The notion of coercivity (combined with the lower semi-continuity property) is used to prove that there exists $u \in U$ such that:

$$f(u) = \inf_{v \in U} f(v),$$

i.e. that a function f has a minimum.

Exercise 2: Gradient descent with optimal step

The aim of this exercise is to prove the following result:

Theorem 1 *Let $f : \mathbb{R}^n \mapsto \mathbb{R}$, a continuously differentiable and $\alpha-$elliptical function. Then, the Gradient Descent algorithm with Optimal step converges.*

To prove this result, we will consider a the sequence $(u_k)_{k \in \mathbb{N}}$ and u the point where the function f reaches its minimum. We aim to show that $\lim_{k \to \infty} u_k = u$.

- Show that $f(u_k) - f(u_{k+1}) \geq \frac{\alpha}{2} \|u_k - u_{k+1}\|_2^2$.
- Explain why $\|u_k - u_{k+1}\|_2^2 \to 0$
- Assume that $\|u_k - u_{k+1}\|_2^2 \to 0$ implies $\|\nabla f(u_k) - \nabla f(u_{k+1})\|_2^2 \to 0$ and show that $\|\nabla f(u_k)\|_2 \to 0$.
- Conclude by showing that:

  $$\lim_{k \to \infty} \|u_k - u\|_2 = 0.$$

Exercise 3: Convergence analysis of the gradient descent with a variable (or fixed) step

The aim of the exercise is to prove the following result:

Theorem 2 *Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be an $\alpha-$elliptical function such that $\forall\ u, v \in \mathbb{R}^n$:*

$$\|\nabla f(v) - \nabla f(u)\|_2^2 \leq M\|v - u\|_2$$

and

$$\langle \nabla f(v) - \nabla f(u), v - u\rangle \geq \alpha\|v - u\|_2^2,$$

where $\alpha, M > 0$. Then the Gradient descent with a variable (or fixed) step ρ (or ρ_k) converges for:

$$0 < a \leq \rho \leq b < \frac{2\alpha}{M}$$

We will denote by $u \in \mathbb{R}^n$ the point for which the function f reaches its minimum.

- Find the value of γ such that $\|u_{k+1} - u\|_2^2 \leq \gamma^2\|u_k - u\|_2^2$.
- Give a condition on γ such that $\underset{k\to\infty}{lim}\ \|u_k - u\| = 0$ and conclude.

Exercise 4: Properties of the Conjuguate Gradient Descent

Before starting this exercise, read in your slides what the method consists of. In the following we consider $A \in \mathcal{S}_n^{++}(\mathbb{R})$ (the set of symetric and PD matrices).
Try to prove, by induction, the following result:

Proposition 1 *Let $1 \leq k \leq n$ be such that $\nabla f(u_0), ..., \nabla f(u_k)$ are non zero. Then we have the following relations:*

$$\langle \nabla f(u_k), \nabla f(u_l)\rangle = 0, \quad \forall l = 0, ..., k-1$$

and

$$\langle Ad_k, d_l\rangle = 0, \quad \forall l = 0, ..., k-1.$$

The following Theorem is then a consequence of the above Proposition.

Theorem 3 *The Conjuguate Gradient Descent converges in, at most, n iterations.*

Proof 1 *Indeed, we have shown that $\langle \nabla f(u_k), \nabla f(u_l)\rangle = 0, \quad \forall l = 0, ..., n$. So the set of derivatives $\nabla f(u_k)$ is a* ***base*** *of $\mathbb{R}^n$, thus u can be expressed as linear combination of these derivatives.* $\square$

Other exercises

Exercise 1: About the Gradient Descent with optimal step

Let $A \in \mathcal{S}_n^{++}(\mathbb{R})$, i.e. a symetric and definite positive matrix. Let $(u_k)_{k\in\mathbb{N}}$ be a sequence obtained using the gradient descent with optimal step applied to the quadratic function:

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

- Show that we have:

$$\|u_{k+1} - u\|_A^2 \leq \left(1 - \frac{\lambda_1}{\lambda_n}\right)^2 \|u_k - u\|_A^2,$$

where λ_1 and λ_n are respectively the smallest and the largest eigenvalue of A.

