Exercises for Part 2

Exercise 2.1. (1) *We let* $A = \begin{pmatrix} 1 & 3 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, $b = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ *and* f *be represented by* (A, b). *Can the gradient descent initialised at a given* $x_0 \in \mathbb{R}^d$ *with fixed step size* $\tau > 0$ *converge?*

(2) *Assume that* A *is symmetric and that for any* $b \in \mathbb{R}^d$, *for any* $x_0 \in \mathbb{R}^d$ *there exists* $\tau > 0$ *such that the gradient descent generated at* x_0 *with step size* $\tau > 0$ *converges. Show that* $A \in S_d^{++}(\mathbb{R})$.

Exercise 2.2. *We let* $A \in S_d(\mathbb{R})$ *be matrix with (at least) two eigenvalues of opposite signs. We let* $b = 0$. *Show that for any* $\tau > 0$ *the set* $\{x_0 \in \mathbb{R}^d :$ *the gradient descent initialised at* x_0 *with fixed step size* τ *converges*$\}$ *has measure zero.*

Exercise 2.3. *[Some basic properties of the conditioning number]*

(1) *Show that, for any symmetric positive definite matrix* M, $\mathrm{cond}(M) \geq 1$.
(2) *Show that for any symmetric definite positive matrix* $\mathrm{cond}(M) = \|M\|_{\mathrm{op}} \cdot \|M^{-1}\|_{\mathrm{op}}$. *We use this expression to define the conditioning number of any invertible matrix* $M \in Gl_d(\mathbb{R})$.
(3) *Show that for any* $M \in Gl_d(\mathbb{R})$ $\mathrm{cond}(M) \geq 1$ *and that, for any orthogonal matrix* P, $\mathrm{cond}(PM) = \mathrm{cond}(M)$.
(4) *For any* $M \in Gl_d(\mathbb{R})$ *show that* $\|M\|_{\mathrm{op}} = \|M^T\|_{\mathrm{op}}$.
(5) *Let* $M \in Gl_d(\mathbb{R})$ *be such that* $\mathrm{Cond}(M) = 1$. *Show that there exists* $x \in \mathbb{R}^*$ *such that* xM *is an orthogonal matrix.*

Exercise 2.4. *Prove Theorem 4.1.*

Exercise 2.5. *Let* $f \in \mathscr{C}^1(\mathbb{R}^d; \mathbb{R})$ *be bounded from below, satisfy the Polyak-Lojasiewicz condition with constant* α:

$$\forall x \in \mathbb{R}^d \,, f(x) - \inf_{\mathbb{R}^d} f \leq \frac{1}{2\alpha} \|\nabla f(x)\|^2.$$

Assume that ∇f *is* μ*-Lipschitz. For any* $\tau \in \left(0; \frac{1}{2\mu}\right)$ *any* $x_0 \in \mathbb{R}^n$, *let* $\{x_k\}_{k \in \mathbb{N}}$ *be the sequence generated by the gradient descent initialised at* x_0 *with fixed step size* τ. *Show that*

$$\forall k \in \mathbb{N} \,, f(x_{k+1}) - \inf f \leq (1 - \tau\alpha)^{k+1} (f(x_0) - \inf f).$$

Exercise 2.6. *The goal of this exercise is to show the convergence of the line-search gradient descent for quadratic functions.*

(1) *Preliminary: Kantorovich inequality* *Let* $A \in S_d^{++}(\mathbb{R})$ *with eigenvalues* $0 < \lambda_1 \leq \cdots \leq \lambda_d$. *Show that*

$$\forall x \in \mathbb{R}^d \setminus \{0\} \,, \|x\|^4 \leq \langle Ax, x \rangle \cdot \langle A^{-1}x, x \rangle \leq \frac{\|x\|^4}{4} \cdot \frac{(\lambda_1 + \lambda_d)^2}{\lambda_1 \lambda_d}.$$

(2) *Let* $A \in S_d^{++}(\mathbb{R})$ *and* $b \in \mathbb{R}^d$. *Let* $x \in \mathbb{R}^d$. *Solve the optimisation problem*[1]

$$\min_{\tau > 0} f(x - \tau \nabla f(x)).$$

[1] In particular, show existence and uniqueness of the optimiser

(3) *We now consider the sequence generated by the line search algorithm. Using the explicit expression of the step size obtained at the previous question and defining, for any $k \in \mathbb{N}$, $y_k := A(x_k - x^*)$, show that*

$$\forall k \in \mathbb{N}\,, \langle y_{k+1}, x_{k+1} - x^* \rangle = \langle y_k, x_k - x^* \rangle \cdot \left(1 - \frac{\|y_k\|^4}{\langle A y_k, y_k \rangle \langle A^{-1} y_k, y_k \rangle}\right).$$

(4) *Conclude the proof.*