

Sheet 1

For the exercise class on the 01.10.2026 at 12:00.

Exercises to be solved in the exercise class

Exercise 1 (Convex Examples).

(0 Points)

Prove the following functions are convex

- (i) affine linear functions, i.e. $f(x) = a^T x + c$ for $a \in \mathbb{R}^d, c \in \mathbb{R}$,
- (ii) norms, i.e. $x \mapsto \|x\|$,
- (iii) sums of convex functions f_k , i.e. $f(x) = \sum_{k=1}^n f_k(x)$,
- (iv) $F(x) := \sup_{f \in \mathcal{F}} f(x)$ for a set of convex functions $\mathcal{F}$.

Exercise 2 (Finite Jensen).

(0 Points)

Let φ be convex, and $\sum_{i=1}^n \lambda_i = 1$ for $\lambda_i \geq 0$. Prove

$$\varphi\left(\sum_{i=1}^n \lambda_i x_i\right) \leq \sum_{i=1}^n \lambda_i \varphi(x_i)$$

and deduce $(\frac{1}{n} \sum_{i=1}^n x_i)^2 \leq \frac{1}{n} \sum_{i=1}^n x_i^2$.

Exercise 3 (Convexity and Minima).

(0 Points)

Prove the following statements

- (i) If f is convex, then every local minimum is also a global minimum.
- (ii) If f is strictly convex, then there exists at most one minimum.
- (iii) If f convex and differentiable and $\nabla f(x^*) = 0$, then x^* is a minimum.

Exercise 4 (Descent Directions of a Maximum).

(0 Points)

Let $x_* \in \mathbb{R}^d$ be a strict local maximum of $f : \mathbb{R}^d \rightarrow \mathbb{R}$. Prove that every $d \in \mathbb{R}^d \setminus \{0\}$ is a descent direction of f in x_* .

Exercise 5 (Convergence Speed).

(0 Points)

Proof that

- (i) if we have

$$\limsup_{k \rightarrow \infty} \frac{e(x_{k+1})}{e(x_k)} = 0,$$

then $e(x_k)$ converges super-linearly.

(ii) If for $c \in (0, 1)$ we have

$$\limsup_{k \rightarrow \infty} \frac{e(x_{k+1})}{e(x_k)} < c,$$

then $e(x_k)$ converges linearly with rate c .

(iii) If for $c \in (0, 1)$ we have

$$\limsup_{k \rightarrow \infty} \frac{e(x_{k+1})}{e(x_k)^2} < c,$$

then $e(x_k)$ converges super-linearly with rate c .

Exercise 6 (Momentum Matrix).

(0 Points)

let $D = \text{diag}(\lambda_1, \dots, \lambda_d)$, $\alpha, \beta > 0$ and define

$$T = \begin{pmatrix} (1 + \beta)\mathbb{I} - \alpha D & -\beta\mathbb{I} \\ \mathbb{I} & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{2d \times 2d}$$

Prove there exists a regular $S \in \mathbb{R}^{2d \times 2d}$ such that

$$S^{-1}TS = \hat{T} = \begin{pmatrix} T_1 & & \\ & \ddots & \\ & & T_d \end{pmatrix}$$

with

$$T_i = \begin{pmatrix} 1 + \beta - \alpha\lambda_i & -\beta \\ 1 & 0 \end{pmatrix} \in \mathbb{R}^{2 \times 2}.$$

Exercises to be prepared at home for the next exercise class at 15.10.2026

Exercise 7 (Bregman Divergence).

(0 Points)

The Bregman Divergence $D_f^{(B)}$ of a continuously differentiable function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is defined as the error of the linear approximation and is related to μ -strong convexity and Lipschitz continuous gradients as follows

$$\frac{\mu}{2} \|x - x_0\|^2 \stackrel{\substack{\mu\text{-strongly convex} \\ \text{(definition)}}}{\leq} \underbrace{f(x) - f(x_0) - \langle \nabla f(x_0), x - x_0 \rangle}_{=: D_f^{(B)}(x, x_0)} \stackrel{\substack{L\text{-Lipschitz gradient} \\ \text{(Descent Lemma)}}}{\leq} \frac{L}{2} \|x - x_0\|^2.$$

For $\mu = 0$ this is simply the convexity condition by Prop. A.1.8. So non-negativity of the Bregman divergence implies convexity. The L -Lipschitz gradients provide us with an upper bound on the Bregman divergence on the other hand which immediately results in an upper bound on f

$$f(x) = f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \underbrace{D_f^{(B)}(x, x_0)}_{\leq \frac{L}{2} \|x - x_0\|^2}. \quad (1)$$

(i) Prove for functions f with L -Lipschitz gradients, we have for all x_0

$$\min_x f(x) \leq f(x_0) - \frac{1}{2L} \|\nabla f(x_0)\|^2.$$

By minimizing the upper bound (1). What is the minimizer of the upper bound?

Hint. First minimize over the direction $x - x_0$ subject to the length $\|x - x_0\| = r$ being constant. Then minimize over r .

- (ii) Prove for convex functions f with L -Lipschitz gradients

$$D_f^{(B)}(x, x_0) \geq \frac{1}{2L} \|\nabla f(x) - \nabla f(x_0)\|^2.$$

Hint. Apply (i) to

$$\phi(x) := D_f^{(B)}(x, x_0).$$

Due to convexity you should already know the global minimum of ϕ .