The ratio $\dfrac{\lambda_n}{\lambda_1} = Cond(A)$ is the condition number of the matrix A. It is used to measure how the error of the output evolves when a small error or change is introduced in the input. A simple application or study can be made by solving a linear system $Au = b$ such as in *Linear Regression*.

Hint: We will assume that for any matrix $A \in \mathcal{S}_n^{++}(\mathbb{R})$, there exists one and only one matrix $B \in \mathcal{S}_n^{++}(\mathbb{R})$ such that $A = B^2$. This matrix is usually denoted by $\sqrt{A}$.

Exercise 2: Inequality of Kantorovich

The aim of this exercise is give a rate of convergence of the Gradient Descent with Optimal step that depends on the Condition number of the matrix A.
Let $(u_k)_{k\in\mathbb{N}}$ be a sequence obtained using the gradient descent with optimal step applied to the quadratic function:

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

- Show that, for all $u \in \mathbb{R}^n$:

  $$\frac{\|u\|^4}{\|u\|_A^2\|u\|_{A^{-1}}^2} \geq \frac{4\lambda_1\lambda_n}{(\lambda_1+\lambda_n)^2}.$$

  Hint: Use the fact that: $\|u\|_A^2\|u\|_{A^{-1}}^2 = \|u\|_{\frac{1}{t}A}^2\|u\|_{tA^{-1}}^2$ for any value of $t > 0$. You will also have to use the inequality: $(a+b)^2 \geq 4ab$.

- Using the previous result, show that:

  $$\|u_{k+1} - u\|_A^2 \leq \left(\frac{Cond(A) - 1}{Cond(A) + 1}\right)^2 \|u_k - u\|_A^2,$$

  where $Cond(A) = \dfrac{\lambda_n}{\lambda_1}$ and λ_1, λ_n have the same meaning as in the previous exercise.

Exercise 3: The Davidon-Fletcher-Powell Algorithm:

The Davidon-Fletcher-Powell Algorithm is described as follows:

1. Choose $S_0 = \mathbf{I}$ and a point $u_0 \in \mathbb{R}^n$
2. For $k \geq 0$, do:

   (a)
   $$f(u_k - \rho_k S_k \nabla f(u_k)) = \underset{\rho>0}{inf}\ f(u_k - \rho S_k \nabla f(u_k)),$$

   (b)
   $$u_{k+1} = u_k - \rho_k S_k \nabla f(u_k),$$

   (c)
   $$S_{k+1} = S_k + \frac{\delta_k \delta_k^T}{\delta_k^T \gamma_k} - \frac{S_k \gamma_k \gamma_k^T S_k}{\gamma_k^T S_k \gamma_k}.$$

   (d) Until $\|\nabla f(u_{k+1})\|_2^2 \leq \varepsilon$.

   with the usual notations:

   $$\gamma_k = \nabla f(u_{k+1}) - \nabla f(u_k), \quad \delta_k = u_{k+1} - u_k.$$

- Show that the *Quasi-Newton's Condition* holds for all matrices S_{k+1}, i.e. we have:

$$S_{k+1}\gamma_k = \delta_k.$$

- Let $u \in \mathbb{R}^n$. Show that:

$$u^T S_{k+1} u = \frac{(u^T S_k u)(\gamma_k^T S_k \gamma_k) - (\gamma_k^T S_k u)^2}{\gamma_k^T S_k \gamma_k} + \frac{(u^T \delta_k)^2}{\gamma_k^T, \delta_k},$$

$$\gamma_k^T \delta_k = \rho_k (\nabla f(u_k)^T S_k \nabla f(u_k)).$$

- Show that the matrices S_k are symetric and positive definite.