- (iii) Prove for convex functions f with L -Lipschitz gradients, we have for all x, y

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{1}{L} \|\nabla f(x) - \nabla f(y)\|^2.$$

Hint. Use (ii) twice.

- (iv) Prove for convex f that the upper bound

$$D_f^{(B)}(x, x_0) \leq \frac{L}{2} \|x - x_0\|^2$$

is sufficient for Lipschitz continuity of the gradient ∇f .

Hint. Argue why you can use (iii) without circular reasoning.

- (v) In this last part we want to assume f is μ -strongly convex (and its gradient L -Lipschitz). Prove for all x, y

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{\mu L}{L + \mu} \|x - y\|^2 + \frac{1}{L + \mu} \|\nabla f(x) - \nabla f(y)\|^2.$$

Hint. Note that strong convexity provides an alternative lower bound to (ii). Using this alternative lower bound we could directly obtain a modified version of (iii). But we are greedy. We want to use both lower bounds for an even tighter bound on the scalar product. We therefore want to use

$$g_x(y) := f(y) - \frac{\mu}{2} \|x - y\|^2$$

to break down the Bregman divergence of f , so we can have our cake and eat it:

$$D_f^{(B)}(y, z) = \underbrace{D_{g_x}^{(B)}(y, z)}_{\text{use (ii)}} + \underbrace{\frac{\mu}{2} \|y - z\|^2}_{\text{strong convexity}}.$$

Remember to check convexity of g_x and Lipschitz continuity of ∇g_x before applying (ii). When applying (ii), the selection $z = x$ might be helpful.

Exercise 8 (Optimizing Quadratic Functions).

(0 Points)

In this exercise we consider functions of type

$$f(x) = x^T A x + b^T x + c,$$

where $x \in \mathbb{R}^d$, $A \in \mathbb{R}^{d \times d}$, $b \in \mathbb{R}^d$, $c \in \mathbb{R}$.

(i) Let $H := A^T + A$ be invertible. Prove that f can be written in the forms

$$f(x) = (x - x_*)^T A(x - x_*) + \tilde{c} \quad (2)$$

$$= \frac{1}{2}(x - x_*)^T \underbrace{(A^T + A)}_{=:H} (x - x_*) + \tilde{c} \quad (3)$$

for some $x_* \in \mathbb{R}^d$ and $\tilde{c} \in \mathbb{R}$. Argue that H is always symmetric. Under which circumstances is x_* a minimum?

(ii) Argue that the Newton Method (with step size $\alpha_n = 1$) applied to f would jump to x_* in one step and then stop moving.

(iii) Let $V = (v_1, \dots, v_d)$ be an orthonormal basis such that

$$H = V \text{diag}[\lambda_1, \dots, \lambda_d] V^T$$

with $0 < \lambda_1 \leq \dots \leq \lambda_d$ and write

$$y^{(i)} := \langle y, v_i \rangle.$$

Express $(x_n - x_*)^{(i)}$ in terms of $(x_0 - x_*)^{(i)}$, where x_n is given by the gradient descent recursion

$$x_{n+1} = x_n - h \nabla f(x_n).$$

For which step size h do all the components $(x_n - x_*)^{(i)}$ converge to zero? Which component has the slowest convergence speed? Find the optimal learning rate h^* and deduce for this learning rate

$$\|x_n - x_*\| \leq (1 - \frac{2}{1+\kappa})^n \|x_0 - x_*\|.$$

with the condition number $\kappa = \frac{\lambda_d}{\lambda_1}$.

Exercise 9 (Lasso).

(0 Points)

Let

$$f(x) = \frac{1}{2} \|x - y\|^2 + \lambda \|x\|_1$$

for $x \in \mathbb{R}^d$ be the Lagrangian form of the least squares LASSO method.

(i) Compute a sub-gradient of f .

(ii) Prove that f is convex.

(iii) Find a global minimum of f .

Exercise 10 (Weak PL-Inequality).

(0 Points)

Assume $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is L -smooth and satisfies the “weak PL inequality”

$$\|\nabla f(x)\| \geq 2c(f(x) - f_*)$$

for some $c > 0$ and all $x \in \mathbb{R}^d$ with $f_* = \min_x f(x) > -\infty$.

- (i) Let $a_0 \in [0, \frac{1}{q}]$ for some $q > 0$ and assume for the sequence $(a_n)_{n \in \mathbb{N}}$ that it is positive and satisfies a diminishing contraction

$$0 \leq a_{n+1} \leq (1 - qa_n)a_n \quad \forall n \geq 0.$$

Prove the convergence rate

$$a_n \leq \frac{1}{nq + 1/a_0} \leq \frac{1}{(n+1)q}.$$

Hint. A useful checkpoint might be the telescoping sum of

$$\frac{1}{a_{n+1}} - \frac{1}{a_n} \geq q.$$

- (ii) Prove that f is bounded. More specifically $e(x) := f(x) - f_* \leq \frac{L}{2c^2}$ for all x .

Hint. Use Exercise 1 (i).

- (iii) For gradient descent $x_{n+1} - x_n = -\alpha_n \nabla f(x_n)$ with constant step size $\alpha_k = \frac{1}{L}$ prove the convergence rate

$$f(x_n) - f_* \leq \frac{L}{2c^2(n+1)}.$$